



Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação
DEPARTAMENTO DE TELEMÁTICA

Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA

Autora: Luzinete Cristina Bonani de Faria

Orientador: Prof. Dr. Reginaldo Palazzo Júnior

Banca Examinadora:

| | |
|--|-----------------|
| Prof. Dr. Reginaldo Palazzo Júnior | FEEC/UNICAMP |
| Prof. Dr. Geraldo Pompeu Júnior | UFSCAR/Sorocaba |
| Prof. Dr. Valdemar Cardoso da Rocha Júnior | DES/UFPE |
| Prof. Dr. Joaquim Aparecido Machado | IGC/UNICAMP |
| Prof. Dr. Romis Ribeiro Faissol Attux | FEEC/UNICAMP |

Tese de Doutorado apresentada à Faculdade de Engenharia Elétrica e de Computação da Universidade Estadual de Campinas, como parte dos requisitos exigidos para a obtenção do título de Doutor em Engenharia Elétrica. Área de concentração: Telecomunicações e Telemática.

Campinas - SP
JULHO de 2011

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

| | |
|-------|---|
| F225e | <p>Faria, Luzinete Cristina Bonani de Existências de códigos corretores de erros e protocolos de comunicação em sequências de DNA / Luzinete Cristina Bonani de Faria. --Campinas, SP: [s.n.], 2011.</p> <p>Orientador: Reginaldo Palazzo Júnior. Tese de Doutorado - Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computação.</p> <p>1. Códigos de controle de erros. 2. Aneis (Álgebra). 3. Genoma. 4. Analise se sequência de DNA. 5. Redes de computadores. I. Palazzo Júnior, Reginaldo. II. Universidade Estadual de Campinas. Faculdade de Engenharia Elétrica e de Computação. III. Título.</p> |
|-------|---|

Título em Inglês: Existence of error-correcting codes and communication protocols in
DNA sequences

Palavras-chave em Inglês: Error control codes, Rings (Algebra), Genome, Sequence
analysis DNA, Computer networks

Área de concentração: Telecomunicações e Telemática

Titulação: Doutor em Engenharia Elétrica

Banca examinadora: Geraldo Pompeu Júnior, Valdemar Cardoso da Rocha Júnior,
Joaquim Aparecido Machado, Romis Ribeiro Faissol Attux

Data da defesa: 08/07/2011

Programa de Pós Graduação: Engenharia Elétrica

COMISSÃO JULGADORA - TESE DE DOUTORADO

Candidata: Luzinete Cristina Bonani de Faria

Data da Defesa: 8 de julho de 2011

Título da Tese: "Existências de Códigos Corretores de Erros e Protocolos de Comunicação em Sequências de DNA"

Prof. Dr. Reginaldo Palazzo Júnior (Presidente): Reginaldo Palazzo Júnior

Prof. Dr. Geraldo Pompeu Júnior: Geraldo Pompeu Júnior

Prof. Dr. Valdemar Cardoso da Rocha Junior: Valdemar C. da Rocha Jr.

Prof. Dr. Joaquim Aparecido Machado: Joaquim Aparecido Machado

Prof. Dr. Romis Ribeiro de Faissol Attux: Romis Ribeiro de Faissol Attux

A todos os meus familiares;

Em especial:

*Ao meu filho, Gustavo, e ao meu marido,
Misael, pela paciência, apoio, estímulo
e amor.*

Dedico

Agradecimentos ¹

Ao Prof. Dr. Reginaldo Palazzo Jr., pela sua dedicação, paciência e compreensão somadas a incomensurável capacidade de orientar. Em especial, à sua confiança depositada em mim e às inúmeras e agradáveis conversas que tivemos. Para sempre ficará na minha memória com muito carinho.

Aos professores membros da banca examinadora pela disponibilidade e atenção dispensada ao trabalho, bem como por suas valiosas sugestões.

Ao Prof. Dr. Geraldo Pompeu Jr., pelo compromisso e dedicação com o ensino nas Universidades e, pela capacidade de despertar nos alunos o interesse pelo ensino da matemática. Serei eternamente grata pelo meu encaminhamento ao mundo acadêmico.

Ao Prof. Dr. Márcio de C. Silva-Filho e ao Prof. Dr. Marcelo M. Brandão, pelas nossas várias conversas sobre o mundo biológico e pela oportunidade de parceria.

Ao Prof. Dr. Michel E. B. Yamagishi e ao Prof. Dr. Roberto H. Herai, pela colaboração na reprodução do gene e do genoma.

Ao Prof. Dr. Carlos Eduardo Câmara pelas boas conversas sobre biologia e comunicação.

Ao meu marido, Misael, que sempre esteve do meu lado com muita paciência, dedicação e amor. O seu companheirismo foi fundamental na concretização deste trabalho. Muito obrigada por fazer parte da minha vida.

Ao meu amado filho, Gustavo - meu maior presente. Ao seu lado, a cada dia aprendo uma nova lição, com muito amor, carinho e dedicação.

Aos meus pais Adhemar e Anna: a minha eterna gratidão - Jamais se esqueçam que eu levarei para sempre um pedaço do ser de cada um dentro do meu ser.

Aos meus irmãos, Ana Cláudia, Lucélia e Samuel, meu enorme carinho.

Às minhas tias, Marlete, Ivani e Alcione, que com ensinamentos de amor, carinho, afeto, lealdade, honestidade e sinceridade me ajudaram a crescer. Meus grandes exemplos de vida.

À minha sogra, dona Nair, muito amável, sábia e prudente. Muito obrigada pelo carinho e confiança e por me presentear com um dos seus filhos, Misael.

A todos os meus familiares, que souberam compreender minhas ausências e sempre me deram a força necessária para seguir em frente.

À minha grande amiga Andréa, minha irmã de coração e parceira de trabalho. A parceria na pesquisa e a realização deste trabalho só foram possíveis em virtude da nossa amizade sólida e sincera. Muito obrigada por todos os momentos que vivenciamos entre estudos,

¹Este trabalho foi financiado pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico - CNPq e, pela Coordenação do Aperfeiçoamento do Pessoal de Nível Superior - CAPES

conversas e alegrias. Muito obrigada pela amizade que também contagiou os nossos maridos e filhos, pela qual nos tornamos uma “grande família”. Deixo com você o meu eterno carinho, respeito e admiração.

Às minhas amigas Juliana e Évelin, que entenderam as minhas ausências em momentos importantes, e, me ajudaram nas dificuldades. Muito obrigada pela amizade e pelo carinho depositado.

Ao amigo João Henrique, pela sua amizade e importante colaboração na parte computacional deste trabalho.

Aos amigos Clarice e Leandro, sempre dedicados e cuidadosos. Muito obrigada por todos os momentos que me deram forças e me incentivaram para que eu não desanimasse.

Aos amigos Cátia e Igor, sempre alegres e determinados. Muito obrigada pelo carinho e importantes momentos, os quais compartilhamos.

Aos amigos do LTIA, em especial, Mércio, Daniel Cunha, Aido, Rodrigo, Giuliano, Júlio, Wanessa, Lucila, Anderson, Cíntia e Maicon pela amizade e pelos ótimos momentos em que passamos durante estes anos.

Aos funcionários da Faculdade de Engenharia Elétrica e de Computação, que de alguma forma contribuíram para a realização deste trabalho.

Resumo

Um dos grandes desafios da comunidade científica em teorias da informação genética, comunicação genética e codificação genética é verificar a existência de uma estrutura matemática relacionada com a estrutura do DNA. Este trabalho propõe modelos para o sistema de comunicação de informação genética e genômica análogos ao modelo de um sistema de comunicação digital. Ambos os modelos são capazes de identificar, reproduzir e classificar matematicamente diferentes sequências de DNA. O primeiro identifica e reproduz a sequência de nucleotídeos de uma fita simples do DNA através da codificação genética e, o segundo identifica e reproduz a sequência das bases complementares da fita dupla do DNA através da codificação genômica. Os objetivos principais do presente trabalho são: a) caracterização matemática dos modelos de codificação genética e codificação genômica, b) proposta de um algoritmo para identificação de sequências de DNA; c) mostrar que sequências de DNA com características biológicas distintas, incluindo proteínas, gene e genoma em termos das fitas simples do DNA e da dupla hélice do DNA são identificadas como palavras-código dos códigos G -linearidade (BCH sobre corpos e BCH sobre anéis), reproduzidas e classificadas matematicamente, d) representação algébrica via polinômios primitivos/geradores e seus polinômios recíprocos das fitas simples do DNA (fita codante e fita não codante) e da dupla hélice do DNA, e) mostrar a existência de códigos concatenados (*nested codes*) entre algumas sequências de direcionamento e suas respectivas proteínas organelares, f) mostrar a arquitetura biológica (*Biological frame*) do genoma do plasmídeo *Lactococcus lactis*. Os resultados apresentados neste trabalho contribuem para o desenvolvimento de uma metodologia que poderá ser aplicada em análises mutacionais e de polimorfismos, produção de novos fármacos, melhoramento genético, entre outros, reduzindo tempo e custos laboratoriais.

Palavras-chave: Codificação genética, codificação genômica, códigos corretores de erros, código BCH sobre corpos e anéis, códigos concatenados, genoma, gene, proteínas, arquitetura biológica.

Abstract

One of the great challenges of the scientific community on theories of genetic information, genetic communication and genetic coding is to determine a mathematical structure related to the structure of DNA. This thesis proposes a model of a communication system for genetic and genomic information similar to the model of a digital communication system. Both models are able to identify, reproduce and classify mathematically different sequences of DNA. The first model identifies and reproduces the nucleotide sequence of a single DNA strand via the genetic encoding. The latter identifies and reproduces the sequence of the complementary bases of the double DNA strand through genomic encoding. The aims of this work are: a) a mathematical characterization of the models of genetic coding and genomic coding, b) the proposal of an algorithm for the identification of DNA sequences, c) to show that DNA sequences with distinct biological characteristics, including protein, gene and genome in terms of the single strands of DNA and of the double helix of DNA are identified as codewords of the *G*-linearity codes (BCH codes over fields and BCH codes over rings), mathematically reproduced and classified, d) algebraic representation via primitive/generator polynomials and their reciprocal polynomials of single DNA strands (coding strand and non-coding strand) and the double DNA strand, e) to show the existence of concatenated codes (*nested codes*) between some targeting sequences and their corresponding organelles proteins, f) to show the biological architecture (*Biological frame*) of the plasmid *Lactococcus lactis* genome. The results presented in this work contribute to the development of a methodology that can be applied to mutational analysis and polymorphisms, production of new drugs, genetic improvement, among others, reducing time and laboratory costs.

Key-words: Genetic coding, genomic coding, error correction code, *G*-linearity, BCH codes over fields and rings, *nested code*, genome, gene, proteins, biological architecture.

Conteúdo

| | |
|--|-------|
| Dedicatória | v |
| Agradecimentos | vii |
| Resumo | ix |
| Abstract | xi |
| Lista de Figuras | xvii |
| Lista de Tabelas | xxiii |
| 1 Introdução | 1 |
| 1.1 Breve Histórico | 3 |
| 1.1.1 Shannon e a teoria da informação | 3 |
| 1.1.2 Os avanços da biologia celular e molecular | 5 |
| 1.1.3 Modelos propostos na literatura | 7 |
| 1.2 Proposta de Pesquisa | 10 |
| 1.2.1 Proposta de Trabalho | 13 |
| 1.3 Descrição do Trabalho | 17 |
| 2 Preliminares sobre o DNA e os Códigos Corretores de Erros | 21 |
| 2.1 Revisão Biológica | 21 |
| 2.1.1 A vida começa nas células | 22 |
| 2.1.2 Estrutura e função das proteínas | 31 |
| 2.1.3 O DNA e a síntese de proteínas | 37 |
| 2.1.4 Estrutura genômica | 42 |
| 2.1.5 Mutações | 48 |
| 2.2 Estruturas Algébricas | 51 |
| 2.2.1 Anéis | 52 |

| | | |
|----------|---|------------|
| 2.2.2 | Corpos | 52 |
| 2.3 | Códigos Corretores de Erros | 54 |
| 2.3.1 | Códigos de bloco | 57 |
| 2.3.2 | Códigos lineares | 59 |
| 2.3.3 | Códigos cíclicos sobre \mathbb{Z}_q | 62 |
| 2.3.4 | Códigos BCH sobre anéis e corpos | 64 |
| 2.3.5 | Códigos geometricamente uniformes | 68 |
| 2.3.6 | G -linearidade | 71 |
| 3 | Modelo de Codificação Genética e Genômica | 75 |
| 3.1 | Analogia entre o Sistema de Comunicação Digital e o Sistema de Informação Genômica | 76 |
| 3.1.1 | Sistema de comunicação digital | 76 |
| 3.1.2 | Sistema de informação genômica | 79 |
| 3.1.3 | Analogias | 82 |
| 3.2 | Modelos de Codificação Genética e Genômica | 84 |
| 3.2.1 | Modelo de codificação genética | 87 |
| 3.2.2 | Modelo de codificação genômica | 105 |
| 3.3 | Similaridades entre o Sistema Biológico e Redes Locais de Computadores | 108 |
| 3.3.1 | Proposta da arquitetura biológica do genoma humano | 122 |
| 3.3.2 | Arquitetura biológica do genoma do plasmídeo <i>Lactococcus lactis</i> pcl 21 | 127 |
| 4 | Algoritmo para a Identificação de Sequências de DNA | 131 |
| 4.1 | Algoritmo de Geração de Códigos BCH sobre $GF(4^r)$ | 133 |
| 4.1.1 | Código BCH $(n, k, d_H) = (63, k, d_H)$ sobre $GF(4^r)$ | 134 |
| 4.2 | Algoritmo de Geração de Códigos BCH sobre $GR(4, r)$ | 146 |
| 4.2.1 | Códigos BCH primitivos sobre $GR(4, r)$: Exemplos | 147 |
| 4.2.2 | Códigos BCH não primitivos sobre $GR(4, r)$: Exemplos | 183 |
| 5 | Análise dos Resultados | 199 |
| 5.1 | Sequências de DNA Reproduzidas pelos Códigos G -linearidade | 200 |
| 5.1.1 | Alto fluxo de informação x baixa redundância | 202 |
| 5.2 | Análise das Sequências de DNA Reproduzidas sobre Corpos | 203 |
| 5.2.1 | Relação entre polinômios primitivos e polinômios geradores do $GF(4^r)$ e do $GR(4, r)$ | 205 |
| 5.2.2 | Rotulamento das sequências de DNA sobre $GF(4)$ | 206 |
| 5.2.3 | Representação algébrica das sequências de DNA no $GF(4^r)$ | 207 |

| | | |
|-----------------------------------|---|------------|
| 5.3 | Análise das Sequências de DNA Reproduzidas sobre Anéis | 208 |
| 5.3.1 | Relação entre os polinômios primitivos e os polinômios geradores sobre $GR(4, r)$ | 213 |
| 5.3.2 | Rotulamento das sequências de DNA sobre $GR(4, r)$ | 217 |
| 5.3.3 | Dependência entre os códigos corretores de erros e os polinômios primitivos | 220 |
| 5.3.4 | A relação entre o BCH primitivo sobre $GR(4, r)$ e o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$ | 221 |
| 5.3.5 | Representação algébrica das sequências sobre $GR(4, r)$ | 225 |
| 5.4 | Reprodução de Sequências de DNA envolvidas em Patologias Clínicas | 228 |
| 5.4.1 | Sequência de direcionamento do vírus <i>Influenza A</i> | 228 |
| 5.4.2 | Proteínas p53 e BCRA1 - Câncer | 229 |
| 5.5 | Reprodução de um Genoma Circular e de um Gene Humano | 234 |
| 5.6 | Sistemas Concatenados | 241 |
| 5.6.1 | Sequências de DNA reproduzidas pelos códigos concatenados | 242 |
| 5.7 | Relação entre o Corpo $GF(4^r)$ e o Anel $GR(4, r)$ | 253 |
| 5.7.1 | Grau das extensões de Galois r e os comprimentos das sequências de DNA. | 257 |
| 6 | Conclusões e Perspectivas Futuras | 259 |
| 6.1 | Desenvolvimento do Trabalho | 260 |
| 6.2 | Contribuições do Trabalho | 262 |
| 6.3 | Sugestões para Trabalhos Futuros | 263 |
| 6.4 | Comentário Final | 265 |
| A | Códigos BCH $C(n, k, d_H)$ sobre $GF(4^r)$ e sobre $GR(4, r)$ | 267 |
| B | Sequências de DNA Reproduzidas através dos Códigos G-linearidade | 271 |
| B.1 | BCH Primitivo sobre $GF(4)$ | 271 |
| B.2 | BCH Primitivo sobre $GR(4, r)$ | 274 |
| B.3 | BCH não Primitivo sobre $GR(4, r)$ | 286 |
| Referências Bibliográficas | | 291 |

Listas de Figuras

| | | |
|------|--|----|
| 1.1 | Série de nucleotídeos representando a estrutura primária de uma molécula de DNA em termos da dupla hélice e da fita simples do DNA. | 2 |
| 1.2 | Biologia molecular. | 7 |
| 1.3 | Modelo de um sistema de comunicação de informação genética. | 11 |
| 1.4 | Modelo proposto para decodificação biológica. | 13 |
| 1.5 | Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genônica. | 16 |
| 2.1 | As células procarióticas apresentam uma organização interna mais simples do que as células eucarióticas. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 23 |
| 2.2 | Trifosfato de adenosina (ATP). Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 24 |
| 2.3 | Visão geral dos blocos químicos que constituem as células e as macroestruturas que as formam. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 25 |
| 2.4 | Durante o crescimento as células avançam continuamente pelos quatro estágios do ciclo celular. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 30 |
| 2.5 | Visão geral da estrutura e função das proteínas. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 32 |
| 2.6 | Estrutura primária. | 32 |
| 2.7 | Hélice α | 33 |
| 2.8 | Folha β | 33 |
| 2.9 | Estrutura terciária. | 34 |
| 2.10 | Estrutura quaternária. | 34 |
| 2.11 | Dobramento de proteínas mediado por chaperonas e chaperoninas. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 37 |
| 2.12 | Visão geral dos quatro processos básicos da genética molecular. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 38 |
| 2.13 | Estrutura geral dos nucleotídeos. | 39 |
| 2.14 | Estrutura química das principais bases dos ácidos nucléicos. | 39 |

| | | |
|------|--|----|
| 2.15 | Estrutura tridimensional do DNA e sua complementaridade. | 40 |
| 2.16 | A informação codificada no DNA é convertida nas sequências de aminoácidos. | 41 |
| 2.17 | Visão geral da estrutura dos genes e dos cromossomos. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 44 |
| 2.18 | Modelo de compactação dos cromossomos eucarióticos. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 45 |
| 2.19 | Visualização dos cromossomos humanos. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 46 |
| 2.20 | Aspecto microscópio de um cromossomo característico em metáfase. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 47 |
| 2.21 | Códigos de árvore sob o ponto de vista algébrico. | 56 |
| 2.22 | Conjunto de sinais 4-PSK e 8-PSK. | 71 |
| 2.23 | Rotulamento do espaço de Hamming bidimensional de ordem 2. | 74 |
| 3.1 | Diagrama de blocos de um sistema de comunicação. | 78 |
| 3.2 | Dogma central da teoria de comunicação. | 82 |
| 3.3 | Dogma central da biologia molecular. | 82 |
| 3.4 | Analogia entre o sistema de informação genômica e o sistema de comunicação simplificado. | 83 |
| 3.5 | Modelo de um sistema de comunicação digital. | 83 |
| 3.6 | Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética. | 85 |
| 3.7 | Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genônica. | 86 |
| 3.8 | Modelo de um sistema de comunicação de informação genética. | 87 |
| 3.9 | O alfabeto do código genético e os alfabetos 4-ários dos CCEs. | 89 |
| 3.10 | Associações entre $N \rightarrow GF(4)$ e $N \rightarrow \mathbb{Z}_4$. | 90 |
| 3.11 | Representação da estrutura de espaço vetorial associado ao $GF(4)$. | 90 |
| 3.12 | Representação de espaço vetorial associado ao \mathbb{Z}_4 . | 91 |
| 3.13 | Codificador genético. | 92 |
| 3.14 | Possibilidades de rotulamento entre os elementos dos conjuntos N , $GF(4)$ e \mathbb{Z}_4 . | 92 |
| 3.15 | Rotulamento S_4 . | 94 |
| 3.16 | Elementos do alfabeto 4-ário do \mathbb{Z}_4 e sua representação binária. | 94 |
| 3.17 | Rotulamentos A, B e C. | 95 |
| 3.18 | Casamento entre o contexto biológico e o contexto matemático. | 95 |
| 3.19 | Mapeamentos do alfabeto \mathbb{Z}_4 . | 96 |
| 3.20 | Códigos resultantes no codificador genético. | 97 |

| | | |
|------|---|-----|
| 3.21 | Codificador do código G -linearidade. | 101 |
| 3.22 | Interpretação do processo de identificação de sequências de DNA. | 102 |
| 3.23 | Mutações relacionadas à resistência a inseticidas, [23]. | 103 |
| 3.24 | Modelo de um sistema de comunicação de informação genômica. | 105 |
| 3.25 | O alfabeto do código genômico e os alfabetos 4-ários dos CCEs. | 106 |
| 3.26 | Associações entre $N' \rightarrow GF(4)$ e $N' \rightarrow \mathbb{Z}_4$. | 107 |
| 3.27 | Codificador genômico. | 107 |
| 3.28 | Possibilidades de rotulamento entre os elementos dos conjuntos N' e $GF(4)$ e N' e \mathbb{Z}_4 . | 108 |
| 3.29 | Representação da distribuição cromossônica dentro do núcleo celular. Mirela Andronesco/ <i>Divulgação</i> , [54]. | 109 |
| 3.30 | Comparação do número e dos tipos de proteínas codificadas no genoma de diferentes eucariotos. Lodish <i>et al.</i> , <i>Molecular Cell Biology</i> , 5th Edition. | 111 |
| 3.31 | Genoma humano distribuído em 23 pares de cromossomos. | 112 |
| 3.32 | Modelo de sete camadas. | 117 |
| 3.33 | Procedimento da transmissão de dados. | 118 |
| 3.34 | Modelos de topologias de redes de computadores. | 119 |
| 3.35 | Quadro do formato de enlace da <i>ethernet</i> . | 120 |
| 3.36 | Representação planar do armazenamento de todo o conteúdo do genoma humano masculino em um CD, e a formação do quadro biológico. | 122 |
| 3.37 | Representação planar do armazenamento do conteúdo de um único cromosomo em um único CD. | 123 |
| 3.38 | Representação planar do armazenamento do conjunto de cromossomos de uma célula em um único CD. | 124 |
| 3.39 | Proposta de armazenamento e transmissão da informação genômica análoga ao processo de transmissão de dados, porém no sentido contrário. | 126 |
| 3.40 | Sequência em nucleotídeos do gene “Trav7”. | 126 |
| 3.41 | Ilustração de uma bactéria com plasmídeos no seu interior. (a) DNA cromossômico (b) Plasmídeos. | 127 |
| 3.42 | Sequência genômica em nucleotídeos. | 128 |
| 3.43 | Genoma Plasmídeo com 2047 nucleotídeos de comprimento. | 129 |
| 4.1 | Sequência de direcionamento com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$. | 145 |
| 4.2 | Sequência de DNA de sinal interno com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$. | 158 |
| 4.3 | Sequência de DNA de sinal interno com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$. | 159 |
| 4.4 | Sequência de DNA de uma proteína com 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$. | 167 |

| | | |
|------|--|-----|
| 4.5 | Sequência de DNA de uma proteína com 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ nos rotulamentos A e rotulamento B e C. | 169 |
| 4.6 | Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 175 |
| 4.7 | Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento A. . | 178 |
| 4.8 | Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento B. . | 180 |
| 4.9 | Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C. . | 182 |
| 4.10 | Sequência de DNA de <i>miRNA</i> com 21 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C. | 188 |
| 4.11 | Sequência de DNA de sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C. | 193 |
| 4.12 | Três possibilidades da sequência de direcionamento com 93 nucleotídeos reproduzida com $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C. | 198 |
| 5.1 | Rotulamentos S_4 e S'_4 para o $GF(4)$ | 206 |
| 5.2 | Representação algébrica via polinômios da dupla hélice e das fitas simples de DNA da sequência Seq.13 no $GF(4)$ | 208 |
| 5.3 | Rotulamentos A, B e C resultantes das 24 permutações entre $N \rightarrow \mathbb{Z}_4$ | 217 |
| 5.4 | Rotulamentos A' , B' e C' resultante das 24 permutações entre $N' \rightarrow \mathbb{Z}_4$ | 218 |
| 5.5 | Representação algébrica da dupla hélice e das fitas simples de DNA da sequência Seq.13 no $GR(4, 6)$ reproduzida com $D(\mathbf{a}, \mathbf{b}) = 1$ | 226 |
| 5.6 | Representação algébrica da dupla hélice e das fitas simples de DNA da sequência Seq.13 no $GR(4, 6)$ reproduzida com $D(\mathbf{a}, \mathbf{b}) = 2$ | 227 |
| 5.7 | Sequência de DNA de sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C. | 229 |
| 5.8 | Proteína envolvida no câncer de ovário. | 231 |
| 5.9 | Sequência de DNA da proteína p53 com $D(\mathbf{a}, \mathbf{b}) = 1$ | 233 |
| 5.10 | Reprodução do genoma plamidial e do gene “Trav7” e através dos códigos G -linearidade. | 235 |
| 5.11 | Representação algébrica do genoma do plasmídeo. | 236 |
| 5.12 | Representação algébrica do gene “Trav 7”. | 237 |
| 5.13 | Sequência do gene “Trav 7” reproduzida através do código G -linearidade. . . | 238 |
| 5.14 | Sequência do genoma do plasmídeo <i>Lactococcus lactis</i> pcl 2.1 reproduzida através do código G -linearidade. | 240 |
| 5.15 | As quatro camadas de um sistema <i>nested</i> em [14]. | 241 |
| 5.16 | Sequência de Direcionamento reproduzida em $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ | 244 |
| 5.17 | Proteína reproduzida em $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ | 247 |

| | | |
|------|--|-----|
| 5.18 | Sequência de direcionamento reproduzida em $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ | 248 |
| 5.19 | Proteína reproduzida em $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ | 250 |
| 5.20 | Sequência de DNA n^o 24 gerada pelo código BCH (63, 57, 3) | 251 |
| B.1 | Sequências de direcionamento (SD) com 63 nucleotídeos e $D(a, b) = 1$ | 271 |
| B.2 | Sequências de Direcionamento (SD), sinal interno (SI) e íntron (I) com 63 nucleotídeos e $D(a, b) = 1$ | 272 |
| B.3 | DNA repetitivo (DNARep) e proteína (P) com 255 nucleotídeos e $D(a, b) = 1$ | 273 |
| B.4 | Sinal interno (SI) e Íntron (I) com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 274 |
| B.5 | DNA repetitivo (DNARep) e Proteína (P) com 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 275 |
| B.6 | Proteína codificada por 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 276 |
| B.7 | Proteína codificada por 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 277 |
| B.8 | Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 279 |
| B.9 | Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 281 |
| B.10 | Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 283 |
| B.11 | Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$ | 285 |
| B.12 | MicroRNA com 21 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 286 |
| B.13 | Sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 286 |
| B.14 | Sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 287 |
| B.15 | Sequências de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 288 |
| B.16 | Sequências de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 289 |
| B.17 | Sequências de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ | 290 |

Lista de Tabelas

| | |
|--|-----|
| 1.1 Sequências que serão analisadas no presente trabalho. | 15 |
| 2.1 Arranjo padrão. | 61 |
| 2.2 Definição das funções β e γ . | 73 |
| 2.3 Mapeamento por rotulamento isométrico. | 74 |
| 3.1 Código genético. | 81 |
| 4.1 Adição em $GF(4)$. | 134 |
| 4.2 Multiplicação em $GF(4)$. | 134 |
| 4.3 Comprimento das sequências de DNA e o grau das extensões de Galois sobre $GF(4)$. | 135 |
| 4.4 Polinômios primitivos da extensão de Galois de grau $r = 3$. | 136 |
| 4.5 Quantidade de polinômios primitivos em cada extensão. | 136 |
| 4.6 Elementos de $GF(64)$. | 137 |
| 4.7 Raízes dos polinômios minimais sobre $GF(64)$. | 138 |
| 4.8 Polinômios minimais que são iguais sobre $GF(64)$. | 139 |
| 4.9 Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(63, k, d_H)$ BCH primitivos sobre $GF(64)$. | 139 |
| 4.10 24 permutações para $GF(4)$. | 142 |
| 4.11 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 63$ nucleotídeos sobre $GF(4)$. | 145 |
| 4.12 Adição módulo 4. | 148 |
| 4.13 Multiplicação módulo 4. | 148 |
| 4.14 Polinômios primitivos da extensão de Galois $r = 6$. | 148 |
| 4.15 Elementos de F_{64} em notação de r -uplas com $p_1(x)$. | 149 |
| 4.16 Elementos do grupo cíclico do grupo $GR^*(4, 6)$ em notação de r -uplas com $p_1(x)$. | 150 |
| 4.17 Elementos de G_{63} | 151 |
| 4.18 Raízes dos polinômios minimais sobre G_{63} . | 152 |
| 4.19 Polinômios minimais que são iguais sobre G_{63} . | 152 |

| | |
|---|-----|
| 4.20 Relação entre as linhas da matriz P e as 24 permutações. | 154 |
| 4.21 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 63$ nucleotídeos. | 157 |
| 4.22 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 63$ nucleotídeos. | 159 |
| 4.23 Polinômios primitivos da extensão de Galois $r = 8$ | 160 |
| 4.24 Elementos de F_{256} com $p_{01}(x)$ | 161 |
| 4.25 Elementos do grupo cíclico do grupo $GR^*(4, 8)$ com $p_{01}(x)$ | 162 |
| 4.26 Elementos de G_{255} | 162 |
| 4.27 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 255$ nucleotídeos. | 166 |
| 4.28 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 255$ nucleotídeos. | 168 |
| 4.29 Elementos de F_{1024} com $p_{42}(x)$ | 171 |
| 4.30 Elementos do grupo cíclico do grupo $GR^*(4, 10)$ com $p_{42}(x)$ | 171 |
| 4.31 Elementos de G_{1023} | 172 |
| 4.32 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 1023$ nucleotídeos. | 176 |
| 4.33 Elementos de G_{21} | 184 |
| 4.34 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 21$ nucleotídeos. | 187 |
| 4.35 Elementos de F_{256} com $p_{05}(x)$ | 189 |
| 4.36 Elementos do grupo cíclico do grupo $GR^*(4, 8)$ com $p_{05}(x)$ | 190 |
| 4.37 Elementos de G_{51} | 190 |
| 4.38 As 24 permutações e a quantidade de palavras-código para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 51$ | 192 |
| 4.39 Elementos de F_{1024} com $p_{01}(x)$ | 194 |
| 4.40 Elementos do grupo cíclico do grupo $GR^*(4, 10)$ com $p_{01}(x)$ | 195 |
| 4.41 Elementos de G_{93} | 195 |
| 4.42 As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 93$ nucleotídeos. | 197 |
| | |
| 5.1 Sequências de DNA reproduzidas pelos códigos G -linearidade. | 202 |
| 5.2 Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GF(4^r)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença. . . | 204 |
| 5.3 Relação dos $p(x)$ e dos $g(x)$ em $GF(4^3)$ e $GR(4, 6)$ para o código $(63, 57, 3)$. . | 205 |
| 5.4 Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GR(4, 6)$ e $GR(4, 8)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença. | 209 |
| 5.5 Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença. . | 210 |
| 5.6 Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH não primitivo sobre anel com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença. . | 211 |
| 5.7 Continuação da relação das sequências de DNA reproduzidas pelo código BCH não primitivo sobre anel com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença. | 212 |

| | | |
|------|--|-----|
| 5.8 | Relação entre $p(x)$ e $g(x)$ no código BCH primitivo sobre $GR(4, r)$ | 214 |
| 5.9 | Classe de polinômios primitivos dos códigos BCH não primitivo - (21, 15, 3). | 215 |
| 5.10 | Classe de polinômios primitivos dos códigos BCH não primitivo - (51, 43, 3). | 215 |
| 5.11 | Classe de polinômios primitivos dos códigos BCH não primitivo - (93, 83, 3). | 216 |
| 5.12 | Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros. | 223 |
| 5.13 | Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 8)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros. | 224 |
| 5.14 | Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros. | 225 |
| 5.15 | Interpretando o código concatenado - <i>B.napus</i> | 245 |
| 5.16 | Interpretando o código concatenado - <i>A.thaliana</i> | 248 |
| 5.17 | Interpretando o código concatenado - <i>A.thaliana</i> | 252 |
| 5.18 | Relação entre os códigos BCH sobre $GF(4^r)$ e sobre $GR(4, r)$ | 254 |
| 5.19 | Relação entre os polinômios geradores do corpo com os polinômios primitivos do anel. | 255 |
| 5.20 | Quantidade das sequências de DNA analisadas em anéis e corpos com os padrões de erro $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ | 256 |
| 5.21 | Grau das extensões de Galois r e os comprimentos das sequências de DNA. | 257 |
| A.1 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (63, k , d_H) BCH primitivos sobre $GF(64)$ - extensão de Galois $r = 3$ | 267 |
| A.2 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos BCH primitivos em $GF(256)$ - extensão de Galois $r = 4$ | 268 |
| A.3 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (63, k , d_H) BCH primitivos sobre $GR^*(4, 6)$ - extensão de Galois $r = 6$ | 268 |
| A.4 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (255, k , d_H) BCH primitivos em $GR^*(4, 8)$ - extensão de Galois $r = 8$ | 269 |
| A.5 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (21, k , d_H) BCH não primitivos sobre $GR^*(4, 6)$ - extensão de Galois $r = 6$ | 269 |
| A.6 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (51, k , d_H) BCH não primitivos sobre $GR^*(4, 8)$ - extensão de Galois $r = 8$ | 269 |
| A.7 | Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos (93, k , d_H) BCH não primitivos sobre $GR^*(4, 10)$ - extensão de Galois $r = 10$ | 270 |
| A.8 | Distâncias de Haming de códigos BCH não primitivos relativos aos compri- mentos n | 270 |

Capítulo 1

Introdução

Em teoria de comunicação, os códigos corretores de erros são utilizados sempre que se deseja transmitir ou armazenar informação. O sistema biológico também armazena e transmite a informação através do código genético. Neste capítulo, evidenciaremos os benefícios da associação dessa teoria com a ciência da vida, os quais são enormes e bastante promissores no mundo acadêmico, industrial e comercial.

Inicialmente, na Seção 1.1, apresentamos um breve histórico dos avanços na teoria de comunicação e na biologia celular e molecular. Alguns dos principais modelos propostos na literatura relacionando as teorias da informação e da comunicação ao estudo do genoma são apresentados nesta seção. Em seguida, na Seção 1.2, apresentamos a proposta do presente trabalho com o objetivo de mostrarmos que sequências de DNA, tanto em termos da fita simples do DNA quanto da dupla hélice do DNA, podem ser identificadas e reproduzidas via os códigos corretores de erros (CCEs), denominados códigos *G*-linearidade. Portanto, existe uma estrutura matemática, algébrica e geométrica associada a essas sequências que permite sua identificação, reprodução e classificação. Ainda nesta seção, sugerimos uma nova abordagem relacionando os processos de armazenamento, organização, transmissão e identificação da informação biológica aos protocolos usados em sistemas de redes locais de computadores. Finalmente, na Seção 1.3 mostramos como o presente trabalho está organizado.

Com objetivo de facilitar a compreensão do conceito sobre sequências de DNA mencionado no decorrer deste trabalho, faz-se necessária a seguinte definição: sequência de DNA é a estrutura primária de uma molécula de DNA representada por uma série de nucleotídeos. Estes nucleotídeos quando estiverem dispostos emparelhados (adenina com timina, citosina com guanina, guanina com citosina e timina com adenina) resultam na formação da dupla hélice do DNA, ou resultam formação da fita simples do DNA quando não estiverem emparelhados (adenina, citosina, guanina e timina), como ilustrado na Figura 1.1.

Durante a evolução da célula formou-se uma molécula, que hoje sabemos ser o ácido

desoxirribonucléico (DNA): macromolécula, formada pela junção de um grande número de nucleotídeos, e que contém a informação genética codificada. O DNA constitui uma espécie de código que determina o que uma célula tem, e possui informações de controle não só da célula em si, mas também (e principalmente nos seres multicelulares) da relação com as demais células. O material responsável pelo comando e coordenação de toda a atividade celular e pelas divisões celulares e transmissões das características hereditárias está organizado nas células em unidades conhecidas como cromossomos, que, por sua vez, compõem o genoma de um organismo. Cada cromossomo é constituído por uma longa fita dupla de DNA. O DNA contém genes, promotores, sequências repetitivas, RNA não codantes, mas que são expressos, etc. Os genes são sequências compostas de exons e íntrons que, após a transcrição, são traduzidas em proteínas.

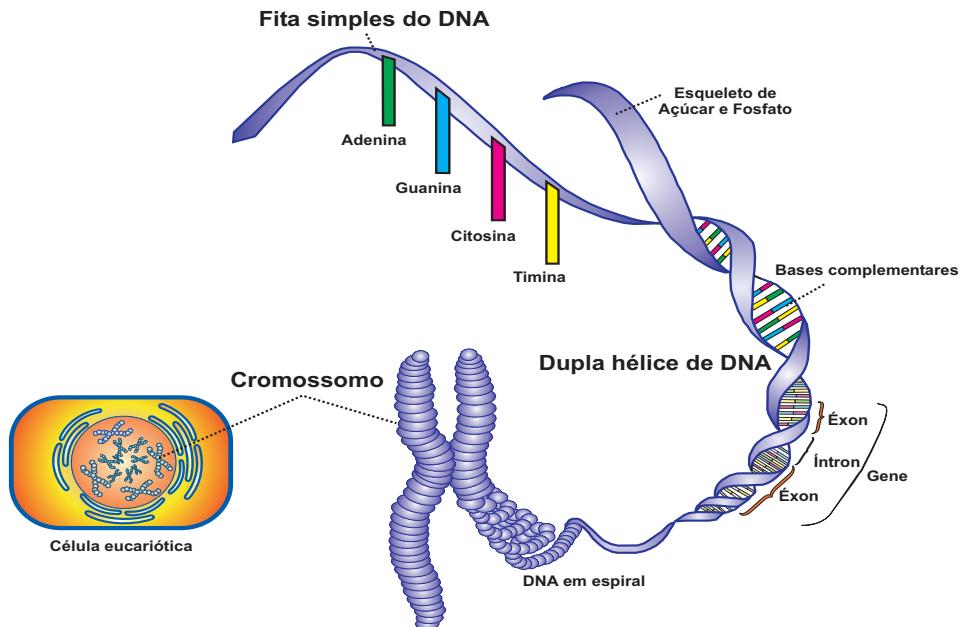


Figura 1.1: Série de nucleotídeos representando a estrutura primária de uma molécula de DNA em termos da dupla hélice e da fita simples do DNA.

Portanto, as sequências de DNA que serão analisadas no presente trabalho através dos códigos G-linearidade estão definidas da seguinte forma:

- Sequências em termos das fitas simples do DNA** - são sequências de nucleotídeos, como sendo, genoma, gene, DNA repetitivo, íntron, RNA, RNA mensageiro (mRNA), proteína, hormônio, e, sequências de direcionamento e de sinal interno de proteínas organelares.
- Sequências em termos da dupla hélice do DNA** - são sequências bases comple-

mentares. Na dupla hélice, as duas fitas de DNA estão em direção opostas, o que significa que são anti-paralelas. O termo anti-paralelas deve-se ao fato de que uma das fitas tem a direção exata da sua síntese ($5'$ - $3'$) enquanto a outra está invertida ($3'$ - $5'$). Todas as sequências citadas no item a) serão analisadas pelo CCE em termos da dupla hélice.

1.1 Breve Histórico

Em [1], Battail menciona que tanto a teoria de comunicação quanto a genética preocupam-se com o envio da informação. Ambas as teorias diferem entre si, principalmente, por não operarem na mesma dimensão - a teoria de comunicação é realizada pelo homem enquanto que a genética é realizada por um processo natural. A teoria de comunicação está programada para enviar mensagens no espaço, de um lugar para outro, já a genética está programada para enviar mensagens no tempo, de um instante a outro (hereditariedade).

A tecnologia de comunicação conta com o progresso significativo na concepção de dispositivos físicos, mas também, embora muito menos perceptível, no desenvolvimento de um poderoso ferramental conceitual, consistente, garantido pela teoria da informação. Embora esse ferramental tenha sido originalmente desenvolvido para a comunicação através do espaço, ele é suficientemente abrangente para aplicação na comunicação através do tempo [1].

Ainda em [1], uma questão central pode ser colocada da seguinte maneira: a estrutura teórica elaborada pelo homem pode contribuir para uma melhor compreensão dos processos naturais que envolvem a comunicação genética? Segundo Battail, a resposta é positiva, provida principalmente pela teoria da informação que renova a visão, que podemos ter do “mundo vivo”.

1.1.1 Shannon e a teoria da informação

A teoria matemática das comunicações ou teoria da informação é a primeira teoria da comunicação que começou a germinar no pós-guerra, no âmbito da matemática e da engenharia elétrica, e ao nível das telecomunicações.

Até a década de 1930, engenheiros eletricistas podiam construir circuitos eletrônicos para resolver problemas lógicos e matemáticos, mas, a maioria o fazia sem qualquer processo, de forma particular, sem rigor teórico para tal. Isso mudou com a tese de mestrado (1937) do matemático Claude E. Shannon (1916-2001), *A Symbolic Analysis of Relay and Switching Circuits*, que deu início ao que mais tarde seria considerado a teoria da informação. Em sua tese de doutorado (1940), Shannon desenvolveu uma proposta sobre relações matemáticas

ligadas à genética Mendeliana, com o objetivo de esclarecer como diferentes combinações de características se propagaram através de várias gerações. Apesar do trabalho ser muito original naquela época, o fato de não ter sido publicado tornou-o pouco conhecido e divulgado. Após ter concluído a sua tese de doutorado, Shannon desviou seu foco para a comunicação digital e criptografia.

Em 1948, Shannon publicou o artigo intitulado *A Mathematical Theory of Communication*, [2]. Em 1949, Shannon, juntamente com o matemático Warren Weaver (1894-1978), publicou o livro *The Mathematical Theory of Communication*, contendo reimpressões do artigo científico anterior de forma também acessível a não-especialistas, cujo conteúdo serviu como fundamento para áreas de estudo como compressão de dados e criptografia.

Nestas publicações, os aspectos fundamentais do sistema de comunicação foram considerados, sendo apresentado um modelo linear de comunicação, simples mas extraordinariamente eficiente na detecção e resolução dos problemas técnicos. Essa teoria enfatizava a teoria da probabilidade e mostrava um primeiro interesse no codificador e decodificador, ambos em termos de seus papéis funcionais e da existência, ou não existência, de codificadores e decodificadores que realizam um certo nível de desempenho. Shannon ainda proporcionou aos engenheiros da comunicação um modo de determinar a capacidade de um canal de comunicação em termos de ocorrência de bits.

A teoria matemática da comunicação visava a precisão e a eficácia do fluxo informacional, procurando não se restringir apenas à área da engenharia, mas, servir de referência a qualquer âmbito da comunicação. Pretendia, assim, ser adaptável a qualquer processo de comunicação, independentemente das características dos seus componentes.

A computação e a comunicação digital transformaram as relações sociais e a forma de se fazer ciência no século XX. Shannon é portanto o formulador de dois princípios fundamentais da “Era Digital”: a teoria que associa a lógica de Boole ao funcionamento dos dispositivos eletrônicos, e a teoria que fundamenta a codificação digital de qualquer forma de comunicação. A primeira proporcionou o desenvolvimento de circuitos cada vez mais elaborados e complexos, e seus princípios ainda hoje são válidos para qualquer computador ou dispositivo eletrônico. A segunda favoreceu o desenvolvimento de aplicações de armazenamento e comunicação de dados cada vez mais eficazes e, é o princípio da convergência digital em curso no século XXI.

Nos últimos anos, tem havido uma procura crescente de serviços altamente eficientes e viáveis para os sistemas de armazenamento e a transmissão de dados digitais, tem sido estendida e trazida para o ponto onde ela está sendo aplicada em sistemas de comunicações práticos. A teoria da informação tornou-se, então, um ramo da teoria da probabilidade e da matemática estatística que lida com sistemas de comunicação, transmissão de dados,

criptografia, codificação, teoria do ruído, correção de erros, compressão de dados, etc. Esta exigência foi acelerada em virtude da emergência, em larga escala e de alta velocidade, do processamento, armazenamento e transmissão da informação digital nas esferas comercial, governamental e militar. Uma preocupação importante do sistema de transmissão é o controle dos possíveis erros para que os dados possam ser reproduzidos fielmente.

1.1.2 Os avanços da biologia celular e molecular

As reflexões sobre o fenômeno da vida, sua origem e processos, emergiram em várias civilizações e culturas, ao longo do tempo. Em um primeiro momento, o objetivo do homem em obter conhecimento sobre o mundo natural voltava-se ao desenvolvimento de técnicas que garantissem sua sobrevivência ¹.

Em 1833, foi sintetizada artificialmente a primeira enzima (diastase): uma nova ciência, a bioquímica, começava a dar os primeiros passos. Em 1839, Mathias Schleiden e Schwann propõem a teoria celular, que estabelecia que a célula é a unidade básica de constituição dos organismos e que as mesmas eram produzidas por células pré-existentes. Em 1859 o britânico Charles Darwin publicava sua grande obra, *A Origem das Espécies*, na qual descreve a seleção natural como mecanismo primário da evolução. Esta teoria se desenvolveu no que é agora considerado o paradigma central para explicação de diversos fenômenos na biologia. Em 1866, a genética dava os seus primeiros passos graças ao trabalho do monge austriaco, Gregor Mendel que formulou as leis da hereditariedade. No entanto, o seu trabalho permaneceu na obscuridade durante 35 anos.

Oswald Avery, em 1943, mostrou concludentemente que era o DNA e não as proteínas, que compunham o material genético dos cromossomos. Em 1953, James Watson, Francis Crick, Maurice Wilkins e Rosalind Franklin demonstraram o funcionamento e a estrutura em dupla hélice do DNA. Com esta descoberta, ficou claro que a informação genética está armazenada na forma de duas fitas diretamente complementares compostas por letras de um alfabeto de quatro símbolos. Até a descoberta das bases moleculares da genética, os pesquisadores concentravam-se na genética clássica, baseada nas leis de Mendel.

Em 1965, foi demonstrado que células normais em cultura dividiam-se apenas um número limitado de vezes (o limite de Hayflick), envelhecendo e morrendo depois. Por volta desta data, descobriu-se que as células-tronco eram uma exceção a esta regra e começou-se o seu estudo exaustivo. O estudo das células-tronco começou a ser crucial para se entender a biologia do desenvolvimento, levando também à esperança de aparecimento de novas aplicações médicas. O estudo dos organismos, da sua reprodução e da função dos seus órgãos, pas-

¹As referências adotadas nesta subseção são [3] e [4].

sou a ser efetuado no nível molecular. O reducionismo constitutivo na análise dos processos biológicos voltados ao funcionamento dos organismos, tornava-se cada vez mais triunfante e promissor. Mesmo os processos de classificação científica dos organismos passaram a utilizar dados moleculares como as sequências de DNA e RNA como caracteres importantes no processo.

As descobertas a partir da década de 50, aliadas à descoberta da primeira enzima de restrição em 1968 e da técnica de PCR (*polymerase chain reaction*) em 1983, proporcionaram o impulso da ciência a que hoje é denominada biologia molecular.

Apesar de o processo de clonagem em plantas já ser conhecido há milênios, somente em 1997 o primeiro animal, a ovelha Dolly, foi clonado através do processo de transferência de um núcleo de célula somática para o citoplasma de um ovócito anucleado (processo de transferência nuclear). Poucos anos mais tarde, outros mamíferos foram clonados pelo mesmo método: cães, gatos, vacas e cavalos.

Avanços espetaculares têm ocorrido em muitas áreas da biologia celular e molecular, contribuindo para a extraordinária beleza da área. A conclusão do projeto genoma, que, através do mapeamento genético, conseguiu identificar cada um dos 100 mil genes do corpo humano, foi, sem dúvida alguma, o maior feito da ciência no início deste século. Essas descobertas revolucionaram a ciência da vida e proporcionaram um desenvolvimento em tecnologias de DNA recombinante e o lançamento das indústrias biotecnológicas. E são ainda maiores as expectativas em torno do conhecimento gerado. Encontrar os mecanismos genéticos que interferem em doenças graves como diabetes, hipertensão e câncer, é um dos principais objetivos da biotecnologia.

A definição de biotecnologia é ampla: tecnologia baseada na biologia com o uso de organismos vivos ou parte deles, para a produção de bens e serviços. Nesta definição se enquadraram um conjunto de atividades que o homem vem desenvolvendo há milhares de anos, como a produção de alimentos fermentados (pão, vinho, iogurte, cerveja, e outros). Por outro lado, a biotecnologia moderna é considerada aquela que faz uso da informação genética, incorporando técnicas de DNA recombinante.

A biotecnologia é uma área bastante promissora e combina disciplinas tais como genética, biologia molecular, bioquímica, embriologia e biologia celular, com a engenharia química, tecnologia da informação, robótica, bioética e o biodireito, entre outras. Pode ser vista como a intersecção entre essas áreas, como mostrado na Figura 1.2. O termo biotecnologia deveria ser empregado num sentido muito mais amplo para descrever uma completa gama de métodos, tanto antigos quanto modernos, usados para manipular organismos visando atender às exigências humanas.

Com a necessidade da redução de tempo e custos nos experimentos laboratoriais, pes-

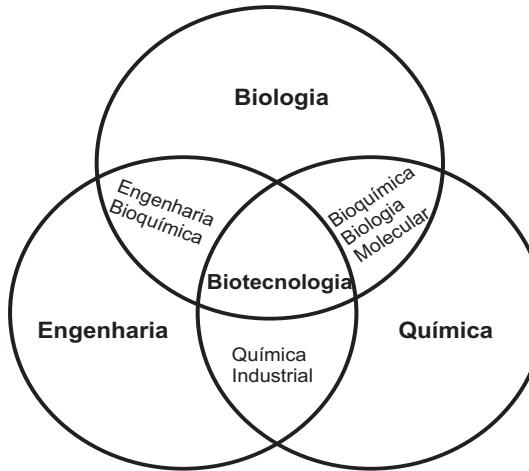


Figura 1.2: Biotecnologia.

quisadores de várias frentes de estudos (matemática, física, química, biologia, engenharias, computação, etc), fixados nas exigências humanas, buscam metodologias científicas capazes de reproduzirem os dados biológicos. A biotecnologia se fortalece com a união dessas frentes de estudos e seu crescimento é bastante promissor neste século.

1.1.3 Modelos propostos na literatura

Historicamente, a aplicação da teoria da informação para análises de dados genéticos iniciou-se na década de 1970, porém esses esforços não tiveram sucesso. Após dez anos, o aumento de dados genéticos despertou novamente o interesse na aplicação da teoria da informação para o estudo do genoma. Esse segundo período de pesquisas continua até o presente momento, porém com um número muito reduzido de pesquisadores. Os trabalhos atualmente buscam analogias entre o fluxo de informação biológica e o sistema de comunicação, dividindo-se basicamente em três linhas de pesquisas: teoria da informação genética, teoria da comunicação genética e a teoria da codificação genética.

A aplicação das teorias da informação, comunicação e codificação em sistemas biológicos contribuem para uma melhor compreensão dos paradigmas biológicos, fazendo com que a biologia, que hoje é uma ciência descritiva, se transforme em uma ciência fundamentada teoricamente. Por outro lado, os avanços das teorias da informação, comunicação e codificação podem ser alcançados através da compreensão do sistema de informação biológico. Esta nova abordagem é muito promissora, podendo proporcionar vários avanços, tais como:

- Identificar sistemas biológicos que podem ser investigados experimentalmente usando as teorias da informação, comunicação e codificação;

- Usar a quantidade de dados e experiências disponíveis para testar a aplicação das teorias da informação, comunicação e codificação;
- Compreender como as interferências afetam os sistemas biológicos;
- Descobrir como as hipóteses da teoria da informação e codificação podem ser modificadas ou flexibilizadas para aplicá-las em sistemas moleculares;
- Compreender como a codificação multidimensional é gerada e usada em proteínas e outras estruturas biológicas;
- Aprender como a codificação tridimensional, como, por exemplo, a codificação encontrada entre as superfícies de interação de proteínas e como elas operam;
- Como as teorias da informação, comunicação e codificação podem explicar os parâmetros de interação biológica entre moléculas, sendo que tais parâmetros podem ser usados na construção do mais alto nível de um sistema biológico;
- A criação de novas técnicas de codificação que aproximam a capacidade do canal para uma aplicação molecular tanto em nível nanotecnológico quanto em nível macroscópico.

Motivados pelo fato de que o código genético é degenerado, isto é, apresenta uma estrutura redundante, vários pesquisadores em teoria da informação e da codificação realizam pesquisas em duas frentes, a saber:

- 1º) Sob o ponto de vista da teoria da informação, aplicar os conceitos inerentes desta com o objetivo de apresentar um método sistemático de determinação das regiões codantes e não-codantes na estrutura do DNA (*source coding problem*);
- 2º) Sob o ponto de vista da teoria da codificação, fornecer a fundamentação necessária para a caracterização de CCEs (*channel coding problem*).

Todavia, sob o ponto de vista da teoria de comunicação, as pesquisas concentram-se mais no aspecto de adaptação do modelo tradicional de um sistema de comunicação digital, em termos de diagrama de blocos, àquele do sistema biológico por considerar que a informação contida no genoma (estrutura do DNA) se realiza através de pacotes de informação contendo regiões com sequências relacionadas a sincronismo, identificação de pacotes, codificantes, não codificantes, etc.

Na literatura, existem alguns trabalhos que relacionam as semelhanças entre um sistema de comunicação e a biologia molecular com o objetivo de modelar os diversos sistemas

biológicos. Schneider em [5], [6] e [7], apresenta um procedimento sistemático para identificar as regiões codantes e não codantes nas sequências de DNA utilizando conceitos da teoria da informação. Yockey, [8], apresentou um modelo de sistema de comunicação digital com aquele associado ao da expressão gênica, ou equivalentemente, ao do fluxo de informação biológica. Forsdyke, [9] e [10], considerou a possibilidade de que os íntrons poderiam ser os dígitos de verificação de paridade associados aos exons. Por outro lado, Rzeszowska-Wolny, [11], propôs que um arranjo apropriado do DNA em nucleosomos pode ser relevante para a operacionalidade deste sistema. Liebovitch, [12], propôs um procedimento que torna possível determinar se um tipo de código corretor de erro está presente ou não na sequência do DNA. Rosen, [13], apresentou um método para a detecção de códigos de bloco lineares que explica a possibilidade de inserções e deleções nas sequências de DNA. Battail, [14], argumenta sobre a existência de códigos entrelaçados no DNA, uma vez que o tamanho do genoma humano é muito maior que o necessário para especificar as características de cada indivíduo. May et. al., [15], propôs o uso de códigos de bloco e convolucionais no processo de inicialização da tradução em organismos procariontes. Mac Donaill, [16], propôs um código de verificação de paridade relacionado à composição dos nucleotídeos. Sánchez et. al., [17], propôs a construção de um espaço vetorial associado ao código genético tendo como estrutura matemática o corpo de Galois com 64 elementos, identificando cada aminoácido com uma sequência binária, possibilitando dessa forma uma caracterização geométrica associada ao código genético. A abordagem destes dois últimos artigos está relacionada exclusivamente com o código genético.

Uma questão sempre presente na maioria dos trabalhos relacionados com a codificação genética e a codificação genômica é a seguinte:

- Existe algum tipo de código corretor de erros na estrutura do DNA?

Embora importantes, os trabalhos citados anteriormente não foram capazes de fornecer subsídios sobre a existência do mais simples dos CCEs nas sequências de DNA, como mencionam [18] e [19].

O modelo proposto em [20] e o presente trabalho indicam, de maneira positiva, uma resposta a essa pergunta com base nas linhas de pesquisa da teoria da comunicação genética e a teoria da codificação genética, e deram origem ao pedido de patente [21], em que o modelo de um sistema de comunicação da informação genética é comum na proposta de pesquisa pertinente aos três trabalhos. É importante ressaltarmos que o presente trabalho dá continuidade ao trabalho descrito em [20], bem como amplia os resultados apresentados em [20] e [21]. Até onde é de nosso conhecimento, pela primeira vez sequências primárias de uma

fita simples do DNA, com características biológicas distintas e comprimentos variados, são identificadas como palavras-código de um CCE e reproduzidas em termos dos nucleotídeos e dos aminoácidos correspondentes. Outro avanço é com relação à identificação da sequência da dupla hélice do DNA como palavra-código de um CCE e sua reprodução em termos das bases complementares.

A seguir, de maneira resumida, comentaremos a proposta de pesquisa e apresentaremos a proposta do presente trabalho, na qual evidenciaremos os avanços alcançados.

1.2 Proposta de Pesquisa

Um dos grandes desafios para os pesquisadores que atuam na área da biologia molecular utilizando os conceitos das teorias da informação, da codificação e da comunicação é mostrar a existência de CCEs na estrutura do DNA. Em [18], Dawy declara: “*However, neither work was able to support the existence of such simple error correcting codes in the DNA*”, quando fazia um breve relato sobre os trabalhos de [8], [12]- -[16] e [19]. E em [1] - página 166, Battail comenta: “*Given a sequence of symbols, say, of nucleotides, consider the problem of determining whether this sequence is a word of some error-correcting code and, if so, of identifying the code to which it belongs. As stated, this problem has no solution*”.

Estes pesquisadores tentam mostrar a existência de CCEs no genoma, o que torna o problema muito complexo. Em [1], Battail apresenta duas hipóteses declarando: “*The survival of an organism necessitates the existence of a reliable information replication process. Therefore error-correcting codes must be used in replication or in another process of information regeneration that precedes replication*”; “*The genetic information undergoes nested encoding, where the result of a previous encoding process is combined with new information and encoded again. The more important genetic information is assumed to be in the primary coded message, regarding nested coding mirrors coding theory’s concept of concatenated codes which are also called nested codes*.”

Modelo proposto em [21]

Em [21], consideramos as duas hipóteses, apresentadas anteriormente, com algumas restrições. Ao invés de analisarmos o processo de replicação do DNA com o objetivo de mostrar a existência de CCEs no genoma, o que torna o problema muito complexo, partimos da seguinte conjectura: Se o genoma é constituído por subregiões consistindo de gene, DNA repetitivo, ítron, RNA, RNA mensageiro (mRNA), proteínas, hormônio, e, sequências de direcionamento e de sinal interno de proteínas organelares, etc, e que cada uma dessas regiões pode

ser reproduzida por um código específico, então o genoma consiste de códigos concatenados (“*nested codes*”), no mínimo justapostos.

Inicialmente, propomos uma analogia entre o dogma central da teoria de comunicação e o dogma central da biologia molecular, na qual os blocos do transmissor, canal e receptor estão relacionados ao DNA, processos de transcrição/tradução e proteína, respectivamente.

Em seguida, propomos um modelo de um sistema de comunicação de informação genética, ilustrado na Figura 1.3, tendo como objetivo mostrar que algumas sequências de DNA (apenas fita simples) possuem uma estrutura matemática e, portanto, podem ser identificadas como palavras-código dos códigos *G*-linearidade e reproduzidas em termos dos nucleotídeos e dos aminoácidos correspondentes.

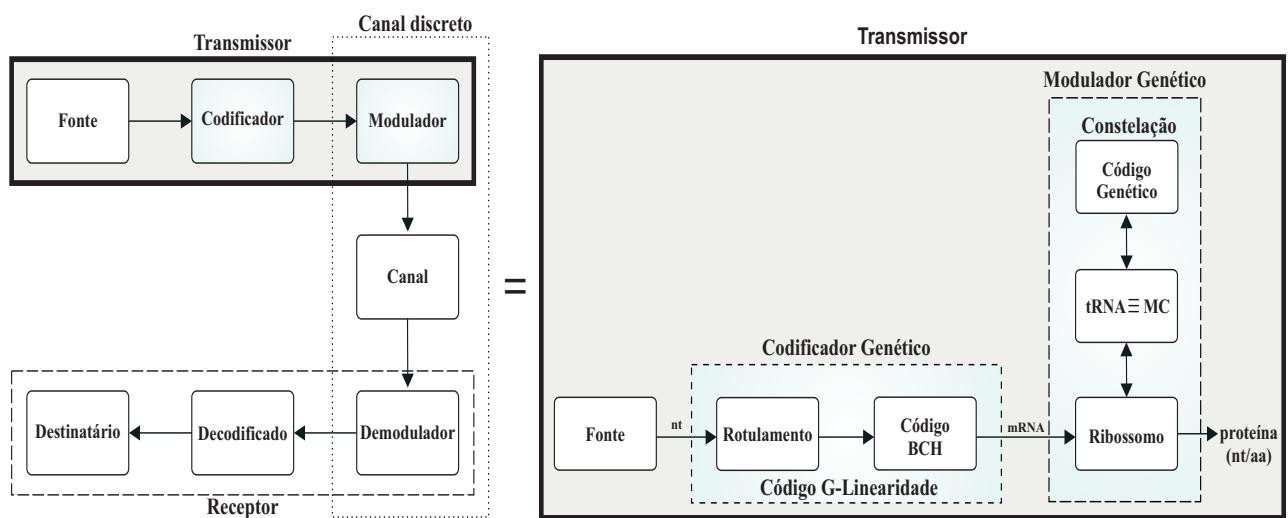
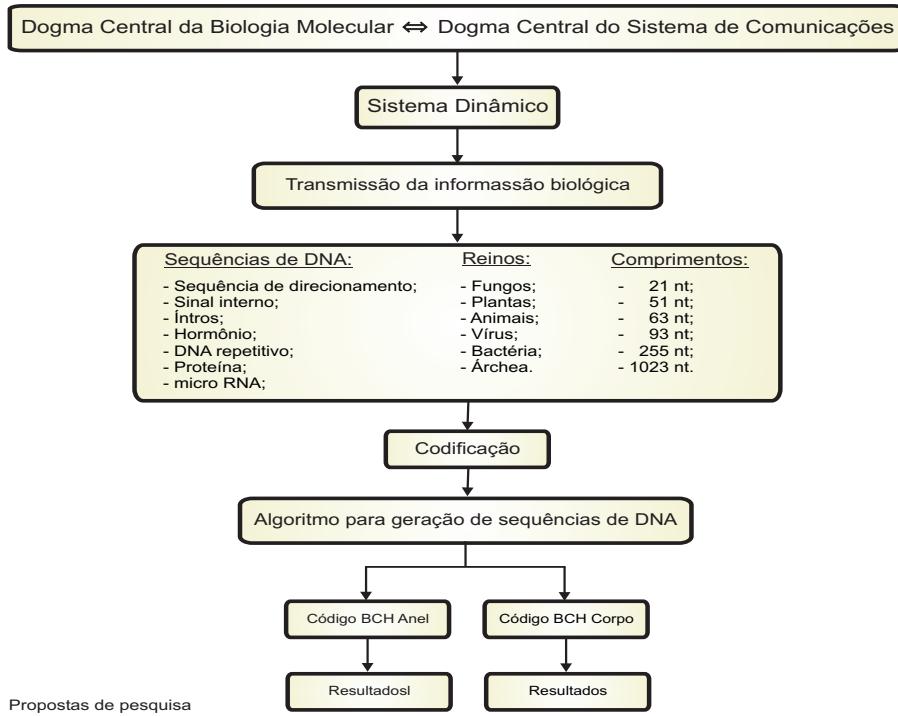


Figura 1.3: Modelo de um sistema de comunicação de informação genética.

Finalmente, desenvolvemos um algoritmo para geração de sequências de DNA, que, através da codificação, identifica e reproduz tais sequências. Os resultados podem ser direcionados em metodologias voltadas em análises mutacionais e de polimorfismos, reduzindo tempo e custos laboratoriais.

Com um leque de possibilidades de uso na indústria, além do significado científico importante, antes de publicarmos em periódicos científicos, depositamos uma patente internacional pelo Tratado de Cooperação em Patentes (PCT), em vários países, e outra nos Estados Unidos, com financiamento integral da FAPESP e gerenciamento da Agência de Inovação da Unicamp e da Agência USP de Inovação, [21]. Com um aprimoramento do modelo e o desenvolvimento de um software, as estruturas matemáticas de qualquer sequência de DNA produtora de proteínas dentro da célula poderão ser testadas em um amplo leque de aplicações e

produtos. Essas informações são importantes para a produção de novos fármacos, criação de novos produtos, aplicação em empresas de agronegócios, melhoramento genético, indústrias têxteis, e, de biotecnologia de um modo geral. A seguir, mostramos um esquema da metodologia considerada em [21], cujo desenvolvimento se deu através da tese de doutorado [20] e da presente tese.



Modelo proposto por Rocha em [20]

Em [20], foi proposto o modelo de um sistema de comunicação para importação de proteínas organelares que consiste na caracterização dos processos de codificação e decodificação de sequências de direcionamento dessas proteínas. O processo de codificação é realizado pelo bloco transmissor, mostrado na Figura 1.3, e o processo de decodificação é realizado pelo bloco receptor, mostrado na Figura 1.4.

A palavra-código na saída do codificador está relacionada à sequência de direcionamento em termos de nucleotídeos, enquanto que a saída do modulador está relacionada à sequência de direcionamento em termos de aminoácidos, Figura 1.3. O modelo de decodificação baseia-se no processo de importação de proteínas mitocondriais. A sequência reproduzida pelo código é a informação gerada pela fonte. Os erros ocorridos no processo de codificação durante a geração da sequência são localizados e corrigidos através do processo de decodificação (Algoritmo de Berlekamp-Massey modificado para anéis), Figura 1.4.

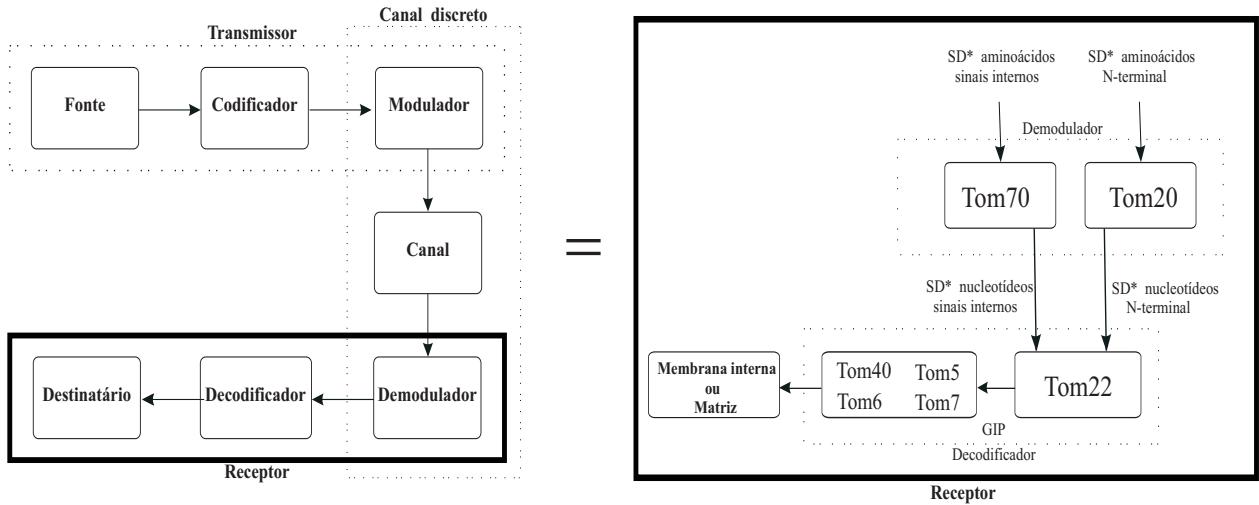


Figura 1.4: Modelo proposto para decodificação biológica.

Em [20], foram analisadas vinte (20) sequências de direcionamento, SD's, de proteínas organelares com comprimentos iguais a sessenta e três (63) nucleotídeos, das quais dez foram identificadas e reproduzidas, diferindo em um (1) nucleotídeo das sequências postadas no NCBI², através de códigos *G*-linearidade ((63,57,3) BCH primitivos, rotulamentos A, B e C) sobre $GR(4, 6)$ - $r = 6$. Essas sequências foram analisadas em termos da fita simples do DNA, tanto no processo de codificação quanto no processo de decodificação. Os resultados apresentados foram relevantes, apontando uma classificação matemática para essas sequências de direcionamento.

Contudo, inicialmente, foram analisadas sequências de DNA com funcionalidade biológica específica (sequências de direcionamento de proteínas organelares - SD's), comprimento específico (63 nucleotídeos) e uma única estrutura algébrica (BCH primitivo sobre $GR(4, 6)$).

1.2.1 Proposta de Trabalho

No presente trabalho, estamos interessados nos processos de identificação, reprodução e classificação matemática de sequências de DNA através dos CCEs, sendo os objetivos principais:

- Ampliar os resultados apresentados em [20], tanto no contexto biológico quanto no contexto matemático, para outras sequências de DNA com características e funções biológicas distintas, incluindo sequências relacionadas a vírus e bactérias causadoras de patologias clínicas, além de sequências de proteínas, gene e genoma;

²National Center for Biotechnology Information

- b) Mostrar evidências da existência de uma estrutura matemática em sequências de DNA que contenham informações genética e genômica, em termos das fitas simples e da dupla hélice do DNA, respectivamente;
- c) Apontar para uma nova abordagem relacionando os processos de armazenamento, organização, transmissão e identificação da informação biológica aos protocolos usados em sistemas de redes locais de computadores.

Inicialmente, buscamos sequências de DNA postadas no *NCBI* com características biológicas distintas e com diferentes comprimentos, como mostrado na Tabela 1.1. Ao invés de iniciarmos a análise do genoma como um todo, focalizamos primeiramente em suas partes em termos da fita simples do DNA (genomas eucarióticos e procarióticos) e, posteriormente focalizamos em um genoma procarioto como um todo e, por último, focalizamos na dupla hélice do DNA.

Todavia, nos deparamos com algumas perguntas cujas respostas serão importantes na validação do modelo matemático em questão:

- 1) Outras sequências de DNA, como as relacionadas na (Tabela 1.1), diferentes das sequências apresentadas em [20], poderão ser reproduzidas pelos códigos G -linearidade (BCH sobre anéis)?
- 2) Como reproduzir as sequências com comprimentos diferentes de 63 nucleotídeos?
- 3) A Tabela 1.1 ilustra sequências de células eucarióticas (*viridiplantae*, *metazoa* e *fungi*) e de células procarióticas (*archea*, *viruses* e *bacteria*). No contexto biológico, existem diferenças enormes entre essas células (ver Subseção 2.1.1, Capítulo 2). Os códigos G -linearidade também serão capazes de reproduzir sequências das células procarióticas?
- 4) Será possível reproduzir uma proteína inteira sem diferença de aminoácido?
- 5) Será possível reproduzir um gene e genoma inteiro?
- 6) As sequências identificadas e reproduzidas através dos códigos G -linearidade (BCH sobre anéis) também poderão ser identificadas e reproduzidas através de outros códigos? Em caso positivo, quais?
- 7) E a dupla helice do DNA? Também será possível reproduzi-la? O código capaz de identificar e reproduzir a fita simples do DNA também é capaz de identificar e reproduzir a dupla hélice do DNA?

| Sequências de DNA | Organismos | Espécies | Comprimentos em nucleotídeos |
|---|--|--|------------------------------|
| <i>SDs</i> - Mitocondria <i>SDs</i> - Cloroplasto <i>SDs</i> - Retículo | Diversos | <i>Viridiplantae</i> <i>Metazoa</i> <i>Fungi</i> <i>Viruses</i> | 39, 45, 51, 63, 93 e 255 nt |
| <i>miRNA</i> | <i>H. Sapiens</i> | <i>Metazoa</i> | 21 nt |
| Sinais internos | <i>S. cerevisiae</i> | <i>Fungi</i> | 63 nt |
| Hormônio | <i>Petúnia</i> | <i>Viridiplantae</i> | 63 nt |
| Íntrons | <i>R. norvegicus</i> <i>E. nidulans</i> | <i>Metazoa</i> <i>Fungi</i> | 63 nt |
| DNA repetitivo | <i>H. vulgare</i> <i>O. sativa</i> | <i>Viridiplantae</i> | 63 e 255 nt |
| Proteínas | Diversos | <i>Archea</i> <i>Bacteria</i> <i>Viruses</i> <i>Viridiplantae</i> <i>Metazoa</i> | 63, 255 e 1023 nt |
| Gene <i>TRAV7</i> | <i>Homo sapiens</i> | <i>Metazoa</i> | 511 nt |
| Genoma | <i>lactococcus lactis</i> | <i>Plasmid</i> | 2047 nt |

Tabela 1.1: Sequências que serão analisadas no presente trabalho.

- 8) Existem relatos na literatura ([22] e [23]) de que quando uma sequência de DNA sofre uma mutação em um determinado nucleotídeo, por algum motivo, outro nucleotídeo também é alterado na sequência. Os códigos *G*-linearidade são capazes de gerar e reproduzir sequências com essas características? Em caso positivo, como?
- 9) Será possível reproduzir uma sequência de direcionamento por um código e a proteína (parte madura da proteína³ mais a sequência de direcionamento) por outro código? Esse fato pode apresentar indícios de códigos concatenados nessas sequências de DNA? Como?

No conjunto de sequências de DNA, identificadas na Tabela 1.1, temos as sequências em termos das fitas simples do DNA que contêm informações genéticas (genoma, gene, DNA

³Parte da proteína sem a sequência de direcionamento

repetitivo, ítron, RNA, mRNA, proteínas, hormônio, sequência de direcionamento (SD) e de sinal interno (SI)) e sequências que contêm informações genômicas (dupla hélice do DNA). Sendo assim, propomos a caracterização de dois modelos, um no contexto de codificação genética (Figura 1.3) e outro no contexto de codificação genômica (Figura 1.5).

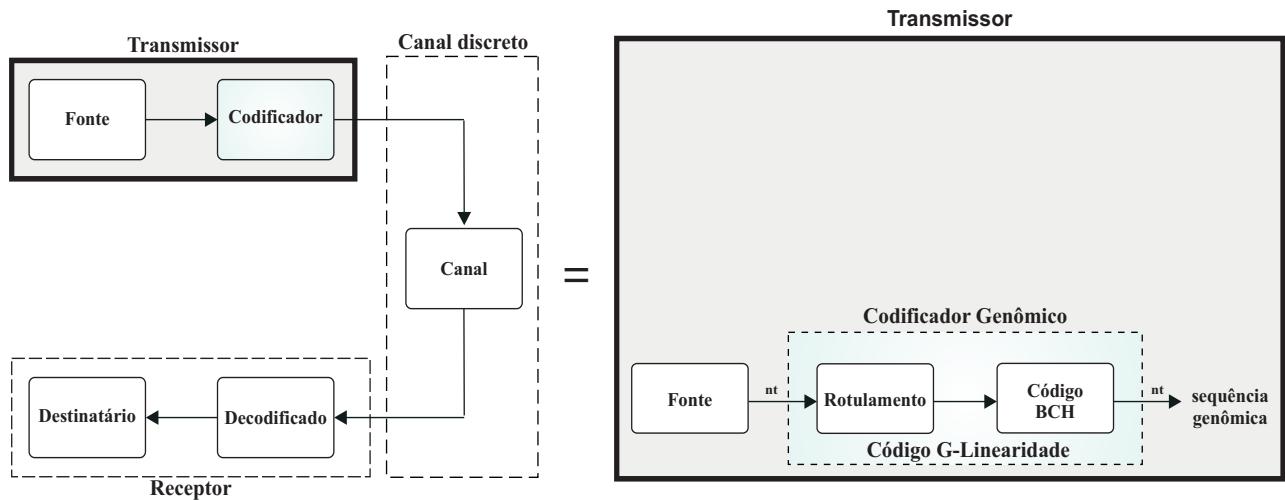


Figura 1.5: Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genônica.

A principal diferença entre eles está relacionada ao codificador de canal. No contexto de codificação genética, o canal é separado em modulador, canal e demodulador. No contexto de codificação genômica, o canal passa a ser uma única entidade chamada “canal discreto podendo ter ou não memória”, como mostra a Figura 1.5. O processo de codificação está relacionado com os possíveis erros que serão introduzidos pelo canal em ambos os modelos.

Para a identificação, reprodução e classificação matemática das sequências de DNA (em termos das fitas simples e dupla) descritas na Tabela 1.1, alguns elementos devem ser considerados, tais como: a estrutura algébrica, o alfabeto, o rotulamento, o mapeamento, o polinômio primitivo e o polinômio gerador, por serem fundamentais na determinação dos CCEs resultantes no codificador. Deste fato decorre a necessidade do desmembramento do codificador genético/genômico e sua caracterização matemática. Portanto, propomos que essas sequências de DNA foram concebidas obedecendo uma estrutura matemática linear, cíclica (abeliana), e não geradas de maneira aleatória.

A caracterização matemática da identificação, reprodução e classificação de sequências genéticas e genômicas deverá ampliar consideravelmente a capacidade de compreensão do funcionamento dos sistemas biológicos e, eventualmente, as possibilidades de sua manipulação serem analisadas por métodos quantitativos.

A associação entre as teorias de informação, comunicação e codificação e a ciência da vida proporciona um desenvolvimento crescente e fundamental para o melhor entendimento dessas ideias. Como uma tendência e uma nova frente de pesquisa, num futuro muito próximo, também poderá ser citado como mais um exemplo das várias aplicações em sistema de comunicação, o sistema de comunicação biológico. Tendo em vista a riqueza de propriedades e características que envolvem o sistema biológico com o objetivo da continuidade da vida, será possível no futuro aprender com esse sistema e, quem sabe, exemplos práticos e mais eficientes possam ser tirados do mundo biológico e usados no mundo das comunicações, e vice-versa.

No presente trabalho, ainda apontamos uma nova abordagem com relação aos processos de armazenamento, organização, transmissão e identificação da informação biológica. Baseados nos sistemas de comunicação intercelular e intracelular existentes no mundo biológico, notamos que esses sistemas são altamente organizados, capazes e eficientes em armazenar e transmitir a informação biológica. Quando comparados com as redes locais de computadores, apresentam fortes evidências em suas semelhanças. Por exemplo - tanto no sistema biológico quanto na comunicação em redes locais de computadores, um conjunto de regras e convenções no processo de armazenamento e transmissão da informação são construídos de acordo com a necessidade de cada um com o objetivo de regulamentar a troca de informações entre as partes, os protocolos. Nessa mesma direção, o modelo de camadas, a arquitetura da rede, a distribuição topológica das entidades envolvidas, a formatação do quadro de enlace, dentre outros, apresentam semelhanças com o sistema biológico. Com isso, o conjunto de conceitos e propriedades pré-estabelecidos em redes locais de computadores podem ser utilizados no contexto biológico e portanto, deve ser estudado futuramente, melhor caracterizado em ambos aspectos, tornando possível uma modelagem matemática apropriada. Dessa maneira, propomos o *Biological frame* do genoma humano sugerindo que a informação genômica pode ser armazenada e organizada de maneira análoga às informações que são armazenadas e organizadas em CD's e, que o armazenamento e a transmissão da informação genômica ocorrem de maneira inversa ao procedimento de transmissão de dados utilizado em redes de computadores. Relacionado ao genoma de um plasmídeo *Lactococcus lactis plasmid* pcl 2.1, propomos a formatação da arquitetura biológica, denominada *Biological frame of Lactococcus lactis plasmid* pcl 2.1 *Genomic*.

1.3 Descrição do Trabalho

O primeiro capítulo deste trabalho apresenta resumidamente, sob uma perspectiva atual e futura, a associação entre as teorias da informação, comunicação e codificação e a biologia

molecular e celular. Neste capítulo é introduzido o contexto dos modelos de codificação genética e codificação genômica de modo a situar o leitor, bem como motivá-lo para a leitura dos capítulos seguintes concernentes ao trabalho envolvido.

No Capítulo 2 é feita uma revisão dos principais conceitos biológicos e de CCEs, em virtude deste trabalho abordar áreas interdisciplinares. O conhecimento dos processos naturais no sistema biológico é fundamental para o entendimento da perpetuação das espécies e até mesmo para a compreensão de inúmeras patologias relacionadas aos erros ocorridos durante a transmissão do DNA, conhecidos como mutações. A introdução à teoria de CCEs permitirá ao leitor, além de observar que vários CCEs podem ser usados no processo de codificação de uma sequência de informação, uma síntese dos conceitos e definições importantes para a construção dos códigos envolvidos neste trabalho.

O terceiro capítulo deste trabalho trata, principalmente, dos modelos de codificação genética e codificação genômica. Embora o grau de complexidade que envolve o sistema biológico e a teoria dos CCEs seja alto, usaremos a comparação entre o sistema biológico e o sistema de comunicação para identificarmos a informação contida em algumas sequências de DNA (fitas simples e dupla hélice), como palavras-código de CCEs, reproduzí-las e classificá-las matematicamente. Nestes modelos, os elementos, tais como: a estrutura algébrica, o alfabeto, o rotulamento, o mapeamento, o polinômio primitivo e o polinômio gerador devem ser considerados, pois a composição desses elementos são fundamentais e determinantes na caracterização dos CCEs resultantes (códigos *G*-linearidade) nos codificadores genéticos e genômicos. Ainda neste capítulo, sugerimos uma nova abordagem de estudo inerente às semelhanças entre os processos biológicos e os protocolos utilizados em sistemas de redes locais de computadores.

O Capítulo 4 tem como objetivo apontar a existência de CCEs em sequências de DNA com características biológicas distintas e comprimentos variados. Para isso, é necessário o desenvolvimento de algoritmos que descrevem os passos da construção e geração dos códigos *G*-linearidade, tais como: os códigos BCH primitivos sobre o corpo $GF(4^r)$ e, os códigos BCH primitivos e não primitivos sobre o anel \mathbb{Z}_4 através das suas extensões de Galois, respectivamente. Como resultados dos algoritmos, palavras-código são geradas e identificadas como sequências de DNA contendo até duas diferenças em nucleotídeos (fita simples do DNA) ou bases complementares (dupla hélice do DNA). Essas diferenças podem estar associadas à mutações e polimorfismos, denominados SNPs (*single nucleotide polymorphism*).

O Capítulo 5 é dedicado à análise dos resultados decorrentes dos modelos de codificação genética e codificação genômica, dos processos de construção e geração dos códigos *G*-linearidade (BCH primitivo sobre corpos e BCH primitivo e não primitivo sobre anéis e suas extensões de Galois, respectivamente), e da identificação, reprodução e classificação

de sequências de DNA através dos códigos \mathbb{Z}_4 -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade. As interpretações desses resultados são realizadas sob o ponto de vista de CCEs, o que possibilita uma nova abordagem e uma classificação matemática dessas sequências identificadas e reproduzidas através dos resultados obtidos com o processo de codificação, incluindo a identificação da existência de códigos concatenados (*nested codes*) na estrutura do DNA, em algumas sequências. A reprodução de sequências relacionadas a patologias importantes, tais como o câncer de mama e a gripe decorrente do vírus *Influenza A*, são mostradas neste capítulo, assim como a reprodução de várias proteínas, de um gene e de um genoma completo.

O Capítulo 6 resume as conclusões dos processos de identificação, reprodução e classificação de sequências de DNA, bem como apresenta propostas e tendências de trabalhos futuros.

Finalmente, no Apêndice A, são mostradas as tabelas referentes a todas as distâncias mínimas relativas aos polinômios geradores de todos os códigos construídos neste trabalho. No Apêndice B mostramos as sequências de DNA que foram identificadas e reproduzidas através dos códigos \mathbb{Z}_4 -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade em termos de nucleotídeos, aminoácidos, polinômios primitivos, polinômios geradores e rotulamentos, que, por uma questão de volume de dados, não estão incluídas nos capítulos pertinentes.

Capítulo 2

Preliminares sobre o DNA e os Códigos Corretores de Erros

Em virtude de este trabalho abordar áreas interdisciplinares, este capítulo introduz alguns conceitos da biologia molecular e da teoria de CCEs, indispensáveis para o desenvolvimento do presente trabalho. Salientamos que, embora o grau de complexidade que envolve o sistema biológico e a teoria dos CCE seja alto, usaremos a comparação entre o sistema biológico e o sistema de comunicações para identificarmos a informação contida nas sequências primárias de algumas sequências de DNA como palavras-código de CCE e reproduzí-las.

O presente capítulo está organizado da seguinte maneira: Na Seção 2.1, fazemos uma breve revisão de alguns conceitos introdutórios sobre o DNA e uma breve elucidação do dogma central da biologia molecular, com o objetivo de compreender o processo de geração de proteínas, realizado no sistema biológico. Entender esse processo natural no sistema biológico é fundamental para a compreensão de inúmeras patologias relacionadas aos erros ocorridos durante a transmissão do DNA, conhecidos como mutações. Na Seção 2.2, revisamos os conceitos básicos sobre álgebra abstrata necessários para o entendimento do algoritmo. Finalmente, na Seção 2.3, apresentamos uma introdução à teoria de CCE e, vemos que vários CCEs podem ser usados no processo de codificação de uma sequência de informação. Porém, neste trabalho, focamos no processo de codificação de sequências contendo a informação primária das sequências de DNA através dos códigos da *G*-linearidade. Assim, os conceitos necessários para a construção desses códigos serão abordados nesta seção.

2.1 Revisão Biológica

Os conceitos apresentados nesta seção podem ser encontrados em [24]-[32].

2.1.1 A vida começa nas células

A **célula** é a unidade básica da vida em todas as formas de organismos vivos, da mais simples bactéria ao mais complexo animal.

Assim como seres humanos, cada célula que forma nosso corpo deve crescer, reproduzir-se, processar informações, responder a estímulos e realizar uma série impressionante de reações químicas. A vida é definida por essas capacidades. Os seres humanos e os outros organismos multicelulares são compostos por bilhões ou trilhões de células organizadas em estruturas complexas, mas outros organismos consistem de uma única célula. Mesmo os organismos unicelulares simples exibem todas as propriedades características da vida, indicando que a célula é a unidade fundamental, [24] e [25].

Diversidade e semelhança entre as células

As células se apresentam sob uma grande variedade de tamanhos e formas. Algumas movem-se rapidamente e têm estruturas que se alteram com muita rapidez. Outras são preferencialmente estacionárias e estruturalmente estáveis. O oxigênio pode matar algumas células, sendo, no entanto, essencial para outras. As células dos organismos multicelulares, em sua maioria, estão intimamente relacionadas com outras células. Apesar de alguns organismos unicelulares viverem em isolamento, outros formam colônias ou vivem sob forte associação com diferentes tipos de organismos.

Apesar dessas e de muitas outras diferenças, todas as células compartilham determinadas características estruturais e realizam diversos processos complexos por vias bastante semelhantes.

O universo biológico é constituído de dois tipos de células - **procarióticas** e **eucarióticas**. As células procarióticas consistem de um único compartimento fechado, que é delimitado por uma membrana plasmática; essas células não possuem um núcleo definido e apresentam uma organização interna relativamente simples (Figura 2.1a). As bactérias, organismos procariotos mais numerosos, são organismos unicelulares.

As células eucarióticas, diferentemente das células procarióticas, contêm um núcleo definido delimitado por membrana e uma grande quantidade de membranas internas que delimitam outros compartimentos, as organelas (Figura 2.1b). A região da célula existente entre a membrana plasmática e o núcleo é o citoplasma, o qual compreende o citosol (fase aquosa) e as organelas. Os eucariotos englobam todos os membros dos reinos animal e vegetal, incluindo os fungos, que ocorrem sob formas multicelulares (bolores e cogumelos) e unicelulares (leveduras), e os protozoários, que são exclusivamente unicelulares. Outra diferença importante entre procariotos e eucariotos está na maneira como a própria **molécula de DNA** está

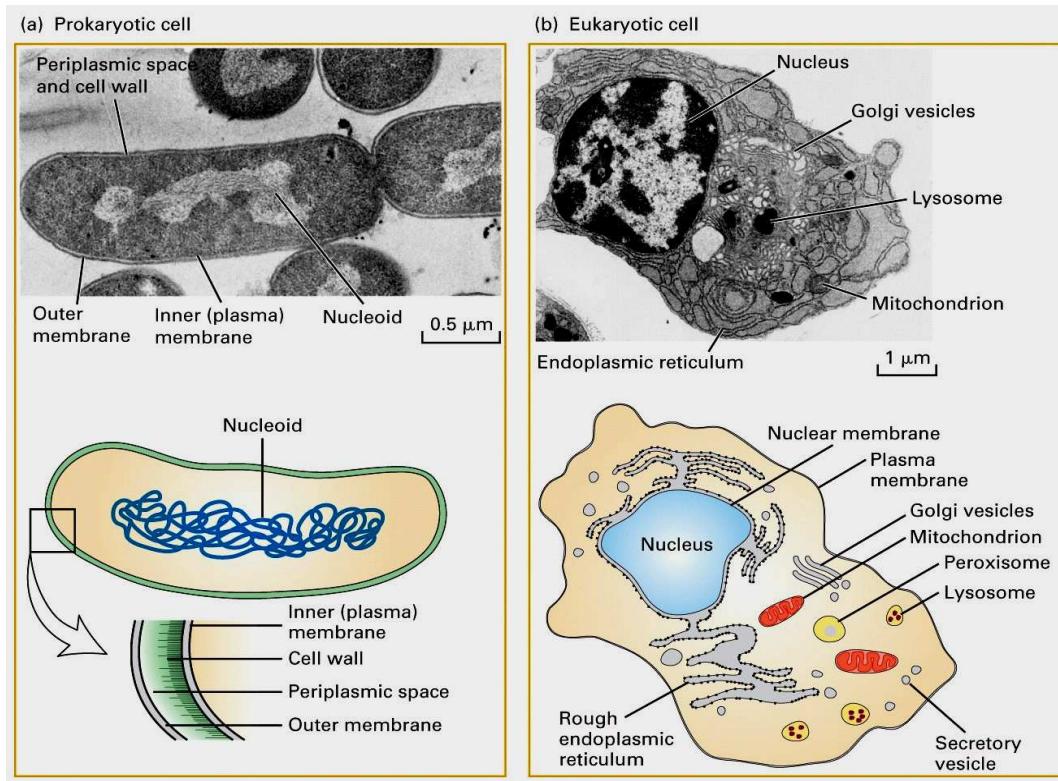


Figura 2.1: As células procarióticas apresentam uma organização interna mais simples do que as células eucarióticas. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

organizada. O **DNA** (ácido desoxirribonucléico) é uma molécula formada por duas cadeias na forma de uma dupla hélice. Por exemplo, o DNA dos eucariotos está altamente associado com proteínas específicas para formar nucleoproteínas (que são reorganizadas para formar cromossomos), enquanto que o DNA dos procariotos está livre de tais proteínas estruturais [26].

Neste trabalho, utilizaremos **sequências de DNA** presentes em células eucarióticas e procarióticas, identificadas na Tabela 1.1 do Capítulo 1, que, serão reproduzidas através de CCEs. No processo da modelagem matemática, analisaremos apenas a estrutura primária dessas sequências de DNA.

As moléculas de uma célula

Em [26] e [27], é explorada a forma como todas as propriedades da célula surgem dos eventos moleculares subjacentes: a montagem de grandes moléculas e suas interligações, os efeitos catalíticos que promovem determinadas reações químicas e a distribuição da informação transportada pelas moléculas gigantes.

A seguir, revisamos os mais importantes tipos de moléculas que são o fundamento químico

do funcionamento e da estrutura da célula.

1. Moléculas pequenas transportam energia, transmitem sinais e se ligam para criar macromoléculas

Muito do conteúdo celular consiste de um meio aquosa adicionada de pequenas moléculas (por exemplo, açúcares simples, aminoácidos, vitaminas) e íons (por exemplo, sódio, cloreto e íons cálcio). O posicionamento e as concentrações das moléculas pequenas e dos íons no interior das células são controlados por numerosas proteínas inseridas nas membranas celulares. Essas bombas, transportadores e canais iônicos transportam praticamente todas as pequenas moléculas e íons para o interior ou exterior das células e de suas organelas.

Uma das mais conhecidas entre as moléculas pequenas é o **trifosfato de adenosina (ATP)**, que estoca energia química facilmente disponível em duas de suas ligações químicas (ver Figura 2.2). Quando as células quebram essas ligações ATP, ricas em energia, a energia liberada pode ser utilizada em um processo que necessite de energia, como a contração muscular ou a biossíntese de proteínas.

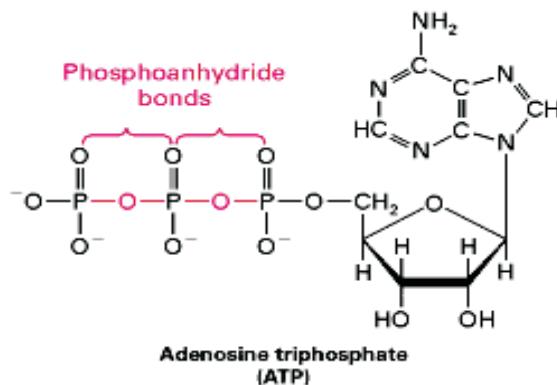


Figura 2.2: Trifosfato de adenosina (ATP). Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

Para obter a energia necessária para a síntese de ATP, as células quebram as moléculas dos alimentos. Por exemplo, quando o açúcar é degradado em dióxido de carbono e água, a energia estocada nas ligações químicas originais é liberada e grande parte dessa energia pode ser “capturada” sob a forma de ATP. As células bacterianas, das plantas ou dos animais, podem produzir ATP por esse processo; papel exercido pelas **mitocôndrias**. Além disso, as plantas e alguns poucos outros organismos podem obter energia a partir da luz solar, formando ATP pela fotossíntese; papel que é exercido pelos plastídeos, mais precisamente os **cloroplastos**.

Certas moléculas pequenas, **monômeros**, presentes na célula podem se unir para a formação de **polímeros** (qualquer molécula grande composta de unidades, monômeros), pela repetição de um único tipo de reação de ligação química (ver Figura 2.3). As células produzem três tipos de grandes polímeros, denominados **macromoléculas**: os **polissacarídeos**, as **proteínas** e os **ácidos nucléicos**.

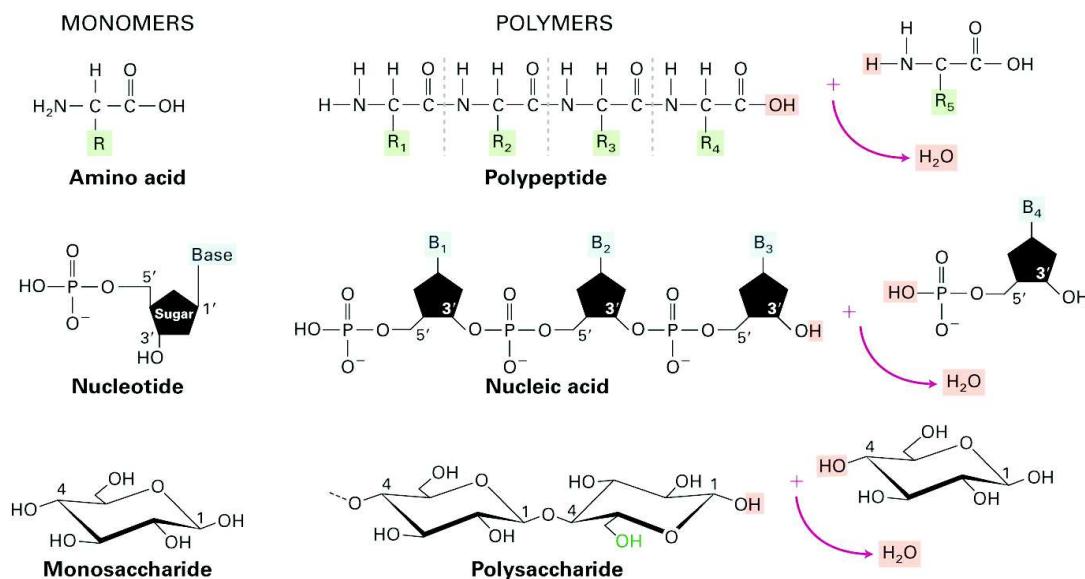


Figura 2.3: Visão geral dos blocos químicos que constituem as células e as macroestruturas que as formam. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

As proteínas são polímeros lineares contendo de dezenas a vários milhares de aminoácidos ligados por **ligação peptídica**. Os ácidos nucléicos são polímeros lineares que contêm de centenas a milhões de nucleotídeos ligados por **ligações fosfodiéster**. Os polissacarídeos são polímeros lineares ou ramificados de monossacarídeos (açúcares), como a glicose, ligados por meio de **ligações glicosídicas**.

2. As proteínas dão estrutura às células e realizam grande parte das tarefas celulares

As diversidades e intrincadas estruturas das proteínas possibilitam que estas desempenhem muitas atividades. As células encadeiam linearmente 20 diferentes **aminoácidos** para formar uma proteína. As proteínas, normalmente, têm um comprimento que varia de 100 a 1000 aminoácidos, podendo, no entanto, ser menores ou maiores do que esses limites. Obtemos aminoácidos através de sua síntese a partir de outras moléculas, ou pela degradação das proteínas que ingerimos. Uma vez que uma cadeia de aminoácidos

tenha sido formada, ela se dobra em uma conformação complexa, conferindo uma estrutura tridimensional característica e uma função determinada a cada proteína.

Algumas proteínas apresentam semelhanças entre si, podendo, portanto, ser classificadas em **famílias protéicas**. Algumas centenas dessas famílias já foram identificadas. A maior parte das proteínas foi projetada para atuar em determinados locais dentro das células ou para ser liberada no espaço extracelular. Vias celulares elaboradas asseguram que as proteínas sejam transportadas para suas localizações intracelulares adequadas ou que sejam secretadas.

As proteínas podem servir como componentes estruturais para as células, por exemplo, formando um esqueleto interno. Podem atuar como sensores que alteram sua forma quando ocorrem mudanças na temperatura, na concentração iônica ou em outras propriedades celulares. As proteínas podem ainda importar ou exportar substâncias através da membrana plasmática. Elas podem ser **enzimas** que provocam determinadas reações químicas mais rapidamente do que seria possível sem o auxílio dessas proteínas **catalisadoras**. Elas podem se ligar a genes específicos, ativando-os ou desligando-os. Elas podem ser sinalizadores extracelulares, liberados por uma célula para a comunicação com outras células ou sinalizadores intracelulares, transportando informações dentro das células. Elas podem ser motores que movimentam outras moléculas, queimando energia química (ATP).

Uma pergunta que surge naturalmente: como podem 20 aminoácidos formar todas as diferentes proteínas necessárias para a realização de tantas tarefas diferentes? Em um primeiro momento, isso parece impossível. No entanto, se uma proteína “padrão” é composta de aproximadamente 400 aminoácidos, existem 20^{400} possíveis sequências protéicas diferentes. Mesmo considerando-se que muitas dessas proteínas sejam funcionalmente equivalentes, instáveis ou por algum motivo descartáveis, o número possível de proteínas ainda é muito grande.

A função das células

Em essência, qualquer célula é simplesmente um compartimento com um interior aquoso que está separado do ambiente externo por uma membrana de superfície (a membrana plasmática), a qual evita que ocorra o livre fluxo de moléculas entre a parte externa e interna da mesma. Além disso, como já foi salientado, as células dos eucariotos possuem uma grande quantidade de membranas internas que as subdividem em vários compartimentos, denominados organelas [26]. A membrana plasmática e outras membranas celulares são compostas principalmente de duas camadas de moléculas **fosfolipídicas**. Estas moléculas

apresentam duas porções, uma extremidade de “atração à água” (**hidrofílica**) e uma extremidade de “repulsão” à água (**hidrofóbica**). As duas camadas fosfolipídicas da membrana estão orientadas de tal forma que as extremidades hidrofílicas se direcionam para o exterior da bicamada, ao passo que as extremidades hidrofóbicas encontram-se escondidas no interior da bicamada. Quantidades menores de outros lipídeos, como o colesterol e diversos tipos de proteínas, encontram-se intercaladas na estrutura fosfolipídica. As moléculas de lipídeos e algumas proteínas podem movimentar-se lateralmente sobre o plano da membrana, dando a esta uma característica fluida. Essa fluidez permite que as células alterem sua conformação e, inclusive, que se movimentem.

O citosol e os espaços internos das organelas diferem tanto entre si quanto diferem do exterior da célula em termos de acidez, composição iônica e conteúdo protéico. As funções específicas e os microambientes dos diferentes compartimentos celulares são determinados, em grande parte, pelas proteínas que residem em suas membranas e em seu interior. Uma grande parte do trabalho celular é desenvolvido por maquinarias moleculares, algumas residentes no citosol e outras nas diversas organelas.

1. As células constroem e degradam um grande número de moléculas e estruturas

As células produzem um número enorme de moléculas complexas a partir de blocos de construção químicos mais simples. Todo esse trabalho de síntese é propelido pela energia química extraída principalmente de açúcares e gorduras ou da luz solar, no caso das células vegetais, e é estocado principalmente sob a forma de ATP, a “moeda” universal de energia química. Em células animais e vegetais, a maior parte do ATP é produzida por grandes máquinas moleculares localizadas em duas organelas, as **mitocôndrias** e os **cloroplastos**.

Acredita-se que as mitocôndrias e os cloroplastos tenham-se originado como bactérias que se estabeleceram no interior das células eucarióticas e tornaram-se colaboradoras bem-vindas.

As células necessitam quebrar matérias inaprováveis ou obsoletas em moléculas pequenas que possam ser descartadas ou recicladas. Essa tarefa de limpeza da casa é função dos **lisossomos**, organelas preenchidas por enzimas de degradação. Os lisossomos são auxiliados no trabalho de limpeza da célula pelos **peroxissomos**. Estas pequenas organelas são especializadas na degradação dos componentes lipídicos das membranas e na inativação de diversas toxinas.

A maior parte das propriedades funcionais e estruturais das células depende das proteínas. Assim, para que as células funcionem adequadamente, as diversas proteínas que compõem

os vários compartimentos de trabalho devem ser transportadas do local onde são fabricadas para sua posição apropriada. Algumas proteínas são produzidas em ribossomos que se encontram livres no citosol. As proteínas secretadas e a maioria das proteínas de membrana, no entanto, são produzidas em ribossomos associados ao **retículo endoplasmático (RE)**. Esta organela produz, processa e promove o transporte das proteínas e dos lipídeos. As cadeias protéicas produzidas no RE são transportadas para o **complexo de Golgi**, onde serão ainda modificadas antes de serem encaminhadas para o seu destino final. As proteínas que utilizam essa rota contêm pequenas sequências de aminoácidos ou de cadeias de açúcares (oligossacarídeos) que os direcionam ao destino correto. Esses sinalizadores funcionam porque são reconhecidos e ligam-se a outras proteínas que os selecionam e reencaminham para os diversos compartimentos celulares.

No presente trabalho, sequências de DNA que possuem tais características serão utilizadas na modelagem.

2. As células alteram suas formas e se movimentam

As células geralmente se apresentam sob formas mais elaboradas devido ao seu esqueleto interno e conexões externas. Nas células animais, o **citoesqueleto** interno é formado por três tipos de filamentos protéicos, organizados em redes e feixes. O citoesqueleto evita que a membrana plasmática das células animais assuma uma conformação relaxada esférica e também participa da locomoção celular e no transporte intracelular de vesículas, cromossomos e macromoléculas. O citoesqueleto pode ser conectado pela superfície celular à matriz extracelular ou ao citoesqueleto de outras células, auxiliando, dessa maneira, a formação dos tecidos, [26] e [27].

A locomoção celular é usada durante o período de desenvolvimento embrionário dos animais multicelulares para dar forma aos tecidos e durante a vida adulta para a defesa contra infecções, para o transporte de nutrientes e para o reparo e a cicatrização de lesões. Esse processo não é utilizado no crescimento e desenvolvimento das plantas multicelulares, pois as novas células vegetais são geradas pela divisão de células preexistentes que compartilham paredes celulares. Desse modo, o desenvolvimento vegetal envolve o aumento do tamanho celular, mas não o movimento das células de uma posição para outra, [27].

3. As células recebem e emitem informações

Uma célula viva monitora continuamente sua vizinhança e ajusta suas atividades e sua composição de acordo com a necessidade. As células também se comunicam por meio do

envio deliberado de sinais que podem ser recebidos e interpretados por outras células. Tais sinais são comuns não apenas no interior de um indivíduo, mas também entre diferentes indivíduos. Os sinais utilizados pelas células incluem pequenas moléculas químicas simples, gases, proteínas, luz e movimentos mecânicos. As células têm numerosas proteínas receptoras (para a detecção de sinais) e rotas elaboradas para a transmissão desses sinais em seu interior, para provocar uma resposta. Em um momento determinado, uma célula pode ser capaz de perceber apenas alguns dos sinais que existem ao seu redor e sua resposta pode mudar, dependendo do momento. Em alguns casos, a recepção de um primeiro sinal indicará à célula o caminho específico a ser seguido em resposta a um sinal diferente subsequente.

Tanto as alterações no ambiente (por exemplo, um aumento ou diminuição nos níveis de um nutriente em particular ou nos padrões de luminosidade) quanto os sinais enviados por outras células representam informações externas que as células devem processar. As respostas mais rápidas a tais sinais incluem alterações na localização ou na atividade de proteínas preexistentes.

A capacidade das células de emitir e responder a sinais é essencial para o desenvolvimento. Muitos sinais importantes para o desenvolvimento são compostos de proteínas secretadas por células específicas em locais e momentos específicos do organismo em desenvolvimento. Frequentemente, uma célula receptora deve integrar múltiplos sinais para determinar o comportamento a ser seguido, por exemplo, para diferenciar-se em um tecido em particular, dar continuidade a um processo, morrer, enviar um sinal de confirmação ou migrar. Sem dúvida, a sinalização e a transdução de sinais são atividades primordiais para as células, [26] e [27].

4. As células crescem e se dividem

A característica mais impressionante das células e do organismo como um todo é a sua capacidade de reprodução. Os sistemas mais simples de reprodução envolvem a divisão de uma célula “parental” em duas células “filhas”. Isso ocorre como parte do **ciclo celular**, uma série de eventos que preparam a célula para a divisão, seguida de uma etapa que se caracteriza realmente como o processo de divisão, denominada de **mitose**. O ciclo celular eucarioto geralmente é representado como composto por quatro estágios, como ilustrado na Figura 2.4.

Os cromossomos e o DNA que eles carregam são copiados durante a **fase S (síntese)**. Os cromossomos replicados são separados durante a **fase M (mitótica)** na qual, durante a divisão celular, cada célula-filha receberá uma cópia de cada um dos cromossomos. As fases M e S são separadas por duas lacunas, a **fase G₁** e a **fase G₂**, que são

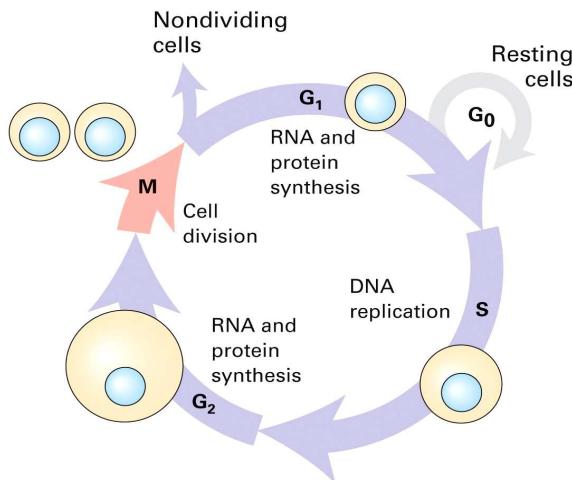


Figura 2.4: Durante o crescimento as células avançam continuamente pelos quatro estágios dos ciclo celular. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

as estapas durante as quais os mRNAs e proteínas são produzidos.

Nos organismos unicelulares, frequentemente (mas nem sempre), as duas células-filhas se assemelham à célula parental. Nos organismos multicelulares, as **células-tronco** podem dar origem a dois diferentes tipos de células, um semelhante à célula parental e outro não. Esse tipo de divisão assimétrica é essencial para a geração de diferentes tipos celulares no organismo. A maioria das células eucarióticas precisa de um tempo consideravelmente maior do que as bactérias, para crescer e se dividir. Além disso, o ciclo celular em plantas e animais adultos encontra-se normalmente sob intensa regulação. Esse forte controle evita o crescimento excessivo e desequilibrado dos tecidos, ao mesmo tempo que assegura que as células danificadas ou com problemas sejam substituídas, e que células adicionais sejam formadas em resposta a novas circunstâncias ou necessidades do desenvolvimento.

A mitose é um processo **assexual**, uma vez que as células-filhas possuirão exatamente a mesma informação genética que a célula parental. Na reprodução **sexual**, a fusão de duas células produz uma terceira, que contém informação genética proveniente de ambas as células parentais. Como essas fusões celulares acarretariam um aumento progressivo no número de cromossomos, o ciclo reprodutivo sexual emprega um tipo especial de divisão celular, denominado **meiose**, que reduz o número de cromossomos na preparação para a posterior fusão. As células que contêm um conjunto completo de cromossomos são chamadas de **diplóides**. Durante a meiose, uma célula diplóide replica seus cromossomos da mesma forma que o faria na mitose mas, a seguir sofre

duas divisões sem que ocorra, neste período intermediário, nova cópia dos cromossomos. Cada uma das células-filha resultantes, as quais possuem apenas a metade do número completo de cromossomos, é chamada de **haplóide**, [27].

5. As células morrem em consequência de agressões ou devido a um programa interno

Quando as células dos organismos multicelulares sofrem lesões muito sérias ou são infectadas por vírus, elas morrem. A morte celular resultante de tais eventos traumáticos é desorganizada e geralmente acompanhada da liberação de constituintes celulares potencialmente tóxicos que podem, por sua vez, lesar as células adjacentes. As células também podem morrer se não receberem os sinais de manutenção de vida ou se receberem a sinalização da morte. Neste tipo de morte celular programada, chamada de **apoptose**, uma célula que está morrendo produzativamente as proteínas necessárias para sua auto-destruição. A morte por apoptose evita a liberação de constituintes celulares potencialmente tóxicos. A morte celular programada é essencial para o desenvolvimento e o funcionamento adequados de nosso organismo.

2.1.2 Estrutura e função das proteínas

A vida está intimamente ligada às proteínas - somos seres protéicos. Estas moléculas especiais realizam as mais variadas funções no nosso organismo, desde o transporte de nutrientes e metabólitos à catálise de reações biológicas. Apesar da complexidade de suas funções, as proteínas são relativamente simples: repetições de 20 unidades básicas, os aminoácidos. A versatilidade extraordinária das proteínas é incrível e muito complexa, podendo estas funcionar como máquinas moleculares, comutadores, catalisadores celulares e componentes das estruturas celulares, Figura 2.5.

Estrutura hierárquica das proteínas

As proteínas podem ter 4 tipos de estrutura, dependendo do tipo de aminoácidos que possuem, do tamanho da cadeia polipeptídica e da sua configuração espacial. As estruturas são:

- **Estrutura primária** - É dada pela sequência de aminoácidos ao longo da cadeia polipeptídica, Figura 2.6. É o nível estrutural mais simples e mais importante, pois dele deriva todo o arranjo espacial da molécula. São específicas para cada proteína, sendo geralmente determinados geneticamente. A estrutura primária da proteína resulta em

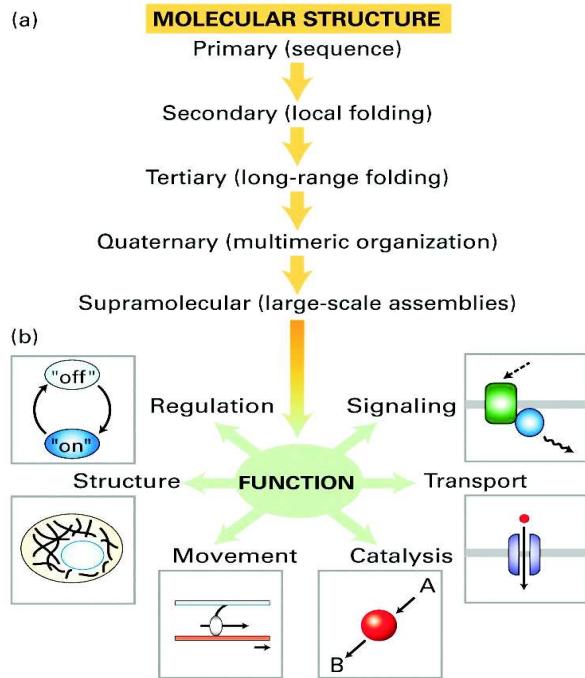


Figura 2.5: Visão geral da estrutura e função das proteínas. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

uma longa cadeia de aminoácidos semelhante a um "colar de contas". Uma extremidade da proteína possui um grupo amino livre (**N-terminal**) e a outra extremidade possui um grupo carboxila livre (**C-terminal**). A sequência de uma cadeia protéica é, por convenção, escrita com a extremidade N-terminal à esquerda, e a C-terminal à direita. Sua estrutura é somente a sequência dos aminoácidos, sem se preocupar com a orientação espacial da molécula.

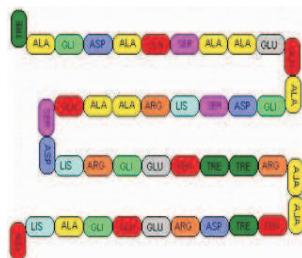
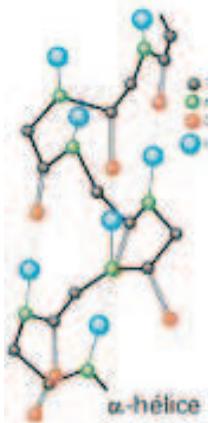
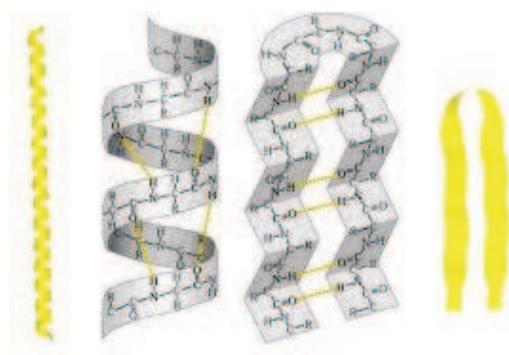


Figura 2.6: Estrutura primária.

- **Estrutura secundária** - É dada pelo arranjo espacial de aminoácidos próximos entre si na sequência primária da proteína. É o último nível de organização das proteínas

fibrosas mais simples estruturalmente. Ocorre graças à possibilidade de rotação das ligações entre os carbonos e dos aminoácidos e seus grupamentos amina e carboxila. O arranjo secundário de um polipeptídeo pode ocorrer de forma regular; isso acontece quando os ângulos das ligações entre carbonos e seus ligantes são iguais e se repetem ao longo do segmento da molécula. São dois os tipos principais de arranjo secundário regular: **hélice α** (Figura 2.7) e **folha β** (Figura 2.8), ou uma **volta** curta em forma de U.

Figura 2.7: Hélice α .Figura 2.8: Folha β .

Em uma proteína média, 60% da cadeia polipeptídica apresenta-se como hélices α e folhas β pregueadas; o restante da molécula apresenta-se em espirais aleatórias e voltas. Portanto, as hélices α e folhas β pregueadas são os principais elementos internos de suporte das proteínas.

- **Estrutura terciária** - Resulta do enrolamento da hélice α ou da folha pregueada, sendo mantido por pontes de hidrogênio e dissulfeto. Esta estrutura confere a atividade biológica às proteínas. A estrutura terciária descreve o dobramento final de uma cadeia, por interações de regiões com estrutura regular ou de regiões sem estrutura definida, Figura 2.9, podendo haver interações de segmentos distantes de estrutura primária, por ligações não covalentes. Enquanto a estrutura secundária é determinada pelo relacionamento estrutural de curta distância, a terciária é caracterizada pelas interações de longa distância entre aminoácidos. Todas têm sequências de aminoácidos diferentes, refletindo estruturas e funções distintas.
- **Estrutura quaternária** - Também conhecida como o quarto nível de organização estrutural, descreve o número (esquiometria) e as posições relativas das subunidades

nas proteínas multiméricas. As proteínas multiméricas consistem de dois ou mais polipeptídeos ou subunidades. Elas são guiadas e estabilizadas pelas mesmas interações da terciária. A junção de cadeias polipeptídicas pode produzir diferentes funções para os compostos. Um dos principais exemplos de estrutura quaternária é a hemoglobina, Figura 2.10. Sua estrutura é formada por quatro cadeias polipeptídicas.



Figura 2.9: Estrutura terciária.

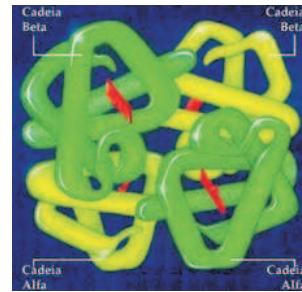


Figura 2.10: Estrutura quaternária.

A organização **espacial** da proteína - sua forma tridimensional - é a chave para entender sua função. Somente quando uma proteína está na estrutura tridimensional correta, ou **conformação**, é capaz de funcionar de modo eficiente. Um conceito importante para o entendimento das proteínas é que *a função é derivada da estrutura tridimensional, e a estrutura tridimensional é especificada pela sequência de aminoácidos*.

A incorporação de proteínas diferentes, durante a evolução, produziu a diversidade de estrutura e função das proteínas. As células contêm grandes arranjos macromoleculares nos quais todos os elementos necessários em um processo celular complexo (por exemplo, síntese de DNA, RNA e de proteínas; fotossíntese; transdução de sinais) estão integrados, formando máquinas moleculares.

Função das proteínas

- **Estrutural ou plástica** - São aquelas que participam dos tecidos dando-lhes rigidez, consistência e elasticidade. São proteínas estruturais: colágeno (constituínte das cartilagens), actina e miosina (presentes na formação das fibras musculares), queratina (principal proteína do cabelo), fibrinogênio (presente no sangue), albumina (encontrada em ovos) e outras;
- **Hormonal** - Exercem alguma função específica sobre algum órgão ou estrutura de um organismo como, por exemplo, a insulina (embora tecnicamente a insulina seja considerada apenas um polipeptídeo, devido a seu pequeno tamanho);

- **Defesa** - Os anticorpos são proteínas que realizam a defesa do organismo, especializados no reconhecimento e neutralização de vírus, bactérias e outras substâncias estranhas. O fibrinogênio e a trombina são outras proteínas responsáveis pela coagulação do sangue e prevenção de perda sanguínea em casos de cortes e machucados;
- **Energética** - Obtenção de energia a partir dos aminoácidos que compõem as proteínas;
- **Enzimática** - Enzimas são proteínas capazes de catalizar reações bioquímicas como, por exemplo, as lípases. As enzimas não reagem, são reutilizadas (sempre respeitando o sítio ativo) e são específicas. As enzimas reduzem a energia de ativação das reações químicas. A função da enzima depende diretamente de sua estrutura. Proteínas altamente especializadas e com atividade catalítica. Mais de 2000 enzimas são conhecidas, cada uma capaz de catalisar um tipo diferente de reação química;
- **Condutoras de gases** - O transporte de gases (principalmente do oxigênio e um pouco do gás carbônico) é realizado por proteínas como a hemoglobina e hemocianina. Proteínas trans-membrânicas são responsáveis pelo transporte de substâncias para dentro e fora da célula através da membrana de composição fosfolipídica e isolante.

É importante salientar que, neste trabalho, usaremos apenas o conteúdo de informação das sequências de DNA, ou seja, a sequência primária de nucleotídeos e de aminoácidos dessas sequências (genoma, gene, DNA repetitivo, ítron, RNA, mRNA, proteínas, hormônio, SD e SI), sendo que, através do processo de modelagem, reproduziremos tais sequências. Agora, a determinação das demais estruturas da proteína requerirá o conhecimento de propriedades e conceitos no contexto bioquímico e matemático.

Dobramento, modificação e degradação das proteínas

Uma cadeia polipeptídica é sintetizada por um processo chamado **tradução**, no qual a polimerização dos aminoácidos em uma sequência específica é determinada pelo **RNA mensageiro (mRNA - messenger RNA)**. A célula promove o dobramento adequado de uma cadeia polipeptídica nascente e, em muitos casos, modifica os resíduos, ou cliva o esqueleto peptídico para produzir uma proteína final. Além disso, a célula tem mecanismos de verificação de erros, que eliminam proteínas sintetizadas ou dobradas incorretamente. As proteínas dobradas incorretamente geralmente não possuem atividade biológica e, em alguns casos, podem estar associadas a doenças. O dobramento incorreto das proteínas é evitado por dois mecanismos distintos. Primeiro, as células têm sistemas que reduzem as chances de formar proteínas mal dobradas. Segundo, qualquer proteína incorretamente dobrada, bem

como as proteínas citosólicas que não são mais necessárias à célula, são degradadas por um sistema celular especializado de descarte.

- **A informação para o dobramento de uma proteína está codificada na sequência**

A sequência de aminoácidos de uma proteína determina seu dobramento em uma conformação tridimensional específica chamada de **estado nativo**; para a grande maioria das proteínas, o estado nativo é a forma mais estável de dobramento da molécula.

- **O dobramento *in Vivo* das proteínas é promovido por chaperonas**

Apesar de o dobramento de proteínas ocorrer *in vitro*, apenas uma minoria das moléculas **desnaturadas** (tratamento laboratorial que faz com que a proteína perca sua conformação nativa e sua função biológica) sofre o dobramento completo e volta à conformação nativa em alguns minutos. Obviamente, as células necessitam de um mecanismo mais rápido e eficiente para o dobramento das proteínas nas formas corretas, pois de outro modo as células gastariam muita energia na síntese das proteínas não-funcionais e na degradação de proteínas não-dobradas ou dobradas incorretamente.

A explicação para a incrível eficiência das células em promover o dobramento correto das proteínas está, provavelmente, nas **chaperonas**, uma classe de proteínas encontradas em todos os organismos, de bactérias a humanos. As chaperonas estão localizadas em todos os compartimentos celulares, se ligam a uma enorme variedade de proteínas e atuam no mecanismo geral de dobramento de proteínas. Duas famílias gerais de chaperonas são conhecidas:

- **Chaperonas moleculares**, que ligam e estabilizam proteínas não-dobradas ou com dobramento parcial, evitando sua agregação e degradação. O dobramento de proteínas *in vivo* ocorre com o auxílio de chaperonas moleculares (proteínas Hsp70), que se ligam ao polipeptídeo nascente que emerge dos ribossomos, impedindo seu dobramento incorreto (Figura 2.11).
- **Chaperoninas**, que atuam diretamente, facilitando o dobramento das proteínas. As chaperoninas, grandes complexos de proteínas semelhantes ao Hsp60, protegem algumas proteínas parcialmente dobradas ou com dobramentos incorretos em uma cavidade em forma de barril, fornecendo um tempo adicional para que ocorra o dobramento correto.

Algumas doenças neurodegenerativas, incluindo a doença de *Alzheimer* e de *Parkinson* em humanos e a encefalopatia espongiforme transmissível (doença da “vaca louca”)

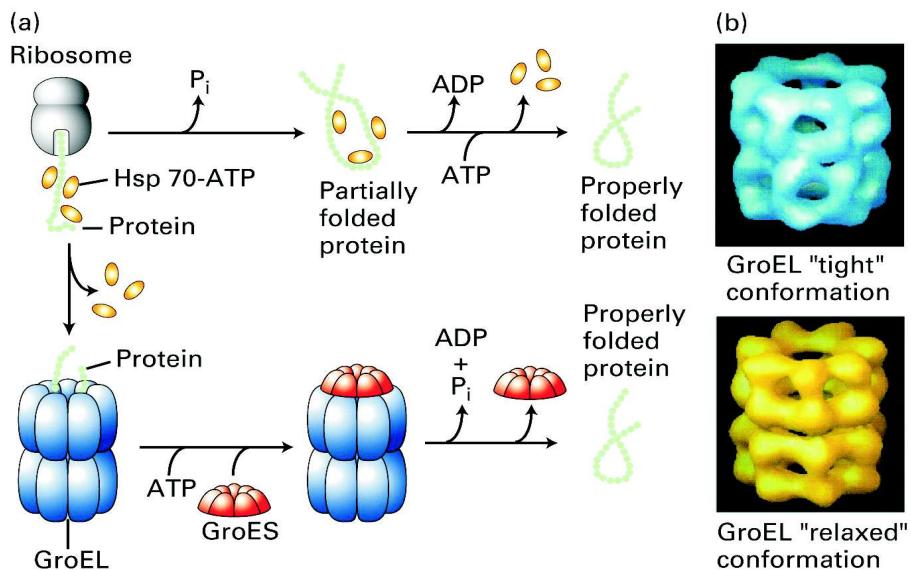


Figura 2.11: Dobramento de proteínas mediado por chaperonas e chaperoninas. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

em bovinos e ovinos, são causadas por agregados protéicos que sofrem dobramentos estáveis em uma conformação alternativa.

2.1.3 O DNA e a síntese de proteínas

Nesta seção resumimos inicialmente a estrutura básica e as propriedades do DNA e do RNA. Abreviadamente, apresentamos os processos básicos resumidos na Figura 2.12: transcrição do DNA em precursores do RNA, processamento desses precursores em moléculas funcionais de RNA, tradução de mRNAs em proteínas e replicação do DNA.

A determinação da estrutura do DNA por James Watson e Francis Crick em 1953, e a subsequente elucidação de como o DNA dirige a síntese de RNA, que, então, orientará a montagem das proteínas - o chamado **dogma central** - foi um evento monumental, que marcou o início da era da biologia molecular. No entanto, a representação simplificada do dogma central sob a forma de DNA → RNA → proteína não reflete a função das proteínas dos ácidos nucléicos. Além disso, as proteínas estão muito envolvidas na **regulação** da **expressão gênica**, o processo completo pelo qual a informação codificada no DNA é decodificada nas proteínas que caracterizam os diversos tipos de células.

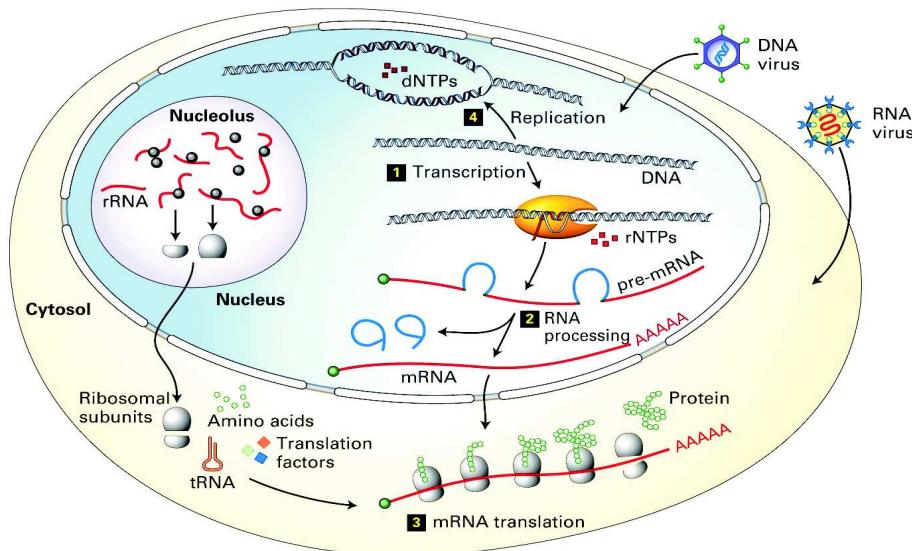


Figura 2.12: Visão geral dos quatro processos básicos da genética molecular. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

A estrutura dos ácidos nucléicos

Os **ácidos nucléicos** são macromoléculas que contêm a informação que determina a sequência de aminoácidos e, consequentemente, a estrutura e a função de todas as proteínas de uma célula, as quais fazem parte das estruturas celulares e alinham os aminoácidos de forma correta, quando uma cadeia polipeptídica está sendo sintetizada e catalisam uma série de reações químicas fundamentais nas células, inclusive a formação das pontes peptídicas entre os aminoácidos, durante a síntese de proteínas.

Os ácidos nucléicos contêm as informações para a produção das proteínas no local e momento adequados. A informação referente a como, quando e onde deve ser produzido cada tipo de proteína está contida no material genético. Dois tipos de ácidos nucléicos quimicamente semelhantes, o **DNA** (ácido desoxirribonucléico) e o **RNA** (ácido ribonucléico), são as principais moléculas carreadoras de informação das células. Os monômeros que formam o DNA e o RNA, denominados **nucleotídeos**, têm uma estrutura comum: um grupo fosfato ligado por uma ligação fosfodiéster a uma pentose (uma molécula de açúcar com cinco carbonos) que, por sua vez, está ligada a um anel cuja estrutura contém nitrogênio e carbono, a qual normalmente é conhecida como “base” (Figura 2.13a). No RNA, a pentose é a ribose, e no DNA, é a desoxirribose (Figura 2.13b).

As bases **adenina**, **guanina** e **citosina** são encontradas tanto no DNA como no RNA. A **timina** é encontrada apenas no DNA e a **uracila**, apenas no RNA. A adenina e a guanina

são **purinas**, contêm um par de anéis fusionados. A citosina, a timina e a uracila são **pirimidinas**, contêm um anel único (Figura 2.14). As bases são geralmente abreviadas por A, G, C, T e U, respectivamente.

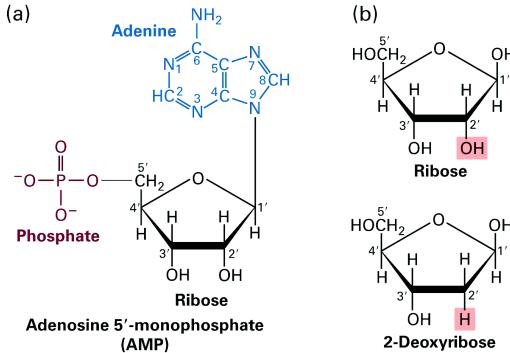


Figura 2.13: Estrutura geral dos nucleotídeos.

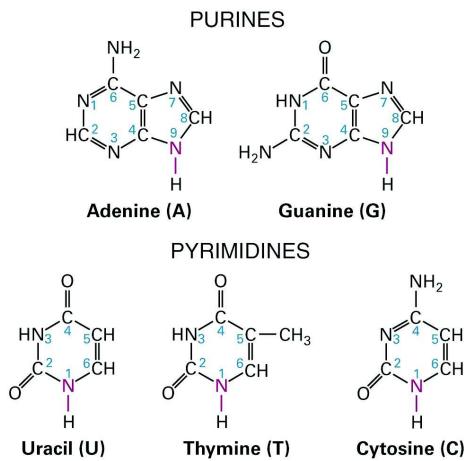


Figura 2.14: Estrutura química das principais bases dos ácidos nucléicos.

O modelo de Watson e Crick possui as seguintes características principais:

1. Duas cadeias polinucleotídicas circundam um eixo comum formando a **dupla hélice** (Figura 2.15).
2. As duas fitas de DNA são **antiparalelas** (possuem direções opostas), mas cada uma forma uma hélice para o lado direito.
3. As bases ocupam o centro da hélice, e as cadeias de açúcar-fosfato estão dispostas na periferia, minimizando a repulsão entre os grupos fosfato carregados. A superfície da dupla hélice forma dois sulcos de largura desigual: a **cavidade maior** e a **cavidade menor**.
4. Cada base está ligada a uma base da fita oposta por meio de pontes de hidrogênio, formando um **par de base** planar. A estrutura de Watson e Crick pode acomodar apenas dois tipos de pares de base. Cada resíduo de adenina deve formar o par com um resíduo de timina e vice-versa, e cada resíduo de guanina deve formar par com um resíduo de citosina e vice-versa (Figura 2.15 b)). Essas interações por pontes de hidrogênio, um fenômeno denominado como **pareamento das bases complementares**, resulta na associação específica das duas cadeias da fita dupla.

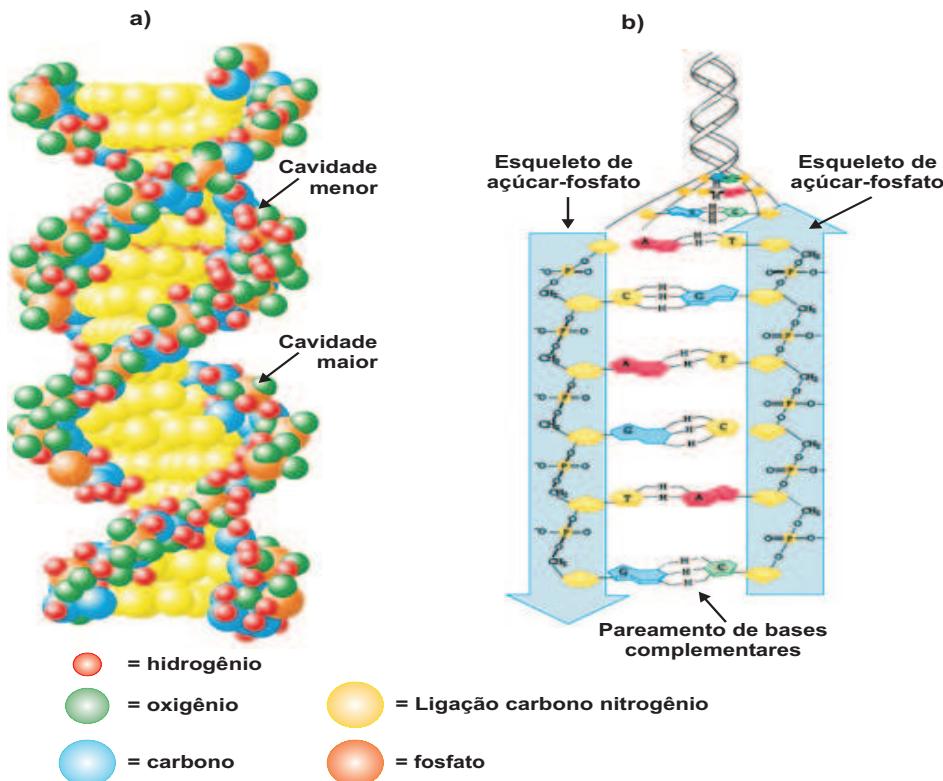


Figura 2.15: Estrutura tridimensional do DNA e sua complementaridade.

A estrutura de Watson e Crick poderá acomodar qualquer sequência de bases em uma fita polinucleotídica se a fita oposta possuir a sequência de bases complementares a ela. Isso explica a regra de Chargaff. Mais importante ainda, sugere que cada fita de DNA pode atuar como um **molde** para a síntese de sua fita complementar e, consequentemente, a informação hereditária está codificada na sequência de bases em qualquer fita.

A maioria das moléculas de DNA é extremamente grande, de acordo com sua função de conter toda a informação genética da célula. Com raras exceções, os organismos mais complexos contêm mais DNA. O **genoma** de um organismo, que é seu conteúdo específico de DNA, pode estar distribuído em diversos **cromossomos** (do grego, *chromos*, cor + *soma*, corpo), cada um contendo uma molécula de DNA separada.

Devido a seu comprimento muito longo, as moléculas de DNA são descritas em termos do número de pares de bases (**pb**) por milhares de pares de bases (**quilobases em pares** ou **kb**). Apesar de cada molécula de DNA ser longa e relativamente firme, ela não é completamente rígida. A dupla hélice de DNA forma espirais e voltas quando compactada dentro da célula. Além disso, dependendo da sequência de nucleotídeos, o DNA pode adotar conformações helicoidais levemente distintas. Por fim, na presença de outros componentes celulares, o

DNA pode dobrar-se ou suas duas fitas podem ser parcialmente desenroladas.

A dupla hélice existe em várias geometrias designadas como DNA A, DNA B, DNA C e DNA Z. A formação dessas diferentes conformações depende da composição em bases do DNA e das condições físicas. O modelo descrito por Watson e Crick possui a conformação do DNA B.

A síntese protéica

A informação genética contida no DNA reside em sua sequência, a ordenação linear dos nucleotídeos ao longo da fita. A parcela de DNA que contém informações encontra-se dividida em unidades funcionais separadas, os **genes**, que, em geral, têm um tamanho entre 5 mil e 100 mil nucleotídeos. A maioria das bactérias possui poucos milhares de genes, enquanto os humanos têm, aproximadamente, 40 mil genes. Os genes que contêm instruções para o preparo das proteínas normalmente podem ser divididos em duas partes: a **região codificante**, que determina a sequência de aminoácidos da proteína, e a **região reguladora**, que controla quando e em que tipos de células essa proteína será produzida.

As células fazem uso de dois processos em série para converter a informação codificada no DNA em proteínas, Figura 2.16.

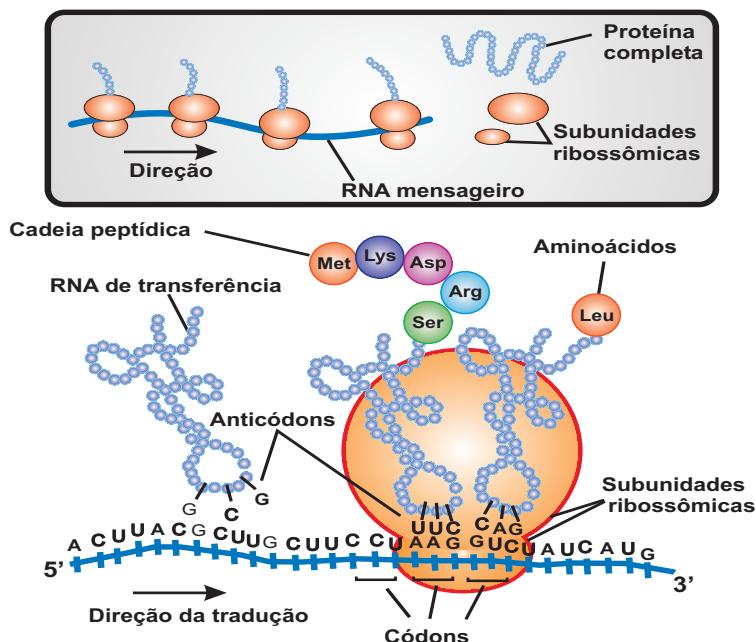


Figura 2.16: A informação codificada no DNA é convertida nas sequências de aminoácidos.

No primeiro, denominado de **transcrição**, a região codificante de um gene é copiada sob a forma de uma versão em fita simples de **ácido ribonucléico (RNA)** a partir da dupla fita de

DNA. Uma grande enzima, a **RNA polimerase**, catalisa a ligação dos nucleotídeos na cadeia de RNA, usando o DNA como molde. Em células eucarióticas, o produto inicial de RNA é processado em uma molécula de **RNA mensageiro (mRNA)** menor, a qual é transportada para o citoplasma. Neste compartimento, o **ribossomo**, uma enorme máquina molecular complexa, composta de RNAs de proteínas, se encarrega de efetuar o segundo processo, denominado **tradução**. Durante a tradução, o ribossomo, organiza e liga os aminoácidos seguindo uma ordem estabelecida, a qual é ditada pela sequência do mRNA, de acordo com um **código genético** praticamente universal.

Todos os organismos têm meios de controlar quando e onde seus genes devem ser transcritos. Na verdade, praticamente todas as células de nosso organismo contêm um conjunto completo dos genes humanos, mas em cada tipo de célula apenas alguns desses genes estão ativos, ou ligados, para a produção de proteínas. Além disso, muitas células podem responder a sinais externos ou a alterações das condições externas ligando ou desligando genes específicos e, dessa forma, adaptando o seu repertório de proteínas às necessidades do momento. Tal controle da atividade gênica depende das proteínas de ligação ao DNA denominadas de **fatores de transcrição**, as quais são capazes de se ligar ao DNA e atuar como disjuntores, ativando ou reprimindo a transcrição de determinados genes.

Os fatores de transcrição são tão precisos que podem se ligar preferencialmente a regiões reguladoras de um número restrito de genes entre os milhares que estão presentes no DNA da célula. Eles frequentemente trabalham como complexos multiprotéicos, nos quais mais de uma proteína contribui com sua especificidade de ligação ao DNA para a seleção dos genes controlados. Em organismos complexos, centenas de diferentes fatores de transcrição são empregados, formando um incrível sistema de controle que ativa os genes corretos, na célula correta, no momento adequado.

2.1.4 Estrutura genômica

No início do século XXI, foi completado o sequenciamento de genomas inteiros de centenas de vírus, de bactérias e da levedura *S.cerevisiae*, um eucarioto unicelular. Além dessas, a maior parte das sequências genômicas de vários eucariotos multicelulares, como o nematódeo *C.elegans*, a mosca-das-frutas *D.melanogaster*, e dos humanos, já são conhecidos. A análise detalhada dos dados obtidos pelo sequenciamento revelou que grande parte dos genomas dos organismos superiores não codificam mRNAs ou qualquer outro RNA necessário ao organismo. Surpreendentemente, este DNA não-codificante constitui mais de 95% do DNA cromossomal humano.

O DNA não-codificante presente nos organismos multicelulares contém diversas regiões

semelhantes, mas não idênticas. As variações nas porções desses **DNAs repetitivos** são tão impressionantes que cada pessoa pode ser distinguida por uma “impressão digital” de DNA baseada nas variações dessas sequências. Além disso, algumas sequências de DNA repetitivas não são encontradas em posições fixas do DNA de indivíduos da mesma espécie. Esses elementos de DNA “móvel”, presentes tanto nos organismos procariotos como nos eucariotos, podem provocar mutações quando se deslocam para outros locais no genoma. Em [27], as regiões de DNA que codificam as proteínas - isso é, os **genes** - localizam-se entre essas áreas do DNA aparentemente não-funcional. Além do DNA não-funcional entre os genes, os **ítrons**, são comumente encontrados nas regiões não-codificantes dentro dos genes dos animais e das plantas multicelulares. Segundo [27], o gene inclui mais do que os nucleotídeos que codificam a sequência dos aminoácidos de uma proteína, chamada **região codificante**. O gene inclui, também, todas as sequências necessárias para a síntese de um determinado transcrito de RNA. O sequenciamento dos genes que codificam a mesma proteína em várias espécies de eucariotos mostrou que a pressão evolutiva favorece a manutenção de sequências relativamente semelhantes nas regiões codificantes, os **éxons**.

A seguir, abordamos resumidamente alguns detalhes da estrutura genômica - em especial a morfologia dos cromossomos, pois, no Capítulo 3, inferimos que o processo de armazenamento da informação genômica pode ocorrer de maneira análoga àquele utilizado em redes de computadores.

1. Organização estrutural dos cromossomos eucarióticos

O comprimento total do DNA celular impõe um problema importante para as células. O DNA de uma única célula humana, com cerca de 2 metros de comprimento total, deve ser acondicionado dentro de uma célula com um diâmetro menor de $10\ \mu m$, em uma proporção de compactação de mais de 10^5 . Proteínas eucarióticas especializadas, associadas ao DNA nuclear, fazem a compactação das estruturas de DNA e proteínas observadas como **cromossomos** individuais durante a mitose. As mitocôndrias e os cloroplastos também apresentam DNAs, provavelmente resquícios evolucionários de suas origens, que codificam componentes essenciais dessas organelas vitais, [26] e [27].

A Figura 2.17 fornece uma visão geral da estrutura dos genes e dos cromossomos. O DNA dos eucariotos superiores consiste de sequências únicas e repetidas. Apenas uns 5% do DNA humano codificam as proteínas e os RNAs funcionais, bem como as sequências reguladoras que controlam sua expressão; o restante é simplesmente DNA entre os genes e ítrons dentro das famílias de genes. A maior parte desse DNA, aproximadamente 50% em humanos, é derivada de elementos de DNA móvel, simbiontes genéticos que contribuíram para a evolução dos genomas contemporâneos. Cada cro-

mossomo consiste de uma única molécula longa de DNA, de até aproximadamente 280 Mb nos humanos, organizadas em níveis crescentes de condensação com as quais está intimamente associada. Moléculas de DNA muito menores estão localizadas nas mitocôndrias e nos cloroplastos.

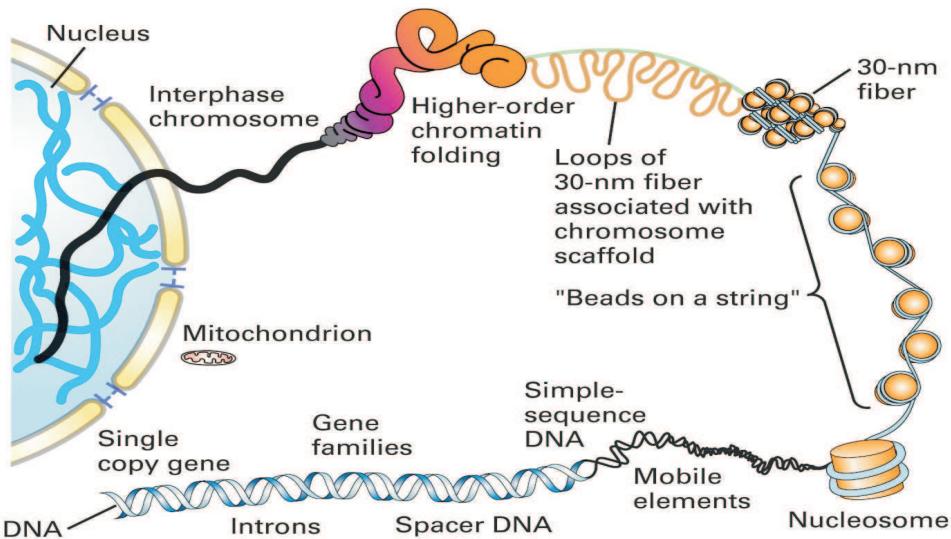


Figura 2.17: Visão geral da estrutura dos genes e dos cromossomos. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

O empacotamento do DNA é fundamental para a arquitetura celular. Durante a interfase, quando as células não estão em divisão, o material genético está na forma de uma nucleoproteína complexa chamada **cromatina**, dispersa por quase todo o núcleo. Como representado esquematicamente na Figura 2.17, nos cromossomos em interfase, longos segmentos de chromatina com 30 nm projetam-se para fora do suporte estendido. O dobramento e a compactação adicional da chromatina, durante a mitose, produz os **cromossomos em metáfase**.

2. O DNA eucariótico é associado a proteínas histonas para formar a chromatina

A estrutura geral da chromatina é incrivelmente semelhante nas células de todos os eucariotos, incluindo fungos, plantas e animais. As proteínas mais abundantes associadas ao DNA eucariótico são as **histonas**, uma família de pequenas proteínas, de caráter básico, presente em todos os núcleos eucarióticos. Segundo [26], as semelhanças entre as sequências das histonas de todos os eucariotos sugere que as suas conformações tridimensionais são bastante similares e foram otimizadas para a sua função em um ancestral comum a todos os eucariotos modernos, no início do processo evolucionário.

3. A cromatina na forma distendida e condensada

A cromatina existe nas formas distendida e condensada. Na forma distendida, o cordão é composto por DNA livre, chamado DNA de ligação (*linker*), que une as estruturas em forma de contas, chamadas de **nucleossomos**. Os nucleossomos, compostos de DNA e histonas, têm um diâmetro de aproximadamente 10 nm e são as unidades estruturais primárias da cromatina. Se a cromatina for isolada em concentrações salinas fisiológicas pré-estabelecida, assumirá uma forma mais condensada, como uma fibra, de 30 nm de diâmetro. Cada uma das proteínas histonas que compõem o centro dos nucleossomos contém uma sequência N-terminal flexível composta por 11 a 37 resíduos, projetando-se para fora da estrutura fixa do nucleossomo; essas extremidades são chamadas **caudas das histonas**. As caudas das histonas são necessárias para a condensação da cromatina, da conformação do “colar de contas” para fibras de 30 nm.

Apesar de as histonas serem as proteínas predominantes nos cromossomos, as proteínas também estão envolvidas na organização da estrutura cromossômica, Figura 2.18.

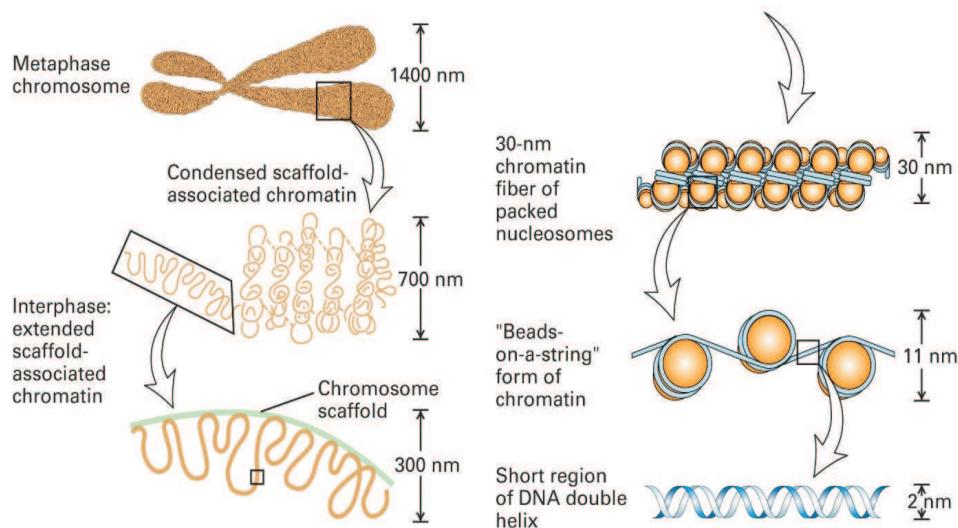


Figura 2.18: Modelo de compactação dos cromossomos eucarióticos. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

As micrografias eletrônicas de cromossomos em metáfase revelaram longas alças de DNA ancoradas a um **suporte cromossômico** composto por proteínas não histônicas. Esse suporte tem a forma de um cromossomo na metáfase e persiste mesmo quando o DNA é digerido por nucleases. Como representado esquematicamente na Figura 2.18, foi proposto pelos geneticistas que as longas alças de fibra de 30 nm da cromatina, com algumas megabases de comprimento se associam a um suporte flexível, produzindo a forma

distendida característica dos **cromossomos da interfase**. Segundo [26], também foi proposto que o dobramento desse suporte produza a estrutura altamente condensada característica dos **cromossomos da metáfase**, cuja geometria exata ainda não foi determinada.

4. Morfologia dos cromossomos eucarióticos

Um método recentemente desenvolvido para a visualização de cada um dos cromossomos humanos em cores vibrantes e distintas, chamado *fish* ou pintura de cromossomos, simplificou enormemente a diferenciação de cromossomos com tamanho e formas, como mostrado na Figura 2.19. Essa técnica, uma variação de hibridização *in situ* por fluorescência, utiliza sondas específicas para os sítios distribuídos ao longo de cada cromossomo.

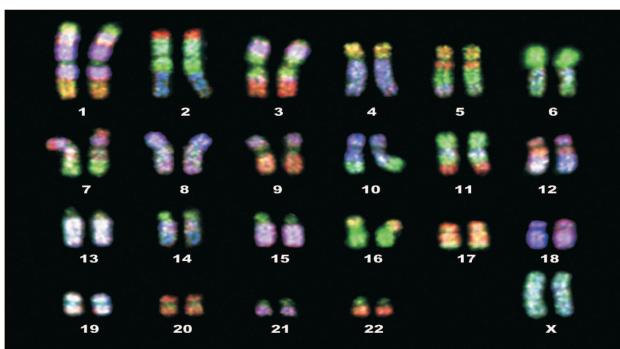


Figura 2.19: Visualização dos cromossomos humanos. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

Os cromossomos individuais das células que não estão em divisão não são visíveis, mesmo com o auxílio de corantes histológicos para DNA (como corantes de *Feulgen* ou *Giemsa*) ou da microscopia eletrônica. Durante a mitose e a meiose, porém, os cromossomos são condensados e tornam-se visíveis ao microscópio óptico. Dessa forma, quase todo trabalho citogenético (ou seja, estudo da morfologia cromossômica) foi realizado em cromossomos em metáfase condensados obtidos de células em divisão - células somáticas, durante a mitose, e gametas, durante a meiose. A espécie humana possui 46 cromossomos, sendo 44 autossomos e 2 sexuais. Todos estes cromossomos encontram-se pareados, logo temos 22 pares de cromossomos autossomos e 1 par de cromossomos sexuais. Em cada espécie há um número definido de cromossomos. Na mitose, as células já passaram pela fase S do ciclo celular e já replicaram todo o seu DNA.

Portanto, os cromossomos visíveis na metáfase são estruturas duplicadas. Cada cromossomo é replicado na metáfase e consiste de duas **cromátides** irmãs unidas pelo

centrômero (Figura 2.20).

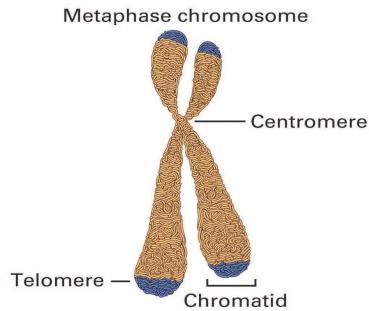


Figura 2.20: Aspecto microscópico de um cromossomo característico em metáfase. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

O número, o tamanho e a forma dos cromossomos durante a metáfase constituem o **cariótipo**, que é diferente para cada espécie. Na maioria dos organismos, todas as células de organismo apresentam o mesmo cariótipo. Por outro lado, espécies que parecem bastante semelhantes podem ter cariótipos muito diferentes, indicando que o potencial genético semelhante pode ser organizado nos cromossomos de várias formas diferentes. Os **telômeros** são sequências de DNA especiais formando a extremidade dos cromossomos. Sua principal função é manter a estabilidade estrutural dos cromossomos. Cada vez que a célula se divide, os telômeros são ligeiramente encurtados, e não se regeneram, chegando a um ponto que de tão curtos não permitem mais a correta replicação dos cromossomos e a célula perde completamente ou parcialmente a sua capacidade de divisão.

Descobertas sobre telômeros podem ser fundamentais no tratamento de várias doenças. Os geneticistas: Elizabeth Blackburn (professora de biologia e fisiologia da University of California em San Francisco); Carol Greider (professora do departamento biologia molecular e genética Escola de Medicina da Johns Hopkins University em Baltimore) e Jack Szostak (professor de genética do Hospital Geral de Massachusetts em Boston) - descobriram os telômeros, código genético que protege as extremidades dos cromossomos, e a telomerase, enzima que auxilia nesse processo. Suas descobertas são importantes no estudo do câncer, envelhecimento e células-tronco. Reconhecidos pela pesquisa com telômeros e telomerase, dividiram o Prêmio Nobel de fisiologia e medicina de 2009. O trabalho realizado pelos pesquisadores esclarece aspectos importantes sobre a replicação do DNA. À medida que o material genético é copiado, a partir do cromossomo, durante a divisão celular, todo o filamento do DNA precisa ser duplicado de ponta a ponta, caso contrário, partes da informação genética se perderão. Até os anos 80, era considerado

um mistério o fato de os cromossomos se replicarem de forma confiável, sem perder nenhuma ponta dos filamentos no processo todo. Os pesquisadores demonstraram que o DNA poderia ser encurtado e cortado no processo de replicação, se partes da capa que recobre as extremidades do telômero estivesse faltando. As descobertas, desde então, têm sido aplicadas a estudos de envelhecimento, células-tronco e câncer.

2.1.5 Mutações

Ocasionalmente, podem ocorrer erros espontâneos durante a replicação do DNA provocando alterações na sequência dos nucleotídeos. Estas alterações, ou **mutações**, podem ser causadas por erros de cópia do material durante a divisão celular, por exposição a radiação ultravioleta ou ionizante, mutagênicos químicos, ou vírus. A célula pode também causar mutações deliberadamente durante processos conhecidos como hipermutação. Em organismos multicelulares, as mutações podem ser divididas entre **mutação de linhagem germinativa**, que pode ser passada aos descendentes, e **mutações somáticas**, que não são transmitidas aos descendentes em animais. Em alguns casos, plantas podem transmitir mutações somáticas aos seus descendentes, de forma assexuada ou sexuada (em casos em que as gemas de flores se desenvolvam numa parte que sofreu mutação somática). Assim, essa classificação é pouco eficiente para plantas, se ajustando melhor a animais. Uma nova mutação que não foi herdada de nenhum dos pais é chamada de **mutação de novo**. A fonte da mutação não se relaciona com seus efeitos, apesar de seus efeitos estarem relacionados com quais células são afetadas pela mutação.

Mutações geram variações no conjunto de genes da população. **Mutações desfavoráveis** (ou deletérias) podem ter sua frequência reduzida na população por meio da seleção natural, enquanto **mutações favoráveis** (benéficas ou vantajosas) podem se acumular, resultando em mudanças evolutivas adaptativas. Por exemplo, uma borboleta pode produzir uma prole com novas mutações. A maioria dessas mutações não terá efeito. No entanto, uma delas pode mudar a cor dos descendentes desse indivíduo, tornando-os mais difíceis (ou fáceis) de serem vistos por predadores. Se essa mudança de cor for vantajosa, a chance dessa borboleta sobreviver e produzir sua própria prole será um pouco maior, e com o tempo o número de borboletas com essa mutação constituir uma maior proporção da população.

Mutações neutras são definidas como mutações cujos efeitos não influenciam a aptidão dos indivíduos. Essas mutações podem se acumular ao longo do tempo devido à deriva genética. Acredita-se que a imensa maioria das mutações não tem efeito significativo na aptidão dos organismos. Além disso, mecanismos de reparo de DNA são capazes de corrigir a maior parte das mudanças antes que elas se tornem mutações permanentes, e muitos organismos têm mecanismos para eliminar células somáticas que sofreram mutações.

As mutações são consideradas o mecanismo que permite a ação da seleção natural, já que insere a variação genética sobre a qual ela irá agir, fornecendo as novas características vantajosas que sobrevivem e se multiplicam nas gerações subsequentes ou as características deletérias que desaparecem em organismos mais fracos.

A sequência de um gene pode ser alterada de diversas maneiras. Mutações genéticas têm diferentes efeitos na saúde, dependendo de onde ocorrem e se alteram a função de proteínas essenciais. Estruturalmente, as mutações podem ser classificadas em:

1. **Mutações de pequena escala, como aquelas que afetam um pequeno gene em um ou poucos nucleotídeos, incluindo:**

- **Mutação de ponto** - Geralmente causada por substâncias mutagênicas ou erros na replicação do DNA, há a troca de um único nucleotídeo por outro [28]. A mais comum, conhecida por transição, ocorre quando há a troca de uma purina por outra purina ($A \leftrightarrow G$) ou uma pirimidina por outra pirimidina ($C \leftrightarrow T$). Transições podem ser causadas por Ácido Nítrico, erro de pareamento entre as bases, ou mutagênicos análogos, como 5-bromo-2-desoxiuridina (BrdU). Um tipo de mutação pontual menos comum é a transversão, em que há a troca de uma purina por uma pirimidina, ou vice-versa ($C/T \leftrightarrow A/G$). Uma mutação pontual pode ser revertida por outra mutação pontual em que o nucleotídeo é mudado de volta ao seu estado original ou por uma reversão a partir de outra mutação (uma mutação complementar em outro local que resulta no retorno do gene à função anterior) [29]. Mutações pontuais que ocorrem dentro da região codificadora da proteína podem ser classificadas em três tipos, dependendo do tipo de expressão apresentado pelo códon mutado [26]:
 - a) **Mutação silenciosa:** O códon codifica para o mesmo aminoácido;
 - b) “**Missense**”, sentido trocado: Codifica para um aminoácido diferente;
 - c) **Mutação sem sentido:** Codifica para um códon de parada, que interrompe a proteína antes de seu término
- **Inserção** - Ocorre pela adição de um ou mais nucleotídeos na sequência de DNA. Geralmente, esse tipo de mutação é causado por transposons ou erros durante a replicação de elementos repetitivos (sequências AT, por exemplo). Inserções na região codificadora de um gene podem alterar o corte (*splicing*) do mRNA, ou causar mudança no quadro de leitura dos códons. Ambas alterações podem modificar significativamente o produto gênico.

- **Deleção** - Há a remoção de um ou mais nucleotídeos da sequência de DNA. Assim como inserções, essas mutações podem modificar o quadro de leitura do gene. Geralmente elas são irreversíveis; apesar de teoricamente a mesma sequência poder ser restaurada por inserção, elementos de transposição capazes de reverter uma deleção muito curta (com uma ou duas bases) em um dado local são muito improváveis ou mesmo inexistentes. É importante notar que uma deleção não é o oposto exato de uma inserção. Enquanto deleções são aleatórias, inserções consistem de uma sequência específica sendo inserida em locais que não são completamente aleatórios.

2. Mutações de grande escala da estrutura do cromossomo, incluindo:

- **Amplificação** (ou duplicação gênica) - Dada pela criação de várias cópias de uma região cromossômica, aumentando a dosagem dos genes dentro dela;
- **Deleção de regiões cromossômicas** - Leva à perda dos genes presentes nessas regiões.
- **Inserção** - Mutações cujo efeito é unir partes do DNA anteriormente separadas, potencialmente unindo genes de tal forma que surjam genes fundidos funcionalmente distintos (por exemplo, bcr-abl). Esse tipo de mutação inclui:
 - a) **Translocação cromossônica**: Ocorre a troca de porções de cadeias de DNA entre cromossomos não homólogos;
 - b) **Deleção do interstício**: Há a deleção de um segmento de DNA de um cromossomo, agrupando, assim, genes anteriormente distantes;
 - c) **Inversão cromossônica**: Ocorre a inversão da orientação de um segmento do cromossomo.
- **Perda de heterozigozidade** - Há a perda de um alelo por deleção ou recombração num organismo que originalmente possuia dois alelos.

As mudanças no DNA causadas por **mutações desfavoráveis** podem causar erros na sequência das proteínas, criando proteínas parciais ou completamente não-funcionais. Para funcionar corretamente, cada célula depende de milhares de proteínas para funcionar nos sítios apropriados. Quando uma mutação altera uma proteína que tem um papel importante no corpo, isso pode resultar numa doença. Uma enfermidade causada por mutações em um ou mais genes é chamada de doença genética. Contudo, apenas uma pequena percentagem de mutações causa doenças genéticas; a maioria não tem impacto na saúde. Por exemplo, algumas mutações alteram a sequência de bases de DNA de um gene mas não mudam a função

da proteína produzida por esse gene. Estudos na mosca da fruta *Drosophila melanogaster* sugerem que se uma mutação muda de fato uma proteína, esta mudança será provavelmente maléfica, com 70 por cento destas mutações tendo efeitos negativos e sendo as restantes neutras ou fracamente benéficas [30].

Se uma mutação estiver presente numa célula germinal, pode dar origem a descendentes portadores dessa mutação em todas as suas células. Este é o caso de doenças hereditárias. Por outro lado, uma mutação pode ocorrer numa célula somática de um organismo. Algumas mutações podem estar presentes em todos os descendentes desta célula e certas mutações podem provocar que a célula se torne maligna, e consequentemente cause cancro [31].

Muitas vezes, mutações gênicas que poderiam provocar uma doença genética são reparadas pelo sistema celular de reparação do DNA. Cada célula tem um certo número de vias bioquímicas através do qual enzimas reconhecem e reparam erros no DNA. Como o DNA pode ser danificado ou mutado de diversas maneiras, o processo de reparação do DNA é uma maneira importante do corpo se proteger de doenças.

Uma percentagem muito pequena de todas as mutações tem na verdade um efeito positivo, que são consideradas **mutações benéficas**. Estas mutações levam a novas versões de proteínas que ajudam o organismo e futuras gerações a adaptarem-se melhor a mudanças no seu ambiente. Por exemplo, uma deleção específica de 32 pares de base no CCR5 humano confere resistência ao HIV a homozigóticos e atrasa o despoletar do SIDA em heterozigóticos. A mutação CCR5 é mais comum em pessoas com ascendência europeia. Uma teoria para a etiologia da relativa alta frequência do CCR5- $\Delta 32$ na população europeia é que esta confere resistência à peste bubônica que flagelou a Europa em meados do Século XIV. Pessoas que tinham esta mutação foram capazes de sobreviver à infecção; por isso, a sua frequência na população aumentou. Isso pode também explicar porque esta mutação não se encontra na África, que não foi afetada pela peste bubônica. Uma teoria mais recente diz que pressão seletiva na mutação CCR5- $\Delta 32$ foi causada pela varíola em vez da peste bubônica [32].

2.2 Estruturas Algébricas

Nesta seção, apresentamos as principais definições e propriedades das estruturas algébricas de anéis e corpos. Estas estruturas são fundamentais na teoria de CCEs que veremos na próxima seção, pois facilitam os processos de codificação, decodificação e análise de desempenho destes.

Os conceitos e definições apresentados nesta seção podem ser encontrados em [33] e [34].

2.2.1 Anéis

Definição 2.2.1. Um **anel** $\langle R, +, \cdot \rangle$ é um conjunto não-vazio R juntamente com duas operações binárias $+$ e \cdot definidas sobre R , as quais chamamos de adição e multiplicação, tal que os seguintes axiomas são satisfeitos:

1. $\langle R, + \rangle$ é um grupo abeliano;
2. A operação de multiplicação é associativa, isto é, $(ab)c = a(bc)$, $\forall a, b, c \in R$;
3. Para todo $a, b, c \in R$, é válida a lei distributiva à esquerda, $a(b + c) = (ab) + (ac)$, e a lei distributiva à direita, $(a + b)c = (ac) + (bc)$.

Definição 2.2.2. Se a e b são elementos não nulos de um anel R tais que $ab = 0$ ou $ba = 0$, então a e b são **divisores de zero**.

Definição 2.2.3. Seja R um anel. Um **R -módulo** consiste de um grupo abeliano G e uma operação de multiplicação de cada elemento de G por todo elemento de R pela esquerda, tais que para todo $\alpha, \beta \in G$ e $r, s \in R$, as seguintes condições são satisfeitas:

1. $(r\alpha) \in G$;
2. $r(\alpha + \beta) = r\alpha + r\beta$;
3. $(r + s)\alpha = r\alpha + s\alpha$;
4. $(rs)\alpha = r(s\alpha)$.

2.2.2 Corpos

Definição 2.2.4. Um **corpo** é um anel comutativo com unidade e tal que todo elemento não-nulo é inversível.

Portanto, dizemos que F é um corpo sob as operações binárias $+$ e \cdot se, e somente se, F constitui um grupo abeliano sob estas operações e, para a operação \cdot , é válida a lei distributiva. Assim, podemos dizer que um corpo apresenta no mínimo dois elementos: as identidades das operações $+$ e \cdot . O número de elementos num corpo é a **ordem** do mesmo e um corpo onde este número é finito é chamado **corpo finito**.

Teorema 2.2.1. As classes residuais de polinômios módulo um polinômio $f(x)$ de grau n formam uma álgebra de dimensão n sobre o corpo dos coeficientes.

Teorema 2.2.2. Seja $p(x)$ um polinômio com coeficientes em um corpo F . Se $p(x)$ for irreduzível em F , i.e., se $p(x)$ não possuir fatores com coeficientes em F , então a álgebra de polinômios sobre F módulo $p(x)$ será um corpo.

Os corpos finitos são usados na maioria das construções dos códigos conhecidos, estes corpos são também conhecidos como **corpos algébricos de Galois** ou **corpos de Galois** e são denotados por $GF(q)$ ou F_q onde $q \geq 2$ é o número de elementos do corpo.

O corpo formado por polinômios sobre um corpo F módulo um polinômio irreduzível $p(x)$ de grau r é chamado **corpo de extensão** de grau r sobre F .

Teorema 2.2.3. Seja F^* o conjunto dos $q - 1$ elementos não-nulos de $GF(q)$, onde $q = p^r$. Então, F^* é um grupo cíclico multiplicativo de ordem p^{r-1} .

Definição 2.2.5. Um **polinômio** de grau $n - 1$ sobre um corpo F_q é escrito como:

$$p(x) = p_{n-1}x^{n-1} + p_{n-2}x^{n-2} + \cdots + p_1x + p_0,$$

onde x é uma variável e os coeficientes $p_i, 0 \leq i \leq n - 1, i \in \mathbb{Z}$, são elementos de F_q .

Definição 2.2.6. Um **polinômio mônico** é aquele cujo coeficiente líder (coeficiente da variável de maior expoente) p_{n-1} é igual a 1, a identidade multiplicativa de F_q .

Sabemos que o conjunto de todos os polinômios sobre $GF(q)$ forma um anel sob as operações usuais de soma e multiplicação de polinômios. Este anel é denotado por $GF(q)[x]$ ou $F_q[x]$.

Definição 2.2.7. Um elemento $\beta \in F_q$ é uma **raiz** ou **zero** do polinômio $p(x) \in F_q[x]$ se $p(\beta) = 0$.

Teorema 2.2.4. Se G é um subgrupo multiplicativo do grupo (F^*, \cdot) de elementos não nulos de um corpo F , então G é cíclico.

Teorema 2.2.5. O anel de polinômios módulo um polinômio $p(x)$ sobre F_q é um corpo se, e somente se, $p(x)$ é um polinômio primo.

Definição 2.2.8. Um gerador do grupo multiplicativo de F_q é denominado um **elemento primitivo** de F_q .

Corolário 2.2.1. Todo corpo finito F contém um elemento primitivo.

Uma consequência imediata do Corolário 2.2.1 é a de que todo corpo de Galois contém um elemento β , tal que todo elemento pertencente ao grupo multiplicativo do corpo finito pode ser expresso como uma potência de β .

Definição 2.2.9. Seja $GF(q')$ um corpo finito e $GF(q)$ um subcorpo de $GF(q')$. Seja $\beta \in GF(q')$. O polinômio primo $p(x)$ de menor grau sobre $GF(q)$, tal que $p(\beta) = 0$, é chamado **polinômio minimal de β sobre $GF(q)$** .

Teorema 2.2.6. Considere os corpos $GF(q')$ e $GF(q)$ como definidos acima. Cada elemento β de $GF(q')$ tem um único polinômio minimal sobre $GF(q)$. Mais do que isso, se β tem $p(x)$ como seu polinômio minimal e um polinômio $g(x)$ tem β como um zero, então $p(x)$ divide $g(x)$.

2.3 Códigos Corretores de Erros

Um código corretor de erros é, em essência, um modo organizado de acrescentar algum dado adicional a cada informação que se queira transmitir ou armazenar e que permita, ao recuperar a informação, detectar e corrigir erros. Os CCEs participam do nosso cotidiano de inúmeras formas, estando presentes, por exemplo, sempre que fazemos uso de informações digitalizadas, tais como assistir a um programa de televisão, falar ao telefone, ouvir um CD de música, assistir a um filme em DVD, mandar um recado a alguém via pager ou navegar pela Internet.

A teoria dos códigos originou-se do trabalho do matemático Claude E. Shannon [2], na década de 1940, sendo que a teoria de CCEs teve início nesta mesma década com os trabalhos de Shannon [2], Golay [35] e Hamming [36]. A grande descoberta da época surgiu, principalmente devido a Shannon, com os modelos de códigos capazes de detectar e corrigir erros num sistema de comunicações. Shannon mostrou em 1948 que, através de uma codificação apropriada da informação, os erros introduzidos por um canal ruidoso a nível desejado sem sacrificar a taxa de transmissão (desde que a taxa de informação R (expressa em dígitos por segundo) for menor que um valor C (também expressa em dígitos por segundo)). A partir de então, pesquisadores vêm procurando encontrar famílias de bons códigos (previstos pela teoria de Shannon) e bons conjuntos de sinais associados a esses códigos, bem como projetar decodificadores eficientes para os mesmos. A finalidade dos códigos é detectar e talvez corrigir esses erros. A pesquisa em busca de tais códigos deu origem à **teoria da codificação**, [37]-[40].

Na linha de pesquisa de códigos, surgiram as classes de códigos lineares e não-lineares e, na linha de conjunto de sinais, foram propostas constelações de sinais sob diversas restrições, como, por exemplo, potência média, potência de pico, faixa de frequência e algumas combinações destas, como exemplos citamos os códigos de Slepian, seus variantes obtidos através de transformações ortogonais e as constelações tendo como base reticulados, etc.

Esses sistemas de comunicações tratavam separadamente os processos de codificação e modulação. Todavia, Ungerboeck [41] mostrou que, através do conceito de particionamento de conjunto de sinais, ganhos de codificação significativos eram obtidos: surgiu assim a modulação codificada.

Dentro dessa linha de pesquisa, Forney [42] apresentou uma nova classe de códigos denominada códigos geometricamente uniformes que, além de englobar os códigos de Slepian e os códigos reticulados, estende o procedimento proposto por Ungerboeck. A procura de bons códigos continua relevante, porém tendo que satisfazer, sempre que possível, a propriedade de serem geometricamente uniformes.

Historicamente, a teoria dos códigos se divide em códigos detectores de erros (CDEs) e códigos corretores de erros (CCEs). Os CCEs tem sido classificados como códigos de árvores, dividindo-se em duas classes principais: a classe dos códigos de blocos e a classe dos códigos de treliças, em geral construídos sobre as extensões de Galois de corpos. Cada uma dessas classes pode ser portanto classificada como linear ou não linear. A classe de códigos mais utilizada na prática é a classe dos códigos lineares. Esta classe está fortemente fundamentada nas estruturas algébricas de grupo, anel, corpos finitos e suas extensões. Estas estruturas são fundamentais, pois sistematizam os processos de codificação, decodificação e análise de desempenho dos códigos. Uma classe dos códigos de treliça lineares que é bem conhecida na literatura é a classe dos códigos convolucionais. A principal característica para esta classificação particular está na presença ou ausência de memória no codificador. A classe dos códigos não-lineares, em geral, não possui uma estrutura algébrica como a dos lineares, a menos que a cardinalidade desses códigos seja uma potência do alfabeto, isto é, q^m , onde $q = p^\alpha$ e α e m são inteiros positivos. Esta flexibilidade possibilita obter códigos com distâncias de Hamming maiores do que as encontradas com os códigos lineares. Porém, a falta de uma estrutura algébrica aumenta a complexidade do processo de decodificação. A Figura 2.21 mostra de maneira simplificada uma classificação dos principais códigos.

Algumas aplicações dos códigos corretores de erros

Com a necessidade do aumento da confiabilidade nas comunicações digitais e a emergência do computador digital como ferramenta essencial na sociedade tecnológica, os CCEs vêm conquistando uma posição proeminente. Para ilustrar a praticidade e importância do uso de CCEs temos:

- a) Uso do bit de paridade como um mecanismo detector de erro - é um dos esquemas mais simples e conhecidos na comunicação computacional;

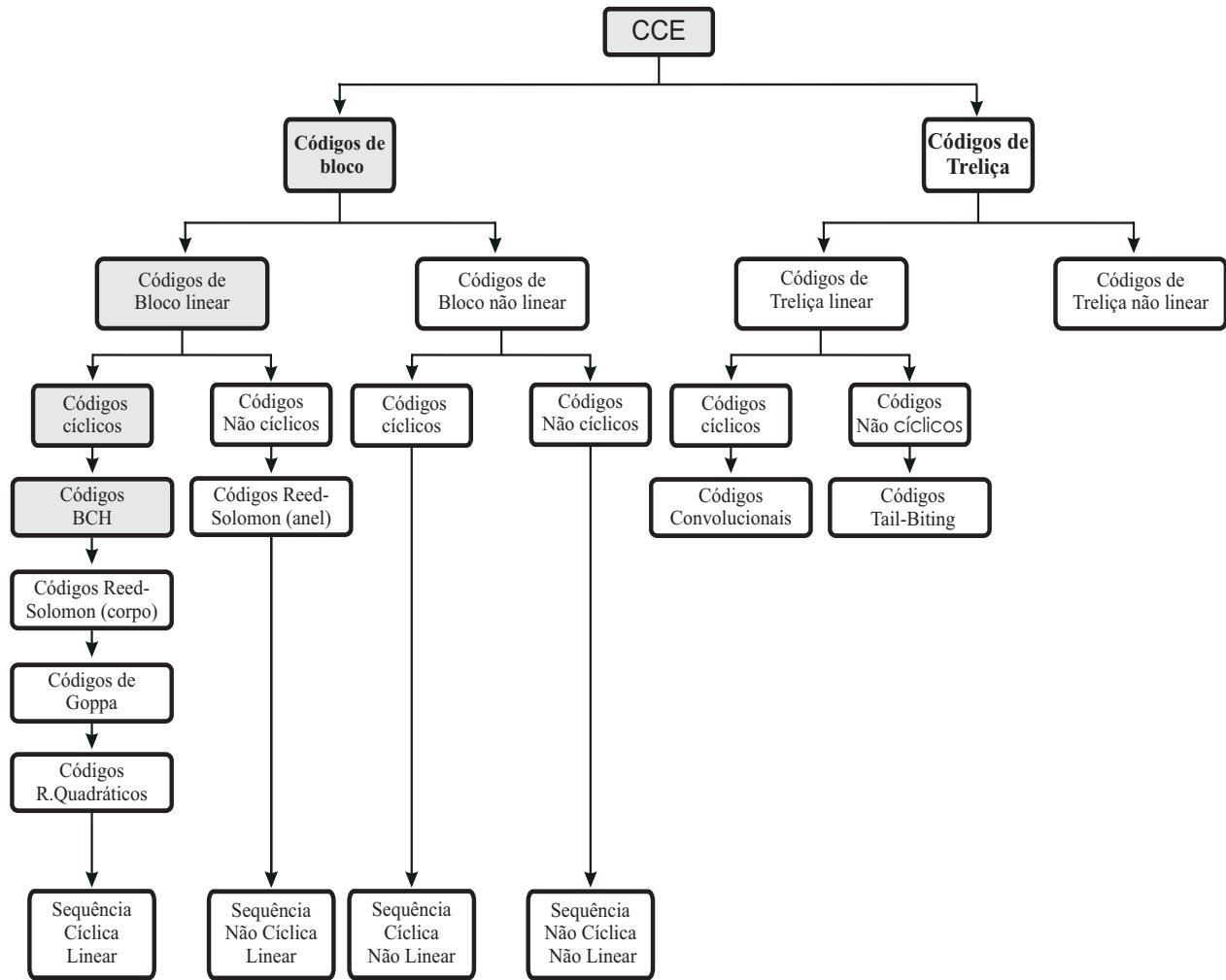


Figura 2.21: Códigos de árvore sob o ponto de vista algébrico.

- b) Armazenamento em discos - CCEs estão sendo muito utilizados devido ao aumento da densidade de armazenamento. Quanto maior a densidade, a probabilidade de ocorrência de erros também aumenta;
- c) Transmissão de informação pelas naves espaciais: - em 1972, a espaçonave Mariner transmitiu figuras de Marte para a Terra com 64 tonalidades de cinza. Atividade solar e outras condições atmosféricas podem introduzir erros em sinais fracos vindos do espaço. O código utilizado foi o de Reed-Muller; - em 1979, a espaçonave Voyager começou a enviar imagens com 4096 tonalidades de cores. O código utilizado foi o de Golay;
- d) Áudio digital - o aumento da popularidade do áudio digital deve-se ao desenvolvimento dos CCEs que facilita o processo de digitalização. Ao inicializar a leitura do CD, o sis-

tema corrige os erros produzidos por marcas de dedos, arranhões e outras imperfeições, para logo em seguida transformar em sinais sonoros. O código utilizado é o de Reed-Solomon.

Nesta seção, faremos uma breve revisão dos CCEs, que estão demarcados na cor cinza na Figura 2.21, necessários para o desenvolvimento deste trabalho. Na Subseção 2.3.1, revisamos os conceitos relacionados a códigos de bloco e suas principais características. As Subseções 2.3.2 e 2.3.3 apresentam uma revisão dos códigos lineares e dos códigos cíclicos, respectivamente. Na Subseção 2.3.4, apresentamos os conceitos e propriedades de códigos BCH sobre anéis e sobre corpos e suas extensões de Galois. Estes conceitos são importantes no entendimento do algoritmo para identificação de sequências de DNA desenvolvido no Capítulo 4. Na Subseção 2.3.5, fazemos uma breve introdução aos códigos geometricamente uniformes, descritos por Forney em [42] e, aos conjuntos de sinais casados a grupos, propostos por Loeliger em [43]. Finalmente, na Subseção 2.3.6, apresentamos conceitos da G -linearidade.

2.3.1 Códigos de bloco

As definições e teoremas a seguir podem ser encontrados em [33], [34] e [44].

Os códigos de bloco são caracterizados por serem códigos sem memória. O ponto de partida é um conjunto A , que pode ser finito ou infinito, chamado **alfabeto**.

Definição 2.3.1. Um **código** \mathcal{C} sobre um alfabeto A é qualquer subconjunto não-vazio do espaço de sequências A^I , onde A é chamado **alfabeto do código** e I é o **conjunto de índices** das sequências $c = \{c_i \mid i \in I\}$. Chamamos de **palavra-código** os elementos, ou **símbolos**, no alfabeto A que compõem o código \mathcal{C} .

Dessa definição, identificamos o alfabeto A com os elementos do corpo \mathbb{F}_q . O codificador para um código de bloco divide a sequência de informação em blocos de k símbolos, onde cada um desses blocos é representado por uma k -upla $\mathbf{u} = (u_1, \dots, u_k)$ chamada **mensagem**. No total, existem q^k mensagens diferentes. Após a divisão da sequência de informação, o codificador transforma cada mensagem \mathbf{u} em uma n -upla $\mathbf{v} = (v_1, \dots, v_n)$ de símbolos discretos chamado **palavra-código**. Se cada uma das q^k mensagens distintas é transformada em uma palavra-código, então existem também q^k palavras-código diferentes.

Neste trabalho, estamos interessados em alfabetos finitos. Entretanto, muitas vezes é conveniente que o mesmo seja “estruturado” a fim de que a codificação e a decodificação sejam simplificadas. Por alfabetos “estruturados”, entendemos aqueles que formam alguma estrutura algébrica, tal como corpo, anel ou grupo. Quando isto acontece e, além disso, o conjunto de índices é finito (por exemplo de tamanho n), dizemos que o conjunto formado

por q^k palavras-código de **comprimento** n é chamado **código de bloco**. Assim, temos a seguinte definição:

Definição 2.3.2. Um **código de bloco** \mathcal{C} de comprimento n sobre um alfabeto A é qualquer subconjunto A^n das sequências $c = \{c_i \mid 1 \leq i \leq n\}$.

Um código de bloco é caracterizado por três parâmetros principais: seu comprimento, sua dimensão e sua distância mínima.

Definição 2.3.3. A **dimensão** de um código \mathcal{C} é dada por

$$k = \log_{|A|} |\mathcal{C}|,$$

onde $|\cdot|$ denota a cardinalidade do conjunto.

Definição 2.3.4. Seja \mathcal{C} um código de comprimento n tal que $|\mathcal{C}| \geq 2$. A **distância mínima de Hamming** de \mathcal{C} , denotada por $d_{min}(\mathcal{C})$ é dada por:

$$d_{min}(\mathcal{C}) = \min_{x,y \in \mathcal{C}, x \neq y} d(x,y).$$

A distância d utilizada na caracterização do código depende da métrica utilizada no alfabeto em questão.

Assim, um código de bloco \mathcal{C} de comprimento n , dimensão k e distância mínima de Hamming $d = d_{min}(\mathcal{C})$ é representado por (n, k, d_{min}) . O seguinte teorema dá um limitante superior para a distância mínima em função dos parâmetros n e k .

Teorema 2.3.1. Para qualquer código de bloco (n, k, d_{min}) , vale a seguinte desigualdade:

$$d \leq n - k + 1.$$

Um outro parâmetro muito importante na caracterização de um código de bloco, indicador de desempenho deste, é a chamada **taxa** do código, definida pela razão entre a dimensão do código e seu comprimento, ou seja,

$$r_{\mathcal{C}} = \frac{k}{n}.$$

Códigos de bloco podem ser usados como CCEs. A **capacidade de correção de erros** de um código (n, k, d) , denotada por t , está relacionada à distância mínima deste código da seguinte forma:

$$d_{min} \leq 2t + 1.$$

Logo, quanto maior a distância mínima do código, maior é a capacidade deste de corrigir erros. Em geral, bons códigos são longos e, por isso, torna-se impraticável descrevê-los através

de listas de palavras-código. Para contornar esse problema, o caminho usual é associar aos códigos estruturas matemáticas que facilitem a execução das operações de codificação e decodificação. Neste sentido, a principal classe dos códigos de bloco é a dos *códigos lineares*.

2.3.2 Códigos lineares

Considere um código de bloco com q^k palavras-código e comprimento n . Se k e n são relativamente grandes, então teremos dificuldade quanto ao espaço para armazenar essas palavras-código. Nesse sentido, códigos de bloco com uma estrutura linear são mais práticos e reduzem a complexidade do codificador. A maioria dos códigos conhecidos até hoje pertencem à classe dos códigos lineares.

Definição 2.3.5. Um código (n, k, d) é dito **linear** se, e somente se, todas as suas palavras-código formam um subespaço vetorial de dimensão k do espaço vetorial F_q^n , o conjunto das n -uplas do corpo F_q .

Observe que um código linear pode ser visto como um subgrupo do grupo aditivo $(\mathbb{F}_q^n, +)$, por esse motivo os códigos lineares são também conhecidos como **códigos de grupo**.

A estrutura linear nos permite fazer algumas observações muito úteis. Uma delas, é que a distância mínima d entre palavras-código distintas é igual ao peso mínimo de todas as palavras-código não nulas de \mathcal{C} , isto é, $d = \omega(\mathcal{C})$.

Outra vantagem de trabalhar com códigos lineares é que, como \mathcal{C} é um subespaço vetorial de dimensão k , podemos exibir uma base $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ para \mathcal{C} . Assim, qualquer palavra-código $\mathbf{v} = (a_1, a_2, \dots, a_k) \in \mathcal{C}$ pode ser escrita como combinação linear dos vetores da base, ou seja,

$$\mathbf{v} = a_1 \mathbf{v}_1 + a_2 \mathbf{v}_2 + \dots + a_k \mathbf{v}_k,$$

onde $a_i \in \{0, 1, 2, \dots, q - 1\}$ e a soma é realizada módulo q .

Qualquer conjunto de vetores formando uma base para o subespaço \mathcal{C} pode ser usado como as linhas de uma matriz G , chamada de **matriz geradora** do código (n, k, d) . O espaço das linhas de G é o código linear \mathcal{C} , e qualquer palavra-código é uma combinação linear das linhas de G . Se a dimensão do espaço vetorial \mathcal{C} é k , então o número de linhas de G é igual a k , pois as linhas de G são linearmente independentes, ou seja, o posto de G é k . Assim G é uma matriz $k \times n$.

Se $B = \{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k\}$ é uma base ordenada de um código \mathcal{C} , então a matriz geradora G de \mathcal{C} associada a B , cujas linhas são os vetores $\mathbf{v}_i = (v_{i1}, \dots, v_{in})$, $i = 1, \dots, k$, é dada por

$$G = \begin{pmatrix} v_{11} & v_{12} & \cdots & v_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ v_{k1} & v_{k2} & \cdots & v_{kn} \end{pmatrix}.$$

Dessa forma, o processo de codificação pode ser escrito como:

$$\mathbf{v} = \mathbf{u}G,$$

onde \mathbf{u} é a palavra a ser codificada ou informação e \mathbf{v} é a palavra-código correspondente.

Como essa matriz depende da escolha da base, então ela não é univocamente determinada por \mathcal{C} . Entretanto, duas matrizes geradoras de um mesmo código podem ser obtidas uma da outra por meio de uma sequência de operações, tais como: permutação de linhas, multiplicação de uma linha por um escalar não nulo e adição de um múltiplo escalar de uma linha a outra. Assim, toda matriz geradora G é equivalente a uma matriz na forma escalonada das linhas, e portanto, equivalente a uma matriz da forma

$$G = (I_k \mid P)$$

chamada **matriz geradora** na forma sistemática, onde I_k é a matriz identidade de ordem k e P é uma matriz $k \times (n - k)$.

Para toda palavra-código \mathbf{v} vale a relação

$$\mathbf{v}H^T = 0,$$

onde a matriz $(n - k) \times n$, denotada por H , é chamada **matriz verificação de paridade** de \mathcal{C} , e qualquer vetor ortogonal a suas linhas pertence ao espaço vetorial das linhas da matriz geradora G associada e vice-versa. O código gerado pela matriz H é chamado **código dual** do código \mathcal{C} , denotado por \mathcal{C}^\perp .

Existe uma maneira simples de achar uma matriz verificação de paridade para um código se uma matriz geradora é dada na forma sistemática. Se \mathcal{C} é o espaço linha da matriz $G = (I_k \mid P)$, então \mathcal{C} é o espaço ortogonal de $H = (-P^T \mid I_{n-k})$, onde I_{n-k} é a matriz identidade de ordem $n - k$ e P^T é a matriz transposta de P .

Definição 2.3.6. Dado um código \mathcal{C} com matriz verificação de paridade H , a síndrome de um vetor $\mathbf{v} \in \mathbb{F}_q$ é o vetor $\mathbf{v}H^T = \mathbf{s}$.

A síndrome é um argumento usado para fazer a correção de erros em códigos lineares.

A expressão **padrão de erro** denomina a diferença entre a palavra-código recebida e a palavra-código enviada. Em um código linear \mathcal{C} com parâmetros (n, k) , considere um padrão

de erro $\mathbf{e} \in \mathbb{F}_q^n$. Como \mathcal{C} é um subgrupo, então $\mathbf{e} + \mathcal{C} = \{\mathbf{e} + \mathbf{v} \mid \mathbf{v} \in \mathcal{C}\}$ é uma classe lateral de \mathbb{F}_q^n .

Estabeleça uma tabela da seguinte maneira:

- a primeira linha da tabela deve conter todas as palavras-código de \mathcal{C} começando com a palavra toda nula;
- Das n -uplas de \mathbb{F}_q^n que não foram usadas, escolha aquela com menor peso e chame-a de \mathbf{e}_1 . A segunda linha da tabela será composta pela classe lateral $\mathbf{e}_1 + \mathcal{C}$;
- A j -ésima linha da tabela é formada pela classe $\mathbf{e}_j + \mathcal{C}$, onde \mathbf{e}_j é sempre escolhido como a n -upla em \mathbb{F}_q^n de menor peso que ainda não foi usada;
- Esse procedimento termina quando todas as palavras de \mathbb{F}_q^n tenham sido usadas.

A Tabela 2.1 assim determinada é chamada **arranjo padrão**. Algumas observações importantes devem ser feitas sobre o arranjo padrão. Cada palavra aparece uma única vez na tabela. Duas palavras estão na mesma classe lateral se, e somente se, possuem a mesma síndrome. A primeira coluna da tabela é formada pelas palavras de peso mínimo dentro de cada classe, e são denominadas os **líderes** das classes laterais.

| $\mathbf{v}_1 = 0$ | \mathbf{v}_2 | \mathbf{v}_3 | \cdots | \mathbf{v}_q^k |
|------------------------|---------------------------------------|---------------------------------------|----------|---|
| \mathbf{e}_1 | $\mathbf{e}_1 + \mathbf{v}_2$ | $\mathbf{e}_1 + \mathbf{v}_3$ | \cdots | $\mathbf{e}_1 + \mathbf{v}_q^k$ |
| \mathbf{e}_2 | $\mathbf{e}_2 + \mathbf{v}_2$ | $\mathbf{e}_2 + \mathbf{v}_3$ | \cdots | $\mathbf{e}_2 + \mathbf{v}_q^k$ |
| \vdots | \vdots | \vdots | \cdots | \vdots |
| $\mathbf{e}_{q^{n-k}}$ | $\mathbf{e}_{q^{n-k}} + \mathbf{v}_2$ | $\mathbf{e}_{q^{n-k}} + \mathbf{v}_3$ | \cdots | $\mathbf{e}_{q^{n-k}} + \mathbf{v}_q^k$ |

Tabela 2.1: Arranjo padrão.

Uma regra de decodificação por **máxima verossimilhança** para um código linear é completamente descrita pelo arranjo padrão. O receptor utiliza o arranjo padrão para decodificar uma palavra recebida da seguinte maneira:

- recebido \mathbf{v} , calcule sua síndrome;
- ache o padrão de erro \mathbf{e} correspondente a essa síndrome na tabela;
- $\mathbf{v} - \mathbf{e}$ é a palavra-código.

Para um código (n, k) sobre \mathbb{F}_q^n , uma lista completa consiste de q^n palavras, ao passo que a lista dada no arranjo padrão tem apenas q^{n-k} palavras e suas síndromes. Mesmo assim, na prática, bons códigos são longos, e fazer o arranjo padrão ainda é impraticável para alguns casos.

2.3.3 Códigos cíclicos sobre \mathbb{Z}_q

Nesta subseção, apresentamos definições e teoremas relacionados a códigos cíclicos sobre anéis \mathbb{Z}_q ($q \geq 4$ e inteiro). Esta teoria servirá de base para o desenvolvimento da construção do código BCH sobre as estruturas algébricas de anéis e corpos e suas extensões de Galois, que será apresentado no Capítulo 4. As referências adotadas aqui são [37], [38], [45]-[47].

Definição 2.3.7. Seja \mathbb{R} um anel. Um **módulo livre** é um R -módulo gerado por um conjunto de vetores linearmente independentes.

Definição 2.3.8. Um **código linear** (n, k) sobre \mathbb{Z}_q é definido como um módulo livre de dimensão k no espaço de todas as n -uplas de \mathbb{Z}_q^n .

Definição 2.3.9. Um código linear \mathcal{C} com parâmetros (n, k) sobre \mathbb{Z}_q é **cíclico** se, para $v = (v_0, v_1, v_2, \dots, v_{n-1}) \in \mathcal{C}$, todo deslocamento cíclico $v^{(1)} = (v_{n-1}, v_0, v_1, v_2, \dots, v_{n-2}) \in \mathcal{C}$, com $v_i \in \mathbb{Z}_q$, $0 \leq i \leq n - 1$.

Os códigos cíclicos são geralmente representados na forma polinomial. Assim, considere a palavra-código $v = (v_0, v_1, v_2, \dots, v_{n-1})$ de um código cíclico \mathcal{C} . Podemos representá-la pelo polinômio:

$$v(x) = v_0 + v_1x + v_2x^2 + \dots + v_{n-1}x^{n-1}.$$

O produto entre x e $v(x)$ módulo $x^n - 1$ é dado por:

$$v^{(1)}(x) = v_{n-1} + v_0x + v_1x^2 + \dots + v_{n-2}x^{n-1},$$

que corresponde à palavra-código:

$$\underline{v}^{(1)} = (v_{n-1}, v_0, v_1, \dots, v_{n-2}),$$

a qual é um deslocamento cíclico da palavra:

$$\underline{v} = (v_0, v_1, v_2, \dots, v_{n-1}).$$

Portanto, $v^{(1)}(x)$ é obtido através do produto $x.v(x)$ no anel quociente $R_n = \frac{\mathbb{Z}_q[x]}{\langle x^n - 1 \rangle}$, onde $\langle x^n - 1 \rangle$ representa o ideal gerado por $x^n - 1$. A adição de duas palavras-código é feita em $\mathbb{Z}_q[x]$.

Note que o conjunto de todas as palavras pertencentes a um código cíclico \mathcal{C} forma um subconjunto do anel R_n , isto é, o conjunto de todos os polinômios cujo grau é menor que n .

Teorema 2.3.2. Um conjunto S de elementos em R_n é um código cíclico se, e somente se, S é um ideal em R_n .

Exemplo 2.3.1. Considere o anel \mathbb{Z}_4 e o anel quociente $R_2 = \frac{\mathbb{Z}_4[x]}{\langle x^2 - 1 \rangle}$. Pelo Teorema 2.3.2, temos que o ideal $\mathcal{C} = \{0, 2, 2x, 2 + 2x, 1 + 3x, 3 + x, 1 + x\}$ de R_2 é um código cíclico.

Proposição 2.3.1. Seja \mathcal{C} um ideal em $R_n = \frac{\mathbb{Z}_q[x]}{\langle x^n - 1 \rangle}$, isto é, um código cíclico de comprimento n . Se existir um polinômio de grau mínimo em \mathcal{C} , cujo coeficiente dominante é um elemento inversível em \mathbb{Z}_q , então o polinômio mônico (ou seja, aquele cujo coeficiente dominante é um) de grau mínimo em \mathcal{C} é único.

Teorema 2.3.3. Seja \mathcal{C} um ideal em $R_n = \frac{\mathbb{Z}_q[x]}{\langle x^n - 1 \rangle}$ e $g(x)$ um polinômio mônico com o menor grau em \mathcal{C} . Assim, $\mathcal{C} = \langle g(x) \rangle$, e portanto, o código \mathcal{C} consiste de todos os múltiplos de $g(x)$. Dizemos então que \mathcal{C} é um **ideal principal**.

Teorema 2.3.4. Seja \mathcal{C} um ideal em R_n . Se o coeficiente dominante do polinômio de menor grau em \mathcal{C} , $g(x)$, é um elemento inversível, então $g(x)$ divide $(x^n - 1)$. Note que se este polinômio for mônico, então $g(x)$ divide $(x^n - 1)$.

O Teorema 2.3.4 fornece um método de construção de códigos cílicos sobre anéis de inteiros residuais análogo ao método de construção de códigos cílicos sobre corpos finitos, ou seja, através da fatoração do polinômio $(x^n - 1)$ sobre o anel de interesse para então tomar um fator (ou produto de fatores) como polinômio gerador do código em questão.

O próximo teorema está relacionado à representação matricial dos códigos cílicos sobre anéis que possuem uma matriz geradora.

Teorema 2.3.5. Se $g(x)$ divide $(x^n - 1)$ e o grau de $g(x)$ é $(n - k)$, então a dimensão de $\mathcal{C} = \langle g(x) \rangle$ é k . Se

$$g(x) = g_0 + g_1x + g_2x^2 + \dots + x^{n-k},$$

então a matriz geradora do código \mathcal{C} é dada por:

$$G = \begin{pmatrix} g_0 & g_1 & g_2 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{n-k-1} & 1 & 0 & \dots & 0 \\ 0 & 0 & g_0 & \dots & g_{n-k-2} & g_{n-k-1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & g_0 & g_1 & g_2 & \dots & 1 \end{pmatrix}$$

Proposição 2.3.2. Se \mathcal{C} é um código cíclico sobre \mathbb{Z}_q onde $q = p_1^{k_1}p_2^{k_2}\dots p_q^{k_q}$, então

$$\mathcal{C} = \bigoplus_{i=1}^q \mathcal{C}_i,$$

onde \mathcal{C}_i é um código cíclico sobre $\mathbb{Z}_{p_i^{k_i}}$, $1 \leq i \leq q$.

2.3.4 Códigos BCH sobre anéis e corpos

Os códigos BCH formam uma importante classe de códigos cílicos devido, principalmente, à simplicidade dos processos de codificação e decodificação associados, o que os torna bons candidatos a serem utilizados em aplicações práticas. Os códigos BCH foram descobertos por R. C. Bose, D. K. Chaudhuri e A. Hocquenghem e representam uma excelente generalização dos códigos de Hamming, permitindo a múltipla correção de erros. Formam assim a classe dos melhores códigos construtivos para canais onde os erros afetam os símbolos de forma independente.

Apesar de ser sempre possível projetar um código BCH que corrija até t erros, para um t qualquer, devemos interpretar esta informação com uma certa restrição, pois as taxas desses códigos são assintoticamente ruins. Ou seja, quando o comprimento das palavras-código não é grande, existem bons códigos BCH, caso contrário, o desempenho destes é prejudicado devido às baixas taxas de transmissão. Contudo, a real importância dos códigos BCH vem da facilidade de implementação do algoritmo de correção de erros, algoritmo de Berlekamp-Massey modificado [46]. A seguir, fazemos algumas considerações sobre extensões de anéis e corpos de Galois e, em seguida, sobre os códigos BCH.

Extensão de Anéis de Galois

A motivação para se utilizar o conceito de extensão de Galois em teoria da codificação está diretamente relacionada com a construção de códigos cílicos sobre anéis locais \mathbb{Z}_q , onde q é uma potência de um primo, $q = p^k$, $k \geq 2$.

A principal diferença da construção de códigos cílicos sobre anéis para a construção de códigos cílicos sobre corpos está no fato de que as raízes do polinômio gerador dos códigos cílicos sobre anéis encontram-se na extensão do anel \mathbb{Z}_q , ao invés de serem encontradas na extensão do corpo $\mathbb{F}_q \cong GF(p^r)$.

Definição 2.3.10. Um código cíclico sobre \mathbb{Z}_q com comprimento $n = q^r - 1$, onde $q = p^k$ e r é o grau da extensão de Galois, é denominado **código cíclico primitivo**.

Vamos assumir que p e n são relativamente primos, isto é, o máximo divisor comum é um, denotado por $mdc(p, n) = 1$, pois assim garantimos que $(x^n - 1)$ não apresenta fatores

quadráticos. Da Subseção 2.3.3, sabemos que um código cíclico de comprimento n sobre \mathbb{Z}_q é o ideal principal no anel de polinômios sobre \mathbb{Z}_q módulo $(x^n - 1)$ e que este ideal é gerado por qualquer polinômio $g(x)$ que divide $(x^n - 1)$.

Seja $\mathbb{Z}_q[x]$ o anel de polinômios na variável x sobre \mathbb{Z}_q onde $p(x)$ é um polinômio primitivo de grau r , irredutível sobre $GF(p)$ e, consequentemente, sobre \mathbb{Z}_q . Representamos por $GR(p^k, r)$ o quociente $\mathbb{Z}_q[x]$ pelo ideal gerado por $p(x)$, ou seja,

$$\mathbb{R} \cong GR(p^k, r) \cong \frac{\mathbb{Z}_q[x]}{\langle p(x) \rangle}.$$

Assim, o anel R é formado por todas as classes laterais de polinômios em x sobre \mathbb{Z}_q mod $p(x)$, isto é, consiste do conjunto dos polinômios de grau menor ou igual a $r - 1$ cujas operações binárias de adição e multiplicação são realizadas módulo $p(x)$. Além disso, R é um anel comutativo com identidade denominado extensão de Galois de dimensão r de \mathbb{Z}_q . Esta extensão é única a menos de isomorfismo [45].

O anel $\mathbb{R} \cong GR(p^k, r)$ é um anel local [45], isto é, seus elementos divisores de zero formam um grupo abeliano aditivo e consistem dos polinômios de grau menor ou igual a $r - 1$ cujos coeficientes são divisores de zero em \mathbb{Z}_q . Um polinômio $p(x) \in R$ com pelo menos um coeficiente inversível em \mathbb{Z}_q não é um divisor de zero em R e, portanto, pertence a R^* (grupo das unidades de R), ou seja, é sempre possível encontrar um polinômio $q(x) \in R$, tal que $p(x) \cdot q(x) = 1$.

Vale lembrar que, da Definição 2.3.8, temos:

Definição 2.3.11. [47] Um polinômio não nulo $p(x)$ é um **divisor de zero** em $\mathbb{Z}_q[x]$ se existe um polinômio $q(x) \in \mathbb{Z}_q[x]$, $q(x) \neq 0$, tal que $p(x) \cdot q(x) = 0$.

Definição 2.3.12. [47] Um polinômio $p(x)$ é dito **regular** se ele não é um divisor de zero no anel $\mathbb{Z}_q[x]$.

Definição 2.3.13. [47] Um polinômio regular $p(x)$ é chamado **local** se $\frac{\mathbb{Z}_q[x]}{\langle p(x) \rangle}$ é uma extensão local de \mathbb{Z}_q .

A irredutibilidade do polinômio $p(x)$ sobre \mathbb{Z}_q é garantida pelo seguinte teorema:

Teorema 2.3.6. [47] Seja $p(x)$ um polinômio regular em \mathbb{Z}_q . Se existe uma aplicação μ , chamada projeção natural, tal que $\mu(p(x))$ seja diferente de zero e irredutível em $GF(p)$, então $p(x)$ é irredutível em \mathbb{Z}_q .

Como, neste momento, estamos interessados na classe dos códigos cíclicos, nosso objetivo é fornecer um procedimento para a construção de tais códigos. O primeiro passo está relacionado com a fatoração de $(x^n - 1)$. Como o grupo das unidades de R , R^* , é um grupo

abeliano multiplicativo, ele pode ser expresso como um produto de grupos cíclicos. Uma vez encontrado este grupo multiplicativo, o problema da construção de códigos cíclicos se reduz à escolha de determinados elementos deste grupo que sejam raízes do polinômio gerador $g(x)$, que divide $(x^n - 1)$.

Os resultados a seguir fornecem os elementos necessários para a construção do subgrupo cíclico G_n do grupo multiplicativo R^* , que contém todas as raízes de $(x^n - 1)$.

Teorema 2.3.7. [45] Existe um único subgrupo cíclico de R^* cuja ordem é relativamente prima a p . Este subgrupo tem ordem $p^r - 1$.

Teorema 2.3.8. [48] Suponha que $f \in R$ gere um subgrupo de ordem n em R^* , onde $\text{mdc}(n, p) = 1$. Então o polinômio $(x^n - 1)$ pode ser fatorado como $x^n - 1 = (x - f)(x - f^2) \dots (x - f^n)$ se, e somente se, $R_p(f)$ tem ordem n em F^* (grupo multiplicativo de $GF(p^r)$), onde $R_p(f)$ é o resto da divisão de f por p (redução de f módulo p).

Corolário 2.3.1. [48] Um polinômio $h(x)$, que divide $(x^n - 1)$ e tem coeficientes em \mathbb{Z}_q , pode ser fatorado sobre G_n como:

$$h(x) = (x - \beta^{e1})(x - \beta^{e2}) \dots (x - \beta^{ej}),$$

se, e somente se, $R_p(h(x))$ pode ser fatorado sobre $GF(p^r)$ como:

$$R_p(h(x)) = (x - (R_p(\beta))^{e1})(x - (R_p(\beta))^{e2}) \dots (x - (R_p(\beta))^{ej}),$$

onde β é um elemento primitivo de G_n e $e_j \in \mathbb{Z}$.

Teorema 2.3.9. [48] Suponha que $f_1 = R_p(f)$ gere um subgrupo cíclico de ordem n em F^* . Então f gera um subgrupo cíclico de ordem nd em R^* , onde d é um inteiro maior ou igual a um , e f^d gera o subgrupo cíclico g_n de R^* .

O subgrupo cíclico G_n é obtido do Teorema 2.3.9, enquanto pelo Corolário 2.3.1, o polinômio minimal $M_i(x)$ associado ao elemento β^i sobre R^* (onde β é um elemento primitivo em G_n) tem como suas raízes todos os elementos na sequência

$$\beta^i, (\beta^i)^p, (\beta^i)^{p^2}, \dots, (\beta^i)^{p^{r-1}}.$$

Portanto, o polinômio minimal $M_i(x)$ pode ser construído de forma similar à construção do polinômio minimal $m_i(x)$ de $R_p(\beta^i)$ sobre $GF(p)$.

Temos ainda a seguinte propriedade:

Teorema 2.3.10. [46] Seja β um elemento primitivo em G_n , onde $n = p^r - 1$. Então o elemento $\delta = \beta^{l_1} - \beta^{l_2}$ possui inverso em R se $0 \leq l_1 \neq l_2 \leq n - 1$.

Definição 2.3.14. Um código cíclico de comprimento n sobre $GF(p)$ é denominado um **código BCH com distância de projeto** d se o seu gerador $g(x)$ for o mínimo múltiplo comum dos polinômios minimais de

$$\beta^m, \beta^{m+1}, \beta^{m+2}, \dots, \beta^{m+d-2},$$

para algum m inteiro não negativo, onde β é uma raiz primitiva (elemento primitivo) de $(x^n - 1)$, em alguma extensão $GF(p^r)$ de $GF(p)$.

Assim, analogamente à Definição 2.3.6, temos:

Definição 2.3.15. Se $n = p^r - 1$, ou seja, se β for um elemento primitivo em F_q , então o código BCH é chamado **primitivo**.

Normalmente, consideramos $m = 1$, o que nos fornece o chamado código BCH no **sentido estrito**.

Os códigos BCH no sentido estrito definidos sobre anéis de inteiros, com distância de projeto d e comprimento n , apresentam $\beta, \beta^2, \beta^3, \dots, \beta^{2t}$ e seus conjugados como raízes de cada um de seus polinômios. Esta propriedade, juntamente com a Definição 2.3.9 de códigos cíclicos sobre anéis \mathbb{Z}_q , nos permite especificar a seguinte matriz:

$$H = \begin{pmatrix} 1 & \beta & \beta^2 & \dots & \beta^{n-1} \\ 1 & \beta^2 & (\beta^2)^2 & \dots & (\beta^2)^{n-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \beta^{2t} & (\beta^{2t})^2 & \dots & (\beta^{2t})^{n-1} \end{pmatrix}$$

A matriz H acima é a matriz verificação de paridade para um código BCH. Note que os elementos β^i , $1 \leq i \leq 2t$ de H pertencem a G_n , e portanto, os coeficientes de β são tomados módulo n . Substituindo os elementos β^i pelos vetores linha de comprimento $r(r - uplas)$ correspondentes, temos a matriz H sobre \mathbb{Z}_q .

A construção de códigos BCH sobre anéis \mathbb{Z}_q , para $q = p^k$ e $k \geq 2$, é análoga à construção de códigos BCH sobre corpos [48]. A diferença entre essas duas construções reside no fato de que, na primeira, as raízes do polinômio gerador do código BCH encontram-se na extensão do anel \mathbb{Z}_q , ao invés de serem encontradas na extensão do corpo F_q . Vale lembrar também que iremos considerar o caso no qual $mdc(n, p) = 1$.

Podemos especificar um código BCH de comprimento n sobre \mathbb{Z}_q , onde $n = p^r - 1$, em termos das raízes de seu polinômio gerador $g(x)$, que pertencem ao subgrupo cíclico G_n . Seja β um elemento primitivo de G_n . Se $\beta^{e_1}, \beta^{e_2}, \dots, \beta^{e_j}$ são raízes de $g(x)$, então podemos gerar um código BCH com símbolos de \mathbb{Z}_q se escolhermos $g(x)$ como:

$$g(x) = \text{mmc} (M_{e_1}(x), M_{e_2}(x), \dots, M_{e_j}(x)),$$

onde $M_{e_i}(x)$ é o polinômio minimal de β^{e_i} . Além disso,

$$\bar{g}(x) = R_p(g(x)) = \text{mmc} (m_{e_1(x)}, m_{e_2(x)}, \dots, m_{e_j(x)}),$$

onde $m_{e_i}(x)$ é o polinômio minimal de $R_p(\beta^{e_i})$, gera um código BCH em $GF(p)$.

Portanto, a construção de códigos BCH cíclicos sobre o anel \mathbb{Z}_q reduz-se à escolha de elementos do subgrupo cíclico G_n para serem raízes do polinômio gerador $g(x)$.

Observação 2.3.1. *O método sistemático para o cálculo do mínimo múltiplo comum de um conjunto de polinômios $p_1(x), p_2(x), \dots, p_n(x)$ é computar o máximo divisor comum, mdc , através do Algoritmo de Euclides e então utilizar a seguinte relação:*

$$\text{mmc} (p_1(x), p_2(x), \dots, p_n(x)) = \frac{\prod_{i=1}^n p_i(x)}{\text{mdc} (p_1(x), p_2(x), \dots, p_n(x))}.$$

Os próximos teoremas estabelecem um limitante inferior para a distância de Hamming do código BCH construído:

Teorema 2.3.11. *Seja $g(x)$ o polinômio gerador de um código cíclico de comprimento n com símbolos de \mathbb{Z}_q e sejam também $\beta^{e_1}, \beta^{e_2}, \dots, \beta^{e_j}$ as raízes de $g(x)$ em G_n , onde β tem ordem n . Então, a distância mínima do código é maior que o número máximo de inteiros consecutivos módulo n no conjunto e_1, e_2, \dots, e_j .*

Teorema 2.3.12. *A distância de Hamming mínima de um código BCH satisfaz a relação:*

$$d \geq 2t + 1,$$

onde t é a capacidade de correção do código.

Note que os polinômios geradores dos códigos BCH cíclicos são construídos de forma a respeitar o limitante para a distância mínima indicado nos Teoremas 2.3.11 e 2.3.12.

No Capítulo 4, apresentaremos em detalhes a construção dos códigos BCH sobre as estruturas algébricas de anéis e corpos e suas extensões de Galois.

2.3.5 Códigos geometricamente uniformes

Forney em [42], ao estender os conceitos de códigos de Slepian e de códigos reticulados, introduziu o conceito de uma importante classe de códigos: os códigos geometricamente uniformes, definidos a seguir.

Definição 2.3.16. [42] Um conjunto de sinais \mathcal{S} definido sobre um espaço métrico (M, d) é um código **geometricamente uniforme** se, dados s_1 e s_2 em \mathcal{S} , existe uma isometria u_{s_1, s_2} que transforma s_1 em s_2 mantendo \mathcal{S} invariante, ou seja,

$$u_{s_1, s_2}(s_1) = s_2,$$

$$u_{s_1, s_2}(\mathcal{S}) = \mathcal{S}.$$

Em outras palavras, \mathcal{S} é geometricamente uniforme se a ação do grupo de simetrias $\Gamma(\mathcal{S})$ de \mathcal{S} é transitiva. Se \mathcal{S} for finito, dizemos que \mathcal{S} é uma **constelação uniforme** se \mathcal{S} for infinito dizemos que \mathcal{S} é um **arranjo regular**. Uma constelação uniforme no espaço Euclidiano é um código de grupo (código de Slepian).

Definição 2.3.17. [42] Seja \mathcal{S} um código geometricamente uniforme. Um **grupo gerador mínimo** $U(\mathcal{S})$ de \mathcal{S} , é um subgrupo do grupo de simetrias de \mathcal{S} que satisfaz

$$\forall s_0 \in \mathcal{S}, \quad \mathcal{S} = \{\mu(s_0), \mu \in U(\mathcal{S})\},$$

é a função $m : U(\mathcal{S}) \rightarrow \mathcal{S}$, dada por $m(\mu) = \mu(s_0)$ é injetora.

Teorema 2.3.13. [42] O produto cartesiano de conjuntos de sinais geometricamente uniformes é um conjunto de sinais geometricamente uniformes.

Um subgrupo normal U' de um grupo gerador mínimo $U(\mathcal{S})$ induz uma partição de um conjunto de sinais geometricamente uniformes.

Definição 2.3.18. [42] Seja \mathcal{S} um conjunto de sinais geometricamente uniformes com grupo gerador mínimo $U(\mathcal{S})$. Uma **partição geometricamente uniforme** \mathcal{S}/\mathcal{S}' , é uma partição de \mathcal{S} , induzida por um subgrupo normal U' de $U(\mathcal{S})$. Os elementos de \mathcal{S}/\mathcal{S}' são subgrupos de \mathcal{S} que correspondem às classes laterais de U' em $U(\mathcal{S})$.

Definição 2.3.19. [42] Sejam \mathcal{S}/\mathcal{S}' uma partição geometricamente uniforme e G um grupo isomorfo a $U(\mathcal{S})/U'(\mathcal{S})$. Um **rotulamento isométrico** é uma função injetora $\mathbf{m} : G \rightarrow \mathcal{S}/\mathcal{S}'$ dada pela composição do isomorfismo entre G e $U(\mathcal{S})/U'(\mathcal{S})$ e a função injetora induzida por \mathbf{m} de $U(\mathcal{S})/U'(\mathcal{S})$ em \mathcal{S}/\mathcal{S}' .

Para um código \mathcal{S} geometricamente uniforme, podemos definir, para cada ponto $s \in \mathcal{S}$, uma região formada por todos os pontos pertencentes ao espaço métrico onde está definido o código que se encontram, no mínimo, tão próximos a s quanto qualquer outro ponto $s' \in \mathcal{S}$.

Definição 2.3.20. [42] Seja \mathcal{S} um conjunto de sinais geometricamente uniforme em um espaço métrico (M, d) . A **região de Voronoi** associada a um ponto $s \in \mathcal{S}$, denotada por $\mathcal{V}(\mathcal{S})$, é o conjunto

$$\mathcal{V}_{\mathcal{S}}(s) = \{x \in M \mid d(x, s) \leq \min_{s' \in \mathcal{S}} d(x, s')\}.$$

A uniformidade geométrica é uma forma mais forte de simetria, apresentando propriedades como: a distância entre quaisquer duas palavras-código de \mathcal{S} é a mesma, todas as regiões de Voronoi são congruentes, todas as palavras-código possuem a mesma probabilidade de erro, e, o grupo gerador $U(\mathcal{S})$ é isomorfo a um grupo de permutações transitivo sobre as palavras-código. Todas essas características são buscadas na construção de novas classes de códigos, pois facilitam o processo de decodificação dos mesmos, no sentido de que não é necessário conhecer a região de decisão da cada palavra-código; basta conhecer a região de Voronoi associada a uma das palavras do código e determinar as demais região a partir de translações da região conhecida.

Com relação aos códigos já existentes utilizados em comunicações digitais, a maioria é geometricamente uniforme, como por exemplo as constelações de sinais M -PSK.

Conjunto de sinais casados a grupos

A principal motivação para considerar o codificador e o modulador como um bloco só é estabelecer a melhor forma de associar uma palavra-código a um sinal a ser transmitido. Conjunto de sinais casado a um grupo, [43], constitui a forma mais adequada de estabelecer esta associação.

Definição 2.3.21. [43] Seja (M, d) um espaço métrico. Dizemos que um **conjunto de sinais** finito \mathcal{S} em M está **casado a um grupo** G se existe uma função sobrejetora $\mu : G \rightarrow \mathcal{S}$, tal que,

$$d(\mu(g), \mu(g')) = d(\mu(g^{-1} * g'), \mu(e)), \quad \forall g, g' \in G,$$

onde e é o elemento neutro de G . A função μ é denominada **mapeamento casado**. Se μ é uma injetora, então μ^{-1} é chamada **rotulamento casado**.

Observe que neste caso temos em \mathcal{S} uma estrutura de grupo na qual as operações são isometrias.

Exemplo 2.3.2. A Figura 2.22 ilustra uma constelação de sinais do tipo 4-PSK e 8-PSK, respectivamente.

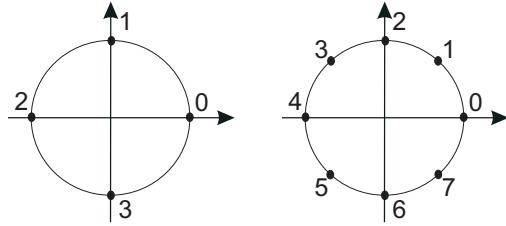


Figura 2.22: Conjunto de sinais 4-PSK e 8-PSK.

Definição 2.3.22. [43] Um mapeamento casado $\mu : G \rightarrow \mathcal{S}$, tal que H é um subgrupo de G (ou seja, $H = \mu^{-1}\mu(e)$, onde e é o elemento neutro em G), e não contém subgrupos normais não triviais de G , é chamado de **mapeamento efetivamente casado**. Nesse caso, dizemos que \mathcal{S} está **efetivamente casado** a G .

2.3.6 G -linearidade

G -linearidade é uma extensão da \mathbb{Z}_4 -linearidade centrada em grupos de simetria. Esta extensão é feita considerando-se um código quaternário mais como um rotulamento do que a imagem de um código por isometria entre módulos. Este conceito foi introduzido em [49] para códigos em espaços métricos em geral.

Todos os códigos binários não-lineares estudados em [50] são imagens de códigos lineares sobre \mathbb{Z}_4 através de um mapeamento adequado.

Para estender este mapeamento para alfabetos não necessariamente binários precisamos conhecer a estrutura do domínio e da imagem do mapeamento $\phi : \mathbb{Z}_4^n \rightarrow (\mathbb{Z}_2 \times \mathbb{Z}_2)^n$. Desse modo, temos as seguintes considerações:

- O domínio básico \mathbb{Z}_4 será visto como um grupo e a distância de Lee associada a \mathbb{Z}_4 é compatível com sua estrutura de grupo, ou seja, é uma métrica de grupo em \mathbb{Z}_4 .
- A imagem básica de $\mathbb{Z}_2 \times \mathbb{Z}_2$ será vista como um espaço métrico onde a métrica associada é a métrica de Hamming.

Tendo como base estas considerações, a questão que se coloca é a seguinte: para um grupo G (como o \mathbb{Z}_4) e um espaço métrico M (como o $\mathbb{Z}_2 \times \mathbb{Z}_2$) quais devem ser as condições de existência do mapeamento $\phi : G^n \rightarrow M^n$, como no caso da \mathbb{Z}_4 -linearidade?

A resposta a esta questão poderá fornecer uma técnica de construção de classes de códigos geometricamente uniformes sobre o alfabeto M , através de códigos de grupo sobre o grupo G . Além de ser possível a construção de códigos sob uma determinada estrutura algébrica a

partir de códigos sob uma estrutura mais adequada, permitirá também fornecer uma técnica de associação das palavras-código aos elementos do conjunto de sinais.

Definição 2.3.23. *Sejam G um grupo, d uma métrica em G e \mathcal{C} um código de comprimento n sobre o alfabeto \mathcal{A} e cuja métrica é d' . Diremos que \mathcal{C} é G -linear se \mathcal{C} , ou um código equivalente \mathcal{C}' , for imagem de um código de grupo \mathcal{C} sobre o grupo G , isto é, $\mathcal{C} = \phi(\mathcal{C})$, onde $\phi : G^n \rightarrow \mathcal{A}^n$ é uma isometria entre os espaços métricos.*

Com esta definição, temos as seguintes propriedades do código \mathcal{C} .

Proposição 2.3.3. *Se um código \mathcal{C} é G -linear, então:*

1. *O alfabeto \mathcal{A} está definitivamente casado ao grupo G , e consequentemente, o código \mathcal{C} está casado ao código de grupo correspondente obtido pelo mapeamento estendido;*
2. *O código \mathcal{C} é geometricamente uniforme.*

Encontrar o mapeamento $\phi : G \rightarrow \mathcal{A}$ é, em princípio, um problema difícil. Todavia, como o alfabeto \mathcal{A} está casado ao grupo G e ϕ é uma bijeção, a procura por este mapeamento é equivalente a determinar um subgrupo transitivo isomorfo ao grupo de simetrias de \mathcal{A} conforme o Teorema 2.3.14.

Teorema 2.3.14. [43] *Seja Θ um grupo que atua transitivamente sobre \mathcal{S} em um espaço métrico (M, d) , ou seja, \mathcal{S} é a órbita de um dado ponto sob Θ . Então \mathcal{S} está casado a Θ e, para todo $s \in \mathcal{S}$, a transformação*

$$\mu_{\mathcal{S}} : \Theta \rightarrow \mathcal{S}; \quad \mu_{\mathcal{S}}(f) = f(s).$$

O caso \mathbb{Z}_4 -linearidade

A \mathbb{Z}_4 -linearidade é um conceito inovador e importante em teoria da codificação porque possibilita que certas classes de códigos não-lineares de comprimento par possam ser vistos como códigos lineares sobre \mathbb{Z}_4 . Obtém-se, com isso, uma redução significativa na complexidade do processo de decodificação dos códigos não-lineares. A seguir apresentaremos os principais conceitos relacionados com a \mathbb{Z}_4 -linearidade.

Definição 2.3.24. [50] *Um mapeamento é uma função*

$$\phi : \mathbb{Z}_4^n \rightarrow \mathbb{Z}_2^{2n},$$

definida por

$$\phi(\underline{c}) = (\beta(\underline{c}), \gamma(\underline{c}), \dots, (\beta(c_1), \beta(c_2), \dots, \beta(c_n), \gamma(c_1), \gamma(c_2), \dots, \gamma(c_n)), \forall \underline{c} = (c_1, \dots, c_n) \in \mathbb{Z}_4^n;$$

onde as funções $\beta : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2$ e $\gamma : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2$ são especificadas na Tabela 2.2.

| c | $\beta(c)$ | $\gamma(c)$ |
|-----|------------|-------------|
| 0 | 0 | 0 |
| 1 | 0 | 1 |
| 2 | 1 | 1 |
| 3 | 1 | 0 |

Tabela 2.2: Definição das funções β e γ .

O mapeamento ϕ foi obtido com a característica de associar um código quaternário linear a um código binário que, mesmo não sendo necessariamente linear, satisfaz a propriedade de que o perfil de distâncias é o mesmo para qualquer palavra-código considerada. Em outras palavras, o mapeamento da \mathbb{Z}_4 -linearidade transportou de uma forma bastante sutil para o código binário a “linearidade” do código quaternário. Esta propriedade é uma consequência imediata da uniformidade geométrica do código binário.

Para que esta condição seja satisfeita, é necessário que a distância de Lee seja compatível com a operação do grupo \mathbb{Z}_4 e que haja uma isometria entre os espaços considerados. Isto implica diretamente na propriedade de casamento (no sentido proposto por Loeliger [43]) entre o grupo \mathbb{Z}_4 e o conjunto de sinais $\mathbb{Z}_2^2 = \mathbb{Z}_2 \times \mathbb{Z}_2$.

O caso $\mathbb{Z}_2 \times \mathbb{Z}_2$

Mais precisamente, considere \mathbb{Z}_2^2 como o espaço métrico de Hamming bidimensional de ordem 2. Note que a propriedade relevante neste caso é o casamento efetivo entre \mathbb{Z}_4 e \mathbb{Z}_2 . Em [43], temos que \mathbb{Z}_2^2 é casado a \mathbb{Z}_4 e a função

$$\mu_{\mathbb{Z}_2^2} : \mathbb{Z}_4 \rightarrow \mathbb{Z}_2^2, \quad \mu_{\mathbb{Z}_2^2}(f) = f(s)$$

é um rotulamento casado para todo $s \in \mathbb{Z}_2^2$. Escolhendo o ponto inicial $s = (0, 0)$ obtemos o rotulamento isométrico, (Tabela 2.3):

A escolha do ponto inicial dependerá da aplicação sendo considerada. No caso em consideração, isto é, \mathbb{Z}_4 e \mathbb{Z}_2 , o interesse é relacionar códigos sobre \mathbb{Z}_4 com códigos sobre \mathbb{Z}_2 . Como em \mathbb{Z}_4 a distância de interesse é a distância de Lee e em \mathbb{Z}_2 a distância natural é a distância de

| | | |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 1 | 0 |
| 2 | 1 | 1 |
| 3 | 0 | 1 |

Tabela 2.3: Mapeamento por rotulamento isométrico.

Hamming vemos que o rotulamento preserva o perfil de distâncias e o peso de cada elemento em cada espaço. Assim temos uma isometria entre os espaços métricos (\mathbb{Z}_4, d_{Lee}) e (\mathbb{Z}_2^2, d_H) .

Exemplo 2.3.3. Seja (\mathbb{Z}_2^n, d_H) o espaço métrico como mostra a Figura 2.3.3. Este é o espaço de Hamming bidimensional. Assim, o grupo de simetrias de \mathbb{Z}_2^n é

$$\Gamma(\mathbb{Z}_2^n) \cong \mathbb{D}^4 \supseteq \mathbb{Z}_4.$$

Além disso, \mathbb{Z}_4 atua transitivamente sobre \mathbb{Z}_2^n e, portanto, \mathbb{Z}_2^n está efetivamente casado a \mathbb{Z}_4 . Se em \mathbb{Z}_4 utilizarmos a métrica de Lee, então concluímos que (\mathbb{Z}_4, d_L) é isométrico a (\mathbb{Z}_4, d_H) segundo o rótulo escolhido a partir do ponto inicial $(0,0)$. Portanto, existem códigos \mathbb{Z}_4 -lineares. Em [50], algumas classes importantes de códigos binários não-lineares são obtidas através de códigos quaternários lineares.

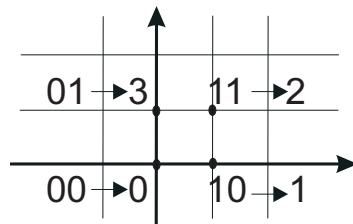


Figura 2.23: Rotulamento do espaço de Hamming bidimensional de ordem 2.

Capítulo 3

Modelo de Codificação Genética e Genômica

Por tratar-se de um assunto interdisciplinar, neste capítulo, abordamos três temas básicos. O primeiro, apenas introdutório, relaciona o sistema de comunicação digital ao sistema biológico. O segundo estabelece os elementos essenciais para a determinação e a caracterização dos modelos de codificação genética e genômica semelhantes ao do sistema de comunicação digital, e o terceiro sugere uma proposta de armazenamento e transmissão da informação genômica análoga àquela usada em redes locais de computadores.

Este capítulo está organizado da seguinte maneira: inicialmente, na Seção 3.1, fazemos uma analogia entre o sistema de comunicação digital e o sistema biológico, importante para o entendimento e estabelecimento dos modelos de codificação genética e genômica descritos na Seção 3.2. Em seguida, nas Subseções 3.2.1 e 3.2.2, apresentamos os modelos de codificação genética e codificação genômica, responsáveis pela identificação e reprodução da informação genética contida na sequência de nucleotídeos de uma fita simples de DNA e da informação genômica contida na dupla hélice do DNA, respectivamente. A caracterização destes modelos está relacionada ao processo de identificação da estrutura matemática tanto do alfabeto quanto do código corretor de erros com o objetivo de reproduzir, identificar e classificar sequências de DNA. Finalmente, na Seção 3.3, sugerimos que o processo de armazenamento da informação genômica ocorre de maneira análoga àquele utilizado em redes de computadores (descrição do quadro utilizado na transmissão da informação). Neste quesito, na Subseção 3.3.1, estamos interessados em relacionar os processos de armazenamento, organização, transmissão e identificação da informação biológica aos protocolos usados em sistemas de redes locais de computadores. Dessa forma, propomos a arquitetura biológica (*Biological frame*) do genoma humano sugerindo que a informação genômica pode ser armazenada e organizada de maneira análoga as informações que são armazenadas e organizadas em

CD's e, que o armazenamento e a transmissão da informação genômica ocorrem de maneira inversa ao procedimento de transmissão de dados utilizado em redes de computadores. Na Subseção 3.3.2, propomos uma formatação da arquiteruta genômica relacionada ao sequenciamento do genoma de um plasmídeo, denominado *Biological frame of Lactococcus lactis plasmid pcl 2.1 Genomic*.

3.1 Analogia entre o Sistema de Comunicação Digital e o Sistema de Informação Genômica

O objetivo desta seção é apontar as semelhanças existentes entre o sistema de comunicação e o sistema de informação genômica. Iniciamos com alguns exemplos de aplicações pertinentes aos sistemas de comunicações.

A Engenharia de Telecomunicações é uma área de especialização da Engenharia Elétrica e da Engenharia Eletrônica responsável por estudo, especificação, projeto, implementação e manutenção de uma variedade de sistemas de comunicações a longa distância através de equipamentos e sistemas elétricos, eletrônicos e ópticos. Sua abrangência se estende pelos seguintes ramos de aplicação e seus respectivos exemplos mais comuns: sistemas de telefonia móvel e fixa; Sistemas de propagação via rádio; sistemas de comunicações via satélite; sistemas de comunicações ópticas; redes de telecomunicações.

Na Seção 2.1 do Capítulo 2 abordamos, de maneira resumida, algumas das principais e importantes funções que as células exercem, dentre elas a capacidade de receber e emitir informações monitorando continuamente a vizinhança celular e ajustando suas atividades e sua composição de acordo com a necessidade. Dessa maneira, os diversos sistemas biológicos podem ser interpretados e estudados como um sistema de comunicação.

Como uma tendência e uma nova frente de pesquisa, num futuro muito próximo, também poderá ser citado como mais um exemplo das várias aplicações em sistema de comunicação o **sistema biológico ou sistema de informação genômica**. Tendo em vista a riqueza de propriedades e características que envolvem o sistema biológico com o objetivo da continuidade da vida, será possível no futuro aprender com esse sistema e, quem sabe, exemplos práticos e mais eficientes possam ser tirados do mundo biológico e usados no mundo das comunicações, e vice-versa.

3.1.1 Sistema de comunicação digital

Devido ao fato de que a motivação para o desenvolvimento dos CCEs ser o da confiabilidade na transmissão da informação, a realização e a aplicação dessa teoria encontra-se

na teoria das comunicações. Aplicações em problemas de comunicações são diversificadas. Dados binários são comumente transmitidos entre terminais de computadores, entre aeronaves e entre espaçonaves. CCEs são usados frequentemente em aplicações militares para proteção contra interferência inimiga intencional. As transmissões entre sistemas computacionais usualmente são intolerantes até mesmo a baixas taxas de erros, porque um simples erro pode alterar um programa de computador.

Podemos considerar um **sistema de comunicação** como sendo um conjunto de equipamentos, meios físicos, ou até mesmo um organismo, que tem por objetivo transmitir a informação de uma **fonte** a um **destinatário** via um **canal** de comunicação. Se o canal não tem ruído, então a informação enviada será recebida sem alteração, porém, na prática, interferências e ruídos, em menor ou maior grau, estão presentes e sendo adicionados à informação, resultando na introdução de erros. De um modo geral, podemos trabalhar com dois tipos de sistema de comunicação:

1. **Sistemas analógicos** - São aqueles que conservam a forma dos sinais desde a fonte ao destinatário;
2. **Sistemas digitais** - São aqueles em que a forma do sinal transmitido pode ser diferente do sinal original, por exemplo: a forma do sinal pode variar em amplitude e/ ou fase e/ ou frequência em intervalos fixos de tempo.

Nos últimos anos, tem havido uma procura crescente de serviços altamente eficientes e viáveis para os sistemas de armazenamento e de transmissão de dados digitais. Esta exigência foi acelerada em virtude da necessidade, em larga escala e de alta velocidade, do processamento, armazenamento e transmissão da informação na forma digital nas esferas comercial, governamental e militar. Uma preocupação importante do sistema é o controle dos possíveis erros para que os dados possam ser reproduzidos fielmente.

O modelo do sistema de comunicação digital representado através de um diagrama de blocos, Figura 3.1, será descrito a seguir. Este sistema conecta uma fonte a um destinatário, onde, cada um dos blocos componentes de um sistema de comunicação típico é definido da seguinte maneira:

- **Fonte:** gerador da informação a ser transmitida (máquina, ser humano ou até mesmo um organismo), podendo gerar um sinal ou uma sequência de símbolos discretos;
- **Codificador de Fonte:** converte o sinal da saída da fonte em uma sequência de dígitos binários. Se a fonte é contínua, uma conversão A/D (análogica para digital) é feita. É projetado de forma a minimizar o número de bits por unidade de tempo necessário para

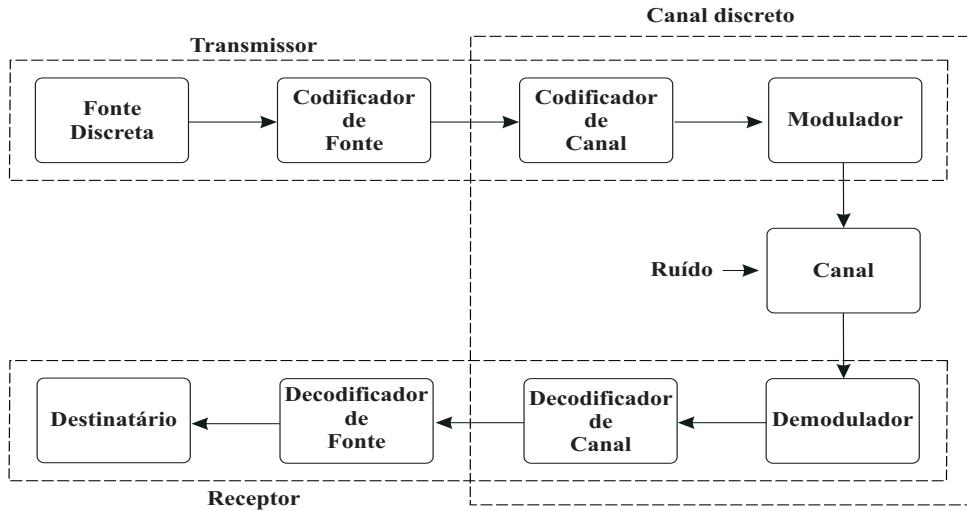


Figura 3.1: Diagrama de blocos de um sistema de comunicação.

representar o sinal de saída da fonte e de forma que este sinal possa ser reconstituído sem ambiguidades;

- **Codificador de Canal:** transforma a sequência da saída do codificador de fonte em uma sequência codificada (**palavra-código**) pela adição de redundância para combater os efeitos do ruído introduzido através do canal. Cada símbolo na palavra-código é representado por bits (dígitos binários) no caso de sinalização binária. Caso use-se mais do que dois sinais (por exemplo q sinais), não temos bits, e sim dígitos de um alfabeto q -ário;
- **Modulador:** converte a saída do codificador de canal em uma forma de onda adequada para a transmissão através do canal. O modulador digital transforma símbolos discretos da saída do codificador de canal em um sinal contínuo com duração t segundos, de tal forma que a amplitude e/ ou frequência e/ ou fase seja(m) alterada(s) de acordo com a necessidade. Algumas destas técnicas são conhecidas como:
 - **PAM** (*pulse amplitude modulation*) ou **ASK** (*amplitude shift-keying*): alteração de amplitude.
 - **FSK** (*frequency shift-keying*): alteração de frequência.
 - **PSK** (*phase shift-keying*): alteração de fase.
 - **QAM** (*quadrature amplitude modulation*): alteração de amplitude e fase;
- **Canal:** meio físico que conecta o transmissor ao receptor. O sinal modulado a ser transmitido através do canal pode ou não sofrer a ação do ruído;

- **Demodulador:** a partir do sinal recebido do canal, estima o sinal transmitido e envia para o decodificador de canal a versão digital correspondente;
- **Decodificador de Canal:** tenta corrigir possíveis erros (se houver) nos dígitos fornecidos pelo demodulador, produzindo uma estimativa dos dígitos na saída do codificador de fonte;
- **Decodificador de Fonte:** transforma a sequência estimada na saída do decodificador de canal em uma estimativa na saída da fonte. Se a fonte for contínua este processo pode envolver uma conversão D/A (digital para analógico);
- **Destinatário:** receptor da informação transmitida (máquina, ser humano ou até mesmo um organismo).

3.1.2 Sistema de informação genômica

Elementos químicos envolvidos na transmissão de caracteres hereditários e na produção de proteínas são os principais constituintes dos seres vivos. Estes elementos químicos são os ácidos desoxirribonucléico e ribonucléico. De acordo com a moderna Biologia, o **DNA** fabrica **RNA**, que fabrica **proteína** conhecido como um paradigma denominado **dogma central da biologia molecular** (embora existam exceções, os retrovírus, como o vírus da Aids).

O DNA é uma molécula formada por duas cadeias na forma de uma dupla hélice. Essas cadeias são constituídas por um açúcar (desoxirribose), um grupo fosfato e uma base nitrogenada: (T) timina, (A) adenina, (C) citosina ou (G) guanina. A dupla hélice é essencial na replicação do DNA durante a divisão celular, pois, cada hélice serve de molde para a construção de uma nova. O RNA é uma molécula também formada por um açúcar (ribose), um grupo fosfato e uma base nitrogenada: (U) uracila, (A) adenina, (C) citosina ou (G) guanina. Um grupo reunindo um açúcar, um fosfato e uma base é denominado “nucleotídeo”.

A informação genômica é perpetuada através da **replicação** do DNA e é estabelecida através de dois processos: A **transcrição** que converte a informação do DNA em uma forma mais acessível (uma fita de RNA) e a **tradução** que converte a informação contida no RNA em proteínas. A seguir descrevemos, de forma sucinta, como ocorrem os processos de duplicação, transcrição e tradução, respectivamente.

Como o DNA se duplica

Para o DNA duplicar-se (ou replicar), há necessidade de uma enzima especial, a DNA polimerase. Estando presente essa enzima, ocorrem as seguintes etapas:

1. As pontes de hidrogênio que ligam as bases nitrogenadas rompem-se e as duas fitas se afastam;
2. Nucleotídeos livres, já existentes na célula, encaixam-se nas duas fitas que se afastaram. O encaixe só ocorre se as bases forem complementares, adenina com timina (A-T ou T-A) e, citosina com guanina (C-G ou G-C);
3. Quando as duas fitas originais forem completadas por novos nucleotídeos, teremos duas moléculas de DNA idênticas entre si.

Em cada molécula, existe um filamento antigo, que pertencia à molécula-mãe, e um novo, que se formou sobre o antigo. Cada filamento antigo atuou como molde, já que sua sequência de bases funcionou como “guia” para a produção da fita nova. O processo de duplicação é também denominado semi-conservativo, já que cada molécula-filha conserva metade da molécula-mãe.

Como o DNA fabrica o mRNA - transcrição

O DNA produz moléculas de **mRNA** (RNA mensageiro), que migram para o citoplasma e controlam a construção das proteínas, aminoácido por aminoácido, garantindo a produção daquela proteína especial no momento correto. A sequência de DNA é que condiciona a sequência da molécula de RNA. Uma diferença importante em relação à duplicação é que apenas **uma fita** de DNA funciona como molde. O RNA produzido será, portanto, uma fita simples e não dupla. Esse processo segue os seguintes passos:

1. É necessária a presença de uma enzima: a RNA polimerase;
2. As pontes de hidrogênio se desfazem e as duas fitas de DNA se afastam;
3. Nucleotídeos livres de RNA encaixam-se apenas numa das fitas, chamada **fita ativa**;
4. A molécula de RNA (fita única) destaca-se de seu molde de DNA e migra para o citoplasma;
5. As duas fitas de DNA tornam a parear, reconstituindo a molécula original.

Os **genes**, forma de unidades consistindo dos elementos do código genético, está no DNA, no núcleo da célula. Já a “fábrica” de proteínas fica no citoplasma celular em estruturas específicas, os **ribossomos**, para onde se dirige o RNA mensageiro. Na transcrição, apenas o gene relacionado à proteína que se quer produzir é copiado na forma de RNA mensageiro. No ribossomo, o RNA mensageiro é lido por moléculas de RNA de transferência, responsável pelo transporte dos aminoácidos até o local onde será montada a cadeia protéica.

Síntese de proteínas - tradução

O DNA presente no núcleo controla toda a síntese de proteínas da célula. Esse controle é efetuado por meio de moléculas de RNA que o DNA fabrica e que passam para o citoplasma. A correspondência entre o DNA e o RNA ocorre base por base, sendo que, entra a uracila no RNA no lugar da timina (proveniente do DNA). Na correspondência entre RNA e proteína, cada três bases do RNA codifica um aminoácido específico da proteína. Cada trinca de bases no DNA ou no RNA é denominada **códon**, onde as trincas representam “palavras” do código genético, cada “palavra” corresponde a um “objeto”, no caso o aminoácido, como mostra a Tabela 3.1. O **código genético** é constituído pelos 64 possíveis códons.

| 1 ^a base | 2 ^a base | | | | 3 ^a base | | | |
|------------------------|--------------------------|-------------------------------|--------------------------|--------------------------|-------------------------------|--------------------------|----------------------------|------------------|
| | U | C | A | G | | | | |
| U | UUU UUC UUA UUG | Phe - F Ser - S Leu - L | UCU UCC UCA UCG | UAU UAC UAA UAG | Tyr - Y stop | UGU UGC UGA UGG | Cys - C stop Trp - W | U C A G |
| | CUU CUC CUA CUG | Leu - L | CCU CCC CCA CCG | CAU CAC CAA CAG | His - H Pro - P Gln - Q | CGU CGC CGA CGG | Arg - R | U C A G |
| | AUU AUC AUU AUG | Ile - I Met - M | ACU ACC ACA ACG | AAU AAC AAA AAG | Asn - N Thr - T Lys - K | AGU AGC AGA AGG | Ser - S Arg - R | U C A G |
| | GUU GUC GUA GUG | Val - V | GCU GCC GCA GCG | GAU GAC GAA GAG | Asp - D Ala - A Glu - E | GGU GGC GGA GGG | Gly - G | U C A G |

- | | | | |
|---|--|--|---|
| █ Aminoácidos não-polares F = Fenilalanina L = Leucina I = Isoleucina M = Metionina V = Valina P = Prolina A = Alanina W = Triptofano | █ Aminoácidos polares S = Serina T = Treonina Y = Tirosina Q = Glutamina N = Asparagina C = Cisteína G = Glicina | █ Aminoácidos básicos H = Histidina K = Lisina R = Arginina | █ Aminoácidos ácidos D = Ácido Aspártico E = Ácido Glutâmico |
|---|--|--|---|

Tabela 3.1: Código genético.

No código genético existem **códons de finalização** (UAA, UGA e UAG) que indicam à célula que a sequência de aminoácidos destinada àquela proteína deve ser finalizada. Existe ainda um **códon de iniciação** (AUG) que indica que a sequência de aminoácidos da proteína deve ser inicializada. Este códon (AUG) codifica o aminoácido metionina (M) de forma que todas as proteínas começam com esse amonoácido. Observe na Tabela 3.1 que existem 64 possíveis trincas, ou códons, que correspondem a apenas 20 aminoácidos. Assim, é fácil entender que mais de um códon pode corresponder a um mesmo aminoácido e, é por isso que o código genético é dito **degenerado** ou **redundante**. Note que existe mais de uma

trinca associada para determinados aminoácidos. Apenas a metionina e o triptofano (W) são codificados por um único códon, representados por AUG e UGG, respectivamente. A glicina (G), por exemplo, é codificada por GGG, GGC, GGA e GGU. O código genético é **universal** devido ao fato de a mesma trinca codificar o mesmo aminoácido em quase todos os organismos.

3.1.3 Analogias

O diagrama de blocos de um sistema de comunicação como mostrado na Figura 3.1 representa de maneira simplificada o **dogma central da teoria de comunicação**, Figura 3.2

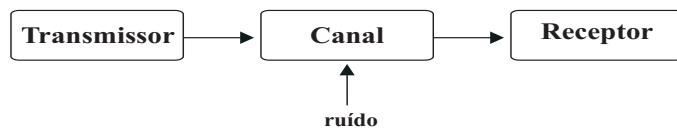


Figura 3.2: Dogma central da teoria de comunicação.

Do **dogma central da biologia molecular** temos que o **DNA** implica em **RNA**, que por sua vez implica em **Proteína** e, portanto pode ser caracterizado através de um diagrama de blocos mostrado na Figura 3.3.



Figura 3.3: Dogma central da biologia molecular.

Fazendo uma analogia entre o dogma central da teoria de comunicação e o dogma central da biologia molecular, temos que:

1. Em um sistema de comunicação, o responsável pela geração das informações a serem transmitidas é o transmissor. Biologicamente, quem exerce esta mesma função é o DNA;
2. Os processos de transcrição/tradução têm como objetivo a transmissão da informação. Durante estes processos podem ocorrer alguns erros que irão interferir na informação, como a não-leitura de um códon como consequência da perda do pareamento do RNA de transferência. Do ponto de vista da comunicação, podemos visualizar os processos de

transcrição e tradução como sendo os processos de codificação de canal e de modulação em um sistema de comunicação, e os eventuais erros cometidos durante estes processos como sendo o ruído introduzido no canal;

3. O receptor pode ser modelado como o local onde a informação está sendo enviada. Neste caso específico, a informação é uma proteína.

A Figura 3.4 ilustra a analogia entre os dois dogmas através de um diagrama de blocos, estando cada bloco no sistema de comunicação relacionado com cada bloco no sistema de informação genômica.

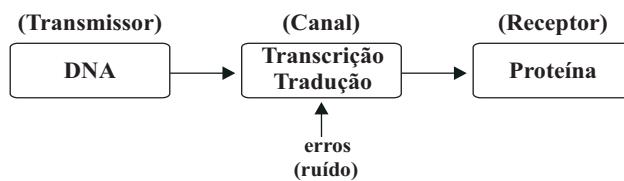


Figura 3.4: Analogia entre o sistema de informação genômica e o sistema de comunicação simplificado.

Diante das semelhanças apontadas anteriormente, fica claro que diversos sistemas biológicos podem ser modelados como sistemas de comunicação, uma vez que o sistema biológico armazena e transmite a informação. Tendo como base essas considerações, caracterizamos o sistema de transmissão da informação genética (em termos da fita simples do DNA) e o sistema de transmissão da informação genômica (em termos da fita dupla do DNA), ambos, análogos ao sistema de comunicação digital (Figura 3.5), através das seguintes associações:

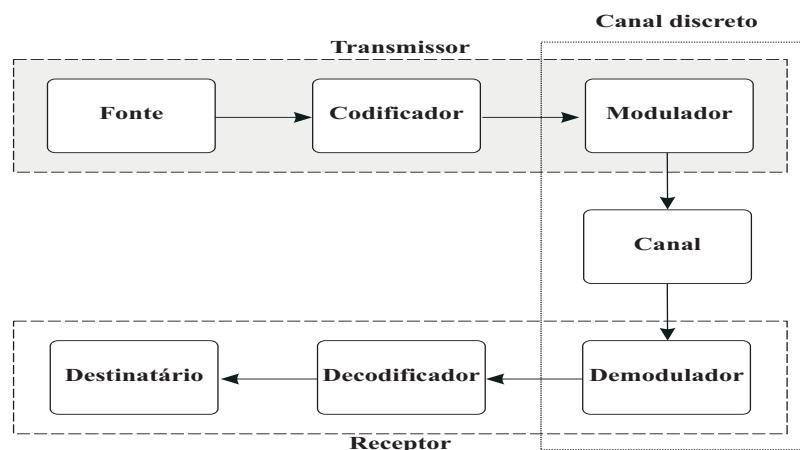


Figura 3.5: Modelo de um sistema de comunicação digital.

1. **Fonte:** Em um sistema de comunicação, a fonte é o lugar onde a mensagem é gerada. Em um sistema biológico, entretanto, o DNA e o RNA mensageiro são responsáveis pela geração e a transmissão da informação e genética, respectivamente;
2. **Transmissor:** Caracterizado mais precisamente, pelos processos de transcrição e tradução (sequência de nucleotídeos (nt)) que têm como objetivo garantir a continuidade da informação genômica e genética a ser transmitida;
3. **Canal:** É o meio pelo qual a informação é transmitida em um sistema de comunicação, e onde erros podem ocorrer durante a transmissão da informação. Os erros que podem ocorrer durante os processos de transcrição/tradução serão considerados no citosol, caracterizado pelo canal;
4. **Receptor:** O receptor pode ser interpretado como qualquer compartimento intracelular ou extracelular, representando o local para onde a informação está sendo enviada. Neste caso específico, a informação é dada pelas várias sequências de DNA que compõem o genoma.

O bloco transmissor do modelo de um sistema de comunicação digital, ilustrado na Figura 3.5, é o objeto de consideração neste trabalho e está relacionado ao sistema de transmissão da informação genética e genômica. Com isso, propomos os modelos de codificação genética e genômica descritos a seguir.

3.2 Modelos de Codificação Genética e Genômica

No Capítulo 2, vimos que os CCEs são utilizados sempre que se deseja transmitir ou armazenar informação. Uma questão sempre presente em trabalhos relacionados com codificação genética e codificação genômica é se existe alguma forma de CCE na estrutura do DNA. Em [1], Battail considera a identificação de algum CCE que seja capaz de reproduzir uma determinada sequência de DNA como um problema que não tem solução. Neste trabalho, apontamos um encaminhamento positivo a esse problema, dando continuidade à proposta descrita em [20]. Até onde é de nosso conhecimento, pela primeira vez sequências de nucleotídeos e dos correspondentes aminoácidos de uma fita simples do DNA com características biológicas distintas e comprimentos variados (incluindo sequências de genoma plasmidial, gene humano, DNA repetitivo, ítron, RNA, mRNA, proteína, hormônio, e, sequências de direcionamento e de sinal interno de proteínas organelares) são identificadas como palavras-código de CCEs. Outro avanço é com relação à identificação da sequência das bases complementares da dupla hélice do DNA como palavra-código de um CCE.

Mediante as analogias citadas na seção anterior, propomos um modelo de um sistema de comunicação de informação genética e um modelo de um sistema de comunicação de informação genômica análogos ao modelo de um sistema de comunicação digital, Figuras 3.6 e 3.7, respectivamente. O modelo de um sistema de comunicação de informação genética está diretamente relacionado com o desmembramento do canal discreto, podendo ter ou não memória, sendo similar àquele utilizado em sistemas de comunicação digital eficientes em faixa e potência (codificação combinada com modulação). A especificação do modulador e do demodulador fundamenta o contexto da informação biológica trafegar de um ponto para outro (intracelular ou intercelular), por exemplo: o transporte de proteínas entre o núcleo de uma célula eucarótica e uma organela. O codificador e o modulador, juntos, são responsáveis pela identificação e reprodução da informação genética contida no RNA mensageiro (mRNA).

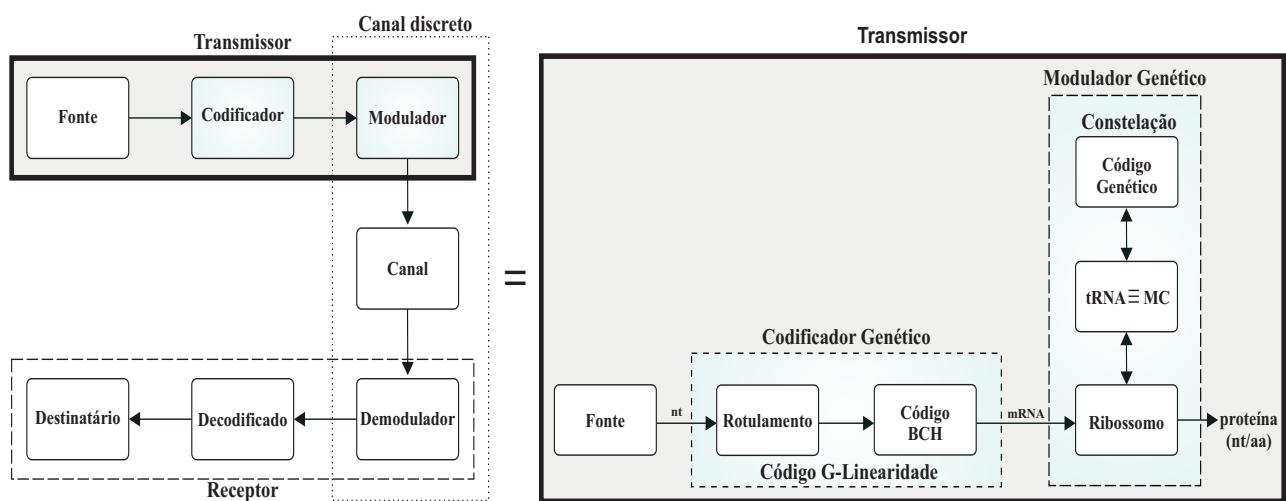


Figura 3.6: Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genética.

Logo, este modelo será usado no processo de identificação e reprodução de sequências de uma fita simples do DNA, tais como: sequências de direcionamento, sinais internos de uma proteína, hormônio, enzima e proteína, sendo que todas essas sequências serão analizadas na direção 5' – 3' e suas respectivas sequências complementares na direção 3' – 5', separadamente.

O modelo de um sistema de comunicação de informação genômica tem a ver somente com o bloco do codificador de canal. O modulador, canal e demodulador passam a ser uma única entidade chamada “canal discreto podendo ter ou não memória”. Sendo assim, o processo de codificação está relacionado com os possíveis erros que serão introduzidos por esse canal. Logo, este modelo será responsável pela identificação e reprodução da informação genômica contida dupla hélice do DNA (as duas fitas do DNA 5' – 3' e 3' – 5' analisadas conjuntamente).

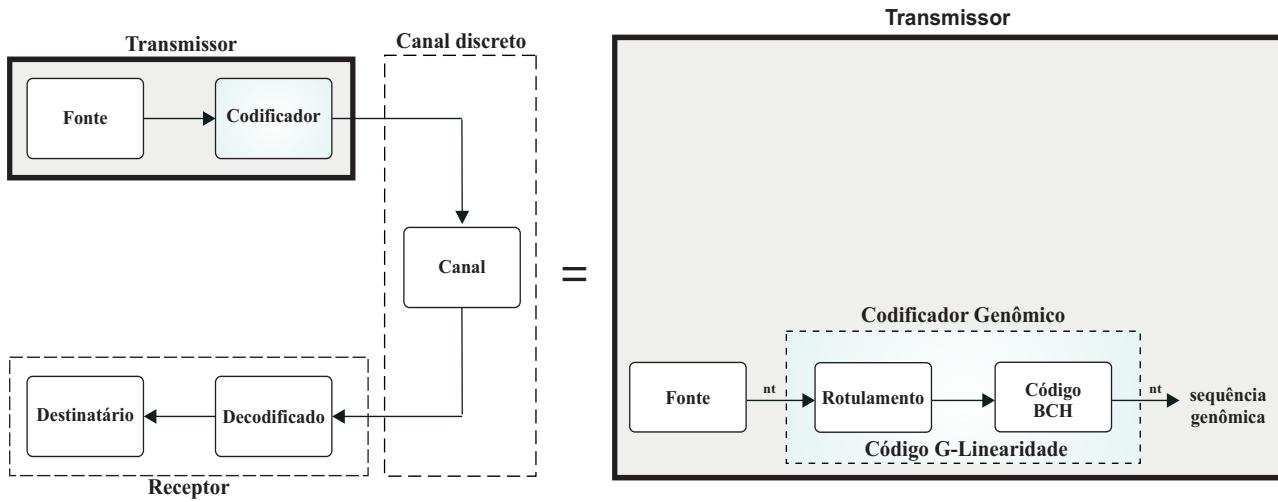


Figura 3.7: Modelo de um sistema de comunicação digital e o modelo de um sistema de comunicação de informação genômica.

Os modelos propostos baseiam-se na seguinte hipótese: se o genoma é constituído por regiões consistindo de genes, exons, íntrons, sequências de direcionamento, sequências de sinais internos de proteínas, DNAs repetitivos, micro RNAs, proteínas, etc; então cada uma dessas regiões pode ser reproduzida por um código específico. Sendo assim, o genoma consiste de códigos entrelaçados e ao invés de iniciarmos a análise do genoma como um todo, focalizamos primeiramente em suas partes em termos da fita simples do DNA (genomas eucarióticos e procarióticos). Em seguida, focalizamos em um genoma procarioto como um todo e, por último, focalizamos na dupla hélice do DNA, da mesma maneira, inicialmente em suas partes (genomas eucarióticos e procarióticos) e posteriormente em um genoma procarioto como um todo. Neste processo, naturalmente surgem as seguintes perguntas:

- 1) Dentre os diversos CCEs usados para a transmissão da informação, existe algum CCE capaz de reproduzir sequências de DNA (regiões do genoma em termos da fita simples do DNA) e suas respectivas fitas complementares?
- 2) Existe algum CCE capaz de reproduzir a dupla hélice do DNA?
- 3) Se existe, qual será a estrutura matemática apropriada para a construção deste CCE?
- 4) Como os alfabetos dos códigos genético e genômico devem ser rotulados no processo de identificação da estrutura matemática?

Diante destas perguntas, iniciamos o processo de caracterização desses modelos.

3.2.1 Modelo de codificação genética

Neste modelo encontramos um **codificador genético** e um **modulador genético**. A palavra-código na saída do codificador está relacionada à sequência de nucleotídeos, e na saída do modulador está relacionada à sequência de aminoácidos (proteína), Figura 3.8. Embora o mapeamento do código genético realizado pelo RNA transportador seja bem conhecido no contexto biológico, o mesmo necessita de uma caracterização matemática no contexto de um sistema de comunicação digital, onde os 64 possíveis códons (as trincas) representam os sinais da constelação de sinais.

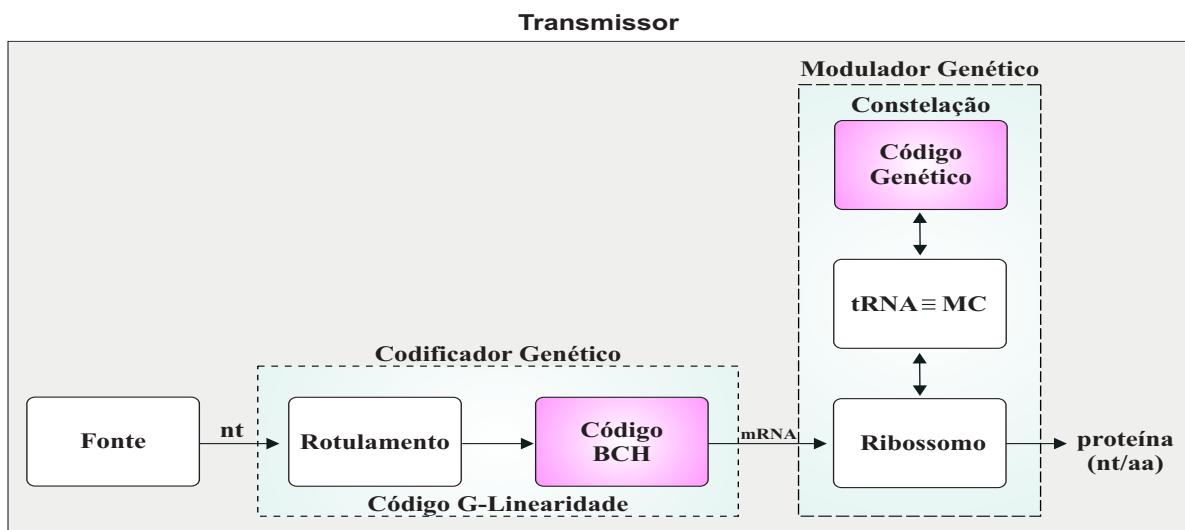


Figura 3.8: Modelo de um sistema de comunicação de informação genética.

Em um sistema de comunicação digital, quando as estruturas algébricas do codificador e do modulador (constelação de sinais) são isomórficas, então diz-se que esse isomorfismo é um **mapeamento casado** (MC). A classe de códigos satisfazendo essa propriedade é bem conhecida e denominada **códigos geometricamente uniformes**. Uma subclasse importante é a **G-linearidade**, onde G denota uma estrutura algébrica, que incorpora todas as vantagens inerentes ao processo de codificação dos códigos lineares, bem como, a dos códigos não lineares através da inserção do bloco rotulamento. Dependendo da classificação desse mapeamento como linear ou não-linear, o código resultante será linear ou não-linear, respectivamente.

O **codificador genético** consiste de um mapeamento associado a um código corretor de erros (bloco rotulamento e bloco código BCH). Logo, os códigos resultantes desse codificador pertencem a subclasse da G -linearidade onde G é \mathbb{Z}_4 -linear ou $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear ou Klein-linear. Como consequência, os códigos resultantes do codificador genético são denominados: **código \mathbb{Z}_4 -linearidade** ou **código $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade** ou **código Klein-linearidade**. O código

\mathbb{Z}_4 -linearidade é um código não-linear e os códigos $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade são códigos lineares, respectivamente. O **modulador genético** consiste do código genético, do RNA transportador e do ribossomo. O código genético pode ser visto como uma **constelação de sinais**, onde cada códon é considerado como um sinal nessa constelação, o RNA transportador realiza a operação de mapeamento casado, enquanto que o RNA ribossômico se comporta como um processador digital de sinais, dando origem à proteína.

A seguir descrevemos o desdobramento do bloco codificador genético, com o objetivo de facilitar o entendimento do processo de caracterização do modelo de um sistema de comunicação de informação genética, ilustrado na Figura 3.8.

Desdobramento do Codificador Genético

Dentre os diversos CCEs usados para a transmissão da informação (ilustrados na Figura 2.21 do Capítulo 2), os códigos BCH [51], usados na transmissão de informação de pacotes em redes de computadores e geração de sequências, formam uma importante classe de códigos cíclicos devido, principalmente, à simplicidade dos processos de codificação e decodificação associados, os que os tornam também bons candidatos a serem utilizados na aplicação para a geração de sequências de DNA, sendo que as estruturas matemáticas mais utilizadas para a construção destes códigos são as estruturas algébricas de corpo e anel e suas extensões de Galois [47] e [52].

Sendo assim, iniciamos o processo de construção do código BCH com o objetivo de identificar e reproduzir diferentes sequências de DNA. Primeiramente, construímos o código BCH sobre a estrutura algébrica de anel. Em seguida, construímos o código BCH sobre a estrutura algébrica de corpo e suas extensões de Galois, respectivamente.

À princípio, buscávamos a identificação e a reprodução das sequências de DNA apenas através da construção do código BCH sobre as estruturas algébricas de corpo e anel e suas extensões de Galois, respectivamente. Neste procedimento de construção do código BCH, utilizamos todos os polinômios primitivos e seus respectivos polinômios geradores relacionados às extensões de Galois de grau r e, todas as possíveis distâncias mínimas dos códigos, d_H . No entanto, os resultados desse procedimento de identificação e reprodução das sequências de DNA, nos apontaram que apenas determinados polinômios (primitivos e geradores) com a distância mínima $d_H = 3$ e, em determinados rotulamentos é que foram capazes de identificar e reproduzir tais sequências. Com isso, ficou claro a existência de um mapeamento casado entre o código genético (modulador) e o código BCH (codificador) (ilustrados na cor lilás na Figura 3.8).

Portanto, os elementos, tais como: a estrutura algébrica, o alfabeto, o rotulamento, o mapeamento, o polinômio primitivo e o polinômio gerador, devem ser considerados. A seguir,

descrevemos a composição desses elementos que foram fundamentais e determinantes na caracterização dos CCEs resultantes pertencentes a subclasse G -linearidade.

Primeiro - Determinação da Estrutura Algébrica do Alfabeto do Código Genético

Como a estrutura algébrica do alfabeto do código genético é desconhecida se faz necessária a identificação de tal estrutura pois a mesma, como usual em teoria da codificação, será utilizada na construção do CCE. Nesse processo de identificação da estrutura algébrica são pertinentes:

a) O Alfabeto do Código Genético e o Alfabeto do Código BCH:

No processo de codificação genética relacionamos o alfabeto do código genético em termos dos nucleotídeos: adenina (A), citosina (C), guanina (G) e timina (T) ou uracila (U), denotado pelo conjunto $N = \{A, C, G, T/U\}$ aos alfabetos 4-ários de CCEs, denotados por $GF(4) = \{0, 1, \alpha, \alpha^2\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como mostrado na Figura 3.9.

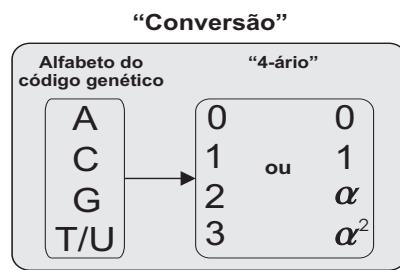
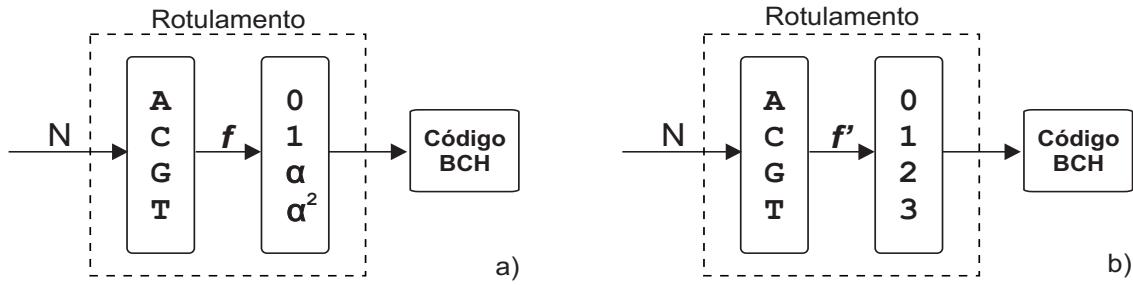


Figura 3.9: O alfabeto do código genético e os alfabetos 4-ários dos CCEs.

O alfabeto de um CCE, em geral, é especificado de acordo com a aplicação em consideração de modo que, a estrutura matemática associada facilite o processo de codificação e decodificação. Como a estrutura algébrica do alfabeto do código genético das sequências de DNA é desconhecida, então se faz necessário a utilização de um processo de “conversão” do alfabeto do código genético para o alfabeto então utilizado nos CCEs, onde os alfabetos 4-ários dos CCEs estão relacionados ao conjunto $N = \{A, C, G, T/U\}$.

Primeiramente, associamos os elementos do conjunto $N = \{A, C, G, T/U\}$ com os elementos dos conjuntos $GF(4) = \{0, 1, \alpha, \alpha^2\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como mostrado na Figura 3.10 a) e b), respectivamente.

Como f e f' são funções sobrejetoras, suas funções inversas identificam estruturas de espaços vetoriais associadas ao código genético, onde $f^{-1} \rightarrow$ identifica a estrutura de espaço vetorial sobre corpo e, $f'^{-1} \rightarrow$ identifica o módulo (espaço vetorial)


 Figura 3.10: Associações entre $N \rightarrow GF(4)$ e $N \rightarrow \mathbb{Z}_4$.

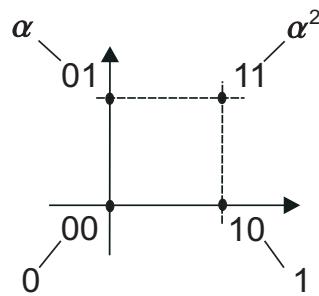
sobre anel. A associação de estruturas algébricas ao conjunto $N = \{A, C, G, T/U\}$, alfabeto do código genético, se faz necessária. Para isso, consideramos as mais simples das estruturas relacionadas ao alfabeto N . Logo, na estrutura de anel, consideramos o alfabeto do código $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, este obedecendo as operações de soma e produto módulo 4. Já na estrutura de corpo consideramos o alfabeto do código $GF(4) = \{0, 1, \alpha, \alpha^2\}$, obedecendo as operações de soma e produto módulo $(x^2 + x + 1)$.

b) A Estrutura de Espaço Vetorial sobre Corpo:

Considere $GF(4) = \{0, 1, \alpha, \alpha^2\}$ como a extensão de Galois de grau 2 de $GF(2)$, isto é,

$$GF(4)[x] \cong \frac{GF(2)[x]}{\langle p(x) \rangle} = \frac{GF(2)[x]}{\langle x^2 + x + 1 \rangle} = \{a_0 + a_1x; \quad a_0, a_1 \in GF(2)\}.$$

Nesse processo de extensão, note que o par a_0a_1 assume os valores 00, 10, 01 e 11. Esses elementos binários estão associados aos elementos de $GF(4)$, sendo $00 - 0$, $10 - 1$, $01 - \alpha$ e $11 - \alpha^2$. Portanto, o espaço vetorial associado ao $GF(4)$ é ilustrado na Figura 3.11.


 Figura 3.11: Representação da estrutura de espaço vetorial associado ao $GF(4)$.

c) **O Módulo (Espaço Vetorial) sobre Anel:**

A associação de espaço vetorial ao conjunto $N = \{A, C, G, T/U\}$ proveniente dos números complexos se verifica através da raiz quarta da unidade, isto é, $x^4 - 1 = 0$. As soluções dessa equação são dadas por

$$S = \{e^{j\frac{0\pi}{2}}, e^{j\frac{1\pi}{2}}, e^{j\frac{2\pi}{2}}, e^{j\frac{3\pi}{2}}\}.$$

Note que S é um grupo multiplicativo, cujo elemento gerador é $e^{j\frac{\pi}{2}}$. Como S é um grupo multiplicativo, e tem um gerador, o que se nota é que existe um isomorfismo entre o grupo multiplicativo S e o grupo aditivo (módulo 4) S' , denotado por

$$S = \{e^{j\frac{0\pi}{2}}, e^{j\frac{1\pi}{2}}, e^{j\pi}, e^{j\frac{3\pi}{2}}\} \quad \rightarrow \quad S' = \{0, 1, 2, 3\}.$$

Como S e S' são isomorfos segue que, o espaço vetorial associado ao \mathbb{Z}_4 é como ilustrado na Figura 3.12.

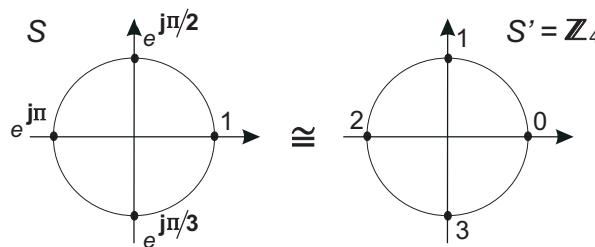


Figura 3.12: Representação de espaço vetorial associado ao \mathbb{Z}_4 .

Em termos de espaço vetorial, estamos introduzindo no bloco codificador estruturas algébricas (corpo e anel), não que elas sejam as únicas, mas, estamos apontando possibilidades de estruturas matemáticas associadas ao código genético que sejam capazes de identificar algumas sequências de DNA, que nada mais é do que inserir um alfabeto associado a um código.

Dessa maneira, o codificador genético é composto pelos blocos rotulamento (composição mapeamento (\mathbb{Z}_4 -linear, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear e Klein-linear)/mapeamento) e código BCH (estruturas de corpo e anel), como ilustrado na Figura 3.13.

Para a determinação do código CCE resultante no codificador genético, se faz necessário o desdobramento do rotulamento, bem como, a identificação dos mapeamentos $GF(4)$ e \mathbb{Z}_4 .

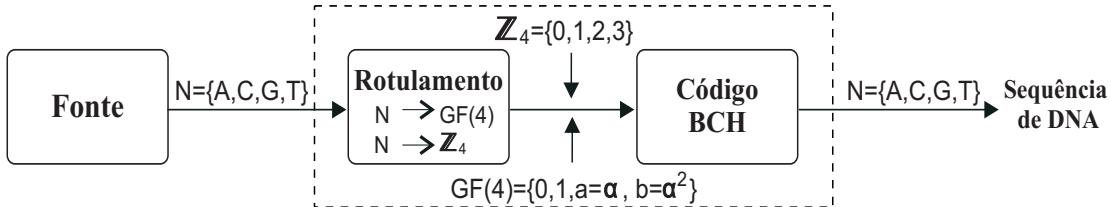


Figura 3.13: Codificador genético.

Segundo - Determinação do Rotulamento/Mapeamento no Codificador Genético

Todas as possibilidades de associações dos elementos do conjunto $N = \{A, C, G, T/U\}$ com os elementos dos conjuntos $GF(4) = \{0, 1, \alpha = a, \alpha^2 = b\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ denominamos **rotulamentos**, como mostrado na Figura 3.14. Estes rotulamentos têm como objetivo determinar qual a melhor associação de cada um dos símbolos no conjunto N com os correspondentes símbolos nos conjuntos \mathbb{Z}_4 e $GF(4)$ e vice-versa.

| $N! = 4! = 24$ possibilidades de rotulamento para $GF(4)=\{0,1,a,b\}$ | | | | | | | |
|---|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[0 \ 1 \ b \ a]$ | $[a \ 1 \ b \ 0]$ | $[0 \ 1 \ a \ b]$ | $[a \ 1 \ 0 \ b]$ | $[0 \ a \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[0 \ b \ 1 \ a]$ | $[a \ b \ 1 \ 0]$ | $[0 \ b \ a \ 1]$ | $[a \ b \ 0 \ 1]$ | $[0 \ a \ b \ 1]$ | $[a \ 0 \ b \ 1]$ | $[a \ 0 \ b \ 1]$ | $[a \ 0 \ b \ 1]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[1 \ 0 \ a \ b]$ | $[b \ 0 \ a \ 1]$ | $[1 \ 0 \ b \ a]$ | $[b \ 0 \ 1 \ a]$ | $[1 \ b \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[1 \ a \ 0 \ b]$ | $[b \ a \ 0 \ 1]$ | $[1 \ a \ b \ 0]$ | $[b \ a \ 1 \ 0]$ | $[1 \ b \ a \ 0]$ | $[b \ 1 \ a \ 0]$ | $[b \ 1 \ a \ 0]$ | $[b \ 1 \ a \ 0]$ |

| $N! = 4! = 24$ possibilidades de rotulamento para $\mathbb{Z}_4=\{0,1,2,3\}$ | | | | | | | |
|--|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|-------------------|
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[0 \ 1 \ 3 \ 2]$ | $[2 \ 1 \ 3 \ 0]$ | $[0 \ 1 \ 2 \ 3]$ | $[2 \ 1 \ 0 \ 3]$ | $[0 \ 2 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[0 \ 3 \ 1 \ 2]$ | $[2 \ 3 \ 1 \ 0]$ | $[0 \ 3 \ 2 \ 1]$ | $[2 \ 3 \ 0 \ 1]$ | $[0 \ 2 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[1 \ 0 \ 2 \ 3]$ | $[3 \ 0 \ 2 \ 1]$ | $[1 \ 0 \ 3 \ 2]$ | $[3 \ 0 \ 1 \ 2]$ | $[1 \ 3 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ |
| $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ | $[A \ C \ G \ T]$ |
| $[1 \ 2 \ 0 \ 3]$ | $[3 \ 2 \ 0 \ 1]$ | $[1 \ 2 \ 3 \ 0]$ | $[3 \ 2 \ 1 \ 0]$ | $[1 \ 3 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ |

 Figura 3.14: Possibilidades de rotulamento entre os elementos dos conjuntos N , $GF(4)$ e \mathbb{Z}_4 .

Lembramos que a proteção desigual do códon não será analizada neste modelo, pois cada nucleotídeo que compõe o códon terá associado a mesma proteção a erros pela própria construção do código.

A identificação da estrutura algébrica das sequências de DNA e do CCE está relacionada à estrutura do alfabeto do código genético, porém, a identificação da estrutura

geométrica das sequências de DNA está relacionada aos alfabetos binários dos CCEs e, portanto é necessário a realização do mapeamento entre os alfabetos 4-ário e binário. A decomposição binária dos elementos de $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ e $GF(4) = \{0, 1, \alpha, \alpha^2\}$ denominamos **mapeamento**, como ilustrado abaixo.

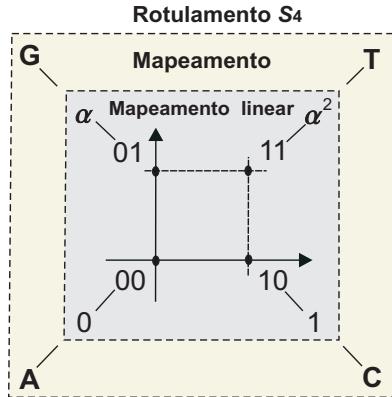
| Mapeamento $GF(4)$ | Mapeamento \mathbb{Z}_4 |
|-------------------------------|---------------------------|
| $00 \leftrightarrow 0$ | $00 \leftrightarrow 0$ |
| $10 \leftrightarrow 1$ | $10 \leftrightarrow 1$ |
| $01 \leftrightarrow \alpha$ | $11 \leftrightarrow 2$ |
| $11 \leftrightarrow \alpha^2$ | $01 \leftrightarrow 3$ |

No modelo do codificador genético (Figura 3.13), observamos que a informação na saída da fonte é formada pelos elementos do conjunto N . Assumindo que a probabilidade de ocorrência é a mesma para cada um desses elementos, ou seja $1/4$, a fonte terá a maior entropia e, portanto a maior incerteza. Como os alfabetos dos códigos BCH sobre $GF(4)$ e \mathbb{Z}_4 são distintos do alfabeto do código genético, é então necessário a realização do rotulamento entre $N \leftrightarrow GF(4)$ e $N \leftrightarrow \mathbb{Z}_4$. Com isso, relacionamos todas as possibilidade de rotulamentos dos elementos do conjunto N com os elementos dos conjuntos $GF(4)$ e \mathbb{Z}_4 . Portanto, as composições mapeamento/mapeamento linear no $GF(4)$, ou, mapeamento/mapeamento não-linear no \mathbb{Z}_4 caracteriza o bloco rotulamento no codificador genético.

Para analisarmos estes rotulamentos é necessário levarmos em consideração a complementaridade que o código genético exerce, de modo que, a adenina (A) se liga com a timina (T) ou a uracila (U) (ou vice-versa) e a guanina (G) se liga a citosina (C) (ou vice-versa), da seguinte maneira:

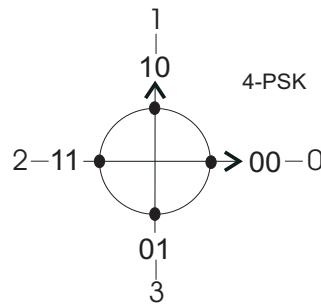
a) **Rotulamento/Mapeamento no $GF(4) = \{0, 1, \alpha, \alpha^2\}$:**

A estrutura matemática proveniente da extensão do $GF(2)$ para o $GF(4)$ é um mapeamento linear (Figura 3.11). Como o rotulamento entre $N \rightarrow GF(4)$ é desconhecido, toda sequência de DNA será rotulada através de cada uma das 24 permutações entre $N \rightarrow GF(4)$. Para cada sequência reproduzida pelo código notamos a existência de 24 palavras-código correspondentes a 24 permutações. Empregando o rotulamento recíproco, $GF(4) \rightarrow N$, em cada uma dessas 24 palavras-código teremos como resultado 24 palavras-código iguais em termos de nucleotídeos e aminoácidos. O rotulamento no $GF(4)$ é análogo ao grupo das permutações denominado S_4 , Figura 3.15.


 Figura 3.15: Rotulamento S_4 .

b) **Rotulamento/Mapeamento no $\mathbb{Z}_4 = \{0, 1, 2, 3\}$:**

Utilizando do código de Gray (rotulado em ponto ou sinal da constelação, o rótulo dos vizinhos correspondentes diferem em um bit ao sinal da constelação) para o rotulamento do 4-PSK têm-se a representação binária do \mathbb{Z}_4 , como ilustrado na Figura 3.16. Observe que a representação binária está associada a cada um dos elementos do conjunto \mathbb{Z}_4 , sendo $00 - 0$; $10 - 1$; $11 - 2$; $01 - 3$. Note que esse mapeamento é não linear.


 Figura 3.16: Elementos do alfabeto 4-ário do \mathbb{Z}_4 e sua representação binária.

A associação de complementaridade dos nucleotídeos A - T/U e C - G com os rotulamentos e o mapeamento não linear é o que diferenciam. Como o rotulamento entre $N \rightarrow \mathbb{Z}_4$ é desconhecido, toda sequência de DNA será rotulada utilizando-se cada uma das 24 permutações entre $N \rightarrow \mathbb{Z}_4$. Para cada sequência reproduzida pelo código notamos que somente um dos três conjuntos contendo oito palavras-código correspondentes a 8 permutações está relacionado. Empregando o rotulamento recíproco, $\mathbb{Z}_4 \rightarrow N$, em cada uma dessas 8 palavras-código teremos como resultado 8 palavras-código iguais em termos de nucleotídeos e aminoácidos. Este

fato resulta em três conjuntos contendo oito permutações cada um, rotulamento A, B e C, como ilustrados na Figura 3.17 que resultará em três mapeamentos distintos (Figura 3.19).

| Rotulamento A | Rotulamento B | Rotulamento C |
|---|---|---|
| $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 1 & 3 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 1 & 3 & 0 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 1 & 2 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 1 & 0 & 3 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 2 & 1 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 0 & 1 & 3 \end{bmatrix}$ |
| $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 3 & 1 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 3 & 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 3 & 2 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 3 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 0 & 2 & 3 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 2 & 0 & 3 & 1 \end{bmatrix}$ |
| $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 0 & 2 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 0 & 2 & 1 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 0 & 3 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 0 & 1 & 2 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 3 & 0 & 2 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 1 & 0 & 2 \end{bmatrix}$ |
| $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 2 & 0 & 3 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 2 & 0 & 1 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 2 & 3 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 2 & 1 & 0 \end{bmatrix}$ | $\begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 1 & 3 & 2 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{A} & \mathbf{C} & \mathbf{G} & \mathbf{T} \\ 3 & 1 & 2 & 0 \end{bmatrix}$ |

Figura 3.17: Rotulamentos A, B e C.

A determinação dos rotulamentos (composição mapeamento/mapeamento) é proveniente da complementaridade biológica (A-T) e (C-G), que por sua vez deve prevalecer sobre complementaridade matemática (00-11) e (01-10) e, consequentemente sobre os códigos resultantes capazes de identificar e reproduzir as sequências de DNA.

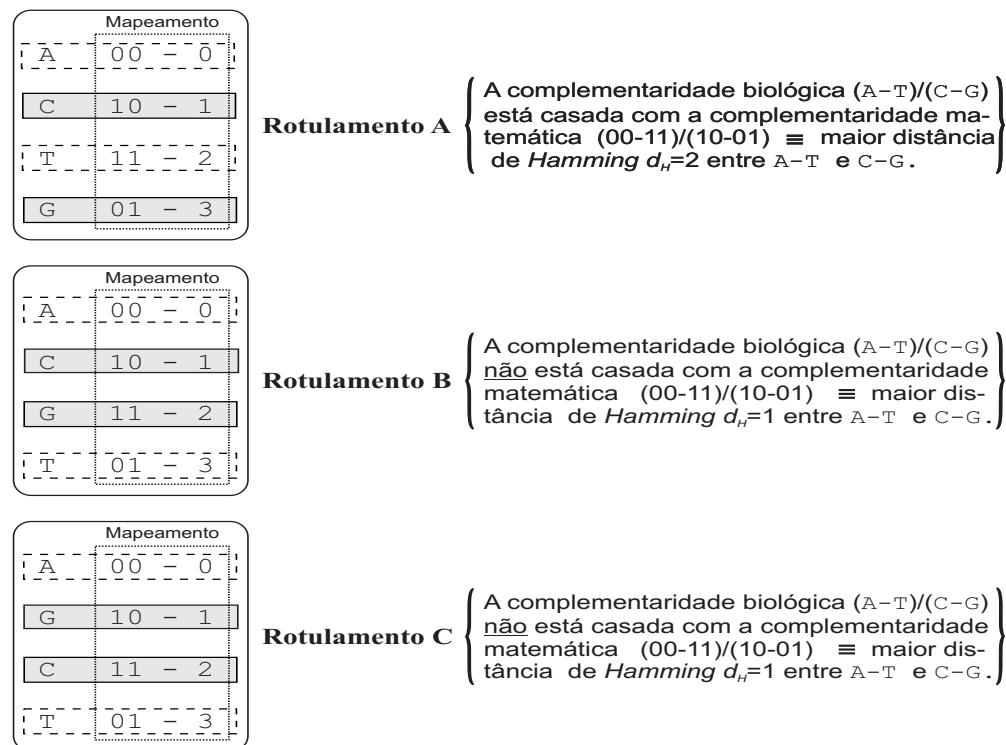


Figura 3.18: Casamento entre o contexto biológico e o contexto matemático.

No entanto, a complementaridade biológica pode ou não estar casada com a complementaridade matemática como mostrado na Figura 3.18. Note na figura que o mapeamento (relação entre os alfabetos binário/4-ário) é o mesmo nos três rotulamentos e é não-linear, porém não é conhecido se os rotulamentos A, B e C (permutações entre o alfabeto do código genético e o alfabeto 4-ário) são lineares ou não. No entanto, a associação entre os rotulamentos e o mapeamento não-linear resultará em mapeamentos não-lineares e lineares.

No caso do rotulamento A, como ocorre o casamento entre a complementaridade biológica e matemática, temos como resultado um **mapeamento não-linear**. Já nos casos dos rotulamentos B e C, não ocorre o casamento entre a complementaridade biológica e matemática e como a complementaridade biológica deve prevalecer, temos como resultado **mapeamentos lineares**.

Tendo em vista as componentes biológica, algébrica e geométrica associadas, como mostrado na Figura 3.19, denominamos o mapeamento não-linear como **\mathbb{Z}_4 -linear** e, os mapeamentos lineares como **$\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear** e **Klein-linear**, respectivamente.

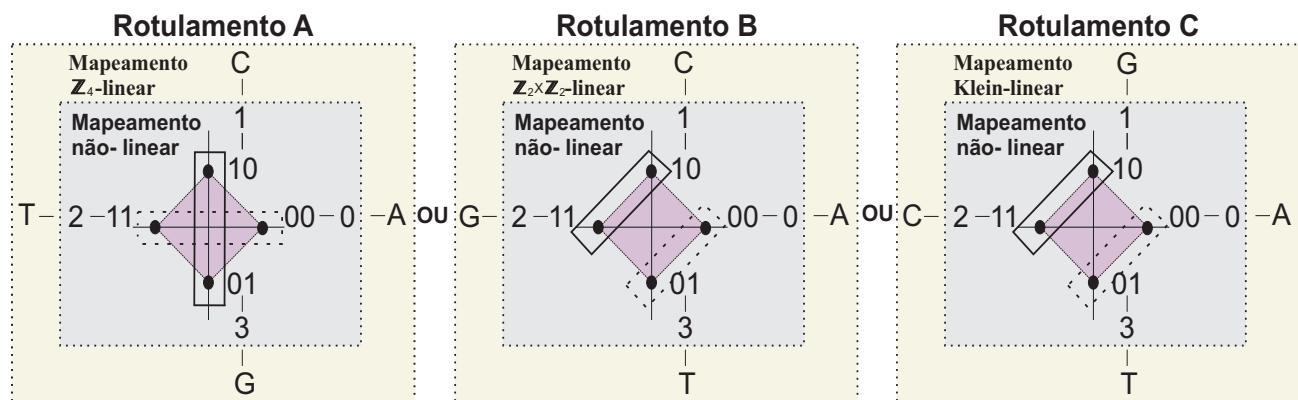


Figura 3.19: Mapeamentos do alfabeto \mathbb{Z}_4 .

Observe na Figura 3.19 que, no caso do rotulamento A qualquer um dos nucleotídeos para alcançar o seu complementar necessita caminhar duas arestas, enquanto que nos dois rotulamentos restantes basta caminhar uma aresta somente. Todas as permutações associadas ao rotulamento A caracterizam o mapeamento como \mathbb{Z}_4 -linear; as permutações associadas ao rotulamento B caracterizam o mapeamento como $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear; enquanto que as permutações associadas ao rotulamento C caracterizam o mapeamento como Klein-linear.

Terceiro - Determinação dos CCEs resultantes no Codificador Genético:

A denominação \mathbb{Z}_4 -linear foi dado por coincidir com a representação \mathbb{Z}_4 pelo código de Gray. Portanto, \mathbb{Z}_4 -linear e \mathbb{Z}_4 -linearidade são conceitos distintos. Na Figura 3.20, veja os códigos CCEs resultantes no codificador genético.

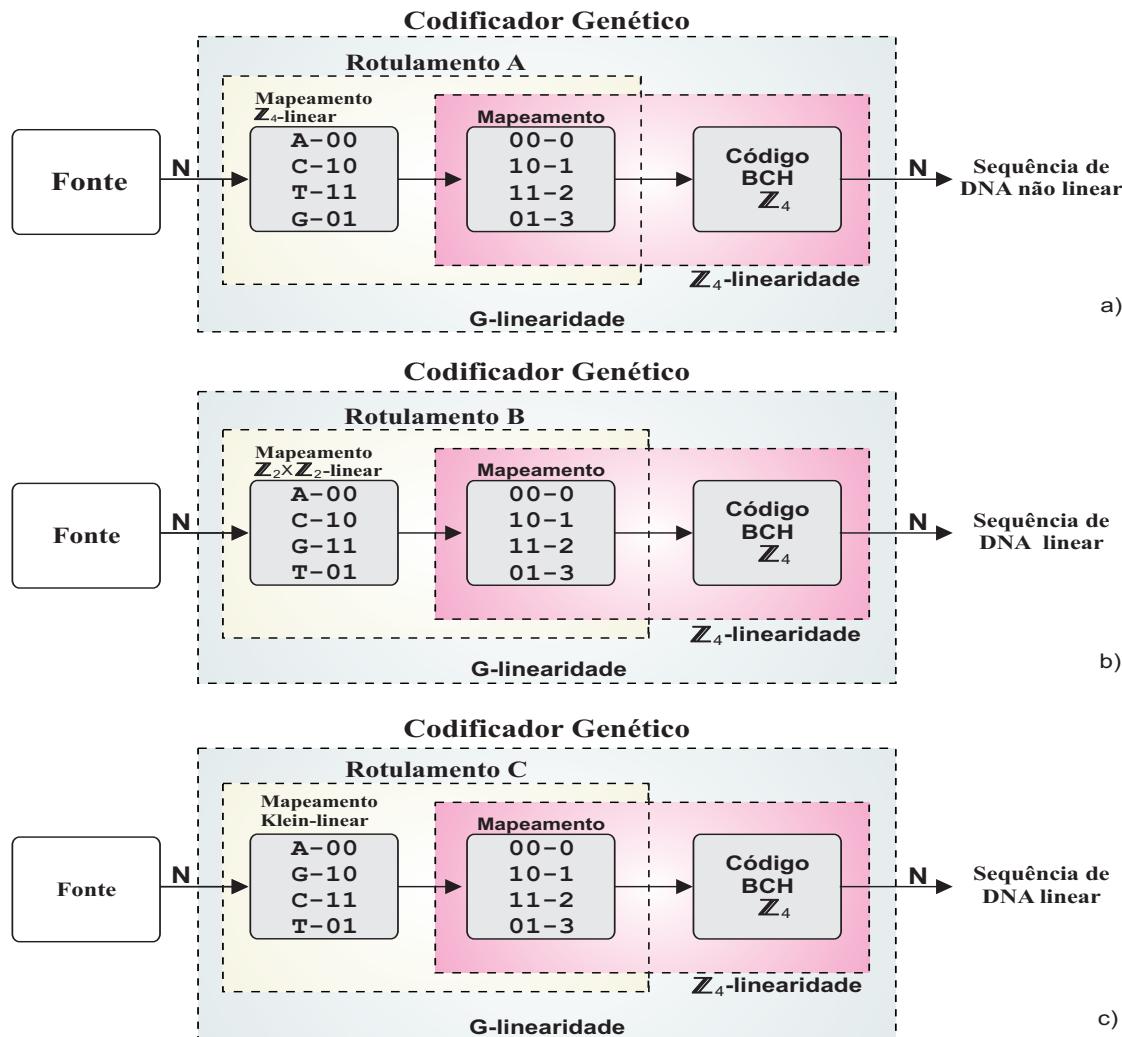


Figura 3.20: Códigos resultantes no codificador genético.

a) \mathbb{Z}_4 -linearidade:

O fato de o mapeamento-não linear ser o único mapeamento associado ao código BCH sobre a estrutura de anel (que é linear), temos como código resultante do tipo \mathbb{Z}_4 -linearidade presente nos três codificadores genéticos ilustrados na Figura 3.20. Ou seja, \mathbb{Z}_4 -linearidade significa que o código resultante (no bloco lilás do codificador genético) é não linear. Em outras palavras, a \mathbb{Z}_4 -linearidade transportou

de uma forma bastante sutil para o código binário a “linearidade” do código 4-ário, onde certas classes de códigos não-lineares podem ser vistos como códigos lineares sobre \mathbb{Z}_4 . Como consequência deste fato, os códigos não-lineares têm distância mínima maior do que os códigos lineares.

Observe na Figura 3.20a que o mapeamento \mathbb{Z}_4 -linear é não-linear e o código BCH sobre a estrutura de anel é linear, logo o código resultante será não-linear (por prevalecer o contexto biológico) e toda sequência de DNA reproduzida por este código será não-linear. Já nos casos dos mapeamentos $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear e Klein-linear (Figuras 3.20b e c), são lineares e o código BCH também é linear, logo os códigos resultantes serão lineares e toda sequência de DNA reproduzida por estes códigos serão lineares.

b) **G -linearidade:**

Na Subseção 2.3.6, vimos que os códigos pertencentes a subclasse G -linearidade são uma extensão da \mathbb{Z}_4 -linearidade. Como resultado do mapeamento entre os conjuntos N e \mathbb{Z}_4 , ilustrado na Figura 3.19, obtemos os conjuntos \mathbb{Z}_4 -linear, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear e Klein-linear. Como cada um desses conjuntos faz parte do grupo de simetrias do quadrado, temos que, a composição rotulamento/mapeamento/ \mathbb{Z}_4 -linearidade (Figura 3.20) leva a um código resultante com as características de um código pertencente a subclasse G -linearidade, agora denominado **código G -linearidade** como ilustrado no bloco do codificador genético da Figura 3.20.

Em [49] e [50], sabe-se que se um código é pertencente à subclasse da G -linearidade, então:

1. O alfabeto do código é efetivamente casado ao grupo de simetrias, e consequentemente, o código está casado ao grupo correspondente obtido pelo mapeamento estendido;
2. O código é geometricamente uniforme, isto é

seja S uma figura consistindo de elementos s_i , $1 \leq i \leq k$. Uma isometria u_{s_1, s_2} é tal que quando aplicada em s_1 conduz ao elemento s_2 , isto é,

$$u_{s_1, s_2}(s_1) = s_2,$$

$$u_{s_1, s_2}(\mathcal{S}) = \mathcal{S}.$$

\mathcal{S} é dito geometricamente uniforme se a ação do grupo de simetrias $\Gamma(\mathcal{S})$ de \mathcal{S} é transitiva. Se \mathcal{S} for finito, dizemos que \mathcal{S} é uma constelação uniforme e, se

\mathcal{S} for infinito dizemos que \mathcal{S} é um arranjo regular. Uma constelação uniforme no espaço Euclidiano é um código de grupo.

Temos ainda que, o grupo gerador mínimo $U(\mathcal{S})$ de \mathcal{S} , é um subgrupo do grupo de simetrias de \mathcal{S} , e, que o produto cartesiano de conjuntos de sinais geometricamente uniformes é um conjunto de sinais geometricamente uniforme.

Em outras palavras, a figura S proveniente do grupo multiplicativo (isomorfo ao grupo aditivo S' representado na Figura 3.12 e, equivalente a Figura 3.16), fica invariante quando aplicadas as isometrias (rotações e reflexões), observadas na Figura 3.19, caracterizando que o código seja **geometricamente uniforme**.

No processo de determinação do modelo de codificação genética, ilustrado na Figura 3.8, é que existe um sistema de codificação combinado com a modulação, isto é, um **mapeamento casado** (MC) implicando que a estrutura algébrica a ser utilizada no codificador é a mesma, a menos de um isomorfismo, que a da constelação de sinais, garantindo assim, a menor complexidade possível do sistema. A classe de códigos satisfazendo essa propriedade é a classe dos códigos geometricamente uniformes e, a subclasse é a da G -linearidade.

Portanto, os códigos que serão usados na identificação e reprodução das sequências de DNA, no presente trabalho, consistem de códigos **G -linearidade = $\{\mathbb{Z}_4$ -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade, Klein-linearidade**.

Quarto - Parâmetros dos Códigos BCH e os Polinômios Primitivos e Geradores

Os parâmetros do código BCH são denotados da seguinte maneira: n = comprimento das palavras-código (comprimento das sequências de DNA em nucleotídeos); k = dimensão do código (comprimento da sequência de informação responsável pela geração da sequência de DNA) e d_H = distância mínima do código (o menor número de posições em que quaisquer duas palavras-código diferem). O código BCH com parâmetros (n, k, d_H) possui uma capacidade de correção de erros estabelecida através da relação $d_H = 2t + 1$, onde t denota a quantidade de erros.

No caso em que o código é sobre anel, para que a fatoração de $x^n - 1$ na extensão $GR(p^k, r) = GR(4, r)$ seja única, é necessário que o comprimento da palavra-código seja ímpar ($n = p^r - 1$). Analisaremos sequências de DNA que possuem comprimentos n iguais a 21, 51, 63, 93, 255, 511, 1023 e 2047 nucleotídeos, respectivamente. Para isso, serão construídos códigos BCH primitivos e não primitivos sobre anéis nas extensões de Galois de grau $r = 6, r = 8, r = 9, r = 10$ e $r = 11$ ($GR(4, r)$) e, códigos BCH

primitivos sobre corpos nas extensões de Galois de grau $r = 3$ e $r = 4$ ($GF(4^r)$).

Os códigos BCH, tanto sobre anel quanto sobre corpo, serão construídos em todas as distâncias mínimas (d_H) e, em todos os polinômios (primitivos/geradores) de cada extensão de Galois de grau r , com o objetivo de encontrar um código capaz de gerar as sequências de DNA sem nenhuma diferença de nucleotídeo ou no máximo diferindo em até dois nucleotídeos. Chamamos a atenção para o seguinte fato, para cada d_H , teremos um polinômio gerador $g(x)$ diferente e, consequentemente, um novo código. Sendo assim, devemos considerar cada um destes códigos como um novo código a ser analisado.

Devido ao isomorfismo existente entre corpos estendidos utilizando polinômio primitivo de mesmo grau, algebraicamente, não existe a preferência por um determinado polinômio primitivo. Todavia, como às sequências de DNA (proteínas) existe uma associação geométrica é razoável considerar todos os polinômios primitivos. Como não conhecemos a estrutura algébrica e geométrica das sequências de DNA, decidimos realizar a construção do código BCH sobre anel e sobre corpo para cada um dos polinômios primitivos de cada extensão de Galois, com o objetivo de verificar quais polinômios primitivos e os correspondentes polinômios geradores resultarão em códigos capazes de reproduzirem as correspondentes sequências de DNA.

Ressaltamos que, no caso dos códigos BCH não primitivos, as classes dos polinômios geradores decorrentes dos polinômios primitivos deverão ser determinadas para cada código analisado.

Considerando ainda o fato de que as sequências de DNA são distintas biologicamente e de que o comprimento das palavras-código deve ser igual ao comprimento dessas sequências, teremos para cada um dos comprimentos uma correspondente extensão de Galois sobre corpo e sobre anel. Cada extensão de Galois possui uma quantidade de polinômios primitivos, de forma que, quanto maior o grau da extensão maior a quantidade de polinômios primitivos. A dificuldade que se apresenta para a solução desse problema está em, quanto maior o grau da extensão de Galois maior será a quantidade de polinômios primitivos que devem ser utilizados na construção dos códigos, aumentando significativamente a complexidade computacional.

Dessa maneira, concluímos o desdobramento do codificador genético, apresentando os elementos (estrutura algébrica, alfabeto, rotulamento, mapeamento, polinômio primitivo e polinômio gerador) fundamentais e determinantes dos códigos G -linearidade, que serão responsáveis pela identificação e reprodução das sequências de DNA. Ainda no contexto de

codificação genética, a seguir, apresentamos o modelo de identificação e o procedimento de classificação matemática das sequências de DNA.

Modelo de identificação das sequências de DNA

Aqui mostramos uma interpretação do processo de identificação de sequências de DNA relacionado ao codificador genético. Neste processo as sequências de DNA são identificadas através dos códigos G -linearidade: BCH primitivo e não primitivo sobre corpo e sobre anel e suas extensões de Galois. Sabemos que a matriz geradora de código linear com parâmetros (n, k, d_H) é dada por

$$G = \begin{pmatrix} g_{11} & g_{12} & \cdots & g_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ g_{k1} & g_{k2} & \cdots & g_{kn} \end{pmatrix}.$$

As linhas da matriz geradora formam uma base do espaço vetorial identificado como o código linear C . Assim, as combinações lineares das linhas de G são palavras-código de C . Dessa forma, o processo de codificação pode ser escrito como $\mathbf{v} = \mathbf{u} \cdot G$, onde \mathbf{u} é a informação e \mathbf{v} é a palavra-código correspondente, no nosso caso as sequências de DNA a serem analisadas. Para toda palavra-código \mathbf{v} vale a relação $\mathbf{v} \cdot H^T = 0$, onde H^T é a transposta da matriz verificação de paridade. O processo de identificação de uma sequência de DNA qualquer é análogo ao processo de construção do codificador de um código linear, cíclico e abeliano descrito anteriormente, como mostrado na Figura 3.21.

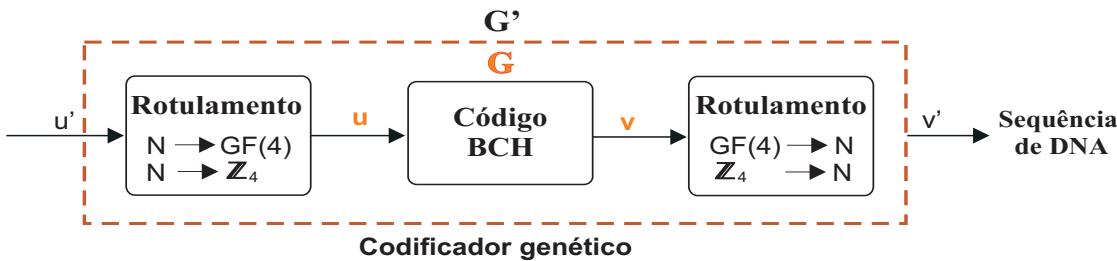


Figura 3.21: Codificador do código G -linearidade.

Observe na figura como ocorre o processo de codificação genética. Na entrada do codificador, observe a sequência de informação \mathbf{u}' associada ao alfabeto do código genético N . É realizado o mapeamento entre $N \rightarrow \mathbb{Z}_4$, e, portanto, a sequência de informação \mathbf{u}' em termos do código genético N agora está relacionada aos alfabetos dos códigos BCH sobre corpo ou sobre anel, sendo denotada por \mathbf{u} . No processo de codificação, $\mathbf{u} \cdot G$, temos a palavra-código \mathbf{v} , em termos dos alfabetos dos códigos $GF(4)$ ou \mathbb{Z}_4 . Um novo mapeamento é realizado,

agora entre $\mathbb{Z}_4 \rightarrow N$ ou $GF(4) \rightarrow N$, para a determinação da sequência em termos do alfabeto do código genético N e portanto, a correspondente sequência de DNA denotada por v' .

No processo de identificação das sequências de DNA, o procedimento natural seria determinarmos todas as sequências de informação u' . Porém, este processo é inviável em virtude do comprimento das sequências de DNA e, consequentemente, do número muito grande de comparações a ser realizadas: no caso em questão, temos 4^k comparações. Observe que, quanto maior for o valor de k , maior será o número de comparações. Para contornarmos este problema, que é classificado como um problema NP-completo, ao invés de gerarmos todas as palavras-código para compararmos com a sequência de DNA, consideramos que a sequência, sob a aplicação de cada uma das 24 permutações, é uma palavra-código, sendo que a diferença entre a palavra-código e a sequência de DNA do NCBI é nula. Assim, para determinarmos se cada uma dessas 24 possibilidades é de fato uma palavra-código usamos a relação $\mathbf{v} \cdot H^T = 0$, onde \mathbf{v} é a possível palavra-código e H^T é a transposta da matriz verificação de paridade. Na Figura 3.22 fazemos uma interpretação do processo de codificação genética no contexto biológico.

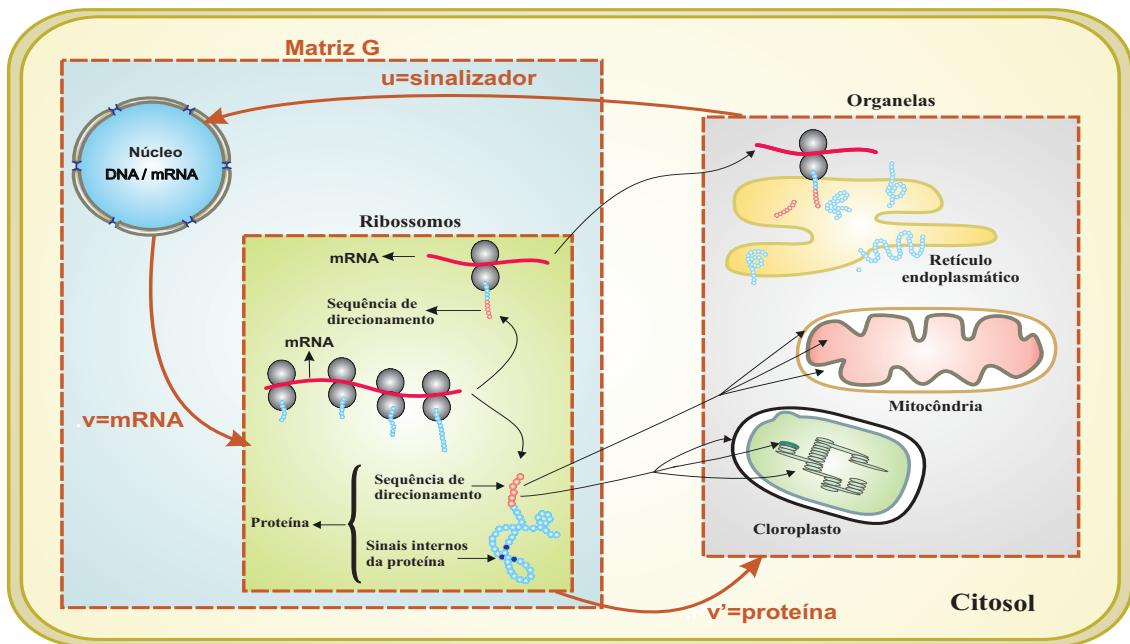


Figura 3.22: Interpretação do processo de identificação de sequências de DNA.

Neste sistema de comunicação biológico mensagens são enviadas e recebidas, de maneira que a maquinaria celular realiza o processamento dessas mensagens de acordo com a necessidade e local apropriado. Por exemplo, quando uma organela (mitocôndria, cloroplasto,

retículo endoplasmático) em um dos seus subcompartimentos necessita de proteína nuclear, essa organela envia uma mensagem ao núcleo da célula através da informação **u**. O núcleo celular, por sua vez, recebe essa informação **u**, localiza no genoma a parte solicitada, extraí a informação do genoma e a codifica em uma palavra-código **v**. No contexto biológico a palavra-código **v** é vista como o RNA mensageiro (mRNA). O mRNA é então processado pelo ribossomo e traduzido em proteína denotada pela palavra-código **v'** que é encaminhada à organela destino. No contexto de codificação genética, os casos em que ocorrem diferenças de um ou dois nucleotídeos entre as sequências reproduzidas e as sequências originais também são de interesse.

Para analisarmos as sequências de DNA que apresentam **1 nucleotídeo de diferença** relativamente à sequência do *NCBI*, consideramos as três outras possibilidades de nucleotídeos em cada posição na sequência de DNA para cada permutação. Para analisarmos as sequências de DNA que apresentam **2 nucleotídeos de diferença** relativamente à sequência do *NCBI*, nos vinte e quatro casos de permutações, consideramos todas as combinações simples tomadas dois a dois dos n nucleotídeos de comprimento da sequência.

A relevância em se considerar até dois nucleotídeos de diferença é retratada, por exemplo, em [22], onde técnicas laboratoriais mostraram que, quando há uma troca de um aminoácido em um ponto da cadeia protéica, para manter a sua estabilidade, ocorre outra troca em outro ponto da cadeia. Na tese de doutorado [23], foram investigados dois mecanismos de resistência a piretróides, um conhecido como *kdr* (*knockdown resistance*), associado a mutações no canal de sódio, e a mutação W251S no gene da carboxilesterase E3, que têm sido associada à hidrólise de piretróides (inseticidas organofosforados e carbamatos). Os resultados indicaram diferentes espécies com mutações que ocorrem simultaneamente em duas posições na sequência de DNA, e estão associadas à resistência a piretróides como ilustra a Figura 3.23.

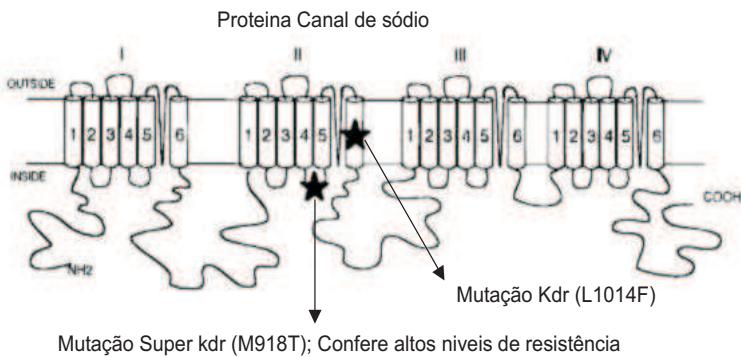


Figura 3.23: Mutações relacionadas à resistência a inseticidas, [23].

Fatos como estes podem ser investigados através da geração e da reprodução de sequências de DNA com 2 nucleotídeos diferindo das sequências originais, possibilitando uma análise mutacional nessas sequências de DNA. Do ponto de vista matemático, existem padrões de erros que podem ser detectados e corrigidos no processo da decodificação, como mostramos na Subseção 2.3.2. Assim, é possível que as sequências de DNA reproduzidas com 2 nucleotídeos de diferença da sequência original apresentem padrões de erros que poderão ser detectados e corrigidos no processo da decodificação.

Classificação matemática das sequências de DNA

Sob o ponto de vista biológico, nos preocupamos em selecionar para a análise sequências de DNA com características e funções celulares específicas e já conhecidas cientificamente. Estas sequências foram mencionadas na Tabela 1.1 do Capítulo 1 e estão relacionadas na Tabela 5.1, do Capítulo 5.

Outro dado importante é que, dentre as sequências de DNA, temos sequências de células eucariontes e células procariontes, as quais diferem, entre outras coisas, quanto à complexidade.

A abordagem do rotulamento, mapeamento, estrutura algébrica e CCE associados ao modelo de codificação genética permitirá uma classificação algébrica e geométrica dessas sequências. Como visto na Figura 3.20, o mapeamento \mathbb{Z}_4 classifica as sequências como não-lineares, enquanto os mapeamentos $\mathbb{Z}_2 \times \mathbb{Z}_2$ e Klein as classificam como lineares. Esses mapeamentos identificam as melhores associações entre cada símbolo no conjunto N e os seus correspondentes símbolos no conjunto \mathbb{Z}_4 e vice-versa. Sob o ponto de vista topológico, as estruturas secundárias das sequências de DNA (alpha-hélice, folha beta paralela e folha beta anti-paralela) guardam semelhanças com a representação geométrica dos códigos \mathbb{Z}_4 -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade. Essas semelhanças devem ser levadas em consideração e estudadas futuramente com o objetivo de avançar a caracterização matemática da estrutura secundária e terciária de certas sequências de DNA.

Uma outra possibilidade de classificação geométrica das sequências de DNA está relacionada aos polinômios geradores dos CCEs. Os coeficientes dos polinômios geradores reduzidos módulo-2 reproduzem os correspondentes polinômios primitivos. Portanto, geometricamente, cada polinômio gerador estará associado a uma superfície de Riemann [53]. Este fato evidencia uma estrutura geométrica associada às sequências de DNA que deverá ser estudada futuramente.

3.2.2 Modelo de codificação genômica

Durante todo o período de concepção e formatação do presente trabalho, estivemos focados no procedimento de identificação e reprodução de sequências de DNA com característica biológicas distintas e comprimentos variados através dos códigos BCH sobre corpo e anel e suas extensões de Galois, respectivamente. No contexto da dupla hélice do DNA, inserido em codificação genômica os códigos BCH sobre corpo e anel e suas extensões de Galois, respectivamente, foram capazes de identificar e reproduzir a dupla hélice de todas as sequências de DNA identificadas no modelo de codificação genética, incluindo o gene e o genoma circular do plasmídeo.

As sequências em termos da fita simples do DNA que são transcritas (RNA mensageiro) e traduzidas em proteínas (nucleotídeos e a correspondência em aminoácidos = código genético), tais como: sequências de direcionamento, sinais internos de uma proteína, hormônio, enzima e proteína são analisadas através do modelo de um sistema de comunicação de informação genética. Como essas sequências são escritas na linguagem dos nucleotídeos e dos correspondentes aminoácidos, o sistema de codificação combinado com modulação é fundamental para a identificação e reprodução dessas sequências de DNA.

Tanto as sequências em termos da fita simples do DNA - tais como íntron, DNA repetitivo, gene e genoma - quanto as sequências em termos da dupla hélice do DNA estão escritas somente na linguagem dos nucleotídeos, e, portanto, é mais apropriado usarmos o modelo de um sistema de comunicação de informação genômica para a análise dessas sequências. Uma vez que essas sequências não têm a correspondência com os aminoácidos, o transmissor no modelo de um sistema de codificação genômica tem a ver somente com a codificação de canal (Figura 3.7). O modulador, canal e demodulador passam a ser uma única entidade chamada “canal discreto podendo ter ou não memória”. O bloco transmissor do modelo de um sistema de comunicação de informação genômica é então caracterizado apenas pelo bloco do **codificador genômico**, Figura 3.24.

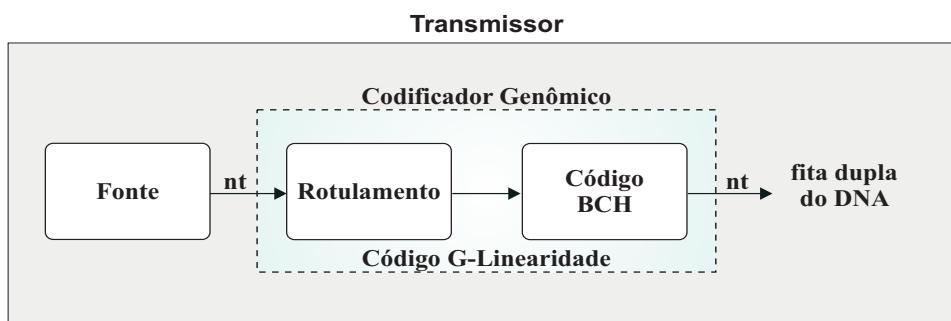


Figura 3.24: Modelo de um sistema de comunicação de informação genômica.

Todo o procedimento de identificação e reprodução das sequências de ítron, DNA repetitivo, gene e genoma, em termos da fita simples do DNA, é realizado igualmente ao procedimento apresentado no modelo de codificação genética, apenas essas sequências não são convertidas em aminoácidos. Sendo assim, fica desnecessária a caracterização do modelo de codificação genômica para essas sequências. No caso do procedimento de identificação e reprodução das sequências da dupla hélice do DNA, também são determinados de maneira análoga aos procedimentos realizados no modelo de codificação genética, porém, apresentando algumas diferenças relacionadas ao alfabeto e ao rotulamento da dupla hélice.

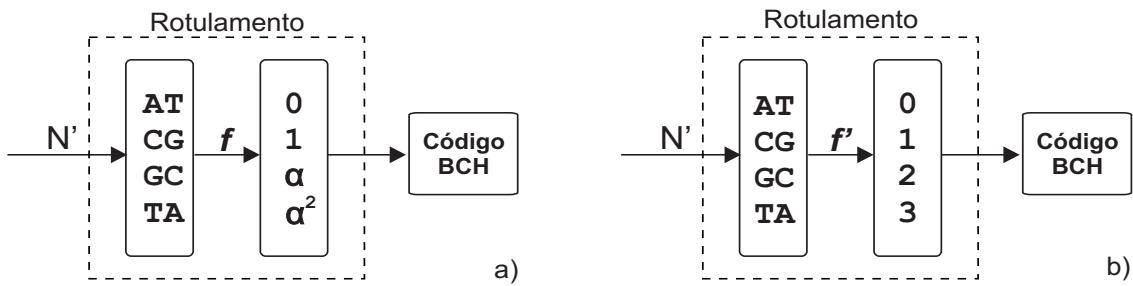
A palavra-código na saída do codificador genômico (Figura 3.24) está relacionada à sequência da dupla hélice do DNA (bases complementares: adenina com timina (AT), cito-sina com guanina (CG), guanina com citosina (GC) e timina com adenina (TA)) e, portanto o conjunto do alfabeto do código genômico é denotado por $N' = \{AT, CG, GC, TA\}$. Da mesma maneira que a estrutura algébrica do alfabeto do código genético é desconhecida, a estrutura algébrica do alfabeto do código genômico também o é. Logo, se faz necessária a identificação de tal estrutura, pois a mesma, como usual em teoria da codificação, será utilizada na construção do CCE. No processo de codificação genômica, utilizamos o alfabeto do código genômico em termos das bases complementares, denotado pelo conjunto $N' = \{AT, CG, GC, TA\}$, relacionado aos alfabetos dos códigos BCH sobre as estruturas algébricas de corpo e anel, como mostrado na Figura 3.25.

| “Conversão” | | | |
|-----------------------------|---|----------|------------|
| Alfabeto do código genético | | “4-ário” | |
| A-T | | 0 | 0 |
| T-A | → | 1 | 1 |
| C-G | | 2 | α |
| G-C | | 3 | α^2 |

Figura 3.25: O alfabeto do código genômico e os alfabetos 4-ários dos CCEs.

Como a estrutura algébrica do alfabeto do código genômico da dupla hélice é desconhecida, então se faz necessário a utilização de um processo de “conversão” do alfabeto do código genético para o alfabeto então utilizado nos CCEs, onde os alfabetos 4-ários dos CCEs estão relacionados ao conjunto $N' = \{AT, CG, GC, TA\}$. Logo, associamos os elementos do conjunto $N' = \{AT, CG, GC, TA\}$ aos elementos dos conjuntos $GF(4) = \{0, 1, \alpha, \alpha^2\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, como mostrado na Figura 3.26 a) e b), respectivamente.

As funções inversas identificam estruturas de espaços vetoriais associadas ao código genômico, onde $f^{-1} \rightarrow$ identifica a estrutura de espaço vetorial sobre corpo e, $f'^{-1} \rightarrow$ identifica o módulo


 Figura 3.26: Associações entre $N' \rightarrow GF(4)$ e $N' \rightarrow \mathbb{Z}_4$.

(espaço vetorial) sobre anel.

Com essas possibilidades de identificação de estruturas algébricas, temos que o codificador genômico é composto pelos blocos rotulamento (composição mapeamento/mapeamento) e código BCH (estruturas de corpo e anel), como ilustrado na Figura 3.27.

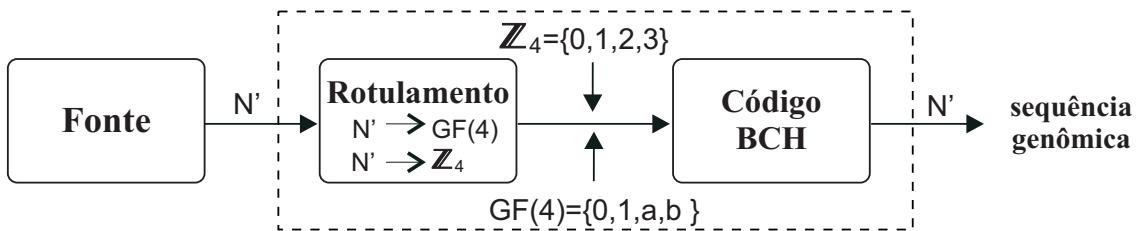


Figura 3.27: Codificador genômico.

Na Figura 3.28, mostramos todas as possibilidades de associações dos elementos do conjunto $N' = \{AT, CG, GC, TA\}$ com os elementos dos conjuntos $GF(4) = \{0, 1, \alpha = a, \alpha^2 = b\}$ e $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, denominados rotulamentos da dupla hélice.

Para determinação do código CCE resultante no codificador genômico, também se faz necessário o desdobramento do rotulamento, bem como a identificação dos mapeamentos $GF(4)$ e \mathbb{Z}_4 , que é a associação entre rotulamento e mapeamento. No entanto, das análises feitas com as sequências genômicas da dupla hélice do DNA, temos que a composição mapeamento/mapeamento no codificador genômico (bloco rotulamento) é igual à composição mapeamento/mapeamento no codificador genético, alterando-se apenas o alfabeto do código genético $N = \{A, C, G, T/U\}$ para o alfabeto do código genômico $N' = \{AT, CG, GC, TA\}$. Assim como no modelo de codificação genética, no modelo de codificação genômica temos: para cada sequência da dupla hélice reproduzida no rotulamento $GF(4)$, vinte e quatro palavras-código iguais em termos das bases complementares do DNA e, para cada sequência reproduzida no rotulamento \mathbb{Z}_4 , temos, como resultado, três conjuntos contendo oito palavras-código iguais em termos das bases complementares do DNA (rotulamentos A', B' e C') re-

| $N'! = 4! = 24$ possibilidades de rotulamento para $GF(4) = \{0, 1, a, b\}$ | | | | | | | |
|---|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[0 \ 1 \ b \ a]$ | $[a \ 1 \ b \ 0]$ | $[0 \ 1 \ a \ b]$ | $[a \ 1 \ 0 \ b]$ | $[0 \ a \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ | $[a \ 0 \ 1 \ b]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[0 \ b \ 1 \ a]$ | $[a \ b \ 1 \ 0]$ | $[0 \ b \ a \ 1]$ | $[a \ b \ 0 \ 1]$ | $[0 \ a \ b \ 1]$ | $[a \ 0 \ b \ 1]$ | $[a \ 0 \ b \ 1]$ | $[a \ 0 \ b \ 1]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[1 \ 0 \ a \ b]$ | $[b \ 0 \ a \ 1]$ | $[1 \ 0 \ b \ a]$ | $[b \ 0 \ 1 \ a]$ | $[1 \ b \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ | $[b \ 1 \ 0 \ a]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[1 \ a \ 0 \ b]$ | $[b \ a \ 0 \ 1]$ | $[1 \ a \ b \ 0]$ | $[b \ a \ 1 \ 0]$ | $[1 \ b \ a \ 0]$ | $[b \ 1 \ a \ 0]$ | $[b \ 1 \ a \ 0]$ | $[b \ 1 \ a \ 0]$ |

| $N'! = 4! = 24$ possibilidades de rotulamento para $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ | | | | | | | |
|--|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[0 \ 1 \ 3 \ 2]$ | $[2 \ 1 \ 3 \ 0]$ | $[0 \ 1 \ 2 \ 3]$ | $[2 \ 1 \ 0 \ 3]$ | $[0 \ 2 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ | $[2 \ 0 \ 1 \ 3]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[0 \ 3 \ 1 \ 2]$ | $[2 \ 3 \ 1 \ 0]$ | $[0 \ 3 \ 2 \ 1]$ | $[2 \ 3 \ 0 \ 1]$ | $[0 \ 2 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ | $[2 \ 0 \ 3 \ 1]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[1 \ 0 \ 2 \ 3]$ | $[3 \ 0 \ 2 \ 1]$ | $[1 \ 0 \ 3 \ 2]$ | $[3 \ 0 \ 1 \ 2]$ | $[1 \ 3 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ | $[3 \ 1 \ 0 \ 2]$ |
| $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ | $[AT \ CG \ GC \ TA]$ |
| $[1 \ 2 \ 0 \ 3]$ | $[3 \ 2 \ 0 \ 1]$ | $[1 \ 2 \ 3 \ 0]$ | $[3 \ 2 \ 1 \ 0]$ | $[1 \ 3 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ | $[3 \ 1 \ 2 \ 0]$ |

Figura 3.28: Possibilidades de rotulamento entre os elementos dos conjuntos N' e $GF(4)$ e N' e \mathbb{Z}_4 .

sultando em nos mapeamentos: \mathbb{Z}_4 -linear, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear e Klein-linear. Assim como no codificador genético, o código resultante no codificador genômico também pertence à subclasse *G*-linearidade e, portanto, a dupla hélice do DNA pode ser identificada e reproduzida através dos códigos *G*-linearidade.

Devido ao fato de o procedimento de identificação e reprodução da dupla hélice do DNA ser análogo ao procedimento de identificação e reprodução da fita simples do DNA, no próximo capítulo apresentaremos um algoritmo para a identificação e classificação somente de sequências de DNA em termos da fita simples do DNA, que realizará exaustivamente a construção dos códigos *G*-linearidade, tais como: **códigos BCH primitivo sobre o corpo $GF(4^r)$** e, dos **códigos BCH primitivos e não primitivos sobre o anel \mathbb{Z}_4** para todos os polinômios primitivos de cada extensão de Galois de grau r , respectivamente, bem como para as distâncias de Hamming variando entre $3 \leq d_H \leq n$.

3.3 Similaridades entre o Sistema Biológico e Redes Locais de Computadores

Os sistemas celulares são tão complexos que, mesmo utilizando de sofisticados recursos, ainda não foi possível à comunidade científica apresentar as devidas respostas ao mais variados questionamentos que estes sistemas impõem. No presente trabalho, conjecturamos que o

armazenamento da informação genômica em forma de quadro (*frame*) segue um conjunto de procedimentos em sua estruturação que lembra os procedimentos utilizados em redes locais de computadores. Por razão de completitude e diversidade dos assuntos abordados, inicialmente, incluímos na presente seção alguns temas de maneira resumida. Entretanto, os conceitos apresentados podem ser encontrados em [26], [27] e [54]-[59].

Armazenamento da informação genômica

A Figura 3.29 ilustra a mais nítida representação já obtida para a distribuição de cromossomos dentro de um núcleo celular. Seus autores a compararam com um nenúfar, a flor aquática imortalizada nos quadros do impressionista Claude Monet. “*Pode não ser tão bonita quanto as dele, mas tem lá sua beleza. Para admirar plenamente a flor do genoma, é preciso contemplá-la com os olhos do intelecto - armados com os binóculos da história da ciência*” [54]. Essa estrutura doméstica e misteriosa do núcleo é o que se pode vislumbrar na “flor” da imagem. Nesse caso específico, cada filamento colorido representa um dos 16 cromossomos de células de levedura *Saccharomyces cerevisiae* (micro-organismo do fermento). Todos estão ancorados pelo centrômero (a “cintura” do X ou Y). Para ele convergem todas as cores, como as pétalas de uma flor se reúnem no pedúnculo. A exceção é o cromossomo de número 12, que se projeta para fora da flor em direção ao nucléolo, uma estrutura específica do núcleo.

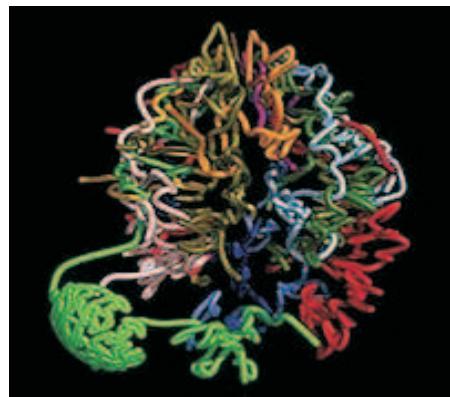


Figura 3.29: Representação da distribuição cromossônica dentro do núcleo celular. Mirela Andronesco/*Divulgação*, [54].

Não é o caso de tentar descrever o método usado pela equipe de William S. Noble, da Universidade de Washington, para reconstituir essa organização floral, publicado em maio de 2010 na revista *Nature*. Cabe assinalar que essa é a flor característica de uma espécie particular de levedura. Outros organismos, com genomas de sequências e arranjos cromossômicos diversos, terão estruturas específicas. “*Quando essas estruturas estiverem todas decifradas,*

poderão ser comparadas em busca de princípios gerais de organização e relações com processos fisiológicos importantes”, comenta William S. Noble, [54] e [55].

O DNA celular é extremamente compacto, sugerindo um alto grau de organização estrutural. O mecanismo de enovelamento não consiste somente em empacotar o DNA, mas também permite o acesso a informação do DNA. Lembrando que, juntamente com os processos de replicação e transcrição uma propriedade da estrutura do DNA, muito importante, é conhecida como *supercoiling* (super-helicoidização). *Supercoiling* significa dar voltas em um fio enrolado. Por exemplo, o fio que liga o fone de um aparelho de telefone a base é enrolado, as voltas que ocorrem frequentemente nesse fio já enrolado são *supercoils*. Nesse alto grau de organização estrutural celular, descrevemos de forma sucinta as noções de genoma, cromossomo e gene da seguinte maneira:

Genoma - Em biologia, o genoma é toda a informação hereditária (passada para seus descendentes) de um organismo que está codificada em seu DNA (ou, em alguns vírus, no RNA), ou seja, é seu manual de instruções. Isto inclui tanto os genes como as sequências não-codificadoras, que são muito importantes para a regulação gênica, dentre outras funções. O genoma é transmitido, com variações individuais, de geração em geração, e determina a espécie do ser vivo. Neste programa genético se encontram gravadas nossas características hereditárias encarregadas de dirigir o desenvolvimento biológico de cada indivíduo. As doenças hereditárias também estão escritas no genoma. Todos os seres vivos, desde os maiores, como o elefante e a baleia, até os minúsculos, como as bactérias, além de plantas, árvores e, claro, o ser humano, têm genoma. A metade do genoma que se herda provém do macho e a outra metade, da fêmea. Assim se reconstrói a árvore genealógica de todo ser vivo e a herança recebida de seus antepassados aparece no genoma de cada ser. O genoma está escrito na linguagem química do ácido desoxirribonucleico (DNA). O tamanho do genoma refere-se à quantidade total de genes contidos no DNA, sendo tipicamente medido em termos de massa ou pelo número total de pares de bases de nucleótidos (tipicamente em milhões de pares de bases ou megabases - abreviadamente Mb ou Mbp). Portanto, a soma total dos genes é chamada de genoma.

A Figura 3.30 mostra aproximadamente o número de genes que codificam as proteínas de vários genomas eucariotos já completamente sequenciados. As funções de cerca de metade desses genomas são conhecidas ou foram previstas com base na comparação de sequências.

Todavia, a complexidade de um organismo não é diretamente proporcional ao tamanho do genoma: observe que o número de genes que codificam proteínas em diferentes orga-

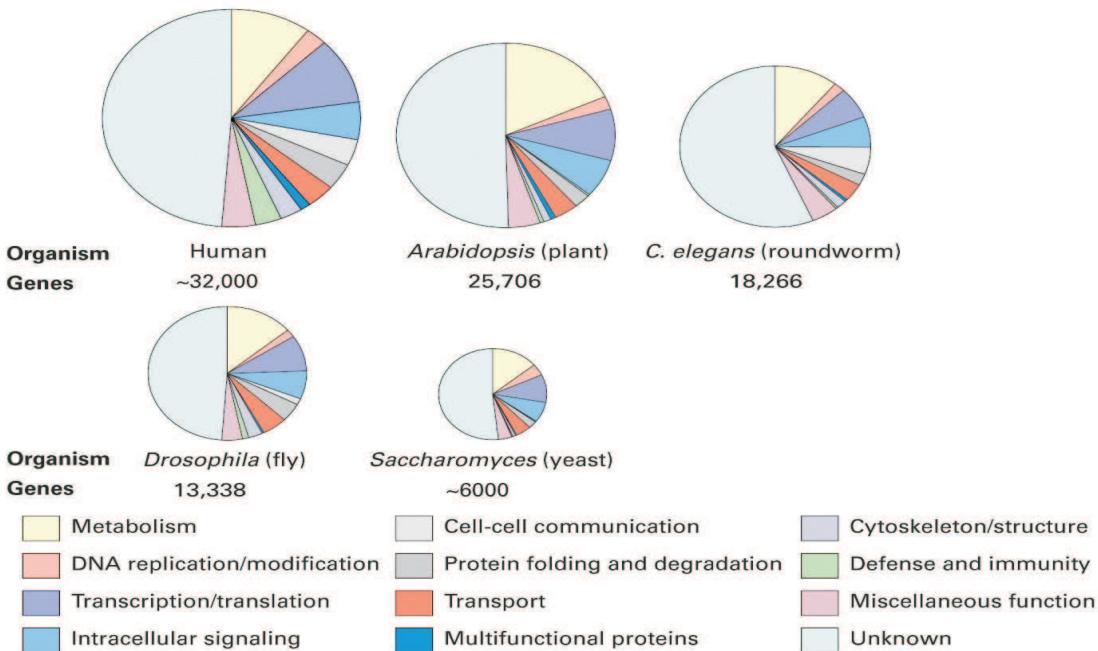


Figura 3.30: Comparação do número e dos tipos de proteínas codificadas no genoma de diferentes eucariotos. Lodish *et al.*, *Molecular Cell Biology*, 5th Edition.

nismos não é proporcional ao esperado por sua complexidade biológica. O homem tem pouco menos do que o dobro dos genes do *C.elegans*, o que parece ser completamente inexplicável devido à grande diferença entre esses organismos, [26].

Um novo genoma se origina pela fecundação, durante a reprodução. As células sexuais, diferentemente de todas as outras, têm apenas metade do genoma do indivíduo. Essa metade não é nem masculina nem feminina, mas uma combinação delas. Para isso, os órgãos reprodutores do organismo adulto combinam aleatoriamente partes do genoma de sua metade masculina com partes de sua metade feminina, em um processo diferente das outras células do corpo, gerando assim células sexuais. No ser humano, as combinações possíveis são da ordem de 3 bilhões elevados ao quadrado. A probabilidade de repetir uma combinação é praticamente nula. Cada célula sexual de um mesmo indivíduo leva uma sequência única, que nunca se formou antes e, depois, jamais voltará a formar-se, [57].

O sequenciamento do genoma é a técnica utilizada para determinar em que ordem as bases (letras) contidas no DNA se encontram. Quando se diz que um genoma foi sequenciado, queremos dizer que foi determinada a ordem em que as informações (genes) estão colocadas no genoma. Com o sequenciamento dos genomas, podemos obter in-

formações sobre a linha evolutiva dos organismos (saber quem tem o DNA mais parecido com quem). A utilidade mais evidente e imediata para o genoma humano é a de permitir conhecer as causas da maioria das doenças. O seu conhecimento poderá permitir diagnosticar e curar muitas delas, poderá resultar em novos métodos de diagnóstico, na formulação de novos medicamentos, vacinas, na prevenção e tratamentos mais eficazes contra doenças ou pragas. Os principais benefícios destas investigações só chegarão quando forem descobertas as funções de cada gene humano.

Cromossomo - O cromossomo é constituído por uma longa fita dupla de DNA. O DNA é o material que constitui os genes. Os genes se enfileiram nos cromossomos, quase sempre desenhados na forma de X ou Y, como bastonetes listrados unidos por uma espécie de cintura (os centrômeros), veja Figura 3.31. Os cromossomos só assumem essa conformação durante uma curta fase da divisão celular, como abordado no Capítulo 2. São poucos minutos, contra muitas horas em que os bastonetes desaparecem. Nesse período, os cromossomos ficam num estado menos compactado. A fita de DNA, que pode ter metros de comprimento se esticada, em lugar de se enrolar sucessivas vezes sobre si mesma para formar os bastonetes, fica “solta” dentro do núcleo. A espécie humana possui 46 cromossomos, onde 22 pares são semelhantes em ambos os sexos e são denominados autossomos. O par restante compreende os cromossomos sexuais, de morfologia diferente entre si, que recebem o nome de X e Y. No sexo feminino existem dois cromossomos X e no masculino existem um cromossomo X e um Y, como representados na Figura 3.31.

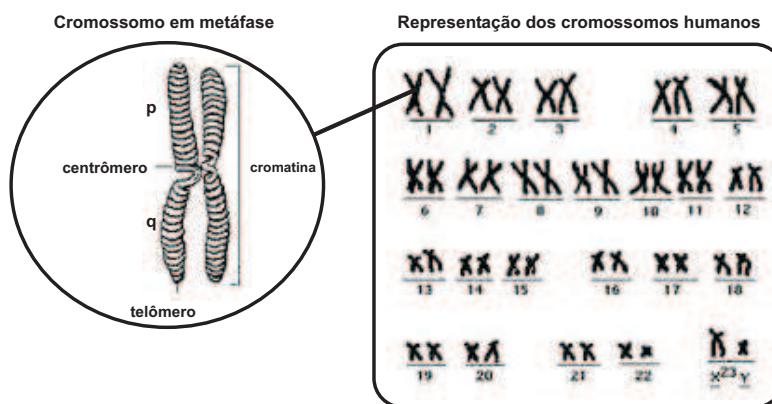


Figura 3.31: Genoma humano distribuído em 23 pares de cromossomos.

O estudo morfológico dos cromossomos mostrou que há dois exemplares idênticos de cada um em cada célula diplóide. Portanto, nos núcleos existem pares de cromossomos homólogos. Denomina-se **n** o número básico de cromossomos de uma espécie, temos

que as células diplóides apresentarão em seu núcleo **2n** cromossomos e as haplóides **n** cromossomos. O centrômero ou constrição primária é o ponto de referência citológico básico, dividindo os cromossomos em dois braços: **p** para o braço curto e **q** para o longo. Os braços são indicados pelo número do cromossomo seguido de **p** ou **q**; por exemplo, **1p** é o braço curto do cromossomo 1.

Os representantes de cada par desses cromossomos são chamados de cromossomos homólogos. Em cada espécie, há um número definido de cromossomos. Alterações em seu número ou disposição de genes podem resultar em mutações genéticas. Quando ocorrem mutações em células germinativas (óvulo ou espermatozóide), as mudanças podem ser transmitidas para as gerações futuras. As mutações que afetam as células somáticas podem resultar em certos tipos de câncer.

Gene - O gene é a unidade fundamental da hereditariedade. Cada gene é formado por uma sequência específica de ácidos nucléicos (biomoléculas mais importantes do controle celular, pois contêm a informação genética). Os genes controlam não só a estrutura e as funções metabólicas das células, mas também todo o organismo. Quando localizados em células reprodutivas, eles passam sua informação para a próxima geração.

O gene geralmente localiza-se intercalado com as sequências de DNA não codificadas por proteínas. Em organismos eucariotos superiores, os genes são compostos por múltiplas regiões codificadoras relativamente curtas denominadas **éxons** separadas por regiões não codificadoras denominadas **ítrons**. Não se sabe exatamente qual a função biológica das regiões não codificadoras. O DNA não codificado compõe cerca de 95% do genoma humano e, apesar de seu nome, ele é necessário para o funcionamento adequado dos genes.

O biólogo Tom Misteli, do Instituto Nacional do Câncer dos EUA, fez uma declaração sobre a importância da flor de Washington e da visualização da estrutura tridimensional do DNA dentro do núcleo, [55]. Ele comentou: “*É uma propriedade fundamental do genoma se organizar, dobrar-se de alguma maneira dentro do núcleo. Agora está ficando claro que há mais coisas no genoma que a sua sequência. Temos de descrever como o genoma se organiza, desvendar os mecanismos envolvidos na organização, e aí descobrir como a organização contribui para a função. Estamos ainda desenvolvendo as ferramentas para realmente enfrentar essas questões de modo sistemático*”.

É consenso geral da comunidade científica dizer que esse mundo de informações genéticas está armazenado em todas as células do organismo, e é como se cada célula fosse um “livro de receitas” completo de cada indivíduo.

Portanto, descrever o processo de como o genoma se organiza, armazena e transmite a informação genética não é uma tarefa fácil, e, inicialmente, vários questionamentos são levantados, dentre eles:

- 1) Existe uma “propriedade fundamental” do genoma se organizar?
- 2) Como essa propriedade simplesmente surgiu numa natureza caótica primordial?
- 3) Como explicar essa capacidade genômica de armazenar informação de maneira ultra-compacta?

Evidentemente, no presente trabalho, não temos a pretensão de responder os questionamentos levantados anteriormente. Tratam-se de problemas com alto grau de complexidade e envolvimento de várias áreas. Porém, de maneira bastante simples, fazemos uma conjectura do armazenamento da informação genômica: no exemplo em questão, fazemos uma representação planar do armazenamento do genoma humano, similar as informações que são armazenadas e organizadas em CD's, DVD's, *pen drives*, *hardwares*, descrita na Subseção 3.3.1. Nessa mesma direção, estamos interessados em relacionar os processos de armazenamento, organização, transmissão e identificação da informação biológica em formato de quadros similares àqueles utilizados em sistemas de redes locais de computadores.

Armazenamento da informação em CD-ROM

Na informática, é chamado de armazenamento o ato de gravar informações em algum dispositivo físico. Essa gravação de dados pode ser feita virtualmente, usando qualquer forma de energia. Um dispositivo de armazenamento retém informação, processa informação, ou ambos. Um dispositivo que somente guarda informação é chamado mídia de armazenamento. Dispositivos que processam informações (equipamento de armazenamento de dados) podem tanto acessar uma mídia de gravação portátil, ou podem ter um componente permanente que armazena e obtém dados. Tipos de dispositivos de armazenamento: por meios magnéticos (HDs, disquetes); por meios ópticos (CDs, DVDs, etc); por meios eletrônicos (SSDs) - *chip* - Exemplos: cartão de memória e *pen drive*. Frisando que: Memória RAM é um dispositivo de armazenamento temporário de informações.

O disco rígido (do inglês *Hard Disk*) é uma unidade de disco que armazena mais dados e é mais rápido que o disco flexível. É um dos principais dispositivos de armazenamento. Como esses dispositivos são capazes de armazenar muitos dados, às vezes são chamados de dispositivos de armazenamento de massa ou de memória de massa. É caracterizado como uma memória física, que é aquela na qual as informações não são perdidas quando o computador é desligado. As unidades de disco rígido são utilizadas para o armazenamento de dados que

devem estar disponíveis ao sistema de computação e ao usuário no momento em que a tarefa tiver que ser realizada, pois normalmente são unidades fixas. O sistema operacional e os programas a serem executados permanecem armazenados no disco rígido, pois podem ser acessados rapidamente, sem a necessidade da colocação de um disco flexível para cada tarefa. Para a ordenação dos dados no HD, é utilizado um esquema conhecido como geometria dos discos. Nele, o disco é dividido em **cilindros**, **trilhas** e **setores**. As **trilhas** são círculos que começam no centro do disco e vão até sua borda, como se estivessem um dentro do outro. Cada trilha é dividida em trechos regulares chamados **setores**. Cada setor possui uma determinada capacidade de armazenamento (geralmente, 512 *bytes*). Já os **cilindros** são a posição das cabeças sobre as mesmas trilhas de seus respectivos discos (lembrando que um HD possui vários discos). Os discos rígidos foram criados originalmente para serem usados em computadores em geral. Mas, atualmente as aplicações para esse tipo de disco foram expandidas e agora são usados em câmeras filmadoras, tocadores de música como *IPod*, *mp3 player*; PDAs; videogames, e até em celulares. Para exemplos em videogames temos o *Xbox360* e o *Playstation 3*.

O disquete já foi considerado um dispositivo com grande capacidade de armazenamento, especialmente devido ao pequeno tamanho dos arquivos. Atualmente, devido ao tamanho cada vez maior dos arquivos, e devido à existência de mídias não voláteis de maior capacidade (*zip disks*, cartões de memória, *flash drivers* USB, CD-R); além de existirem outras maneiras de guardar arquivos, como armazenamento distribuído e/ou compartilhamento de arquivos em redes locais, e-mail e disco virtual.

A seguir, faremos uma breve elucidação de redes locais de computadores, com o objetivo de identificarmos as principais convenções e regras inerentes no processo de comunicação e observarmos que este processo guarda semelhanças com processos de armazenamento, organização, transmissão e identificação da informação biológica.

Redes locais de computadores

Redes de computadores são estruturas físicas (equipamentos) e lógicas (programas, protocolos) que permitem que dois ou mais computadores possam compartilhar informações. Quando falamos em redes nos vem em mente, a internet, pois esta é a maior rede já desenvolvida pelo homem, e, muitos outros exemplos podem ser citados: caixa eletrônico, caixa de mercado, etc.

Um computador sozinho, sem estar conectado a nenhum outro computador, só terá acesso às suas informações (presentes em seu disco rígido) ou a informações que porventura venham a ele através de disquetes e CD's/DVD's/*pen drives*. Quando um computador está conectado a uma rede de computadores, ele pode ter acesso às informações que chegam a ele e às

informações presentes nos outros computadores ligados a ele na mesma rede.

As redes de computadores podem ser classificadas como: LAN (Rede Local) - Uma rede que liga computadores próximos (normalmente em um mesmo prédio ou, no máximo, entre prédios próximos); WAN (Rede Extensa) - Redes que se estendem além das proximidades físicas dos computadores. Como, por exemplo, redes ligadas por conexão telefônica, por satélite, ondas de rádio, etc. (Ex: A Internet, as redes dos bancos internacionais).

Uma definição mais técnica, enfatizando o aspecto de interconexão, considera como rede local de computador uma rede de comunicação de dados para interconectar computadores e recursos computacionais numa área geograficamente limitada. O processo de comunicação entre os diversos sistemas usuários (*hardware* e *software*) de uma rede local de computadores pressupõe a existência de um conjunto de regras e convenções que permita disciplinar a troca de informações. Essas regras comuns constituem os chamados **protocolos de comunicação**, ou, simplesmente, **protocolos**.

Com o intuito de reduzir a complexidade do projeto, a maioria das redes foi organizada como uma série de níveis ou camadas, que são colocadas uma sobre a outra. O número, o nome, o conteúdo e a função de cada camada difere de uma rede para outra. Em todas as redes, no entanto, o objetivo de cada camada é oferecer determinados serviços para as camadas superiores. A camada **n** de uma máquina se comunica com a camada **n** da outra máquina. Para isso acontecer, ela se baseia em um conjunto de convenções e regras que vão permitir gerenciar esta comunicação a qual foi denominada protocolo da camada **n**, ou, simplesmente, **protocolo n**.

Os protocolos de comunicação regulamentam diversos aspectos relativos à troca de informações entre sistemas usuários da rede, tais como:

- a) A unidade de informação trocada (bit, caracter, pacote, mensagem, etc.);
- b) A velocidade de transferência da informação;
- c) A identificação da origem e destino da informação (endereçamento);
- d) O controle de anomalia na recepção da informação, etc.

Além de garantir uma certa confiabilidade no processo de comunicação à distância, os protocolos permitem disciplinar, de maneira eficiente, o compartilhamento dos vários recursos (*hardware* e *software*) comuns envolvidos na comunicação.

Para facilitar a interconexão de sistemas de computadores, a ISO (*International Standards Organization*) desenvolveu um modelo de referência chamado OSI (*Open Systems Interconnection*), para que os fabricantes pudessem criar protocolos a partir desse modelo.

Portanto, protocolo define um conjunto de regras que permitem especificar aspectos da realização do serviço, particularmente, o significado dos quadros, pacotes ou mensagens trocadas entre as entidades pares de uma dada camada. O modelo de protocolos OSI é um modelo de sete camadas, apresentadas na Figura 3.32. As entidades que ocupam as mesmas camadas em diferentes máquinas são denominadas **pares**. São os pares que se comunicam utilizando o protocolo. Os dados não são transferidos diretamente entre os pares, pois não existe meio físico entre eles. Então cada camada transfere os dados para a camada inferior a ela, até alcançar a última camada. Após a última camada está o meio físico (meio de transmissão) através do qual se dá a comunicação.

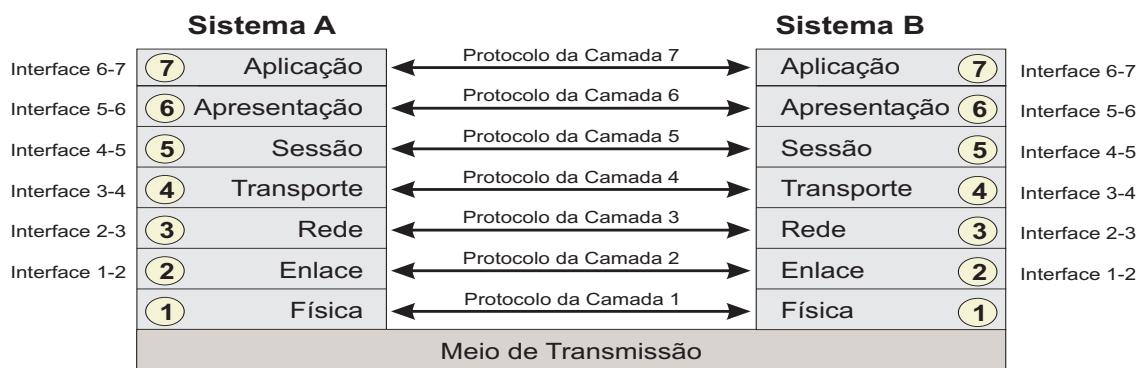


Figura 3.32: Modelo de sete camadas.

Em cada par de camadas adjacentes, há uma **interface** que define as operações e serviços que a camada inferior tem a oferecer para a camada superior a ela. O modo como estão estruturados os protocolos de comunicação em uma rede de computadores (de âmbito local ou a longa distância) define a chamada **arquitetura da rede**. As especificações da arquitetura devem conter informações suficientes para permitir o correto desenvolvimento da rede, tanto do ponto de vista do *software* quanto do *hardware*. Por outro lado, os detalhes de implementação dos mecanismos em cada camada, assim como as especificações detalhadas das interfaces, não fazem parte da definição da arquitetura da rede.

Na transmissão de dados, cada camada recebe/processa as informações passadas pela camada superior, acrescenta informações pelas quais ela seja responsável e passa os dados para a camada imediatamente inferior, como mostra a Figura 3.33. Esse processo é conhecido como encapsulamento. Na camada 4, Transporte, o dado enviado pelo aplicativo é dividido em pacotes. Na camada 2, Link de Dados, o pacote é dividido em vários quadros. Na recepção de um dado, o processo é o inverso. Um usuário que pede para o seu programa de e-mail baixar os emails, na verdade está fazendo com que o seu programa de e-mail inicie uma transmissão de dados com a camada 7 - Aplicação - do protocolo usado, pedindo para baixar

os e-mails do servidor de e-mails. Essa camada processa esse pedido, acrescenta informações de sua competência, e passa os dados para a camada imediatamente inferior, a camada 6 (Apresentação). Esse processo continua até a camada 1 (Física) enviar o quadro de dados para o cabeamento da rede, quando, então, atingirá o dispositivo receptor, que fará o processo inverso, até a sua aplicação, nesse exemplo, um programa servidor de e-mail.

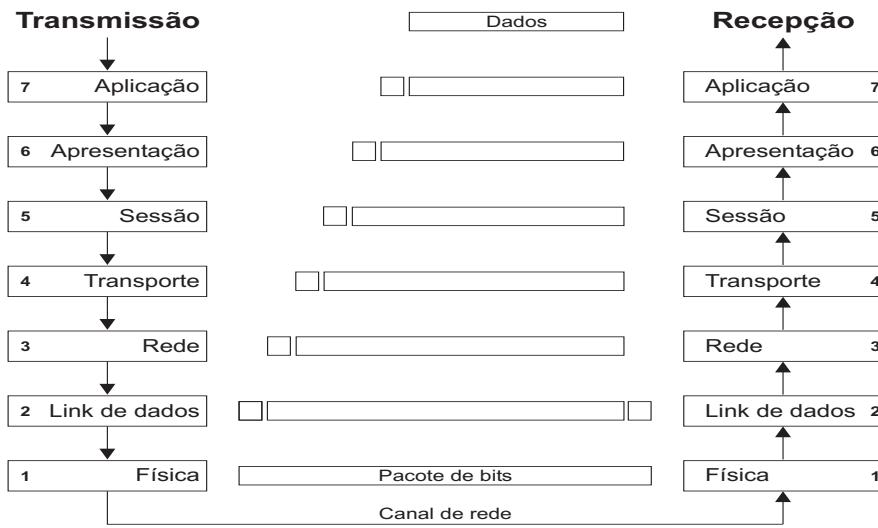


Figura 3.33: Procedimento da transmissão de dados.

Na prática, uma determinada camada do transmissor comunica-se diretamente com a mesma camada do dispositivo receptor. Por exemplo, a camada 4, Transporte, do dispositivo transmissor comunica-se diretamente com a camada 4 do dispositivo receptor e simplesmente ignora as comunicações efetuadas pelas camadas inferiores existentes. E assim por diante. Essa comunicação virtual é possível porque cada camada, durante a criação do pacote que será enviado através da rede, acrescentou o seu próprio cabeçalho, como está ilustrado na Figura 3.33.

As camadas do modelo OSI podem ser divididas em três grupos: aplicação, transporte e rede. As camadas de rede se preocupam com a transmissão e recepção dos dados através da rede e, portanto, são **camadas de nível inferior**. A camada de transporte é responsável por pegar os dados recebidos pela rede e repassá-los para as camadas de aplicação de uma forma compreensível, isto é, ela processa os pacotes de dados e transforma-os em dados quase prontos para serem usados pela aplicação. As camadas de aplicação, que são **camadas de alto nível**, colocam o dado recebido em um padrão que seja compreensível pelo programa (aplicação) que fará uso desse dado.

Um conjunto de dados enviado através da rede é denominado **quadro**. Dentro de um quadro, encontramos informações de endereçamento físico, como, por exemplo, o endereço

real de uma placa de rede. Logo, um quadro está associado às camadas mais baixas (1 e 2) do modelo OSI. Um conjunto de dados manipulados nas camadas 3 e 4 do modelo OSI é denominado **pacote**. No pacote, há informações de endereçamento virtual. Por exemplo, a camada 4 cria um pacote de dados para ser enviado pela rede e a camada 2 divide esse pacote em vários quadros que serão efetivamente enviados através da rede. Um pacote, portanto, contém a informação proveniente de vários quadros.

A **topologia** de rede descreve como é o *layout* de uma rede de computadores através da qual há o tráfego de informações, e também como os dispositivos estão conectados a ela. Essa estrutura de interconexão pode refletir tanto a localização geográfica das estações como o fluxo de informação gerado entre elas. No caso de redes locais de computadores, a interconexão das estações é realizada através das interfaces ou **nós** de comunicação. Em princípio, existem várias maneiras de se configurar a interconexão dos nós num ambiente localizado. Cada uma dessas maneiras apresenta características próprias, com diferentes implicações quanto ao desenvolvimento, operação e manutenção da rede. Essas implicações podem ser fatores determinantes da relação custo-desempenho global da rede. Topologias podem ser descritas fisicamente e logicamente. A topologia física é a verdadeira aparência ou *layout* da rede, enquanto a lógica descreve o fluxo dos dados através da rede. A diversidade da estrutura/organização das redes é composta basicamente pelas topologias de barramento, anel, estrela, híbrida, e malha, como ilustrado na Figura 3.34.

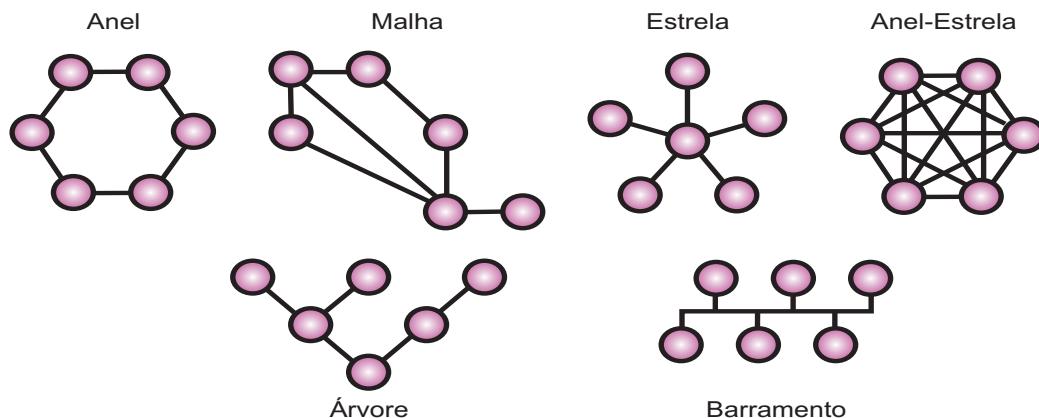


Figura 3.34: Modelos de topologias de redes de computadores.

Estabelecida a topologia da rede, em cada micro, o emulador de terminais, quando tem uma linha completa do usuário ou uma linha de controle (pedido de estabelecimento de ligação, pedido de fim de transmissão, etc.), a entrega ao protocolo do nível de enlace que, após “envelopá-la” com o cabeçalho desse nível, a envia ao servidor de comunicação. O formato da mensagem do nível de enlace é detalhado na Figura 3.35, que ilustra a estrutura

do quadro *ethernet*.

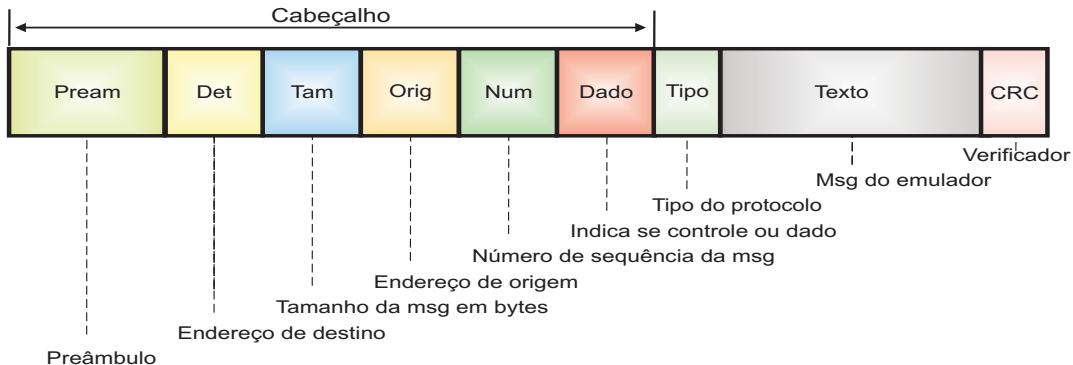


Figura 3.35: Quadro do formato de enlace da *ethernet*.

A seguir, faremos uma conjectura levantando algumas das similaridades entre o sistema biológico e o armazenamento de dados em CD-Rom e os conceitos e procedimentos usados na comunicação em redes locais de computadores, dentre eles, os protocolos usados na transmissão, a arquitetura e a topologia das redes e por fim, o entendimento de como os quadros são estruturados para posteriormente serem transmitidos.

Similaridades entre o sistema biológico e uma rede local de computadores

Como já mencionamos, o sistema biológico possui uma organização estrutural incrível. É formidável a sua capacidade em organizar e armazenar a informação biológica, bem como a capacidade de transmissão dessa informação.

Do simples para o complexo, moléculas organizadas formam as células, que, unidas, formam os tecidos, que, unidos, formam os órgãos, os quais, unidos, formam os sistemas orgânicos que criam e mantêm a vida. A vida de todos os organismos pluricelulares baseia-se na comunicação e nas interações entre as células que os compõem: corresponde à habilidade adquirida pelas células de comunicarem-se entre si. Esta comunicação é exigida pela regulação do funcionamento e organização dos tecidos, pelo controle de crescimento e divisão, bem como pela coordenação das diversas atividades. A existência de tão diversificados mecanismos de comunicação intercelular implica, na maior parte dos casos, em uma profunda especialização celular. Surgem assim, nos organismos superiores, grupos de células responsáveis pela produção e pela transmissão de mensagens: células nervosas, células neurosecretoras, e células secretoras de hormônios, entre outras. Necessita-se de um emissor, de uma mensagem transmitida, de uma via (canal ou rota metabólica), de um transmissor, e, por fim, de um receptor. Erros existentes no processamento de informação celular são responsáveis por doenças como o cancro, a autoimunidade e diabetes.

Com o melhor entendimento dos processos de sinalização celular, muitas doenças poderão ser tratadas de maneira mais eficaz, e, em teoria, tecidos artificiais poderão ser fabricados. O estudo de sistemas biológicos ajuda na compreensão da estrutura subjacente às redes de sinalização e a perceber como as mudanças nessas redes afetam a transmissão de informação [26] e [27].

No sistema biológico, existe a comunicação célula-célula, ou entre multiplas células ao mesmo tempo, e a comunicação interna da célula. Nesses processos de comunicação, as células constroem e degradam um grande número de moléculas e estruturas; alteram suas formas e se movimentam; recebem e emitem informações; crescem e se dividem; fabricam proteínas, morrem em consequência de agressões ou devido a um programa interno, etc. Enfim, são muitos os fatores responsáveis pela comunicação intercelular e intracelular. As estratégias de comunicação desenvolvidas pelas células de um organismo dependem de diversos fatores, dentre os quais se destaca a distância que separa as células comunicantes. Nesta perspectiva, podem distinguir-se três situações distintas: a) comunicação entre células adjacentes ou de contato; b) comunicação local, a curta distância; c) comunicação entre células muito distanciadas. Em [26] e [27], vimos que a comunicação por meio de sinais extracelulares geralmente envolve as seguintes etapas: (1) síntese e (2) liberação da molécula sinalizadora pela célula sinalizadora; (3) transporte do sinal até a célula-alvo; (4) ligação do sinal a uma proteína receptora específica, provocando sua ativação; (5) iniciação, pelo receptor ativado, de uma ou mais vias (rotas) intracelulares de transdução de sinal; (6) alterações específicas no funcionamento celular, no metabolismos ou no desenvolvimento; e (7) remoção do sinal, o que frequentemente leva ao encerramento da resposta celular.

Baseados nos sistemas de comunicação intercelular e intracelular existentes no mundo biológico, notamos que esses sistemas são altamente organizados, capazes e eficientes em armazenar e transmitir a informação biológica. Quando comparados com as redes locais de computadores, apresentam fortes evidências em suas semelhanças. Por exemplo - tanto no sistema biológico quanto na comunicação em redes locais de computadores, um conjunto de regras e convenções no processo de armazenamento e transmissão da informação são construídos de acordo com a necessidade de cada um com o objetivo de regulamentar a troca de informações entre as partes, os protocolos. Nessa mesma direção, o modelo de camadas, a arquitetura da rede, a distribuição topológica das entidades envolvidas, a formatação do quadro de enlace, dentre outros, apresentam semelhanças com o sistema biológico. Com isso, o conjunto de conceitos e propriedades pré-estabelecidos em redes locais de computadores pode ser utilizado no contexto biológico, e, portanto, deve ser estudado futuramente, melhor caracterizado em ambos aspectos, tornando possível uma modelagem matemática apropriada.

3.3.1 Proposta da arquitetura biológica do genoma humano

No presente trabalho, observamos que existem semelhanças entre o armazenamento da informação genômica e o armazenamento de dados em CD-ROM e portanto, de maneira bastante simples, conjecturamos que a informação genômica pode ser organizada e armazenada de maneira similar à informação armazenada em CD's (e seus descendentes) - em **setores**, **subsetores** e **trilhas**, como representado esquematicamente na Figura 3.36.

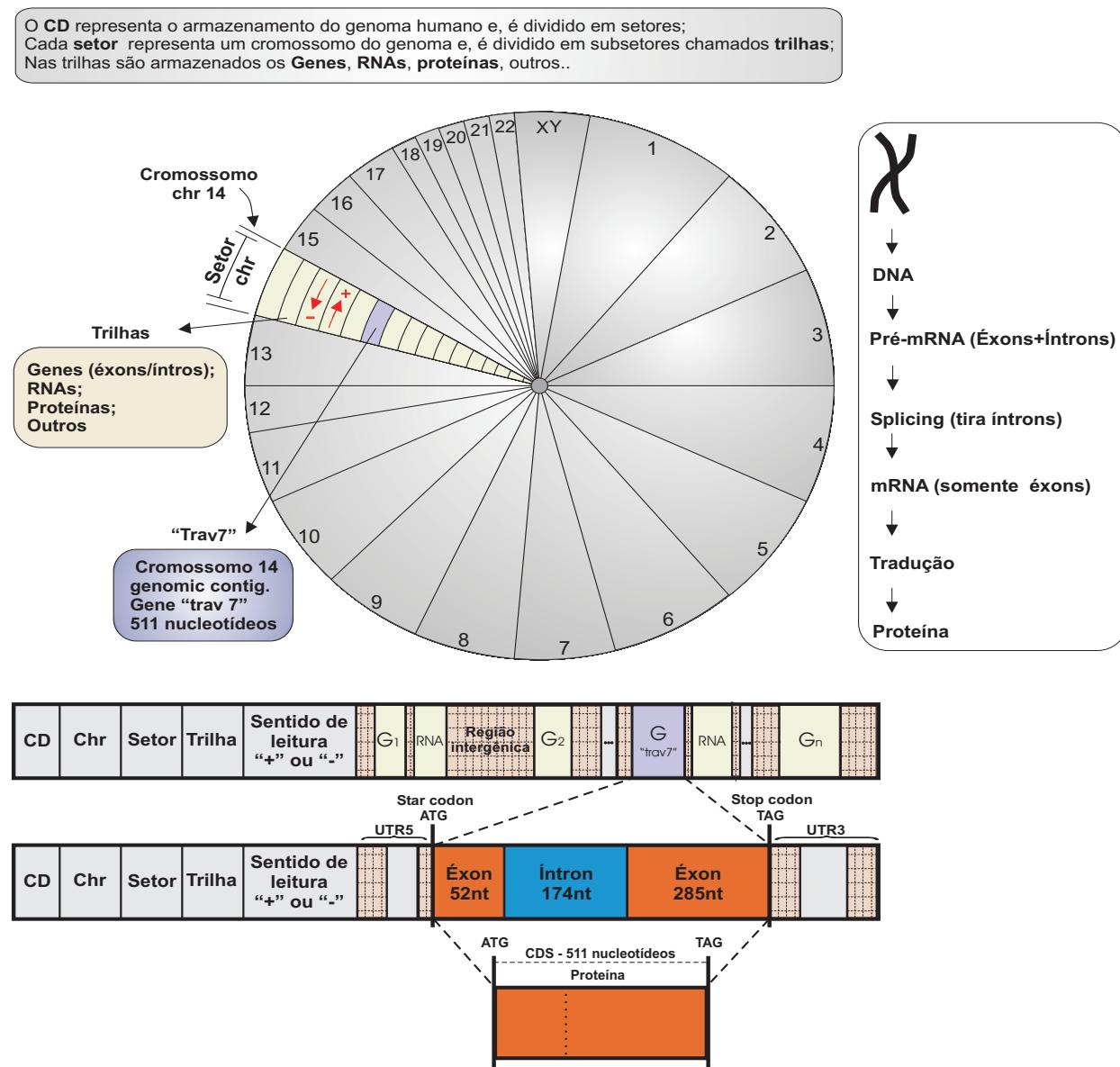


Figura 3.36: Representação planar do armazenamento de todo o conteúdo do genoma humano masculino em um CD, e a formação do quadro biológico.

A princípio, inferimos que o armazenamento da informação genômica pode ser realizado de

duas maneiras (descritas a seguir em I) e II)), tendo como ponto de partida as representações planares da estrutura primária do DNA (Figuras 3.37 a) e 3.38a)) já conhecidas na literatura.

I) Armazenamento do conteúdo de um único cromossomo em um único CD:

Na parte a) da Figura 3.37, temos uma representação das unidades primárias de cromossomos específicos identificadas via bandas G (produzidas pela coloração Giemsa e demarcadas em azul na figura), [26].

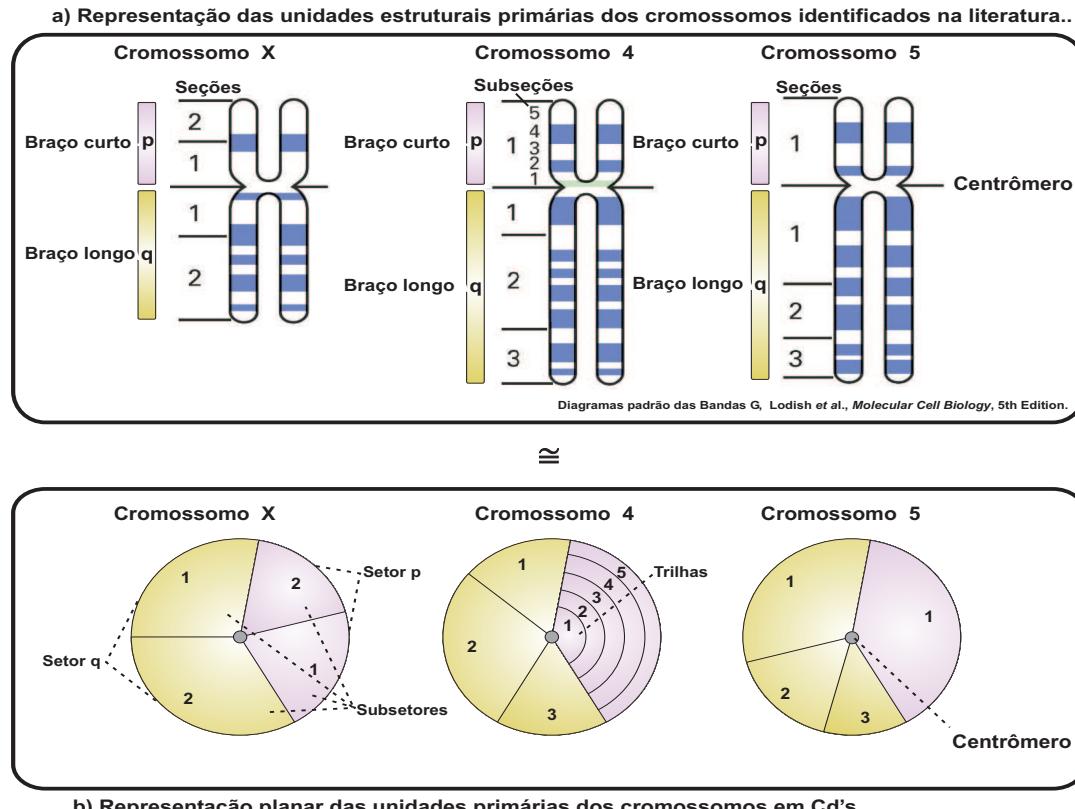


Figura 3.37: Representação planar do armazenamento do conteúdo de um único cromossomo em um único CD.

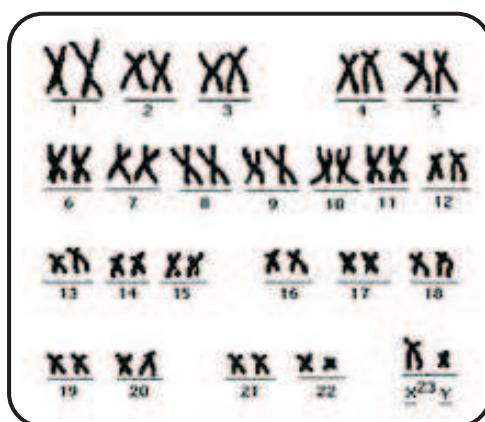
Observa-se que os cromossomos têm comprimentos e tamanhos variados e, por convenção, p denota o braço curto do cromossomo e q denota o braço longo. Cada braço é dividido em seções maiores (1, 2, etc) e subseções. O braço curto do cromossomo 4, por exemplo, tem 5 subseções. Na parte b) da Figura 3.37, temos uma representação planar de três CD's, na qual fazemos uma analogia aos três cromossomos representados na parte a) da figura. Cada CD está dividido em setores p e q (representando os braços do cromossomo), subsetores (representando as seções maiores do cromossomo) e trilhas (representando as subseções). Dessa maneira inferimos que cada CD tem a capacidade de armazenar

de maneira organizada o conteúdo das partes de cada cromossomo, tais como: genes, RNAs, proteínas e outros.

II) Armazenamento do conjunto de cromossomos de uma célula em um único CD:

Na parte a) da Figura 3.38 temos uma representação planar do conjunto cromossômico típico masculino humano, que são os 22 pares de cromossomos autossomos e 1 par de cromossomos sexuais (X e Y). Como já comentado no capítulo anterior, durante a mitose e a meiose, os cromossomos estão condensados e tornam-se visíveis ao microscópio óptico e, portanto essa representação planar do conjunto de cromossomos de várias espécies é bem conhecida na literatura. Na parte b) da Figura 3.38, temos uma representação planar de um CD, a partir da qual inferimos que o conjunto de cromossomos de uma célula pode ser compartimentalizado neste CD. O CD é dividido em setores de acordo com a quantidade de cromossomos (ou pares de cromossomos) da espécie em estudo, sendo que cada setor armazena o conteúdo de cada cromossomo (ou o par de cromossomos), e que novamente, os setores são divididos em subsetores chamados trilhas, facilitando assim o armazenamento e a organização do conteúdo do genoma celular com relação aos genes, RNAs, proteínas e outros.

a) Representação planar do conjunto de cromossomos humanos masculino (X e Y)



b) Representação planar do armazenamento do conjunto de cromossomos humanos masculino (X e Y) em CD.

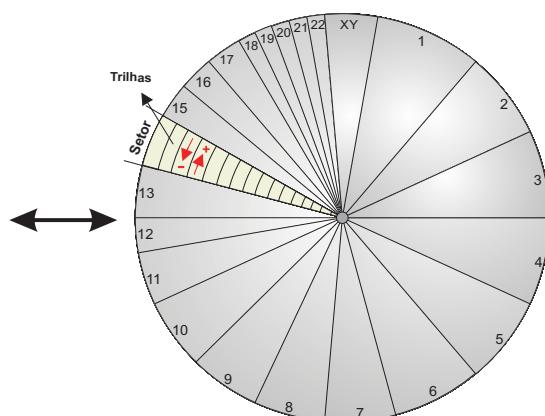


Figura 3.38: Representação planar do armazenamento do conjunto de cromossomos de uma célula em um único CD.

Como o tamanho do genoma, bem como o número de cromossomos, genes e proteínas, varia entre as espécies, essas variações devem ser consideradas durante a modelagem. Com isso, qualquer espécie pode ser modelada usando as conjecturas anteriormente citadas, devendo obedecer as suas particularidades.

No exemplo em questão, sugerimos o armazenamento de todo o conteúdo do **genoma humano masculino** em CD, como ilustrado na Figura 3.36. Os 22 pares de cromossomos autossomos e os cromossomos sexuais (X e Y) são armazenados separadamente cada um em um setor deste CD. Os setores são divididos em trilhas com a capacidade de armazenar e organizar todo conteúdo das partes de cada par de cromossomo: os genes, RNAs, ítrons, éxons, proteínas e outras sequências. Nos subsetores, também está contida a informação do sentido da leitura da sequência genômica “positiva (+)” que se refere a fita do DNA no sentido 5' – 3' ou “negativa (-)” que se refere a fita do DNA no sentido 3' – 5'. Na Figura 3.38, podem ser observados os diferentes tamanhos entre os cromossomos do genoma humano. Com isso, durante a modelagem, o tamanho de cada setor do CD representa o tamanho de cada cromossomo ou par de cromossomo do genoma (em termos de quantidade de conteúdo) que deve ser levado em consideração.

Outra conjectura é que o procedimento de armazenamento e transmissão da informação genômica (Figura 3.39) segue similarmente o procedimento da transmissão de dados (Figura 3.33), tornando possível a montagem de um quadro da informação genômica semelhante a montagem do quadro *ethernet* em redes locais de computadores (Figura 3.35), sendo que os protocolos são estabelecidos de acordo com o interesse.

Como o sistema biológico é detentor da informação, a comunicação intercelular e intracelular pode ser vista como uma comunicação de rede locais de computadores. O quadro biológico do genoma humano denominado **arquitetura biológica** ou **biological frame** pode ser uma alternativa. Observe na Figura 3.36 que a determinação de cada pacote na arquitetura biológica obedece “certas regras”, os chamados protocolos, tornando possível armazenar, localizar e transmitir determinada informação do genoma.

Neste caso, está contido na arquitetura biológica a identificação da célula (CD); a identificação do par de cromossomos (setor); a identificação do sentido de leitura da fita do DNA (trilha); a identificação dos genes e das regiões intergênicas (trilhas). No processo de identificação dos genes, é acrescentado ao quadro biológico os pacotes com informações relacionadas às ORFs (fase aberta de leitura dentro do mRNA - UTR5' e UTR3') do gene desejado, bem como, os éxons e os ítrons desse gene. Como estamos interessados em reproduzir, através dos CCEs, o gene “Trav7” localizado no cromossomo 14, inserimos essa informação no *Biological frame*. Finalmente, o pacote relacionado à proteína desejada pode ser montado com a indicação de quais éxons serão responsáveis para a montagem dessa proteína e com a indicação dos códons de inicialização (*start codon*) e finalização (*stop codon*). Logo, inferimos que o armazenamento, a localização e a transmissão da informação genômica podem ser modelados matematicamente de maneira análoga aos protocolos utilizados em redes de computadores.

A Figura 3.39 ilustra que o armazenamento e a transmissão da informação genômica

seguem no sentido contrário ao procedimento da transmissão de dados, parte a) da figura. Note na figura, que a arquitetura biológica é montada de acordo com os procedimentos biológicos realizados pela célula, como ilustrado na parte b) da figura. A caracterização apropriada para se determinar quais informações biológicas estão contidas em quais camadas da rede deve ser estudado futuramente, assim como a arquitetura e a topologia das redes da comunicação intracelular e intercelular. Porém, sabemos que nem toda informação biológica é conhecida, o que dificulta a modelagem matemática.

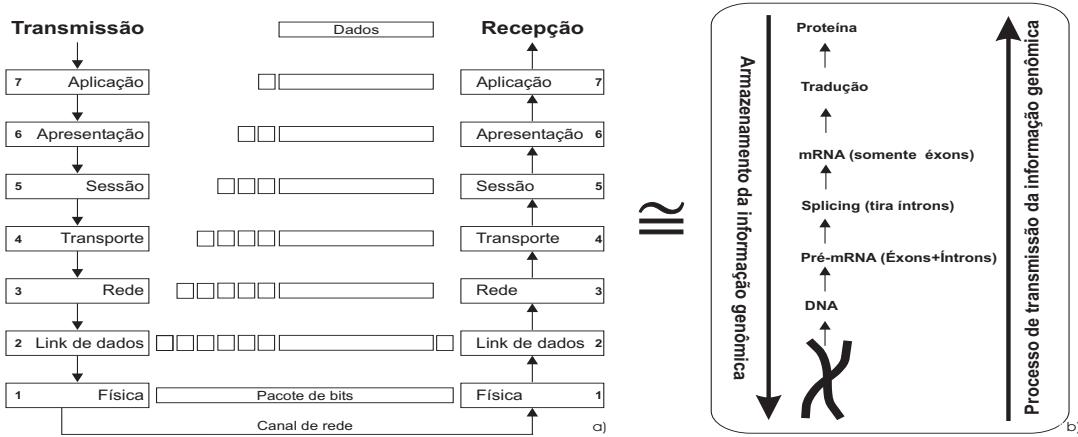


Figura 3.39: Proposta de armazenamento e transmissão da informação genômica análoga ao processo de transmissão de dados, porém no sentido contrário.

A sequência em nucleotídeos do gene “Trav7”, contida no cromossomo 14 (ilustrado no CD e na arquitetura biológica da Figura 3.36), é mostrada na Figura 3.40. No Capítulo 5, mostraremos que a sequência em nucleotídos do gene “Trav7” será identificada e reproduzida pelo CCE. Observe na sequência abaixo os códons de inicialização (ATG) e finalização (TAG) presentes no pré-mRNA mensageiro do gene “Trav7” indicando os sítios desse gene. Em seguida, o gene sofrerá um *splicing* (retirada do íntron) tornando-o mRNA “maduro” (somente os éxons) e, finalmente, com o processo de tradução, o(s) éxon(s) solicitado(s) pelo sistema biológico será(ão) responsável(eis) pela formação da proteína desejada pelo sistema biológico.

```
Homo sapiens - Trav7 predicted gene - chro mosome 14 genomic contig - 132 63_NT_026437
atggagaagatgcggagacctgtcctaattatattttgtctatgtctggct (52nt-E) gtaagttggggttctaa
agaactggggaccccaggagacatttattcaagtcccttggggatggggatgttagtctggacttacttgtcattt
cttggtttagatataaggaaaatttatgaaaggctcaaataatgtacatattgtacactgttttcgtatgtctttctgaata
g (174nt-I) gggcaaatggagaaaaccagggtggagcacgcctcatttctgggaccccagcaggagacgttgcc
tccatgagctgcacgtactctgtcaagtgtttacaatttgcagtggatcaggcaaatacaggatgggtccccaa
caaccttattccatgttccatgttgcagatggatgtggaaaggcagaaggactaaatgtacattactgttagatggaaagc
agttgtacattacagccgtgcagcctgaagattcagccacctatttctgtgttagatg (285nt-E)
```

Figura 3.40: Sequência em nucleotídeos do gene “Trav7”.

3.3.2 Arquitetura biológica do genoma do plasmídeo *Lactococcus lactis* pcl 21

Os plasmídeos (DNA extra-cromossomais), Figura 3.41 parte (b), são moléculas circulares duplas de DNA capazes de se reproduzir independentemente do DNA cromossômico parte (a). Ocorrem geralmente em bactérias e por vezes também em organismos eucarióticos unicelulares (ex: o anel de 2-micra em *Saccharomyces cerevisiae*) e células de eucariotas superiores.

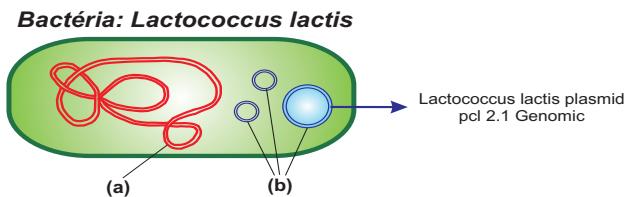


Figura 3.41: Ilustração de uma bactéria com plasmídeos no seu interior. (a) DNA cromossômico (b) Plasmídeos.

Os plasmídeos podem ser pequenos com baixo peso molecular, ou grandes, com alto peso molecular; quanto maior o peso molecular do plasmídeo, menor será o número de cópias possíveis de serem realizadas (cerca de três cópias por células); já os plasmídeos de baixo peso molecular podem realizar até cem cópias por célula. Eles replicam-se de forma independente do DNA cromossômico. Sua replicação acontece a cada divisão celular de forma a conservar pelo menos uma cópia em cada célula-filha, [26] e [27].

A presença dos plasmídeos e sua transferência entre as bactérias proporcionam a estes microorganismos uma transferência de caracteres essenciais à sua sobrevida. Citam-se como exemplos os plasmídeos, responsáveis pela resistência antimicrobiana, os plasmídeos responsáveis pela degradação de metais pesados, ou mesmo aqueles que codificam as toxinas bacterianas. Nos dias atuais, a transferência de resistência aos antibióticos entre as bactérias tem causado sérios problemas no meio hospitalar, gerando microorganismos multirresistentes; a cada dia, surgem novos antibióticos para combater estes microorganismos, o que gera novas expectativas quanto a atividade e possível desenvolvimento de resistência. Esta resistência é transferida entre as bactérias pelos plasmídeos, levando informações genéticas e novas características às bactérias receptoras, [63].

A seguir, mostramos o quadro biológico do genoma do plasmídeo *Lactococcus lactis* a partir da sequência em nucleotídos postada no NCBI, como mostrado na Figura 3.42.

Primeiramente, identificamos todas as partes desse genoma que estão ilustradas por cores diferentes. Em seguida, relacionamos cada uma dessas partes a um setor do CD, e, então, montamos o quadro biológico, denominado arquitetura biológica do genoma do plasmídeo

Figura 3.42: Sequência genômica em nucleotídeos.

Lactococcus lactis, como mostra a Figura 3.43.

As partes dos nucleotídos que estão em preto na sequência genômica (Figura 3.42) e em branco no quadro biológico (Figura 3.43) são sequências cuja funcionalidade biológica ainda é desconhecida. Porém, já são conhecidos o início e o fim dessas sequências - que são as regiões: L_1 com 715 nucleotídeos, L_3 com 104 nucleotídeos, L_5 com 12 nucleotídeos e L_9 com 113 nucleotídeos. A região L_2 com 168 nucleotídeos é identificada pela origem da replicação de uma fita do DNA, a região L_4 com 129 nucleotídeos é identificada pela origem da replicação da fita dupla do DNA, a região L_6 com 138 nucleotídeos é identificada pelo gene “Cob G”, a região L_7 com 76 nucleotídeos é identificada pelo RNA e, a região L_8 com 612 nucleotídeos é identificada gene “Rep B”.

Note que a leitura das sequências nas regiões identificadas, exceto a região L_8 , é realizada no sentido positivo 5' para 3' e, a leitura da sequência identificada na região L_8 é realizada no sentido negativo 3' para 5'.

Outra importante observação é com relação às partes identificadas na cor cinza. Em ambas as figuras, observe que essas partes fazem parte de duas regiões ao mesmo tempo. Por exemplo, o nucleotídeo **a** pertence tanto a sequência de nucleotídeo da região L_6 quanto a sequência de nucleotídeo da região L_7 . O mesmo acontece com a sequência **atgacagaaaaaaaaacta**, que é parte integrante das regiões L_7 e L_8 .

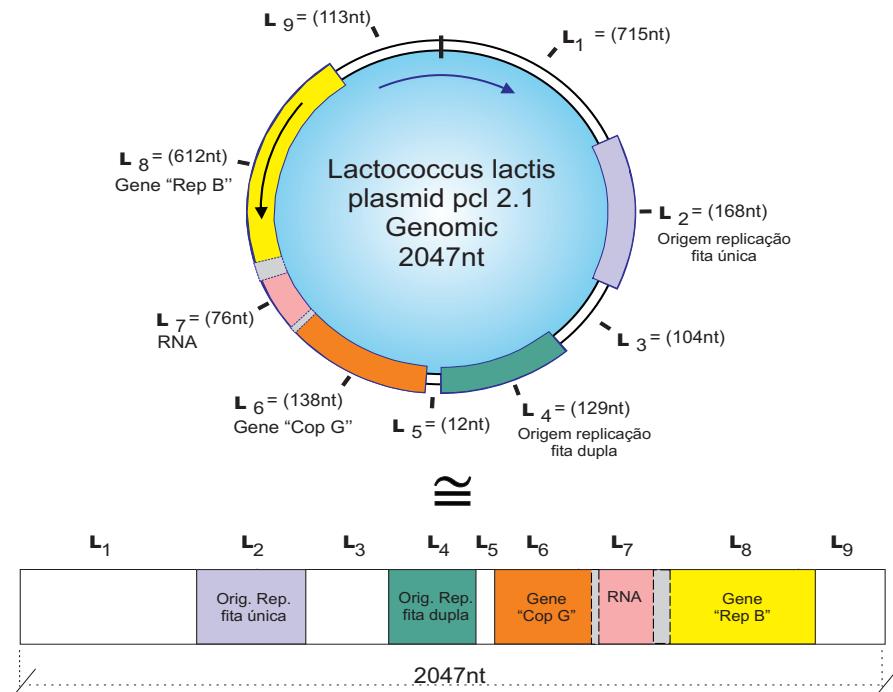


Figura 3.43: Genoma Plasmídeo com 2047 nucleotídeos de comprimento.

Devido a relevância dos plasmídeos, anteriormente citada, julgamos de suma importância a sua análise via CCEs, que mostraremos no Capítulo 5, bem como, a constatação da arquitetura biológica do genoma do plasmídeo *Lactococcus lactis*.

Capítulo 4

Algoritmo para a Identificação de Sequências de DNA

Como o objetivo deste trabalho é mostrar a existência de CCEs em sequências de DNA com diferentes comprimentos (21, 51, 63, 93, 255, 511, 1023 e 2047 nucleotídeos) e com características biológicas distintas (sequências de direcionamento, sequência de direcionamento ambígua, enzima, sinal interno, hormônio, ítron, DNA repetitivo, proteínas e *miRNA*), será necessário descrever a construção de códigos *G-linearidade*, tais como: os **códigos BCH sobre o corpo $GF(4^r)$** e, os **códigos BCH sobre o anel \mathbb{Z}_4** através das suas extensões de Galois, respectivamente.

Os códigos BCH sobre $GF(4^r)$ possuem parâmetros (n, k, d_H) , e são capazes de reproduzir sequências de DNA com comprimentos iguais ou submúltiplos de $n = (4^r - 1)$. Já os códigos BCH sobre \mathbb{Z}_4 possuem os mesmos parâmetros, mas são capazes de reproduzir sequências de DNA com comprimentos iguais ou submúltiplos de $n = (2^r - 1)$. Os parâmetros do código BCH, tanto para o corpo $GF(4^r)$ quanto para o anel \mathbb{Z}_4 , são denotados da seguinte maneira: n = comprimento das palavras-código (comprimento das sequências de DNA); k = dimensão do código (comprimento da sequência de informação responsável pela geração da sequência de DNA) e d_H = distância mínima do código (o menor número de posições em que quaisquer duas palavras-código diferem).

Durante a construção desses códigos, mostramos o procedimento para a reprodução de sequências de DNA com 1 e com 2 nucleotídeos diferindo da sequência original (*NCBI*). Todas as sequências de DNA utilizadas neste trabalho estão postadas no *NCBI* através do número do GI, independentemente da distância mínima d_H dos códigos. Assim, as sequências de direcionamento que foram identificadas e reproduzidas através de códigos *G-linearidade* (BCH primitivos sobre anéis diferindo da sequência original em 1 nucleotídeo) em [20], serão analisadas através dos códigos *G-linearidade* (BCH primitivos sobre anéis diferindo da sequência

original em 2 nucleotídeos) e, através dos códigos G -linearidade (BCH primitivos sobre corpos). Os conceitos e propriedades essenciais utilizados na construção destes códigos foram descritos no Capítulo 2.

Antes de iniciarmos a construção dos códigos anteriormente mencionados, estabelecemos algumas definições com o intuito de facilitar a leitura deste trabalho, da seguinte maneira:

- Denotamos por **palavra-código** a sequência de dígitos $\mathbf{a} = (a_1, a_2, \dots, a_n)$, onde a_i 's $\in GF(4)$ ou \mathbb{Z}_4 ;
- Denotamos por **Seq** a sequência de DNA do NCBI $\mathbf{b} = (b_1, b_2, \dots, b_n)$, onde a_i 's $\in GF(4)$ ou \mathbb{Z}_4 ;
- Denotamos por **padrões de erros** a diferença entre a palavra-código e a sequência de DNA do NCBI $D(\mathbf{a}, \mathbf{b})$, ou seja,
 - a) Quando $D(\mathbf{a}, \mathbf{b}) = 0$, onde $\mathbf{a} = \mathbf{b}$ (todos os componentes são iguais);
 - b) Quando $D(\mathbf{a}, \mathbf{b}) = 1$, onde $\mathbf{a} \neq \mathbf{b}$ (difere em uma única componente);
 - c) Quando $D(\mathbf{a}, \mathbf{b}) = 2$, onde $\mathbf{a} \neq \mathbf{b}$ (diferem em duas componentes).

Este capítulo está organizado da seguinte forma. Na Seção 4.1, apresentamos o algoritmo de identificação de sequências de DNA sobre corpos e suas extensões de Galois. Um exemplo da construção de código G -linearidade (BCH sobre corpo) pode ser visto na Subseção 4.1.1. Já na Seção 4.2, apresentamos o algoritmo de identificação de sequências de DNA sobre o anel \mathbb{Z}_4 através de extensões de Galois de graus 6, 8 e 10. As sequências com comprimentos iguais a $2^r - 1$ serão analisadas através dos **códigos G -linearidade (BCH primitivos)**: $(63, k, d_H)$ sobre $GR(4, 6)$, $(255, k, d_H)$ sobre $GR(4, 8)$ e $(1023, k, d_H)$ sobre $GR(4, 10)$ (Subseção 4.2.1), e as sequências cujos comprimentos são submúltiplos de $2^r - 1$ e divisíveis pela ordem dos elementos de $GR^*(4, r)$ serão analisadas através dos **códigos G -linearidade (BCH não primitivos)**: $(21, k, d_H)$ sobre $GR(4, 6)$, $(51, k, d_H)$ sobre $GR(4, 8)$ e $(93, k, d_H)$ sobre $GR(4, 10)$ (Subseção 4.2.2). Com isso, verificaremos se as sequências de DNA mencionadas na Tabela 1.1, do Capítulo 3, apresentam uma estrutura matemática de corpo e ou de anel, e, portanto, podem ser identificadas e reproduzidas através dos códigos G -linearidade. Os exemplos da reprodução de algumas sequências de DNA através dos códigos G -linearidade serão mostrados no decorrer deste capítulo. Os demais exemplos serão mostrados no Capítulo 5 e no Apêndice B.

4.1 Algoritmo de Geração de Códigos BCH sobre $GF(4^r)$

O algoritmo para a identificação de sequências de DNA tendo como base a construção de códigos G -linearidade (BCH sobre $GF(4^r)$), passo-a-passos, é descrito da seguinte maneira:

- Passo 1 - Especificar a estrutura matemática e o alfabeto do código;
- Passo 2 - Determinar a extensão de Galois;
- Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois;
- Passo 4 - Determinar a extensão do corpo $GF(4)$;
- Passo 5 - Determinar o grupo das unidades para o código BCH primitivo, quando o comprimento da sequência de DNA for igual a $n = (4^r - 1)$, ou, determinar o subgrupo das unidades para o código BCH não primitivo, quando o comprimento da sequência de DNA for um submúltiplo de $n = (4^r - 1)$;
- Passo 6 - Determinar o polinômio gerador da matriz G , $g(x)$:
 - 1º) Cálculo das raízes dos polinômios minimais;
 - 2º) Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \dots, n - 1$;
 - 3º) Cálculo dos polinômios geradores para todos os valores de t relacionados à distância de Hamming $d_H \geq 2t + 1$;
- Passo 7 - Determinar o polinômio gerador da matriz H , $h(x)$;
- Passo 8 - Determinar a matriz G e a sua transposta G^T ;
- Passo 9 - Determinar a matriz H e a sua transposta H^T ;
- Passo 10 - Rotular a sequência de DNA utilizando o Passo 1;
- Passo 11 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a, b) = 0$, $D(a, b) = 1$ e $D(a, b) = 2$;
- Passo 12 - Comparar todas as palavras-código armazenadas no Passo 11 com a sequência de DNA do NCBI e mostrar onde os erros ocorreram;
- Passo 13 - Voltar para o Passo 6 e determinar outro $g(x)$;
- Passo 14 - Repetir os Passos 7 ao Passo 11 para o $g(x)$ obtido no Passo 13, até que se esgotem todas as possibilidades de $g(x)$;
- Passo 15 - Voltar para o Passo 3 e escolher outro $p(x)$, e, então, repetir os Passos 4 ao 14 até esgotar todos os $p(x)$ do Passo 3;
- Passo 16 - Fim.

A seguir, apresentamos a construção de códigos BCH sobre $GF(4^r)$ de ordem $n = (4^r - 1)$, onde r é o grau da extensão de Galois. Em [66], mostramos que algumas sequências de DNA foram identificadas e reproduzidas através dos códigos G -linearidade (BCH sobre $GF(4^r)$) diferindo em 1 nucleotídeo das sequências do NCBI. No caso das sequências da fita simples de DNA, as bases do alfabeto do código genético (adenina (A), citosina (C), guanina (G) e timina (T) ou uracila(U)) são usadas para construir o código BCH sobre $GF(4^r)$, relacionando-as

ao alfabeto 4-ário, conforme detalhado no **Passo 1** do algoritmo.

No caso da dupla hélice de DNA, as bases complementares do alfabeto do código genômico é que são usadas para construir o código (AT, CG, GC, TA). Sendo assim, consideramos $q = p^k$, logo $GF(q) = GF(4)$. Note, no **Passo 5** do algoritmo, que as sequências de DNA com comprimentos iguais a n deverão ser analisadas através dos **códigos BCH primitivos** e as sequências de DNA cujos comprimentos são submúltiplos de n deverão ser analisadas através dos **códigos BCH não primitivos**. A seguir, descrevemos a construção do código BCH primitivo sobre $GF(4^r)$ apenas para o comprimento $n = 63$. O procedimento adotado para a construção dos demais comprimentos (255, 511, 1023 e 2047) é análogo ao desenvolvido no exemplo $(63, k, d_H)$ sobre $GF(4^r)$. Vale lembrar que o procedimento usado para a identificação da dupla hélice do DNA nos comprimentos citados também é análogo ao exemplo desenvolvido a seguir.

4.1.1 Código BCH $(n, k, d_H) = (63, k, d_H)$ sobre $GF(4^r)$

Considere a construção do código BCH primitivo sobre a estrutura algébrica de corpo e suas extensões de Galois com parâmetros $(n, k, d_H) = (63, k, d_H)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = (4^r - 1) = (4^3 - 1) = 63$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo diferindo da sequência original, onde r é o grau da extensão de Galois.

Passo 1 - Especificar a estrutura matemática e o alfabeto do código - O alfabeto 4-ário do código genético está relacionado ao conjunto de nucleotídeos correspondendo às bases adenina (A), citosina (C), guanina (G) e timina (T) ou uracila (U) denotado por $N = \{A, C, G, T/U\}$. Por este motivo, propomos a utilização do alfabeto 4-ário $GF(4) = \{0, 1, a, b\}$, obedecendo as operações de adição e multiplicação, como mostram as Tabelas 4.1 e 4.2, respectivamente.

| + | 0 | 1 | a | b |
|---|---|---|---|---|
| 0 | 0 | 1 | a | b |
| 1 | 1 | 0 | b | a |
| a | a | b | 0 | 1 |
| b | b | a | 1 | 0 |

Tabela 4.1: Adição em $GF(4)$.

| . | 0 | 1 | a | b |
|---|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | a | b |
| a | 0 | a | b | 1 |
| b | 0 | b | 1 | a |

Tabela 4.2: Multiplicação em $GF(4)$.

Passo 2 - Determinar a extensão de Galois - A condição necessária para que a fatoração de $x^n - 1$ em $GF(p^r)$, grupo das unidades, é que o comprimento da sequência de DNA

seja da forma $n = (4^r - 1)$, onde n contabiliza o número de elementos não nulos no corpo de Galois, $GF(4^r)$, ou equivalentemente, os elementos que possuem inverso e r denota o grau de extensão do corpo de Galois. Os valores de r serão utilizados na extensão do corpo $GF(4)$ no **Passo 4**. Nos casos em que as sequências de DNA possuem comprimento par da forma $n = (4^r + 2)$ o aminoácido metionina ou *stop* pode ser desconsiderado sem perda de generalidade.

Neste exemplo, apresentaremos a construção de um código BCH primitivo com comprimento $n = 63$. Logo, o grau r do polinômio primitivo a ser usado na extensão de Galois do corpo $GF(4)$ é $r = 3$, pois, $GF(4^r) = GF(64)$ e $n = (4^r - 1) = (4^3 - 1) = 63$. Este código será usado para gerar e reproduzir sequências de DNA com 63 nucleotídeos de comprimento. Para outros comprimentos n de sequências de DNA que satisfazem a relação $n = (4^r - 1)$, ou são submúltiplos de n , o grau da extensão de Galois e a ordem do corpo $GF(4^r)$ são determinados, de maneira análoga a este exemplo, como mostra a Tabela 4.3.

| Comprimento das sequências | Condição $n = (p^r - 1)$ | Grau da extensão de Galois r |
|----------------------------|--------------------------|--------------------------------|
| $n = 21, n = 63$ | $n = (4^3 - 1) = 63$ | $r = 3$ |
| $n = 51, n = 255$ | $n = (4^4 - 1) = 255$ | $r = 4$ |
| $n = 93, n = 1023$ | $n = (4^5 - 1) = 1023$ | $r = 5$ |

Tabela 4.3: Comprimento das sequências de DNA e o grau das extensões de Galois sobre $GF(4)$.

Assim, os códigos BCH primitivos com parâmetros $(63, k, d_H)$, $(255, k, d_H)$ e $(1023, k, d_H)$ e os códigos BCH não primitivos com parâmetros $(21, k, d_H)$, $(51, k, d_H)$ e $(93, k, d_H)$ sobre corpos nas extensões de Galois de grau $r = 3$, $r = 4$ e $r = 5$, respectivamente, são construídos analogamente ao exemplo que está sendo desenvolvido neste algoritmo. Note que as sequências de DNA com comprimentos iguais a 63, 255 e 1023 satisfazem a condição n , já as sequências de DNA com comprimentos iguais a 21, 51 e 93 são submúltiplos de n .

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Neste passo, os polinômios primitivos relacionados ao grau r da extensão de Galois são informados. Neste exemplo vamos usar o $p_{05}(x) = x^3 + ax^2 + ax + a$, da Tabela 4.4, para realizarmos a extensão do corpo no **Passo 4**.

Quanto maior for o grau da extensão de Galois, maior será a quantidade de $p(x)$ associados, aumentando assim a complexidade computacional para os cálculos, veja Tabela 4.5. Em [37], estes polinômios são encontrados.

| Polinômios Primitivos $p(x)$ | |
|-----------------------------------|-----------------------------------|
| $p_{01}(x) = x^3 + x^2 + ax + b$ | $p_{07}(x) = x^3 + bx^2 + bx + b$ |
| $p_{02}(x) = x^3 + x^2 + bx + a$ | $p_{08}(x) = x^3 + bx^2 + x + a$ |
| $p_{03}(x) = x^3 + ax^2 + bx + a$ | $p_{09}(x) = x^3 + x^2 + x + a$ |
| $p_{04}(x) = x^3 + bx^2 + ax + b$ | $p_{10}(x) = x^3 + ax^2 + bx + b$ |
| $p_{05}(x) = x^3 + ax^2 + ax + a$ | $p_{11}(x) = x^3 + ax^2 + x + b$ |
| $p_{06}(x) = x^3 + x^2 + x + b$ | $p_{12}(x) = x^3 + bx^2 + ax + a$ |

 Tabela 4.4: Polinômios primitivos da extensão de Galois de grau $r = 3$.

| Grau da extensão | Quantidade de polinômios |
|------------------|--------------------------|
| $r = 3$ | 12 |
| $r = 4$ | 32 |
| $r = 5$ | 120 |
| $r = 6$ | 288 |

Tabela 4.5: Quantidade de polinômios primitivos em cada extensão.

Passo 4 - Determinar a extensão do corpo $GF(4)$ - O corpo $GF(4^r)$ é obtido através do quociente de $GF(4)[x]$ por um ideal gerado por qualquer um dos $p(x)$ de grau $r = 3$. Considere o corpo de Galois $GF(4^r) = GF(4^3) = GF(64) = F_{64}$, dado por

$$GF(4^r)[x] \cong \frac{GF(4)[x]}{\langle p(x) \rangle} = \{a_0 + a_1x + a_2x^2 + \dots + a_rx^{r-1} : a_i \in GF(4)\}, \quad \delta p(x) = r.$$

Neste exemplo, $p(x)$ é o polinômio $p_{05}(x)$ do **Passo 3**, logo

$$GF(64)[x] = \frac{GF(4)[x]}{\langle x^3 + ax^2 + ax + a \rangle} = \{a_0 + a_1x + a_2x^2 : a_i \in GF(4)\},$$

onde

- i) a é raiz de $y^2 + y + 1$, então consideramos $\alpha = a$;
- ii) $b = a + 1$, então consideramos $\alpha^2 = b$.

Passo 5 - Determinar o grupo das unidades para o código BCH primitivo, quando o comprimento da sequência de DNA for igual a $n = (4^r - 1)$, ou, determinar o subgrupo das unidades para o código BCH não primitivo, quando o comprimento da sequência de DNA for um submúltiplo de $n = (4^r - 1)$ - Como, neste exemplo, o comprimento da sequência de DNA satisfaz a condição $n = (4^r - 1)$, então vamos determinar o grupo das unidades de $GF(64)$ para o código BCH primitivo.

Seja α uma raiz de $p_{05}(x) = x^3 + ax^2 + ax + a$, então, $\alpha^3 + a\alpha^2 + a\alpha + a = 0$, ou seja, $\alpha^3 = -a\alpha^2 - a\alpha - a$ módulo polinômio primitivo, temos que $\alpha^3 = a\alpha^2 + a\alpha + a$.

Apenas para facilitar os cálculos do grupo das unidades, substituimos α por f , logo: $\alpha^3 = a\alpha^2 + a\alpha + a \rightarrow f^3 = af^2 + af + a$. Assim, os elementos de $GF(64)$ mostrados na Tabela 4.6, são gerados sempre obedecendo as operações soma e produto módulo $f(x)$ (Tabelas 4.1 e 4.2) da seguinte maneira:

$$\begin{aligned}
 f^3 &= af^2 + af + a \rightarrow f^3 = \alpha^3 \rightarrow (a \ a \ a) \\
 f^4 &= f \cdot f^3 = f(af^2 + af + a) = af^3 + af^2 + af \\
 f^4 &= a(af^2 + af + a) + af^2 + af = bf^2 + bf + b + af^2 + af \\
 f^4 &= f^2(b + a) + f(a + b) + b = f^2 + f + b \rightarrow f^4 = \alpha^4 \rightarrow (b \ 1 \ 1) \\
 \vdots &= \vdots \\
 f^{63} &= f \cdot f^{62} = f(bf^2 + f + 1) \rightarrow f^{63} = \alpha^{63} \rightarrow (1 \ 0 \ 0)
 \end{aligned}$$

| $(\alpha^0 \alpha^1 \alpha^2)$ | $(\alpha^0 \alpha^1 \alpha^2)$ | $(\alpha^0 \alpha^1 \alpha^2)$ | $(\alpha^0 \alpha^1 \alpha^2)$ |
|---------------------------------------|---------------------------------------|---------------------------------------|---------------------------------------|
| $0 \rightarrow (0 \ 0 \ 0)$ | $\alpha^{16} \rightarrow (1 \ 0 \ 1)$ | $\alpha^{33} \rightarrow (a \ 0 \ b)$ | $\alpha^{50} \rightarrow (b \ b \ 0)$ |
| $1 \rightarrow (1 \ 0 \ 0)$ | $\alpha^{17} \rightarrow (a \ b \ a)$ | $\alpha^{34} \rightarrow (1 \ b \ 1)$ | $\alpha^{51} \rightarrow (0 \ b \ b)$ |
| $\alpha \rightarrow (0 \ 1 \ 0)$ | $\alpha^{18} \rightarrow (b \ 1 \ 0)$ | $\alpha^{35} \rightarrow (a \ b \ 1)$ | $\alpha^{52} \rightarrow (1 \ 1 \ a)$ |
| $\alpha^2 \rightarrow (0 \ 0 \ 1)$ | $\alpha^{19} \rightarrow (0 \ b \ 1)$ | $\alpha^{36} \rightarrow (a \ 0 \ 1)$ | $\alpha^{53} \rightarrow (b \ a \ a)$ |
| $\alpha^3 \rightarrow (a \ a \ a)$ | $\alpha^{20} \rightarrow (a \ a \ 1)$ | $\alpha^{37} \rightarrow (a \ 0 \ a)$ | $\alpha^{54} \rightarrow (b \ 0 \ 1)$ |
| $\alpha^4 \rightarrow (b \ 1 \ 1)$ | $\alpha^{21} \rightarrow (a \ 0 \ 0)$ | $\alpha^{38} \rightarrow (b \ 1 \ b)$ | $\alpha^{55} \rightarrow (a \ 1 \ a)$ |
| $\alpha^5 \rightarrow (a \ 1 \ b)$ | $\alpha^{22} \rightarrow (0 \ a \ 0)$ | $\alpha^{39} \rightarrow (1 \ a \ 0)$ | $\alpha^{56} \rightarrow (b \ 1 \ a)$ |
| $\alpha^6 \rightarrow (1 \ b \ 0)$ | $\alpha^{23} \rightarrow (0 \ 0 \ a)$ | $\alpha^{40} \rightarrow (0 \ 1 \ a)$ | $\alpha^{57} \rightarrow (b \ 0 \ a)$ |
| $\alpha^7 \rightarrow (0 \ 1 \ b)$ | $\alpha^{24} \rightarrow (b \ b \ b)$ | $\alpha^{41} \rightarrow (b \ b \ a)$ | $\alpha^{58} \rightarrow (b \ 0 \ b)$ |
| $\alpha^8 \rightarrow (1 \ 1 \ 0)$ | $\alpha^{25} \rightarrow (1 \ a \ a)$ | $\alpha^{42} \rightarrow (b \ 0 \ 0)$ | $\alpha^{59} \rightarrow (1 \ a \ 1)$ |
| $\alpha^9 \rightarrow (0 \ 1 \ 1)$ | $\alpha^{26} \rightarrow (b \ a \ 1)$ | $\alpha^{43} \rightarrow (b \ 0 \ 0)$ | $\alpha^{60} \rightarrow (a \ b \ 0)$ |
| $\alpha^{10} \rightarrow (a \ a \ b)$ | $\alpha^{27} \rightarrow (a \ 1 \ 0)$ | $\alpha^{44} \rightarrow (0 \ 0 \ b)$ | $\alpha^{61} \rightarrow (0 \ a \ b)$ |
| $\alpha^{11} \rightarrow (1 \ b \ b)$ | $\alpha^{28} \rightarrow (0 \ a \ 1)$ | $\alpha^{45} \rightarrow (1 \ 1 \ 1)$ | $\alpha^{62} \rightarrow (1 \ 1 \ b)$ |
| $\alpha^{12} \rightarrow (1 \ 0 \ a)$ | $\alpha^{29} \rightarrow (a \ a \ 0)$ | $\alpha^{46} \rightarrow (a \ b \ b)$ | $\alpha^{63} \rightarrow (1 \ 0 \ 0)$ |
| $\alpha^{13} \rightarrow (b \ a \ b)$ | $\alpha^{30} \rightarrow (0 \ a \ a)$ | $\alpha^{47} \rightarrow (1 \ b \ a)$ | |
| $\alpha^{14} \rightarrow (1 \ a \ b)$ | $\alpha^{31} \rightarrow (b \ b \ 1)$ | $\alpha^{48} \rightarrow (b \ a \ 0)$ | |
| $\alpha^{15} \rightarrow (1 \ 0 \ b)$ | $\alpha^{32} \rightarrow (a \ 1 \ 1)$ | $\alpha^{49} \rightarrow (0 \ b \ a)$ | |

Tabela 4.6: Elementos de $GF(64)$.

Passo 6 - Determinar o polinômio gerador da matriz G , $g(x)$ - Podemos agora construir códigos BCH de comprimento n sobre $GF(4)$, considerando que a distância mínima (distância de Hamming) do código, é no máximo, igual ao comprimento do código, ou seja, $d_H \leq n$. O algoritmo irá analisar todos os valores possíveis de d_H que estão relacionados com a capacidade de correção de erros estabelecida através da relação $d_H \leq 2t + 1$, onde t denota a quantidade de erros que o código será capaz de corrigir no processo de decodificação.

O polinômio gerador $g(x)$ do código de comprimento n tem como raízes os elementos na sequência $\{(\alpha^i), (\alpha^i)^p, (\alpha^i)^{p^2}, (\alpha^i)^{p^3}, \dots, (\alpha^i)^{p^{r/2-1} \pmod{n}}\}$, e, são dados por

$$g(x) = \text{mmc}(M_1(x), M_2(x), \dots, M_{n-1}(x)),$$

onde $M_i(x)$ é o polinômio minimal associado ao elemento α_i , $i = 1, 2, \dots, n - 1$ (α é um elemento primitivo em G_n), e mmc denota mínimo múltiplo comum.

No caso da palavra-código em questão, cujo comprimento é $n = 63$, os valores de $1 \leq t \leq 31$ serão analisados. Para cada valor de t , teremos uma distância equivalente e seus respectivos polinômios minimais envolvidos nos cálculos dos polinômios geradores $g(x)$, da seguinte maneira:

1º) Cálculo das raízes dos polinômios minimais:

Para cada polinômio minimal $M_i(x) = M_i$, com $i = 1, 2, \dots, 62$, temos:

$$\begin{aligned} M_1(x) &= \{(\alpha^1), (\alpha^1)^4, \dots, (\alpha^1)^{46/2-1 \pmod{63}}\} \rightarrow M_1(x) = \{\alpha, \alpha^4, \alpha^{16}\}, \\ M_2(x) &= \{(\alpha^2), (\alpha^2)^4, \dots, (\alpha^2)^{46/2-1 \pmod{63}}\} \rightarrow M_2(x) = \{\alpha^2, \alpha^8, \alpha^{32}\}, \\ \vdots &= \vdots \\ M_{62}(x) &= \{(\alpha^{62}), (\alpha^{62})^4, \dots, (\alpha^{62})^{46/2-1 \pmod{63}}\} \rightarrow M_{62}(x) = \{\alpha^{62}, \alpha^{59}, \alpha^{47}\}. \end{aligned}$$

| | | |
|--|--|--|
| $M_1 = (\alpha, \alpha^4, \alpha^{16})$ | $M_{22} = (\alpha^{22}, \alpha^{25}, \alpha^{37})$ | $M_{43} = (\alpha^{43}, \alpha^{46}, \alpha^{58})$ |
| $M_2 = (\alpha^2, \alpha^8, \alpha^{32})$ | $M_{23} = (\alpha^{23}, \alpha^{29}, \alpha^{53})$ | $M_{44} = (\alpha^{44}, \alpha^{50}, \alpha^{11})$ |
| $M_3 = (\alpha^3, \alpha^{12}, \alpha^{48})$ | $M_{24} = (\alpha^{24}, \alpha^{33}, \alpha^6)$ | $M_{45} = (\alpha^{45}, \alpha^{54}, \alpha^{27})$ |
| $M_4 = (\alpha^4, \alpha^{16}, \alpha)$ | $M_{25} = (\alpha^{25}, \alpha^{37}, \alpha^{22})$ | $M_{46} = (\alpha^{46}, \alpha^{58}, \alpha^{43})$ |
| $M_5 = (\alpha^5, \alpha^{20}, \alpha^{17})$ | $M_{26} = (\alpha^{26}, \alpha^{41}, \alpha^{38})$ | $M_{47} = (\alpha^{47}, \alpha^{62}, \alpha^{59})$ |
| $M_6 = (\alpha^6, \alpha^{24}, \alpha^{33})$ | $M_{27} = (\alpha^{27}, \alpha^{45}, \alpha^{54})$ | $M_{48} = (\alpha^{48}, \alpha^3, \alpha^{12})$ |
| $M_7 = (\alpha^7, \alpha^{28}, \alpha^{49})$ | $M_{28} = (\alpha^{28}, \alpha^{49}, \alpha^7)$ | $M_{49} = (\alpha^{49}, \alpha^7, \alpha^{28})$ |
| $M_8 = (\alpha^8, \alpha^{32}, \alpha^2)$ | $M_{29} = (\alpha^{29}, \alpha^{53}, \alpha^{23})$ | $M_{50} = (\alpha^{50}, \alpha^{11}, \alpha^{44})$ |
| $M_9 = (\alpha^9, \alpha^{36}, \alpha^{18})$ | $M_{30} = (\alpha^{30}, \alpha^{57}, \alpha^{39})$ | $M_{51} = (\alpha^{51}, \alpha^{15}, \alpha^{60})$ |
| $M_{10} = (\alpha^{10}, \alpha^{40}, \alpha^{34})$ | $M_{31} = (\alpha^{31}, \alpha^{61}, \alpha^{55})$ | $M_{52} = (\alpha^{52}, \alpha^{19}, \alpha^{13})$ |
| $M_{11} = (\alpha^{11}, \alpha^{44}, \alpha^{50})$ | $M_{32} = (\alpha^{32}, \alpha^2, \alpha^8)$ | $M_{53} = (\alpha^{53}, \alpha^{23}, \alpha^{29})$ |
| $M_{12} = (\alpha^{12}, \alpha^{48}, \alpha^3)$ | $M_{33} = (\alpha^{33}, \alpha^6, \alpha^{24})$ | $M_{54} = (\alpha^{54}, \alpha^{27}, \alpha^{45})$ |
| $M_{13} = (\alpha^{13}, \alpha^{52}, \alpha^{19})$ | $M_{34} = (\alpha^{34}, \alpha^{10}, \alpha^{40})$ | $M_{55} = (\alpha^{55}, \alpha^{31}, \alpha^{61})$ |
| $M_{14} = (\alpha^{14}, \alpha^{56}, \alpha^{35})$ | $M_{35} = (\alpha^{35}, \alpha^{14}, \alpha^{56})$ | $M_{56} = (\alpha^{56}, \alpha^{35}, \alpha^{14})$ |
| $M_{15} = (\alpha^{15}, \alpha^{60}, \alpha^{51})$ | $M_{36} = (\alpha^{36}, \alpha^{18}, \alpha^9)$ | $M_{57} = (\alpha^{57}, \alpha^{39}, \alpha^{30})$ |
| $M_{16} = (\alpha^{16}, \alpha, \alpha^4)$ | $M_{37} = (\alpha^{37}, \alpha^{22}, \alpha^{25})$ | $M_{58} = (\alpha^{58}, \alpha^{43}, \alpha^{46})$ |
| $M_{17} = (\alpha^{17}, \alpha^5, \alpha^{20})$ | $M_{38} = (\alpha^{38}, \alpha^{26}, \alpha^{41})$ | $M_{59} = (\alpha^{59}, \alpha^{47}, \alpha^{62})$ |
| $M_{18} = (\alpha^{18}, \alpha^9, \alpha^{36})$ | $M_{39} = (\alpha^{39}, \alpha^{30}, \alpha^{57})$ | $M_{60} = (\alpha^{60}, \alpha^{51}, \alpha^{15})$ |
| $M_{19} = (\alpha^{19}, \alpha^{13}, \alpha^{52})$ | $M_{40} = (\alpha^{40}, \alpha^{34}, \alpha^{10})$ | $M_{61} = (\alpha^{61}, \alpha^{55}, \alpha^{31})$ |
| $M_{20} = (\alpha^{20}, \alpha^{17}, \alpha^9)$ | $M_{41} = (\alpha^{41}, \alpha^{38}, \alpha^{26})$ | $M_{62} = (\alpha^{62}, \alpha^{59}, \alpha^{47})$ |
| $M_{21} = (\alpha^{21})$ | $M_{42} = ((\alpha)^{42})$ | |

Tabela 4.7: Raízes dos polinômios minimais sobre $GF(64)$.

2º) Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \dots, 62$:

Os polinômios minimais são calculados da seguinte maneira:

$$\begin{aligned} M_1(x) &= \{(x - \alpha)(x - \alpha^4)(x - \alpha^{16})\} = x^3 + ax^2 + ax + a, \\ M_2(x) &= \{(x - \alpha^2)(x - \alpha^8)(x - \alpha^{32})\} = x^3 + bx^2 + bx + b. \end{aligned}$$

De maneira análoga, os demais polinômios minimais são determinados. Note na Tabela 4.7, que vários polinômios minimais possuem as mesmas raízes. Portanto, esses polinômios minimais são iguais, veja Tabela 4.8.

| | | |
|-------------------------------------|-------------------------------------|-------------------------------------|
| $M_1(x) = M_4(x) = M_{16}(x)$ | $M_{11}(x) = M_{44}(x) = M_{50}(x)$ | $M_{27}(x) = M_{45}(x) = M_5(x)$ |
| $M_2(x) = M_8(x) = M_{32}(x)$ | $M_{13}(x) = M_{52}(x) = M_{19}(x)$ | $M_{30}(x) = M_{57}(x) = M_{39}(x)$ |
| $M_3(x) = M_{12}(x) = M_{48}(x)$ | $M_{14}(x) = M_{56}(x) = M_{35}(x)$ | $M_{31}(x) = M_{61}(x) = M_{55}(x)$ |
| $M_5(x) = M_{20}(x) = M_{17}(x)$ | $M_{15}(x) = M_{60}(x) = M_{51}(x)$ | $M_{42}(x)$ |
| $M_6(x) = M_{24}(x) = M_{33}(x)$ | $M_{21}(x)$ | $M_{43}(x) = M_{43}(x) = M_{58}(x)$ |
| $M_7(x) = M_{28}(x) = M_{49}(x)$ | $M_{22}(x) = M_{25}(x) = M_{37}(x)$ | $M_{47}(x) = M_{62}(x) = M_{59}(x)$ |
| $M_9(x) = M_{36}(x) = M_{18}(x)$ | $M_{23}(x) = M_{29}(x) = M_{53}(x)$ | |
| $M_{10}(x) = M_{40}(x) = M_{34}(x)$ | $M_{26}(x) = M_{41}(x) = M_{38}(x)$ | |

Tabela 4.8: Polinômios minimais que são iguais sobre $GF(64)$.

3º) Cálculo dos polinômios geradores para $1 \leq t \leq 31$:

O polinômio gerador para cada valor de t é dado por $g(x) = \text{mmc}\{M_1(x), M_2(x), \dots, M_{n-1}(x)\}$, formado pelos polinômios minimais que são diferentes entre si e possuem raízes $\alpha, \dots, \alpha^{2t}$, veja Tabela 4.9, onde $M_i(x) = M_i$.

| t | $d_H \geq 2t + 1$ | $\alpha^1, \dots, \alpha^{2t}$ | $g(x) = \text{mmc}(M_1(x), M_2(x), \dots, M_{n-1}(x))$ |
|----------|-------------------|----------------------------------|--|
| 31 | $d_H \geq 63$ | $(\alpha^1, \dots, \alpha^{62})$ | $g(x) = \text{mmc}(M_1, \dots, M_{62})$ |
| 30 | $d_H \geq 61$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = \text{mmc}(M_1, \dots, M_{60})$ |
| \vdots | \vdots | \vdots | \vdots |
| 05 | $d_H \geq 11$ | $(\alpha^1, \dots, \alpha^{10})$ | $g(x) = \text{mmc}(M_1, M_2, M_3, M_5, M_6, M_7, M_9, M_{10})$ |
| 04 | $d_H \geq 09$ | $(\alpha^1, \dots, \alpha^8)$ | $g(x) = \text{mmc}(M_1, M_2, M_3, M_5, M_6, M_7)$ |
| 03 | $d_H \geq 07$ | $(\alpha^1, \dots, \alpha^6)$ | $g(x) = \text{mmc}(M_1, M_2, M_3, M_5, M_6)$ |
| 02 | $d_H \geq 05$ | $(\alpha^1, \dots, \alpha^4)$ | $g(x) = \text{mmc}(M_1, M_2, M_3)$ |
| 01 | $d_H \geq 03$ | (α^1, α^2) | $g(x) = \text{mmc}(M_1, M_2)$ |

Tabela 4.9: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(63, k, d_H)$ BCH primitivos sobre $GF(64)$.

Considerando que a distância mínima do código seja $d_H = 3$, $g(x)$ é dado por:

$$\begin{aligned}
 g(x) &= \text{mmc}(M_1(x), M_2(x)) \\
 g(x) &= (x^3 + ax^2 + ax + a)(x^3 + bx^2 + bx + b) \\
 g(x) &= x^6 + x^5 + x^3 + x^2 + 1.
 \end{aligned}$$

O polinômio gerador $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ está relacionado com a matriz geradora G do código BCH sobre $GF(4)$ com parâmetros $(n, k, d_H) = (63, 57, 3)$.

De maneira análoga, os demais polinômios geradores para outros valores de t correspondentes a outras distâncias são determinados de acordo com a Tabela 4.9.

Passo 7 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H é obtido através da relação:

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{63} - 1}{x^6 + x^5 + x^3 + x^2 + 1}$$

$$h(x) = x^{57} + x^{56} + x^{55} + x^{50} + x^{47} + x^{43} + x^{42} + x^{40} + x^{39} + x^{36} + x^{34} + x^{33} + x^{31} + x^{29} + x^{28} + x^{27} + x^{25} + x^{24} + x^{23} + x^{22} + x^{19} + x^{18} + x^{14} + x^{12} + x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$$

onde os coeficientes do polinômio $h(x)$ pertencem a $GF(4)$.

Passo 8 - Determinar a matriz G e a sua transposta G^T - Determinado o polinômio gerador no **Passo 6**, construímos a matriz geradora G . Um código (n, k, d_H) é dito linear se, e somente se, todas as suas palavras-código formam um subespaço vetorial de dimensão k do espaço vetorial F_q^n , o conjunto das n -uplas do corpo F_q . Portanto, podemos representar este código matricialmente. O Teorema 2.3.5 estabelece que se $g(x)$ divide $(x^n - 1)$ e o grau de $g(x)$ é $(n - k)$, então a dimensão de $C = \langle g(x) \rangle$ é k . Se

$$g(x) = g_0 + g_1x + g_2x^2 + \dots + x^{n-k},$$

então a matriz geradora do código C é dada por:

$$G = \begin{pmatrix} g_0 & g_1 & g_2 & \dots & 1 & 0 & 0 & \dots & 0 \\ 0 & g_0 & g_1 & \dots & g_{n-k-1} & 1 & 0 & \dots & 0 \\ 0 & 0 & g_0 & \dots & g_{n-k-2} & g_{n-k-1} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & g_0 & g_1 & g_2 & \dots & 1 \end{pmatrix}.$$

Realizando os deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita, obtemos a matriz G com dimensão 57×63 :

$$G = \begin{pmatrix} 101101100 \\ 01011011000 \\ 00101101100 \\ 0001011011000 \\ \vdots & \ddots & \ddots & \vdots \\ 000101101100 \\ 000101101100 \\ 00010110110 \\ 00010110111 \end{pmatrix}.$$

A matriz G^T com dimensão 63×57 é determinada como sendo a troca da linha pela coluna.

Passo 9 - Determinar a matriz H e a sua transposta H^T - A matriz H é chamada **matriz verificação de paridade de** (n, k, d_H) , se qualquer vetor ortogonal a suas linhas pertence ao espaço vetorial das linhas da matriz geradora G associada e vice-versa. O código gerado pela matriz H é chamado **código dual** do código C , denotado por C^\perp . Determinado o polinômio $h(x)$ no **Passo 7**, obtemos a matriz H realizando os deslocamentos dos coeficientes do polinômio gerador $h(x)$ da direita para a esquerda.

Matriz H com dimensão 6×63 :

$$H = \begin{pmatrix} 00000111000010010001101100101101011101100010101001111111 \\ 00001110000100100011011001011010111011101100010101001111110 \\ 000111000010010001101100101101011101111011000101010011111100 \\ 001110000100100011011001011010111011110110001010100111111000 \\ 011100001001000110110010110101110111101100010101001111110000 \\ 111000010010001101100101101011101111011000101010011111100000 \end{pmatrix}.$$

A matriz H^T com dimensão 63×6 é determinada pela troca da linha pela coluna.

Passo 10 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, analisaremos se o código BCH sobre $GF(64)$ é capaz de reproduzir a sequência de DNA correspondente à sequência de direcionamento (SD = Seq.13) da proteína mitocondrial: *F1-ATPase delta subunit* - organismo: *Ipomoea batatas* - GI número: 217937 com comprimento $n = 63$ nucleotídeos.

Este passo determina as 24 permutações entre o alfabeto do código genético $N = \{A, C, G, T/U\}$ e o alfabeto do código BCH $GF(4) = \{0, 1, a, b\}$. Uma vez que o mapeamento entre $N \rightarrow GF(4)$ não é conhecido, a SD deve ser rotulada de acordo

com as 24 permutações apresentadas na Tabela 4.10, resultando em 24 possibilidades de SD's representada através das linhas da matriz P .

| $N \rightarrow GF(4)$ | $N \rightarrow GF(4)$ | $N \rightarrow GF(4)$ |
|-----------------------------|-----------------------------|-----------------------------|
| Caso 01-(A,C,G,T)=(0,1,b,a) | Caso 09-(A,C,G,T)=(0,1,a,b) | Caso 17-(A,C,G,T)=(0,a,1,b) |
| Caso 02-(A,C,G,T)=(0,b,1,a) | Caso 10-(A,C,G,T)=(0,b,a,1) | Caso 18-(A,C,G,T)=(0,a,b,1) |
| Caso 03-(A,C,G,T)=(1,0,a,b) | Caso 11-(A,C,G,T)=(1,0,b,a) | Caso 19-(A,C,G,T)=(1,b,0,a) |
| Caso 04-(A,C,G,T)=(1,a,0,b) | Caso 12-(A,C,G,T)=(1,a,b,0) | Caso 20-(A,C,G,T)=(1,b,a,0) |
| Caso 05-(A,C,G,T)=(a,1,b,0) | Caso 13-(A,C,G,T)=(a,1,0,b) | Caso 21-(A,C,G,T)=(a,0,1,b) |
| Caso 06-(A,C,G,T)=(a,b,1,0) | Caso 14-(A,C,G,T)=(a,b,0,1) | Caso 22-(A,C,G,T)=(a,0,b,1) |
| Caso 07-(A,C,G,T)=(b,0,a,1) | Caso 15-(A,C,G,T)=(b,0,1,a) | Caso 23-(A,C,G,T)=(b,1,0,a) |
| Caso 08-(A,C,G,T)=(b,a,0,1) | Caso 16-(A,C,G,T)=(b,a,1,0) | Caso 24-(A,C,G,T)=(b,1,a,0) |

Tabela 4.10: 24 permutações para $GF(4)$.

Seja a SD do NCBI igual a

{5'- ATGTTCAAGGCACCTCTCGACTCCTAGCTCGCGCCACCACAATGGGGTGGCGTCGCCCTTC -3'}

então

$$P = \left(\begin{array}{l} 0abaa10bb101a1aa1a1b01a11a0b1a1b1b110110100abbabb1ba1b1111aa1 \\ 0a1aab011b0bababab10babba01bab1b1bb0bb0b00a1111a11b1ab1bbbbaab \\ 1bab01aa010b0bb0b0a10b00b1a0b0a0a001001011baaaaabaa0ab0a0000bb0 \\ 1b0bba100a1ababba01abaab10aba0a0aa1aa11b0000b00a0ba0aaaabba \\ a0b001abb1a10100101ba10110ab101b1b11a1aa0bbbb0bb1b01b1111001 \\ a0100ba11bab0b0b0b1ab0bb0a1b0b1b1bbabbabaa01111011b10b1bbbb00b \\ b1a110baa0b01011010ab01001ba010a0a00b00b0bb1aaaa1aa0a10a0000110 \\ b1011ab00aba1a11a1a0ba1aa1b0a1a0a0aabaaabb10000100a01a0aaaa11a \\ 0bab10aa101b1bb1b1a01b11b0a1b1a1a110110100baaaaabaa1ab1a1111bb1 \\ 01a11b0aab0b1b11b1ba0b1b10ab1bababb0bb0b001aaaa1aab1babbb11b \\ 1abaa01bb010a0aa0a0b10a00a1b0a0b0b001001011abbabb0ba0b0000aa0 \\ 10b00a1bba1a0a00a0ab1a0aa01ba0ababaa1aa1a110bb0bb0bab0aaaa00a \\ ab0bb1a001a1b1bb1b10a1b11ba01b101011a11a1aab0000b0010b101111bb1 \\ a1011ba00bab1b1b1ba0ab1b1a0b1b0bbabbabaa10000100b01b0bb11b \\ ba1aa0b110b0a0aa0a01b0a00ab10a010100b0bb1111a1101a010000aa0 \\ b0100ab11aba0a00a0a1ba0aa0b1a0a1aabaaabb01111011a10a1aaaa00a \\ 0b1bba011a0ababbaba10aabab01aba1a1aa0aa0a00b111b11a1ba1aaaabba \\ 01b11a0bba0a1a11a1ab0a1aa10ba1ababaa0aa0a001bb1bbab1abaaaa11a \\ 1a0aab100b1bababab01babba10bab0b0bb1b11a0000a00b0ab0bbbaab \\ 10a00b1aab1b0b0b0ba1b0bb01ab0bababb1bb1b110aaaa0aab0babbb00b \\ ab1bb0a110a0b0bb0b01a0b00ba10b010100a00a0aab111b1101b010000bb0 \\ a1b110abb0a01011010ba01001ab010b0b00a00a0aa1bb1bb0b10b0000110 \\ ba0aa1b001b1a1aa1a10b1a11ab01a101011b1b1bb0000a0010a101111aa1 \\ b0a001baa1b10100101ab10110ba101a1a11b1bb0aaaa0aa1a01a1111001 \end{array} \right).$$

Passo 11 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a, b) = 0$, $D(a, b) = 1$ e $D(a, b) = 2$ - As linhas da matriz geradora do código linear (n, k, d_H) do Passo 8, formam uma base do espaço vetorial identificado como o código linear C . Assim, as combinações lineares das linhas de G são palavras-código de C . Dessa forma, o processo de codificação pode ser escrito como

$\mathbf{v} = \mathbf{u} \cdot G$, onde \mathbf{u} é a informação e \mathbf{v} é a palavra-código correspondente, no nosso caso as sequências de DNA a serem analisadas. Para toda palavra-código \mathbf{v} vale a relação

$$\mathbf{v} \cdot H^T = 0. \quad (4.1)$$

A capacidade de correção de erros de um código está relacionada com o número de palavras-código: no caso em questão temos 4^k palavras-código, onde $k = n - r$. Observe que, quanto maior for o valor de k , maior será o número de palavras-código, implicando assim em uma maior complexidade computacional para gerar todas as 4^k palavras-código. Para contornarmos este problema, que é classificado como um problema NP-completo, ao invés de gerarmos todas as palavras-código para compararmos com a sequência de DNA, consideramos que a sequência sob a aplicação de cada uma das 24 permutações do **Passo 10** é uma palavra-código, que corresponde ao padrão de erro denotado por $D(\mathbf{a}, \mathbf{b}) = 0$, ou seja, a diferença entre a palavra-código e a sequência de DNA do *NCBI* é nula. Assim, para determinarmos se cada uma dessas 24 possibilidades é de fato uma palavra-código usamos a relação $\mathbf{v} \cdot H^T = 0$, onde \mathbf{v} é a possível palavra-código e H^T é a transposta da matriz verificação de paridade determinada no **Passo 9**.

Ainda neste passo, vamos determinar quais das sequências de DNA apresentando os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença da sequência de DNA do *NCBI*, respectivamente, são palavras-código dos códigos (n, k, d_H) usando a equação (4.1), da seguinte maneira:

a) $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Para analisarmos as sequências de DNA que apresentam o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$, consideramos as 3 outras possibilidades de nucleotídeos em cada posição na sequência de DNA para cada permutação, resultando em um total de 3 possibilidades de nucleotídeo em cada posição, multiplicados pelo comprimento da sequência n e pelas 24 possibilidades de permutações, neste exemplo $3 \times 63 \times 24 = 4536$ possibilidades de para cada sequência de DNA, então usamos a equação (4.1). As palavras-código encontradas são armazenadas.

b) $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Para analisarmos as sequências de DNA que apresentam o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$, nos 24 casos de permutação, consideramos todas as combinações simples tomados 2 a 2 dos n nucleotídeos de comprimento da sequência, isto é, $C_{n,m} = \frac{n!}{m!(n-m)!}$. No exemplo onde $n = 63$ e $m = 2$ temos 1953×9 possibilidades resultando em 17577 palavras-código

para cada caso de permutação de cada sequência de DNA analisada. As palavras-código encontradas também são armazenadas.

Passo 12 - Comparar todas as palavras-código armazenadas no Passo 11 com a sequência de DNA do NCBI e mostrar onde os erros ocorreram - Neste passo, todas as palavras-código armazenadas no passo anterior estão rotuladas na forma do alfabeto do código, $GF(4) = \{0, 1, a, b\}$, e serão convertidas em nucleotídeos usando o rotulamento do alfabeto do código genético $N = \{A, C, G, T\}$. Após o rotulamento as palavras código são comparadas, uma-a-uma, com a sequência de DNA do NCBI mostrando onde os nucleotídeos diferem, e armazena os resultados.

Passo 13 - Voltar para o Passo 6 e determinar outro $g(x)$ - Neste passo, determinamos outro valor da distância mínima d_H , por exemplo $d_H = 5$ (Tabela 4.9) e utilizamos o mesmo procedimento, apresentado no **Passo 6**, para calcular o polinômio gerador relativo a esta distância.

Passo 14 - Repetir os Passos 7 ao Passo 11 para o $g(x)$ obtido no Passo 13, até que se esgotem todas as possibilidades de $g(x)$ - Neste passo, o algoritmo determina todas as palavras-código encontradas com nenhum, 1 e 2 nucleotídeos de diferença através de todos os polinômios geradores relativos à distância mínima $1 \leq d_H \leq n$, neste exemplo $1 \leq d_H \leq 63$, e armazena os resultados.

Passo 15 - Voltar para o Passo 3 e escolher outro $p(x)$, e, então, repetir os Passos 4 ao 14 até esgotar todos os $p(x)$ do Passo 3

Passo 16 - Fim.

Resultados

Como resultado do algoritmo no procedimento de geração da SD para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$ (diferindo em um nucleotídeo da sequência do NCBI), obtemos 24 palavras-código que correspondem às 24 permutações apresentadas na Tabela 4.10 do **Passo 10**. Essas palavras-código são diferentes em termos do alfabeto do código, $GF(4) = \{0, 1, a, b\}$, porém são iguais quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$ resultando em uma única sequência DNA como mostra a Tabela 4.11. Como as sequências que foram reproduzidas nos 24 casos são iguais, é suficiente considerarmos apenas um desses casos (Figura 4.1). Definimos esses casos de permutações como o **grupo das permutações**, S_4 . Veja, na Figura 4.1, a SD (Seq.13)¹ reproduzida pelo código G -linearidade ((63, 57, 3)

¹O número das sequências, por exemplo Seq.13, serve para facilitar a identificação da sequência no Capítulo 5 e no Apêndice B. As abreviações nas sequências devem ser lidas como: aaO = aminoácidos originais do;

| Sequência de DNA do NCBI Seq.13 <i>I.batatas</i> - Mitocondria - F1-ATPase delta subunit - GI número 217937 SD={5'- ATGTTCAGGCACTCTTCTCGACTCCTAGCTCGGCCACCACAATGGGGTG G CGTCGCCCTTC -3'} | | |
|---|-----------------------------|--|
| 24 casos de permutações | Palavras-código $n = 63$ | Sequência de DNA reproduzida |
| Caso 01-(A,C,G,T)=(0,1,b,a) | 0abaa ... 11aa1 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 02-(A,C,G,T)=(0,b,1,a) | 0a1aa ... bbaab | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 03-(A,C,G,T)=(1,0,a,b) | 1bab ... 00bb0 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 04-(A,C,G,T)=(1,a,0,b) | 1b0bb ... aabba | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 05-(A,C,G,T)=(a,1,b,0) | a0b00 ... 11001 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 06-(A,C,G,T)=(a,b,1,0) | a0100 ... bb00b | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 07-(A,C,G,T)=(b,0,a,1) | b1a11 ... 00110 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 08-(A,C,G,T)=(b,a,0,1) | b1011 ... aa11a | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 09-(A,C,G,T)=(0,1,a,b) | 0bab ... 11bb1 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 10-(A,C,G,T)=(0,b,a,1) | 01a11 ... bb11b | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 11-(A,C,G,T)=(1,0,b,a) | 1abaa ... 00aa0 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 12-(A,C,G,T)=(1,a,b,0) | 10b00 ... aa00a | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 13-(A,C,G,T)=(a,1,0,b) | ab0bb ... 11bb1 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 14-(A,C,G,T)=(a,b,0,1) | a1011 ... bb11b | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 15-(A,C,G,T)=(b,0,1,a) | balaa ... 00aa0 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 16-(A,C,G,T)=(b,0,1,a) | b0100 ... aa00a | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 17-(A,C,G,T)=(0,a,1,b) | 0b1bb ... aabba | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 18-(A,C,G,T)=(0,a,b,1) | 01b11 ... aa11a | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 19-(A,C,G,T)=(1,b,0,a) | 1a0aa ... bbaab | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 20-(A,C,G,T)=(1,b,a,0) | ab1bb ... 00bb0 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 21-(A,C,G,T)=(a,0,1,b) | ab1bb ... 00bb0 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 22-(A,C,G,T)=(a,0,b,1) | a1b11 ... 00110 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 23-(A,C,G,T)=(b,1,0,a) | ba0aa ... 11aa1 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |
| Caso 24-(A,C,G,T)=(b,1,a,0) | b0a00 ... 11001 | ATGTTCAGGCACTCTTCT ... TG A CGTCGCCCTTC |

 Tabela 4.11: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 63$ nucleotídeos sobre $GF(4)$.

BCH primitivo sobre $GF(4)$, S_4) através do polinômio primitivo $p_{05}(x) = x^3 + ax^2 + ax + a$ e do polinômio gerador $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$.

Seq.13 | *I. batatas* – Mitocondria - F1-ATPase delta subunit – GI número 217937

$$p_{05}(x) = x^3 + ax^2 + ax + a \quad - \quad g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--------------|-----|-----|-----|-----|
| aa0: | M | F | R | H | S | S | R | L | L | A | R | A | T | T | M | G | W | R | R | P | F |
| ntO: | ATG | TTC | AGG | CAC | TCT | TCT | CGA | CTC | CTA | GCT | CGC | GCC | ACC | ACA | ATG | GGG | TGG | CGT | CGC | CCC | TTC |
| RtO: | 0ba | bb1 | 0aa | 101 | b1b | b1b | 1a0 | 1b1 | 1b0 | a1b | 1a1 | a11 | 011 | 010 | 0ba | aaa | baa | 1ab | 1a1 | 111 | bb1 |
| RtG: | 0ba | bb1 | 0aa | 101 | b1b | b1b | 1a0 | 1b1 | 1b0 | a1b | 1a1 | a11 | 011 | 010 | 0ba | aaa | ba 0 | 1ab | 1a1 | 111 | bb1 |
| ntG: | ATG | TTC | AGG | CAC | TCT | TCT | CGA | CTC | CTA | GCT | CGC | GCC | ACC | ACA | ATG | GGG | T G A | CGT | CGC | CCC | TTC |
| aaG: | M | F | R | H | S | S | R | L | L | A | R | A | T | T | M | G | sto | R | R | P | F |

 Figura 4.1: Sequência de direcionamento com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Note que, na reprodução dessa SD, ocorreu uma troca de nucleotídeo na posição da trinca 17, alterando o aminoácido Triptofano (W) para o aminoácido de finalização (stop), denotado por (W17stop).

ntO = nucleotídeos originais; RtO = rotulamento original; RtG = rotulamento gerado; ntG = nucleotídeos gerados; aaG = aminoácidos gerados.

No Apêndice B, mostramos outras sequências de DNA com característica biológica diversa (sequências de direcionamento, sinal interno, íntron e proteína) e com comprimentos iguais a 63 e 255 nucleotídeos que foram identificadas e reproduzidas pelos códigos G -linearidade, sendo $((63, 57, 3), \text{BCH primitivo sobre } GF(4), S_4)$ e $(255, 247, 3), \text{BCH primitivo sobre } GF(4), S_4$ nas extensões de Galois de grau $r = 3$ e $r = 4$, respectivamente. Os polinômios primitivos e os polinômios geradores relacionados à geração dos códigos e à reprodução das sequências estão mencionados nas figuras.

Em virtude da complexidade computacional envolvida nos cálculos da construção em corpos ser ainda maior do que em anéis, focamos apenas na reprodução das sequências de DNA através dos códigos BCH sobre $GF(4)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença.

4.2 Algoritmo de Geração de Códigos BCH sobre $GR(4, r)$

Nesta seção, apresentamos a construção de códigos BCH primitivos e não primitivos, ambos, sobre anel local \mathbb{Z}_q de ordem $n = (p^r - 1)$, onde $q = p^k$, $p = k = 2$ e r é o grau da extensão de Galois. A principal diferença da construção de códigos cíclicos sobre anéis para a construção de códigos cíclicos sobre corpos está no fato de que as raízes do polinômio gerador dos códigos cíclicos sobre anéis encontram-se na extensão do anel \mathbb{Z}_q , ao invés de serem encontradas na extensão do corpo $\mathbb{F}_q \cong GF(p^r)$. Se a ordem do corpo base, p , e o comprimento das palavras-código, n , são relativamente primos, isto é, $mdc(p, n) = 1$, então $x^n - 1$ não apresenta multiplicidade de raízes.

A seguir, mostramos o algoritmo de codificação para geração de sequências de DNA que descreve a construção de códigos BCH sobre o anel de Galois \mathbb{Z}_4 , passo-a-passo:

- Passo 1 - Especificar a estrutura matemática e o alfabeto do código;**
- Passo 2 - Determinar a extensão de Galois;**
- Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois;**
- Passo 4 - Determinar a extensão do corpo $GF(2)$;**
- Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 ;**
- Passo 6 - Determinar o grupo das unidades para o código BCH primitivo, quando o comprimento da sequência de DNA for igual a $n = (2^r - 1)$, ou, determinar o subgrupo das unidades para o código BCH não primitivo, quando o comprimento da sequência de DNA for um submúltiplo de $n = (2^r - 1)$;**
- Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$:**
 - 1º) Cálculo das raízes dos polinômios minimais;

- 2º) Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \dots, n - 1$;
- 3º) Cálculo dos polinômios geradores para todos os valores de t relacionados à distância de Hamming $d_H \geq 2t + 1$;
- Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$;
- Passo 9 - Determinar a matriz G e a sua transposta G^T ;
- Passo 10 - Determinar a matriz H e a sua transposta H^T ;
- Passo 11 - Rotular a sequência de DNA utilizando o Passo 1;
- Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(a, b) = 0$, $D(a, b) = 1$ e $D(a, b) = 2$;
- Passo 13 - Comparar todas as palavras-código armazenadas no Passo 12 com a sequência de DNA do NCBI e mostrar onde os erros ocorreram;
- Passo 14 - Voltar para o Passo 7 e determinar outro $g(x)$;
- Passo 15 - Repetir os Passos 8 ao Passo 12 para o $g(x)$ obtido no Passo 14, até que se esgotem todas as possibilidades de $g(x)$;
- Passo 16 - Voltar para o Passo 3 e escolher outro $p(x)$, e, então, repetir os Passos 4 ao 14 até esgotar todos os $p(x)$ do Passo 3;
- Passo 17 - Fim.

No caso de sequências de DNA que possuem comprimentos iguais ou submúltiplos de $n = (2^r + 2)$, a metionina da primeira posição ou *stop* da última posição pode ser desconsiderada, uma vez que a matriz geradora possui uma coluna com os mesmos elementos.

4.2.1 Códigos BCH primitivos sobre $GR(4, r)$: Exemplos

As sequências com comprimentos $n = 2^r - 1$ serão analisadas através dos **códigos G-linearidade (BCH primitivos)**: $(63, k, d_H)$ sobre $GR(4, 6)$, $(255, k, d_H)$ sobre $GR(4, 8)$ e $(1023, k, d_H)$ sobre $GR(4, 10)$. A seguir, os exemplos da construção destes códigos serão apresentados.

Código BCH $(n, k, d_H) = (63, k, d_H)$ sobre $GR(4, 6)$

O código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (63, k, d_H)$ é capaz de gerar e reproduzir sequências de DNA com comprimentos $n = (2^r - 1) = (2^6 - 1) = 63$, com $D(\mathbf{a}, \mathbf{b}) = 1$ e com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos diferindo da sequência de DNA do NCBI, onde r é o grau da extensão de Galois.

Passo 1 - Especificar a estrutura matemática e o alfabeto do código - O alfabeto 4-ário do código genético está relacionado ao conjunto formado pelos nucleotídeos denotado por $N = \{A, C, G, T/U\}$ correspondendo a adenina, citosina, guanina e timina ou uracila. Por este motivo, utilizamos o alfabeto 4-ário denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$

obedecendo às operações de adição e multiplicação módulo 4, (Tabelas 4.12 e 4.13), o que lhe confere uma estrutura algébrica de anel.

| $+$ | 0 | 1 | 2 | 3 |
|-----|---|---|---|---|
| 0 | 0 | 1 | 2 | 3 |
| 1 | 1 | 2 | 3 | 0 |
| 2 | 2 | 3 | 0 | 1 |
| 3 | 3 | 0 | 1 | 2 |

Tabela 4.12: Adição módulo 4.

| \cdot | 0 | 1 | 2 | 3 |
|---------|---|---|---|---|
| 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 2 | 3 |
| 2 | 0 | 2 | 0 | 2 |
| 3 | 0 | 3 | 2 | 1 |

Tabela 4.13: Multiplicação módulo 4.

Passo 2 - Determinar a extensão de Galois - A condição necessária para que a fatoração de $x^n - 1$ em $GR^*(4, r)$, grupo das unidades, seja única, é que o comprimento da sequência de DNA seja ímpar da forma $n = (2^r - 1)$, onde n contabiliza o número de elementos não nulos no corpo de Galois, $GF(2^r)$, ou, equivalentemente, os elementos que possuem inverso, 2^r denota a cardinalidade do conjunto $GF(2^r)$, e r denota o grau de extensão do corpo de Galois. Os valores de r serão utilizados na extensão do corpo $GF(2)$ no **Passo 4**.

Neste exemplo, analisaremos a sequência de DNA de sinal interno, SI, de uma proteína mitocondrial, GI número 832917, cujo comprimento é $n = 63$ nucleotídeos (Seq.36). Logo, o grau r do polinômio primitivo a ser usado na extensão de Galois do corpo $GF(2)$ é $r = 6$, pois $n = 2^r - 1 = 2^6 - 1 = 63$. Portanto, esse valor de $r = 6$ será utilizado na extensão do corpo $GF(2)$ no Passo 4.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Neste passo, os $p(x)$ relacionados ao grau da extensão de Galois $r = 6$, (Tabela 4.14), são informados. Em [37], estes polinômios podem ser encontrados.

| Polinômios Primitivos $p(x)$ | | | | | |
|------------------------------------|--|--|--------------------------------------|--|--|
| $p_1(x) = x^6 + x + 1$ | | | $p_4(x) = x^6 + x^5 + x^2 + x + 1$ | | |
| $p_2(x) = x^6 + x^4 + x^3 + x + 1$ | | | $p_5(x) = x^6 + x^5 + x^3 + x^2 + 1$ | | |
| $p_3(x) = x^6 + x^5 + 1$ | | | $p_6(x) = x^6 + x^5 + x^4 + x + 1$ | | |

Tabela 4.14: Polinômios primitivos da extensão de Galois $r = 6$.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - O corpo $GF(2^r)$ é obtido através do quociente do anel de todos os polinômios com coeficientes em $GF(2)$, $GF(2)[x]$ por um

ideal gerado por qualquer um dos polinômios primitivos de grau $r = 6$. Neste passo, realizamos a extensão do corpo $GF(2)$ da seguinte maneira:

Considere o corpo de Galois $GF(2^r) = GF(2^6) = GF(64) = F_{64}$ dado por

$$\frac{F_2[x]}{\langle p(x) \rangle} \cong \frac{F_2[x]}{\langle x^6 + x + 1 \rangle} = \{a_0 + a_1x + a_2x^2 + \cdots + a_5x^5 : a_i \in F_2\}, \quad (4.2)$$

onde $p(x)$ é o polinômio primitivo $p_1(x)$ do **Passo 3**.

Seja α um elemento primitivo em F_{64} , equivalentemente, α é uma raiz de $x^6 + x + 1 = 0$, ou seja, $\alpha^6 + \alpha + 1 = 0$, implicando em $\alpha^6 = -\alpha - 1$. Como os coeficientes dos polinômios que formam o conjunto dos elementos de F_{64} pertencem a F_2 , fazemos a redução módulo 2 destes coeficientes e obtemos $\alpha^6 = \alpha + 1$. Os elementos de F_{64} são mostrados na Tabela 4.15.

| Elementos de F_{64} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ | Elementos de F_{64} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ |
|-------------------------------------|---|---|---|
| 0 | (000000) | $\alpha^{10} = \alpha \times \alpha^9$ | (000011) |
| 1 | (100000) | \vdots | \vdots |
| α | (010000) | $\alpha^{55} = \alpha \times \alpha^{54}$ | (011101) |
| α^2 | (001000) | $\alpha^{56} = \alpha \times \alpha^{55}$ | (111110) |
| α^3 | (000100) | $\alpha^{57} = \alpha \times \alpha^{56}$ | (011111) |
| α^4 | (000010) | $\alpha^{58} = \alpha \times \alpha^{57}$ | (111111) |
| α^5 | (000001) | $\alpha^{59} = \alpha \times \alpha^{58}$ | (101111) |
| α^6 | (110000) | $\alpha^{60} = \alpha \times \alpha^{59}$ | (100111) |
| $\alpha^7 = \alpha \times \alpha^6$ | (011000) | $\alpha^{61} = \alpha \times \alpha^{60}$ | (100011) |
| $\alpha^8 = \alpha \times \alpha^7$ | (001100) | $\alpha^{62} = \alpha \times \alpha^{61}$ | (100001) |
| $\alpha^9 = \alpha \times \alpha^8$ | (000110) | $\alpha^{63} = \alpha \times \alpha^{62}$ | (100000) |

Tabela 4.15: Elementos de F_{64} em notação de r -uplas com $p_1(x)$.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Considere o anel $GR(p^k, r) = GR(4, 6)$ como sendo dado pelo quociente do anel $\mathbb{Z}_4[x]$ (conjunto de todos os polinômios com coeficientes em \mathbb{Z}_4) pelo ideal gerado pelo mesmo $p(x)$ utilizado para realizar a extensão do corpo no **Passo 4**, isto é,

$$\frac{\mathbb{Z}_4[x]}{\langle p(x) \rangle} \cong \frac{\mathbb{Z}_4[x]}{\langle x^6 + x + 1 \rangle} = \{b_0 + b_1x + b_2x^2 + \cdots + b_5x^5 : b_i \in \mathbb{Z}_4\} \quad (4.3)$$

A seguir, determinaremos a ordem do grupo cíclico pertencente ao grupo das unidades. Sabemos que as operações em $GR^*(4, 6)$ são realizadas módulo $(x^6 + x + 1)$. Como α é uma raiz do polinômio primitivo usado tanto na extensão do corpo como na do

anel, então $\alpha^6 = -\alpha - 1$. Como os coeficientes dos polinômios em $GR(4, 6)$ estão em \mathbb{Z}_4 , então $\alpha^6 = 3\alpha + 3$. Considerando $f = (010000) = \alpha$, todos os elementos não nulos e inversíveis do grupo cíclico do grupo $GR^*(4, 6)$ são determinados através da potenciação de f , como mostra a Tabela 4.16.

| $GR^*(4, 6)$ | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ | $GR^*(4, 6)$ | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ |
|------------------------|---|------------------------------------|---|
| 1 | (100000) | $f^{10} = x^{10} = \alpha^{10}$ | (000033) |
| $f = x = \alpha$ | (010000) | $f^{11} = x^{11} = \alpha^{11}$ | (110003) |
| $f^2 = x^2 = \alpha^2$ | (001000) | \vdots | \vdots |
| $f^3 = x^3 = \alpha^3$ | (000100) | $f^{120} = x^{120} = \alpha^{120}$ | (013131) |
| $f^4 = x^4 = \alpha^4$ | (000010) | $f^{121} = x^{121} = \alpha^{121}$ | (331313) |
| $f^5 = x^5 = \alpha^5$ | (000001) | $f^{122} = x^{122} = \alpha^{122}$ | (103131) |
| $f^6 = x^6 = \alpha^6$ | (330000) | $f^{123} = x^{123} = \alpha^{123}$ | (300313) |
| $f^7 = x^7 = \alpha^7$ | (033000) | $f^{124} = x^{124} = \alpha^{124}$ | (100031) |
| $f^8 = x^8 = \alpha^8$ | (003300) | $f^{125} = x^{125} = \alpha^{125}$ | (300003) |
| $f^9 = x^9 = \alpha^9$ | (000330) | $f^{126} = x^{126} = \alpha^{126}$ | (100000) |

Tabela 4.16: Elementos do grupo cíclico do grupo $GR^*(4, 6)$ em notação de r -uplas com $p_1(x)$.

Passo 6 - Determinar o grupo das unidades - Do **Passo 5**, resulta que f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 6)$, onde $d \geq 1 \in \mathbb{Z}$, e f^d gera o subgrupo cíclico cuja ordem é 63 em $GR^*(4, 6)$ (Teorema 2.3.9). Sendo assim, temos que $n \cdot d = 63 \cdot d = 126$, implicando que $d = 2$. Consequentemente, $f^2 = (001000) = \alpha^2$ gera um subgrupo cíclico de ordem 63 em $GR^*(4, 6)$. Logo, $\beta = \alpha^2$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{63}$, mostrado na Tabela 4.17. Esse elemento primitivo será utilizado na construção de um código BCH de comprimento $n = 63$ sobre \mathbb{Z}_4 . Quando o comprimento n da palavra-código desejada for igual a cardinalidade de G_n , faremos então a construção de um código **código BCH primitivo**, onde f gera um grupo cíclico de ordem $n \cdot 2$ em $GR^*(4, r)$, ver Definição 2.3.14.

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Neste passo, vamos calcular os polinômios geradores $g(x)$ das matrizes geradoras G dos códigos.

Os polinômios geradores dos códigos de comprimento n , tem como raízes os elementos na sequência, $\{(\beta^i), (\beta^i)^p, (\beta^i)^{p^2}, (\beta^i)^{p^3}, \dots, (\beta^i)^{p^{r-1} \pmod{n}}\}$.

Estes polinômios são dados por

$$g(x) = mmc(M_1(x), M_2(x), \dots, M_{n-1}(x))$$

onde $M_i(x)$ é o polinômio minimal associado ao elemento β_i , $\{i = 1, 2, \dots, n-1\}$ (β é um elemento primitivo em G_n).

| $G_{63} \rightarrow (\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$ | $G_{63} \rightarrow (\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$ | $G_{63} \rightarrow (\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5)$ |
|---|---|---|
| $\beta \rightarrow (001000)$ | $\beta^{22} \rightarrow (303101)$ | $\beta^{43} \rightarrow (100303)$ |
| $\beta^2 \rightarrow (000010)$ | $\beta^{23} \rightarrow (032031)$ | $\beta^{44} \rightarrow (012003)$ |
| $\beta^3 \rightarrow (330000)$ | $\beta^{24} \rightarrow (103320)$ | $\beta^{45} \rightarrow (011120)$ |
| $\beta^4 \rightarrow (003300)$ | $\beta^{25} \rightarrow (221033)$ | $\beta^{46} \rightarrow (220111)$ |
| $\beta^5 \rightarrow (000033)$ | $\beta^{26} \rightarrow (123210)$ | $\beta^{47} \rightarrow (321201)$ |
| $\beta^6 \rightarrow (121000)$ | $\beta^{27} \rightarrow (331232)$ | $\beta^{48} \rightarrow (032212)$ |
| $\beta^7 \rightarrow (001210)$ | $\beta^{28} \rightarrow (131312)$ | $\beta^{49} \rightarrow (312322)$ |
| $\beta^8 \rightarrow (330012)$ | $\beta^{29} \rightarrow (313313)$ | $\beta^{50} \rightarrow (201123)$ |
| $\beta^9 \rightarrow (311300)$ | $\beta^{30} \rightarrow (300133)$ | $\beta^{51} \rightarrow (233011)$ |
| $\beta^{10} \rightarrow (003113)$ | $\beta^{31} \rightarrow (120001)$ | $\beta^{52} \rightarrow (321330)$ |
| $\beta^{11} \rightarrow (301031)$ | $\beta^{32} \rightarrow (030200)$ | $\beta^{53} \rightarrow (113213)$ |
| $\beta^{12} \rightarrow (102010)$ | $\beta^{33} \rightarrow (000302)$ | $\beta^{54} \rightarrow (302132)$ |
| $\beta^{13} \rightarrow (331020)$ | $\beta^{34} \rightarrow (022003)$ | $\beta^{55} \rightarrow (131021)$ |
| $\beta^{14} \rightarrow (223310)$ | $\beta^{35} \rightarrow (011220)$ | $\beta^{56} \rightarrow (210310)$ |
| $\beta^{15} \rightarrow (332233)$ | $\beta^{36} \rightarrow (220112)$ | $\beta^{57} \rightarrow (332103)$ |
| $\beta^{16} \rightarrow (120322)$ | $\beta^{37} \rightarrow (310201)$ | $\beta^{58} \rightarrow (010321)$ |
| $\beta^{17} \rightarrow (203203)$ | $\beta^{38} \rightarrow (032102)$ | $\beta^{59} \rightarrow (213103)$ |
| $\beta^{18} \rightarrow (013032)$ | $\beta^{39} \rightarrow (022321)$ | $\beta^{60} \rightarrow (013131)$ |
| $\beta^{19} \rightarrow (132130)$ | $\beta^{40} \rightarrow (213223)$ | $\beta^{61} \rightarrow (103131)$ |
| $\beta^{20} \rightarrow (111321)$ | $\beta^{41} \rightarrow (233132)$ | $\beta^{62} \rightarrow (100031)$ |
| $\beta^{21} \rightarrow (210113)$ | $\beta^{42} \rightarrow (130331)$ | $\beta^{63} \rightarrow (100000)$ |

 Tabela 4.17: Elementos de G_{63}

No caso da palavra-código em questão, cujo comprimento é $n = 63$, os valores de $1 \leq t \leq 31$ serão analisados. Para cada valor de t , teremos uma distância equivalente e seus respectivos polinômios minimais envolvidos nos cálculos dos $g(x)$, da seguinte maneira:

1º) Cálculo das raízes dos polinômios minimais:

Para cada polinômio minimal $M_i(x) = M_i$, com $i = 1, 2, \dots, 62$, temos:

$$\begin{aligned}
 M_1(x) &= \{(\beta^1), (\beta^1)^2, \dots, (\beta^1)^{2^{6-1} \pmod{63}}\} \rightarrow M_1 = \{\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}\}, \\
 M_2(x) &= \{(\beta^2), (\beta^2)^2, \dots, (\beta^2)^{2^{6-1} \pmod{63}}\} \rightarrow M_2 = \{\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta\}, \\
 &\vdots = \vdots \\
 M_{62}(x) &= \{(\beta^{62}), (\beta^{62})^2, \dots, (\beta^{62})^{2^{6-1} \pmod{63}}\} \rightarrow M_{62} = \{\beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}\}.
 \end{aligned}$$

2º) Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \dots, 62$:

Os polinômios minimais são calculados da seguinte maneira:

$$\begin{aligned}
 M_1(x) &= \{(x - \beta)(x - \beta^2)(x - \beta^4)(x - \beta^8)(x - \beta^{16})(x - \beta^{32})\} \\
 M_1(x) &= x^6 + 2x^3 + 3x + 1.
 \end{aligned}$$

| | | |
|---|---|---|
| $M_1 = (\beta, \beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32})$ | $M_{22} = (\beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11})$ | $M_{43} = (\beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53})$ |
| $M_2 = (\beta^2, \beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta)$ | $M_{23} = (\beta^{23}, \beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43})$ | $M_{44} = (\beta^{44}, \beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22})$ |
| $M_3 = (\beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33})$ | $M_{24} = (\beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12})$ | $M_{45} = (\beta^{45}, \beta^{27}, \beta^{54})$ |
| $M_4 = (\beta^4, \beta^8, \beta^{16}, \beta^{32}, \beta, \beta^2)$ | $M_{25} = (\beta^{25}, \beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44})$ | $M_{46} = (\beta^{46}, \beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23})$ |
| $M_5 = (\beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34})$ | $M_{26} = (\beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13})$ | $M_{47} = (\beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55})$ |
| $M_6 = (\beta^6, \beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3)$ | $M_{27} = (\beta^{27}, \beta^{54}, \beta^{45})$ | $M_{48} = (\beta^{48}, \beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24})$ |
| $M_7 = (\beta^7, \beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35})$ | $M_{28} = (\beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14})$ | $M_{49} = (\beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56})$ |
| $M_8 = (\beta^8, \beta^{16}, \beta^{32}, \beta, \beta^2, \beta^4)$ | $M_{29} = (\beta^{29}, \beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46})$ | $M_{50} = (\beta^{50}, \beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25})$ |
| $M_9 = (\beta^9, \beta^{18}, \beta^{36})$ | $M_{30} = (\beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}, \beta^{15})$ | $M_{51} = (\beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57})$ |
| $M_{10} = (\beta^{10}, \beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^5)$ | $M_{31} = (\beta^{31}, \beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47})$ | $M_{52} = (\beta^{52}, \beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26})$ |
| $M_{11} = (\beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50}, \beta^{37})$ | $M_{32} = (\beta^{32}, \beta, \beta^2, \beta^4, \beta^8, \beta^{16})$ | $M_{53} = (\beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29}, \beta^{58})$ |
| $M_{12} = (\beta^{12}, \beta^{24}, \beta^{48}, \beta^{33}, \beta^3, \beta^6)$ | $M_{33} = (\beta^{33}, \beta^3, \beta^6, \beta^{12}, \beta^{24}, \beta^{48})$ | $M_{54} = (\beta^{54}, \beta^{45}, \beta^{27})$ |
| $M_{13} = (\beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19}, \beta^{38})$ | $M_{34} = (\beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40}, \beta^{17})$ | $M_{55} = (\beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61}, \beta^{59})$ |
| $M_{14} = (\beta^{14}, \beta^{28}, \beta^{56}, \beta^{49}, \beta^{35}, \beta^7)$ | $M_{35} = (\beta^{35}, \beta^7, \beta^{14}, \beta^{28}, \beta^{56}, \beta^{49})$ | $M_{56} = (\beta^{56}, \beta^{49}, \beta^{35}, \beta^7, \beta^{14}, \beta^{28})$ |
| $M_{15} = (\beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51}, \beta^{39})$ | $M_{36} = (\beta^{36}, \beta^9, \beta^{18})$ | $M_{57} = (\beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30}, \beta^{60})$ |
| $M_{16} = (\beta^{16}, \beta^{32}, \beta, \beta^2, \beta^4, \beta^8)$ | $M_{37} = (\beta^{37}, \beta^{11}, \beta^{22}, \beta^{44}, \beta^{25}, \beta^{50})$ | $M_{58} = (\beta^{58}, \beta^{53}, \beta^{43}, \beta^{23}, \beta^{46}, \beta^{29})$ |
| $M_{17} = (\beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20}, \beta^{40})$ | $M_{38} = (\beta^{38}, \beta^{13}, \beta^{26}, \beta^{52}, \beta^{41}, \beta^{19})$ | $M_{59} = (\beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62}, \beta^{61})$ |
| $M_{18} = (\beta^{18}, \beta^{36}, \beta^9)$ | $M_{39} = (\beta^{39}, \beta^{15}, \beta^{30}, \beta^{60}, \beta^{57}, \beta^{51})$ | $M_{60} = (\beta^{60}, \beta^{57}, \beta^{51}, \beta^{39}, \beta^{15}, \beta^{30})$ |
| $M_{19} = (\beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{16}, \beta^{41})$ | $M_{40} = (\beta^{40}, \beta^{17}, \beta^{34}, \beta^5, \beta^{10}, \beta^{20})$ | $M_{61} = (\beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31}, \beta^{62})$ |
| $M_{20} = (\beta^{20}, \beta^{40}, \beta^{17}, \beta^{34}, \beta^{16}, \beta^{10})$ | $M_{41} = (\beta^{41}, \beta^{19}, \beta^{38}, \beta^{13}, \beta^{26}, \beta^{52})$ | $M_{62} = (\beta^{62}, \beta^{61}, \beta^{59}, \beta^{55}, \beta^{47}, \beta^{31})$ |
| $M_{21} = (\beta^{21}, \beta^{42})$ | $M_{42} = (\beta^{42}, \beta^{21})$ | |

 Tabela 4.18: Raízes dos polinômios minimais sobre G_{63} .

De maneira análoga, os demais polinômios minimais são determinados. Lembrando que as operações módulo 4 devem ser obedecidas nos cálculos dos polinômios minimais. Na Tabela 4.18, observe que alguns polinômios minimais possuem as mesmas raízes. Portanto, estes polinômios minimais são iguais, como mostrados na Tabela 4.19.

| | |
|---|---|
| $M_1 = M_2 = M_4 = M_8 = M_{16} = M_{32}$ | $M_{13} = M_{26} = M_{52} = M_{41} = M_{19} = M_{38}$ |
| $M_3 = M_6 = M_{12} = M_{24} = M_{48} = M_{33}$ | $M_{15} = M_{30} = M_{60} = M_{57} = M_{51} = M_{39}$ |
| $M_5 = M_{10} = M_{20} = M_{40} = M_{17} = M_{34}$ | $M_{21} = M_{42}$ |
| $M_7 = M_{14} = M_{28} = M_{56} = M_{49} = M_{35}$ | $M_{23} = M_{46} = M_{29} = M_{58} = M_{53} = M_{43}$ |
| $M_9 = M_{18} = M_{36}$ | $M_{27} = M_{54} = M_{45}$ |
| $M_{11} = M_{22} = M_{44} = M_{25} = M_{50} = M_{37}$ | $M_{31} = M_{62} = M_{61} = M_{59} = M_{55} = M_{47}$ |

 Tabela 4.19: Polinômios minimais que são iguais sobre G_{63} .

3º) Cálculo dos polinômios geradores para $1 \leq t \leq 31$:

O polinômio gerador para cada valor de t é dado por $g(x) = \text{mmc}\{M_1(x), M_2(x), \dots, M_{n-1}(x)\}$, formado pelos polinômios minimais que são diferentes entre si e possuem raízes β, \dots, β^{2t} , como mostra a Tabela A.3 (Apêndice A).

Considerando que a distância mínima do código seja $d_H = 3$, então o polinômio gerador do código é dado por $g_1(x) = x^6 + 2x^3 + 3x + 1$, que gera o código desejado e está relacionado com a matriz geradora G do código BCH sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (63, 57, 3)$. De maneira análoga, os demais polinômios geradores para outros valores de t correspondentes a outras distâncias são determinados de acordo com a Tabela A.3 (Apêndice A).

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H é obtido através da relação:

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{63} - 1}{x^6 + 2x^3 + 3x + 1}$$

$$h(x) = x^{57} + 2x^{54} + x^{52} + 3x^{51} + x^{47} + 2x^{46} + x^{45} + 2x^{44} + 3x^{42} + x^{41} + 3x^{40} + 3x^{39} + x^{37} + 2x^{35} + 2x^{34} + x^{33} + x^{32} + 3x^{31} + 2x^{29} + x^{28} + 2x^{26} + 3x^{25} + 2x^{24} + 3x^{23} + x^{22} + 2x^{21} + 3x^{20} + x^{19} + x^{18} + 3x^{16} + 3x^{15} + 2x^{14} + 2x^{13} + x^{12} + 3x^{11} + x^9 + 2x^8 + 3x^7 + x^5 + 3x^4 + x^3 + 3x^2 + 3x + 3$$

onde os coeficientes do polinômio $h(x)$ pertencem a \mathbb{Z}_4 .

Passo 9 - Determinar a matriz G e a sua transposta G^T - O polinômio gerador $g_1(x) = x^6 + 2x^3 + 3x + 1$ está relacionado à matriz geradora G . Realizando os deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita, obtemos a matriz G com dimensão 57×63 :

$$G = \begin{pmatrix} 1302001000 \\ 0130200100 \\ 001302001000 \\ 0001302001000 \\ \vdots & \vdots & \vdots \\ 00 \\ 00 \\ 00 \\ 00 \end{pmatrix}.$$

A matriz G^T com dimensão 63×57 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - Determinado o polinômio $h(x)$ no **Passo 8**, neste passo, realizamos os deslocamentos dos coeficientes do polinômio gerador $h(x)$ da direita para a esquerda e obtemos a matriz H com dimensão 6×63 :

$$H = \begin{pmatrix} 000001002013000121203133010221130210232312311033221301230131333 \\ 000010020130001212031330102211302102323123110332213012301313330 \\ 000100201300012120313301022113021023231231103322130123013133300 \\ 001002013000121203133010221130210232312311033221301230131333000 \\ 010020130001212031330102211302102323123110332213012301313330000 \\ 100201300012120313301022113021023231231103322130123013133300000 \end{pmatrix}.$$

A matriz H^T com dimensão 63×6 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, analisaremos se o código BCH primitivo sobre $GR(4, 6)$ é capaz de reproduzir a sequência

de sinal interno (SI = Seq.36) de uma proteína mitocondrial - GI número 832917 com comprimento $n = 63$ nucleotídeos.

Este passo determina as 24 permutações entre o alfabeto do código genético $N = \{A, C, G, T/U\}$ e o alfabeto do código BCH $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ da sequência de DNA a ser analisada, no exemplo em questão a SI. Uma vez que o mapeamento entre $N \rightarrow \mathbb{Z}_4$ não é conhecido, consideramos todas as permutações entre esses dois conjuntos. As 24 linhas da matriz P correspondem às 24 permutações da SI. Seja a SI do NCBI igual a

{5'-GCCGTTCATGTTACTCTGGGTTGCCCTGGTGGGAACATATCGCGGCCACCACCATCCTCATT-3'}, então

$$P = \left(\begin{array}{c} 21123310323330131322233211332232220013031212211011011031131033 \\ 3113221023220121233322311223323330012021313311011011021121022 \\ 1221332031333023231113122331131110023032121122022022032232033 \\ 3223112013111021213331322113313333002101232322022022012212011 \\ 23321130121110313122211233112212220031013232233033033013313011 \\ 13312230212220323211122133221121110032023131133033033023323022 \\ 20023301323331030322233200332232221103130202200100100130030133 \\ 3003220123222102023332300223323331102120303300100100120020122 \\ 02203321303331232300033022330030001123132020022122122132232133 \\ 32230021030001202033300322003303331120102323322122122102202100 \\ 0330223120222132320002203322002000113212303003313313123323122 \\ 233200310200013030222002330022022221130103232233133133103303100 \\ 10013302313332030311133100331131112203230101100200200230030233 \\ 300311021311120101333113001133133332201210303300200200210010211 \\ 01103312303332131300033011330030002213231010011211211231131233 \\ 311300120300021010333003110033033332210201313311211211201101200 \\ 03301132101112313100011033110010002213213030033233233213313211 \\ 133100320100023030111001330011011112230203131133233233203303200 \\ 10012203212223020211122100221121113302320101100300300320020322 \\ 2002110312111301012221120011221222330131020220030030010010311 \\ 01102213202223121200022011220020000331232101001131131321121322 \\ 21120013020003101022200211002202223310301212211311311301101300 \\ 02201123101113212100011022110010003321312020022322322312212311 \\ 12210023010003202011100122001101113320302121122322322302202300 \end{array} \right).$$

As linhas da matriz P estão relacionadas com as 24 permutações entre $N \rightarrow \mathbb{Z}_4$, cada uma das 24 permutações foi definida como um caso, como mostrado na Tabela 4.20.

| Linha = Caso | $N \rightarrow \mathbb{Z}_4$ | Linha = Caso | $N \rightarrow \mathbb{Z}_4$ | Linha = Caso | $N \rightarrow \mathbb{Z}_4$ |
|---------------|------------------------------|----------------|------------------------------|----------------|------------------------------|
| L 1 = Caso 01 | (A,C,G,T)=(0,1,2,3) | L 9 = Caso 09 | (A,C,G,T)=(1,2,0,3) | L 17 = Caso 17 | (A,C,G,T)=(2,3,0,1) |
| L 2 = Caso 02 | (A,C,G,T)=(0,1,3,2) | L 10 = Caso 10 | (A,C,G,T)=(1,2,3,0) | L 18 = Caso 18 | (A,C,G,T)=(2,3,1,0) |
| L 3 = Caso 03 | (A,C,G,T)=(0,2,1,3) | L 11 = Caso 11 | (A,C,G,T)=(1,3,0,2) | L 19 = Caso 19 | (A,C,G,T)=(3,0,1,2) |
| L 4 = Caso 04 | (A,C,G,T)=(0,2,3,1) | L 12 = Caso 12 | (A,C,G,T)=(1,3,2,0) | L 20 = Caso 20 | (A,C,G,T)=(3,0,2,1) |
| L 5 = Caso 05 | (A,C,G,T)=(0,3,2,1) | L 13 = Caso 13 | (A,C,G,T)=(2,0,1,3) | L 21 = Caso 21 | (A,C,G,T)=(3,1,0,2) |
| L 6 = Caso 06 | (A,C,G,T)=(0,3,1,2) | L 14 = Caso 14 | (A,C,G,T)=(2,0,3,1) | L 22 = Caso 22 | (A,C,G,T)=(3,1,2,0) |
| L 7 = Caso 07 | (A,C,G,T)=(1,0,2,3) | L 15 = Caso 15 | (A,C,G,T)=(2,1,0,3) | L 23 = Caso 23 | (A,C,G,T)=(3,2,0,1) |
| L 8 = Caso 08 | (A,C,G,T)=(1,0,3,2) | L 16 = Caso 16 | (A,C,G,T)=(2,1,3,0) | L 24 = Caso 24 | (A,C,G,T)=(3,2,1,0) |

Tabela 4.20: Relação entre as linhas da matriz P e as 24 permutações.

Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(\mathbf{a}, \mathbf{b}) = 0$, $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ - O procedimento usado para determinar quais das sequências (com 0, 1 e 2 nucleotídeos de diferença da sequência original) são palavras-código dos códigos $(63, k, d_H)$, é o mesmo adotado na construção dos códigos sobre corpos, da seguinte maneira:

a) $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Para analisarmos as sequências de DNA com até 1 nucleotídeo de diferença da sequência de DNA do *NCBI*, $D(\mathbf{a}, \mathbf{b}) = 1$, consideramos 4356 possíveis palavras-código para cada sequência de DNA analisada e, então usamos a relação $v \cdot H^T = 0$. As palavras-código encontradas são armazenadas.

Como resultado da geração de $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença, obtemos a matriz R onde cada linha é uma palavra-código encontrada.

$$R = \begin{pmatrix} 122133203133302323111331223111311110023032121122022022032232033 \\ 322311201311102121333113221333133330021012323322022022012212011 \\ 033022312022213232000220332000200001132123030033133133123323122 \\ 233200310200013030222002330222022221130103232233133133103303100 \\ 10013302313332030311133100311131112203230101100200200230030233 \\ 300311021311120101333113001333133332201210303300200200210010211 \\ 01102213202223121200022011200020003312321010011311311321121322 \\ 211200130200031010222002110222022223310301212211311311301101300 \end{pmatrix}.$$

b) $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Para analisarmos as sequências de DNA com até 2 nucleotídeos de diferença da sequência de DNA do *NCBI*, $D(\mathbf{a}, \mathbf{b}) = 2$, consideramos 17577 possíveis palavras-código para cada sequência de DNA analisada e, então, usamos a relação $\mathbf{v} \cdot H^T = 0$. As palavras-código encontradas são armazenadas.

Como resultado da geração de $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença, obtemos várias matrizes diferentes entre si, por exemplo, R' , onde cada linha é uma palavra-código encontrada.

$$R' = \left(\begin{array}{c} 21123310323331133322233211332232220013031212211011011031131033 \\ 31232210232220121233322311233323330012021313311011011021121022 \\ 32213320313330232331133122331131110023032121122022022032232033 \\ 122311201311102121133113221133133330021012323322022022012212011 \\ 23321130121113311122211233112212220031013232233033033013313011 \\ 1321223021222032321122133211121110032023131133033033023323022 \\ 20323301323331030322233200322232221103130202200100100130030133 \\ 3003220123222002223322300223323331102120303300100100120020122 \\ 023033213033312323000330223000300001123132020022122122132232133 \\ 322300210300022000333003220033033331120102323322122122102202100 \\ 233022312022213232200220332200200001132123030033133133123323122 \\ 033200310200013030022002330022022221130103232233133133103303100 \\ 30013302313332030331133100331131112203230101100200200230030233 \\ 10031102131112010113311300113313333220121030330020020010211 \\ 0110331230333113330033011330030002213231010011211211231131233 \\ 310300120300021010330031103303332210201313311211211201101200 \\ 033011321011133111000110331100100022312130300332332331331211 \\ 130100320100023030111001330111011112230203131133233233203303200 \\ 1001220321222002221122100221121113302320101100300300320020322 \\ 2012110312111301012221120012221222330131020220030030010010311 \\ 2110221320222312122002201122002000331232101001131131321121322 \\ 0112001302000310100220021100220222233103012122113113101101300 \\ 02101123101113212100011022100010003321312020022322322312212311 \\ 122100230100022000111001220011011113320302121122322322302202300 \end{array} \right)$$

Passo 13 - Comparar todas as palavras-código armazenadas no Passo 12 com a sequência de DNA original e mostrar onde os erros ocorreram - Neste passo, todas as palavras-código armazenadas no passo anterior estão rotuladas na forma do alfabeto do código, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, e serão convertidas em nucleotídeos usando o rotulamento do alfabeto do código genético $N = \{A, C, G, T\}$. Após o rotulamento, as palavras-código são comparadas, uma-a-uma, com a sequência de DNA original mostrando onde os nucleotídeos diferem, e armazena os resultados.

Passo 14 - Voltar para o Passo 7 e determinar outro $g(x)$ - Neste passo, determinamos outro valor da distância mínima d_H , por exemplo $d_H = 5$ (Tabela A.3), e utilizamos o mesmo procedimento, apresentado no **Passo 7**, para calcular o polinômio gerador relativo a esta distância.

Passo 15 - Repetir os Passos 8 ao Passo 12 para o $g(x)$ obtido no Passo 14, até que se esgotem todas as possibilidades de $g(x)$ - Neste passo, o algoritmo determina todas as palavras-código encontradas com nenhum, 1 e 2 nucleotídeos de diferença através de todos os polinômios geradores relativos à distância mínima $1 \leq d_H \leq n$, neste exemplo $1 \leq d_H \leq 63$, e armazena os resultados.

Passo 16 - Voltar para o Passo 3 e escolher outro $p(x)$, e então, repetir os Passos 4 ao 14 até esgotar todos os $p(x)$ do Passo 3;

Passo 17 - Fim.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

| Sequência de DNA do NCBI Seq.36 <i>S.cerevisiae</i> - OXA 1 gene - GI número 832917 SI={5'- GCCGTTCATGTTACTCTGGGTTGCCTTGGTGGGAACATCGCGGCCACCACCATCCTCATT -3'} | | | |
|---|-----------------------------|-------------------------------|--|
| 24 casos de permutações | Palavras-código $n = 63$ | Quantidade de palavras-código | Sequência de DNA reproduzida |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 00000 ... 00000 | 0 | - |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 00000 ... 00000 | 0 | - |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 12213 ... 32033 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 32231 ... 12011 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 00000 ... 00000 | 0 | - |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 00000 ... 00000 | 0 | - |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 00000 ... 00000 | 0 | - |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 00000 ... 00000 | 0 | - |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 00000 ... 00000 | 0 | - |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 00000 ... 00000 | 0 | - |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 03302 ... 23122 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 23320 ... 03100 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 10013 ... 30233 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 30031 ... 10211 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 00000 ... 00000 | 0 | - |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 00000 ... 00000 | 0 | - |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 00000 ... 00000 | 0 | - |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 00000 ... 00000 | 0 | - |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 00000 ... 00000 | 0 | - |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 00000 ... 00000 | 0 | - |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 01102 ... 21322 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 21120 ... 01300 | 1 | GCCGTTCATGTTACTCTGGGTTGCCTGGG ... |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 00000 ... 00000 | 0 | - |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 00000 ... 00000 | 0 | - |

Tabela 4.21: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 63$ nucleotídeos.

Como resultado do algoritmo de geração da sequência de DNA para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença da sequência do NCBI, obtemos 8 palavras-código, sendo 1 palavra-código para cada caso de permutação que corresponde aos 8 casos de permutações, representados através da matriz R do **Passo 12**. Essas palavras-código são diferentes em termos do alfabeto do código, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, porém são iguais quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$, resultando em uma única sequência DNA, Tabela 4.21. Como as sequências que foram reproduzidas nos casos 03, 04, 11, 12, 13, 14, 21 e 22 são iguais, é suficiente considerarmos a sequência correspondente ao caso 03 mostrada na Figura 4.2. Ressaltamos que a diferença do dígito da palavra-código correspondente ao nucleotídeo da sequência de DNA sempre ocorre na mesma posição nessas 8 palavras-código.

Em [20], as sequências de direcionamento com $n = 63$ nucleotídeos reproduzidas pelos códigos G -linearidade já mostraram que, para cada 8 palavras-código correspondentes aos 8

Seq.36 | *S. cerevisiae* - OXA 1 – Sinal interno – GI número 832917

Código klein-linearidade((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento C)

$$p_1(x) = x^6 + x + 1 \quad g_1(x) = x^6 + 2x^3 + 3x + 1 \quad \text{Caso 3-}(\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |
| ntO: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| RtO: | 122 | 133 | 203 | 133 | 302 | 323 | 111 | 331 | 223 | 311 | 311 | 110 | 023 | 032 | 121 | 122 | 022 | 022 | 032 | 232 | 033 |
| RtG: | 122 | 133 | 203 | 133 | 302 | 323 | 111 | 331 | 223 | 111 | 311 | 110 | 023 | 032 | 121 | 122 | 022 | 022 | 032 | 232 | 033 |
| ntG: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | GGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| aaG: | A | V | H | V | Y | S | G | L | P | G | W | G | T | I | A | A | T | T | I | L | I |

Figura 4.2: Sequência de DNA de sinal interno com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

casos de permutações, obtinha-se uma única sequência de direcionamento. Assim, os casos de permutações mencionadas na Tabela 4.20 foram definidos como: **rotulamento A** - os casos 02, 06, 07, 09, 16, 18, 20 e 23; **rotulamento B** - os casos 01, 05, 08, 10, 15, 17, 19 e 24 e **rotulamento C** - os casos 03, 04, 11, 12, 13, 14, 21 e 22, (Figura 3.17).

A Figura 4.2 mostra a sequência de SI (Seq.36) reproduzida pelo código G -linearidade: Klein-linearidade ((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento C) através do polinômio primitivo $p_1(x) = x^6 + x + 1$ e do polinômio gerador $g_1(x) = x^6 + 2x^3 + 3x + 1$. Observe que, na posição da décima trinca, ocorreu uma troca de nucleotídeo ocasionando a troca do aminoácido nesta posição, sendo o Triptofano (W) substituído pela Glicina (G), indicado por (W10G). Essa alteração de aminoácidos implica na mudança de classe, da classe de aminoácido não-polar para classe de aminoácido polar (Tabela 3.1). Dessa forma, observamos que os resultados dos rotulamentos apresentados em [20] para sequências de direcionamento com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença da sequência do NCBI foram confirmados para outras sequências de DNA.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Como resultado do algoritmo para $D(\mathbf{a}, \mathbf{b}) = 2$, obtemos várias palavras-código para cada uma das 24 permutações, como mostra a Tabela 4.22. Neste exemplo, obtemos 5 palavras-código, distintas entre si, em cada caso de permutação dos rotulamentos A e B, respectivamente, e 93 palavras-código em cada caso do rotulamento C, enquanto para $D(\mathbf{a}, \mathbf{b}) = 1$, obtemos apenas uma palavra-código para cada caso de permutação.

A matriz R' do **Passo 12** representa uma possibilidade dentre as 24 palavras-código, correspondendo aos 24 casos de permutações nos rotulamentos A, B e C. As palavras-código em termos do alfabeto do código $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ diferem nos 24 casos de permutações, porém quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$, são iguais nos 8 casos dos rotulamentos A, B e C, respectivamente.

| Seq.36 <i>S. cerevisiae</i> - OXA 1 - Sinal interno - GI número 832917 | |
|--|-------------------------------|
| 24 Permutações | Quantidade de palavras-código |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 5 |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 5 |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 93 |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 93 |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 5 |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 5 |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 5 |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 5 |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 5 |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 5 |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 93 |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 93 |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 93 |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 93 |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 5 |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 5 |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 5 |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 5 |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 5 |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 5 |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 93 |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 93 |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 5 |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 5 |

 Tabela 4.22: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 63$ nucleotídeos.

Seq.36 | *S. cerevisiae* - OXA 1 – Sinal interno – GI número 832917

| Código Z_4 -linearidade((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento A) | |
|--|---|
| $p_1(x)=x^6+x+1 - g_1(x)=x^6+2x^3+3x+1 - \text{Caso 2-}(A,C,G,T)=(0,1,3,2)$ | |
| aaO: | A V H V Y S G L P W W G T I A A T T T I L I |
| nto: | GCC GTT CAT GTT TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT |
| Rto: | 211 233 103 233 301 313 222 332 113 322 322 220 013 031 212 211 011 031 131 033 |
| Rtg: | 211 233 103 233 311 333 222 332 113 322 322 220 013 031 212 211 011 031 131 033 |
| ntg: | GCC GTT CAT GTT TCC TTT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT |
| aag: | A V H V S F G L P W W G T I A A T T I L I |

Código $Z_2 \times Z_2$ -linearidade((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento B)

$$p_1(x)=x^6+x+1 - g_1(x)=x^6+2x^3+3x+1 - \text{Caso 1-}(A,C,G,T)=(0,1,2,3)$$

| | |
|------|---|
| aaO: | A V H V Y S G L P W W G T I A A T T T I L I |
| nto: | GCC GTT CAT GTT TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT |
| Rto: | 311 322 102 322 201 212 333 223 112 233 233 330 012 021 313 311 011 021 121 022 |
| Rtg: | 311 321 102 322 201 212 333 223 112 233 233 330 012 021 313 311 011 021 121 022 |
| ntg: | GCC GTC CAT GTT TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACA ATC CTC ATT |
| aag: | A V H V Y S G L P W W G T I A A T T I L I |

Código Klein-linearidade((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento C)

$$p_1(x)=x^6+x+1 - g_1(x)=x^6+2x^3+3x+1 - \text{Caso 3-}(A,C,G,T)=(0,2,1,3)$$

| | |
|------|---|
| aaO: | A V H V Y S G L P W W G T I A A T T T I L I |
| nto: | GCC GTT CAT GTT TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCG GCC ACC ACC ATC CTC ATT |
| Rto: | 122 133 203 133 302 323 111 331 223 311 311 110 023 032 121 122 022 022 032 232 033 |
| Rtg: | 122 133 203 131 302 323 111 331 223 311 311 110 023 032 123 122 022 022 032 232 033 |
| ntg: | GCC GTT CAT GTG TAC TCT GGG TTG CCT TGG TGG GGA ACT ATC GCT GCC ACC ACC ATC CTC ATT |
| aag: | A V H V Y S G L P W W G T I A A T T I L I |

 Figura 4.3: Sequência de DNA de sinal interno com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$.

A Figura 4.3 mostra a SI (Seq.36) reproduzida pelos códigos G -linearidade: \mathbb{Z}_4 -linearidade ((63,57,3) BCH primitivo sobre $GR(4, 6)$, rotulamento A); $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade ((63,57,3) BCH primitivo sobre $GR(4, 6)$, rotulamento B) e Klein-linearidade ((63,57,3) BCH primitivo sobre $GR(4, 6)$, rotulamento C), através do polinômio primitivo $p_1(x) = x^6 + x + 1$ e do polinômio gerador $g_1(x) = x^6 + 2x^3 + 3x + 1$ com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos diferindo da sequência do NCBI. Observe que, os nucleotídeos das posições das trincas 5 e 6 da sequência reproduzida no rotulamento A foram alterados, ocasionando a troca dos aminoácidos nestas posições, (Y5S) e (S6F). Por outro lado, as alterações de nucleotídeos nas posições 2 e 18 da sequência reproduzida no rotulamento B, assim como, as alterações de nucleotídeos nas posições 4 e 15 da sequência reproduzida no rotulamento C, não ocasionaram as trocas dos aminoácidos nessas trincas, respectivamente.

Código BCH $(n, k, d_H) = (255, k, d_H)$ sobre $GR(4, 8)$

Considere agora a construção do código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (255, k, d_H)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = 255$ nucleotídeos. Lembramos que todos os passos do algoritmo descrito no exemplo anterior devem ser seguidos, mas vamos apenas descrever os passos cujos cálculos são necessários para a construção deste código: os demais passos são análogos aos passos descrito no exemplo anterior.

Passo 2 - Determinar a extensão de Galois - Como $n = 255$, logo o grau r do polinômio primitivo a ser usado na extensão de Galois do corpo $GF(2)$ é 8, pois $n = (2^r - 1) = (2^8 - 1) = 255$.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Na Tabela 4.23, mostramos todos os polinômios primitivos relacionados à extensão de Galois de grau $r = 8$ que serão utilizados no **Passo 4**.

| Polinômios Primitivos $p(x)$ | |
|---|---|
| $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ | $p_{09}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ |
| $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ | $p_{10}(x) = x^8 + x^7 + x^6 + x + 1$ |
| $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ | $p_{11}(x) = x^8 + x^7 + x^5 + x^3 + 1$ |
| $p_{04}(x) = x^8 + x^5 + x^3 + x + 1$ | $p_{12}(x) = x^8 + x^7 + x^2 + x + 1$ |
| $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ | $p_{13}(x) = x^8 + x^6 + x^3 + x^2 + 1$ |
| $p_{06}(x) = x^8 + x^6 + x^5 + x + 1$ | $p_{14}(x) = x^8 + x^7 + x^6 + x^3 + x^2 + x + 1$ |
| $p_{07}(x) = x^8 + x^7 + x^3 + x^2 + 1$ | $p_{15}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + 1$ |
| $p_{08}(x) = x^8 + x^5 + x^3 + x^2 + 1$ | $p_{16}(x) = x^8 + x^6 + x^5 + x^4 + 1$ |

Tabela 4.23: Polinômios primitivos da extensão de Galois $r = 8$.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - Considere o corpo de Galois $GF(2^r) = GF(2^8) = GF(256) = F_{256}$, dado por

$$\frac{F_2[x]}{\langle p_{01}(x) \rangle} \cong \frac{F_2[x]}{\langle x^8 + x^4 + x^3 + x^2 + 1 \rangle} = \{a_0 + a_1x + a_2x^2 + \cdots + a_7x^7 : a_i's \in F_2\}.$$

Analogamente à extensão de Galois de grau 6, a Tabela 4.24 mostra alguns elementos de F_{256} usando o polinômio primitivo $p_{01}(x)$.

| Elementos de F_{256} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|---|
| 0 | (00000000) |
| 1 | (10000000) |
| α | (01000000) |
| α^2 | (00100000) |
| α^3 | (00010000) |
| α^4 | (00001000) |
| α^5 | (00000100) |
| α^6 | (00000010) |
| α^7 | (00000001) |
| α^8 | (10111000) |
| $\alpha^9 = \alpha \times \alpha^8$ | (01011100) |
| $\alpha^{10} = \alpha \times \alpha^9$ | (00101110) |
| \vdots | \vdots |
| $\alpha^{253} = \alpha \times \alpha^{252}$ | (11100010) |
| $\alpha^{254} = \alpha \times \alpha^{253}$ | (01110001) |
| $\alpha^{255} = \alpha \times \alpha^{254}$ | (10000000) |

Tabela 4.24: Elementos de F_{256} com $p_{01}(x)$.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Considere o anel $GR(4, 8)$ dado pela extensão de Galois de grau $r = 8$, isto é,

$$GR(4, 8)[x] \cong \frac{\mathbb{Z}_4[x]}{\langle p_{01}(x) \rangle} \cong \frac{\mathbb{Z}_4[x]}{\langle x^8 + x^4 + x^3 + x^2 + 1 \rangle} = \{b_0 + b_1x + b_2x^2 + \cdots + b_7x^7 : b'_is \in \mathbb{Z}_4\}.$$

As operações em $GR^*(4, 8)$ são realizadas módulo $x^8 + x^4 + x^3 + x^2 + 1$. Logo, $\alpha^8 = -\alpha^4 - \alpha^3 - \alpha^2 - 1$; porém, os coeficientes de $GR^*(4, 8)$ estão em \mathbb{Z}_4 , temos que $\alpha^8 = 3\alpha^4 + 3\alpha^3 + 3\alpha^2 + 3$. Assim, considerando $f = (01000000) = \alpha$, a representação dos elementos do grupo cíclico do grupo $GR^*(4, 8)$ é mostrada na Tabela 4.25.

Passo 6 - Determinar o grupo das unidades - Do **Passo 5**, resulta que f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 8)$, onde $d \geq 1 \in \mathbb{Z}$, e f^d gera o subgrupo cíclico cuja ordem é 255 em $GR^*(4, 8)$. Portanto $n \cdot d = 255 \cdot d = 510$, implicando que $d = 2$ e, consequentemente, $f^2 = (00100000) = \alpha^2$ gera um subgrupo cíclico de ordem 255 em $GR^*(4, 8)$. Logo, $\beta = \alpha^2$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{255}$,

| $GR^*(4, 8)$ | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|---|
| 1 | (10000000) |
| $f = x = \alpha$ | (01000000) |
| $f^2 = x^2 = \alpha^2$ | (00100000) |
| $f^3 = x^3 = \alpha^3$ | (00010000) |
| $f^4 = x^4 = \alpha^4$ | (00001000) |
| $f^5 = x^5 = \alpha^5$ | (00000100) |
| $f^6 = x^6 = \alpha^6$ | (00000010) |
| $f' = x' = \alpha'$ | (00000001) |
| $f^8 = x^8 = \alpha^8$ | (30333000) |
| $f^9 = f \times f^8 = x^9 = \alpha^9$ | (03033300) |
| $f^{10} = f \times f^9 = x^{10} = \alpha^{10}$ | (00303330) |
| : | : |
| $f^{508} = f \times f^{507} = x^{508} = \alpha^{508}$ | (33300030) |
| $f^{509} = f \times f^{508} = x^{509} = \alpha^{509}$ | (03330003) |
| $f^{510} = f \times f^{509} = x^{510} = \alpha^{510}$ | (10000000) |

 Tabela 4.25: Elementos do grupo cíclico do grupo $GR^*(4, 8)$ com $p_{01}(x)$.

mostrado na Tabela 4.26. Esse elemento primitivo será utilizado na construção de um código BCH primitivo de comprimento $n = 255$ sobre \mathbb{Z}_4 .

| $G_{255} \rightarrow (\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|
| $(f^2)^1 = f^2 = \alpha^2 = \beta \rightarrow (00100000)$ |
| $(f^2)^2 = f^4 = \alpha^4 = \beta^2 \rightarrow (00001000)$ |
| $(f^2)^3 = f^6 = \alpha^6 = \beta^3 \rightarrow (00000010)$ |
| $(f^2)^4 = f^8 = \alpha^8 = \beta^4 \rightarrow (30333000)$ |
| $(f^2)^5 = f^{10} = \alpha^{10} = \beta^5 \rightarrow (00303330)$ |
| : |
| $(f^2)^{253} = f^{506} = \alpha^{506} = \beta^{253} \rightarrow (02213011)$ |
| $(f^2)^{254} = f^{508} = \alpha^{508} = \beta^{254} \rightarrow (33300030)$ |
| $(f^2)^{255} = f^{510} = \alpha^{510} = \beta^{255} \rightarrow (10000000)$ |

 Tabela 4.26: Elementos de G_{255}

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Neste passo, o algoritmo irá analisar todos os valores possíveis de d_H que estão relacionados com a capacidade de correção de erros estabelecida através da relação $d_H \leq 2t + 1$ (lembrando que $d_H \leq n$). Neste exemplo, cujo comprimento da palavra-código é $n = 255$, os valores de $1 \leq t \leq 127$ serão analisados (Tabela A.4 do Apêndice A). Para cada valor de t , teremos uma distância equivalente e seus respectivos polinômios minimais envolvidos nos cálculos dos polinômios geradores $g(x)$, da seguinte maneira:

1º) Cálculo das raízes dos polinômios minimais:

Para cada polinômio minimal $M_i(x) = M_i$, com $i = 1, 2, \dots, 254$, temos:

$$\begin{aligned}
 M_1(x) &= \{(\beta), (\beta)^2, (\beta)^4, (\beta)^8, (\beta)^{16}, (\beta)^{32}, (\beta)^{64}, (\beta)^{128}\}, \\
 M_2(x) &= \{(\beta)^2, (\beta)^4, (\beta)^8, (\beta)^{16}, (\beta)^{32}, (\beta)^{64}, (\beta)^{128}, (\beta)\}, \\
 M_3(x) &= \{(\beta^3), (\beta)^6, (\beta)^{12}, (\beta)^{24}, (\beta)^{48}, (\beta)^{96}, (\beta)^{192}, (\beta)^{129}\}, \\
 M_4(x) &= \{(\beta)^4, (\beta)^8, (\beta)^{16}, (\beta)^{32}, (\beta)^{64}, (\beta)^{128}, (\beta), (\beta)^2\}, \\
 &\vdots = \vdots \\
 M_{252}(x) &= \{(\beta)^{252}, (\beta)^{249}, (\beta)^{243}, (\beta)^{231}, (\beta)^{207}, (\beta)^{159}, (\beta)^{63}, (\beta)^{126}\}, \\
 M_{253}(x) &= \{(\beta)^{253}, (\beta)^{251}, (\beta)^{247}, (\beta)^{239}, (\beta)^{223}, (\beta)^{191}, (\beta)^{127}, (\beta)^{254}\}, \\
 M_{254}(x) &= \{(\beta)^{254}, (\beta)^{253}, (\beta)^{251}, (\beta)^{247}, (\beta)^{239}, (\beta)^{223}, (\beta)^{191}, (\beta)^{127}\}.
 \end{aligned}$$

Das raízes dos polinômios minimais $M_i(x) = M_i$, com $i = 1, 2, \dots, 254$ temos que:

$$\begin{aligned}
 M_1 &= M_2 = M_4 = M_8 = M_{16} = M_{32} = M_{64} = M_{128}, \\
 M_3 &= M_6 = M_{12} = M_{24} = M_{48} = M_{96} = M_{129} = M_{192}, \\
 &\vdots \\
 M_{254} &= M_{253} = M_{251} = M_{247} = M_{239} = M_{223} = M_{191} = M_{127}.
 \end{aligned}$$

2º) Cálculo dos polinômios minimais $M_i(x)$, para todo $i = 1, 2, \dots, 254$ é dado por:

$$\begin{aligned}
 M_1(x) &= \{(x - \beta)(x - \beta^2)(x - \beta^4)((x - \beta^8)(x - \beta^{16})(x - \beta^{32})(x - \beta^{64})(x - \beta^{128})\} \\
 M_1(x) &= x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1.
 \end{aligned}$$

De maneira análoga, os demais polinômios minimais são determinados.

3º) Cálculo dos polinômios geradores para $1 \leq t \leq 127$:

O polinômio gerador para cada valor de t é dado por $g(x) = mmc\{M_1(x), M_2(x), \dots, M_{n-1}(x)\}$, formado pelos polinômios minimais que são diferentes entre si e possuem raízes β, \dots, β^{2t} (ver os detalhes na Tabela A.4 do Apêndice A).

Considerando que a distância mínima do código seja $d_H = 3$, então o polinômio gerador $g(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1$ gera o código desejado e está relacionado com a matriz geradora G do código BCH sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (255, 247, 3)$.

De maneira análoga, os demais polinômios geradores para outros valores de t correspondentes às outras distâncias são determinados de acordo com a Tabela A.4 (Apêndice A).

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - Seja $h(x)$ o polinômio gerador da matriz verificação de paridade H , isto é,

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{255} - 1}{x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1}$$

$$h(x) = x^{247} + 2x^{245} + 2x^{244} + \cdots + 2x^6 + 2x^5 + 2x^4 + x^3 + 3x^2 + 2x + 3,$$

onde os coeficientes do polinômio $h(x)$ pertencem a \mathbb{Z}_4 , podendo ser explicitado como

$$h(x) = [102213120210212111200020332230232211033120300000123233031010312 \\ 210112020113110132131301231123000100223123122033330231100332023 \\ 211030030001210030121221130133211001131231313303201302332301020 \\ 13200223312301230131132030100021203313133320023031132221323].$$

Passo 9 - Determinar a matriz G e a sua transposta G^T - Através do polinômio gerador $g(x) = 1 + 2x + 3x^2 + x^3 + 3x^4 + 2x^5 + 2x^6 + x^8$, que gera o código BCH desejado, obtemos a matriz G com dimensão 247×255 , isto é,

A matriz G^T com dimensão 255×247 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - A matriz verificação de paridade geradora do código dual associado é dada pela matriz H com dimensão 8×255 , sendo

$$H = \begin{pmatrix} 000000010221312021021211120002033 \cdots 20023031132221323 \\ 00000010221312021021211120002033 \cdots 200230311322213230 \\ 0000010221312021021211120002033 \cdots 2002303113222132300 \\ 000010221312021021211120002033 \cdots 20023031132221323000 \\ 00010221312021021211120002033 \cdots 200230311322213230000 \\ 0010221312021021211120002033 \cdots 2002303113222132300000 \\ 010221312021021211120002033 \cdots 20023031132221323000000 \\ 10221312021021211120002033 \cdots 200230311322213230000000 \end{pmatrix}.$$

A matriz H^T com dimensão 255×8 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, analisaremos se o código BCH sobre anel é capaz de reproduzir sequência de DNA da proteína ($P = \text{Seq.46}$) *EST4002 plasmid pEST4011* - organismo: *Achromobacter denitrificans* - bactéria - GI número: 45368559 com comprimento $n = 255$ nucleotídeos. Esta sequência deve ser rotulada de acordo com as 24 permutações apresentadas na Tabela 4.20, podendo ser representada matricialmente por uma matriz $P_{24 \times 255}$, análoga à matriz P do exemplo anterior, sendo as linhas as 24 permutações entre $N \rightarrow \mathbb{Z}_4$ e as colunas os 255 nucleotídeos. Seja

$\text{Seq.46} = \{5' \text{ ATGTCAGAGGCAAATATCAGGCTTGAATGCCTCGGGCCCGAAACGACGGCTGGAGCA } \\ \text{GCCGACCGCGAAGAGGTGCGCGAGGCCTGAAAGCGCGGGCTTCACGGGAGGCCAAGCCGCGAAAG } \\ \text{CCCTCGGGCTGGGGCAAAGGGCGATAGAACCGTACGCCGCTGGATCGCGGGGATTGGCCATCCCC } \\ \text{TATGCCGCATGGGCCTTGCTGTGCGACTTCGGCAATTGGCCAATCTGGAAGAAGGAC 3'}\}.$

Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(\mathbf{a}, \mathbf{b}) = 0$, $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ - Vamos determinar quais das sequências (sem diferença, com 1 e com 2 nucleotídeos de diferença da sequência do NCBI) são palavras-código dos códigos $(255, k, d_H)$ usando a relação $\mathbf{v} \cdot H^T = 0$.

a) $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Para $D(\mathbf{a}, \mathbf{b}) = 1$, temos 18360 possíveis palavras-código para cada sequência de DNA analisada e, então usamos a relação $\mathbf{v} \cdot H^T = 0$. As palavras-código encontradas são armazenadas.

b) $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Para $D(\mathbf{a}, \mathbf{b}) = 2$, temos 279990 possibilidades de palavras-código para cada caso de permutação de cada sequência de DNA analisada. As palavras-código encontradas também são armazenadas.

Os demais passos do algoritmo são análogos aos passos descritos no exemplo do código $(63, 57, 3)$ sobre $GR(4, 6)$. Portanto não há necessidade em repeti-los.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

| Sequência de DNA do NCBI Seq.46 <i>A.denitrificans</i> - EST4002 plasmid pEST4011 - hypothetical protein - GI número 45368559 | | | |
|--|------------------------------|-------------------------------|-------------------------------|
| 24 casos de permutações | Palavras-código $n = 255$ | Quantidade de palavras-código | Sequência de DNA $n = 255$ |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 0323102 ... 2002201 | 1 | ATGTCAG ... GAAGGAC |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 0000000 ... 0000000 | 0 | - |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 0000000 ... 0000000 | 0 | - |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 0000000 ... 0000000 | 0 | - |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 0121302 ... 2002203 | 1 | ATGTCAG ... GAAGGAC |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 0000000 ... 0000000 | 0 | - |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 0000000 ... 0000000 | 0 | - |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 1232013 ... 3113310 | 1 | ATGTCAG ... GAAGGAC |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 0000000 ... 0000000 | 0 | - |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 1030213 ... 3113312 | 1 | ATGTCAG ... GAAGGAC |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 0000000 ... 0000000 | 0 | - |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 0000000 ... 0000000 | 0 | - |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 0000000 ... 0000000 | 0 | - |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 0000000 ... 0000000 | 0 | - |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 2303120 ... 0220021 | 1 | ATGTCAG ... GAAGGAC |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 0000000 ... 0000000 | 0 | - |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 2101320 ... 0220023 | 1 | ATGTCAG ... GAAGGAC |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 0000000 ... 0000000 | 0 | - |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 3212031 ... 1331130 | 1 | ATGTCAG ... GAAGGAC |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 0000000 ... 0000000 | 0 | - |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 0000000 ... 0000000 | 0 | - |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 0000000 ... 0000000 | 0 | - |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 0000000 ... 0000000 | 0 | - |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 3010231 ... 1331132 | 1 | ATGTCAG ... GAAGGAC |

Tabela 4.27: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ e $n = 255$ nuclóetídeos.

Como resultado das 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$, assim como no exemplo $GR(4, 6)$, também obtemos uma matriz R com 8 palavras-código, sendo uma palavra-código para cada caso de permutação correspondendo aos 8 casos de permutações do rotulamento B, que são diferentes em termos do alfabeto do código, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ (Tabela 4.27). Quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$ os 8 casos de permutações resultam em uma única sequência de DNA, (Figura 4.4).

Veja na Figura 4.4 a proteína, Seq.46, reproduzida pelo código G -linearidade: $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade $((255, 247, 3)$ BCH primitivo sobre $GR(4, 8)$, rotulamento B) através do polinômio primitivo $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ e do polinômio gerador $g_{01}(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1$. Observe que, na posição da trinca 39 ocorreu uma troca de nucleotídeo que não ocasionou a troca de aminoácido nesta posição.

Seq.46| *A.denitrificans*- EST4002 plasmid pEST4011 - hypothetical protein – GI número 45368559

| Código $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade ((255,247,3) BCH primitivo sobre $GR(4,8)$, rotulamento B) | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|---|---|---|--|
| $p_1(x) = x^8 + x^4 + x^3 + x^2 + 1$ – $g_1(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1$ – Caso 01 – (A, C, G, T) = (0, 1, 2, 3) | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | | | | |
| ntO: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCA | AAC | GAC | GCG | TGG | GAG | CAG | CCG | | | | |
| RtO: | 032 | 310 | 202 | 210 | 003 | 031 | 022 | 133 | 200 | 321 | 132 | 122 | 111 | 210 | 001 | 201 | 221 | 322 | 202 | 102 | 112 | | | | |
| RtG: | 032 | 310 | 202 | 210 | 003 | 031 | 022 | 133 | 200 | 321 | 132 | 122 | 111 | 210 | 001 | 201 | 221 | 322 | 202 | 102 | 112 | | | | |
| ntG: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCA | AAC | GAC | GCG | TGG | GAG | CAG | CCG | | | | |
| aaG: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | | | | |
| aaO: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | | | | |
| ntO: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | CTG | AAA | GCG | GCG | GTC | TTC | ACG | GGA | GGC | CAA | GCC | GCG | AAA | | | | |
| RtO: | 011 | 221 | 200 | 202 | 232 | 121 | 202 | 212 | 132 | 000 | 212 | 212 | 221 | 331 | 012 | 220 | 221 | 100 | 211 | 212 | 000 | | | | |
| RtG: | 011 | 221 | 200 | 202 | 232 | 121 | 202 | 212 | 132 | 000 | 212 | 212 | 221 | 331 | 012 | 220 | 221 | 102 | 211 | 212 | 000 | | | | |
| ntG: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | CTG | AAA | GCG | GCG | GTC | TTC | ACG | GGA | GGC | CAG | GCC | GCG | AAA | | | | |
| aaG: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | | | | |
| aaO: | A | L | G | L | G | V | R | E | A | K | G | D | R | T | V | R | R | W | I | G | G | D | S | A | |
| ntO: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GGC | GAT | AGA | ACC | GTA | CGC | CGC | TGG | ATC | GGC | GGG | GAT | TCG | GCC | | | | |
| RtO: | 211 | 131 | 222 | 132 | 222 | 210 | 002 | 221 | 203 | 020 | 011 | 230 | 121 | 121 | 322 | 031 | 221 | 222 | 203 | 312 | 211 | | | | |
| RtG: | 211 | 131 | 222 | 132 | 222 | 210 | 002 | 221 | 203 | 020 | 011 | 230 | 121 | 121 | 322 | 031 | 221 | 222 | 203 | 312 | 211 | | | | |
| ntG: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GGC | GAT | AGA | ACC | GTA | CGC | CGC | TGG | ATC | GGC | GGG | GAT | TCG | GCC | | | | |
| aaG: | A | L | G | L | G | A | K | G | D | R | T | V | R | R | W | I | G | G | D | S | A | | | | |
| aaO: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | | | | |
| ntO: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GGC | AAT | TTG | GGC | CAA | ATC | TGG | AAG | AAG | | | | |
| RtO: | 031 | 111 | 303 | 211 | 210 | 322 | 211 | 332 | 132 | 321 | 201 | 331 | 221 | 003 | 332 | 221 | 100 | 031 | 322 | 002 | 002 | | | | |
| RtG: | 031 | 111 | 303 | 211 | 210 | 322 | 211 | 332 | 132 | 321 | 201 | 331 | 221 | 003 | 332 | 221 | 100 | 031 | 322 | 002 | 002 | | | | |
| ntG: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GGC | AAT | TTG | GGC | CAA | ATC | TGG | AAG | AAG | | | | |
| aaG: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | | | | |
| aaO: | D | | | | | | | | | | | | | | | | | | | | | | | | |
| ntO: | GAC | | | | | | | | | | | | | | | | | | | | | | | | |
| RtO: | 201 | | | | | | | | | | | | | | | | | | | | | | | | |
| RtG: | 201 | | | | | | | | | | | | | | | | | | | | | | | | |
| ntG: | GAC | | | | | | | | | | | | | | | | | | | | | | | | |
| aaG: | D | | | | | | | | | | | | | | | | | | | | | | | | |

Figura 4.4: Sequência de DNA de uma proteína com 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Como resultado das 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$, obtemos várias matrizes diferentes entre si, R' , com 24 palavras-código em cada uma dessas matrizes, correspondente aos 24 casos de permutações divididos nos rotulamentos A, B e C. Assim, como no exemplo do código (63, 57, 3) sobre $GR(4, 6)$, a quantidade de palavras-código geradas para $D(\mathbf{a}, \mathbf{b}) = 2$ também é diferente da quantidade de palavras-código geradas para $D(\mathbf{a}, \mathbf{b}) = 1$ (que é uma palavra-código para cada caso). Neste exemplo, obtemos 381 palavras-código nos casos do rotulamento A, 3 palavras-código nos casos do rotulamento B, 5 palavras-código nos casos do rotulamento C, ver Tabela 4.28.

A Figura 4.5, mostra a proteína Seq.46, reproduzida pelo códigos G -linearidade: \mathbb{Z}_4 -linearidade ((255,247,3) BCH primitivo sobre $GR(4,8)$, rotulamento A); $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade ((255,247,3) BCH primitivo sobre $GR(4,8)$, rotulamento B) e Klein-linearidade ((255,247,3) BCH primitivo sobre $GR(4,8)$, rotulamento C), através do polinômio primitivo $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ e do polinômio gerador $g_{01}(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1$ com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos diferindo da sequênciade NCBI.

| Seq.46 <i>A.denitrificans</i> - protein - GI número 45368559 | |
|--|-------------------------------|
| 24 Permutações | Quantidade de palavras-código |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 381 |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 3 |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 5 |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 5 |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 381 |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 3 |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 3 |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 381 |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 3 |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 381 |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 5 |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 5 |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 5 |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 5 |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 381 |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 3 |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 381 |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 3 |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 381 |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 3 |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 5 |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 5 |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 3 |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 381 |

 Tabela 4.28: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 255$ nucleotídeos.

Seq.46| *A.denitrificans* - EST4002 plasmid pEST4011 - hypothetical protein - GI número 45368559

| Código Z ₄ -linearidade((255,247,3) BCH primitivo sobre GR(4,8), rotulamento A) | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|-----|-----|-----|-----|-----|-----|--|--|--|
| $p_1(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_1(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1 - \text{Caso 02- } (A, C, G, T) = (0, 1, 3, 2)$ | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | | | | |
| ntO: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCA | AAC | GAC | GCG | TGG | GAG | CAG | CCG | | | | |
| RtO: | 023 | 210 | 303 | 310 | 002 | 021 | 033 | 122 | 300 | 231 | 123 | 133 | 111 | 310 | 001 | 301 | 331 | 233 | 303 | 103 | 113 | | | | |
| RtG: | 023 | 210 | 303 | 310 | 002 | 021 | 033 | 122 | 300 | 231 | 123 | 133 | 111 | 310 | 001 | 301 | 331 | 233 | 303 | 103 | 113 | | | | |
| ntG: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCA | AAC | GAC | GCG | TGG | GAG | CAG | CCG | | | | |
| aaG: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | | | | |
| aaO: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | | | | |
| ntO: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | GCG | CTG | AAA | GCG | GCG | GCG | TTC | ACG | GGG | GGC | CAA | GCC | GCG | AAA | | | |
| RtO: | 011 | 331 | 300 | 303 | 323 | 131 | 303 | 313 | 123 | 000 | 313 | 313 | 331 | 221 | 013 | 330 | 331 | 100 | 311 | 313 | 000 | | | | |
| RtG: | 011 | 331 | 300 | 303 | 323 | 131 | 303 | 313 | 123 | 000 | 313 | 313 | 331 | 221 | 013 | 330 | 331 | 100 | 311 | 313 | 000 | | | | |
| ntG: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | GCG | CTG | AAA | GCG | GCG | GCG | TTC | ACG | GGG | GGC | CAA | GCC | GCG | AAA | | | |
| aaG: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | | | | |
| aaO: | A | L | G | L | G | A | K | G | D | R | T | V | R | R | W | I | G | G | D | S | A | | | | |
| ntO: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GCG | GAT | AGA | ACC | GTA | CGC | CGC | TGG | ATC | GGC | GGG | GAT | TCG | GCC | | | | |
| RtO: | 311 | 121 | 333 | 123 | 333 | 310 | 003 | 331 | 302 | 030 | 011 | 320 | 131 | 131 | 233 | 021 | 331 | 333 | 302 | 213 | 311 | | | | |
| RtG: | 311 | 121 | 333 | 123 | 333 | 310 | 003 | 331 | 302 | 030 | 011 | 320 | 131 | 131 | 233 | 121 | 331 | 333 | 302 | 213 | 311 | | | | |
| ntG: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GCG | GAT | AGA | ACC | GTA | CGC | CGC | TGG | CTC | GCG | GGG | GAT | TCG | GCC | | | | |
| aaG: | A | L | G | L | G | A | K | G | D | R | T | V | R | R | W | L | G | G | D | S | A | | | | |
| aaO: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | | | | |
| ntO: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GGC | AAT | TTG | GGC | CAA | ATC | TGG | AAG | AAG | | | | |
| RtO: | 021 | 111 | 202 | 311 | 310 | 233 | 311 | 223 | 123 | 231 | 301 | 221 | 331 | 002 | 223 | 331 | 100 | 021 | 233 | 003 | 003 | | | | |
| RtG: | 021 | 111 | 202 | 311 | 310 | 233 | 311 | 223 | 123 | 231 | 301 | 221 | 331 | 002 | 223 | 330 | 100 | 021 | 233 | 003 | 003 | | | | |
| ntG: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GGC | AAT | TTG | GGA | CAA | ATC | TGG | AAG | AAG | | | | |
| aaG: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | | | | |
| aaO: | D | | | | | | | | | | | | | | | | | | | | | | | | |
| ntO: | GAC | | | | | | | | | | | | | | | | | | | | | | | | |
| RtO: | 301 | | | | | | | | | | | | | | | | | | | | | | | | |
| RtG: | 301 | | | | | | | | | | | | | | | | | | | | | | | | |
| ntG: | GAC | | | | | | | | | | | | | | | | | | | | | | | | |
| aaG: | D | | | | | | | | | | | | | | | | | | | | | | | | |

Seq.46| *A.denitrificans* - EST4002 plasmid pEST4011 - hypothetical protein – GI número 45368559

| Código Z ₂ x Z ₂ -linearidade((255,247,3) BCH primitivo sobre GR(4,8), rotulamento B) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_1(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_1(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1 - \text{Caso } 01 - (A, C, G, T) = (0, 1, 2, 3)$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M S E A N I R L E C L R P A N D G W E Q P ntO: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CTG CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG RtO: 032 310 202 210 003 031 022 133 200 321 132 122 111 210 001 201 221 322 202 102 112 RtG: 032 310 202 210 003 031 022 133 200 321 132 122 111 210 001 201 221 322 202 102 112 ntG: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CTG CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG aaG: M S E A N I R L E C L R P A N D G W E Q P | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: T G E E V R E A L K A A G F T G G Q A A K ntO: ACC GGC GAA GAG GTG CGC GAG GCG CTG AAA GCG CGC GGC TTC ACG GGA GGC CAA GCC GCG AAA RtO: 011 221 200 202 232 121 202 212 132 000 212 212 221 331 012 220 221 100 211 212 000 RtG: 011 221 200 202 232 121 202 212 132 000 212 212 221 331 012 220 221 100 211 212 000 ntG: ACC GGC GAA GAG GTG CGC GAG GCG CTG AAA GCG CGC GGC TTC ACG GGA GGC CAA GCC GCG AAA aaG: T G E E V R E A L K A A G F T G G Q A A K | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: A L G L G A K G D R T V R R W I G G D S A ntO: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC RtO: 211 131 222 132 222 210 002 221 203 020 011 230 121 322 031 221 222 203 312 211 RtG: 211 131 222 132 222 210 002 221 203 020 011 230 121 322 031 221 222 203 312 211 ntG: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC aaG: A L G L G A K G D R T V R R W I G G D S A | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: I P Y A A W A L L C D F G N L G Q I W K K ntO: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CAA ATC TGG AAG AAG RtO: 031 111 303 211 210 322 211 332 132 321 201 331 221 003 332 221 100 031 322 002 002 RtG: 031 111 303 211 210 322 211 332 132 321 201 331 221 003 330 221 120 031 322 002 002 ntG: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CGA ATC TGG AAG AAG aaG: I P Y A A W A L L C D F G N L G R I W K K | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: D ntO: GAC RtO: 201 RtG: 201 ntG: GAC aaG: D | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Seq.46| *A.denitrificans* - EST4002 plasmid pEST4011 - hypothetical protein – GI número 45368559

| Código Klein-linearidade((255,247,3) BCH primitivo sobre GR(4,8), rotulamento C) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1 - \text{Caso } 3 - (A, C, G, T) = (0, 2, 1, 3)$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M S E A N I R L E C L R P A N D G W E Q P ntO: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CTG CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG RtO: 031 320 101 120 003 032 011 233 100 312 231 211 222 120 002 102 112 311 101 201 221 RtG: 031 320 101 120 003 032 011 233 100 312 231 211 222 120 002 102 112 311 101 201 221 ntG: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CTG CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG aaG: M S E A N I R L E C L R P A N D G W E Q P | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: T G E E V R E A L K A A G F T G G Q A A K ntO: ACC GGC GAA GAG GTG CGC GAG GCG CTG AAA GCG CGC GGC TTC ACG GGA GGC CAA GCC GCG AAA RtO: 022 112 100 101 131 212 101 121 231 000 121 121 112 332 021 110 112 200 122 121 000 RtG: 022 112 100 101 131 212 101 121 231 000 121 121 112 332 021 110 112 200 122 121 000 ntG: ACC GGC GAA GAG GTG CGC GAG GCG CGA AAA GCG CGC TTC ACG GGA GGC CAA GCC GCG AAA aaG: T G E E V R E A Q K A A R F T G G Q A A K | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: A L G L G A K G D R T V R R W I G G D S A ntO: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC RtO: 122 232 111 231 111 120 001 112 103 010 022 130 212 212 311 032 112 111 103 321 122 RtG: 122 232 111 231 111 120 001 112 103 010 022 130 212 212 311 032 112 111 103 321 122 ntG: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC aaG: A L G L G A K G D R T V R R W I G G D S A | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: I P Y A A W A L L C D F G N L G Q I W K K ntO: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CAA ATC TGG AAG AAG RtO: 032 222 303 122 120 311 122 331 231 312 102 332 112 003 331 112 200 032 311 001 001 RtG: 032 222 303 122 120 311 122 331 231 312 102 332 112 003 331 112 200 032 311 001 001 ntG: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CAA ATC TGG AAG AAG aaG: I P Y A A W A L L C D F G N L G Q I W K K | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: D ntO: GAC RtO: 102 RtG: 102 ntG: GAC aaG: D | | | | | | | | | | | | | | | | | | | | | | | | | | | |

Figura 4.5: Sequência de DNA de uma proteína com 255 nucleotídeos e $D(a, b) = 2$ nos rotulamentos A e rotulamento B e C.

Observe na proteína reproduzida no rotulamento A que o nucleotídeo da posição 58 foi alterado, ocasionando a troca de aminoácido nesta posição, (I58L), porém o nucleotídeo alterado da posição 79 não ocasionou a troca de aminoácido. Já a proteína reproduzida no rotulamento B, o nucleotídeo da posição 78 foi alterado não ocasionando a troca de aminoácido nesta posição, enquanto o nucleotídeo e o aminoácido da posição 80 foram alterados, (Q80R). E na proteína reproduzida no rolamento C, os dois nucleotídeos alterados nas posições das trincas 30 e 34 ocasionaram a troca dos aminoácidos nestas posições, (L30Q) e (G34R), respectivamente.

Código BCH $(n, k, d_H) = (1023, k, d_H)$ sobre $GR(4, 10)$

Considere ainda a construção do código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (1023, 1013, 3)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = 1023$ nucleotídeos.

Passo 2 - Determinar a extensão de Galois - O grau da extensão de Galois do corpo $GF(2)$ é $r = 10$, pois $n = (2^r - 1) = (2^{10} - 1) = 1023$.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Em [37], encontramos 60 polinômios primitivos relacionados à extensão de Galois de grau $r = 10$, que serão identificados no Capítulo 5. Neste exemplo, vamos usar o polinômio primitivo $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1$ para realizarmos a extensão do corpo no **Passo 4**.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - Considere o corpo $GF(2^r) = GF(2^{10}) = GF(1024) = F_{1024}$ dado por

$$\begin{aligned} \frac{F_2[x]}{\langle p_{42}(x) \rangle} &\cong \frac{F_2[x]}{\langle x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 \rangle} = \\ &= \{a_0 + a_1x + a_2x^2 + \cdots + a_9x^9 : a_i \text{ }' s \in F_2\}. \end{aligned}$$

Assim, F_{1024} é formado pelos elementos contidos na Tabela 4.29.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Considere ainda o anel $GR(4, r) = GR(4, 10)$ dado pela extensão de Galois de grau $r = 10$ abaixo:

$$\begin{aligned} GR(4, 10)[x] &\cong \frac{\mathbb{Z}_4[x]}{\langle p_{42}(x) \rangle} \cong \frac{\mathbb{Z}_4[x]}{\langle x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 \rangle} = \\ &= \{b_0 + b_1x + b_2x^2 + \cdots + b_9x^9 : b_i \text{ }' s \in \mathbb{Z}_4\}. \end{aligned}$$

| Elementos de F_{1024} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7 \alpha^8 \alpha^9)$ |
|---|---|
| 0 | (0000000000) |
| 1 | (1000000000) |
| α | (0100000000) |
| α^2 | (0010000000) |
| α^3 | (0001000000) |
| α^4 | (0000100000) |
| α^5 | (0000010000) |
| α^6 | (0000001000) |
| α^7 | (0000000100) |
| α^8 | (0000000010) |
| α^9 | (0000000001) |
| α^{10} | (1101101111) |
| $\alpha^{11} = \alpha \times \alpha^{10}$ | (?) |
| \vdots | \vdots |
| $\alpha^{1020} = \alpha \times \alpha^{1019}$ | (0000011101) |
| $\alpha^{1021} = \alpha \times \alpha^{1020}$ | (1101100001) |
| $\alpha^{1022} = \alpha \times \alpha^{1021}$ | (1011011111) |
| $\alpha^{1023} = \alpha \times \alpha^{1022}$ | (1000000000) |

 Tabela 4.29: Elementos de F_{1024} com $p_{42}(x)$.

As operações em $GR^*(4, 10)$ são realizadas módulo $(x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1)$. Logo, $\alpha^{10} = 3\alpha^9 + 3\alpha^8 + 3\alpha^7 + 3\alpha^6 + 3\alpha^4 + 3\alpha^3 + 3\alpha + 3$, pois os coeficientes de $GR^*(4, 10)$ estão em \mathbb{Z}_4 . Assim, considerando $f = (01000000) = \alpha$, os elementos do grupo cíclico do grupo $GR^*(4, 10)$ são dados na Tabela 4.30.

| $GR^*(4, 8)$ | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7) \alpha^8 \alpha^9$ |
|---|---|
| 1 | (1000000000) |
| $f = x = \alpha$ | (0100000000) |
| $f^2 = x^2 = \alpha^2$ | (0010000000) |
| $f^3 = x^3 = \alpha^3$ | (0001000000) |
| $f^4 = x^4 = \alpha^4$ | (0000100000) |
| $f^5 = x^5 = \alpha^5$ | (0000010000) |
| $f^6 = x^6 = \alpha^6$ | (0000001000) |
| $f^7 = x^7 = \alpha^7$ | (0000000100) |
| $f^8 = x^8 = \alpha^8$ | (0000000010) |
| $f^9 = x^9 = \alpha^9$ | (0000000001) |
| $f^{10} = x^{10} = \alpha^{10}$ | (3303303333) |
| $f^{11} = f \times f^{10} = x^{11} = \alpha^{11}$ | (1031031000) |
| \vdots | \vdots |
| $f^{2043} = f \times f^{2042} = x^{2043} = \alpha^{2043}$ | (2002033303) |
| $f^{2044} = f \times f^{2043} = x^{2044} = \alpha^{2044}$ | (1301300001) |
| $f^{2045} = f \times f^{2044} = x^{2045} = \alpha^{2045}$ | (3033033333) |
| $f^{2046} = f \times f^{2045} = x^{2046} = \alpha^{2046}$ | (1000000000) |

 Tabela 4.30: Elementos do grupo cíclico do grupo $GR^*(4, 10)$ com $p_{42}(x)$.

Passo 6 - Determinar o grupo das unidades - Como f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 10)$, e f^d gera o subgrupo cíclico cuja ordem é 1023 em $GR^*(4, 10)$, então $n \cdot d = 1023 \cdot d = 2046$ implicando que $d = 2$. Consequentemente, $f^2 = (0010000000) = \alpha^2$ gera um subgrupo cíclico de ordem 1023 em $GR^*(4, 10)$. Logo, $\beta = \alpha^2$ é o elemento

primitivo em G_{1023} , na Tabela 4.31.

| |
|--|
| $G_{255} \rightarrow (\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5\alpha^6\alpha^7\alpha^8\alpha^9)$ |
| $(f^2)^1 = f^2 = \alpha^2 = \beta \rightarrow (001000000)$ |
| $(f^2)^2 = f^4 = \alpha^4 = \beta^2 \rightarrow (000010000)$ |
| $(f^2)^3 = f^6 = \alpha^6 = \beta^3 \rightarrow (0000001000)$ |
| $(f^2)^4 = f^8 = \alpha^8 = \beta^4 \rightarrow (0000000010)$ |
| $(f^2)^5 = f^{10} = \alpha^{10} = \beta^5 \rightarrow (3303303333)$ |
| ⋮ |
| $(f^2)^{1021} = f^{2042} = \alpha^{2042} = \beta^{1021} \rightarrow (2002311212)$ |
| $(f^2)^{1022} = f^{2044} = \alpha^{2044} = \beta^{1022} \rightarrow (1301300001)$ |
| $(f^2)^{1023} = f^{2046} = \alpha^{2046} = \beta^{1023} \rightarrow (1000000000)$ |

Tabela 4.31: Elementos de G_{1023}

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Podemos agora construir um código BCH de comprimento $n = 1023$ sobre \mathbb{Z}_4 . Considerando que a distância mínima do código seja $d_H \geq 3$, o polinômio gerador $g(x)$ do código tem como raízes β e β^2 . Este polinômio é dado por $g(x) = mmc\{M_1(x), M_2(x)\}$, onde $M_i(x)$ é o polinômio minimal de β^i , $i = 1, 2$. O polinômio minimal de β é dado por $M_1(x) =$

$$\{(x-\beta)(x-\beta^2)(x-\beta^4)(x-\beta^8)(x-\beta^{16})(x-\beta^{32})(x-\beta^{64})(x-\beta^{128})(x-\beta^{256})(x-\beta^{512})\} =$$

$$= 1 + 3x + x^3 + x^4 + x^6 + 3x^7 + x^8 + x^9 + x^{10}.$$

Note que os polinômios minimais de β e β^2 são iguais, isto é, $M_1(x) = M_2(x)$.

Assim, $g_{42}(x) = 1 + 3x + x^3 + x^4 + x^6 + 3x^7 + x^8 + x^9 + x^{10}$ gera o código BCH desejado sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (1023, 1013, 3)$.

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H , é dado por:

$$h(x) = \frac{x^n - 1}{g_{42}(x)} = \frac{x^{1023} - 1}{x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1}$$

$$h(x) = x^{1013} + 3x^{1012} + 2x^{1010} + 3x^{1008} + 2x^{2007} + \dots + 2x^6 + 2x^4 + 3x^2 + 3x + 3,$$

onde os coeficientes do polinômio $h(x)$ pertencem a \mathbb{Z}_4 .

Passo 9 - Determinar a matriz G e a sua transposta G^T - O polinômio gerador $g_{42}(x) = 1 + 3x + x^3 + x^4 + x^6 + 3x^7 + x^8 + x^9 + x^{10}$ está relacionado à matriz geradora G com dimensão 1013×1023 , isto é,

$$G = \begin{pmatrix} 13011013111000000000000000 \dots 00 \\ 01301101311100000000000000 \dots 00 \\ \vdots & \ddots & \vdots \\ 00 \\ 00 \end{pmatrix}.$$

A matriz G^T com dimensão 1023×1013 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - A matriz verificação de paridade geradora do código dual associado é dada pela matriz H com dimensão 10×1013 , sendo

$$H = \begin{pmatrix} 0000000001302032110202323110 \dots 11212130132212302020333 \\ 000000001302032110202323110 \dots 112121301322123020203330 \\ 00000001302032110202323110 \dots 1121213013221230202033300 \\ 0000001302032110202323110 \dots 11212130132212302020333000 \\ 000001302032110202323110 \dots 112121301322123020203330000 \\ 00001302032110202323110 \dots 1121213013221230202033300000 \\ 0001302032110202323110 \dots 11212130132212302020333000000 \\ 0001302032110202323110 \dots 112121301322123020203330000000 \\ 001302032110202323110 \dots 1121213013221230202033300000000 \\ 001302032110202323110 \dots 11212130132212302020333000000000 \\ 01302032110202323110 \dots 11212130132212302020333000000000 \\ 1302032110202323110 \dots 11212130132212302020333000000000 \end{pmatrix}.$$

A matriz H^T com dimensão 1023×10 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, analisaremos se o código BCH sobre anel é capaz de reproduzir a sequência de DNA da proteína (P = Seq.51) *Malate dehydrogenase 1* - organismo: *Arabidopsis thaliana* - *Viridiplantae* - GI número: 30695458 com comprimento $n = 1023$ nucleotídeos. Essa proteína deve ser rotulada de acordo com as 24 permutações já apresentadas no exemplo anterior.

Os demais passos do algoritmo são análogos aos passos dos exemplos das extensões de grau 6 e 8.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Assim como nos exemplos anteriores, como resultado das 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$, 8 palavras-código correspondentes aos casos de permutações do rotulamento C, foram encontradas. Logo, a proteína *Malate dehydrogenase 1*, Seq.51, mostrada na Figura 4.6, foi reproduzida com 1 nucleotídeo de diferença da sequência do NCBI pelo código G -linearidade:Klein-linearidade ((1023, 1013, 3) BCH primitivo sobre $GR(4, 10)$, rotulamento C) através do polinômio primitivo $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ e do polinômio gerador

$g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$. Observe que na posição da trinca 289 ocorreu uma troca de nucleotídeo ocasionando a troca de aminoácido nesta posição, sendo (L289F).

Seq.51| A. thaliana - Mitocôndria - Malate dehydrogenase 1 – GI número 30695458

| Código Klein-linearidade ((1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento C) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 - g(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Caso 03 - (A,C,G,T) = (0,2,1,3) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M F R S M L V R S S A S A K Q A V I R R S ntO: ATG TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC Rto: 031 332 010 323 031 232 132 210 323 323 122 322 121 001 201 121 133 032 212 213 012 Rtg: 031 332 010 323 031 232 132 210 323 323 122 322 121 001 201 121 133 032 212 213 012 ntG: ATG TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC aaG: M F R S M L V R S S A S A K Q A V I R R S | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: F S S G S V P E R K V A I L G A A G G I G ntO: TTC TCC TCC GGC TCC GTC CCC GAG CGT AAA GTC GCC ATC CTT GGT GCC GCC GGT GGA ATT GGT Rto: 332 322 322 112 322 132 222 101 213 000 132 122 032 233 113 122 122 113 110 033 113 Rtg: 332 322 322 112 322 132 222 101 213 000 132 122 032 233 113 122 122 113 110 033 113 ntG: TTC TCC TCC GGC TCC GTC CCC GAG CGT AAA GTC GCC ATC CTT GGT GCC GCC GGT GGA ATT GGT aaG: F S S G S V P E R K V A I L G A A G G I G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: Q P L A L M K L N P L V S S L S L Y D I ntO: CAG CCT CTT GCT CTC CTC ATG AAG CTT AAT CCT CTT GTC TCT TCC CTC TCC CTC TAC GAT ATC Rto: 201 223 233 123 232 031 001 233 003 223 233 132 323 322 232 322 232 302 103 032 Rtg: 201 223 233 123 232 031 001 233 003 223 233 132 323 322 232 322 232 302 103 032 ntG: CAG CCT CTT GCT CTC CTC ATG AAG CTT AAT CCT CTT GTC TCT TCC CTC TCC CTC TAC GAT ATC aaG: Q P L A L M K L N P L V S S L S L Y D I | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: A N T P G V A A D V G H I N T R S E V V G ntO: GCT AAC ACT CCT GGA GTT GCT GAT GTT GGT CAC ATC AAC ACC AGA TCT GAG GTT GTT GGA Rto: 123 002 023 223 110 133 123 103 133 113 202 032 002 022 010 323 101 133 133 110 Rtg: 123 002 023 223 110 133 123 103 133 113 202 032 002 022 010 323 101 133 133 110 ntG: GCT AAC ACT CCT GGA GTT GCT GAT GTT GGT CAC ATC AAC ACC AGA TCT GAG GTT GTT GGA aaG: A N T P G V A A D V G H I N T R S E V V G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Oaa: Y M G D D N L A K A L E G A D L V I I P A ntO: TAC ATG GGC GAT GAT AAC TTG GCC AAA GCT CTT GAA GGA GCT GAT CTC GTT ATC ATT CCA GCT Rto: 302 031 112 103 103 002 331 122 000 123 233 100 110 123 103 232 133 032 033 220 123 Rtg: 302 031 112 103 103 002 331 122 000 123 233 100 110 123 103 232 133 032 033 220 123 ntG: TAC ATG GGC GAT GAT AAC TTG GCC AAA GCT CTT GAA GGA GCT GAT CTC GTT ATC ATT CCA GCT aaG: Y M G D D N L A K A L E G A D L V I I P A | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: G V P R K P G M T R D D L F N I N A G I V ntO: GGT GTA CCA AGG AAG CCT GGT ATG ACC CGT GAC GAT CTT TTC AAC ATT AAT GCT GGA ATT GTC Rto: 113 130 220 011 001 223 113 031 022 213 102 103 233 332 002 033 003 123 110 033 132 Rtg: 113 130 220 011 001 223 113 031 022 213 102 103 233 332 002 033 003 123 110 033 132 ntG: GGT GTA CCA AGG AAG CCT GGT ATG ACC CGT GAC GAT CTT TTC AAC ATT AAT GCT GGA ATT GTC aaG: G V P R K P G M T R D D L F N I N A G I V | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: K N L C T A I A K Y C P H A L I N M I S N ntO: AAG AAC CTT TGC ACT GCC ATC GCC AAG TAC TGC CCA CAT GCG CTT ATT AAT ATG ATC AGC AAC Rto: 001 002 233 312 023 122 032 122 001 302 312 220 203 121 233 033 003 031 032 012 002 Rtg: 001 002 233 312 023 122 032 122 001 302 312 220 203 121 233 033 003 031 032 012 002 ntG: AAG AAC CTT TGC ACT GCC ATC GCC AAG TAC TGC CCA CAT GCG CTT ATT AAT ATG ATC AGC AAC aaG: K N L C T A I A K Y C P H A L I N M I S N | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: P V N S T V P I A A E I F K K A G M Y D E ntO: CCT GTG AAC TCT ACT GTT CCA ATT GCA GCT GAG ATA TTT AAG AAG GCT GGT ATG TAC GAT GAA Rto: 223 131 002 323 023 133 220 033 120 123 101 030 333 001 001 123 113 031 302 103 100 Rtg: 223 131 002 323 023 133 220 033 120 123 101 030 333 001 001 123 113 031 302 103 100 ntG: CCT GTG AAC TCT ACT GTT CCA ATT GCA GCT GAG ATA TTT AAG AAG GCT GGT ATG TAC GAT GAA aaG: P V N S T V P I A A E I F K K A G M Y D E | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|
| aao: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| nto: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| Rto: | 001 | 000 | 331 | 333 | 113 | 133 | 022 | 023 | 233 | 102 | 132 | 132 | 011 | 122 | 011 | 023 | 332 | 303 | 123 | 110 | 001 |
| Rtg: | 001 | 000 | 331 | 333 | 113 | 133 | 022 | 023 | 233 | 102 | 132 | 132 | 011 | 122 | 011 | 023 | 332 | 303 | 123 | 110 | 001 |
| ntg: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| aag: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| aao: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| nto: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | ATT | GGT | CAT | GCT | GGG | GTT | ACT | ATT | |
| Rto: | 120 | 003 | 132 | 220 | 133 | 120 | 100 | 133 | 003 | 133 | 221 | 131 | 033 | 113 | 113 | 203 | 123 | 111 | 133 | 023 | 033 |
| Rtg: | 120 | 003 | 132 | 220 | 133 | 120 | 100 | 133 | 003 | 133 | 221 | 131 | 033 | 113 | 113 | 203 | 123 | 111 | 133 | 023 | 033 |
| ntg: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | ATT | GGT | CAT | GCT | GGG | GTT | ACT | ATT | |
| aag: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| aao: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| nto: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT |
| Rto: | 232 | 223 | 232 | 332 | 323 | 201 | 120 | 023 | 223 | 200 | 122 | 002 | 331 | 320 | 013 | 102 | 030 | 233 | 022 | 122 | 233 |
| Rtg: | 232 | 223 | 232 | 332 | 323 | 201 | 120 | 023 | 223 | 200 | 122 | 002 | 331 | 320 | 013 | 102 | 030 | 233 | 022 | 122 | 233 |
| ntg: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT |
| aag: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| aao: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| nto: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| Rto: | 023 | 001 | 213 | 022 | 200 | 103 | 110 | 113 | 020 | 100 | 132 | 131 | 101 | 120 | 000 | 120 | 110 | 000 | 113 | 320 | 123 |
| Rtg: | 023 | 001 | 213 | 022 | 200 | 103 | 110 | 113 | 020 | 100 | 132 | 131 | 101 | 120 | 000 | 120 | 110 | 000 | 113 | 320 | 123 |
| ntg: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| aag: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| aao: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| nto: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| Rto: | 020 | 331 | 322 | 031 | 122 | 303 | 122 | 110 | 120 | 331 | 332 | 123 | 103 | 120 | 312 | 331 | 000 | 110 | 232 | 002 | 113 |
| Rtg: | 020 | 331 | 322 | 031 | 122 | 303 | 122 | 110 | 120 | 331 | 332 | 123 | 103 | 120 | 312 | 331 | 000 | 110 | 232 | 002 | 113 |
| ntg: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| aag: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| aao: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A |
| nto: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC |
| Rto: | 133 | 220 | 103 | 132 | 030 | 100 | 312 | 320 | 302 | 131 | 200 | 323 | 020 | 032 | 022 | 101 | 233 | 223 | 332 | 333 | 122 |
| Rtg: | 133 | 220 | 103 | 132 | 030 | 100 | 312 | 320 | 302 | 131 | 200 | 323 | 020 | 032 | 022 | 101 | 333 | 223 | 332 | 333 | 122 |
| ntg: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | TTT | CCT | TTC | TTT | GCC |
| aag: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | F | P | F | F | A |
| aao: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| nto: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GTT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| Rto: | 321 | 001 | 131 | 011 | 331 | 111 | 001 | 003 | 113 | 131 | 101 | 101 | 133 | 233 | 102 | 331 | 110 | 220 | 232 | 320 | 102 |
| Rtg: | 321 | 001 | 131 | 011 | 331 | 111 | 001 | 003 | 113 | 131 | 101 | 101 | 133 | 233 | 102 | 331 | 110 | 220 | 232 | 320 | 102 |
| ntg: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GTT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| aag: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| aao: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| nto: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| Rto: | 333 | 101 | 001 | 100 | 112 | 331 | 100 | 120 | 331 | 001 | 220 | 100 | 232 | 001 | 322 | 322 | 030 | 100 | 001 | 110 | 132 |
| Rtg: | 333 | 101 | 001 | 100 | 112 | 331 | 100 | 120 | 331 | 001 | 220 | 100 | 232 | 001 | 322 | 322 | 030 | 100 | 001 | 110 | 132 |
| ntg: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| aag: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| aao: | K | F | A | N | Q | | | | | | | | | | | | | | | | |
| nto: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| Rto: | 001 | 333 | 122 | 002 | 201 | | | | | | | | | | | | | | | | |
| Rtg: | 001 | 333 | 122 | 002 | 201 | | | | | | | | | | | | | | | | |
| ntg: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| aag: | K | F | A | N | Q | | | | | | | | | | | | | | | | |

 Figura 4.6: Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Para $D(\mathbf{a}, \mathbf{b}) = 2$, também obtemos várias palavras-código nos 24 casos de permutações que se dividem nos rotulamentos A, B e C. Assim como nos exemplos dos códigos (63, 57, 3) sobre $GR(4, 6)$ e (251, 247, 3) sobre $GR(4, 8)$, a quantidade de palavras-código geradas para $D(\mathbf{a}, \mathbf{b}) = 2$ também é diferente da quantidade de palavras-código geradas para $D(\mathbf{a}, \mathbf{b}) = 1$ (uma palavra-código para cada caso). Neste exemplo, obtemos 5 palavras-código nos casos do rotulamento A, 3 palavras-código nos casos do rotulamento B, 1023 palavras-código nos casos do rotulamento C, ver Tabela 4.32.

| Sequência de DNA do NCBI Seq.51 <i>A.thaliana</i> - Malate dehydrogenase 1 - GI número 30695458 | |
|--|-------------------------------|
| 24 Permutações | Quantidade de palavras-código |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 3 |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 5 |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 1023 |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 1023 |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 3 |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 5 |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 5 |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 3 |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 5 |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 3 |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 1023 |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 1023 |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 1023 |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 1023 |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 3 |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 5 |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 3 |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 5 |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 3 |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 5 |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 1023 |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 1023 |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 5 |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 3 |

Tabela 4.32: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 1023$ nucleotídeos.

As Figuras 4.7, 4.8 e 4.9, mostram a proteína *Malate dehydrogenase 1*, Seq.51, com $n = 1023$ nucleotídeos reproduzida pelo códigos G -linearidade: \mathbb{Z}_4 -linearidade ((1023,1013,3) BCH primitivo sobre $GR(4, 10)$, rotulamento A); $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade ((1023,1013,3) BCH primitivo sobre $GR(4, 10)$, rotulamento B) e Klein-linearidade ((1023,1013,3) BCH primitivo sobre $GR(4, 10)$, rotulamento C), todos através do polinômio primitivo $p_{42}(x) = x^{10}+x^9+x^8+x^7+x^6+x^4+x^3+x+1$ e do polinômio gerador $g_{42}(x) = x^{10}+x^9+x^8+3x^7+x^6+x^4+x^3+3x+1$ com $D(a, b) = 2$ nucleotídeos diferindo da sequência do *NCBI*, da seguinte maneira:

- A proteína reproduzida no rotulamento A, os nucleotídeos nas posições das trincas 70 e 240 foram alterados, porém não ocasionaram a troca dos aminoácidos nestas posições,

respectivamente;

Seq.51| *A. thaliana* - Mitocôndria - Malate dehydrogenase 1 – GI número 30695458

| Código Z ₄ -linearidade ((1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento A) | | | | | | | | | | | | | | | | | | | | | | | | | | |
|---|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|--|--|--|--|
| p ₄₂ (x)=x ¹⁰ +x ⁹ +x ⁸ +x ⁷ +x ⁶ +x ⁴ +x ³ +x+1 - g ₄₂ (x)=x ¹⁰ +x ⁹ +x ⁸ +3x ⁷ +x ⁶ +x ⁴ +x ³ +3x+1 | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Caso 2 – (A,C,G,T) = (0,1,3,2) | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: | M | F | R | S | M | L | V | R | S | S | A | S | A | K | Q | A | V | I | R | R | S | | | | | |
| ntO: | ATG | TTC | AGA | TCT | ATG | CTC | GTC | CGA | TCT | TCT | GCC | TCC | GCG | AAG | CAG | GCG | GTT | ATC | CGC | CGT | AGC | | | | | |
| RtO: | 023 | 221 | 030 | 212 | 023 | 121 | 321 | 130 | 212 | 212 | 311 | 211 | 313 | 003 | 103 | 313 | 322 | 021 | 131 | 132 | 031 | | | | | |
| RtG: | 023 | 221 | 030 | 212 | 023 | 121 | 321 | 130 | 212 | 212 | 311 | 211 | 313 | 003 | 103 | 313 | 322 | 021 | 131 | 132 | 031 | | | | | |
| ntG: | ATG | TTC | AGA | TCT | ATG | CTC | GTC | CGA | TCT | TCT | GCC | TCC | GCG | AAG | CAG | GCG | GTT | ATC | CGC | CGT | AGC | | | | | |
| aaG: | M | F | R | S | M | L | V | R | S | S | A | S | A | K | Q | A | V | I | R | R | S | | | | | |
| aaO: | F | S | S | G | S | V | P | E | R | K | V | A | I | L | G | A | A | G | G | I | G | | | | | |
| ntO: | TTC | TCC | TCC | GGC | TCC | GTC | CCC | GAG | CGT | AAA | GTC | GCC | ATC | CTT | GGT | GCC | GCC | GGT | GGA | ATT | GGT | | | | | |
| RtO: | 221 | 211 | 331 | 211 | 321 | 111 | 303 | 132 | 000 | 321 | 311 | 021 | 122 | 332 | 311 | 311 | 332 | 330 | 022 | 332 | | | | | | |
| RtG: | 221 | 211 | 331 | 211 | 321 | 111 | 303 | 132 | 000 | 321 | 311 | 021 | 122 | 332 | 311 | 311 | 332 | 330 | 022 | 332 | | | | | | |
| ntG: | TTC | TCC | GGC | TCC | GTC | CCC | GAG | CGT | AAA | GTC | GCC | ATC | CTT | GGT | GCC | GCC | GGT | GGA | ATT | GGT | | | | | | |
| aaG: | F | S | S | G | S | V | P | E | R | K | V | A | I | L | G | A | A | G | G | I | G | | | | | |
| aaO: | Q | P | L | A | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I | | | | | |
| ntO: | CAG | CCT | CTT | GCT | CTC | CTC | ATG | AAG | CTT | AAT | CCT | GTC | TCT | TCC | CTC | TCC | CTC | TAC | GAT | ATC | | | | | | |
| RtO: | 103 | 112 | 122 | 312 | 121 | 121 | 023 | 003 | 122 | 002 | 112 | 122 | 321 | 212 | 211 | 121 | 211 | 121 | 201 | 302 | 021 | | | | | |
| RtG: | 103 | 112 | 122 | 312 | 121 | 121 | 023 | 003 | 122 | 002 | 112 | 122 | 321 | 212 | 211 | 121 | 211 | 121 | 201 | 302 | 021 | | | | | |
| ntG: | CAG | CCT | CTT | GCT | CTC | CTC | ATG | AAG | CTT | AAT | CCT | GTC | TCT | TCC | CTC | TCC | CTC | TAC | GAT | ATC | | | | | | |
| aaG: | Q | P | L | A | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I | | | | | |
| aaO: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | E | V | V | G | | | | | |
| ntO: | GCT | AAC | ACT | CCT | GGA | GTT | GCT | GCT | GAT | GTT | GGT | CAC | ATC | AAC | ACC | AGA | TCT | GAG | GTT | GTT | GGA | | | | | |
| RtO: | 312 | 001 | 012 | 112 | 330 | 322 | 312 | 312 | 302 | 322 | 332 | 101 | 021 | 001 | 011 | 030 | 212 | 303 | 322 | 322 | 330 | | | | | |
| RtG: | 312 | 001 | 012 | 112 | 330 | 322 | 311 | 312 | 302 | 322 | 332 | 101 | 021 | 001 | 011 | 030 | 212 | 303 | 322 | 322 | 330 | | | | | |
| ntG: | GCT | AAC | ACT | CCT | GGA | GTT | GCC | GCT | GAT | GTT | GGT | CAC | ATC | AAC | ACC | AGA | TCT | GAG | GTT | GTT | GGA | | | | | |
| aaG: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | E | V | V | G | | | | | |
| aaO: | Y | M | G | D | D | N | L | A | K | A | L | E | G | A | D | L | V | I | I | P | A | | | | | |
| ntO: | TAC | ATG | GGC | GAT | GAT | AAC | TTG | GCC | AAA | GCT | CTT | GAA | GGA | GCT | GAT | CTC | GTT | ATC | ATT | CCA | GCT | | | | | |
| RtO: | 201 | 023 | 331 | 302 | 302 | 001 | 223 | 311 | 000 | 312 | 122 | 300 | 330 | 312 | 302 | 121 | 322 | 021 | 022 | 110 | 312 | | | | | |
| RtG: | 201 | 023 | 331 | 302 | 302 | 001 | 223 | 311 | 000 | 312 | 122 | 300 | 330 | 312 | 302 | 121 | 322 | 021 | 022 | 110 | 312 | | | | | |
| ntG: | TAC | ATG | GGC | GAT | GAT | AAC | TTG | GCC | AAA | GCT | CTT | GAA | GGA | GCT | GAT | CTC | GTT | ATC | ATT | CCA | GCT | | | | | |
| aaG: | Y | M | G | D | D | N | L | A | K | A | L | E | G | A | D | L | V | I | I | P | A | | | | | |
| aaO: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V | | | | | |
| ntO: | GGT | GTA | CCA | AGG | AAG | CCT | GGT | ATG | ACC | CGT | GAC | GAT | CTT | TTC | AAC | ATT | AAT | GCT | GGA | ATT | GTC | | | | | |
| RtO: | 332 | 320 | 110 | 033 | 003 | 112 | 332 | 023 | 011 | 132 | 301 | 302 | 122 | 221 | 001 | 022 | 002 | 312 | 330 | 022 | 321 | | | | | |
| RtG: | 332 | 320 | 110 | 033 | 003 | 112 | 332 | 023 | 011 | 132 | 301 | 302 | 122 | 221 | 001 | 022 | 002 | 312 | 330 | 022 | 321 | | | | | |
| ntG: | GGT | GTA | CCA | AGG | AAG | CCT | GGT | ATG | ACC | CGT | GAC | GAT | CTT | TTC | AAC | ATT | AAT | GCT | GGA | ATT | GTC | | | | | |
| aaG: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V | | | | | |
| aaO: | K | N | L | C | T | A | I | A | K | Y | C | P | H | A | L | I | N | M | I | S | N | | | | | |
| ntO: | AAG | AAC | CTT | TGC | ACT | GCC | ATC | GCC | AAG | TAC | TGC | CCA | CAT | GCG | CTT | ATT | AAT | GCT | GGA | ATT | GTC | | | | | |
| RtO: | 003 | 001 | 122 | 231 | 012 | 311 | 021 | 311 | 003 | 201 | 231 | 110 | 102 | 313 | 122 | 022 | 002 | 023 | 021 | 031 | 001 | | | | | |
| RtG: | 003 | 001 | 122 | 231 | 012 | 311 | 021 | 311 | 003 | 201 | 231 | 110 | 102 | 313 | 122 | 022 | 002 | 023 | 021 | 031 | 001 | | | | | |
| ntG: | AAG | AAC | CTT | TGC | ACT | GCC | ATC | GCC | AAG | TAC | TGC | CCA | CAT | GCG | CTT | ATT | AAT | ATG | ATC | AGC | AAC | | | | | |
| aaG: | K | N | L | C | T | A | I | A | K | Y | C | P | H | A | L | I | N | M | I | S | N | | | | | |
| aaO: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | M | Y | D | E | | | | | |
| ntO: | CCT | GTG | AAC | TCT | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATA | TTT | AAG | AAG | GCT | GGT | ATG | TAC | GAT | GAA | | | | | |
| RtO: | 112 | 323 | 001 | 212 | 012 | 322 | 110 | 022 | 310 | 312 | 303 | 020 | 222 | 003 | 003 | 312 | 332 | 023 | 201 | 302 | 300 | | | | | |
| RtG: | 112 | 323 | 001 | 212 | 012 | 322 | 110 | 022 | 310 | 312 | 303 | 020 | 222 | 003 | 003 | 312 | 332 | 023 | 201 | 302 | 300 | | | | | |
| ntG: | CCT | GTG | AAC | TCT | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATA | TTT | AAG | AAG | GCT | GGT | ATG | TAC | GAT | GAA | | | | | |
| aaG: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | M | Y | D | E | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|--|
| aaO: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K | |
| ntO: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG | |
| RtO: | 003 | 000 | 223 | 222 | 332 | 322 | 011 | 012 | 122 | 301 | 321 | 321 | 033 | 311 | 033 | 012 | 221 | 202 | 312 | 330 | 003 | |
| RtG: | 003 | 000 | 223 | 222 | 332 | 322 | 011 | 012 | 122 | 301 | 321 | 321 | 033 | 311 | 033 | 012 | 221 | 202 | 312 | 330 | 003 | |
| ntG: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG | |
| aaG: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K | |
| aaO: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I | |
| ntO: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | ATT | GGT | GGT | CAT | GCT | GGG | GTT | ACT | ATT | |
| RtO: | 310 | 002 | 321 | 110 | 322 | 310 | 300 | 322 | 002 | 322 | 113 | 323 | 022 | 332 | 332 | 102 | 312 | 333 | 322 | 012 | 022 | |
| RtG: | 310 | 002 | 321 | 110 | 322 | 310 | 300 | 322 | 002 | 322 | 113 | 323 | 022 | 332 | 332 | 102 | 312 | 333 | 322 | 012 | 022 | |
| ntG: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | ATT | GGT | GGT | CAT | GCT | GGG | GTT | ACT | ATT | |
| aaG: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I | |
| aaO: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L | |
| ntO: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT | |
| RtO: | 121 | 112 | 121 | 221 | 212 | 103 | 310 | 012 | 112 | 100 | 311 | 001 | 223 | 210 | 032 | 301 | 020 | 122 | 011 | 311 | 122 | |
| RtG: | 121 | 112 | 121 | 221 | 212 | 103 | 310 | 012 | 112 | 100 | 311 | 001 | 223 | 210 | 032 | 301 | 020 | 122 | 011 | 311 | 122 | |
| ntG: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT | |
| aaG: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L | |
| aaO: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A | |
| ntO: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTC | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT | |
| RtO: | 012 | 003 | 132 | 011 | 100 | 302 | 330 | 332 | 010 | 300 | 321 | 323 | 303 | 310 | 000 | 310 | 330 | 000 | 332 | 210 | 312 | |
| RtG: | 012 | 003 | 132 | 011 | 100 | 302 | 330 | 332 | 013 | 300 | 321 | 323 | 303 | 310 | 000 | 310 | 330 | 000 | 332 | 210 | 312 | |
| ntG: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACG | GAA | GTC | GTC | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT | |
| aaG: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A | |
| aaO: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G | |
| ntO: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT | |
| RtO: | 010 | 223 | 211 | 023 | 311 | 202 | 311 | 330 | 310 | 223 | 221 | 312 | 302 | 310 | 231 | 223 | 000 | 330 | 121 | 001 | 332 | |
| RtG: | 010 | 223 | 211 | 023 | 311 | 202 | 311 | 330 | 310 | 223 | 221 | 312 | 302 | 310 | 231 | 223 | 000 | 330 | 121 | 001 | 332 | |
| ntG: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT | |
| aaG: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G | |
| aaO: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A | |
| ntO: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC | |
| RtO: | 322 | 110 | 302 | 321 | 020 | 300 | 231 | 210 | 201 | 323 | 100 | 212 | 010 | 021 | 011 | 303 | 122 | 112 | 221 | 222 | 311 | |
| RtG: | 322 | 110 | 302 | 321 | 020 | 300 | 231 | 210 | 201 | 323 | 100 | 212 | 010 | 021 | 011 | 303 | 122 | 112 | 221 | 222 | 311 | |
| ntG: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC | |
| aaG: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A | |
| aaO: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D | |
| ntO: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GTT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC | |
| RtO: | 213 | 003 | 323 | 033 | 223 | 333 | 003 | 002 | 332 | 323 | 303 | 303 | 322 | 122 | 301 | 223 | 330 | 110 | 121 | 210 | 301 | |
| RtG: | 213 | 003 | 323 | 033 | 223 | 333 | 003 | 002 | 332 | 323 | 303 | 303 | 322 | 122 | 301 | 223 | 330 | 110 | 121 | 210 | 301 | |
| ntG: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GTT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC | |
| aaG: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D | |
| aaO: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V | |
| ntO: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC | |
| RtO: | 222 | 303 | 003 | 300 | 331 | 223 | 300 | 310 | 223 | 003 | 110 | 300 | 121 | 003 | 211 | 211 | 020 | 300 | 003 | 330 | 321 | |
| RtG: | 222 | 303 | 003 | 300 | 331 | 223 | 300 | 310 | 223 | 003 | 110 | 300 | 121 | 003 | 211 | 211 | 020 | 300 | 003 | 330 | 321 | |
| ntG: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC | |
| aaG: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V | |
| aaO: | K | F | A | N | Q | | | | | | | | | | | | | | | | | |
| ntO: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | | |
| RtO: | 003 | 222 | 311 | 001 | 103 | | | | | | | | | | | | | | | | | |
| RtG: | 003 | 222 | 311 | 001 | 103 | | | | | | | | | | | | | | | | | |
| ntG: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | | |
| aaG: | K | F | A | N | Q | | | | | | | | | | | | | | | | | |

 Figura 4.7: Sequência da proteína com 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento A.

- A proteína reproduzida no rotulado B, o nucleotídeo na posição 142 foi alterado ocasionando a troca de aminoácido nesta posição, (I142S), porém, o nucleotídeo da posição 229 foi alterado, mas, ocasionando na troca de aminoácido nesta posição;

Seq.51 | *A. thaliana* - Mitocôndria - Malate dehydrogenase 1 – GI número 30695458

Código $Z_2 \times Z_2$ -linearidade((1023, 1013, 3), BCH primitivo sobre $GF(4, 10)$, rotulamento B)

$$p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 - g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$$

Caso 1 - (A, C, G, T) = (0, 1, 2, 3)

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | F | R | S | M | L | V | R | S | S | A | S | A | K | Q | A | V | I | R | R | S | |
| ntO: | ATG | TTC | AGA | TCT | ATG | CTC | GTC | CGA | TCT | TCT | GCC | TCC | GCG | AAG | CAG | GCG | GTT | ATC | CGC | CGT | AGC | |
| RtO: | 032 | 331 | 020 | 313 | 032 | 131 | 231 | 120 | 313 | 313 | 211 | 311 | 212 | 002 | 102 | 212 | 233 | 031 | 121 | 123 | 021 | |
| RtG: | 032 | 331 | 020 | 313 | 032 | 131 | 231 | 120 | 313 | 313 | 211 | 311 | 212 | 002 | 102 | 212 | 233 | 031 | 121 | 123 | 021 | |
| ntG: | ATG | TTC | AGA | TCT | ATG | CTC | GTC | CGA | TCT | TCT | GCC | TCC | GCG | AAG | CAG | GCG | GTT | ATC | CGC | CGT | AGC | |
| aaG: | M | F | R | S | M | L | V | R | S | S | A | S | A | K | Q | A | V | I | R | R | S | |
| aaO: | F | S | S | G | S | V | P | E | R | K | V | A | I | L | G | A | A | G | G | I | G | |
| ntO: | TTC | TCC | TCC | GGC | TCC | GTC | CCC | GAG | CGT | AAA | GTC | GCC | ATC | CTT | GGT | GCC | GCC | GGT | GGA | ATT | GGT | |
| RtO: | 331 | 311 | 311 | 221 | 311 | 231 | 111 | 202 | 123 | 000 | 231 | 211 | 031 | 133 | 223 | 211 | 211 | 223 | 220 | 033 | 223 | |
| RtG: | 331 | 311 | 311 | 221 | 311 | 231 | 111 | 202 | 123 | 000 | 231 | 211 | 031 | 133 | 223 | 211 | 211 | 223 | 220 | 033 | 223 | |
| ntG: | TTC | TCC | TCC | GGC | TCC | GTC | CCC | GAG | CGT | AAA | GTC | GCC | ATC | CTT | GGT | GCC | GCC | GGT | GGA | ATT | GGT | |
| aaG: | F | S | S | G | S | V | P | E | R | K | V | A | I | L | G | A | A | G | G | I | G | |
| aaO: | Q | P | L | A | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I | |
| ntO: | CAG | CCT | CTT | GCT | CTC | CTC | ATG | AAG | CTT | AAT | CCT | CTT | GTC | TCT | TCC | CTC | CTC | CTC | TAC | GAT | ATC | |
| RtO: | 102 | 113 | 133 | 213 | 131 | 032 | 002 | 133 | 003 | 113 | 133 | 231 | 313 | 311 | 131 | 311 | 131 | 311 | 301 | 203 | 031 | |
| RtG: | 102 | 113 | 133 | 213 | 131 | 032 | 002 | 133 | 003 | 113 | 133 | 231 | 313 | 311 | 131 | 311 | 131 | 311 | 301 | 203 | 031 | |
| ntG: | CAG | CCT | CTT | GCT | CTC | CTC | ATG | AAG | CTT | AAT | CCT | CTT | GTC | TCT | TCC | CTC | CTC | CTC | TAC | GAT | ATC | |
| aaG: | Q | P | L | A | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I | |
| aaO: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | E | V | V | G | |
| ntO: | GCT | AAC | ACT | CCT | GGA | GTT | GCT | GAT | GAT | GTT | GGT | CAC | ATC | AAC | ACC | AGA | TCT | GAG | GTT | GTT | GGG | |
| RtO: | 213 | 001 | 013 | 113 | 220 | 233 | 213 | 203 | 233 | 223 | 101 | 031 | 001 | 011 | 020 | 313 | 202 | 233 | 233 | 220 | | |
| RtG: | 213 | 001 | 013 | 113 | 220 | 233 | 213 | 203 | 233 | 223 | 101 | 031 | 001 | 011 | 020 | 313 | 202 | 233 | 233 | 220 | | |
| ntG: | GCT | AAC | ACT | CCT | GGA | GTT | GCT | GAT | GAT | GTT | GGT | CAC | ATC | AAC | ACC | AGA | TCT | GAG | GTT | GTT | GGG | |
| aaG: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | E | V | V | G | |
| aaO: | Y | M | G | D | D | N | L | A | K | A | L | E | G | A | D | L | V | I | I | P | A | |
| ntO: | TAC | ATG | GGC | GAT | GAT | AAC | TTG | GCC | AAA | GCT | CTT | GAA | GGA | GCT | GAT | CTC | GTT | ATC | ATT | CCA | GCT | |
| RtO: | 301 | 032 | 221 | 221 | 203 | 203 | 001 | 332 | 211 | 000 | 213 | 133 | 200 | 220 | 213 | 203 | 131 | 233 | 031 | 033 | 110 | 213 |
| RtG: | 301 | 032 | 221 | 221 | 203 | 203 | 001 | 332 | 211 | 000 | 213 | 133 | 200 | 220 | 213 | 203 | 131 | 233 | 031 | 033 | 110 | 213 |
| ntG: | TAC | ATG | GGC | GAT | GAT | AAC | TTG | GCC | AAA | GCT | CTT | GAA | GGA | GCT | GAT | CTC | GTT | ATC | ATT | CCA | GCT | |
| aaG: | Y | M | G | D | D | N | L | A | K | A | L | E | G | A | D | L | V | I | I | P | A | |
| aaO: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V | |
| ntO: | GGT | GTA | CCA | AGG | AAG | CCT | GGT | ATG | ACC | CGT | GAC | GAT | CTT | TTC | AAC | ATT | AAT | GCT | GGA | ATT | GTC | |
| RtO: | 223 | 230 | 110 | 022 | 002 | 113 | 223 | 032 | 011 | 123 | 201 | 203 | 133 | 331 | 001 | 033 | 003 | 213 | 220 | 033 | 231 | |
| RtG: | 223 | 230 | 110 | 022 | 002 | 113 | 223 | 032 | 011 | 123 | 201 | 203 | 133 | 331 | 001 | 033 | 003 | 213 | 220 | 033 | 231 | |
| ntG: | GGT | GTA | CCA | AGG | AAG | CCT | GGT | ATG | ACC | CGT | GAC | GAT | CTT | TTC | AAC | ATT | AAT | GCT | GGA | ATT | GTC | |
| aaG: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V | |
| aaO: | K | N | L | C | T | A | I | A | K | Y | C | P | H | A | L | I | N | M | I | S | AAC | |
| ntO: | AAG | AAC | CTT | TGC | ACT | GCC | ATC | GCC | AAG | TAC | TGC | CCA | CAT | GCG | CTT | ATT | AAT | ATG | ATC | AGC | AAC | |
| RtO: | 002 | 001 | 133 | 321 | 013 | 211 | 031 | 211 | 002 | 301 | 321 | 110 | 103 | 212 | 133 | 033 | 003 | 032 | 031 | 021 | 001 | |
| RtG: | 002 | 001 | 133 | 321 | 013 | 211 | 031 | 211 | 002 | 301 | 321 | 110 | 103 | 212 | 133 | 023 | 003 | 032 | 031 | 021 | 001 | |
| ntG: | AAG | AAC | CTT | TGC | ACT | GCC | ATC | GCC | AAG | TAC | TGC | CCA | CAT | GCG | CTT | ATT | AAT | ATG | ATC | AGC | AAC | |
| aaG: | K | N | L | C | T | A | I | A | K | Y | C | P | H | A | L | S | N | M | I | S | N | |
| aaO: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | M | Y | D | E | |
| ntO: | CCT | GTG | AAC | TCT | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATA | TTT | AAG | AAG | GCT | GTT | ATG | TAC | GAT | GAA | |
| RtO: | 113 | 232 | 001 | 313 | 013 | 233 | 110 | 033 | 210 | 213 | 202 | 030 | 333 | 002 | 002 | 213 | 223 | 032 | 301 | 203 | 200 | |
| RtG: | 113 | 232 | 001 | 313 | 013 | 233 | 110 | 033 | 210 | 213 | 202 | 030 | 333 | 002 | 002 | 213 | 223 | 032 | 301 | 203 | 200 | |
| ntG: | CCT | GTG | AAC | TCT | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATA | TTT | AAG | AAG | GCT | GTT | ATG | TAC | GAT | GAA | |
| aaG: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | M | Y | D | E | |

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| ntO: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| RtO: | 002 | 000 | 332 | 333 | 223 | 233 | 011 | 013 | 133 | 201 | 231 | 231 | 022 | 211 | 022 | 013 | 331 | 303 | 213 | 220 | 002 |
| RtG: | 002 | 000 | 332 | 333 | 223 | 233 | 011 | 013 | 133 | 201 | 231 | 231 | 022 | 211 | 022 | 013 | 331 | 303 | 213 | 220 | 002 |
| ntG: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| aaG: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| aaO: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| ntO: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTG | ATT | GGT | GGT | CAT | GCT | GGG | GTT | ACT | ATT |
| RtO: | 210 | 003 | 231 | 110 | 233 | 210 | 200 | 233 | 003 | 233 | 112 | 232 | 033 | 223 | 223 | 103 | 213 | 222 | 233 | 013 | 033 |
| RtG: | 210 | 003 | 231 | 110 | 233 | 210 | 200 | 233 | 003 | 233 | 112 | 232 | 033 | 223 | 223 | 103 | 213 | 222 | 233 | 013 | 033 |
| ntG: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTG | ATT | GGT | GGT | CAT | GCT | GGG | GTT | ACT | ATT |
| aaG: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| aaO: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| ntO: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT |
| RtO: | 131 | 113 | 131 | 331 | 313 | 102 | 210 | 013 | 113 | 100 | 211 | 001 | 332 | 310 | 023 | 201 | 030 | 133 | 011 | 211 | 133 |
| RtG: | 131 | 113 | 131 | 331 | 313 | 102 | 210 | 013 | 113 | 100 | 211 | 001 | 332 | 310 | 023 | 201 | 030 | 133 | 013 | 211 | 133 |
| ntG: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACT | GCC | CTT |
| aaG: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| aaO: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| ntO: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| RtO: | 013 | 002 | 123 | 011 | 100 | 203 | 220 | 223 | 010 | 200 | 231 | 232 | 202 | 210 | 000 | 210 | 220 | 000 | 223 | 310 | 213 |
| RtG: | 013 | 002 | 123 | 011 | 100 | 203 | 220 | 223 | 010 | 200 | 231 | 232 | 202 | 210 | 000 | 210 | 220 | 000 | 223 | 310 | 213 |
| ntG: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| aaG: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| aaO: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| ntO: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| RtO: | 010 | 332 | 311 | 032 | 211 | 303 | 211 | 220 | 210 | 332 | 331 | 213 | 203 | 210 | 321 | 332 | 000 | 220 | 131 | 001 | 223 |
| RtG: | 010 | 332 | 311 | 032 | 211 | 303 | 211 | 220 | 210 | 332 | 331 | 213 | 203 | 210 | 321 | 332 | 000 | 220 | 131 | 001 | 223 |
| ntG: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| aaG: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| aaO: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A |
| ntO: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC |
| RtO: | 233 | 110 | 203 | 231 | 030 | 200 | 321 | 310 | 301 | 232 | 100 | 313 | 010 | 031 | 011 | 202 | 133 | 113 | 331 | 333 | 211 |
| RtG: | 233 | 110 | 203 | 231 | 030 | 200 | 321 | 310 | 301 | 232 | 100 | 313 | 010 | 031 | 011 | 202 | 133 | 113 | 331 | 333 | 211 |
| ntG: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC |
| aaG: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A |
| aaO: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| ntO: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GGT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| RtO: | 312 | 002 | 232 | 022 | 332 | 222 | 002 | 003 | 223 | 232 | 202 | 202 | 233 | 133 | 201 | 332 | 220 | 110 | 131 | 310 | 201 |
| RtG: | 312 | 002 | 232 | 022 | 332 | 222 | 002 | 003 | 223 | 232 | 202 | 202 | 233 | 133 | 201 | 332 | 220 | 110 | 131 | 310 | 201 |
| ntG: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GGT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| aaG: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| aaO: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| ntO: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| RtO: | 333 | 202 | 002 | 200 | 221 | 332 | 200 | 210 | 332 | 002 | 110 | 200 | 131 | 002 | 311 | 311 | 030 | 200 | 002 | 220 | 231 |
| RtG: | 333 | 202 | 002 | 200 | 221 | 332 | 200 | 210 | 332 | 002 | 110 | 200 | 131 | 002 | 311 | 311 | 030 | 200 | 002 | 220 | 231 |
| ntG: | TTT | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| aaG: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| aaO: | K | F | A | N | Q | | | | | | | | | | | | | | | | |
| ntO: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| RtO: | 002 | 333 | 211 | 001 | 102 | | | | | | | | | | | | | | | | |
| RtG: | 002 | 333 | 211 | 001 | 102 | | | | | | | | | | | | | | | | |
| ntG: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| aaG: | K | F | A | N | Q | | | | | | | | | | | | | | | | |

 Figura 4.8: Sequência da proteína com 1023 nucleotídeos e $D(a, b) = 2$ no rotulamento B.

- A proteína reproduzida no rotulamento C trocou de nucleotídeo na posição 149 não alterando o aminoácido nesta posição, enquanto a troca de nucleotídeo na posição 289 alterou o aminoácido nesta posição, (L289F).

Seq.51| *A. thaliana* - Mitocôndria - Malate dehydrogenase 1 – GI número 30695458

| | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| Código klein-linearidade((1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento C) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| $p(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 - g(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$ | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Caso 03 - (A,C,G,T) = (0,2,1,3) | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M F R S M L V R S S A S A K Q A V I R R S ntO: ATG TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC Rto: 031 332 010 323 031 232 132 210 323 323 122 322 121 001 201 121 133 032 212 213 012 Rtg: 031 332 010 323 031 232 132 210 323 323 122 322 121 001 201 121 133 032 212 213 012 ntG: ATG TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC aaG: M F R S M L V R S S A S A K Q A V I R R S | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: F S S G S V P E R K V A I L G A A G G I G ntO: TTC TCC TCC GGC TCC GTC CCC GAG CGT AAA GTC GCC ATC CTT GGT GCC GCC GGT GGA ATT GGT Rto: 332 322 322 112 322 132 222 101 213 000 132 122 032 233 113 122 122 113 110 033 113 Rtg: 332 322 322 112 322 132 222 101 213 000 132 122 032 233 113 122 122 113 110 033 113 ntG: TTC TCC TCC GGC TCC GTC CCC GAG CGT AAA GTC GCC ATC CTT GGT GCC GCC GGT GGA ATT GGT aaG: F S S G S V P E R K V A I L G A A G G I G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: Q P L A L L M K L N P L V S S L S L Y D I ntO: CAG CCT CTT GCT CTC CTC ATG AAG CTT AAT CCT CTT GTC TCT TCC CTC TCC CTC TAC GAT ATC Rto: 201 223 233 123 232 232 031 001 233 003 223 233 132 323 322 232 322 232 302 103 032 Rtg: 201 223 233 123 232 232 031 001 233 003 223 233 132 323 322 232 322 232 302 103 032 ntG: CAG CCT CTT GCT CTC CTC ATG AAG CTT AAT CCT CTT GTC TCT TCC CTC TCC CTC TAC GAT ATC aaG: Q P L A L L M K L N P L V S S L S L Y D I | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: A N T P G V A A D V G H I N T R S E V V G ntO: GCT AAC ACT CCT GGA GTT GCT GAT GAT GTT GGT CAC ATC AAC ACC AGA TCT GAG GTT GTT GGA Rto: 123 002 023 223 110 133 123 123 103 133 113 202 032 002 022 010 323 101 133 133 110 Rtg: 123 002 023 223 110 133 123 123 103 133 113 202 032 002 022 010 323 101 133 133 110 ntG: GCT AAC ACT CCT GGA GTT GCT GAT GAT GTT GGT CAC ATC AAC ACC AGA TCT GAG GTT GTT GGA aaG: A N T P G V A A D V G H I N T R S E V V G | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Oaa: Y M G D D N L A K A L E G A D L V I I P A ntO: TAC ATG GGC GAT GAT AAC TTG GCC AAA GCT CTT GAA GGA GCT GAT CTC GTT ATC ATT CCA GCT Rto: 302 031 112 103 103 002 331 122 000 123 233 100 110 123 103 232 133 032 033 220 123 Rtg: 302 031 112 103 103 002 331 122 000 123 233 100 110 123 103 232 133 032 033 220 123 ntG: TAC ATG GGC GAT GAT AAC TTG GCC AAA GCT CTT GAA GGA GCT GAT CTC GTT ATC ATT CCA GCT aaG: Y M G D D N L A K A L E G A D L V I I P A | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: G V P R K P G M T R D D L F N I N A G I V ntO: GGT GTA CCA AGG AAG CCT GGT ATG ACC CGT GAC GAT CTT TCC AAC ATT AAT GCT GGA ATT GTC Rto: 113 130 220 011 001 223 113 031 022 213 102 103 233 332 002 033 003 123 110 033 132 Rtg: 113 130 220 011 001 223 113 031 022 213 102 103 233 332 002 033 003 123 110 033 132 ntG: GGT GTA CCA AGG AAG CCT GGT ATG ACC CGT GAC GAT CTT TCC AAC ATT AAT GCT GGA ATT GTC aaG: G V P R K P G M T R D D L F N I N A G I V | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: K N L C T A I A K Y C P H A L I N M I S N ntO: AAG AAC CTT TGC ACT GCC ATC GCC AAG TAC TGC CCA CAT GCG CTT ATT AAT ATG ATC AGC AAC Rto: 001 002 233 312 023 122 032 122 001 302 312 220 203 121 233 033 003 031 032 012 002 Rtg: 001 002 233 312 023 122 032 122 001 302 312 220 203 121 233 033 003 031 032 012 002 ntG: AAG AAC CTT TGC ACT GCC ATC GCC AAG TAC TGC CCA CAT GCG CTT ATT AAT ATG ATC AGC AAC aaG: K N L C T A I A K Y C P H A L I N M I S N | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: P V N S T V P I A A E I F K K A G M Y D E ntO: CCT GTG AAC TCT ACT GTT CCA ATT GCA GCT GAG ATA TTT AAG AAG GCT GGT ATG TAC GAT GAA Rto: 223 131 002 323 023 133 220 033 120 123 101 030 333 001 001 123 113 031 302 103 100 Rtg: 223 131 002 323 023 133 220 033 120 123 101 030 333 001 001 123 113 031 302 103 100 ntG: CCT GTG AAC TCT ACT GTT CCA ATT GCA GCT GAG ATA TTT AAG AAG GCT GGT ATG TAC GAT GAA aaG: P V N S T V P I A A E I F K K A G M Y D E | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| ntO: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| RtO: | 001 | 000 | 331 | 333 | 113 | 133 | 022 | 023 | 233 | 102 | 132 | 132 | 011 | 122 | 011 | 023 | 332 | 303 | 123 | 110 | 001 |
| RtG: | 001 | 000 | 331 | 333 | 113 | 133 | 022 | 023 | 233 | 102 | 132 | 132 | 011 | 122 | 011 | 023 | 332 | 303 | 123 | 110 | 001 |
| ntG: | AAG | AAA | TTG | TTT | GGT | GTT | ACC | ACT | CTT | GAC | GTC | GTC | AGG | GCC | AGG | ACT | TTC | TAT | GCT | GGA | AAG |
| aaG: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| aaO: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| ntO: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | AGT | GCT | GGG | GTT | ACT | ATT | | | |
| RtO: | 120 | 003 | 132 | 220 | 133 | 120 | 100 | 133 | 003 | 133 | 221 | 131 | 033 | 113 | 113 | 203 | 123 | 111 | 133 | 023 | 033 |
| RtG: | 120 | 003 | 132 | 220 | 133 | 120 | 100 | 133 | 003 | 133 | 221 | 131 | 033 | 113 | 113 | 203 | 123 | 111 | 133 | 023 | 033 |
| ntG: | GCA | AAT | GTC | CCA | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GTC | AGT | CAT | GCT | GGG | GTT | ACT | ATT | | |
| aaG: | A | N | V | P | V | A | E | V | N | V | P | V | I | G | G | H | A | G | V | T | I |
| aaO: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| ntO: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT |
| RtO: | 232 | 223 | 232 | 332 | 323 | 201 | 120 | 023 | 223 | 200 | 122 | 002 | 331 | 320 | 013 | 102 | 030 | 233 | 022 | 122 | 233 |
| RtG: | 232 | 223 | 232 | 332 | 323 | 201 | 120 | 023 | 223 | 200 | 122 | 002 | 331 | 320 | 013 | 102 | 030 | 233 | 022 | 122 | 233 |
| ntG: | CTC | CCT | CTC | TTC | TCT | CAG | GCA | ACT | CCT | CAA | GCC | AAC | TTG | TCA | AGT | GAC | ATA | CTT | ACC | GCC | CTT |
| aaG: | L | P | L | F | S | Q | A | T | P | Q | A | N | L | S | S | D | I | L | T | A | L |
| aaO: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| ntO: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| RtO: | 023 | 001 | 213 | 022 | 200 | 103 | 110 | 113 | 020 | 100 | 132 | 131 | 101 | 120 | 000 | 120 | 110 | 000 | 113 | 320 | 123 |
| RtG: | 023 | 001 | 213 | 022 | 200 | 103 | 110 | 113 | 020 | 100 | 132 | 131 | 101 | 120 | 000 | 120 | 110 | 000 | 113 | 320 | 123 |
| ntG: | ACT | AAG | CGT | ACC | CAA | GAT | GGA | GGT | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCA | GGA | AAA | GGT | TCA | GCT |
| aaG: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| aaO: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| ntO: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| RtO: | 020 | 331 | 322 | 031 | 122 | 303 | 122 | 110 | 120 | 331 | 322 | 123 | 103 | 120 | 312 | 331 | 000 | 110 | 232 | 002 | 113 |
| RtG: | 020 | 331 | 322 | 031 | 122 | 303 | 122 | 110 | 120 | 331 | 322 | 123 | 103 | 120 | 312 | 331 | 000 | 110 | 232 | 002 | 113 |
| ntG: | ACA | TTG | TCC | ATG | GCC | TAT | GCC | GGA | GCA | TTG | TTC | GCT | GAT | GCA | TGC | TTG | AAA | GGA | CTC | AAC | GGT |
| aaG: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| aaO: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A |
| ntO: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | CTT | CCT | TTC | TTT | GCC |
| RtO: | 133 | 220 | 103 | 132 | 030 | 100 | 312 | 320 | 302 | 131 | 200 | 323 | 020 | 032 | 022 | 101 | 233 | 223 | 332 | 333 | 122 |
| RtG: | 133 | 220 | 103 | 132 | 030 | 100 | 312 | 320 | 302 | 131 | 200 | 323 | 020 | 032 | 022 | 101 | 333 | 223 | 332 | 333 | 122 |
| ntG: | GTT | CCA | GAT | GTC | ATA | GAA | TGC | TCA | TAC | GTG | CAA | TCT | ACA | ATC | ACC | GAG | TTT | CCT | TTC | TTT | GCC |
| aaG: | V | P | D | V | I | E | C | S | Y | V | Q | S | T | I | T | E | F | P | F | F | A |
| aaO: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| ntO: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| RtO: | 321 | 001 | 131 | 011 | 331 | 111 | 001 | 003 | 113 | 131 | 101 | 101 | 133 | 233 | 102 | 331 | 110 | 220 | 232 | 320 | 102 |
| RtG: | 321 | 001 | 131 | 011 | 331 | 111 | 001 | 003 | 113 | 131 | 101 | 101 | 133 | 233 | 102 | 331 | 110 | 220 | 232 | 320 | 102 |
| ntG: | TCG | AAG | GTG | AGG | TTG | GGG | AAG | AAT | GGT | GTG | GAG | GAG | GT | CTT | GAC | TTG | GGA | CCA | CTC | TCA | GAC |
| aaG: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| aaO: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| ntO: | TTT | GAG | AAG | GAA | GCG | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| RtO: | 333 | 101 | 001 | 100 | 112 | 331 | 100 | 120 | 331 | 001 | 220 | 100 | 232 | 001 | 322 | 322 | 030 | 100 | 001 | 110 | 132 |
| RtG: | 333 | 101 | 001 | 100 | 112 | 331 | 100 | 120 | 331 | 001 | 220 | 100 | 232 | 001 | 322 | 322 | 030 | 100 | 001 | 110 | 132 |
| ntG: | TTT | GAG | AAG | GAA | GCG | TTG | GAA | GCA | TTG | AAG | CCA | GAA | CTC | AAG | TCC | TCC | ATA | GAA | AAG | GGA | GTC |
| aaG: | F | E | K | E | G | L | E | A | L | K | P | E | L | K | S | S | I | E | K | G | V |
| aaO: | K | F | A | N | Q | | | | | | | | | | | | | | | | |
| ntO: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| RtO: | 001 | 333 | 122 | 002 | 201 | | | | | | | | | | | | | | | | |
| RtG: | 001 | 333 | 122 | 002 | 201 | | | | | | | | | | | | | | | | |
| ntG: | AAG | TTT | GCC | AAC | CAG | | | | | | | | | | | | | | | | |
| aaG: | K | F | A | N | Q | | | | | | | | | | | | | | | | |

 Figura 4.9: Sequência da proteína com 1023 nucleotídeos e $D(a, b) = 2$ no rotulamento C.

4.2.2 Códigos BCH não primitivos sobre $GR(4, r)$: Exemplos

Desenvolvemos agora a construção de códigos BCH não primitivos sobre o anel \mathbb{Z}_4 .

A diferença entre os códigos BCH primitivos e os códigos BCH não primitivos está relacionada aos comprimentos n das palavras-código dos códigos BCH. Quando o comprimento da sequência de DNA (palavra-código) for igual a $n = (2^r - 1)$, essa sequência será analisada através dos códigos BCH primitivos. Quando o comprimento da sequência de DNA for um submúltiplo de n , essa sequência poderá ser identificada e reproduzida através dos **códigos BCH não primitivos**. A principal alteração acontece no **Passo 6** do algoritmo, na determinação do valor de d : os demais passos são determinados de maneira análoga ao procedimento adotado nos exemplos da construção de códigos BCH primitivos sobre anel \mathbb{Z}_4 , apresentados na Subseção 4.2.1.

A seguir, mostramos alguns exemplos da construção de códigos BCH não primitivos sobre $GR(4, r)$.

Código BCH $(n, k, d_H) = (21, k, d_H)$ sobre $GR(4, 6)$

Considere a construção do código BCH não primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (21, 15, 3)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = 21$ nucleotídeos.

Passo 2 - Determinar a extensão de Galois - Seja $r = 6$, pois $n = (2^6 - 1) = (2^6 - 1) = 63$, e portanto $n = 21$ é submúltiplo de 63.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Neste exemplo, usaremos novamente o polinômio primitivo $p_1(x) = x^6 + x + 1$, ilustrado na Tabela 4.14, para realizarmos a extensão do corpo.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - O corpo $GF(2^r) = GF(2^6) = GF(64) = F_{64}$, desenvolvido na expressão (4.2), apresenta os elementos dados na Tabela 4.15, sendo α um elemento primitivo deste corpo.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Seja o anel $GR(4, 6)$ gerado pela extensão de grau 6, desenvolvido na expressão (4.3). Os elementos não nulos e inversíveis do grupo cíclico do grupo $GR^*(4, 6)$ mostrados na Tabela 4.16.

Passo 6 - Determinar o subgrupo das unidades - Do **Passo 5** resulta que f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 6)$, onde $d \geq 1 \in \mathbb{Z}$ e f^d gera o subgrupo cíclico cuja ordem é 63 em $GR^*(4, 6)$. Sendo assim, temos que $n \cdot d = 21 \cdot d = 126$ implicando que

$d = 6$. Consequentemente, $f^6 = (330000) = \alpha^6$ gera um subgrupo cíclico de ordem 21 em $GR^*(4, 6)$. Logo, $\beta = \alpha^6$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{21}$, ver Tabela 4.33. Esse elemento primitivo será utilizado na construção do código BCH não primitivo de comprimento $n = 21$ sobre \mathbb{Z}_4 .

| $G_{21} \rightarrow (\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ | $G_{21} \rightarrow (\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5)$ |
|--|--|
| $(f^6)^1 = f^6 = \alpha^6 = \beta \rightarrow (330000)$ | $(f^{72})^{12} = f^{72} = \alpha^{72} = \beta^{12} \rightarrow (331232)$ |
| $(f^{12})^2 = f^{12} = \alpha^{12} = \beta^2 \rightarrow (121000)$ | $(f^{78})^{13} = f^{78} = \alpha^{78} = \beta^{13} \rightarrow (331232)$ |
| $(f^{18})^3 = f^{18} = \alpha^{18} = \beta^3 \rightarrow (311300)$ | $(f^{84})^{14} = f^{84} = \alpha^{84} = \beta^{14} \rightarrow (331232)$ |
| $(f^{24})^4 = f^{24} = \alpha^{24} = \beta^4 \rightarrow (102010)$ | $(f^{90})^{15} = f^{90} = \alpha^{90} = \beta^{15} \rightarrow (331232)$ |
| $(f^{30})^5 = f^{30} = \alpha^{30} = \beta^5 \rightarrow (332233)$ | $(f^{96})^{16} = f^{96} = \alpha^{96} = \beta^{16} \rightarrow (331232)$ |
| $(f^{36})^6 = f^{36} = \alpha^{36} = \beta^6 \rightarrow (013032)$ | $(f^{102})^{17} = f^{102} = \alpha^{102} = \beta^{17} \rightarrow (331232)$ |
| $(f^{42})^7 = f^{42} = \alpha^{42} = \beta^7 \rightarrow (210113)$ | $(f^{108})^{18} = f^{108} = \alpha^{108} = \beta^{18} \rightarrow (331232)$ |
| $(f^{48})^8 = f^{48} = \alpha^{48} = \beta^8 \rightarrow (103320)$ | $(f^{114})^{19} = f^{114} = \alpha^{114} = \beta^{19} \rightarrow (331232)$ |
| $(f^{54})^9 = f^{54} = \alpha^{54} = \beta^9 \rightarrow (331232)$ | $(f^{120})^{20} = f^{120} = \alpha^{120} = \beta^{20} \rightarrow (331232)$ |
| $(f^{60})^{10} = f^{60} = \alpha^{60} = \beta^{10} \rightarrow (300133)$ | $(f^{126})^{21} = f^{126} = \alpha^{126} = \beta^{21} \rightarrow (100000)$ |
| $(f^{66})^{11} = f^{66} = \alpha^{66} = \beta^{11} \rightarrow (000302)$ | |

Tabela 4.33: Elementos de G_{21}

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Para a distância mínima de projeto do código BCH não primitivo sobre \mathbb{Z}_4 igual a 3, o polinômio gerador $g(x)$ tem como raízes β e β^2 . Assim, temos:

$$M_1(x) = M_2(x) = (x - \beta)(x - \beta^2) = 1 + x + 3x^2 + 3x^4 + 2x^5 + x^6$$

Dessa forma, $g(x) = mmc\{M_1(x), M_2(x)\} = 1 + x + 3x^2 + 3x^4 + 2x^5 + x^6$ gera o código BCH desejado sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (21, 15, 3)$.

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H é dado por:

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{21} - 1}{x^6 + 2x^5 + 3x^4 + 3x^2 + x + 1}$$

$$h(x) = x^{15} + 2x^{14} + x^{13} + 2x^{11} + x^{10} + 2x^9 + 2x^8 + 3x^7 + 3x^6 + 2x^4 + 3x^3 + 2x^2 + x + 3$$

onde os coeficientes do polinômio $h(x)$ pertencem a \mathbb{Z}_4 .

Passo 9 - Determinar a matriz G e a sua transposta G^T - Realizando os deslocamentos dos coeficientes do polinômio $g(x)$ da esquerda para a direita, obtemos a matriz G com dimensão 15×21 :

$$G = \begin{pmatrix} 1130321000000000000000 \\ 0113032100000000000000 \\ 0011303210000000000000 \\ 0001130321000000000000 \\ 0000113032100000000000 \\ 0000011303210000000000 \\ 0000001130321000000000 \\ 0000000113032100000000 \\ 0000000011303210000000 \\ 0000000001130321000000 \\ 0000000000113032100000 \\ 0000000000011303210000 \\ 0000000000001130321000 \\ 0000000000000113032100 \\ 0000000000000011303210 \\ 0000000000000001130321 \end{pmatrix}.$$

A matriz G^T com dimensão 21×15 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - A matriz H com dimensão 6×21 :

$$H = \begin{pmatrix} 000001210212233023213 \\ 000012102122330232130 \\ 000121021223302321300 \\ 001210212233023213000 \\ 012102122330232130000 \\ 121021223302321300000 \end{pmatrix}.$$

A matriz H^T com dimensão 21×6 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, vamos verificar se o código BCH não primitivo sobre anel é capaz de reproduzir a sequência de DNA de *miRNA* (Seq.59) - GI número 240153948 com 21 nucleotídeos. Esta sequência deve ser rotulada de acordo com as 24 permutações apresentadas na Tabela 4.20, análogamente aos exemplos dos códigos BCH primitivos, resultando na matriz P . Seja a Seq.59 do NCBI igual {5'-AGGGCCCCCCTCAATCCTGT-3'}, então obtemos

$$P = \left(\begin{array}{c} 02221111113100311323 \\ 03331111112100211232 \\ 011122222223200322313 \\ 033322222221200122131 \\ 022233333331300133121 \\ 011133333332300233212 \\ 122200000003011300323 \\ 133300000002011200232 \\ 10002222223211322303 \\ 13332222220211022030 \\ 100033333332311233202 \\ 12223333330311033020 \\ 211100000003022300313 \\ 233300000001022100131 \\ 20001111113122311303 \\ 23331111110122011030 \\ 200033333331322133101 \\ 211133333330322033010 \\ 311100000002033200212 \\ 322200000001033100121 \\ 30001111112133211202 \\ 32221111110133011020 \\ 30002222221233122101 \\ 31112222220233022010 \end{array} \right).$$

Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões de erros estabelecidos: $D(\mathbf{a}, \mathbf{b}) = 0$, $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ - Novamente, como no caso do BCH primitivo, vamos analisar se as sequências (sem diferença, com 1 ou com 2 nucleotídeos de diferença) são palavras-código dos códigos (n, k, d_H) usando a equação $\mathbf{v} \cdot \mathbf{H}^T = 0$, da seguinte maneira:

a) $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Vamos analisar se a sequência de DNA, Seq.59, pode ser reproduzidas com até 1 nucleotídeo de diferença, considerando as 3 outras possibilidades de nucleotídeos em cada posição na sequência para cada permutação. As palavras-código encontradas devem ser armazenadas.

a) $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Vamos analisar se a sequência, Seq.59, pode ser reproduzidas com até 2 nucleotídeos de diferença em cada permutação, considerando todas as combinações simples tomados 2 a 2 dos n nucleotídeos de comprimento da sequência. As palavras-código encontradas também devem ser armazenadas.

Como os demais passos do algoritmo são análogos aos exemplos apresentados na cons-

trução dos códigos BCH primitivos, podemos ir direto aos resultados desse algoritmo.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Como resultado das 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 1$ citada no **Passo 12** do algoritmo, não encontramos palavras-código nos 24 casos de permutações com até 1 dígito de diferença. Portanto, a sequência de DNA, Seq.59, não foi identificada e reproduzida através do código BCH não primitivo com até 1 nucleotídeo de diferença.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

| Sequência de DNA do NCBI: {5'- A G GGCCCCCCCCTCAATCCTG T -3'} | | | |
|---|------------------------------|-------------------------------|---|
| Seq.59 <i>H.sapiens</i> - miRNA WO2008043521 - GI número 240153948 | | | |
| 24 casos de permutações | Palavras-código $n = 21$ | Quantidade de palavras-código | Sequência de DNA reproduzida |
| Caso 01-(A,C,G,T)=(0,1,2,3) | 00000 ... 00000 | 0 | - |
| Caso 02-(A,C,G,T)=(0,1,3,2) | 00000 ... 00000 | 0 | - |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 02112222223200322311 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 02332222221200122133 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 05-(A,C,G,T)=(0,3,2,1) | 00000 ... 00000 | 0 | - |
| Caso 06-(A,C,G,T)=(0,3,1,2) | 00000 ... 00000 | 0 | - |
| Caso 07-(A,C,G,T)=(1,0,2,3) | 00000 ... 00000 | 0 | - |
| Caso 08-(A,C,G,T)=(1,0,3,2) | 00000 ... 00000 | 0 | - |
| Caso 09-(A,C,G,T)=(1,2,0,3) | 00000 ... 00000 | 0 | - |
| Caso 10-(A,C,G,T)=(1,2,3,0) | 00000 ... 00000 | 0 | - |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 130033333332311233200 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 13223333330311033022 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 201100000003022300311 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 203300000001022100133 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 15-(A,C,G,T)=(2,1,0,3) | 00000 ... 00000 | 0 | - |
| Caso 16-(A,C,G,T)=(2,1,3,0) | 00000 ... 00000 | 0 | - |
| Caso 17-(A,C,G,T)=(2,3,0,1) | 00000 ... 00000 | 0 | - |
| Caso 18-(A,C,G,T)=(2,3,1,0) | 00000 ... 00000 | 0 | - |
| Caso 19-(A,C,G,T)=(3,0,1,2) | 00000 ... 00000 | 0 | - |
| Caso 20-(A,C,G,T)=(3,0,2,1) | 00000 ... 00000 | 0 | - |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 31001111112133211200 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 31221111110133011022 | 1 | ACGGCCCCCCCCTCAATCCTGG |
| Caso 23-(A,C,G,T)=(3,2,0,1) | 00000 ... 00000 | 0 | - |
| Caso 24-(A,C,G,T)=(3,2,1,0) | 00000 ... 00000 | 0 | - |

Tabela 4.34: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 21$ nucleotídeos.

O resultado obtido na reprodução da sequência Seq.59 através dos códigos BCH não primitivos para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença, foi o mesmo resultado obtido nas análises de sequências de DNA identificadas através dos códigos BCH primitivos para $D(\mathbf{a}, \mathbf{b}) = 1$.

nucleotídeo de diferença. Obtemos 8 palavras-código correspondente aos 8 casos de permutações, sendo os casos 03, 04, 11, 12, 13, 14, 21 e 22. Essas palavras-código são diferentes em termos do alfabeto do código, $\mathbb{Z}_4 = \{0, 1, 2, 3\}$, porém são iguais quando rotuladas usando o alfabeto do código genético $N = \{A, C, G, T\}$ resultando em uma única sequência de DNA, ver Tabela 4.34.

Veja na Figura 4.10, a sequência de *miRNA*, Seq.59, reproduzida pelo código G -linearidade: Klein-linearidade ((21, 15, 3) BCH não primitivo sobre $GR(4, 6)$, rotulamento C) através do polinômio primitivo $p_1(x) = x^6 + x + 1$ e do polinômios gerador $g_1(x) = x^6 + 2x^5 + 3x^4 + 3x^2 + x + 1$ com $D(a, b) = 2$ nucleotídeos de diferença da sequência do NCBI. Como essa sequência de DNA não fabrica proteínas (aminoácidos) apenas os nucleotídeos são reproduzidos.

Seq.59| *H.Sapiens* - miRNA - WO2008043521 - GI número 240153948

| |
|--|
| Código Klein-linearidade((21,15,3) BCH não primitivo sobre GR(4,6), rotulamento C) |
| $p_1(x)=x^6+x+1 - g_1(x)=x^6+2x^5+3x^4+3x^2+x+1 - \text{Caso } 03-(A,C,G,T)=(0,2,1,3)$ |
| nt0: AGG GCC CCC CCT CAA TCC TGT Rto: 011 122 222 223 200 322 313 RtG: 021 122 222 223 200 322 311 ntG: ACG GCC CCC CCT CAA TCC TGG |

Figura 4.10: Sequência de DNA de *miRNA* com 21 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C.

Código BCH $(n, k, d_H) = (51, k, d_H)$ sobre $GR(4, 8)$

Considere agora a construção do código BCH não primitivo sobre a estrutura de anel com parâmetros $(n, k, d) = (51, 43, 3)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = 51$.

Passo 2 - Determinar a extensão de Galois - Como $n = 51$, o grau r do polinômio primitivo a ser usado a extensão de Galois do corpo $GF(2)$ é 8, pois $n = (2^r - 1) = 2^8 - 1 = 255$, e, 51 é submúltiplo de 255.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Neste exemplo, vamos usar o polinômio primitivo $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$, mostrado na Tabela 4.23, para estender o corpo $GF(2^r)$ e o anel $GR(4, r)$.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - Considere o corpo de Galois $GF(2^r) = GF(2^8) = GF(256) = F_{256}$ dado por

$$\frac{F_2[x]}{\langle p_{05}(x) \rangle} \cong \frac{F_2[x]}{\langle x^8 + x^6 + x^5 + x^2 + 1 \rangle} = \{a_0 + a_1x + a_2x^2 + \cdots + a_7x^7 : a_i \text{ 's } \in F_2\},$$

Seja α um elemento primitivo em F_{256} , equivalentemente, α é uma raiz de $x^8 + x^6 + x^5 + x^2 + 1 = 0$, ou seja, $\alpha^8 + \alpha^6 + \alpha^5 + \alpha^2 + 1 = 0$ implicando em $\alpha^8 = -\alpha^6 - \alpha^5 - \alpha^2 - 1$. Assim, obtemos $\alpha^8 = \alpha^6 + \alpha^5 + \alpha^2 + 1$. A Tabela 4.35 mostra os elementos de F_{256} .

| Elementos de F_{256} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|---|
| 0 | (00000000) |
| 1 | (10000000) |
| α | (01000000) |
| α^2 | (00100000) |
| α^3 | (00010000) |
| α^4 | (00001000) |
| α^5 | (00000100) |
| α^6 | (00000010) |
| α^7 | (00000001) |
| α^8 | (10100110) |
| $\alpha^9 = \alpha \times \alpha^8$ | (01010011) |
| $\alpha^{10} = \alpha \times \alpha^9$ | (10001111) |
| \vdots | \vdots |
| $\alpha^{253} = \alpha \times \alpha^{252}$ | (10011010) |
| $\alpha^{254} = \alpha \times \alpha^{253}$ | (01001101) |
| $\alpha^{255} = \alpha \times \alpha^{254}$ | (10000000) |

Tabela 4.35: Elementos de F_{256} com $p_{05}(x)$.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Considere o anel $GR(4, 8)$ como sendo dado pelo quociente do anel $\mathbb{Z}_4[x]$ pelo ideal gerado pelo $p_{05}(x)$, isto é,

$$\frac{\mathbb{Z}_4[x]}{\langle p_{05}(x) \rangle} \cong \frac{\mathbb{Z}_4[x]}{\langle x^8 + x^6 + x^5 + x^2 + 1 \rangle} = \{b_0 + b_1x + b_2x^2 + \cdots + b_7x^7 : b_i \text{ 's } \in \mathbb{Z}_4\}$$

As operações em $GR^*(4, 8)$ são realizadas módulo $(x^8 + x^6 + x^5 + x^2 + 1)$. Logo, $\alpha^8 = 3\alpha^6 + 3\alpha^5 + 3\alpha^2 + 3$, pois os coeficientes de $GR^*(4, 8)$ estão em $\mathbb{Z}_4[x]$. Assim, considerando $f = (01000000) = \alpha$, todos os elementos do grupo cíclico do grupo $GR^*(4, 8)$ são determinados, como mostra a Tabela 4.36.

Passo 6 - Determinar o subgrupo das unidades - Sabemos que f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 8)$, onde $d \geq 1 \in \mathbb{Z}$ e f^d gera o subgrupo cíclico cuja ordem é 255

| $GR^*(4, 8)$ | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|---|
| 1 | (10000000) |
| $f = x = \alpha$ | (01000000) |
| $f^2 = x^2 = \alpha^2$ | (00100000) |
| $f^3 = x^3 = \alpha^3$ | (00010000) |
| $f^4 = x^4 = \alpha^4$ | (00001000) |
| $f^5 = x^5 = \alpha^5$ | (00000100) |
| $f^6 = x^6 = \alpha^6$ | (00000010) |
| $f' = x' = \alpha'$ | (00000001) |
| $f^8 = x^8 = \alpha^8$ | (30300330) |
| $f^9 = f \times f^8 = x^9 = \alpha^9$ | (03030033) |
| $f^{10} = f \times f^9 = x^{10} = \alpha^{10}$ | (10003113) |
| : | : |
| $f^{508} = f \times f^{507} = x^{508} = \alpha^{508}$ | (30033030) |
| $f^{509} = f \times f^{508} = x^{509} = \alpha^{509}$ | (03003303) |
| $f^{510} = f \times f^{509} = x^{510} = \alpha^{510}$ | (10000000) |

 Tabela 4.36: Elementos do grupo cíclico do grupo $GR^*(4, 8)$ com $p_{05}(x)$.

em $GR^*(4, 8)$. Portanto, $n \cdot d = 51 \cdot d = 510$ implicando que $d = 10$ consequentemente, $f^{10} = (10003113) = \alpha^{10}$ gera um subgrupo cíclico de ordem 51 em $GR^*(4, 8)$. Logo, $\beta = \alpha^{10}$ é o elemento primitivo que gera o subgrupo cíclico $G_n = G_{51}$ mostrado na Tabela 4.37.

| $G_{255} \rightarrow (\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7)$ |
|---|
| $(f^{10})^1 = f^{10} = \alpha^{10} = \beta \rightarrow (10003113)$ |
| $(f^{10})^2 = f^{20} = \alpha^{20} = \beta^2 \rightarrow (11101302)$ |
| $(f^{10})^3 = f^{30} = \alpha^{30} = \beta^3 \rightarrow (11002120)$ |
| $(f^{10})^4 = f^{40} = \alpha^{40} = \beta^4 \rightarrow (10202023)$ |
| $(f^{10})^5 = f^{50} = \alpha^{50} = \beta^5 \rightarrow (03202023)$ |
| : |
| $(f^{10})^{48} = f^{480} = \alpha^{480} = \beta^{48} \rightarrow (03002010)$ |
| $(f^{10})^{49} = f^{490} = \alpha^{490} = \beta^{49} \rightarrow (22130002)$ |
| $(f^{10})^{50} = f^{500} = \alpha^{500} = \beta^{50} \rightarrow (32313123)$ |
| $(f^{10})^{51} = f^{510} = \alpha^{510} = \beta^{51} \rightarrow (10000000)$ |

 Tabela 4.37: Elementos de G_{51}

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Considerando que a distância mínima do código seja $d_H \leq 3$, o polinômio gerador $g(x)$ do código tem como raízes β e β^2 . Os polinômios minimais de β e β^2 são dados por:

$$M_1(x) = M_2(x) = (x - \beta)(x - \beta^2) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1.$$

Portanto, o polinômio gerador deste código é dado por $g(x) = mmc\{M_1(x), M_2(x)\}$ gera o código desejado e está relacionado com a matriz geradora G do código BCH sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (51, 43, 3)$.

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H é obtido, sendo

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{51} - 1}{x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1}$$

$$h(x) = x^{43} + x^{42} + 3x^{41} + x^{40} + 2x^{39} + 2x^{38} + x^{37} + 3x^{36} + x^{35} + 2x^{34} + 3x^{33} + 2x^{32} + 3x^{31} + 2x^{30} + 2x^{29} + 3x^{28} + 2x^{26} + 2x^{25} + 2x^{24} + 3x^{22} + 2x^{21} + x^{19} + 2x^{18} + 3x^{17} + x^{16} + 2x^{15} + x^{14} + 2x^{13} + 2x^{12} + x^{11} + x^9 + x^8 + 3x^7 + 3x^6 + 3x^5 + x + 3.$$

Passo 9 - Determinar a matriz G e a sua transposta G^T - Através do polinômio gerador $g(x)$ a matriz G com dimensão 43×51 é obtida:

A matriz G^T com dimensão 51×43 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - A matriz H com dimensão 8×51 é dada por:

$$H = \begin{pmatrix} 00000011312213123232230222032312312122101133300013 \\ 00000113122131232322302220323123121221011333000130 \\ 000001131221312323223022203231231212210113330001300 \\ 000011312213123232230222032312312122101133300013000 \\ 000113122131232322302220323123121221011333000130000 \\ 001131221312323223022203231231212210113330001300000 \\ 011312213123232230222032312312122101133300013000000 \\ 113122131232322302220323123121221011333000130000000 \end{pmatrix}.$$

A matriz H^T com dimensão 51×8 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, vamos verificar se o código BCH não primitivo sobre anel é capaz de reproduzir a SD (Seq.11) do organismo: *Influenza A virus* - hemagglutinin[A/Czechoslovakia/2/1988/H1N1] - GI número: 305125 com comprimento $n = 51$ nucleotídeos. Esta sequência deve ser rotulada de acordo com as 24 permutações apresentadas na Tabela 4.20 análoga aos exemplos anteriores.

Passo 12 - Verificar se a sequência de DNA é palavra-código de acordo com os padrões

de erros estabelecidos: $D(\mathbf{a}, \mathbf{b}) = 0$, $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ - Novamente, como nos exemplos anteriores, vamos analisar se as sequências (sem diferença, com 1 e com 2 nucleotídeos de diferença) são palavras-código dos códigos (n, k, d) usando a relação (4.1) para cada uma das 24 permutações mencionadas no **Passo 11**.

Os demais passos do algoritmo são determinados de maneira análoga aos exemplos anteriores.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Para $D(\mathbf{a}, \mathbf{b}) = 1$, não encontramos palavras-código nos 24 casos de permutações com até 1 dígito de diferença. Portanto, a SD, Seq.11, com 51 nucleotídeos não foi identificada e reproduzida através do código BCH não primitivo com até um nucleotídeo de diferença.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

| Sequência de DNA do NCBI Seq.59 <i>H.sapiens</i> - miRNA WO2008043521 - GI número 240153948 SD={5'- ATGAAAGCAAAACTA C TAGCCTGTTATGTGCA T TTACAGCTACAGATGCA -3'} | | |
|---|-----------------------------|--|
| 24 casos de permutações | Palavras Código $n = 51$ | Sequência de DNA reproduzida |
| 01-(A,C,G,T)=(0,1,2,3) | 0000 ... 0000 | - |
| 02-(A,C,G,T)=(0,1,3,2) | 0000 ... 0000 | - |
| 03-(A,C,G,T)=(0,2,1,3) | 0310 ... 3120 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 04-(A,C,G,T)=(0,2,3,1) | 0130 ... 1320 | - |
| 05-(A,C,G,T)=(0,3,2,1) | 0000 ... 0000 | - |
| 06-(A,C,G,T)=(0,3,1,2) | 0000 ... 0000 | - |
| 07-(A,C,G,T)=(1,0,2,3) | 0000 ... 0000 | - |
| 08-(A,C,G,T)=(1,0,3,2) | 0000 ... 0000 | - |
| 09-(A,C,G,T)=(1,2,0,3) | 0000 ... 0000 | - |
| 10-(A,C,G,T)=(1,2,3,0) | 0000 ... 0000 | - |
| 11-(A,C,G,T)=(1,3,0,2) | 1201 ... 2031 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 12-(A,C,G,T)=(1,3,2,0) | 1021 ... 0231 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 13-(A,C,G,T)=(2,0,1,3) | 2312 ... 3102 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 14-(A,C,G,T)=(2,0,3,1) | 2132 ... 1302 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 15-(A,C,G,T)=(2,1,0,3) | 0000 ... 0000 | - |
| 16-(A,C,G,T)=(2,1,3,0) | 0000 ... 0000 | - |
| 17-(A,C,G,T)=(2,3,0,1) | 0000 ... 0000 | - |
| 18-(A,C,G,T)=(2,3,1,0) | 0000 ... 0000 | - |
| 19-(A,C,G,T)=(3,0,1,2) | 0000 ... 0000 | - |
| 20-(A,C,G,T)=(3,0,2,1) | 0000 ... 0000 | - |
| 21-(A,C,G,T)=(3,1,0,2) | 3203 ... 2013 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 22-(A,C,G,T)=(3,1,2,0) | 3023 ... 0213 | ATGAAAGCAAAACTA G TAGCCTGTTATGTGCA C TTACAGCTACAGATGCA |
| 23-(A,C,G,T)=(3,2,0,1) | 0000 ... 0000 | - |
| 24-(A,C,G,T)=(3,2,1,0) | 0000 ... 0000 | - |

Tabela 4.38: As 24 permutações e a quantidade de palavras-código para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 51$.

O resultado obtido na reprodução SD, Seq.11, para $D(\mathbf{a}, \mathbf{b}) = 2$, foi o mesmo resultado obtido no exemplo do código (21, 15, 3) sobre $GR(4, 6)$. Obtemos 8 palavras-código correspondentes aos 8 casos de permutações do rotulamento C, Tabela 4.38.

A Figura 5.7, mostra a SD do vírus *Influenza A*, Seq.59, reproduzida pelo código G -linearidade: Klein-linearidade ((51, 43, 3) BCH não primitivo sobre $GR(4, 8)$, rotulamento C) através do polinômio primitivo $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ e do polinômio gerador $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ com dois nucleotídeos de diferença da sequência do NCBI. As trocas de nucleotídeos nas posições das trincas 6 e 12, ocasionaram nas trocas dos aminoácidos nestas posições, (L6V) e (F12L), respectivamente.

Seq.11| Influenza A - hemagglutinin H1N1 (1998) - vírus - GI número 305125

| Código klein-linearidade((51, 43, 3) BCH não primitivo sobre GR(4, 8), rotulamento C) | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1 - g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1 - \text{Caso 3} - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 2, 1, 3)$ | | | | | | | | | | | | | | | | | | | |
| aaO: M K A K L L V L L C A F T A T D A ntO: ATG AAA GCA AAA CTA CTA GTC CTG TTA TGT GCA TTT ACA GCT ACA GAT GCA RtO: 031 000 120 000 230 230 132 231 330 313 120 333 020 123 020 103 120 RtG: 031 000 120 000 230 130 132 231 330 313 120 233 020 123 020 103 120 ntG: ATG AAA GCA AAA CTA GTA GTC CTG TTA TGT GCA CTT ACA GCT ACA GAT GCA aaG: M K A K L V V L L C A L T A T D A | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | | | | | | |

Figura 4.11: Sequência de DNA de sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$ no rotulamento C.

Código BCH $(n, k, d_H) = (93, k, d_H)$ sobre $GR(4, 10)$

Considere ainda a construção do código BCH primitivo sobre a estrutura de anel com parâmetros $(n, k, d_H) = (93, 83, 3)$ capaz de gerar e reproduzir sequências de DNA com comprimentos $n = 93$ nucleotídeos.

Passo 2 - Determinar a extensão de Galois - O grau da extensão de Galois do corpo $GF(2)$ é $r = 10$, pois $n = (2^r - 1) = (2^{10} - 1) = 1023$ e, 93 é um submúltiplo de 1023.

Passo 3 - Determinar todos os polinômios primitivos $p(x)$, relacionados à extensão de Galois - Neste exemplo, vamos usar o polinômio primitivo $p_{01}(x) = x^{10} + x^3 + 1$ relacionados à extensão de Galois de grau $r = 10$.

Passo 4 - Determinar a extensão do corpo $GF(2)$ - Considere o corpo $GF(2^r) = GF(2^{10}) = GF(1024) = F_{1024}$ dado por

$$\frac{F_2[x]}{\langle p_{01}(x) \rangle} \cong \frac{F_2[x]}{\langle x^{10} + x^3 + 1 \rangle} = \{a_0 + a_1x + a_2x^2 + \dots + a_9x^9 : a_i \in F_2\}.$$

Assim, F_{1024} é formado pelos elementos contidos na Tabela 4.39.

| Elementos de F_{1024} | $(\alpha^0 \alpha^1 \alpha^2 \alpha^3 \alpha^4 \alpha^5 \alpha^6 \alpha^7 \alpha^8 \alpha^9)$ |
|---|---|
| 0 | (0000000000) |
| 1 | (1000000000) |
| α | (0100000000) |
| α^2 | (0010000000) |
| α^3 | (0001000000) |
| α^4 | (0000100000) |
| α^5 | (0000010000) |
| α^6 | (0000001000) |
| α^7 | (0000000100) |
| α^8 | (0000000010) |
| α^9 | (0000000001) |
| α^{10} | (1001000001) |
| $\alpha^{11} = \alpha \times \alpha^{10}$ | (?) |
| \vdots | \vdots |
| $\alpha^{1020} = \alpha \times \alpha^{1019}$ | (1000000100) |
| $\alpha^{1021} = \alpha \times \alpha^{1020}$ | (0100000010) |
| $\alpha^{1022} = \alpha \times \alpha^{1021}$ | (0010000001) |
| $\alpha^{1023} = \alpha \times \alpha^{1022}$ | (1000000000) |

Tabela 4.39: Elementos de F_{1024} com $p_{01}(x)$.

Passo 5 - Determinar a extensão do anel \mathbb{Z}_4 - Considere ainda o anel $GR(4, r) = GR(4, 10)$ dado pela extensão de Galois de grau $r = 10$ abaixo:

$$GR(4, 10)[x] \cong \frac{\mathbb{Z}_4[x]}{\langle p_{01}(x) \rangle} \cong \frac{\mathbb{Z}_4[x]}{\langle x^{10} + x^3 + 1 \rangle} = \{b_0 + b_1x + b_2x^2 + \cdots + b_9x^9 : b_i \in \mathbb{Z}_4\}.$$

As operações em $GR^*(4, 10)$ são realizadas módulo $(x^{10} + x^3 + 1)$. Logo, $\alpha^{10} = 3\alpha^3 + 3$, pois os coeficientes de $GR^*(4, 10)$ estão em \mathbb{Z}_4 . Assim, considerando $f = (01000000) = \alpha$, os elementos do grupo cíclico do grupo $GR^*(4, 10)$ são dados na Tabela 4.40.

Passo 6 - Determinar o subgrupo das unidades - Como f gera um grupo cíclico de ordem $n \cdot d$ em $GR^*(4, 10)$, e f^d gera o subgrupo cíclico cuja ordem é 1023 em $GR^*(4, 10)$, então $n \cdot d = 93 \cdot d = 2046$ implicando que $d = 22$. Consequentemente, $f^{22} = (001002001) = \alpha^2$ gera um subgrupo cíclico de ordem 93 em $GR^*(4, 10)$. Logo, $\beta = \alpha^2$ é o elemento primitivo em G_{93} , na Tabela 4.41.

Passo 7 - Determinar o polinômio gerador da matriz G , $g(x)$ - Podemos agora construir um código BCH de comprimento $n = 93$ sobre \mathbb{Z}_4 . Considerando que a distância mínima do código seja $d_H \geq 3$, o polinômio gerador $g(x)$ do código tem como raízes β e β^2 . Os polinômios minimais de β e β^2 são dados por: $M_1(x) = M_2(x) =$

$$\{(x - \beta)(x - \beta^2)(x - \beta^4)(x - \beta^8)(x - \beta^{16})(x - \beta^{32})(x - \beta^{64})(x - \beta^{35})(x - \beta^{70})(x - \beta^{47})\} =$$

| $GR^*(4, 10)$ | $(\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5\alpha^6\alpha^7)\alpha^8\alpha^9$ |
|---|--|
| 1 | (1000000000) |
| $f = x = \alpha$ | (0100000000) |
| $f^2 = x^2 = \alpha^2$ | (0010000000) |
| $f^3 = x^3 = \alpha^3$ | (0001000000) |
| $f^4 = x^4 = \alpha^4$ | (0000100000) |
| $f^5 = x^5 = \alpha^5$ | (0000010000) |
| $f^6 = x^6 = \alpha^6$ | (0000001000) |
| $f^7 = x^7 = \alpha^7$ | (0000000100) |
| $f^8 = x^8 = \alpha^8$ | (0000000010) |
| $f^9 = x^9 = \alpha^9$ | (0000000001) |
| $f^{10} = x^{10} = \alpha^{10}$ | (3003000000) |
| $f^{11} = f \times f^{10} = x^{11} = \alpha^{11}$ | (0300300000) |
| \vdots | \vdots |
| $f^{2044} = f \times f^{2043} = x^{2044} = \alpha^{2044}$ | (0300000030) |
| $f^{2045} = f \times f^{2044} = x^{2045} = \alpha^{2045}$ | (0030000003) |
| $f^{2046} = f \times f^{2045} = x^{2046} = \alpha^{2046}$ | (1000000000) |

 Tabela 4.40: Elementos do grupo cíclico do grupo $GR^*(4, 10)$ com $p_{01}(x)$

| |
|---|
| $G_{93} \rightarrow (\alpha^0\alpha^1\alpha^2\alpha^3\alpha^4\alpha^5\alpha^6\alpha^7\alpha^8\alpha^9)$ |
| $(f^{22})^1 = f^{22} = \alpha^{22} = \beta \rightarrow (001002001)$ |
| $(f^{22})^2 = f^{44} = \alpha^{44} = \beta^2 \rightarrow (2002103003)$ |
| $(f^{22})^3 = f^{66} = \alpha^{66} = \beta^3 \rightarrow (3113011012)$ |
| $(f^{22})^4 = f^{88} = \alpha^{88} = \beta^4 \rightarrow (213011202)$ |
| \vdots |
| $(f^{22})^{91} = f^{2002} = \alpha^{2002} = \beta^{91} \rightarrow (1322012021)$ |
| $(f^{22})^{92} = f^{2024} = \alpha^{2024} = \beta^{92} \rightarrow (3111303031)$ |
| $(f^{22})^{93} = f^{2046} = \alpha^{2046} = \beta^{93} \rightarrow (1000000000)$ |

 Tabela 4.41: Elementos de G_{93}

$$= 1 + 2x + 3x^2 + 2x^3 + x^4 + x^5 + 2x^6 + 2x^7 + x^{10}.$$

O polinômio gerador deste código é dado por $g(x) = mmc\{M_1(x), M_2(x)\}$, sendo $g(x) = 1 + 2x + 3x^2 + 2x^3 + x^4 + x^5 + 2x^6 + 2x^7 + x^{10}$, e gera o código BCH desejado sobre \mathbb{Z}_4 com parâmetros $(n, k, d_H) = (93, 83, 3)$.

Passo 8 - Determinar o polinômio gerador da matriz H , $h(x)$ - O polinômio gerador da matriz verificação de paridade H , é dado por:

$$h(x) = \frac{x^n - 1}{g(x)} = \frac{x^{93} - 1}{x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1}$$

$$h(x) = x^{83} + 2x^{80} + 2x^{79} + 3x^{78} + 3x^{77} + \dots + 3x^6 + 3x^5 + 2x^3x + 3x^2 + 2x + 3,$$

onde os coeficientes do polinômio $h(x)$ pertencem a \mathbb{Z}_4 .

Passo 9 - Determinar a matriz G e a sua transposta G^T - O polinômio gerador $g(x)$ está

relacionado à matriz geradora G com dimensão 83×93 , isto é,

$$G = \begin{pmatrix} 12321122001000000000000000 \cdots 00000000000000000000000000000000 \\ 012321122001000000000000 \cdots 00000000000000000000000000000000 \\ \vdots & \ddots & \vdots \\ 00000000000000000000000000000000 \cdots 00000000000000000000000000000000123211220010 \\ 00000000000000000000000000000000 \cdots 0000000000000000000000000000000012321122001 \end{pmatrix}.$$

A matriz G^T com dimensão 93×83 é determinada como sendo a troca da linha pela coluna.

Passo 10 - Determinar a matriz H e a sua transposta H^T - A matriz verificação de paridade geradora do código dual associado é dada pela matriz H com dimensão 10×93 , sendo

$$H = \begin{pmatrix} 00000000100223321202100323000100023023223021321302023223121200002102121100123102221123302323 \\ 000000001002233212021003230001000230232230213213020232231212000021021211001231022211233023230 \\ 000000010022332120210032300010002302322302132130202322312120000210212110012310222112330232300 \\ 000000100223321202100323000100023023223021321302023223121200002102121100123102221123302323000 \\ 000001002233212021003230001000230232230213213020232231212000021021211001231022211233023230000 \\ 000010022332120210032300010002302322302132130202322312120000210212110012310222112330232300000 \\ 000100223321202100323000100023023223021321302023223121200002102121100123102221123302323000000 \\ 001002233212021003230001000230232230213213020232231212000021021211001231022211233023230000000 \\ 010022332120210032300010002302322302132130202322312120000210212110012310222112330232300000000 \\ 100223321202100323000100023023223021321302023223121200002102121100123102221123302323000000000 \end{pmatrix}.$$

A matriz H^T com dimensão 93×10 é determinada pela troca da linha pela coluna.

Passo 11 - Rotular a sequência de DNA utilizando o Passo 1 - Neste exemplo, analisaremos se o código BCH sobre anel é capaz de reproduzir a SD (Seq.33) do organismo: *S.tuberosum - Viridiplantae* - GI número 1707998 com comprimento $n = 93$ nucleotídeos. Essa SD deve ser rotulada de acordo com as 24 permutações, analogamente aos exemplos anteriores.

Os demais passos do algoritmo são análogos aos passos dos exemplos anteriores.

Resultados:

Para $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença

Assim como nos exemplos dos códigos $(21, 15, 6)$ sobre $GR(4, 6)$ e $(51, 43, 3)$ sobre $(GR(4, 8))$, nenhuma palavra-código com até $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença da SD do NCBI foi encontrada.

Para $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença

Para $D(\mathbf{a}, \mathbf{b}) = 2$, também obtemos 3 palavras-código distintas para cada um dos 8 casos correspondente ao rotulamento C (Tabela 4.42), que, quando rotuladas em termos do alfabeto do código genético correspondem a 3 SD's distintas reproduzidas no rotulamento C. Veja Figura na 4.12, as 3 possibilidades da SD, Seq.33, com $n = 93$ nucleotídeos reproduzidas pelo código G -linearidade: Klein-linearidade ((93,83,3) BCH primitivo sobre $\text{GR}(4,10)$, rotulamento C), através do polinômio primitivo $p_{01}(x) = x^{10} + x^3 + 1$ e do polinômio gerador $g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos diferindo da SD do NCBI. Observe na primeira SD da Figura 4.12 que os nucleotídeos nas posições das trincas 3 e 17 foram alterados, ocasionando a troca dos aminoácidos nestas posições, (A4E) e (V18F), respectivamente. Na segunda SD, note que o nucleotídeo na posição 7 foi alterado não ocasionando a troca de aminoácido nesta posição, enquanto que, o nucleotídeo da posição 26 foi alterado ocasionando na troca de aminoácido nesta posição, (S26Y). O mesmo aconteceu com a terceira SD, o nucleotídeo da posição 14 foi alterado não alterando o aminoácido nesta posição, porém, o nucleotídeo da posição 16 foi alterado,(K16Q).

| Sequência de DNA do NCBI Seq.33 <i>S.tuberosum</i> - Serine methylase - GI número 1707998 | |
|--|-------------------------------|
| 24 Permutações | Quantidade de palavras-código |
| Caso 01-(A,C,G,T)=(0,1,2,3) | - |
| Caso 02-(A,C,G,T)=(0,1,3,2) | - |
| Caso 03-(A,C,G,T)=(0,2,1,3) | 3 |
| Caso 04-(A,C,G,T)=(0,2,3,1) | 3 |
| Caso 05-(A,C,G,T)=(0,3,2,1) | - |
| Caso 06-(A,C,G,T)=(0,3,1,2) | - |
| Caso 07-(A,C,G,T)=(1,0,2,3) | - |
| Caso 08-(A,C,G,T)=(1,0,3,2) | - |
| Caso 09-(A,C,G,T)=(1,2,0,3) | - |
| Caso 10-(A,C,G,T)=(1,2,3,0) | - |
| Caso 11-(A,C,G,T)=(1,3,0,2) | 3 |
| Caso 12-(A,C,G,T)=(1,3,2,0) | 3 |
| Caso 13-(A,C,G,T)=(2,0,1,3) | 3 |
| Caso 14-(A,C,G,T)=(2,0,3,1) | 3 |
| Caso 15-(A,C,G,T)=(2,1,0,3) | - |
| Caso 16-(A,C,G,T)=(2,1,3,0) | - |
| Caso 17-(A,C,G,T)=(2,3,0,1) | - |
| Caso 18-(A,C,G,T)=(2,3,1,0) | - |
| Caso 19-(A,C,G,T)=(3,0,1,2) | - |
| Caso 20-(A,C,G,T)=(3,0,2,1) | - |
| Caso 21-(A,C,G,T)=(3,1,0,2) | 3 |
| Caso 22-(A,C,G,T)=(3,1,2,0) | 3 |
| Caso 23-(A,C,G,T)=(3,2,0,1) | - |
| Caso 24-(A,C,G,T)=(3,2,1,0) | - |

Tabela 4.42: As 24 permutações para $D(\mathbf{a}, \mathbf{b}) = 2$ e $n = 93$ nucleotídeos.

Seq.33 | S.tuberosum – Midocôndria – Serine methylase – GI número 1707998
Código Klein-linearidade((93,83,3) BCH primitivo sobre GR(4,10), rotulamento C)

$$p_{01}(x) = x^{10} + x^3 + 1 - g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1 - \text{Caso 3-}(A, C, G, T) = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|
| aaO: | M | A | M | A | I | A | L | R | R | L | S | A | T | V | D | K | P | V | K | S | L | Y |
| ntO: | ATG | GCT | ATG | GCA | ATA | GCT | CTT | CGG | AGG | CTT | TCC | GCT | ACA | GTT | GAC | AAA | CCG | GTT | AAG | AGT | CTC | TAC |
| RtO: | 031 | 123 | 031 | 120 | 030 | 123 | 233 | 211 | 011 | 233 | 322 | 123 | 020 | 133 | 102 | 000 | 221 | 133 | 001 | 013 | 232 | 302 |
| RtG: | 031 | 123 | 031 | 100 | 030 | 123 | 233 | 211 | 011 | 233 | 322 | 123 | 020 | 133 | 102 | 000 | 221 | 333 | 001 | 013 | 232 | 302 |
| ntG: | ATG | GCT | ATG | GAA | ATA | GCT | CTT | CGG | AGG | CTT | TCC | GCT | ACA | GTT | GAC | AAA | CCG | TTT | AAG | AGT | CTC | TAC |
| aaG: | M | A | M | E | I | A | L | R | R | L | S | A | T | V | D | K | P | F | K | S | L | Y |
| aaO: | N | G | G | S | L | Y | Y | M | S | | | | | | | | | | | | | |
| ntO: | AAT | GGT | GGC | TCT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | |
| RtO: | 003 | 113 | 112 | 323 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | |
| RtG: | 003 | 113 | 112 | 323 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | |
| ntG: | AAT | GGT | GGC | TCT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | |
| aaG: | N | G | G | S | L | Y | Y | M | S | | | | | | | | | | | | | |

Código Klein-linearidade((93,83,3) BCH primitivo sobre GR(4,10), rotulamento C)

$$p_{01}(x) = x^{10} + x^3 + 1 - g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1 - \text{Caso 3-}(A, C, G, T) = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|------------|-----|-----|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | A | M | A | I | A | L | R | R | L | S | A | T | V | D | K | P | V | K | S | L | Y |
| ntO: | ATG | GCT | ATG | GCA | ATA | GCT | CTT | CGG | AGG | CTT | TCC | GCT | ACA | GTT | GAC | AAA | CCG | GTT | AAG | AGT | CTC | TAC |
| RtO: | 031 | 123 | 031 | 120 | 030 | 123 | 233 | 211 | 011 | 233 | 322 | 123 | 020 | 133 | 102 | 000 | 221 | 133 | 001 | 013 | 232 | 302 |
| RtG: | 031 | 123 | 031 | 120 | 030 | 123 | 231 | 211 | 011 | 233 | 322 | 123 | 020 | 133 | 102 | 000 | 221 | 133 | 001 | 013 | 232 | 302 |
| ntG: | ATG | GCT | ATG | GCA | ATA | GCT | CTG | CGG | AGG | CTT | TCC | GCT | ACA | GTT | GAC | AAA | CCG | GTT | AAG | AGT | CTC | TAC |
| aaG: | M | A | M | A | I | A | L | R | R | L | S | A | T | V | D | K | P | V | K | S | L | Y |
| aaO: | N | G | G | S | L | Y | Y | M | S | | | | | | | | | | | | | |
| ntO: | AAT | GGT | GGC | TCT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | |
| RtO: | 003 | 113 | 112 | 323 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | |
| RtG: | 003 | 113 | 112 | 303 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | |
| ntG: | AAT | GGT | GGC | TAT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | |
| aaG: | N | G | G | Y | L | Y | Y | M | S | | | | | | | | | | | | | |

Código Klein-linearidade((93,83,3) BCH primitivo sobre GR(4,10), rotulamento C)

$$p_{01}(x) = x^{10} + x^3 + 1 - g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1 - \text{Caso 3-}(A, C, G, T) = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|----------|------------|-----|-----|-----|-----|-----|-----|
| aaO: | M | A | M | A | I | A | L | R | R | L | S | A | T | V | D | K | P | V | K | S | L | Y | |
| ntO: | ATG | GCT | ATG | GCA | ATA | GCT | CTT | CGG | AGG | CTT | TCC | GCT | ACA | GTT | GAC | AAA | CCG | GTT | AAG | AGT | CTC | TAC | |
| RtO: | 031 | 123 | 031 | 120 | 030 | 123 | 233 | 211 | 011 | 233 | 322 | 123 | 020 | 133 | 102 | 000 | 221 | 133 | 001 | 013 | 232 | 302 | |
| RtG: | 031 | 123 | 031 | 120 | 030 | 123 | 231 | 211 | 011 | 233 | 322 | 123 | 020 | 131 | 102 | 200 | 221 | 133 | 001 | 013 | 232 | 302 | |
| ntG: | ATG | GCT | ATG | GCA | ATA | GCT | CTT | CGG | AGG | CTT | TCC | GCT | ACA | GTT | G | GAC | CAA | CCG | GTT | AAG | AGT | CTC | TAC |
| aaG: | M | A | M | A | I | A | L | R | R | L | S | A | T | V | D | Q | P | V | K | S | L | Y | |
| aaO: | N | G | G | S | L | Y | Y | M | S | | | | | | | | | | | | | | |
| ntO: | AAT | GGT | GGC | TCT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | | |
| RtO: | 003 | 113 | 112 | 323 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | | |
| RtG: | 003 | 113 | 112 | 323 | 232 | 303 | 302 | 031 | 320 | | | | | | | | | | | | | | |
| ntG: | AAT | GGT | GGC | TCT | CTC | TAT | TAC | ATG | TCA | | | | | | | | | | | | | | |
| aaG: | N | G | G | S | L | Y | Y | M | S | | | | | | | | | | | | | | |

Figura 4.12: Três possibilidades da sequência de direcionamento com 93 nucleotídeos reproduzida com $D(a, b) = 2$ no rotulamento C.

Capítulo 5

Análise dos Resultados

Neste capítulo, fazemos uma análise dos resultados e das observações relevantes que foram obtidos no decorrer deste trabalho. Analisamos um total de 92 sequências de DNA através dos CCEs, sendo sequências com os seguintes comprimentos: 21, 39, 45, 51, 63, 93, 105, 195, 255, 511, 1023 e 2047 nucleotídeos e as seguintes características biológicas: sequências de direcionamento, sequências de direcionamento ambígua, enzima, sinal interno, hormônio, ítron, DNA repetitivo, *miRNA*, proteínas de vírus e de bactérias (células procariontes), proteínas (células eucariontes), gene e genoma procariótico. De maneira resumida, mostramos todas as sequências de DNA que foram identificadas e reproduzidas através dos códigos *G*-linearidade: (BCH sobre o corpo $GF(4^r)$ e sobre o anel $GR(4, r)$). Com isso, podemos afirmar que essas sequências de DNA possuem uma estrutura matemática, e portanto, podem ser identificadas, reproduzidas e classificadas através códigos CCEs.

Até onde é de nosso conhecimento, pela primeira vez diferentes sequências em termos da fita simples do DNA e da dupla hélice do DNA - incluindo gene e genoma, são identificadas e reproduzidas através dos CCEs. Com isso, apontamos na direção da existência de CCEs na estrutura do DNA levantadas em [8], [12], [13], [14], [15], [16] e [19], e ampliamos significativamente os resultados obtidos em [20] para sequências de direcionamento com 63 nucleotídeos. Outro resultado relevante neste trabalho, é mostrarmos a existência de códigos concatenados entre algumas sequências de direcionamento e suas respectivas proteínas organelares, conhecidos como “*nested codes*”. Em [1], Battail sugere fortemente a existência de códigos concatenados no genoma, porém não mostra exemplos de sequências de DNA identificadas e reproduzidas por códigos concatenados.

Este capítulo está organizado da seguinte forma: A Seção 5.1 inicia com a apresentação da Tabela 5.1, ilustramos informações importantes e necessárias sobre as sequências de DNA reproduzidas neste trabalho. Na Subseção 5.1.1, mostramos que apenas a distância $d_H = 3$ dos códigos *G*-linearidade (tanto a estrutura de corpo quanto a estrutura de anel) foi capaz

de identificar e reproduzir as sequências da fita simples do DNA e da dupla hélice do DNA. Nas Seções 5.2 e 5.3, fazemos uma análise das sequências de DNA identificadas no $GF(4^r)$ e no $GR(4, r)$, respectivamente. Tanto no corpo quanto no anel, apontamos o rotulamento dessas sequências e, portanto, a sua classificação matemática, e a sua representação algébrica via polinômios primitivos/geradores das fitas codantes e polinômios recíprocos das fitas não codantes das sequências. Mostramos também a relação entre os polinômios primitivos e os polinômios geradores responsáveis pela geração desses códigos. No caso do $GF(4^r)$, em virtude da complexidade computacional envolvida nos cálculos, fizemos todas as análises somente para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$. Já no anel, fizemos as análises para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$ e para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$, o que nos levou à importantes observações, tais como: as classes de polinômios primitivos e polinômios geradores, que serão discutidas na Subseção 5.3.1; e através das análises em anel para $D(\mathbf{a}, \mathbf{b}) = 2$ determinarmos se uma sequência possui ou não a estrutura de corpo, vista na Subseção 5.3.4. Na Seção 5.4.2, chamamos a atenção para as sequências de DNA reproduzidas neste trabalho envolvidas com a patologia do câncer. A reprodução das sequências completas da fita simples do DNA e da dupla hélice do DNA de um gene e de um genoma é mostrada na Seção 5.5. A Seção 5.6 considera algumas sequências de direcionamento e suas respectivas proteínas reproduzidas por códigos concatenados, o “*nested codes*”. Finalmente na Seção 5.7, fazemos algumas observações da relação entre o $GF(4^r)$ e o $GR(4, r)$ levantadas no processo de construção desses códigos.

5.1 Sequências de DNA Reproduzidas pelos Códigos G -linearidade

A pesquisa foi realizada através das análises das sequências de DNA, sendo noventa e duas (92) sequências de vários organismos, espécies e comprimentos postadas no banco de dados do *NCBI* através do número do GI. A Tabela 5.1 apresenta informações relevantes com relação às sessenta e cinco (65) sequências que foram reproduzidas matematicamente através dos códigos G -linearidade (BCH sobre o corpo $GF(4^r)$ e BCH sobre o anel $GR(4, r)$), sendo 17 sequências diferindo em 1 nucleotídeo das sequências do *NCBI* e 65 sequências diferindo em 2 nucleotídeos das sequências do *NCBI*.

Por uma questão de minimização de espaço na Tabela 5.1, indicamos por \checkmark as sequências de DNA que foram reproduzidas com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$ (um nucleotídeo de diferença das sequências do *NCBI*) e indicamos por $\checkmark\checkmark$ as sequências reproduzidas com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$ (dois nucleotídeos de diferença das sequências do *NCBI*). As sequências

de DNA indicadas por ✓* são as sequências de direcionamento contendo 63 nucleotídeos que foram reproduzidas em [20] através dos códigos *G*-linearidade (BCH primitivo sobre o anel $GR(4, 6)$) diferindo em 1 nucleotídeo das sequências do *NCBI*.

| Seq. nº | Sequências de DNA | GI número | Organismo | Espécie | Comprimentos em nt e aa | Código BCH sobre Anéis | Código BCH sobre Corpos |
|------------|----------------------|--------------|------------------------|----------------------|----------------------------|---------------------------|----------------------------|
| 01 | SD - M | 30686795 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 54 nt - 18 aa | ✓✓ | - |
| 02 | SD - M | 438130 | <i>S.tuberosum</i> | <i>Viridiplantae</i> | 54 nt - 18 aa | ✓✓ | - |
| 03 | SD - RE | 298590 | <i>H.sapiens</i> | <i>Metazoa</i> | 51 nt - 17 aa | ✓✓ | - |
| 04 | SD - M | 114685 | <i>S.cerevisiae</i> | <i>Fungi</i> | 51 nt - 17 aa | ✓✓ | - |
| 05 | SD - M | 114152885 | <i>S.cerevisiae</i> | <i>Fungi</i> | 51 nt - 17 aa | ✓✓ | - |
| 06 | SD - M | 2815500 | <i>X.laevis</i> | <i>Metazoa</i> | 51 nt - 17 aa | ✓✓ | - |
| 07 | SD - M | 548791 | <i>S.cerevisiae</i> | <i>Fungi</i> | 54 nt - 18 aa | ✓✓ | - |
| 08 | SD - M | 547901 | <i>S.cerevisiae</i> | <i>Fungi</i> | 51 nt - 17 aa | ✓✓ | - |
| 09 | SD - M | 417647 | <i>S.cerevisiae</i> | <i>Fungi</i> | 51 nt - 17 aa | ✓✓ | - |
| 10 | SD - M | 126733 | <i>H.sapiens</i> | <i>Metazoa</i> | 54 nt - 18 aa | ✓✓ | - |
| 11 | SD | 305125 | <i>Influenza A</i> | <i>Virus</i> | 51 nt - 17 aa | ✓✓ | - |
| 12 | SD - M | 899225 | <i>B.napus</i> | <i>Viridiplantae</i> | 66 nt - 22 aa | ✓* e ✓✓ | - |
| 13 | SD - M | 217937 | <i>I.batatas</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓* e ✓✓ | ✓ |
| 14 | SD - RE | 186509758 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓* e ✓✓ | - |
| 15 | SD - RE | 632733 | <i>N.tabacum</i> | <i>Viridiplantae</i> | 66 nt - 22 aa | ✓* e ✓✓ | - |
| 16 | SD - RE | 1808650 | <i>H.vulgare</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓✓ | - |
| 17 | SD - RE | 78096542 | <i>T.sativum</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓* e ✓✓ | ✓ |
| 18 | SD - C | 21227 | <i>Soleracea</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓✓ | - |
| 19 | SD - M | 45269853 | <i>S.cerevisiae</i> | <i>Fungi</i> | 63 nt - 21 aa | ✓* e ✓✓ | ✓ |
| 20 | SD - M | 163770 | <i>B.taurus</i> | <i>Metazoa</i> | 66 nt - 22 aa | ✓✓ | - |
| 21 | SD - M | 148878474 | <i>B.taurus</i> | <i>Metazoa</i> | 66 nt - 22 aa | ✓✓ | - |
| 22 | SD - M | 497233 | <i>G.max</i> | <i>Viridiplantae</i> | 66 nt - 22 aa | ✓✓ | - |
| 23 | SD - C | 30912740 | <i>C.sinensis</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓✓ | - |
| 24 | SD - M | 30695458 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 66 nt - 22 aa | ✓* e ✓✓ | - |
| 25 | SD - M | 145338521 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 66 nt - 22 aa | ✓✓ | - |
| 26 | SD - M | 1352036 | <i>R.norvegicus</i> | <i>Metazoa</i> | 66 nt - 22 aa | ✓✓ | - |
| 27 | SD - M | 2493051 | <i>S.cerevisiae</i> | <i>Fungi</i> | 66 nt - 22 aa | ✓✓ | - |
| 28 | SD - M | 12587 | <i>H.sapiens</i> | <i>Metazoa</i> | 66 nt - 22 aa | ✓* e ✓✓ | - |
| 29 | SD - RE | 16740522 | <i>M.martensii</i> | <i>Metazoa</i> | 63 nt - 21 aa | ✓* e ✓✓ | - |
| 30 | SD - RE | 536792 | <i>P.vulgaris</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓✓ | - |
| 31 | SD - RE | 51093376 | <i>P.dominulus</i> | <i>Metazoa</i> | 63 nt - 21 aa | ✓* e ✓✓ | ✓ |
| 32 | SD - M | 773581 | <i>P.sativum</i> | <i>Viridiplantae</i> | 93 nt - 31 aa | ✓✓ | - |
| 33 | SD - M | 1707998 | <i>S.tuberosum</i> | <i>Viridiplantae</i> | 93 nt - 31 aa | ✓✓ | - |
| 34 | SD Ambígua | 38017095 | <i>N.Benthamiana</i> | <i>Viridiplantae</i> | 255 nt - 85 aa | ✓✓ | - |
| 35 | RZ(-) RNA | 30018261 | <i>E.viroid</i> | <i>Virus</i> | 51 nt | ✓✓ | - |
| 36 | Sinal interno | 832917 | <i>S.cerevisiae</i> | <i>Fungi</i> | 63 nt - 21 aa | ✓ e ✓✓ | ✓ |
| 37 | Hormônio | 146762153 | <i>Petúnia</i> | <i>Viridiplantae</i> | 63 nt - 21 aa | ✓✓ | - |
| 38 | Ítron | 600528 | <i>R.norvegicus</i> | <i>Metazoa</i> | 63 nt | ✓ e ✓✓ | ✓ |
| 39 | Ítron | 27901730 | <i>E.nidulans</i> | <i>Fungi</i> | 63 nt | ✓✓ | - |
| 40 | DNA repetitivo | 20513849 | <i>H.vulgare</i> | <i>Viridiplantae</i> | 63 nt | ✓✓ | - |
| 41 | DNA repetitivo | 21734423 | <i>O.sativa</i> | <i>Viridiplantae</i> | 255 nt | ✓ e ✓✓ | ✓ |
| 42 | Proteína | 26556996 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 255 nt - 85 aa | ✓ e ✓✓ | - |
| 43 | Proteína | 38229169 | <i>Y.monkey</i> | <i>Virus</i> | 255 nt - 85 aa | ✓✓ | - |
| 44 | Proteína | 158339871 | <i>A.marina</i> | <i>Bactéria</i> | 255 nt - 85 aa | ✓ e ✓✓ | - |
| 45 | Proteína | 45368547 | <i>A.denitrificans</i> | <i>Bactéria</i> | 255 nt - 85 aa | ✓✓ | - |
| 46 | Proteína | 45368559 | <i>A.denitrificans</i> | <i>Bactéria</i> | 255 nt - 85 aa | ✓ e ✓✓ | ✓ |
| 47 | Proteína | 207258119 | <i>A.hospitalis</i> | <i>Archea</i> | 255 nt - 85 aa | ✓ e ✓✓ | - |
| 48 | Proteína | 158340643 | <i>A.marina</i> | <i>Bactéria</i> | 255 nt - 85 aa | ✓✓ | - |
| 49 | Proteína | 16740522 | <i>M.martensii</i> | <i>Metazoa</i> | 255 nt - 85 aa | ✓✓ | - |
| 50 | Proteína | 899225 | <i>B.napus</i> | <i>Viridiplantae</i> | 1023 nt - 341 aa | ✓✓ | - |
| 51 | Proteína | 30695458 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 1023 nt - 341 aa | ✓ e ✓✓ | - |
| 52 | Proteína | 145338521 | <i>A.thaliana</i> | <i>Viridiplantae</i> | 1023 nt - 341 aa | ✓✓ | - |
| 53 | Proteína | 50812173 | <i>B.subtilis</i> | <i>Bactéria</i> | 1023 nt - 341 aa | ✓ e ✓✓ | - |
| 54 | Proteína | 32455378 | <i>A.pasteurianus</i> | <i>Bactéria</i> | 1023 nt - 341 aa | ✓ e ✓✓ | ✓ |
| 55 | Proteína | 158341503 | <i>A.marina</i> | <i>Bactéria</i> | 1023 nt - 341 aa | ✓ e ✓✓ | - |

| Seq. nº | Sequências de DNA | GI número | Organismo | Espécie | Comprimentos em nt e aa | Código BCH sobre Anéis | Código BCH sobre Corpos |
|------------|----------------------|--------------|----------------------|-----------------|----------------------------|---------------------------|----------------------------|
| 56 | Proteína | 158340280 | <i>A.marina</i> | <i>Bactéria</i> | 1023 nt - 341 aa | ✓✓ | - |
| 57 | Proteína | 38229169 | <i>Y.monkey</i> | <i>Virus</i> | 1023 nt - 341 aa | ✓✓ | - |
| 58 | Proteína | 55376107 | <i>H.marismortui</i> | <i>Archea</i> | 1023 nt - 341 aa | ✓ e ✓✓ | ✓ |
| 59 | miRNA | 240153948 | <i>H.sapiens</i> | <i>Metazoa</i> | 21 nt | ✓✓ | - |
| 60 | miRNA | 240153946 | <i>H.sapiens</i> | <i>Metazoa</i> | 21 nt | ✓✓ | - |
| 61 | Proteína | 1000569 | <i>H.sapiens</i> | câncer | 63 nt - 21 aa | ✓ e ✓✓ | - |
| 62 | Proteína | 25140446 | <i>H.sapiens</i> | câncer | 63 nt - 21 aa | ✓ e ✓✓ | - |
| 63 | Proteína | 1000569 | <i>H.sapiens</i> | câncer | 1023 nt - 341 aa | ✓ e ✓✓ | - |
| 64 | Gene | 13263 | <i>H.sapiens</i> | <i>Metazoa</i> | 511 nt | ✓ e ✓✓ | - |
| 65 | Genoma | 118213250 | <i>H.sapiens</i> | <i>Metazoa</i> | 2047 nt | ✓ e ✓✓ | - |

Tabela 5.1: Sequências de DNA reproduzidas pelos códigos *G*-linearidade.

Observe na Tabela 5.1 que apenas 10 das 65 sequências de DNA foram reproduzidas através dos códigos *G*-linearidade: BCH primitivos sobre a estrutura de corpo e suas extensões de Galois com $D(\mathbf{a}, \mathbf{b}) = 1$. Como mencionamos anteriormente, não fizemos as análises para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$ em corpo devido à complexidade computacional envolvida nos cálculos. Um vez que, as fitas simples das sequências de DNA foram identificadas e reproduzidas através dos CCEs, as respectivas sequências da dupla hélice também serão identificadas e reproduzidas através dos mesmos CCEs, portanto, não há necessidade em repetir a reprodução das 65 sequências em termos da dupla hélice do DNA.

5.1.1 Alto fluxo de informação x baixa redundância

Em [20], foi observado que somente os códigos com a distância $d_H = 3$ foram capazes de identificar e reproduzir as sequências de direcionamento com 63 nucleotídeos identificadas na Tabela 5.1 por \checkmark^* . O mesmo fato aconteceu para todas as sequências de DNA reproduzidas neste trabalho.

No Apêndice A, mostramos as tabelas contendo a quantidade de erros t , as distâncias mínimas d_H , os polinômios geradores $g(x)$ e os códigos BCH com parâmetros (n, k, d_H) sobre corpos e sobre anéis relativos aos comprimentos de sequências de DNA n iguais a 21, 51, 63, 93, 255 e 1023 (Tabelas A.1 a A.7). No caso das sequências iguais a 21, 51, 63, 93 nucleotídeos, analisamos todas as distâncias correspondentes aos códigos (n, k, d_H) com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$. No caso das sequências com 255 e 1023 nucleotídeos analisamos as distâncias $d_H = 3$, $d_H = 5$ e $d_H = 7$ com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$. Dessas análises, observamos que somente os códigos BCH sobre corpos e sobre anéis com a distância $d_H = 3$ foram capazes de identificar e reproduzir as sequências de DNA citadas na Tabela 5.1. Lembrando que para o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$ as análises foram feitas somente em anéis. Nos casos de anel, observamos que o grau dos polinômios primitivos r e dos polinômios geradores $n - k$ são iguais. Nos casos de corpo, o grau dos polinômios primitivos

r e do polinômios geradores $n - k$ são diferentes, porém, o grau dos polinômios geradores no corpo continua sendo o mesmo grau dos polinômios geradores no anel (ver Tabela 5.3). Como consequência deste fato, a redundância está associada com o grau desses polinômios. Então, uma pequena redundância implica em um código de taxa alta, bem como em uma alta entropia (fluxo de informação alto).

5.2 Análise das Sequências de DNA Reproduzidas sobre Corpos

A Tabela 5.2 contempla as informações pertinentes às extensões de Galois (r), os parâmetros dos códigos, os polinômios primitivos e geradores, o rotulamento (Rot), a posição onde ocorreu a diferença (Pos) e, as diferenças entre os nucleotídeos e os aminoácidos das sequências de DNA que foram identificadas e reproduzidas em $GF(4^r)$. As diferenças em nucleotídeos são identificadas na tabela pela cor vermelha e, as alterações dos aminoácidos identificadas pelas cores correspondentes as classes dos aminoácidos. As figuras contendo as sequências de DNA do *NCBI* e as sequências reproduzidas pelos códigos também estão relacionadas na Tabela 5.2. A sequência Seq.13 é mostrada na Figura 5.2 e as demais sequências serão mostradas no Apêndice B (Figuras B.1 à B.3).

Observe na Tabela 5.2 que todas as sequências foram reproduzidas com um nucleotídeo de diferença das sequências de DNA do *NCBI*. As diferenças são indicadas da seguinte maneira: Seq.13 (W17stop), Seq.17 (A14S), Seq.19 (I19T), Seq.31 (S4R), Seq.31 (Q20stop), Seq.36 (Y5S), Seq.36 (T17S), Seq.38 ((GAG)12(GAT)), Seq.41 (F73C) e Seq.46 (R27C). No caso da sequência Seq.38, note que não tem a representação dos aminoácidos, apenas nucleotídeos, pois trata-se de uma sequência DNA de ítron que não fabrica proteína. Observe que as sequências Seq.31 e Seq.36 foram reproduzidas em dois polinômios primitivos/geradores distintos, apresentando erros de nucleotídeos/aminoácidos em posições distintas nas respectivas sequências.

Note nas sequências Seq.36 que os aminoácidos das posições 5 e 17 foram alterados, porém, as alterações ocorreram dentro da mesma classe de aminoácidos, representada na Tabela 5.2 pela mesma cor (ver também a classificação dos aminoácidos na Tabela 3.1 do Capítulo 3). No caso das demais sequências, as alterações nas posições dos aminoácidos provocaram mudanças de classes destes aminoácidos, representadas por cores diferentes. Essas sequências reproduzidas podem ou não manter a funcionalidade da proteína. Na literatura, existem casos conhecidos, cuja alteração de aminoácido dentro da mesma classe não altera a funcionalidade da proteína. Não sabemos se, biologicamente, essas sequências irão desempenhar as suas

funções, por estarem dentro da mesma classe de aminoácidos, pois não temos comprovação laboratorial para esses casos.

| r | Parâmetros do código $C(n, k, d_H)$ | Sequência de DNA (<i>Seq.nº</i>) | Polinômios: $p(x) = \text{primitivo} - g(x) = \text{gerador}$ | Rot Pos | Diferenças em nt e aa |
|---------|-------------------------------------|------------------------------------|---|---------------------------------|---|
| $r = 3$ | (63, 57, 3) | SD - M Seq.13 Figura 4.1 | $p_{05}(x) = x^3 + ax^2 + ax + a$ $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | S_4 17 ^a trinca | W TGG TGA stop |
| | | SD - RE Seq.17 Figura B.1 | $p_{08}(x) = x^3 + bx^2 + x + a$ $g_{08}(x) = x^6 + x^5 + x^4 + x + 1$ | S_4 14 ^a trinca | A GCC TCC S |
| | | SD - M Seq.19 Figura B.1 | $p_{10}(x) = x^3 + ax^2 + bx + b$ $g_{10}(x) = x^6 + x^5 + 1$ | S_4 4 ^a trinca | I ATT ACT T |
| | | SD - RE Seq.31 Figura B.1 | $p_{10}(x) = x^3 + ax^2 + bx + b$ $g_{10}(x) = x^6 + x^5 + 1$ | S_4 4 ^a trinca | S AGT AGA R |
| | | SD - RE Seq.31 Figura B.2 | $p_{08}(x) = x^3 + bx^2 + x + a$ $g_{08}(x) = x^6 + x^5 + x^4 + x + 1$ | S_4 20 ^a trinca | Q CAA TAA stop |
| | | S.interno Seq.36 Figura B.2 | $p_{01}(x) = x^3 + x^2 + ax + b$ $g_{01}(x) = x^6 + x + 1$ | S_4 5 ^a trinca | Y TAC TCC S |
| | | S.interno Seq.36 Figura B.2 | $p_{03}(x) = x^3 + ax^2 + bx + a$ $g_{03}(x) = x^6 + x^5 + x^2 + x + 1$ | S_4 17 ^a trinca | T ACC TCC S |
| | | Ítron Seq.38 Figura B.2 | $p_{05}(x) = x^3 + ax^2 + ax + a$ $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | S_4 12 ^a trinca | GAG GAT |
| $r = 4$ | (255, 247, 3) | DNA rep Seq.41 Figura B.3 | $p_{13}(x) = x^4 + x^3 + x + a$ $g_{13}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ | S_4 73 ^a trinca | TTT TGT |
| | | Proteína Seq.46 Figura B.3 | $p_{01}(x) = x^4 + x^3 + ax^2 + ax + a$ $g_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ | S_4 27 ^a trinca | R CGC TGC C |
| $r = 5$ | (1023, 1013, 3) | Proteína Seq.54 | $p_{06}(x) = x^5 + x^3 + x + b$ $g_{06}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ | - | - |
| | | Proteína Seq.58 | $p_{08}(x) = x^5 + bx^4 + x^3 + bx^2 + a$ $g_{08}(x) = x^{10} + x^9 + x^8 + x^6 + x^3 + x^2 + 1$ | - | - |

Tabela 5.2: Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GF(4^r)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença.

As sequências Seq.13 e Seq.31 tiveram alterações drásticas (*stop*) nos aminoácidos das posições 17 e 20, respectivamente. Isso pode implicar na perda da funcionalidade dessas proteínas do ponto de vista biológico. No caso das sequências Seq.54 e Seq.58, não temos as

sequências reproduzidas pelos códigos (1023, 1013, 3) em virtude da complexidade dos cálculos envolvidos no programa computacional. Sabemos que essas sequências são identificadas no corpo de Galois $GF(4^5)$ devido as análise dessas sequências no anel de Galois $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ que comentaremos mais adiante.

5.2.1 Relação entre polinômios primitivos e polinômios geradores do $GF(4^r)$ e do $GR(4, r)$

Durante as simulações dos algoritmos, usamos todos os polinômios primitivos correspondentes para cada uma das extensões de Galois tanto do $GF(4^r)$ quanto do $GR(4, r)$. No caso do BCH primitivo sobre $GF(4^r)$, observamos que, para cada dois polinômios primitivos, temos um polinômio gerador correspondente, cujo polinômio gerador é igual a um dos polinômios primitivos do anel, ou seja, esses polinômios estão co-relacionados entre si de alguma forma, ver exemplo na Tabela 5.3.

| Grau dos $p(x)$ e dos $g(x)$ para o código (63, 57, 3) | |
|--|--|
| Corpo | Anel |
| $p_{01}(x) = x^3 + x^2 + ax + b$ e $p_{02}(x) = x^3 + x^2 + bx + a$ $g_{01}(x) = g_{02}(x) = x^6 + x + 1$ | $p_{01}(x) = x^6 + x + 1$ $g_{01}(x) = x^6 + 2x^3 + 3x + 1$ |
| $p_{06}(x) = x^3 + x^2 + x + b$ e $p_{09}(x) = x^3 + x^2 + x + a$ $g_{06}(x) = g_{09}(x) = x^6 + x^4 + x^3 + x + 1$ | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ $g_{02}(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1$ |
| $p_{10}(x) = x^3 + ax^2 + bx + b$ e $p_{12}(x) = x^3 + bx^2 + ax + a$ $g_{10}(x) = g_{12}(x) = x^6 + x^5 + 1$ | $p_{03}(x) = x^6 + x^5 + 1$ $g_{03}(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ |
| $p_{03}(x) = x^3 + ax^2 + bx + a$ e $p_{04}(x) = x^3 + bx^2 + ax + b$ $g_{03}(x) = g_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ |
| $p_{05}(x) = x^3 + ax^2 + ax + a$ e $p_{07}(x) = x^3 + bx^2 + bx + b$ $g_{05}(x) = g_{07}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$ |
| $p_{08}(x) = x^3 + bx^2 + x + a$ e $p_{11}(x) = x^3 + ax^2 + x + b$ $g_{08}(x) = g_{11}(x) = x^6 + x^5 + x^4 + x + 1$ | $p_{06}(x) = x^6 + x^5 + x^4 + x + 1$ $g_{06}(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ |

Tabela 5.3: Relação dos $p(x)$ e dos $g(x)$ em $GF(4^3)$ e $GR(4, 6)$ para o código (63, 57, 3).

Note que o grau dos polinômios geradores no corpo é o mesmo que o grau dos polinômios primitivos e dos polinômios geradores no anel, como comentamos na Subseção 5.1.1 deste capítulo. O mesmo fato se repetiu com os códigos (255, 247, 3) e (1023, 1013, 3) construídos sobre o $GF(4^r)$ e o $GR(4, r)$.

5.2.2 Rotulamento das sequências de DNA sobre $GF(4)$

Em geral o alfabeto de um CCE é frequentemente estabelecido *a priori* de modo que tenha uma estrutura matemática bem definida para facilitar o processo de codificação e decodificação das sequências. Este não é o caso para os CCEs genéticos e genômicos, uma vez que existe uma estrutura bioquímica e biofísica bem definida no genoma. Com isso conjecturamos que tais alfabetos e suas possíveis estruturas matemáticas provavelmente já estão determinados nas sequências de DNA.

Para contornar este problema procedemos da seguinte maneira: a) No caso de sequências em termos da fita simples do DNA, o alfabeto 4-ário na saída da fonte está relacionado ao conjunto de nucleotídeos denotado por $N = \{A, C, G, T/U\}$, alfabeto do código genético; b) No caso de sequências em termos da dupla hélice do DNA, o alfabeto 4-ário na saída da fonte está relacionado ao conjunto de bases complementares denotado por $N' = \{AT, CG, GC, TA\}$, alfabeto do código genômico. Similarmente, o alfabeto 4-ário do código de bloco linear BCH sobre corpo é denotado por $GF(4) = \{0, 1, a, b\}$, obedecendo as operações de adição e multiplicação. Como os rotulamentos entre $N \rightarrow GF(4)$ e $N' \rightarrow GF(4)$ são desconhecidos, então toda sequência de DNA foi rotulada por cada uma das 24 permutações entre $N \rightarrow GF(4)$ e $N' \rightarrow GF(4)$, como mostra a Figura 5.1.

$N! = 4! = 24$ possibilidades de rotulamento para $GF(4) = \{0, 1, a, b\}$

$$\begin{bmatrix} A & C & G & T \\ 0 & 1 & b & a \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & 1 & b & 0 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 0 & 1 & a & b \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & 1 & 0 & b \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 0 & a & 1 & b \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & 0 & 1 & b \end{bmatrix}$$

$$\begin{bmatrix} A & C & G & T \\ 0 & b & 1 & a \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & b & 1 & 0 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 0 & b & a & 1 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & b & 0 & 1 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 0 & a & b & 1 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ a & 0 & b & 1 \end{bmatrix}$$

$$\begin{bmatrix} A & C & G & T \\ 1 & 0 & a & b \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & 0 & a & 1 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 1 & 0 & b & a \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & 0 & 1 & a \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 1 & b & 0 & a \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & 1 & 0 & a \end{bmatrix}$$

$$\begin{bmatrix} A & C & G & T \\ 1 & a & 0 & b \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & a & 0 & 1 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 1 & a & b & 0 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & a & 1 & 0 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ 1 & b & a & 0 \end{bmatrix} \begin{bmatrix} A & C & G & T \\ b & 1 & a & 0 \end{bmatrix}$$

$N'! = 4! = 24$ possibilidades de rotulamento para $GF(4) = \{0, 1, a, b\}$

$$\begin{bmatrix} AT & CG & GC & TA \\ 0 & 1 & b & a \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & 1 & b & 0 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 0 & 1 & a & b \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & 1 & 0 & b \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 0 & a & 1 & b \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & 0 & 1 & b \end{bmatrix}$$

$$\begin{bmatrix} AT & CG & GC & TA \\ 0 & b & 1 & a \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & b & 1 & 0 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 0 & b & a & 1 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & b & 0 & 1 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 0 & a & b & 1 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ a & 0 & b & 1 \end{bmatrix}$$

$$\begin{bmatrix} AT & CG & GC & TA \\ 1 & 0 & a & b \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & 0 & a & 1 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 1 & 0 & b & a \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & 0 & 1 & a \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 1 & b & 0 & a \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & 1 & 0 & a \end{bmatrix}$$

$$\begin{bmatrix} AT & CG & GC & TA \\ 1 & a & 0 & b \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & a & 0 & 1 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 1 & a & b & 0 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & a & 1 & 0 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ 1 & b & a & 0 \end{bmatrix} \begin{bmatrix} AT & CG & GC & TA \\ b & 1 & a & 0 \end{bmatrix}$$

Figura 5.1: Rotulamentos S_4 e S'_4 para o $GF(4)$.

Observamos que para cada sequência identificada e reproduzida pelo código existem vinte e quatro palavras-código correspondentes as 24 permutações. Empregando os rotulamentos recíprocos, $GF(4) \rightarrow N$ e $GF(4) \rightarrow N'$, em cada uma dessas 24 palavras-código temos como

resultado 24 palavras-código iguais em termos de nucleotídeos e aminoácidos, (ver exemplo na Tabela 4.11 do Capítulo 4). Este fato resulta em um conjunto contendo as vinte e quatro permutações, denominado grupo das permutações S_4 para o alfabeto genético S'_4 para o alfabeto genômico. Classificamos o rotulamento utilizado no codificador, em termos dos rotulamentos S_4 e S'_4 .

5.2.3 Representação algébrica das sequências de DNA no $GF(4^r)$

Nos processos de identificação e reprodução das sequências de DNA através dos códigos G -linearidade, alguns questionamentos surgiram com relação às fitas codante e não codante do DNA e a dupla hélice. Da Biologia, temos que as fitas da dupla hélice do DNA estão em direções opostas, o que significa que são antiparalelas. Uma vez determinada a estrutura matemática para algumas sequências de direcionamento ([20]) em termos da fita simples, é possível determinarmos a estrutura matemática da dupla hélice também? É possível uma caracterização algébrica via polinômios dessas fitas (simples e dupla)?

Mostramos que os códigos BCH sobre corpos de Galois são capazes de representar algebricamente a sequência de DNA em termos da dupla hélice do DNA e das fitas simples do DNA, como mostrado na Figura 5.2. Durante o processo, consideramos: a) A identificação e reprodução das bases complementares da dupla hélice do DNA através do polinômio primitivo $p_{05}(x)$ e do polinômio gerador $g_{05}(x)$; b) A identificação e reprodução dos nucleotídeos das fitas simples do DNA, tais como: 1º) A reprodução da fita simples na direção 5'-3' através dos mesmos polinômios $p_{05}(x)$ e $g_{05}(x)$; 2º) A reprodução da fita simples na direção 3'-5' (fita complementar) também através dos mesmos polinômios $p_{05}(x)$ e $g_{05}(x)$; 3º) A reprodução da fita não codante na direção 5'-3' através dos polinômios recíprocos $p'(x)$ e $g'(x)$; 4º) A reprodução do mRNA novamente através dos polinômios $p_{05}(x)$ e $g_{05}(x)$.

Primeiramente, observe que os polinômios $p_{05}(x)$ e $g_{05}(x)$ foram capazes de identificar e reproduzir tanto a dupla hélice do DNA quanto a fita codante e a fita complementar, todas com um nucleotídeo de diferença da sequência original (identificado pela cor vermelha) na mesma posição. Segundo os rotulamentos S_4 e S'_4 , mostrados na Figura 5.1, observe que não existem diferenças entre a reprodução da dupla hélice e das fitas simples do DNA. Em seguida, observe que os polinômios recíprocos $p'(x)$ e $g'(x)$ dos polinômios $p_{05}(x)$ e $g_{05}(x)$, foram capazes de identificar e reproduzir a fita não codante. Finalmente, a sequência é transcrita no mRNA sendo alterado o nucleotídeo timina (T) pela uracila (U), o que matematicamente não tem diferença, e, traduzida na sequência de aminoácidos. O RNA transportador (tRNA) realiza o mapeamento casado entre cada um dos códons nesta sequência com os correspondentes aminoácidos. Essa representação algébrica é válida para todas as sequências que foram identificadas e reproduzidas em corpo indicadas na Tabela 5.1.

Seq.13| *I.batatas* - Mitocôndria - F1-ATPase delta subunit – GI número 217937

Dupla hélice no NCBI:

5' -ATG TTC AGG CAC TCT TCT CGA CTC CTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 3' -TAC AAG TCC GTG AGA AGA GCT GAG GAT CGA GCG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'

a) Reprodução da dupla hélice do DNA: $p_{05}(x) = x^3 + ax^2 + ax + a$ - $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$

Código G-linearidade (63,57,3) BCH primitivo sobre GF(4) – Rotulamento S'_4 : (0,1,a,b) = (AT,CG,GC,TA))

5' -ATG TTC AGG CAC TCT TCT CGA CTC CTA GCT CGC GCC ACC ACA ATG GGG TGA CGT CGC CCC TTC-3'
 |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
 3' -TAC AAG TCC GTG AGA AGA GCT GAG GAT CGA GCG CGG TGG TGT TAC CCC ACT GCA GCG GGG AAG-5'
 Oba bbl 0aa 101 b1b b1b 1a0 1b1 1b0 alb 1a1 a11 011 010 Oba aaa ba0 lab 1a1 111 bb1

b) Reprodução das fitas simples do DNA:

Código G-linearidade ((63,57,3) BCH primitivo sobre GF(4) – Rotulamento S_4 : (0,1,a,b) = (A,C,G,T))

1º) Fita simples na direção 5'-3' reproduzida pelo código: $p_{05}(x) = x^3 + ax^2 + ax + a$ - $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$

5' -ATG TTC AGG CAC TCT TCT CGA CTC CTA GCT CGC GCC ACC ACA ATG GGG TGA CGT CGC CCC TTC-3'
 Oba bbl 0aa 101 b1b b1b 1a0 1b1 1b0 alb 1a1 a11 011 010 Oba aaa ba0 lab 1a1 111 bb1

2º) Fita simples na direção 3'-5' reproduzida pelo código: $p_{05}(x) = x^3 + ax^2 + ax + a$ - $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$

3' -TAC AAG TCC GTG AGA AGA GCT GAG GAT CGA GCG CGG TGG TGT TAC CCC ACT GCA GCG GGG AAG-5'
 Oba bbl 0aa 101 b1b b1b 1a0 1b1 1b0 alb 1a1 a11 011 010 Oba aaa ba0 lab 1a1 111 bb1

3º) Fita não codante: $p'(x) = x^3 + x^2 + x + b$ - $g'(x) = x^6 + x^4 + x^3 + x + 1$

5' -GAA GGG GCG ACG TCA CCC CAT TGT GGT GGC GCG AGC TAG GAG TCG AGA AGA GTG CCT GAA CAT-3'
 a00 aaa ala 01a b10 111 10b bab aab aal ala 0a1 b0a a0a bla 0a0 0a0 aba 11b a00 10b

4º) Transcrição-mRNA: $p_{05}(x) = x^3 + ax^2 + ax + a$ - $g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ – Rotulamento S_4 : (0,1,a,b) = (A,C,G,U)

5' -AUG UUC AGG CAC UCU UCU CGA CUC CUA GCU CGC GCC ACC ACA AUG GGG UGA CGU CGC CCC UUC-3'
 Oba bbl 0aa 101 b1b b1b 1a0 1b1 1b0 alb 1a1 a11 011 010 Oba aaa ba0 lab 1a1 111 bb1

Tradução (mapeamento casado – tRNA e rRNA)

Proteína:

M F R H S S R L L A R A T T M G sto R R P F

Figura 5.2: Representação algébrica via polinômios da dupla hélice e das fitas simples de DNA da sequência Seq.13 no GF(4).

5.3 Análise das Sequências de DNA Reproduzidas sobre Anéis

As informações necessárias sobre a identificação e a reprodução das sequências de DNA através dos códigos G -linearidade: BCH primitivos sobre anel com o padrão de erro $D(a, b) = 1$ são mostradas nas Tabelas 5.4 e 5.5. Observe que, as sequências Seq.38 e Seq.41 não têm os aminoácidos correspondentes pois essas sequências não fabricam proteínas. As sequências

Seq.42, Seq.46 (rotulamentos A e B) e Seq.58 foram reproduzidas sem alterações de aminoácidos, ou seja, sofreram mutações silenciosas.

| r | Parâmetros do código $C(n, k, d)$ | Sequência de DNA (Seq.nº) | Polinômios: $p(x) = \text{primitivo} - g(x) = \text{gerador}$ | Rot Pos | Diferenças em nt e aa |
|----------|-----------------------------------|-----------------------------|---|-----------------------------|--|
| $r = 06$ | $(63, 57, 3)$ | S.interno Seq.36 Figura 4.2 | $p_{01}(x) = x^6 + x + 1$ $g_{01}(x) = x^6 + 2x^3 + 3x + 1$ | C 10 ^a trinca | W TGG GGG G |
| | | S.interno Seq.36 Figura B.4 | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ $g_{02}(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1$ | B 16 ^a trinca | A GCC TCC S |
| | | S.interno Seq.36 Figura B.4 | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ | A 12 ^a trinca | G GGA GCA A |
| | | Ítron Seq.38 Figura B.4 | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ | B 3 ^a trinca | AGG CGG |
| | | Ítron Seq.38 Figura B.4 | $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$ | C 21 ^a trinca | AAG CAG |
| $r = 08$ | $(255, 247, 3)$ | DNA rep. Seq.41 Figura B.5 | $p_{09}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ $g_{09}(x) = x^8 + 2x^7 + 3x^6 + x^4 + 3x^3 + x^2 + x + 1$ | C 52 ^a trinca | ACA CCA |
| | | Proteína Seq.42 Figura B.5 | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $g_{02}(x) = x^8 + 2x^7 + x^6 + 3x^5 + x^3 + 1$ | B 65 ^a trinca | A GCT GAT D |
| | | Proteína Seq.42 Figura B.6 | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $g_{02}(x) = x^8 + 2x^7 + x^6 + 3x^5 + x^3 + 1$ | C 63 ^a trinca | G GGC GGT G |
| | | Proteína Seq.44 Figura B.6 | $p_{12}(x) = x^8 + x^7 + x^2 + x + 1$ $g_{12}(x) = x^8 + 3x^7 + 2x^5 + x^2 + x + 1$ | C 47 ^a trinca | A GCA GGA G |
| | | Proteína Seq.46 Figura 4.4 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + 2x^6 + 2x^5 + 3x^4 + x^3 + 3x^2 + 2x + 1$ | B 39 ^a trinca | Q CAA CAG Q |
| | | Proteína Seq.46 Figura B.7 | $p_{11}(x) = x^8 + x^7 + x^5 + x^3 + 1$ $g_{11}(x) = x^8 + 3x^7 + 2x^6 + x^5 + 3x^3 + 1$ | A 14 ^a trinca | A GCA GCG A |
| | | Proteína Seq.47 Figura B.7 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 2x^7 + x^6 + x^5 + 2x^3 + x^2 + 2x + 1$ | A 69 ^a trinca | I ATT CTT L |

Tabela 5.4: Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GR(4, 6)$ e $GR(4, 8)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença.

As sequências Seq.47, Seq.51, Seq.53 e Seq.55 foram reproduzidas com alterações de aminoácidos dentro da mesma classe e, as demais sequências tiveram alterações nas classes dos aminoácidos.

| r | Parâmetros do código $C(n, k, d)$ | Sequência de DNA (<i>Seq.nº</i>) | Polinômios: $p(x) = \text{primitivo} - g(x) = \text{gerador}$ | Rot Pos | Diferenças em nt e aa |
|----------|-----------------------------------|------------------------------------|--|------------------------------|--|
| $r = 10$ | (1023, 1013, 3) | Proteina Seq.51 Figura 4.6 | $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ $g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$ | C 290 ^a trinca | L C T T F |
| | | Proteina Seq.53 Figura B.8 | $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ $g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$ | B 198 ^a trinca | E G A A T D |
| | | Proteina Seq.54 Figura B.9 | $p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ $g_{04}(x) = x^{10} + 2x^8 + 3x^6 + x^5 + 3x^3 + 3x^2 + x + 1$ | A 226 ^a trinca | R C G T |
| | | Proteina Seq.55 Figura B.11 | $p_{60}(x) = x^{10} + x^7 + 1$ $g_{60}(x) = x^{10} + 3x^7 + 2x^5 + 1$ | C 268 ^a trinca | L T T G C F |
| | | Proteina Seq.58 Figura B.11 | $p_{05}(x) = x^{10} + x^9 + x^8 + x^6 + x^3 + x^2 + 1$ $g_{05}(x) = x^{10} + x^9 + 3x^8 + 2x^7 + x^6 + x^3 + x^2 + 2x + 1$ | B 77 ^a trinca | D G A C G A T D |

Tabela 5.5: Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH primitivos sobre $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença.

Nas Tabelas 5.6 e 5.7, mostramos as informações sobre a identificação e a reprodução das sequências de DNA através dos códigos G -linearidade: BCH não primitivos sobre anel com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$. As diferenças em nucleotídeos são mostradas nas tabelas pela cor vermelha e os aminoácidos identificados pelas cores correspondentes as classes dos aminoácidos (hidrofóbicos ou hidrofílicos)¹. As figuras contendo as sequências reproduzidas no anel estão relacionadas nas tabelas anteriormente citadas e, podem ser vistas no Apêndice B. Os resultados das simulações das sequência de DNA identificadas através dos códigos G -linearidade: BCH primitivos sobre $GR(4, r)$ reproduzidas com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$ são mostrados nas Tabelas 5.12, 5.13 e 5.14.

As sequências mencionadas nas Tabelas 5.6 e 5.7 não foram identificadas com 1 nucleotídeo de diferença da sequência do NCBI. Apenas com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$ é que foi possível identificar e reproduzir essas sequências através dos códigos G -linearidade: BCH não primitivos sobre anel.

Observe que as sequências Seq.01 (rotulamento A), Seq.02 (rotulamento C) e Seq.09 (rotulamento C) foram reproduzidas sem alterações de aminoácidos. As demais sequências apresentaram alterações, onde, em algumas sequências uma das posições não sofre alteração de aminoácido e a outra posição sofre alteração de aminoácido, podendo ocorrer dentro da mesma classe ou não.

¹hidrofóbicos = aminoácidos que repelem a água; hidrofílica = aminoácidos atraídos pela água.

| r | Parâmetros do código $C(n, k, d)$ | Sequência de DNA ($Seq.\,n^o$) | Polinômios: $p(x) = \text{primitivo} - g(x) = \text{gerador}$ | Rot Pos (trincas) | Diferenças em nt e aa |
|----------|-----------------------------------|----------------------------------|--|--|--|
| $r = 6$ | (21, 15, 3) | miRNA Seq.59 Figura 4.10 | $p_{01}(x) = x^6 + x + 1$ $g_{01}(x) = x^6 + 2x^5 + 3x^4 + 3x^2 + x + 1$ | C 1 ^a e 7 ^a | A G G T G T A C G T G G |
| | | miRNA Seq.60 Figura B.12 | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ | C 2 ^a | A G T C A T |
| $r = 08$ | (51, 43, 3) | SD - M Seq.01 Figura B.13 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | A 11 ^a e 16 ^a | S P T C A C C A T C G C C G S P |
| | | SD - M Seq.01 Figura B.13 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | C 2 ^a e 9 ^a | S L T C T CTC C C T T C P F |
| | | SD - M Seq.02 Figura B.14 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | A 7 ^a e 11 ^a | A F G C G T T C G T G T T T V F |
| | | SD - M Seq.02 Figura B.14 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | B 4 ^a e 11 ^a | A F G C T T T C G C G T G C A C |
| | | SD - M Seq.02 Figura B.14 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | C 4 ^a e 6 ^a | A S G C T T T C G C G T C A A S |
| | | SD - RE Seq.03 Figura B.15 | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $g_{02}(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$ | A 3 ^a e 15 ^a | Y C T A T T G C T A C T G G Y W |
| | | SD - RE Seq.03 Figura B.15 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | C 9 ^a e 10 ^a | S L T C T CTG T G T CCG C P |
| | | SD - M Seq.04 Figura B.15 | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $g_{02}(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$ | B 5 ^a e 8 ^a | V R G T C A G G G T G A T G V M |
| | | SD - M Seq.05 Figura B.16 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | B 6 ^a e 12 ^a | M G A T G G T A T T C G T I R |
| | | SD - M Seq.05 Figura B.16 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | B 9 ^a e 14 ^a | T Y A C C T A T C C C T T T P F |
| | | SD - M Seq.06 Figura B.16 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | C 10 ^a e 15 ^a | F R T T T A G A T T A A G T L N |
| | | SD - M Seq.06 Figura B.17 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | B 2 ^a e 17 ^a | F Q T T T C A G T A T T A G Y stop |
| | | SD - M Seq.07 Figura B.17 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | A 7 ^a e 8 ^a | C R T G T A G A T A T A C A Y T |
| | | SD - M Seq.08 Figura B.17 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | C 11 ^a e 17 ^a | S P T C C C T T C G A C T S T |

Tabela 5.6: Relação das sequências de DNA reproduzidas pelos códigos G -linearidade: BCH não primitivo sobre anel com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença.

| r | Parâmetros do código $C(n, k, d)$ | Sequência de DNA (<i>Seq.nº</i>) | Polinômios: $p(x) = \text{primitivo} - g(x) = \text{gerador}$ | Rot Pos (trincas) | Diferenças em nt e aa |
|----------|-----------------------------------|------------------------------------|---|--|---|
| $r = 08$ | (51, 43, 3) | SD - M Seq.08 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | B 3 ^a e 6 ^a | S A T C A G C T T G A G C C stop A |
| | | SD - M Seq.09 | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $g_{02}(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$ | C 6 ^a e 13 ^a | Q T C A A A C T C A T C C T H P |
| | | SD - M Seq.09 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | C 5 ^a e 7 ^a | T A A C T G C T A C C G C T A |
| | | SD - M Seq.10 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | A 1 ^a e 12 ^a | L T T T G A C T A T G C C T M P |
| | | SD - M Seq.10 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | B 1 ^a e 16 ^a | L R T T G C G A T G G C A A W Q |
| | | SD - M Seq.11 Figura 5.7 | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ | C 6 ^a e 12 ^a | L F C T A T TT G T A C TT V L |
| | | SD - M Seq.11 | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ | C* 8 ^a e 9 ^a | L L C T G T TA A T G T CA M S |
| | | RNA Seq.35 | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ | B 2 ^a e 9 ^a | G T C C T G C T C C T A |
| | | SD - M Seq.32 | $p_{01}(x) = x^{10} + x^3 + 1$ $g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ | A 15 ^a e 25 ^a | G T C G T A C C T T G CC L A |
| $r = 10$ | (93, 83, 3) | SD - M Seq.33 Figura 4.12 | $p_{01}(x) = x^{10} + x^3 + 1$ $g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ | C 3 ^a e 17 ^a | A V G C A G TT G A A T TT E F |
| | | SD - M Seq.33 Figura 4.12 | $p_{01}(x) = x^{10} + x^3 + 1$ $g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ | C 7 ^a e 26 ^a | L S C T T T C T C T G T A T L Y |
| | | SD - M Seq.33 Figura 4.12 | $p_{01}(x) = x^{10} + x^3 + 1$ $g_{01}(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ | C 14 ^a e 16 ^a | V A G T T A AA G T G C AA V K |

Tabela 5.7: Continuação da relação das sequências de DNA reproduzidas pelo código BCH não primitivo sobre anel com $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença.

Temos ainda casos em que as duas posições sofrem alterações de aminoácidos, podendo estar dentro ou fora da classe de aminoácidos. Não sabemos se biologicamente as sequências, com alterações nos aminoácidos, continuarão exercendo as suas respectivas funções. Análises laboratoriais são necessárias para a comprovação desses fatos. É importante ressaltarmos que em todos os casos de sequências de DNA reproduzidas em anel com $D(\mathbf{a}, \mathbf{b}) = 1$, não tivemos

nenhum caso onde a alteração do nucleotídeo provocou alteração drástica no aminoácido, a inclusão do códon de parada stop, exceto as sequências de câncer que mostraremos mais adiante. Para o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 2$, este fato aconteceu em mais sequências. No caso das sequências reproduzidas em corpo, vimos que duas delas apresentaram essa alteração drástica no aminoácido.

De uma forma geral, podemos relatar alguns aspectos biológicos observados na reprodução dessas sequências. Apesar dos códigos possuirem proteção igual para cada posição do códon, isto é, as três posições de cada códon possuem probabilidades iguais de erros, as trocas de nucleotídeos ocorreram com mais frequência na primeira e na terceira posição dos códons para $D(\mathbf{a}, \mathbf{b}) = 1$. Para $D(\mathbf{a}, \mathbf{b}) = 2$, isso ainda acontece, porém em menor frequência. Este fato infere que a segunda posição foi mais protegida contra os erros durante a reprodução dessas sequências pelo código CCE, o que faz sentido biologicamente, uma vez que a troca de nucleotídeo na segunda posição implica em uma troca de aminoácido.

Os processos de transição² e de transversão³ são observados em todas as sequências reproduzidas pelos códigos G -linearidade. Nas sequências Seq.42 (rotulamento C), Seq.46 (rotulamentos A e B), Seq.51 e Seq.58 da Tabela 5.4 ocorreu o processo de transição. As demais sequências tiveram a ocorrência do processo de transversão. Nas sequências Seq.01 (rotulamento A), Seq.02 (rotulamento C) e Seq.09 (rotulamento C) (Tabelas 5.6 e 5.7) ocorreu o processo de transição. Note que o processo de transversão ocorreu na maioria dos casos, Seq.59, Seq.60, Seq.02 (rotulamentos B e C), Seq.04, Seq.05, Seq.06, Seq.08, Seq.10, Seq.35 e Seq.33. Nas Sequências Seq.03 (rotulamentos A e C), Seq.07, Seq.08, Seq.10, Seq.11 e Seq.32, os processos de transversão e transição (ou vice-versa) ocorreram na mesma sequência.

5.3.1 Relação entre os polinômios primitivos e os polinômios geradores sobre $GR(4, r)$

Nos processos de geração dos códigos responsáveis pela identificação e reprodução das sequências de DNA estudadas neste trabalho, observamos que o comportamento dos polinômios primitivos ($p(x)$) em relação aos polinômios geradores ($g(x)$) nos códigos G -linearidade: BCH primitivos sobre $GR(4, r)$ são diferentes do comportamento dos ($p(x)$) em relação aos ($g(x)$) nos códigos G -linearidade: BCH não primitivos sobre $GR(4, r)$. Os valores de d nas Tabelas 5.8, 5.9, 5.10 e 5.11 indicam a potência do elemento primitivo f responsável por identificar o subgrupo cíclico em $GR^*(4, r)$ (Passo 6 do algoritmo de codificação sobre \mathbb{Z}_4 no Capítulo 4), vejamos:

²Um tipo de substituição de nucleotídeos no qual uma purina (adenina ou guanina) substitui outra purina ou uma pirimidina (citosina ou timina) substitui outra pirimidina

³Um tipo de substituição de nucleotídeos no qual uma purina substitui uma pirimidina ou vice-versa.

- **BCH primitivo** - Nos processos de geração dos códigos BCH primitivos: $(63, 57, 3)$ sobre $GR(4, 6)$, $(255, 247, 3)$ sobre $GR(4, 8)$, $(511, 502, 3)$ sobre $GR(4, 9)$, $(1023, 1013, 3)$ sobre $GR(4, 10)$ e $(2047, 2036, 3)$ sobre $GR(4, 11)$ observamos que cada $p(x)$ de grau r tem apenas um $g(x)$ correspondente, como mostrado na Tabela 5.8.

| Códigos BCH primitivos sobre $GR(4, r)$ | |
|--|-----------------------------|
| Extensão de Galois $r = 6 - G_n = G_{63} - d = 2 - nd = 126$ | |
| Polinômios primitivos $p(x)$ | Polinômios geradores $g(x)$ |
| $p_{01}(x) \dots$ | $g_{01}(x)$ |
| $p_{02}(x) \dots$ | $g_{02}(x)$ |
| \vdots | \vdots |
| $p_{06}(x) \dots$ | $g_{06}(x)$ |
| Extensão de Galois $r = 8 - G_n = G_{255} - d = 2 - nd = 510$ | |
| Polinômios primitivos $p(x)$ | Polinômios geradores $g(x)$ |
| $p_{01}(x) \dots$ | $g_{01}(x)$ |
| $p_{02}(x) \dots$ | $g_{02}(x)$ |
| \vdots | \vdots |
| $p_{16}(x) \dots$ | $g_{16}(x)$ |
| Extensão de Galois $r = 9 - G_n = G_{511} - d = 1 - nd = 511$ | |
| Polinômios primitivos $p(x)$ | Polinômios geradores $g(x)$ |
| $p_{01}(x) \dots$ | $g_{01}(x)$ |
| $p_{02}(x) \dots$ | $g_{02}(x)$ |
| \vdots | \vdots |
| $p_{48}(x) \dots$ | $g_{48}(x)$ |
| Extensão de Galois $r = 10 - G_n = G_{1023} - d = 2 - nd = 2046$ | |
| Polinômios primitivos $p(x)$ | Polinômios geradores $g(x)$ |
| $p_{01}(x) \dots$ | $g_{01}(x)$ |
| $p_{02}(x) \dots$ | $g_{02}(x)$ |
| \vdots | \vdots |
| $p_{60}(x) \dots$ | $g_{60}(x)$ |
| Extensão de Galois $r = 11 - G_n = G_{2047} - d = 1 - nd = 2047$ | |
| Polinômios primitivos $p(x)$ | Polinômios geradores $g(x)$ |
| $p_{01}(x) \dots$ | $g_{01}(x)$ |
| $p_{02}(x) \dots$ | $g_{02}(x)$ |
| \vdots | \vdots |
| $p_{176}(x) \dots$ | $g_{176}(x)$ |

Tabela 5.8: Relação entre $p(x)$ e $g(x)$ no código BCH primitivo sobre $GR(4, r)$

- **BCH não primitivo** - Nos processos de geração dos códigos BCH não primitivos observamos classes de polinômios primitivos/geradores. O comprimento n e a potência d são importantes no processo de determinação dessas classes de polinômios. Na Tabela 5.9, mostramos os seis $p(x)$ do código $(21, 15, 3)$ sobre $GR(4, 6)$ dividindo-se em duas

classes, sendo uma classe determinada pelos polinômios $p_{01}(x)$, $p_{05}(x)$ e $p_{06}(x)$ resultando no polinômio $g_{01}(x)$, e outra classe, determinada pelos polinômios $p_{02}(x)$, $p_{03}(x)$ e $p_{04}(x)$ resultando no polinômio $g_{02}(x)$.

| Código (21, 15, 3) BCH não primitivo sobre $GR(4, 6)$ - $d = 6$ | | |
|---|---|--|
| Sejam $n = (p^r - 1)$, $p = 2$ e $r = 6$ - $G_n = G_{21}$ - $nd = 126$ | | |
| Classes | Polinômios primitivos $p(x)$ | Polinômio gerador $g(x)$ |
| 1) | $p_{01}(x) = x^6 + x + 1$ $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ $p_{06}(x) = x^6 + x^5 + x^4 + x + 1$ | $g_{01}(x) = x^6 + 2x^5 + 3x^4 + 3x^2 + x + 1$ |
| 2) | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ $p_{03}(x) = x^6 + x^5 + 1$ $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | $g_{02}(x) = x^6 + x^5 + 3x^4 + 3x^2 + 2x + 1$ |

Tabela 5.9: Classe de polinômios primitivos dos códigos BCH não primitivo - (21, 15, 3).

No caso do código (51, 43, 3) sobre $GR(4, 8)$, os dezesseis polinômios $p(x)$ dividem-se em quatro classes resultando em quatro $g(x)$, como mostrados na Tabela 5.10.

| Código (51, 43, 3) BCH não primitivo sobre $GR(4, 8)$ - $d = 10$ | | |
|---|--|--|
| Sejam $n = (p^r - 1)$, $p = 2$ e $r = 8$ - $G_n = G_{51}$ - $nd = 510$ | | |
| Classes | Polinômios primitivos $p(x)$ | Polinômio gerador $g(x)$ |
| 1) | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ $p_{04}(x) = x^8 + x^5 + x^3 + x + 1$ $p_{12}(x) = x^8 + x^7 + x^2 + x + 1$ $p_{13}(x) = x^8 + x^6 + x^3 + x^2 + 1$ | $g_1(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ |
| 2) | $p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1$ $p_{07}(x) = x^8 + x^7 + x^3 + x^2 + 1$ $p_{14}(x) = x^8 + x^7 + x^6 + x^3 + x^2 + x + 1$ $p_{15}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + 1$ | $g_2(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$ |
| 3) | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ $p_{06}(x) = x^8 + x^6 + x^5 + x + 1$ $p_{08}(x) = x^8 + x^5 + x^3 + x^2 + 1$ $p_{09}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ | $g_3(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$ |
| 4) | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ $p_{10}(x) = x^8 + x^7 + x^6 + x + 1$ $p_{11}(x) = x^8 + x^7 + x^5 + x^3 + 1$ $p_{16}(x) = x^8 + x^6 + x^5 + x^4 + 1$ | $g_5(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ |

Tabela 5.10: Classe de polinômios primitivos dos códigos BCH não primitivo - (51, 43, 3).

E no caso do código (93, 83, 3) sobre $GR(4, 10)$, os sessenta polinômios $p(x)$ dividem-se em seis classes de polinômios $g(x)$, Tabela 5.11.

Código (93, 83, 3) BCH não primitivo sobre $GR(4, 10)$ - $d = 22$

 Sejam $n = (p^r - 1)$, $p = 2$ e $r = 10$ - $G_n = G_{93}$ - $nd = 2046$

| Classes | Polinômios primitivos $p(x)$ | Polinômio gerador $g(x)$ |
|---------|--|--|
| 1) | $p_{01}(x) = x^{10} + x^3 + 1$ $p_{10}(x) = x^{10} + x^9 + x^4 + x + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{26}(x) = x^{10} + x^6 + x^5 + x^2 + 1$ $p_{27}(x) = x^{10} + x^5 + x^3 + x^2 + 1$ $p_{30}(x) = x^{10} + x^5 + x^2 + x + 1$ $p_{33}(x) = x^{10} + x^9 + x^7 + x^6 + 1$ $p_{37}(x) = x^{10} + x^8 + x^6 + x^5 + x^3 + x + 1$ $p_{46}(x) = x^{10} + x^9 + x^8 + x^6 + x^4 + x^3 + 1$ $p_{53}(x) = x^{10} + x^9 + x^8 + x^4 + x^3 + x^2 + 1$ | $g_1(x) = x^{10} + 2x^7 + 2x^6 + x^5 + x^4 + 2x^3 + 3x^2 + 2x + 1$ |
| 2) | $p_{02}(x) = x^{10} + x^8 + x^3 + x^2 + 1$ $p_{12}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^2 + 1$ $p_{15}(x) = x^{10} + x^9 + x^8 + x^6 + x^4 + x^2 + 1$ $p_{19}(x) = x^{10} + x^9 + x^7 + x^6 + x^4 + x + 1$ $p_{28}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$ $p_{39}(x) = x^{10} + x^9 + x^7 + x^6 + x^4 + x^3 + x^2 + x + 1$ $p_{40}(x) = x^{10} + x^9 + x^8 + x^7 + x^3 + x^2 + 1$ $p_{43}(x) = x^{10} + x^9 + x^8 + x^7 + x^5 + x^4 + 1$ $p_{48}(x) = x^{10} + x^9 + x^6 + x^3 + x^2 + x + 1$ $p_{58}(x) = x^{10} + x^9 + x^8 + x^4 + x^2 + x + 1$ | $g_2(x) = x^{10} + x^9 + x^8 + 2x^7 + 2x^6 + 3x^5 + x^3 + 2x^2 + 3x + 1$ |
| 3) | $p_{03}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + 1$ $p_{06}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1$ $p_{08}(x) = x^{10} + x^8 + x^5 + x + 1$ $p_{17}(x) = x^{10} + x^9 + x^8 + x^5 + x^4 + x^3 + 1$ $p_{29}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + 1$ $p_{35}(x) = x^{10} + x^9 + x^4 + x^2 + 1$ $p_{44}(x) = x^{10} + x^7 + x^3 + x + 1$ $p_{50}(x) = x^{10} + x^7 + x^6 + x^2 + 1$ $p_{56}(x) = x^{10} + x^9 + x^5 + x^4 + x^2 + x + 1$ $p_{59}(x) = x^{10} + x^8 + x^7 + x^4 + x^2 + x + 1$ | $g_3(x) = x^{10} + 2x^9 + x^8 + 2x^5 + 2x^4 + 3x^3 + 2x^2 + 3x + 1$ |
| 4) | $p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ $p_{16}(x) = x^{10} + x^8 + x^7 + x^3 + x^2 + x + 1$ $p_{20}(x) = x^{10} + x^9 + x^8 + x^6 + x^2 + x + 1$ $p_{22}(x) = x^{10} + x^9 + x^8 + x^7 + x^4 + x + 1$ $p_{31}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1$ $p_{32}(x) = x^{10} + x^9 + x^6 + x^4 + x^3 + x + 1$ $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ $p_{49}(x) = x^{10} + x^8 + x^6 + x^4 + x^2 + x + 1$ $p_{55}(x) = x^{10} + x^8 + x^5 + x^4 + x^3 + x^2 + 1$ $p_{57}(x) = x^{10} + x^8 + x^7 + x^2 + 1$ | $g_4(x) = x^{10} + 3x^9 + 2x^8 + x^7 + 3x^5 + 2x^4 + 2x^3 + x^2 + x + 1$ |
| 5) | $p_{05}(x) = x^{10} + x^9 + x^8 + x^6 + x^3 + x^2 + 1$ $p_{11}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x + 1$ $p_{13}(x) = x^{10} + x^8 + x^4 + x^3 + 1$ $p_{21}(x) = x^{10} + x^7 + x^6 + x^5 + x^2 + x + 1$ $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{38}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x + 1$ $p_{41}(x) = x^{10} + x^9 + x^7 + x^3 + 1$ $p_{47}(x) = x^{10} + x^9 + x^5 + x^2 + 1$ $p_{52}(x) = x^{10} + x^9 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$ $p_{54}(x) = x^{10} + x^8 + x^6 + x + 1$ | $g_5(x) = x^{10} + 3x^9 + 2x^8 + 3x^7 + 2x^6 + 2x^5 + x^2 + 2x + 1$ |
| 6) | $p_{07}(x) = x^{10} + x^4 + x^3 + x + 1$ $p_{09}(x) = x^{10} + x^8 + x^5 + x^4 + 1$ $p_{18}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x + 1$ $p_{23}(x) = x^{10} + x^8 + x^7 + x^6 + x^2 + x + 1$ $p_{24}(x) = x^{10} + x^7 + x^6 + x^4 + x^2 + x + 1$ $p_{25}(x) = x^{10} + x^9 + x^7 + x^5 + x^4 + x^2 + 1$ $p_{36}(x) = x^{10} + x^9 + x^8 + x^5 + 1$ $p_{45}(x) = x^{10} + x^8 + x^7 + x^5 + 1$ $p_{51}(x) = x^{10} + x^9 + x^6 + x + 1$ $p_{60}(x) = x^{10} + x^7 + 1$ | $g_7(x) = x^{10} + 2x^9 + 3x^8 + 2x^7 + x^6 + x^5 + 2x^4 + 2x^3 + 1$ |

Tabela 5.11: Classe de polinômios primitivos dos códigos BCH não primitivo - (93, 83, 3).

5.3.2 Rotulamento das sequências de DNA sobre $GR(4, r)$

Todas as sequências de DNA mencionadas na Tabela 5.1 foram rotuladas para cada uma das 24 permutações entre $N \rightarrow \mathbb{Z}_4$ e $N' \rightarrow \mathbb{Z}_4$, Figuras 5.3 e Figura 5.4. Como os rotulamentos entre $N \rightarrow \mathbb{Z}_4$ e $N' \rightarrow \mathbb{Z}_4$ eram desconhecidos para essas sequências, relacionamos os conjuntos denotados por $N = \{A, C, G, T/U\}$ e $N' = \{AT, CG, GC, TA\}$ ao alfabeto 4-ário do código de bloco linear denotado por $\mathbb{Z}_4 = \{0, 1, 2, 3\}$ para a estrutura de anel, satisfazendo as operações de soma e produto módulo 4, como detalhado nos Capítulos 3 e 4.

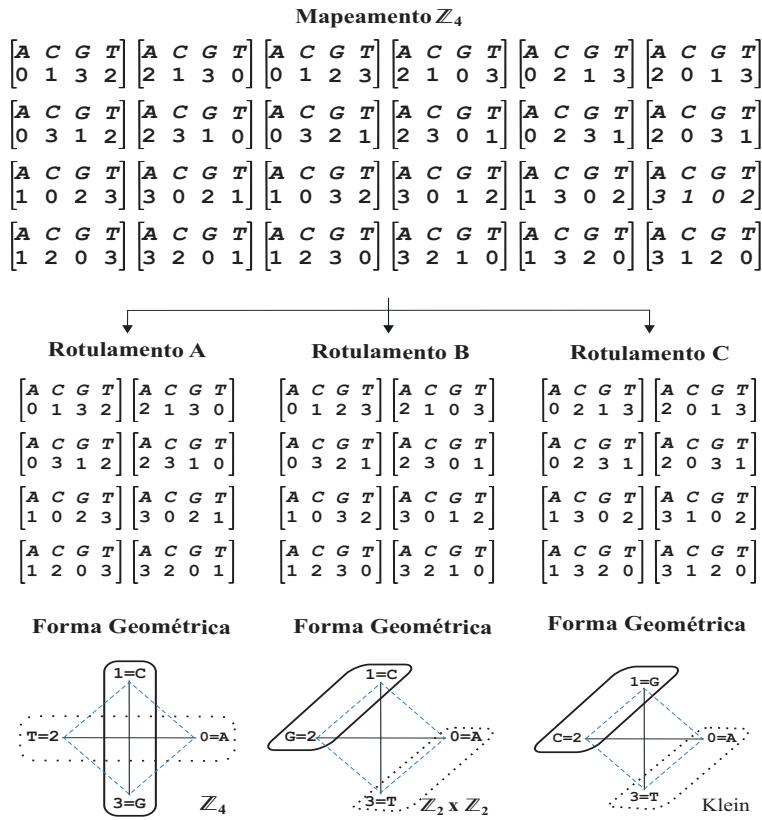


Figura 5.3: Rotulamentos A, B e C resultantes das 24 permutações entre $N \rightarrow \mathbb{Z}_4$.

O procedimento foi usado para todas as sequências da Tabela 5.1, reproduzidas pelos códigos BCH primitivo com $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$, e, BCH não primitivo com $D(\mathbf{a}, \mathbf{b}) = 2$. Para cada sequência, em termos da fita simples do DNA, reproduzida pelos códigos G -linearidade existem pelo menos oito palavras-código correspondentes a oito permutações. Empregando o rotulamento recíproco, $\mathbb{Z}_4 \rightarrow N$, em cada uma dessas oito palavras-código temos como resultado oito palavras-código iguais em termos de nucleotídeos e aminoácidos. Este fato resulta em três conjuntos contendo oito permutações cada um. Classificamos o rotulamento utilizado no codificador, em termos de rotulamentos A, B e C, Figura 5.3. Esses

rotulamentos estão relacionados às formas geométricas que produzem um diferente nível de não-linearidade para as sequências reproduzidas.

A representação binária associada a cada um desses rotulamentos é $0 - 00; 1 - 10; 2 - 11; 3 - 01$. Todavia, a associação de complementaridade dos nucleotídeos $A - T$ e $C - G$ com os rotulamentos é o que os diferenciam. No caso do rotulamento A , vemos que qualquer um dos nucleotídeos para alcançar o seu complementar necessita caminhar duas arestas, enquanto que nos dois rotulamentos restantes basta caminhar uma aresta somente. Todas as permutações associadas ao rotulamento A caracterizam o mapeamento como \mathbb{Z}_4 -linear; as permutações associadas ao rotulamento B caracterizam o mapeamento como $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear; enquanto que as permutações associadas ao rotulamento C caracterizam o mapeamento como Klein-linear. O rotulamento A classifica as sequências como não-lineares, enquanto que os rotulamentos B e C as classificam como lineares. Esses rotulamentos/mapeamentos identificam as melhores associações entre cada símbolo no conjunto N e no conjunto N' e o seu correspondente símbolo no conjunto \mathbb{Z}_4 e vice-versa.

Esses três conjuntos também se confirmaram na identificação e reprodução da dupla hélice do DNA sobre anel mostrados na figura abaixo.

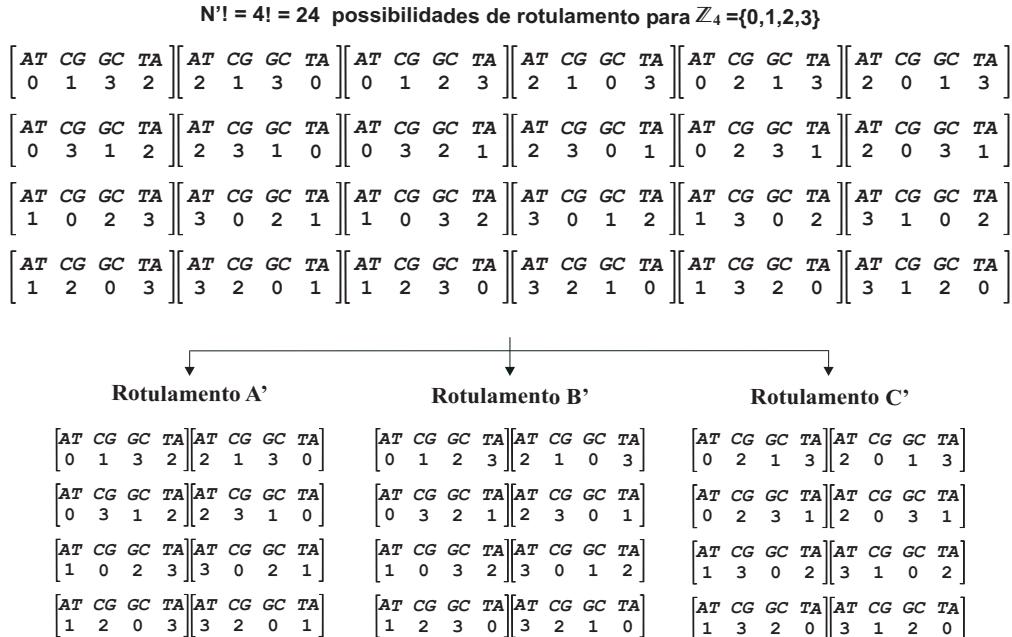


Figura 5.4: Rotulamentos A' , B' e C' resultante das 24 permutações entre $N' \rightarrow \mathbb{Z}_4$.

Em termos dos códigos: BCH primitivos para $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ e, BCH não primitivos para $D(\mathbf{a}, \mathbf{b}) = 2$, a identificação e reprodução das sequências está relacionada a determinados polinômios primitivos e aos rotulamentos. Sendo que, as oito permutações

associadas aos rotulamentos A, B e C não são alteradas em termos destes códigos, o que muda é a quantidade de palavras-código para cada oito permutações, da seguinte forma:

- **Códigos BCH Primitivos com $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$** - No caso das sequências de DNA identificadas e reproduzidas através dos códigos BCH primitivos (63, 57, 3), (255, 247, 3) e (1023, 1013, 3) com $D(\mathbf{a}, \mathbf{b}) = 1$, existem oito palavras-código correspondentes a oito permutações, que, quando empregado o rotulamento recíproco, obtemos oito palavras-código iguais em termos de nucleotídeos e aminoácidos, ou seja, o resultado final é uma sequência de DNA reproduzida no rotulamento A ou B ou C (se necessário, consultar Tabelas 4.21, 4.27 do Capítulo 4). As sequências de DNA mencionadas na Tabela 5.4 foram reproduzidas com $D(\mathbf{a}, \mathbf{b}) = 1$ através de determinados polinômios primitivos associadas aos rotulamentos A, B ou C, com exceção da Seq.42, gerada com dezesseis palavras-código reproduzidas nos rotulamentos B e C pelo mesmo polinômio primitivo.

Já no caso dessas sequências serem identificadas e reproduzidas através dos códigos BCH primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$, sempre existem várias palavras-código para cada uma das vinte e quatro permutações, conforme o comprimento da palavra-código n (se necessário, consultar as Tabelas 4.22, 4.28 4.32 do Capítulo 4). Quando empregamos o rotulamento recíproco em todas essas palavras-código encontradas, novamente os três rotulamentos A, B e C são identificados e associados às sequências de DNA reproduzidas, ou seja, o resultado final são várias sequências de DNA reproduzidas nos rotulamentos A, B e C. Veja nas Tabelas 5.12, 5.13 e 5.14 o resultado das simulações de todas as sequências de DNA com comprimentos iguais a 63, 255 e 1023 nucleotídeos que foram identificadas através dos códigos BCH primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$ e reproduzidas nos três rotulamentos. Nas tabelas anteriormente citadas, mostramos a quantidade de palavras-código relativas aos rotulamentos A, B e C indicadas pelo número superior dentro de cada quadradinho, por exemplo: 5, 3, 63 ou 93.

- **Códigos BCH não Primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$** - As sequências com comprimentos iguais a 21, 51 e 93 nucleotídeos, identificadas através dos códigos BCH não primitivos (21, 15, 3), (51, 43, 3) e (93, 83, 3) com $D(\mathbf{a}, \mathbf{b}) = 2$, também estão associadas aos três rotulamentos A, B ou C, porém, apresentaram três situações distintas com relação a quantidade de palavras-código (se necessário, consultar as Tabelas 4.34, 4.38 e 4.32 do Capítulo 4).

Por exemplo, a sequência Seq.02 da Tabela 5.6 foi reproduzida nos rotulamentos A, B e C pelo mesmo polinômio primitivo, nesse caso obtemos vinte e quatro palavras-código, ou seja, o resultado final são três sequências de DNA reproduzidas nos rotulamentos A,

B e C. A sequência Seq.10 (Tabela 5.7) foi gerada com dezesseis palavras-código, resultando em duas sequências de DNA reproduzidas nos rotulamentos A e B. A sequência Seq.33 foi gerada com três palavras-código distintas em cada uma das oito permutações correspondente ao rotulamento C: por isso temos três sequências distintas da Seq.33 reproduzidas pelo mesmo polinômio no rotulamento C, mostradas na Figura 4.12 do Capítulo 4.

Do ponto de vista matemático, as sequências de DNA mostradas nas Tabelas 5.4, 5.6 e 5.7 com o rotulamento A foram reproduzidas pelos códigos \mathbb{Z}_4 -linearidade: estas sequências são classificadas como não-lineares. As sequências de DNA com rotulamentos B e C foram reproduzidas pelos códigos $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade, respectivamente. Estas sequências são classificadas como sequências lineares. Portanto, os códigos usados na identificação e reprodução das sequências de DNA sobre anéis, no presente trabalho, consistem de códigos **G -linearidade** = $\{\mathbb{Z}_4\text{-linearidade}, \mathbb{Z}_2 \times \mathbb{Z}_2\text{-linearidade}, \text{Klein-linearidade}\}$, se necessário ver os detalhes na Subseção 3.2.1 do Capítulo 3.

5.3.3 Dependência entre os códigos corretores de erros e os polinômios primitivos

Sob o ponto de vista algébrico, em sistemas de transmissão digital, a construção de um CCE sobre anel ou corpo não depende do polinômio primitivo com o grau r usado na extensão de Galois, pois, algebraicamente, os corpos gerados pelos correspondentes polinômios primitivos são isomorfos. Entretanto, a identificação e reprodução das sequências de DNA através dos códigos G -linearidade: códigos BCH primitivos com os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$, e códigos BCH não primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$ mostraram uma relação de dependência na existência de CCE (e seus rotulamentos/mapeamentos associados) com alguns polinômios primitivos. Em [20], esse fato já havia sido identificado para as dez sequências de direcionamento identificadas em anel através dos códigos BCH primitivos sobre $GR(4, 6)$. No presente estudo, observamos os diferentes graus de dependência entre os CCE e os polinômios primitivos da seguinte maneira:

- **Códigos BCH Primitivos com $D(\mathbf{a}, \mathbf{b}) = 1$** - No caso das sequências reproduzidas através do código BCH primitivo com $D(\mathbf{a}, \mathbf{b}) = 1$ existe um alto grau de dependência na existência de CCE (e seus rotulamentos associados) com alguns polinômios primitivos. Observe na Tabela 5.4 que não são todos os polinômios primitivos de grau r que foram capazes de reproduzir as sequências de DNA com $D(\mathbf{a}, \mathbf{b}) = 1$. Apenas determinados polinômios associados a determinados rotulamentos são capazes de reproduzir

algumas sequências de DNA dando uma sustentação forte para a existência de uma forma algébrica associada a forma geométrica e topológica dessas sequências, onde os elementos, tais como: a estrutura algébrica, o alfabeto, o rotulamento, o mapeamento, o polinômio primitivo e o polinômio gerador, são fundamentais e, portanto, devem ser considerados.

- **Códigos BCH Primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$** - Já, no caso da reprodução das sequências através dos códigos BCH primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$, essa relação de dependência entre os CCEs e os polinômios primitivos é quase que totalmente perdida. Não importa o polinômio primitivo de grau r a ser usado no processo de geração do código, todo e qualquer polinômio primitivo de grau r será capaz de identificar e reproduzir as sequências de DNA nos três rotulamentos A, B e C, caracterizados como os mapeamentos \mathbb{Z}_4 -linear, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linear e Klein-linear, respectivamente. Porém, dizemos que é quase que totalmente perdida, pois do ponto de vista biológico, nem todos os polinômios primitivos serão capazes de reproduzir as sequências de DNA sem alteração de aminoácido. Veja nas Tabelas 5.12, 5.13 e 5.14 os resultados das simulações das sequências de DNA com comprimentos iguais a 63, 255 e 1023 nucleotídeos reproduzidas com $D(\mathbf{a}, \mathbf{b}) = 2$ relativos aos polinômios primitivos $p(x)$ e aos rotulamentos A, B e C. Nas tabelas, os quadradinhos na cor cinza indicam em quais $p(x)$ e em quais rotulamentos as sequências de DNA foram identificadas e reproduzidas sem alteração de aminoácido.
- **Códigos BCH não Primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$** - Em termos das sequências reproduzidas através dos códigos BCH não primitivos com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$, novamente temos uma forte relação de dependência na existência de CCEs (e seus rotulamentos associados) com alguns polinômios primitivos, porém em menor grau que os BCH primitivos - $D(\mathbf{a}, \mathbf{b}) = 1$. Nas Tabelas 5.6 e 5.7 podem ser observados alguns casos, onde uma mesma sequência foi gerada e reproduzida em mais de um polinômio primitivo. Por exemplo, a sequência Seq.01 foi gerada e reproduzida nos polinômios primitivos p_{01} e p_{03} , os demais polinômios primitivos da extensão $r = 8$ não foram capazes de identificar e reproduzir essa sequência, mostrando assim a relação de dependência.

5.3.4 A relação entre o BCH primitivo sobre $GR(4, r)$ e o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$

A ideia em identificar e reproduzir as sequências de DNA com até 2 nucleotídeos de diferença das sequências do NCBI ($D(\mathbf{a}, \mathbf{b}) = 2$) segue do questionamento das sequências que

não foram reproduzidas com até 1 nucleotídeo de diferença ($D(\mathbf{a}, \mathbf{b}) = 1$). Então analisamos todas as sequências da Tabela 5.1 com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$.

Curiosamente, nesse processo observamos alguns fatos importantes. No caso das sequências de DNA com comprimentos n iguais a 63, 255, 511, 1023 e 2047 identificadas pelos códigos BCH primitivos e reproduzidas com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença, observamos que todas as sequências analisadas foram reproduzidas em todos os polinômios primitivos de grau r e em todos os rotulamentos, como mostra as Tabelas 5.12, 5.13 e 5.14 para os comprimentos 63, 255 e 1023. Lembrando que, no caso das sequências de DNA com comprimentos submúltiplos de $n = p^r - 1$, ou seja, 21, 51 e 93, apenas algumas das sequências analisadas é que foram identificadas e reproduzidas em alguns polinômios primitivos e em alguns rotulamentos, mostradas nas Tabelas 5.6 e 5.7 (ver também a Tabela 5.20).

Nesse processo, a quantidade de palavras-código, resultados das simulações das sequências identificadas nos códigos BCH primitivos com $D(\mathbf{a}, \mathbf{b}) = 2$ é fundamental para identificarmos quais polinômios e rotulamentos serão capazes de reproduzir as sequências com $D(\mathbf{a}, \mathbf{b}) = 1$, e ainda, quais polinômios que serão capazes de identificar a estrutura de corpo nessas sequências, ou seja, identificar o polinômio gerador em $GF(4)$ que será capaz de reproduzir a sequência no corpo.

Com isso, através das simulações em sequências de DNA com comprimentos $n = p^r - 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ sobre $GR(4, r)$, podemos afirmar que:

- a) Todas as sequências de DNA identificadas através dos códigos BCH primitivos com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$ nucleotídeos de diferença serão reproduzidas em todos os polinômios primitivos de grau r e em todos os rotulamentos A, B e C;
- b) É possível identificar se as sequências de DNA serão identificadas com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$ nucleotídeo de diferença, qual polinômio e qual rotulamento;
- c) É possível identificar se as sequências possuem estrutura algébrica de corpo, ou seja, se elas serão identificadas e reproduzidas também em $GF(4)$ e, o polinômio.

Veja nas Tabelas 5.12, 5.13 e 5.14, os números que estão na parte superior de cada quadradinho correspondem a quantidade de palavras-código identificadas pelos códigos BCH primitivos nos correspondentes polinômios $p(x)$ e nos rotulamentos A, B e C; e os números que estão na parte inferior de cada quadradinho correspondem a quantidade de sequências de DNA reproduzidas sem erros de aminoácidos. Os quadradinhos na cor cinza servem apenas para identificarmos rapidamente quais sequências foram reproduzidas sem erros de aminoácidos, com $D(\mathbf{a}, \mathbf{b}) = 1$ e sobre $GF(4)$. As sequências Seq.38, Seq.39 e Seq.40 não fabricam proteínas: por isso não tem os números na parte inferior de cada quadradinho.

| Seq. nº | Códigos: G-linearidade ((63, 57, 3) BCH primitivo, rotulamentos A, B e C sobre $GR(4, 6)$) | | | | | | | | | | | | | | | | | | |
|------------|---|--------|---------|--------------|---------|--------|--------------|---------|---------|--------------|---------|----------|--------------|---------|--------|--------------|---------|---------|--------|
| | p01(x) | | | p02(x) | | | p03(x) | | | p04(x) | | | p05(x) | | | p06(x) | | | |
| | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | |
| | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C | |
| 12 | 5 0 | 3 0 | 3 0 | 5 1 | 3 0 | 5 0 | 3 0 | 5 0 | 5 1 | 5 1 | 3 0 | 5 0 | 63 62 | 5 0 | 5 0 | 3 1 | 3 0 | 3 0 | |
| 13 | 5 1 | 5 0 | 5 0 | 3 0 | 3 0 | 3 0 | 5 0 | 5 0 | 3 0 | 3 0 | 5 0 | 5 0 | 3 0 | 93 4 | 5 1 | 5 0 | 3 1 | 3 0 | 3 0 |
| 14 | 5 0 | 5 0 | 5 0 | 3 0 | 3 0 | 5 0 | 5 2 | 5 1 | 3 0 | 3 1 | 5 0 | 5 0 | 63 63 | 5 0 | 5 0 | 3 0 | 3 0 | 3 0 | |
| 15 | 5 0 | 5 0 | 3 0 | 3 0 | 3 0 | 5 0 | 5 0 | 5 1 | 5 0 | 5 0 | 3 0 | 63 63 | 3 0 | 5 1 | 5 0 | 3 1 | 3 0 | 3 1 | |
| 16 | 5 1 | 3 1 | 5 0 | 3 1 | 5 0 | 3 0 | 3 3 | 5 0 | 5 1 | 5 0 | 3 0 | 3 1 | 3 0 | 3 0 | 5 0 | 3 0 | 3 0 | 3 0 | |
| 17 | 5 0 | 5 0 | 5 0 | 5 1 | 3 0 | 3 0 | 3 0 | 5 0 | 3 0 | 5 0 | 3 0 | 5 3 | 3 0 | 3 1 | 3 0 | 5 0 | 5 0 | 93 3 | |
| 18 | 3 0 | 3 0 | 5 0 | 5 0 | 5 1 | 5 0 | 5 0 | 3 0 | 5 1 | 3 0 | 3 0 | 5 0 | 3 0 | 3 0 | 3 0 | 5 0 | 3 0 | 5 1 | |
| 19 | 3 0 | 5 0 | 3 0 | 3 1 | 5 0 | 3 0 | 3 0 | 93 2 | 3 0 | 5 0 | 3 0 | 5 0 | 3 0 | 3 0 | 5 0 | 5 0 | 5 0 | 5 0 | |
| 20 | 5 1 | 5 0 | 3 0 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 0 | 3 0 | 3 0 | 5 1 | 5 1 | 5 0 | |
| 21 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 3 0 | 5 1 | 3 0 | 3 0 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 3 0 | |
| 22 | 5 1 | 3 0 | 5 0 | 3 1 | 3 0 | 3 0 | 3 1 | 3 3 | 1 | 3 0 | 3 0 | 5 0 | 5 0 | 5 1 | 3 0 | 5 0 | 3 0 | 3 0 | |
| 23 | 5 1 | 3 0 | 5 0 | 5 0 | 3 0 | 3 0 | 3 0 | 1 | 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 3 | 5 1 | 5 0 | 5 1 | 3 0 | |
| 24 | 5 2 | 5 1 | 5 0 | 3 0 | 5 0 | 5 2 | 5 0 | 3 0 | 5 0 | 5 0 | 63 0 | 5 0 | 3 0 | 5 0 | 5 0 | 3 0 | 3 0 | 5 0 | |
| 25 | 5 3 | 5 0 | 5 0 | 3 0 | 3 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 0 | 5 0 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 3 0 | |
| 26 | 3 1 | 5 0 | 3 0 | 3 0 | 1 | 0 | 3 0 | 3 0 | 5 1 | 3 0 | 5 0 | 3 0 | 3 0 | 1 | 0 | 5 0 | 3 0 | 3 0 | |
| 27 | 3 1 | 5 0 | 5 0 | 5 1 | 3 0 | 3 0 | 3 0 | 0 | 1 | 5 0 | 3 0 | 3 0 | 3 0 | 0 | 0 | 5 2 | 3 0 | 3 0 | |
| 28 | 5 0 | 3 0 | 5 0 | 5 0 | 3 1 | 5 0 | 3 0 | 5 0 | 63 1 | 3 0 | 3 0 | 3 0 | 3 0 | 1 | 0 | 3 1 | 3 0 | 5 0 | |
| 29 | 3 0 | 5 0 | 3 0 | 5 0 | 3 0 | 5 0 | 3 0 | 3 0 | 3 0 | 3 0 | 3 0 | 3 0 | 5 0 | 3 0 | 3 0 | 63 1 | 3 0 | 3 0 | |
| 30 | 5 3 | 3 0 | 3 0 | 5 1 | 3 0 | 5 0 | 3 0 | 5 1 | 3 0 | 5 0 | 3 0 | 3 0 | 3 0 | 3 0 | 5 1 | 3 0 | 3 0 | 5 0 | |
| 31 | 5 0 | 5 0 | 5 0 | 3 0 | 5 1 | 3 0 | 93 2 | 5 0 | 5 0 | 3 0 | 5 0 | 3 0 | 3 0 | 1 | 0 | 3 0 | 93 3 | 3 0 | |
| 36 | 5 1 | 5 0 | 93 2 | 5 0 | 63 1 | 5 0 | 3 0 | 5 0 | 5 0 | 93 3 | 3 1 | 5 0 | 3 0 | 5 0 | 5 0 | 3 0 | 5 0 | 5 1 | |
| 37 | 3 0 | 3 0 | 5 0 | 5 0 | 5 1 | 5 0 | 3 0 | 0 | 0 | 5 0 | 5 0 | 5 0 | 5 1 | 0 | 0 | 5 1 | 3 0 | 5 0 | |
| 38 | 3 0 | 5 3 | 5 0 | 5 3 | 3 | 5 3 | 3 | 3 | 3 | 3 63 | 3 | 3 | 3 | 5 | 93 | 3 | 3 | 3 | |
| 39 | 5 0 | 3 3 | 5 5 | 5 5 | 3 | 3 | 5 3 | 3 | 3 | 3 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | |
| 40 | 5 0 | 5 5 | 5 5 | 3 3 | 5 0 | 3 3 | 3 1 | 3 0 | 5 0 | 5 3 | 3 | 5 | 3 | 3 | 3 | 5 3 | 3 | 3 | |

Tabela 5.12: Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros.

Observe, na Tabela 5.12, a sequência Seq.38 gerada no polinômio $p_4(x)$ - rotulamento B com 63 palavras-código - disso confirmamos que essa sequência também foi gerada com $D(\mathbf{a}, \mathbf{b}) = 1$ no mesmo polinômio e no mesmo rotulamento, já mostrada na Tabela 5.4.

Veja ainda, a sequência Seq.38 gerada no polinômio $p_5(x)$ - rotulamento C com 93 palavras-código, onde confirmamos que essa sequência também foi gerada com $D(\mathbf{a}, \mathbf{b}) = 1$ no mesmo polinômio e no mesmo rotulamento, mostrada na Tabela 5.4, e, também foi gerada no corpo de Galois $GF(4^3)$, mostrada Tabela 5.2.

| Códigos: G-linearidade ((255, 247, 3) BCH primitivo, rotulamentos A, B e C sobre $GR(4, 8)$) | | | | | | | | | | | | | | | |
|---|--------------|---|---|--------------|-----|-----|--------------|-----|-----|--------------|---|---|--------------|---|-----|
| Pol. $p(x)$ | Seq. 34 | | | Seq. 41 | | | Seq. 42 | | | Seq. 43 | | | Seq. 44 | | |
| | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | |
| | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C |
| $p_{01}(x)$ | 3 | 3 | 5 | 5 | 3 | 3 | 5 | 3 | 3 | 3 | 5 | 3 | 3 | 3 | 3 |
| $p_{02}(x)$ | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 255 | 255 | 5 | 5 | 5 | 5 | 3 | 3 |
| $p_{03}(x)$ | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 3 | 5 | 3 | 5 | 3 | 3 | 3 | 3 |
| $p_{04}(x)$ | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 5 | 3 | 5 | 5 | 3 |
| $p_{05}(x)$ | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 3 | 3 |
| $p_{06}(x)$ | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 3 | 3 | 3 | 3 | 5 | 3 | 3 | 3 |
| $p_{07}(x)$ | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 5 | 5 | 3 | 5 | 5 | 3 | 3 | 5 |
| $p_{08}(x)$ | 3 | 5 | 3 | 3 | 5 | 5 | 5 | 5 | 3 | 3 | 5 | 3 | 3 | 5 | 3 |
| $p_{09}(x)$ | 5 | 3 | 3 | 5 | 5 | 381 | 3 | 5 | 3 | 3 | 3 | 5 | 5 | 3 | 5 |
| $p_{10}(x)$ | 5 | 5 | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 5 | 3 |
| $p_{11}(x)$ | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 5 | 5 | 3 | 3 | 3 | 3 | 5 | 5 |
| $p_{12}(x)$ | 3 | 5 | 5 | 5 | 3 | 5 | 3 | 3 | 3 | 5 | 3 | 3 | 3 | 5 | 255 |
| $p_{13}(x)$ | 5 | 3 | 3 | 5 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 3 | 3 | 3 | 5 |
| $p_{14}(x)$ | 3 | 5 | 5 | 3 | 5 | 3 | 5 | 5 | 3 | 3 | 5 | 5 | 5 | 5 | 5 |
| $p_{15}(x)$ | 3 | 5 | 3 | 3 | 5 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 3 |
| $p_{16}(x)$ | 3 | 5 | 3 | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 3 | 3 | 3 |
| Pol. $p(x)$ | Seq. 45 | | | Seq. 46 | | | Seq. 47 | | | Seq. 48 | | | Seq. 49 | | |
| | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | |
| | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C |
| $p_{01}(x)$ | 3 | 5 | 5 | 3 | 381 | 5 | 5 | 3 | 3 | 5 | 3 | 5 | 5 | 3 | 3 |
| $p_{02}(x)$ | 3 | 5 | 3 | 3 | 3 | 5 | 3 | 5 | 3 | 3 | 5 | 3 | 5 | 5 | 3 |
| $p_{03}(x)$ | 5 | 3 | 3 | 3 | 5 | 5 | 5 | 3 | 3 | 5 | 3 | 5 | 3 | 3 | 5 |
| $p_{04}(x)$ | 3 | 5 | 3 | 3 | 3 | 3 | 3 | 5 | 5 | 5 | 5 | 5 | 3 | 5 | 3 |
| $p_{05}(x)$ | 5 | 5 | 3 | 5 | 3 | 3 | 255 | 3 | 5 | 5 | 3 | 3 | 3 | 3 | 3 |
| $p_{06}(x)$ | 5 | 3 | 3 | 3 | 5 | 3 | 5 | 5 | 3 | 5 | 5 | 3 | 3 | 5 | 5 |
| $p_{07}(x)$ | 3 | 3 | 3 | 5 | 3 | 3 | 3 | 3 | 5 | 5 | 3 | 5 | 3 | 3 | 3 |
| $p_{08}(x)$ | 3 | 3 | 3 | 5 | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 5 |
| $p_{09}(x)$ | 5 | 3 | 3 | 3 | 5 | 3 | 3 | 3 | 5 | 3 | 5 | 5 | 3 | 3 | 3 |
| $p_{10}(x)$ | 3 | 5 | 5 | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 5 | 5 | 3 | 3 | 3 |
| $p_{11}(x)$ | 3 | 3 | 5 | 255 | 3 | 3 | 5 | 5 | 3 | 3 | 5 | 5 | 5 | 5 | 5 |
| $p_{12}(x)$ | 3 | 3 | 3 | 3 | 5 | 3 | 5 | 5 | 3 | 3 | 5 | 5 | 3 | 3 | 3 |
| $p_{13}(x)$ | 3 | 5 | 3 | 5 | 3 | 3 | 5 | 3 | 3 | 5 | 5 | 3 | 3 | 3 | 5 |
| $p_{14}(x)$ | 5 | 3 | 3 | 3 | 3 | 3 | 3 | 5 | 3 | 3 | 5 | 3 | 3 | 5 | 5 |
| $p_{15}(x)$ | 3 | 4 | 5 | 5 | 3 | 5 | 3 | 5 | 5 | 3 | 5 | 3 | 5 | 3 | 3 |
| $p_{16}(x)$ | 3 | 4 | 3 | 5 | 5 | 3 | 5 | 5 | 3 | 5 | 3 | 5 | 3 | 5 | 5 |

Tabela 5.13: Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 8)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros.

Assim, o número 255 na Tabela 5.13 indica o polinômio e o rotulamento no qual a sequência foi reproduzida também com $D(\mathbf{a}, \mathbf{b}) = 1$ e o número 381 indica a reprodução da sequência com $D(\mathbf{a}, \mathbf{b}) = 1$ e também sobre $GF(4^4)$. Veja a sequência Seq.41 reproduzida no polinômio $p_{09}(x)$ - rotulamento C com $D(\mathbf{a}, \mathbf{b}) = 2$, também foi reproduzida no mesmo polinômio e mesmo rotulamento com $D(\mathbf{a}, \mathbf{b}) = 1$, mostrado na Tabela 5.4, e, reproduzida na estrutura de corpo mostrado na Tabela 5.2. O mesmo pode ser observado com as sequências Seq.42, Seq.44, Seq.46 e Seq.47. Na Tabela 5.14, a reprodução da sequência com $D(\mathbf{a}, \mathbf{b}) = 1$

Códigos: G-linearidade ((1023, 1013, 3) BCH primitivo, rotulamentos A, B e C sobre $GR(4, 10)$)

| Pol. $p(x)$ | Seq. 51 | | | Seq. 53 | | | Seq. 54 | | | Seq. 55 | | | Seq. 58 | | |
|----------------|--------------|---|------|--------------|------|---|--------------|---|---|--------------|---|------|--------------|------|---|
| | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | | Rotulamentos | | |
| | A | B | C | A | B | C | A | B | C | A | B | C | A | B | C |
| $p_{04}(x)$ | 3 | 3 | 3 | 5 | 3 | 5 | 1533 | 5 | 5 | 5 | 3 | 5 | 5 | 5 | 3 |
| $p_{05}(x)$ | 3 | 5 | 5 | 5 | 3 | 3 | 3 | 5 | 5 | 3 | 5 | 3 | 3 | 1533 | 5 |
| $p_{42}(x)$ | 5 | 3 | 1023 | 5 | 1023 | 5 | 3 | 5 | 3 | 5 | 3 | 5 | 3 | 5 | 3 |
| $p_{60}(x)$ | 5 | 3 | 5 | 5 | 3 | 5 | 5 | 3 | 5 | 3 | 5 | 1023 | 5 | 3 | 5 |

 Tabela 5.14: Sequências de DNA reproduzidas pelo código BCH primitivo sobre $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$ erros.

é indicada por 1023 e 1533, sendo que os casos de 1533 indica que a sequência também possui a estrutura de corpo, ou seja, ela será reproduzida sobre $GF(4^5)$. As sequências Seq.51, Seq.53, Seq.54, Seq.55 e Seq.58 foram mencionadas na Tabela 5.4 com $D(\mathbf{a}, \mathbf{b}) = 1$. As sequências Seq.54 e Seq.58 foram mencionadas na Tabela 5.2 indicando que as mesmas possuem a estrutura algébrica de corpo.

Note que os números: 63, 93, 255, 381, 1023 e 1533 estão relacionados entre si, da seguinte maneira:

- $93 - 63 = 30$ corresponde a cardinalidade do grupo cíclico do grupo $GR^*(4, 4)$;
- $381 - 255 = 126$ corresponde a cardinalidade do grupo cíclico do grupo $GR^*(4, 6)$;
- $1533 - 1023 = 510$ corresponde a cardinalidade do grupo cíclico do grupo $GR^*(4, 8)$;

Portanto, o número correspondente a quantidade de palavras-código, onde, através dele identificamos se uma sequência de DNA possui também a estrutura de corpo, $GF(4)$ é dado por

$$\psi = n + [2(2^{r-2} - 1)],$$

onde n é o comprimento da sequência de DNA e r é igual ao grau da extensão de Galois.

5.3.5 Representação algébrica das sequências sobre $GR(4, r)$

Aqui, mostramos que os códigos BCH primitivos sobre $GR(4, r)$ também foram capazes de representar algebricamente as sequências de DNA sob a caracterização em termos da dupla hélice do DNA e das fitas simples do DNA com os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$, como exemplo da sequência e direcionamento Seq.13 mostrado nas Figuras 5.5 e 5.6, respectivamente.

Assim como no $GF(4)$, durante o processo, tanto para $D(\mathbf{a}, \mathbf{b}) = 1$ quanto para $D(\mathbf{a}, \mathbf{b}) = 2$, consideramos: a) A identificação e reprodução das bases complementares da dupla hélice do

DNA através do polinômio primitivo $p_{05}(x)$ e do polinômio gerador $g_{05}(x)$; b) A identificação e reprodução dos nucleotídeos das fitas simples do DNA, tais como: 1º) A reprodução da fita simples na direção 5'-3' através dos mesmos polinômios $p_{05}(x)$ e $g_{05}(x)$; 2º) A reprodução da fita simples na direção 3'-5' (fita complementar) também através dos mesmos polinômios $p_{05}(x)$ e $g_{05}(x)$; 3º) A reprodução da fita não codante na direção 5'-3' através dos polinômios recíprocos $p'(x)$ e $g'(x)$; 4º) A reprodução do mRNA novamente através dos polinômios $p_{05}(x)$ e $g_{05}(x)$.

Seq.13 | *I.batatas* - Mitocôndria - F1-ATPase delta subunit - GI número 217937

Dupla hélice no NCBI:

$$D(a,b)=1$$

5'-ATG TTC AGG CAC TCT TCT CGA CTC CTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3'-TAC AAG TCC GTG AGA AGA GCT GAG GAT CGA GCG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'

a) Reprodução da dupla hélice do DNA: $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1 = g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$

Código G-linearidade ((63,57,3) BCH primitivo sobre GR(4,6), Rotulamento B: (0,1,2,3) – (AT,CG,GC,TA))

b) Reprodução das fitas simples do DNA:

Código G-linearidade ((63,57,3) BCH primitivo sobre GR(4,6), Rotulamento B: (0,1,2,3) – (A,C,G,T))

1º) Fita simples na direção 5'-3' reproduzida pelo código: $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ - $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$

5'-ATG TTC AGG CAC TCT TCT CGA **TTC** CTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
 032 331 022 101 313 313 120 **331** 130 213 121 211 011 010 032 222 322 123 121 111 331

2º) Fita simples na direção 3'-5' reproduzida pelo código: $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ – $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$

3'-TAC AAG TCC GTG AGA AGA GCT **AAG** GAT CGA GCG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'
 301 002 311 232 020 020 213 **002** 203 120 212 122 322 323 301 111 011 210 212 222 002

3º) Fita não codante: $p'(x) = x^6 + x^4 + x^3 + x + 1 - g'(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1$

5'-GAA GGG GCG ACG CCA CCC CAT TGT GGT GGC GCG AGC TAG GAA TCG AGA AGA GTG CCT GAA CAT-3'
 200 222 212 012 110 111 103 323 223 221 212 021 302 200 312 020 020 232 113 200 103

$$4^{\circ}) \text{ Transcrição-mRNA: } p_{05}(x) = x^3 + ax^2 + ax + a - g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$$

5'-AUG UUC AGG CAC UCU UCU CGA **UUC** CUA GCU CGC GCC ACC ACA AUG GGG UGG CGU CGC CCC UUC-3'
 032 331 022 101 313 313 120 **331** 130 213 121 211 011 010 032 222 322 123 121 111 331

Tradução (mapeamento casado – tRNA e rRNA)

Proteina:

M F R H S S R F L A R A T T M G W R R P E

Figura 5.5: Representação algébrica da dupla hélice e das fitas simples de DNA da sequência Seq.13 no $GR(4, 6)$ reproduzida com $D(\mathbf{a}, \mathbf{b}) = 1$.

Primeiramente, observe o caso $D(\mathbf{a}, \mathbf{b}) = 1$. Note que os polinômios $p_{05}(x)$ e $g_{05}(x)$ foram capazes de identificar e reproduzir tanto a dupla hélice do DNA quanto a fita codante e a fita complementar, todas com um nucleotídeo de diferença da sequência original (identificado pela cor vermelha) na mesma posição. Seguindo os rotulamentos A, B, C e A', B' e C' mostrados

nas Figuras 5.3 e 5.4, respectivamente, observe que não existem diferenças entre a reprodução da dupla hélice e das fitas simples do DNA. Em seguida, observe que os polinômios recíprocos $p'(x)$ e $g'(x)$ dos polinômios $p_{05}(x)$ e $g_{05}(x)$ foram capazes de identificar e reproduzir a fita não codante. Finalmente, a sequência é transcrita no mRNA sendo alterado o nucleotídeo timina (T) pela uracila (U), o que matematicamente não tem diferença, e, traduzida na sequência de aminoácidos. Veja que o tRNA realiza o mapeamento casado entre cada um dos códons nesta sequência com os correspondentes aminoácidos. Observe que na posição da oitava trinca ocorreu uma alteração do nucleotídeo citosina (C) para a timina (T)/uracila(U) acarretando uma alteração de aminoácido nesta posição, (L8F). O mesmo procedimento foi adotado no caso $D(\mathbf{a}, \mathbf{b}) = 2$ com os mesmos polinômios $p(x)$ e $g(x)$, veja figura abaixo.

Seq.13| I.batatas - Mitocôndria - F1-AT Pase delta subunit – GI número 217937

Dupla hélice no NCBI:

D (\mathbf{a}, \mathbf{b})=2

```

5'-ATG TTC AGG CAC TCT TCT CGA CTC CTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||
3'-TAC AAG TCC GTG AGA AGA GCT GAG GAT CGA CGG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'
      32 1 022 103 313 313 120 131 330 213 121 211 011 010 032 222 322 123 121 111 331
    
```

a) Reprodução da dupla hélice do DNA: $p_{05}(x)=x^6+x^5+x^3+x^2+1$ – $g_{05}(x)=x^6+3x^5+x^3+x^2+2x+1$

Código G-linearidade ((63,57,3) BCH primitivo sobre GR(4,6) , Rotulamento B: (0,1,2,3) – (AT,CG,GC,TA))

```

5'-ATG TTC AGG CAT TCT TCT CGA CTC TTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
      |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||
3'-TAC AAG TCC GTA AGA AGA GCT GAG AAT CGA CGG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'
      32 1 022 103 313 313 120 131 330 213 121 211 011 010 032 222 322 123 121 111 331
    
```

b) Reprodução das fitas simples do DNA:

Código G-linearidade ((63,57,3) BCH primitivo sobre GR(4,6) , Rotulamento B: (0,1,2,3) – (A,C,G,T))

1º) Fita simples na direção 5'-3' reproduzida pelo código: $p_{05}(x)=x^6+x^5+x^3+x^2+1$ – $g_{05}(x)=x^6+3x^5+x^3+x^2+2x+1$

```

5'-ATG TTC AGG CAT TCT TCT CGA CTC TTA GCT CGC GCC ACC ACA ATG GGG TGG CGT CGC CCC TTC-3'
      032 331 022 103 313 313 120 131 330 213 121 211 011 010 032 222 322 123 121 111 331
    
```

2º) Fita simples na direção 3'-5' reproduzida pelo código: $p_{05}(x)=x^6+x^5+x^3+x^2+1$ – $g_{05}(x)=x^6+3x^5+x^3+x^2+2x+1$

```

3'-TAC AAG TCC GTA AGA AGA GCT GAG AAT CGA CGG CGG TGG TGT TAC CCC ACC GCA GCG GGG AAG-5'
      032 331 022 103 313 313 120 131 330 213 121 211 011 010 032 222 322 123 121 111 331
    
```

3º) Fita não codante: $p'(x)=x^6+x^4+x^3+x+1$ – $g'(x)=x^6+2x^5+x^4+x^3+3x+1$

```

5'-GAA GGG GCG ACG CCA CCC CAT TGT GGT GGC GCG AGC TAA GAG TCG AGA AGA ATG CCT GAA CAT-3'
      200 222 212 012 110 111 103 323 223 221 212 021 300 202 312 020 020 032 113 200 103
    
```

4º) Transcrição-mRNA: $p_{05}(x)=x^3+ax^2+ax+a$ – $g_{05}(x)=x^6+x^5+x^3+x^2+1$

```

5'-AUG UUC AGG CAU UCU UCU CGA CUC UUA GCU CGC GCC ACC ACA AUG GGG UGG CGU CGC CCC UUC-3'
      032 331 022 103 313 313 120 131 330 213 121 211 011 010 032 222 322 123 121 111 331
    
```

Tradução (mapeamento casado – tRNA e rRNA)

Proteína:

```

M   M   F   R   H   S   S   R   L   L   A   R   A   T   T   M   G   W   R   R   P   F

```

Figura 5.6: Representação algébrica da dupla hélice e das fitas simples de DNA da sequência Seq.13 no $GR(4,6)$ reproduzida com $D(\mathbf{a}, \mathbf{b}) = 2$.

Como resultado os polinômios $p_{05}(x)$ e $g_{05}(x)$ identificaram e reproduziram a dupla hélice

e as fitas codante e complementar e, os polinômios recíprocos $p'(x)$ e $g'(x)$ dos polinômios $p_{05}(x)$ e $g_{05}(x)$ identificaram e reproduziram a fita não codante. Observe na Figura 5.6, que a quarta e a nona trinca sofreram alterações na dupla hélice e nas fitas, em termos das bases complementares e dos nucleotídeos, respectivamente, porém os correspondentes aminoácidos dessas trincas não sofreram alterações.

Essa representação algébrica é válida para todas as sequências que foram identificadas e reproduzidas em anel indicadas na Tabela 5.1. Portanto, todas as sequências identificadas e reproduzidas neste trabalho, tanto no corpo $GF(4^r)$ quanto no anel $GR(4, r)$ com os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$, a representação algébrica em termos da dupla hélice do DNA e das fitas simples do DNA foi satisfeita, onde os polinômios primitivos $p(x)$ e $g(x)$ foram capazes de identificar e reproduzir as duplas hélices e as fitas codantes e complementar do DNA, e, os polinômios recíprocos $p'(x)$ e $g'(x)$ foram capazes de identificar e reproduzir as fitas não codantes do DNA.

5.4 Reprodução de Sequências de DNA envolvidas em Patologias Clínicas

No presente trabalho, dentre as sequências de DNA identificadas e reproduzidas através dos códigos G -linearidade, temos várias sequências relacionadas a vírus e bactérias causadoras de patologias clínicas, Tabela 5.1. Dentre elas, temos uma sequência de direcionamento (SD) de uma das variações da proteína *H1N1* do vírus *Influenza A* e três sequências referente as proteínas p53 e BCRA1 relacionadas ao câncer. Uma vez observada e determinada a estrutura algébrica nas sequências de DNA, futuramente, metodologias podem ser direcionadas voltadas em análises mutacionais e de polimorfismos, tanto em patologias clínicas quanto nos diversos ramos da biotecnologia, reduzindo tempo e custos laboratoriais, conforme descrito em [21].

5.4.1 Sequência de direcionamento do vírus *Influenza A*

Influenza A subtipo *H1N1* também conhecido como *A(H1N1)*, é um subtipo de *Influenzavirus A* e a causa mais comum da *influenza* (gripe) em humanos. A letra *H* refere-se à proteína *hemaglutinina* e a letra *N* à proteína *neuraminidase*. Este subtipo deu origem, por mutação, a várias estirpes, incluindo a da gripe espanhola (atualmente extinta), estirpes moderadas de gripe humana, estirpes endêmicas de gripe suína e várias estirpes encontradas em aves. Segundo a Organização Mundial da Saúde (OMS), em maio de 2009, o número de pessoas contaminadas pelo vírus e as mortes registradas no mundo foi assustador.

A Figura 5.7, mostra a SD do vírus *Influenza A*, Seq.11, reproduzida pelo código *G-linearidade*: Klein-linearidade ((51, 43, 3) BCH não primitivo sobre $GR(4, 8)$, rotulamento C) através do polinômio primitivo $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ e do polinômio gerador $g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$ com dois nucleotídeos de diferença da sequência do NCBI. As trocas de nucleotídeos nas posições das trincas 6 e 12, ocasionaram nas trocas dos aminoácidos nestas posições, (L6V) e (F12L), respectivamente.

Seq.11| Influenza A - hemagglutinin H1N1 (1998) - vírus – GI número 305125

| Código klein-linearidade((51,43,3) BCH não primitivo sobre GR(4,8), rotulamento C) | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_{05}(x)=x^8+x^6+x^5+x^2+1 - g_{05}(x)=x^8+3x^7+2x^6+x^4+x^3+x^2+1 - \text{Caso 3} - (A,C,G,T)=(0,2,1,3)$ | | | | | | | | | | | | | | | | | | | |
| aaO: M K A K L L V L L C A F T A T D A ntO: ATG AAA GCA AAA CTA CTA GTC CTG TTA TGT GCA TTT ACA GCT ACA GAT GCA RtO: 031 000 120 000 230 230 132 231 330 313 120 333 020 123 020 103 120 RtG: 031 000 120 000 230 130 132 231 330 313 120 233 020 123 020 103 120 ntG: ATG AAA GCA AAA CTA GTA GTC CTG TTA TGT GCA CTT ACA GCT ACA GAT GCA aaG: M K A K L V V L L C A L T A T D A | | | | | | | | | | | | | | | | | | | |

Figura 5.7: Sequência de DNA de sequência de direcionamento com 51 nucleotídeos e $D(a, b) = 2$ no rotulamento C.

5.4.2 Proteínas p53 e BCRA1 - Câncer

Pesquisadores do Hospital A.C. Camargo, em São Paulo, vêm estudando um tipo de mutação que cria explosão de câncer raro no Brasil. A entrevista da médica e pesquisadora Maria Isabel W. Achatz ao Jornal Folha de São Paulo em 15/11/2009 conclui que o problema remonta, de fato, a um ancestral comum - provavelmente um tropeiro que deixou descendentes país afora no século XVIII.

Segundo Achatz, certa parcela de tumores da região Sul e Sudeste do país, que ainda não se sabe qual é, está ligada a essa mutação. A alteração no DNA, típica de algumas das famílias do Sul e Sudeste, se encaixa num conjunto mais amplo de mutações ligadas a formas severas de câncer. Esse grupo maior, conhecido como síndrome de *Li-Fraumeni*, se caracteriza por vários tumores na mesma pessoa - de mama, do cérebro e da glândula suprarrenal, por exemplo (antes dos 45 anos de idade).

Os cânceres da síndrome de *Li-Fraumeni* têm a mesma causa: mutações no trecho de DNA que carrega a receita para a produção da proteína p53. Essa proteína, apelidada de “guardiã do genoma”, tem como principal função justamente impedir os erros da cópia do DNA que levam ao surgimento do câncer. As alterações do gene da p53 que produzem a síndrome de *Li-Fraumeni* são raras, atingindo uma a cada 5.000 pessoas. Mas, segundo Achatz, em 2001 percebeu que o número de pacientes era bem maior do que o esperado,

logo desconfiou que estava acontecendo alguma coisa estranha. Resumindo, Achatz e seus colegas analisaram 12 famílias com essa mutação, em princípio sem relação de parentesco. Resultado: todas carregavam o mesmo conjunto de 29 trocas de nucleotídeos no gene da p53. “*A chance de todas essas trocas acontecerem juntas em famílias diferentes é baixíssima, o melhor jeito é imaginar que todas herdaram o conjunto típico de alterações de um ancestral comum distante*”, disse Achatz em entrevista ao jornal.

O próximo passo para os pesquisadores do Hospital A.C. Camargo é tentar estimar a data de origem da mutação, em colaboração com a *University College* de Londres. Os pesquisadores defendem que valeria a pena testar a mutação de maneira mais ampla na população, para enfrentar esse tipo de câncer com a maior precocidade possível.

Em [20], um estudo de filogenia em colaboração com o Dr Marcelo M. Brandão do laboratório de Biologia Molecular de Plantas da ESALQ/USP, para testar a hipótese de que a proteína *Malate dehydrogenase 1* gerada pelo código⁴ é uma ancestral da sequência disponível no *NCBI*. Essa hipótese decorreu da interpretação para a diferença de nucleotídeo apresentada em todas as sequências identificadas e reproduzidas em [20] e neste trabalho. No contexto biológico, esse descasamento (“*mismatch*”) é conhecido como polimorfismo de um único nucleotídeo (SNP’s - *single nucleotide polymorphism*). Diante deste fato, sugerimos que as sequências do *NCBI* são SNP’s das sequências reproduzidas pelos códigos \mathbb{Z}_4 -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade, ou vice-versa. Novos estudos de filogenia para outras proteínas identificadas e reproduzidas neste trabalho devem ser realizados com o objetivo de explicarmos os SNP’s identificados no processo de reprodução de sequências de DNA.

Motivados pelos resultados apresentados no processo da identificação e reprodução das sequências de DNA, no estudo de filogenia e na entrevista da médica Achatz, buscamos no *NCBI* por sequências de DNA envolvidas com câncer, com comprimentos $n = p^r - 1$ ou submútiplos de n . Existem centenas de sequências envolvidas com algum tipo de câncer, a dificuldade está em encontrarmos o comprimento que satisfaz as condições de n . Encontramos três sequências envolvidas com o câncer, sendo duas sequências da proteína BRCA1 (Seq.61 e Seq.62) e uma sequência da proteína p53 (Seq.63).

Como resultado das simulações para as proteínas BCRA1, o código $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade ((63, 57, 3) BCH primitivo sobre $GR(4, 6)$, rotulamento B) através dos polinômios $p_{04}(x)$ e $g_{04}(x)$ reproduziu-as com $D(\mathbf{a}, \mathbf{b}) = 1$. Note na Figura 5.8, que duas proteínas, Seq.61 e Seq.62, foram reproduzidas com o códon de parada *stop*. Note ainda que a fita não codante da Seq.62 também foi reproduzida com os códons de parada stop (TGA; TAA e TAG).

⁴Seq.51, mostrada na Figura 4.6, reproduzida com $D(\mathbf{a}, \mathbf{b}) = 1$ pelo código *G*-linearidade: Klein-linearidade ((1023, 1013, 3) BCH primitivo sobre $GR(4, 10)$, rotulamento C) através do polinômio primitivo $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ e do polinômio gerador $g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1$.

Seq.61| *H.sapiens* – BRCA1 {exon 23, internal fragment} [human, serous papillary ovarian adenocarcinoma, patient sample 61, Genomic Mutant] – GI número 1000568

Fita codante: 5' – 3'

Código $Z_2 \times Z_2$ -linearidade((63,57,3) BCH primitivo sobre GR(4,6), rotulamento B)

$$p_{04}(x) = x^6 + x^5 + x^2 + x + 1 - g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$$

| | P | D | P | G | Q | R | T | M | A | S | M | Q | L | G | R | C | V | R | H | L | W |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | P | D | P | G | Q | R | T | M | A | S | M | Q | L | G | R | C | V | R | H | L | W |
| ntO: | CCA | GAT | CCT | GGA | CAG | AGG | ACA | ATG | GCT | TCC | ATG | CAA | TTG | GCG | AGA | TGT | GTG | AGG | CAC | CTG | TGG |
| RtO: | 110 | 203 | 113 | 220 | 102 | 022 | 010 | 032 | 213 | 311 | 032 | 100 | 332 | 221 | 020 | 323 | 232 | 022 | 101 | 132 | 322 |
| RtG: | 110 | 203 | 113 | 220 | 102 | 022 | 010 | 032 | 213 | 311 | 032 | 100 | 332 | 221 | 320 | 323 | 232 | 022 | 101 | 132 | 322 |
| ntG: | CCA | GAT | CCT | GGA | CAG | AGG | ACA | ATG | GCT | TCC | ATG | CAA | TTG | GCG | TGA | TGT | GTG | AGG | CAC | CTG | TGG |
| aaG: | P | D | P | G | Q | R | T | M | A | S | M | Q | L | G | sto | C | V | R | H | L | W |

Fita não codante: 3' – 5'

Código $Z_2 \times Z_2$ -linearidade((63,57,3) BCH primitivo sobre GR(4,6), rotulamento B)

$$p'(x) = x^6 + x^5 + x^4 + x + 1 - g'(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ntO: | CCA | CAG | GTG | CCT | CAC | ACA | TCT | GCC | CAA | TTG | CAT | GGA | AGC | CAT | TGT | CCT | CTG | TCC | AGG | ATC | TGG |
| RtO: | 110 | 102 | 232 | 113 | 101 | 010 | 313 | 211 | 100 | 332 | 103 | 220 | 021 | 103 | 323 | 113 | 132 | 311 | 022 | 031 | 322 |
| RtG: | 110 | 102 | 232 | 113 | 101 | 010 | 310 | 211 | 100 | 332 | 103 | 220 | 021 | 103 | 323 | 113 | 132 | 311 | 022 | 031 | 322 |
| ntG: | CCA | CAG | GTG | CCT | CAC | ACA | TCA | GCC | CAA | TTG | CAT | GGA | AGC | CAT | TGT | CCT | CTG | TCC | AGG | ATC | TGG |

Seq.62| *H.sapiens* – truncated breast and ovarian cancer susceptibility protein (BRCA1) gene – GI número 25140445

Fita codante: 5' – 3'

Código $Z_2 \times Z_2$ -linearidade((63,57,3) BCH primitivo sobre GR(4,6), rotulamento B)

$$p_{01}(x) = x^6 + x^5 + 1 - g_{01}(x) = x^6 + 3x^5 + 2x^3 + 1$$

| | E | A | A | S | G | C | E | S | E | T | S | V | S | E | D | C | S | G | L | S | E |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | E | A | A | S | G | C | E | S | E | T | S | V | S | E | D | C | S | G | L | S | E |
| ntO: | GAA | GCA | GCA | TCT | GGG | TGT | GAG | AGT | GAA | ACA | AGC | GTC | TCT | GAA | GAC | TGC | TCA | GGG | CTA | TCA | GAG |
| RtO: | 200 | 210 | 210 | 313 | 222 | 323 | 202 | 023 | 200 | 010 | 021 | 231 | 313 | 200 | 201 | 321 | 310 | 222 | 130 | 310 | 202 |
| RtG: | 200 | 210 | 210 | 313 | 222 | 320 | 202 | 023 | 200 | 010 | 021 | 231 | 313 | 200 | 201 | 321 | 310 | 222 | 130 | 310 | 202 |
| ntG: | GAA | GCA | GCA | TCT | GGG | TGA | GAG | AGT | GAA | ACA | AGC | GTC | TCT | GAA | GAC | TGC | TCA | GGG | CTA | TCA | GAG |
| aaG: | E | A | A | S | G | sto | E | S | E | T | S | V | S | E | D | C | S | G | L | S | E |

Fita não codante: 3' – 5'

Código $Z_2 \times Z_2$ -linearidade((63,57,3) BCH primitivo sobre GR(4,6), rotulamento B)

$$p'(x) = x^6 + x^5 + x^4 + x + 1 - g'(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ntO: | CTC | TGA | TAG | CCC | TGA | GCA | GTC | TTC | AGA | GAC | GCT | TGT | TTC | ACT | CTC | ACA | CCC | AGA | TGC | TGC | TTC |
| RtO: | 131 | 320 | 302 | 111 | 320 | 210 | 231 | 331 | 020 | 201 | 213 | 323 | 331 | 013 | 131 | 010 | 111 | 020 | 321 | 321 | 331 |
| RtG: | 131 | 320 | 302 | 111 | 320 | 210 | 231 | 331 | 020 | 201 | 213 | 323 | 331 | 013 | 131 | 310 | 111 | 020 | 321 | 321 | 331 |
| ntG: | CTC | TGA | TAG | CCC | TGA | GCA | GTC | TTC | AGA | GAC | GCT | TGT | TTC | ACT | CTC | TCA | CCC | AGA | TGC | TGC | TTC |

Figura 5.8: Proteína envolvida no câncer de ovário.

De todas as análises feitas neste trabalho e em [20], nenhuma sequência foi reproduzida em anel com $D(\mathbf{a}, \mathbf{b}) = 1$ com o códon *stop*, como no caso das sequências Seq.61 e Seq.62. Inferimos que os *stops* possam estar relacionados com os SNP's, nesses casos do câncer de ovário, ou até mesmo, indicando que a sequência ancestral já apresentava problemas. Isso é apenas uma hipótese que deve ser investigada. Já, as simulações feitas na proteína p53, mostraram que o código Kein-linearidade ((1023,1013,3) BCH primitivo sobre $GR(r, 10)$, rotulamento C) foi capaz de identificar e reproduzir a proteína com $D(\mathbf{a}, \mathbf{b}) = 1$ na posição da trinca 188 sem alteração de aminoácido, através dos polinômios $p_{07}(x) = x^{10} + x^4 + x^3 + x + 1$

e $g_{07}(x) = x^{10} + 2x^7 + 2x^5 + x^4 + 3x^3 + 3x + 1$, como mostra a Figura 5.9. Veja que a alteração do nucleotídeo é uma mutação silenciosa.

O fato de terem comprimentos dentro da capacidade de restrição dos códigos geradores, $n = (2^r - 1)$, faz com que essas sequências são reproduzidas através dos G -linearidade: códigos BCH primitivos. Como já mencionamos, por serem códigos BCH primitivos, isso implica que todos os polinômios primitivos e todos os rotulamentos serão capazes de identificar e reproduzir essas sequências.

Seq.63| *H.sapiens* – p53 beta isoform (TP53) mRNA, complete cds – GI número 75914680

| Código klein-linearidade(1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento C | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_{07}(x) = x^{10} + x^7 + 2x^5 + x^4 + 3x^3 + 3x + 1$ – $g(x) = x^{10} + 2x^5 + x^4 + 3x^3 + 3x + 1$ | | | | | | | | | | | | | | | | | | | | | | | | | |
| Caso 03 – (A,C,G,T) = (0,2,1,3) | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M E E P Q S D P S V E P P L S Q E T F S D ntO: ATG GAG GAG CCG CAG TCA GAT CCT AGC GTC GAG CCC CCT CTG AGT CAG GAA ACA TTT TCA GAC Rto: 031 101 101 221 201 320 103 223 012 132 101 222 223 231 013 201 100 020 333 320 102 RtG: 031 101 101 221 201 320 103 223 012 132 101 222 223 231 013 201 100 020 333 320 102 rtG: ATG GAG GAG CCG CAG TCA GAT CCT AGC GTC GAG CCC CCT CTG AGT CAG GAA ACA TTT TCA GAC aaG: M E E P Q S D P S V E P P L S Q E T F S D | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: L W K L L P E N V L S P P L P S Q A M D Ont: CTA TGG AAA CTA CTT CCT GAA AAC AAC GTT CTG TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT Rto: 230 311 000 230 233 223 100 002 002 133 231 322 222 331 221 322 200 120 031 103 103 RtG: 230 311 000 230 233 223 100 002 002 133 231 322 222 331 221 322 200 120 031 103 103 rtG: CTA TGG AAA CTA CTT CCT GAA AAC AAC GTT CTG TCC CCC TTG CCG TCC CAA GCA ATG GAT GAT aaG: L W K L L P E N V L S P P L P S Q A M D D | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: L M L S P D D I E Q W F T E D P G P D E A ntO: TTG ATG CTG TCC CCG GAC GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAA GCT Rto: 331 031 231 322 221 102 103 033 100 200 311 332 023 100 102 220 113 220 103 100 123 RtG: 331 031 231 322 221 102 103 033 100 200 311 332 023 100 102 220 113 220 103 101 123 rtG: TTG ATG CTG TCC CCG GAC GAT ATT GAA CAA TGG TTC ACT GAA GAC CCA GGT CCA GAT GAG GCT aaG: L M L S P D D I E Q W F T E D P G P D E A | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: P R M P E A A P R V A P A P A A P T P A A A ntO: CCC AGA ATG CCA GAG GCT GCT CCC CGC GTG GCC CCG CCT GCA CCA GCA GCT CCT ACA CCG GCG GCC Rto: 222 010 031 220 101 123 123 222 212 131 122 223 120 220 120 123 223 020 221 121 122 RtG: 222 010 031 220 101 123 123 222 212 131 122 223 120 220 120 123 223 020 221 121 122 rtG: CCC AGA ATG CCA GAG GCT GCT CCC CGC GTG GCC CCG CCT GCA CCA GCA GCT CCT ACA CCG GCG GCC aaG: P R M P E A A P R V A P A P A A P T P A A A | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: P A P A P S W P L S S S V P S Q K T Y Q G ntO: CCT GCA CCA GCC CCC TCC TGG CCC CTG TCA TCT TCT GTC CCT TCC CAG AAA ACC TAC CAG GGC Rto: 223 120 220 122 222 322 311 222 231 320 323 323 132 223 322 201 000 022 302 201 112 RtG: 223 120 220 122 222 322 311 222 231 320 323 323 132 223 322 201 000 022 302 201 112 rtG: CCT GCA CCA GCC CCC TCC TGG CCC CTG TCA TCT TCT GTC CCT TCC CAG AAA ACC TAC CAG GGC aaG: P A P A P S W P L S S S V P S Q K T Y Q G | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: S Y G F R L G F L H S G T A K S V T C T Y ntO: AGC TAC GGT TTC CGT CTG GGC TTC TTG CAT TCT GGG ACA GCC AAG TCT GTG ACT TGC ACG TAC Rto: 012 302 113 332 213 231 112 332 331 203 323 111 020 122 001 323 131 023 312 021 302 RtG: 012 302 113 332 213 231 112 332 331 203 323 111 020 122 001 323 131 023 312 021 302 rtG: AGC TAC GGT TTC CGT CTG GGC TTC TTG CAT TCT GGG ACA GCC AAG TCT GTG ACT TGC ACG TAC aaG: S Y G F R L G F L H S G T A K S V T C T Y | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: S P A L N K M F C Q L A K T C P V Q L W V ntO: TCC CCT GCC CTC AAC AAG ATG TTT TGC CAA CTG GCC AAG ACC TGC CCT GTG CAG CTG TGG GTT Rto: 322 223 122 232 002 001 031 333 312 200 231 122 001 022 312 223 131 201 231 311 133 Glb: 322 223 122 232 002 001 031 333 312 200 231 122 001 022 312 223 131 201 231 311 133 Gnt: TCC CCT GCC CTC AAC AAG ATG TTT TGC CAA CTG GCC AAG ACC TGC CCT GTG CAG CTG TGG GTT Gaa: S P A L N K M F C Q L A K T C P V Q L W V | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: D S T P P G T R V R A M A I Y K Q S Q H ntO: GAT TCC ACA CCC CCG CCC GGC ACC CGC GTC GCC ATG GCC ATC TAC AAG CAG TCA CAG CAC Rto: 103 322 020 222 221 222 112 022 212 132 212 122 031 122 032 302 001 201 320 201 202 Glb: 103 322 020 222 221 222 112 022 212 132 212 122 031 122 032 302 001 201 320 201 202 Gnt: GAT TCC ACA CCC CCG CCC GGC ACC CGC GTC GCC ATG GCC ATC TAC AAG CAG TCA CAG CAC Gaa: D S T P P G T R V R A M A I Y K Q S Q H | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M T E V V R R C P H H E R C S D S D G L A ntO: ATG ACG GAG GTT GTG AGG CGC TGC CCC CAC CAT GAG CGC TGC TCA GAT AGC GAT GGT CTG GCC Rto: 031 021 101 133 131 011 212 312 222 202 203 101 212 312 320 103 012 103 113 231 122 Glb: 031 021 101 133 131 011 212 312 222 202 203 101 212 312 320 103 012 103 113 231 122 Gnt: ATG ACG GAG GTT GTG AGG CGC TGC CCC CAC CAT GAG CGC TGC TCA GAT AGC GAT GGT CTG GCC Gaa: M T E V V R R C P H H E R C S D S D G L A | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | P | P | Q | H | L | I | R | V | E | G | N | L | R | V | E | Y | L | D | D | R | N |
| ntO: | CCT | CCT | CAG | CAC | CTT | ATC | CGA | GTG | GAA | GGA | AAT | TTG | CGT | GTG | GAG | TAT | TTG | GAT | GAC | AGA | AAC |
| RtO: | 223 | 223 | 201 | 202 | 233 | 032 | 210 | 131 | 100 | 110 | 003 | 331 | 213 | 131 | 101 | 303 | 331 | 103 | 102 | 010 | 002 |
| RtG: | 223 | 223 | 201 | 202 | 233 | 032 | 210 | 131 | 100 | 110 | 003 | 331 | 213 | 131 | 101 | 303 | 331 | 103 | 102 | 010 | 002 |
| ntG: | CCT | CCT | CAG | CAC | CTT | ATC | CGA | GTG | GAA | GGA | AAT | TTG | CGT | GTG | GAG | TAT | TTG | GAT | GAC | AGA | AAC |
| aaG: | P | P | Q | H | L | I | R | V | E | G | N | L | R | V | E | Y | L | D | D | R | N |
| aaO: | T | F | R | H | S | V | V | V | P | Y | E | P | P | E | V | G | S | D | C | T | T |
| ntO: | ACT | TTT | CGA | CAT | AGT | GTG | GTG | GTG | CCC | TAT | GAG | CCG | CCT | GAG | GTT | GTC | TCT | GAC | TGT | ACC | ACC |
| RtO: | 023 | 333 | 210 | 203 | 013 | 131 | 131 | 131 | 222 | 303 | 101 | 221 | 223 | 101 | 133 | 112 | 323 | 102 | 313 | 022 | 022 |
| RtG: | 023 | 333 | 210 | 203 | 013 | 131 | 131 | 131 | 222 | 303 | 101 | 221 | 223 | 101 | 133 | 112 | 323 | 102 | 313 | 022 | 022 |
| ntG: | ACT | TTT | CGA | CAT | AGT | GTG | GTG | GTG | CCC | TAT | GAG | CCG | CCT | GAG | GTT | GTC | TCT | GAC | TGT | ACC | ACC |
| aaG: | T | F | R | H | S | V | V | V | P | Y | E | P | P | E | V | G | S | D | C | T | T |
| aaO: | I | H | Y | N | Y | M | C | N | S | S | C | M | G | G | M | N | R | R | P | I | L |
| ntO: | ATC | CAC | TAC | AAC | TAC | ATG | TGT | AAC | AGT | TCC | TGC | ATG | GGC | GGC | ATG | AAC | CGG | AGG | CCC | ATC | CTC |
| RtO: | 032 | 202 | 302 | 002 | 302 | 031 | 313 | 002 | 013 | 322 | 312 | 031 | 112 | 112 | 031 | 002 | 211 | 011 | 222 | 032 | 232 |
| RtG: | 032 | 202 | 302 | 002 | 302 | 031 | 313 | 002 | 013 | 322 | 312 | 031 | 112 | 112 | 031 | 002 | 211 | 011 | 222 | 032 | 232 |
| ntG: | ATC | CAC | TAC | AAC | TAC | ATG | TGT | AAC | AGT | TCC | TGC | ATG | GGC | GGC | ATG | AAC | CGG | AGG | CCC | ATC | CTC |
| aaG: | I | H | Y | N | Y | M | C | N | S | S | C | M | G | G | M | N | R | R | P | I | L |
| aaO: | T | I | I | T | L | E | D | S | S | G | N | L | L | G | R | N | S | F | E | V | R |
| ntO: | ACC | ATC | ATC | ACA | CTG | GAA | GAC | TCC | AGT | GGT | AAT | CTA | CTG | GGA | CGG | AAC | AGC | TTT | GAG | GTG | CGT |
| RtO: | 022 | 032 | 032 | 020 | 231 | 100 | 102 | 322 | 013 | 113 | 003 | 230 | 231 | 110 | 211 | 002 | 012 | 333 | 101 | 131 | 213 |
| RtG: | 022 | 032 | 032 | 020 | 231 | 100 | 102 | 322 | 013 | 113 | 003 | 230 | 231 | 110 | 211 | 002 | 012 | 333 | 101 | 131 | 213 |
| ntG: | ACC | ATC | ATC | ACA | CTG | GAA | GAC | TCC | AGT | GGT | AAT | CTA | CTG | GGA | CGG | AAC | AGC | TTT | GAG | GTG | CGT |
| aaG: | T | I | I | T | L | E | D | S | S | G | N | L | L | G | R | N | S | F | E | V | R |
| aaO: | V | C | A | C | P | G | R | D | R | R | T | E | E | E | N | L | R | K | K | G | E |
| ntO: | GTT | TGT | GCC | TGT | CCT | GGG | AGA | GAC | CGG | CGC | ACA | GAG | GAA | GAG | AAT | CTC | CGC | AAG | AAA | GGG | GAG |
| RtO: | 133 | 313 | 122 | 313 | 223 | 111 | 010 | 102 | 211 | 212 | 020 | 101 | 100 | 101 | 003 | 232 | 212 | 001 | 000 | 111 | 101 |
| RtG: | 133 | 313 | 122 | 313 | 223 | 111 | 010 | 102 | 211 | 212 | 020 | 101 | 100 | 101 | 003 | 232 | 212 | 001 | 000 | 111 | 101 |
| ntG: | GTT | TGT | GCC | TGT | CCT | GGG | AGA | GAC | CGG | CGC | ACA | GAG | GAA | GAG | AAT | CTC | CGC | AAG | AAA | GGG | GAG |
| aaG: | V | C | A | C | P | G | R | D | R | R | T | E | E | E | N | L | R | K | K | G | E |
| aaO: | P | H | H | E | L | P | P | G | S | T | K | R | A | L | P | N | N | T | S | S | S |
| ntO: | CCT | CAC | CAC | GAG | CTG | CCC | CCA | GGG | AGC | ACT | AAG | CGA | GCA | CTG | CCC | AAC | AAC | ACC | AGC | TCC | TCT |
| RtO: | 223 | 202 | 202 | 101 | 231 | 222 | 220 | 111 | 012 | 023 | 001 | 210 | 120 | 231 | 222 | 002 | 002 | 022 | 012 | 322 | 323 |
| RtG: | 223 | 202 | 202 | 101 | 231 | 222 | 220 | 111 | 012 | 023 | 001 | 210 | 120 | 231 | 222 | 002 | 002 | 022 | 012 | 322 | 323 |
| ntG: | CCT | CAC | CAC | GAG | CTG | CCC | CCA | GGG | AGC | ACT | AAG | CGA | GCA | CTG | CCC | AAC | AAC | ACC | AGC | TCC | TCT |
| aaG: | P | H | H | E | L | P | P | G | S | T | K | R | A | L | P | N | N | T | S | S | S |
| aaO: | P | Q | P | K | K | P | L | D | G | E | Y | F | T | L | Q | D | Q | T | S | F | |
| ntO: | CCC | CAG | CCA | AAG | AAG | AAA | CCA | CTG | GAT | GGA | GAA | TAT | TTC | ACC | CTT | CAG | GAC | CAG | ACC | AGC | TTT |
| RtO: | 222 | 201 | 220 | 001 | 001 | 000 | 220 | 231 | 103 | 110 | 100 | 303 | 332 | 022 | 233 | 201 | 102 | 201 | 022 | 012 | 333 |
| RtG: | 222 | 201 | 220 | 001 | 001 | 000 | 220 | 231 | 103 | 110 | 100 | 303 | 332 | 022 | 233 | 201 | 102 | 201 | 022 | 012 | 333 |
| ntG: | CCC | CAG | CCA | AAG | AAG | AAA | CCA | CTG | GAT | GGA | GAA | TAT | TTC | ACC | CTT | CAG | GAC | CAG | ACC | AGC | TTT |
| aaG: | P | Q | P | K | K | P | L | D | G | E | Y | F | T | L | Q | D | Q | T | S | F | |
| aaO: | Q | K | E | N | C | | | | | | | | | | | | | | | | |
| ntO: | CAA | AAA | GAA | AAT | TGT | | | | | | | | | | | | | | | | |
| RtO: | 200 | 000 | 100 | 003 | 313 | | | | | | | | | | | | | | | | |
| RtG: | 200 | 000 | 100 | 003 | 313 | | | | | | | | | | | | | | | | |
| ntG: | CAA | AAA | GAA | AAT | TGT | | | | | | | | | | | | | | | | |
| aaG: | Q | K | E | N | C | | | | | | | | | | | | | | | | |

 Figura 5.9: Sequência de DNA da proteína p53 com $D(\mathbf{a}, \mathbf{b}) = 1$.

Logo para cada sequência com comprimento 63 nucleotídeos, temos no mínimo 18 sequências de DNA diferentes e, para a sequência com 1023 nucleotídeos, temos no mínimo 180 sequências diferentes. Com o auxílio de estudos mutacionais e filogenéticos as diferenças de nucleotídeos nessas sequências podem ser estudadas.

Mais sequências envolvidas em câncer devem ser futuramente estudadas, para que as interpretações sejam também mais concretas e apropriadas. Acreditamos que a colaboração com biólogos, médicos, engenheiros, matemáticos, bioinformáticos, proporcionará ganhos em várias frentes de estudos relacionadas não somente às patologias do câncer, mas, em outras

patologias também. Esforços em busca de colaborações futuras devem ser dispendidos nesse sentido.

5.5 Reprodução de um Genoma Circular e de um Gene Humano

No presente trabalho, mostramos que os códigos G -linearidade foram capazes de identificar, reproduzir e classificar matematicamente várias sequências de DNA com funcionalidades biológicas específicas e comprimentos variados nos contextos de codificação genética (fita simples do DNA) e codificação genômica (dupla hélice do DNA).

Desde o início da pesquisa, estávamos interessados em reproduzir um genoma. No entanto, o comprimento de sequências de genomas são relativamente grandes - o que aumenta significativamente a complexidade computacional. Apenas os genomas circulares de bactéria, plasmídeos e vírus possuem comprimentos menores. Porém, ainda nos deparamos com a dificuldade em encontrar no *NCBI* essas sequências menores e que estejam dentro da restrição de comprimento do código $n = 2^r - 1$. As mesmas dificuldades também se apresentam para encontrar sequências de genes. Na busca dessas sequências contamos com a colaboração do Prof. Dr. Michel E. B. Yamagishi (Embrapa Informática Agropecuária) e do Prof. Dr. Roberto H. Herai (*Postdoctoral fellowship - University of California San Diego*).

Pela primeira vez sequências de um genoma plasmidial e de um gene humano são identificadas como palavras-código de códigos G -linearidade e reproduzidas em termos das fitas simples do DNA e das duplas hélices, como ilustrado na Figura 5.10. Com isso, identificamos a existência de uma estrutura matemática, algébrica e geométrica, associada a essas sequências permitindo a sua identificação, reprodução e classificação.

Tanto a sequência do genoma quanto a sequência do gene foram reproduzidas através dos códigos G -linearidade: ((2047, 2036, 3) BCH primitivo sobre $GR(4, 11)$, rotulamento B) e ((511, 502, 3) BCH primitivo sobre $GR(4, 9)$, rotulamento C), respectivamente. A construção destes códigos não foi detalhada no presente trabalho, uma vez que é análoga a construção dos códigos BCH primitivos: (63, 57, 3) sobre $GR(4, 6)$, (255, 247, 3) sobre $GR(4, 8)$ e (1023, 1013, 3) sobre $GR(4, 10)$.

No processo de geração do genoma com 2047 nucleotídeos, existem 176 polinômios primativos que devem ser usados na extensão de Galois de grau $r = 11$ para a geração dos códigos. Dentro desses polinômios, apenas um polinômio primitivo $p_{157}(x) = x^{11} + x^{10} + x^7 + x^2 + 1$ associado ao polinômio gerador $g_{157}(x) = x^{11} + 3x^{10} + 2x^9 + x^7 + 2x^6 + 2x^5 + 3x^2 + 2x + 3$ foi capaz de identificar a sequência do genoma como palavra-código e reproduzi-la em termos da

fita simples do DNA e da dupla hélice do DNA, como mostrado na representação algébrica da Figura 5.11. Observe que os SNP's ocorrem na mesma posição, tanto na dupla hélice quanto nas fitas simples, conforme o esperado.

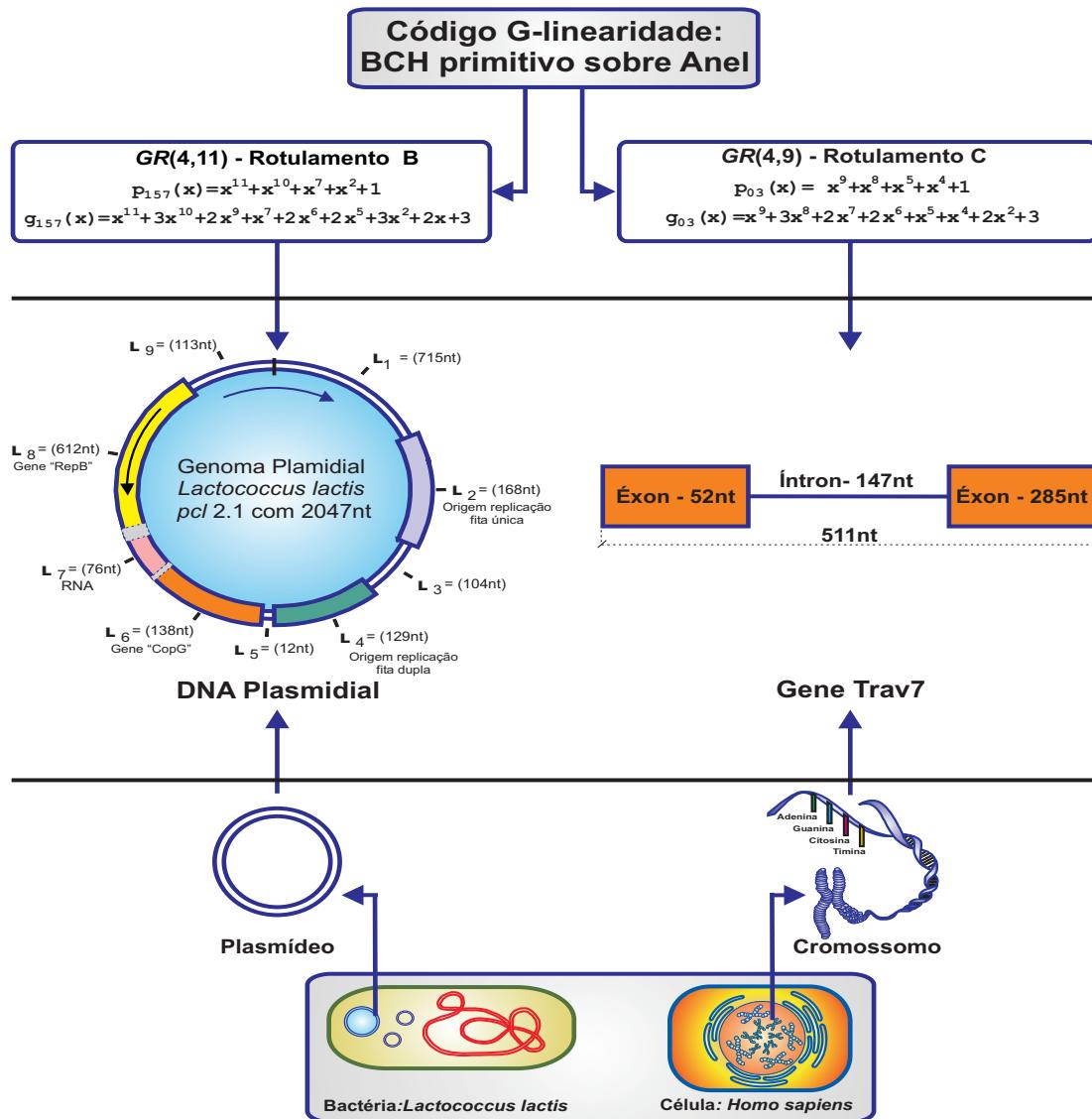


Figura 5.10: Reprodução do genoma plamidial e do gene “Trav7” e através dos códigos G-linearidade.

Na parte a) da figura mostramos a reprodução da dupla hélice indicando onde ocorreu uma troca das bases complementares adenina com a timina para citosina com guanina na posição 1547 (A-T 1547 C-G). Na parte b) mostramos a reprodução das fitas simples geradas separadamente, sendo a fita 5'-3' com a alteração do nucleotídeo adenina para citosina na mesma posição (A1547C) e, a fita 3'-5' com a alteração do nucleotídeo timina para guanina com alteração do nucleotídeo adenina para o nucleotídeo citosina na posição 1547 (A1547C).

Por uma questão de espaço, realçamos apenas a região do genoma onde ocorrem os SNP's (Figura 5.11). Note que a alteração de nucleotídeos na posição 1547 ocorreu no Gene "RebB" na região L_8 do genoma, como representado na Figura 5.10.

**Seq.65 | Homo sapiens - Lactococcus lactis plasmid pCL2.1, sequência completa
GI número: 118213250 – ref.: NC_004981.2**

Dupla hélice postada no NCBI:

```
1501 5' - attggcatat aataatttct ttttcaaata aaaaaaggtta taagcaagta ttaaaaattt - 3'
      ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |
      3' - taaccgtata ttattaaaga aaaagtttat tttttcaat attcggtcat aatttttaaa - 5'
```

a) Sequência da dupla hélice gerada pelo código:

Código $Z_2 \times Z_2$ -linearidade: ((2047,2036,3) - BCH primitivo sobre GR(4,11), Rotulamento B: (0,1,2,3)=(AT,CG,GC,TA)

$$p_{157}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad g_{157}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
1501 5' - attggcatat aataatttct ttttcaaata aaaaaaggtta taagcacgta ttaaaaattt - 3'
      ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| ||||||||| |
      3' - taaccgtata ttattaaaga aaaagtttat tttttcaat attcgtgcat aatttttaaa - 5'
```

Rto:0332210303 0030033313 3333100030 0000002330 3002100230 3300000333
Rtg:0332210303 0030033313 3333100030 0000002330 300210**1**230 3300000333

b) Sequência das fitas simples gerada pelo código:

Código $Z_2 \times Z_2$ -linearidade: ((2047,2036,3) - BCH primitivo sobre GR(4,11), Rotulamento B: (0,1,2,3)=(A,C,G,T)

$$1^{\circ}) \text{ Reprodução da fita } 5'-3': \quad p_{157}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad g_{157}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
1501 5' - attggcatat aataatttct ttttcaaata aaaaaaggtta taagcaagta ttaaaaattt - 3'
      Rto:0332210303 0030033313 3333100030 0000002330 3002100230 3300000333
      Rtg:0332210303 0030033313 3333100030 0000002330 3002101230 3300000333
      5' - attggcatat aataatttct ttttcaaata aaaaaaggtta taagcacgta ttaaaaattt - 3'
```

$$2^{\circ}) \text{ Reprodução da fita } 3'-5': \quad p_{157}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad g_{157}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
1501 3' - taaccgtata ttattaaaga aaaagtttat tttttcaat attcggtcat aatttttaaa - 5'
      Rto:3001123030 3303300020 0000233303 3333331003 0331233103 0033333000
      Rtg:3001123030 3303300020 0000233303 3333331003 0331232103 0033333000
      3' - taaccgtata ttattaaaga aaaagtttat tttttcaat attcgtgcat aatttttaaa - 5'
```

Figura 5.11: Representação algébrica do genoma do plasmídeo.

No processo de geração do gene com 511 nucleotídeos, existem 48 polinômios primitivos que devem ser usados na extensão de Galois de grau $r = 9$ para a geração dos códigos. Neste caso, apenas o polinômio primitivo $p_{03}(x) = x^9 + x^8 + x^5 + x^4 + 1$ associado ao polinômio gerador $g_{03}(x) = x^9 + 3x^8 + 2x^7 + 2x^6 + x^5 + x^4 + 2x^2 + 3$ foi capaz de identificar e reproduzir a sequência do gene "Trav 7" em termos da fita simples do DNA e da dupla hélice do DNA, como mostrado na representação algébrica da Figura 5.12.

Na parte a) da figura mostramos a reprodução da dupla hélice indicando onde ocorreu uma troca das bases complementares adenina com timina para guanina com citosina na

Seq.64 | *Homo sapiens* - Trav7 predicted gene - chromosome 14 genomic contig – 13263_NT_026437

Dupla hélice postada no NCBI:

```
121 5'- gatgttagtct ggacttaatt gtcattgctt gtttgagatt aagaataaaa attatgaaag - 3'  
      ||||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
      3'- ctacatcaga cctgaatgaa cagtaacgaa caaactctaa ttcttttattt taataacttcc - 5'
```

a) Sequência da dupla hélice gerada pelo código:

Código Klein-linearidade: ((511,502,3) - BCH primitivo sobre GR(4,9), Rotulamento C: (0,2,1,3)=(AT,CG,GC,TA)

$$p_{03}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad - \quad g_{03}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
121 5'- ggtgttagtct ggacttaatt gtcattgctt gtttgagatt aagaataaaa attatgaaag - 3'  
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||  
      3'- ccacatcaga cctgaatgaa cagtaacgaa caaactctaa ttcttttattt taataacttcc - 5'
```

RtO: 1031301323 1102330233 1320331233 1333101033 0010003000 0330310001
RtG: 1131301323 1102330233 1320331233 1333101033 0010003000 0330310001

b) Sequência das fitas simples gerada pelo código:

Código Klein-linearidade: ((511,502,3) - BCH primitivo sobre GR(4,9), Rotulamento C: (0,2,1,3)=(A,C,G,T)

$$1^{\circ}) \text{ Reprodução da fita } 5'-3': \quad p_{03}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad - \quad g_{03}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
121 5'- gatgttagtct ggacttaatt gtcattgctt gtttgagatt aagaataaaa attatgaaag - 3'  
RtO: 1031301323 1102330233 1320331233 1333101033 0010003000 0330310001  
RtG: 1131301323 1102330233 1320331233 1333101033 0010003000 0330310001  
5' - ggtgttagtct ggacttaatt gtcattgctt gtttgagatt aagaataaaa attatgaaag - 3'
```

$$2^{\circ}) \text{ Reprodução da fita } 3'-5': \quad p_{03}(x)=x^{11}+x^{10}+x^7+x^2+1 \quad - \quad g_{03}(x)=x^{11}+3x^{10}+2x^9+x^7+2x^6+2x^5+3x^2+2x+3$$

```
121 3'- ctacatcaga cctgaatgaa cagtaacgaa caaactctaa ttcttttattt taataacttcc - 5'  
RtO: 2302032010 2231003100 2013002100 2000232300 3323330333 3003023332  
RtG: 2202032010 2231003100 2013002100 2000232300 3323330333 3003023332  
3' - ccacatcaga cctgaatgaa cagtaacgaa caaactctaa ttcttttattt taataacttcc - 5'
```

Figura 5.12: Representação algébrica do gene “Trav 7”.

posição 122 (A-T 122 G-C). Na parte b) mostramos a reprodução das fitas simples geradas separadamente, sendo a fita 5'-3' com a alteração do nucleotídeo adenina para guanina na mesma posição 122 (A122G) e, a fita 3'-5' com a alteração do nucleotídeo timina para citocina (T121C).

Veja nas Figuras 5.13 e 5.14 a reprodução completa do gene “Trav 7” com 511 nucleotídeos e do genoma do plasmídeo *Lactococcus lactis* pcl 2.1 com 2047 nucleotídeos em termos da fita simples do DNA, respectivamente. Portanto, mostramos que a dupla hélice do DNA e as fitas simples são identificadas, reproduzidas e classificadas matematicamente através dos códigos *G*-linearidade.

Observe que as sequências do gene e do genoma foram identificadas e reproduzidas somente na estrutura algébrica sobre anel, uma vez que a estrutura algébrica sobre corpo não

foi capaz de reproduzi-las. Observe também que, como o código gerador dessas sequências é o BCH primitivo, temos que todos os polinômios primitivos e todos os rotulamentos são capazes de reproduzir essas sequências com o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$. No caso da sequência do gene teremos no mínimo 144 possibilidades de sequências reproduzidas diferindo em dois nucleotídeos do *NCBI* e, no caso do genoma teremos no mínimo 528 possibilidades. Essas possibilidades de sequências indicando as diferenças de nucleotídeos podem estar disponíveis para novos estudos em análises mutacionais.

Seq.64 | Homo sapiens - Trav7 predicted gene - chromosome 14 genomic contig - 13263_NT_026437

Código Klein-linearidade: BCH primitivo sobre GR(4,9), rotulamento C

$$p_{03}(x) = x^9 + x^8 + x^5 + x^4 + 1 \quad - \quad g_{03}(x) = x^9 + 3x^8 + 2x^7 + 2x^6 + x^5 + x^4 + 2x^2 + 3 \quad - \quad \text{Caso 03 (A,C,G,T)=(0,2,1,3)}$$

| | |
|-----|---|
| 1 | ntO: atggagaaga tgccgagacc tgtcctaatt atatttgc tatgtcttgg ctgttaagttg RtO: 0311010010 3121101022 3132230033 0303333132 3031323311 2313001331 RtG: 0311010010 3121101022 3132230033 0303333132 3031323311 2313001331 ntG: atggagaaga tgccgagacc tgtcctaatt atatttgc tatgtcttgg ctgttaagttg |
| 61 | ntO: agggttctaa gaactgggga ccccaggaga catttattca agtccttttg gggagatggg RtO: 0111332300 1002311110 2222011010 2033303320 0132233331 1110103111 RtG: 0111332300 1002311110 2222011010 2033303320 0132233331 1110103111 ntG: agggttctaa gaactgggga ccccaggaga catttattca agtccttttg gggagatggg |
| 121 | ntO: gatgttagtct ggacttactt gtcattgctt gtttgagatt aagaataaaa attatgaaag RtO: 1031301323 1102330233 1320331233 1333101033 0010003000 0330310001 RtG: 1131301323 1102330233 1320331233 1333101033 0010003000 0330310001 ntG: gggtgttagtct ggacttactt gtcattgctt gtttgagatt aagaataaaa attatgaaag |
| 181 | ntO: gtctaaatta aaatgtacat attgtacctg atgtctttct gaataggggc aaatggagaa RtO: 1323000330 0003130203 0331302231 0313233323 1003011112 0003110100 RtG: 1323000330 0003130203 0331302231 0313233323 1003011112 0003110100 ntG: gtctaaatta aaatgtacat attgtacctg atgtctttct gaataggggc aaatggagaa |
| 241 | ntO: aaccagggtgg agcacagccc tcattttctg ggaccaggc agggagacgt tgcctccatg RtO: 0022011311 0120201222 3203333231 1102222012 0111010213 3122322031 RtG: 0022011311 0120201222 3203333231 1102222012 0111010213 3122322031 ntG: aaccagggtgg agcacagccc tcattttctg ggaccaggc agggagacgt tgcctccatg |
| 301 | ntO: agctgcacgt actctgtcag tcgttttaac aatttgcagt ggtacaggca aaatacaggg RtO: 0123120213 0232313201 3213333002 0033312013 1130201120 0003020111 RtG: 0123120213 0232313201 3213333002 0033312013 1130201120 0003020111 ntG: agctgcacgt actctgtcag tcgttttaac aatttgcagt ggtacaggca aaatacaggg |
| 361 | ntO: atgggtcccc aacaccttatt atccatgtat tcagctggat atgagaagca gaaaggaaga RtO: 0311132220 0020223033 0322031303 3201231103 0310100120 1000110010 RtG: 0311132220 0020223033 0322031303 3201231103 0310100120 1000110010 ntG: atgggtcccc aacaccttatt atccatgtat tcagctggat atgagaagca gaaaggaaga |
| 421 | ntO: ctaaatgcta cattactgaa gaatggaagc agcttgtaca ttacagccgt gcagcctgaa RtO: 2300031230 2033023100 1003110012 0123313020 3302012213 1201223100 RtG: 2300031230 2033023100 1003110012 0123313020 3302012213 1201223100 ntG: ctaaatgcta cattactgaa gaatggaagc agcttgtaca ttacagccgt gcagcctgaa |
| 481 | ntO: gattcagcca cctatttctg tgctgttagat g RtO: 1033201220 2230333231 3123130103 1 RtG: 1033201220 2230333231 3123130103 1 ntG: gattcagcca cctatttctg tgctgttagat g |

Figura 5.13: Sequência do gene “Trav 7” reproduzida através do código *G*-linearidade.

Seq.65 | *Homo sapiens - Lactococcus lactis plasmid pCL2.1, sequência completa*
GI número: 118213250 – ref.: NC_004981.2

```

1 Ont: cctacatttt tttattgctc tgctatgatt gtttatcgat agtttttat acagataagc
Olb: 1130103333 3330332131 3213032033 2333031203 0233333303 0102030021
Glb: 1130103333 3330332131 3213032033 2333031203 0233333303 0102030021
Gnt: cctacatttt tttattgctc tgctatgatt gtttatcgat agtttttat acagataagc
61 Ont: gtgcgacgct tgctcttcc gaggaggaag tcatgtcgac aagcacggca gagcctccgc
Olb: 2321201213 3213133311 2022022002 3103213201 0021012210 2021131121
Glb: 2321201213 3213133311 2022022002 3103213201 0021012210 2021131121
Gnt: gtgcgacgct tgctcttcc gaggaggaag tcatgtcgac aagcacggca gagcctccgc
121 Ont: atgaaatgct ctaatgaaa ttgccggcg agcttttg agcttgtgcc acttgtcgaaa
Olb: 0320003213 1310032000 3321122122 0213333332 0213323211 0133212000
Glb: 0320003213 1310032000 3321122122 0213333332 0213323211 0133212000
Gnt: atgaaatgct ctaatgaaa ttgccggcg agcttttg agcttgtgcc acttgtcgaaa
181 Ont: aaaacaagaa caaaagagac agggaaactgt cttttttgc ttgcttgggg attggggcaa
Olb: 0000100200 1000020201 0220001323 1333333321 3321332222 0332222100
Glb: 0000100200 1000020201 0220001323 1333333321 3321332222 0332222100
Gnt: aaaacaagaa caaaagagac agggaaactgt cttttttgc ttgcttgggg attggggcaa
241 Ont: cgccccaaaa ataaaaagaa tcgtctgaaa cgaggaacaa actaaaatgt aaattttagt
Olb: 1211110000 0300000200 3123132000 1202200100 0130000323 0003333023
Glb: 1211110000 0300000200 3123132000 1202200100 0130000323 0003333023
Gnt: cgccccaaaa ataaaaagaa tcgtctgaaa cgaggaacaa actaaaatgt aaattttagt
301 Ont: tgttaccgag tggaaatgta atactttta acctatgtgt atacacacat agtaagctcg
Olb: 3233011202 3220020320 0301333330 0113032323 0301010103 0230021312
Glb: 3233011202 3220020320 0301333330 0113032323 0301010103 0230021312
Gnt: tgttaccgag tggaaatgta atactttta acctatgtgt atacacacat agtaagctcg
361 Ont: ctataatact ttataacgtt tttatttaca tgagcaagc gagttttcc aacacgttta
Olb: 1303003013 3303001233 3330333010 3202100021 2023333311 0010123330
Glb: 1303003013 3303001233 3330333010 3202100021 2023333311 0010123330
Gnt: ctataatact ttataacgtt tttatttaca tgagcaagc gagttttcc aacacgttta
421 Ont: atctaaaata ttggcaattt ataccatgat tttcatgtta tctaagtgcg cccttaggaa
Olb: 0313000030 3322100333 0301103203 3331032230 3230023212 1113302200
Glb: 0313000030 3322100333 0301103203 3331032230 3230023212 1113302200
Gnt: atctaaaata ttggcaattt ataccatgat tttcatgtta tctaagtgcg cccttaggaa
481 Ont: aataatttga atatatttca gatttcaat ctgactgtc ctgtcatcg gcagaccat
Olb: 0030033320 0303033310 2033331003 1320132131 1323103120 2102011203
Glb: 0030033320 0303033310 2033331003 1320132131 1323103120 2102011203
Gnt: aataatttga atatatttca gatttcaat ctgactgtc ctgtcatcg gcagaccat
541 Ont: gaggaaaaca aaaagaggac taaacaaaaa agtttagtcc tctttttgtt ttgaatagtt
Olb: 2022000010 0000202201 3000100000 0233302311 3133333233 3320030233
Glb: 2022000010 0000202201 3000100000 0233302311 3133333233 3320030233
Gnt: gaggaaaaca aaaagaggac taaacaaaaa agtttagtcc tctttttgtt ttgaatagtt
601 Ont: ctagaacgtc atattttgcg ttttaagcaa ttttgactaa ctaggcgggg atttttactt
Olb: 1302001231 0303333212 3333002100 3333201300 1302212222 0333330133
Glb: 1302001231 0303333212 3333002100 3333201300 1302212222 0333330133
Gnt: ctagaacgtc atattttgcg ttttaagcaa ttttgactaa ctaggcgggg atttttactt
661 Ont: agaaatttatt caaaacgtct gtaaaatgtct taaaatcggt tctaagagct tttagcggtt
Olb: 0200033033 1000012313 2300023213 3000031233 3130020213 3330212333
Glb: 0200033033 1000012313 2300023213 3000031233 3130020213 3330212333
Gnt: agaaatttatt caaaacgtct gtaaaatgtct taaaatcggt tctaagagct tttagcggtt
721 Ont: atttcggtta gttatcggtca taatcggtta aacaggcggtt atcgttagcg aaaagccctt
Olb: 0333123330 2330312210 3003123300 0010221233 0312302122 0000211133
Glb: 0333123330 2330312210 3003123300 0010221233 0312302122 0000211133
Gnt: atttcggtta gttatcggtca taatcggtta aacaggcggtt atcgttagcg aaaagccctt
781 Ont: gagcgttagcg tggcttgca gtgaagatgt tgcgtttagg attatgaaag ccgataactg
Olb: 2021230212 3221333210 2320020323 3231323302 0330320002 1120300132
Glb: 2021230212 3221333210 2320020323 3231323302 0330320002 1120300132
Gnt: gagcgttagcg tggcttgca gtgaagatgt tgcgtttagg attatgaaag ccgataactg
841 Ont: aatgaaataa taagcgttagc gccccttatt tcggtcggag gaggtcaag ggagtttgag
Olb: 0032000300 3002123021 2111133033 3122312202 2022131002 2202333202
Glb: 0032000300 3002123021 2111133033 3122312202 2022131002 2202333202
Gnt: aatgaaataa taagcgttagc gccccttatt tcggtcggag gaggtcaag ggagtttgag
901 Ont: ggaatgaaat tccctcatgg ttttaaaaatt gcttgcatt ttggcgagcg gtagcgctgg
Olb: 2200320003 3111310322 3333000033 2133210033 3321120212 2302121322
Glb: 2200320003 3111310322 3333000033 2133210033 3321120212 2302121322
Gnt: ggaatgaaat tccctcatgg ttttaaaaatt gcttgcatt ttggcgagcg gtagcgctgg
961 Ont: aaaatttttg aaaaaaaaaaattt ggaatttggaa aaaatggggg ggtactacga ccccccccta
Olb: 0000333332 0000000333 2200333220 0000322222 2230130120 1111111130
Glb: 0000333332 0000000333 2200333220 0000322222 2230130120 1111111130
Gnt: aaaatttttg aaaaaaaaaaattt ggaatttggaa aaaatggggg ggtactacga ccccccccta

```

| | | | | | | | |
|------|------|-------------|-------------|--------------|-------------|------------------------|-------------|
| 1021 | Ont: | tgtggtaatt | tggtaacttg | gtcaaaattg | atactaataat | atattaaaac | agcacaaaaac |
| | Olb: | 3232230033 | 3223001332 | 2310000332 | 0301300303 | 0303300001 | 0210100001 |
| | Glb: | 3232230033 | 3223001332 | 2310000332 | 0301300303 | 0303300001 | 0210100001 |
| | Gnt: | tgtggtaatt | tggtaacttg | gtcaaaattg | atactaataat | atattaaaac | agcacaaaaac |
| 1081 | Ont: | agaatcttat | gatataataa | gatatactga | aatttgaagg | agtaaaaaat | ggcagaagag |
| | Olb: | 0200313303 | 2030300300 | 2030301320 | 0033320022 | 0230000003 | 2210200202 |
| | Glb: | 0200313303 | 2030300300 | 2030301320 | 0033320022 | 0230000003 | 2210200202 |
| | Gnt: | agaatcttat | gatataataa | gatatactga | aatttgaagg | agtaaaaaat | ggcagaagag |
| 1141 | Ont: | aaaaaaagag | tttgctaac | tttgcgtt | gacaaggcg | aagaattaga | aactatatac |
| | Olb: | 0000000202 | 3333213001 | 3332312332 | 2010002102 | 0020033020 | 0013030310 |
| | Glb: | 0000000202 | 3333213001 | 3332312332 | 2010002102 | 0020033020 | 0013030310 |
| | Gnt: | aaaaaaagag | tttgctaac | tttgcgtt | gacaaggcg | aagaattaga | aactatatac |
| 1201 | Ont: | aaagaatgg | gaatttagtaa | atctgcttt | gttagttt | ggattgcgga | aaattctaga |
| | Olb: | 0002000322 | 2003302300 | 0313213133 | 2330233303 | 2203321220 | 0003313020 |
| | Glb: | 0002000322 | 2003302300 | 0313213133 | 2330233303 | 2203321220 | 0003313020 |
| | Gnt: | aaagaatgg | gaatttagtaa | atctgcttt | gttagttt | ggattgcgga | aaattctaga |
| 1261 | Ont: | aaataaaaaa | agagccacgg | cgaatggctc | tagtatattt | acggtagga | atattatagc |
| | Olb: | 0003000000 | 0202110122 | 1200322131 | 3023030333 | 0122330220 | 0303303021 |
| | Glb: | 0003000000 | 0202110122 | 1200322131 | 3023030333 | 0122330220 | 0303303021 |
| | Gnt: | aaataaaaaa | agagccacgg | cgaatggctc | tagtatattt | acggtagga | atattatagc |
| 1321 | Ont: | atatgacaga | aaaaaaacta | gaaaaaaaaatg | acccagttag | aaactggagt | tgggttgtt |
| | Olb: | 0303201020 | 0000000130 | 20000000032 | 0111023302 | 0001322023 | 3222332333 |
| | Glb: | 0303201020 | 0000000130 | 20000000032 | 0111023302 | 0001322023 | 3222332333 |
| | Gnt: | atatgacaga | aaaaaaacta | gaaaaaaaaatg | acccagttag | aaactggagt | tgggttgtt |
| 1381 | Ont: | atccagagtc | tgccctgaa | aatttggagaa | cattgtttaga | cgaacttggaa | gaaaaatggaa |
| | Olb: | 0311020231 | 3213113200 | 0033220200 | 1033233020 | 1200013220 | 2000003220 |
| | Glb: | 0311020231 | 3213113200 | 0033220200 | 1033233020 | 1200013220 | 2000003220 |
| | Gnt: | atccagagtc | tgccctgaa | aatttggagaa | cattgtttaga | cgaacttggaa | gaaaaatggaa |
| 1441 | Ont: | ttgagagtc | tttgcgtat | aaagatatta | acgaaacaac | aaacgaaccg | aaaaaggcac |
| | Olb: | 3320202311 | 2332103203 | 0002030330 | 0120001001 | 0001200112 | 0000022101 |
| | Glb: | 3320202311 | 2332103203 | 0002030330 | 0120001001 | 0001200112 | 0000022101 |
| | Gnt: | ttgagagtc | tttgcgtat | aaagatatta | acgaaacaac | aaacgaaccg | aaaaaggcac |
| 1501 | Ont: | attggcatat | aataattttc | tttccaaata | aaaaaagttt | taagcaagta | ttaaaaattt |
| | Olb: | 0332210303 | 0030033313 | 3333100030 | 0000002330 | 3002100230 | 3300000333 |
| | Glb: | 0332210303 | 0030033313 | 3333100030 | 0000002330 | 3002101230 | 3300000333 |
| | Gnt: | attggcatat | aataattttc | tttccaaata | aaaaaagttt | taagca cgta | ttaaaaattt |
| 1561 | Ont: | ctgaaatgtt | aatatgcacca | gagcctgttaa | aaacaaaaaaa | tttacaaggg | tcagttaat |
| | Olb: | 1320003233 | 0003210110 | 2021132300 | 0001000000 | 3330100222 | 3102331003 |
| | Glb: | 1320003233 | 0003210110 | 2021132300 | 0001000000 | 3330100222 | 3102331003 |
| | Gnt: | ctgaaatgtt | aatatgcacca | gagcctgttaa | aaacaaaaaaa | tttacaaggg | tcagttaat |
| 1621 | Ont: | atttggcga | cagaaacaat | cctgaaaaat | atcagtatata | taaaagcgat | tttgcgttc |
| | Olb: | 0333232210 | 1020001003 | 1132000003 | 0310230300 | 3000021203 | 2332332131 |
| | Glb: | 0333232210 | 1020001003 | 1132000003 | 0310230300 | 3000021203 | 2332332131 |
| | Gnt: | atttggcga | cagaaacaat | cctgaaaaat | atcagtatata | taaaagcgat | tttgcgttc |
| 1681 | Ont: | ataatgggtt | taaatataga | caatattttaa | cagatattgg | agttgatact | gattctattt |
| | Olb: | 0300322233 | 3000303020 | 1003033300 | 1020303322 | 0233203013 | 2033130333 |
| | Glb: | 0300322233 | 3000303020 | 1003033300 | 1020303322 | 0233203013 | 2033130333 |
| | Gnt: | ataatgggtt | taaatataga | caatattttaa | cagatattgg | agttgatact | gattctattt |
| 1741 | Ont: | tacaagaatgt | tatagaatgg | ataaaagaaaa | ctggatgttc | tgaatataga | gatttagtcg |
| | Olb: | 3010020023 | 3030200322 | 0300002000 | 1322032331 | 3200030320 | 2033302312 |
| | Glb: | 3010020023 | 3030200322 | 0300002000 | 1322032331 | 3200030320 | 2033302312 |
| | Gnt: | tacaagaatgt | tatagaatgg | ataaaagaaaa | ctggatgttc | tgaatataga | gatttagtcg |
| 1801 | Ont: | attatgcagt | atcagaacgt | ttcgatgatt | ggtttctac | agtcagaagt | caaaccatat |
| | Olb: | 0330321023 | 0310200123 | 3312032033 | 2233311301 | 0231020023 | 1000110303 |
| | Glb: | 0330321023 | 0310200123 | 3312032033 | 2233311301 | 0231020023 | 1000110303 |
| | Gnt: | attatgcagt | atcagaacgt | ttcgatgatt | ggtttctac | agtcagaagt | caaaccatat |
| 1861 | Ont: | tttttaattt | tttatttacgc | tcaaatcgtc | atagtcagaa | aaaatataat | ccagaaaacag |
| | Olb: | 3333000331 | 3303330121 | 3100031231 | 0302310200 | 0000303003 | 1102000102 |
| | Glb: | 3333000331 | 3303330121 | 3100031231 | 0302310200 | 0000303003 | 1102000102 |
| | Gnt: | tttttaattt | tttatttacgc | tcaaatcgtc | atagtcagaa | aaaatataat | ccagaaaacag |
| 1921 | Ont: | gagaggtgtt | atgaaatgtt | aaattatagc | tagtggtttt | agtgaaaaat | cagttcagaa |
| | Olb: | 2020223233 | 0320002332 | 0003303021 | 3023233333 | 0232000003 | 1023310200 |
| | Glb: | 2020223233 | 0320002332 | 0003303021 | 3023233333 | 0232000003 | 1023310200 |
| | Gnt: | gagaggtgtt | atgaaatgtt | aaattatagc | tagtggtttt | agtgaaaaat | cagttcagaa |
| 1981 | Ont: | aaaagtaaat | aattttatttg | attattttaa | tgacaataat | tttgaagttat | tttgaagttca |
| | Olb: | 0000230003 | 0033330332 | 0330333000 | 3201003003 | 3332002303 | 3220023310 |
| | Glb: | 0000230003 | 0033330332 | 0330333000 | 3201003003 | 3332002303 | 3220023310 |
| | Gnt: | aaaagtaaat | aattttatttg | attattttaa | tgacaataat | tttgaagttat | tttgaagttca |
| 2041 | Ont: | atataagg | | | | | |
| | Olb: | 0303022 | | | | | |
| | Glb: | 0303022 | | | | | |
| | Gnt: | atataagg | | | | | |

Figura 5.14: Sequência do genoma do plasmídeo *Lactococcus lactis* pcl 2.1 reproduzida através do código G-linearidade.

5.6 Sistemas Concatenados

Códigos concatenados têm o significado de um sistema que combina vários códigos em camadas. Por exemplo: primeiramente, uma mensagem de informação é codificada de acordo com alguns códigos. Em seguida, uma segunda mensagem de informação é acrescentada à palavra-código resultado da primeira mensagem codificada. A mensagem resultante é codificada novamente por um outro código. Este processo é repetido várias vezes. Nesse processo existe uma proteção eficaz das informações mais antigas. Isso significa que a multiplicidade de códigos fornece um grau muito maior de segurança do que cada um deles separadamente.

Uma ilustração simples e intuitiva da conjectura de Battail, usa a metáfora da fortaleza: um código é representado como um muro que protege o que está dentro dos invasores externos. Vários muros são sucessivamente construídos para proteger as informações separadamente, assim, as informações mais antiga e central é muito mais protegida do que a informação mais recente e periférica, ver Figura 5.15. Note que uma proteção muito eficiente da informação mais central não demanda códigos individuais muito eficientes: a multiplicidade dos muros é muito mais seguros do que cada um deles separadamente.

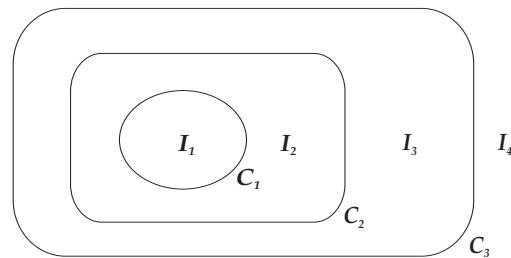


Figura 5.15: As quatro camadas de um sistema *nested* em [14].

O conceito de códigos concatenados surgiu da constatação de que certas partes do genoma, como a dos genes *Hox* são conservados com espantosa fidelidade em muitas espécies animais. Em desacordo com esta permanência extrema, no entanto, verifica-se que a variabilidade genômica tem algumas vantagens como a prova do sucesso evolutivo para a criação de novas combinações de alelos. Assim, é provável que a informação genômica é desigualmente protegida contra erros, de maneira simples e eficiente através dos códigos *nested*. Outro ponto importante, é assumir que os códigos apareceram sucessivamente no tempo, a informação genômica sendo a mais bem protegida, de modo que a variabilidade refere-se essencialmente às camadas mais periféricas do sistema de códigos concatenados, veja declaração a seguir de Battail em [14]: “*The subsidiary hypothesis is intended to account for this fact. It assumes that the genomic error-correcting code actually consists of a combination of codes which were*

successively laid down during the geological ages. Every time a new coding layer has been appended, the information inside it has acquired a much smaller probability of regeneration error. Francis Crick's 'frozen accident' finds here a substantial content."

Contudo, Battail faz importantes conjecturas, porém não mostra exemplos concretos de sequências de DNA reproduzidas através dos CCEs e do sistema concatenado.

5.6.1 Sequências de DNA reproduzidas pelos códigos concatenados

Ainda em [14], Battail afirma: “*The survival of an organism necessitates the existence of a reliable information replication process. Therefore error-correcting codes must be used in replication or in another process of information regeneration that precedes replication. The genetic information undergoes nested encoding, where the result of a previous encoding process is combined with new information and encoded again. The more important genetic information is assumed to be in the primary coded message, regarding nested coding mirrors coding theory's concept of concatenated codes which are also called nested codes.”*

Neste trabalho, consideramos as hipóteses apresentadas anteriormente por Battail, com algumas restrições. Ao invés de analisarmos o processo de replicação do DNA com o objetivo de mostrarmos a existência de CCEs no genoma, o que torna o problema muito complexo, fazemos a seguinte conjectura: Se o genoma é constituído por regiões consistindo de exons, íntrons, sequências de direcionamento, *promoters*, DNA repetitivos, etc, e que cada uma dessas regiões pode ser reproduzida por um código CCE específico, então o genoma consiste de códigos concatenados (*nested codes*) no mínimo justapostos.

Aqui, mostramos que as sequências de DNA (várias partes do genoma) são palavras-código de CCEs com $D(\mathbf{a}, \mathbf{b}) = 1$ e, em alguns casos com $D(\mathbf{a}, \mathbf{b}) = 2$, identificadas e reproduzidas através dos códigos *G*-linearidade. Portanto, conjecturamos que a natureza utiliza CCEs genômicos.

No Capítulo 2, vimos que cada sequência de direcionamento (SD) possui uma única proteína organelar, exercendo o importante papel de direcionar a proteína à organela correta e ao local adequado no interior desta organela. As sequências de DNA Seq.12, Seq.24 e Seq.25 mencionadas na Tabela 5.1 são as sequências de direcionamento das proteínas Seq.50, Seq.51 e Seq.52, respectivamente. O nosso objetivo é apontar um indício da presença de códigos concatenados nestas sequências. Para alcançarmos este objetivo procedemos da seguinte maneira:

Primeiro - Iniciamos as análises com as SD's: Seq.12, Seq.24 e Seq.25. na extensão de Galois de

grau $r = 6$:

- a) Analisamos todas as possibilidades de geração destas SD's com 1 e 2 nucleotídeos de diferenças - $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ (nos 6 polinômios primitivos da extensão $r = 6$);
- b) Observamos quais polinômios primitivos foram capazes de identificar estas sequências e em quais rotulamentos;
- c) Identificamos os casos gerados sem diferenças em aminoácidos;

Segundo - Continuamos as análises com as proteínas: Seq.50, Seq.51 e Seq.52. na extensão de Galois de grau $r = 10$:

- a) Analisamos todas as possibilidades de geração destas proteínas com $D(\mathbf{a}, \mathbf{b}) = 1$ (nos 60 polinômios primitivos da extensão $r = 10$);
- b) Observamos quais polinômios primitivos foram capazes de identificar estas sequências e em quais rotulamentos;
- c) Identificamos os casos em que a parte da SD é gerada sem alteração de aminoácido;
- d) Tendo em vista a quantidade de polinômios primitivos na extensão $r = 10$, e, que essas sequências são identificadas com $D(\mathbf{a}, \mathbf{b}) = 2$ em todos os polinômios e nos três rotulamentos A, B e C, optamos por fazer uma seleção dos polinômios primitivos que julgamos possíveis candidatos ao código concatenado;
- e) Selecionamos apenas os polinômios, onde, os monômios dos polinômios primitivos da extensão $r = 10$ apresentavam semelhanças com os monômios dos polinômios primitivos da extensão $r = 6$, cujas SD's já tinham sido reproduzidas sem erros de aminoácidos;
- f) Fizemos as simulações com $D(\mathbf{a}, \mathbf{b}) = 2$ das proteínas em todos os polinômios (e rotulamentos) escolhidos no **item e**;
- g) Identificamos os casos em que a parte da SD é gerada sem alteração de aminoácido.

Sistema parcialmente concatenado

Veja na Figura 5.16 que a SD Seq.12 foi reproduzida com $D(\mathbf{a}, \mathbf{b}) = 1$ sem alteração de aminoácido somente no polinômio $p_{05}(x)$. Com $D(\mathbf{a}, \mathbf{b}) = 2$, a SD foi reproduzida sem alteração de aminoácido nos polinômios $p_{02}(x)$, $p_{03}(x)$, $p_{04}(x)$, $p_{05}(x)$ e $p_{06}(x)$ (veja a Tabela 5.12 para confirmação desses dados). A proteína Seq.50 não foi gerada com $D(\mathbf{a}, \mathbf{b}) = 1$. Nas

análises com $D(\mathbf{a}, \mathbf{b}) = 2$, selecionamos (entre os 60 polinômios primitivos) os polinômios primitivos com monômios semelhantes aos monômios de cada polinômio primitivo da extensão $r = 6$, como mostra a Tabela 5.15.

Seq.12| *B. napus* - Mitochondrial - Malate dehydrogenase* - GI número 899225

| | |
|-----------------------|--|
| 1^a) | $P_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ - $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$ - Rot.A: (0, 1, 3, 2) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| RtG: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 222 |
| ntG: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |
| 2^a) | $P_{02}(x) = x^6 + x^4 + x^3 + x + 1$ - $g_{02}(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1$ - Rot.A: (0, 1, 3, 2) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| RtG: | 222 030 212 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| ntG: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |
| 3^a) | $P_{03}(x) = x^6 + x^5 + 1$ - $g_{03}(x) = x^6 + 3x^5 + 2x^3 + 1$ - Rot.C: (0, 2, 1, 3) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 332 010 322 121 233 132 210 322 322 122 321 121 001 201 321 233 232 212 212 012 332 |
| RtG: | 332 010 322 123 232 132 210 322 322 122 321 121 001 201 321 233 232 212 212 012 332 |
| ntG: | TTC AGA TCC GCT CTC GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |
| 4^a) | $P_{04}(x) = x^6 + x^5 + x^2 + x + 1$ - $g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ - Rot.A: (0, 1, 3, 2) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| RtG: | 221 030 211 311 122 321 130 211 311 213 313 003 103 213 123 121 131 131 031 221 |
| ntG: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTG CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |
| 5^a) | $P_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ - $g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1$ - Rot.A: (0, 1, 3, 2) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| RtG: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 121 121 131 131 031 221 |
| ntG: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTG CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |
| 6^a) | $P_{06}(x) = x^6 + x^5 + x^4 + x + 1$ - $g_{06}(x) = x^6 + x^5 + x^4 + 2x^2 + 3x + 1$ - Rot.A: (0, 1, 3, 2) - (A, C, G, T) |
| aaO: | F R S A L V R S S A S A K Q S L L R R S F |
| ntO: | TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| RtO: | 221 030 211 313 122 321 130 211 311 213 313 003 103 213 122 121 131 131 031 221 |
| RtG: | 221 030 211 313 122 321 030 211 311 213 313 003 103 213 122 121 131 131 031 222 |
| ntG: | TTC AGA TCC GCG CTT GTC AGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC TTC |
| aaG: | F R S A L V R S S A S A K Q S L L R R S F |

Figura 5.16: Sequência de Direcionamento reproduzida em $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$.

Por exemplo, os monômios do polinômio primitivo $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ (extensão $r = 6$) devem constar nos polinômios primitivos da extensão $r = 10$. Esse mesmo procedimento se repete para os demais polinômios das extensões $r = 6$ e $r = 10$, e também para as demais SD's e proteínas.

| Sistema parcialmente concatenado: Seq.12 e Seq.50 - BCH sobre Anéis | | | |
|---|---|-----|---|
| Diferença em nucleotídeos | Polinômios primitivos da extensão de Galois $r = 6$ responsáveis pela geração da Seq.12 | Rot | Polinômios primitivos da extensão de Galois $r = 10$ responsáveis pela geração da Seq.50 |
| $D(\mathbf{a}, \mathbf{b}) = 1$ | $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | A | não foi gerado nos 60 polinômios |
| | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ | A | $p_{06}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{32}(x) = x^{10} + x^9 + x^6 + x^4 + x^3 + x + 1$ $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{39}(x) = x^{10} + x^9 + x^7 + x^6 + x^4 + x^3 + x^2 + x + 1$ $\textcolor{red}{p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1}$ - Rot. C |
| $D(\mathbf{a}, \mathbf{b}) = 2$ | $p_{03}(x) = x^6 + x^5 + 1$ | C | $p_{03}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + 1$ $p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ $p_{06}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1$ $p_{11}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x + 1$ $p_{12}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^2 + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{18}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x + 1$ $p_{21}(x) = x^{10} + x^7 + x^6 + x^5 + x^2 + x + 1$ $p_{26}(x) = x^{10} + x^6 + x^5 + x^2 + 1$ $\textcolor{red}{p_{28}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x^4 + x^3 + x^2 + 1}$ - Rot.A $p_{29}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + 1$ $p_{31}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1$ $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{37}(x) = x^{10} + x^8 + x^6 + x^5 + x^3 + x + 1$ $p_{38}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x + 1$ $p_{52}(x) = x^{10} + x^9 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$ |
| | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | A | $p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{21}(x) = x^{10} + x^7 + x^6 + x^5 + x^2 + x + 1$ $p_{31}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1$ $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ |
| | $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | A | $p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $\textcolor{red}{p_{28}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x^4 + x^3 + x^2 + 1}$ - Rot.A $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{52}(x) = x^{10} + x^9 + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$ |
| | $p_{06}(x) = x^6 + x^5 + x^4 + x + 1$ | A | $p_{06}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1$ $p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{18}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x + 1$ $p_{31}(x) = x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1$ $p_{34}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$ $p_{38}(x) = x^{10} + x^7 + x^6 + x^5 + x^4 + x + 1$ |

 Tabela 5.15: Interpretando o código concatenado - *B.napus*.

Como resultado das análises para as sequências Seq.12 e Seq.50, observe na Tabela 5.15 que o polinômio responsável pela reprodução da SD, $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$, está contido no polinômio responsável pela reprodução da proteína, $p_{28}(x) = x^{10} + x^9 + x^8 + x^6 + x^5 + x^4 + x^3 + x^2 + 1$, e, somente nesse caso, tanto a SD quanto a proteína foram reproduzidas

nos mesmos rotulamentos sem alteração de aminoácidos.

Logo, dizemos que o código \mathbb{Z}_4 -linearidade ((1023,1013,3) BCH primitivo sobre $GR(4,10)$, rotulamento A) contém o código \mathbb{Z}_4 -linearidade ((63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento A) e, como eles apresentam a propriedade de códigos entrelaçados portanto, são códigos concatenados. Como o monômio x^4 não pertence aos dois polinômios ($p_5(x)$ e $p_{28}(x)$), inferimos que estes códigos são parcialmente concatenados.

Seq.50 | *B. napus* - Mitochondrial - Malate dehydrogenase* - GI número 899225

| Código \mathbb{Z}_4 -linearidade ((1023,1013,3) BCH primitivo sobre $GR(4,10)$, rotulamento A) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| $p_{28}(x)=x^{10}+x^9+x^8+x^6+x^5+x^4+x^3+x^2+1 - g_{42}(x)=x^{10}+x^9+3x^8+2x^7+3x^6+x^5+3x^4+3x^3+3x^2+2x+1$ | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| Caso 2 - (A,C,G,T) = (0,1,3,2) | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: M F R S A L V R S S A S A K Q S L L R R S ntO: ATG TTC AGA TCC GCG CTT GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC RtO: 023 221 030 211 313 122 321 130 211 211 311 213 313 003 103 213 122 121 131 131 031 RtG: 023 221 030 211 313 123 321 130 211 211 311 213 313 003 103 213 122 121 131 131 031 ntG: ATG TTC AGA TCC GCG CTG GTC CGA TCC TCC GCC TCG GCG AAG CAG TCG CTT CTC CGC CGC AGC aaG: M F R S A L V R S S A S A K Q S L L R R S | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: F S S G S V P E R K V A I L G A A G G I G ntO: TTC TCT TCC GGA TCC GTC CCG GAG CGT AAA GTC GCC ATC TTG GGA GCC GCC GGT GGA ATC GGT RtO: 221 212 211 330 211 321 111 303 132 000 321 311 021 223 330 311 311 332 330 021 332 RtG: 221 212 211 330 211 321 111 303 132 000 321 311 021 223 330 311 311 332 330 021 332 ntG: TTC TCT TCC GGA TCC GTC CCC GAG CGT GTC GCC ATC TTG GGA GCC GCC GGT GGA ATC GGT aaG: F S S G S V P E R K V A I L G A A G G I G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: Q P L A L L M K L N P L V S S L S L Y D I ntO: CAG CCT CTG GCT CTC CTC ATG AAA CTC AAT CCC CTC GTC TCA AGC CTC TCC CTT TAC GAT ATC RtO: 103 112 123 312 121 121 023 000 121 002 111 121 321 210 031 121 211 122 201 302 021 RtG: 103 112 123 312 121 121 023 000 121 002 111 121 321 210 031 121 211 122 201 302 021 ntG: CAG CCT CTG GCT CTC CTC ATG AAA CTC AAT CCC CTC GTC TCA AGC CTC TCC CTT TAC GAT ATC aaG: Q P L A L L M K L N P L V S S L S L Y D I | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: A N T P G V A A D V G H I N T R S Q V V G ntO: GCC AAC ACA CCC GGA GTC GCC GCT GAC GTC GGT CAC ATC AAC ACC AGA TCT CAG GTG GTT GGG RtO: 311 001 010 111 330 321 311 312 301 321 332 101 021 001 011 030 212 103 323 322 333 RtG: 311 001 010 111 330 321 311 312 301 321 332 101 021 001 011 030 212 103 323 322 333 ntG: GCC AAC ACA CCC GGA GTC GCC GCT GAC GTC GGT CAC ATC AAC ACC AGA TCT CAG GTG GTT GGG aaG: A N T P G V A A D V G H I N T R S Q V V G | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: Y M G D D N L A K A L E G A D L V I I P A ntO: TAC ATG GGA GAT GAT AAC TTG GCG AAA GCT CTT GAA GGA GCT GAC CTC GTT ATC ATT CCA GCT RtO: 201 023 330 302 302 001 223 313 000 312 122 300 330 312 301 121 322 021 022 110 312 RtG: 201 023 330 302 302 001 223 313 000 312 122 300 330 312 301 121 322 021 022 110 312 ntG: TAC ATG GGA GAT GAT AAC TTG GCG AAA GCT CTT GAA GGA GCT GAC CTC GTT ATC ATT CCA GCT aaG: Y M G D D N L A K A L E G A D L V I I P A | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: G V P R K P G M T R D D L F N I N A G I V ntO: GGT GTC CCC AGG AAG CCT GGT ATG ACC CGT GAT GAT CTT TTC AAC ATC AAT GCT GGC ATT GTC RtO: 332 321 111 033 003 112 332 023 011 132 302 302 122 221 001 021 002 312 331 022 321 RtG: 332 321 111 033 003 112 332 023 011 132 302 302 122 221 001 021 002 312 331 022 321 ntG: GGT GTC CCC AGG AAG CCT GGT ATG ACC CGT GAT GAT CTT TTC AAC ATC AAT GCT GGC ATT GTC aaG: G V P R K P G M T R D D L F N I N A G I V | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: K N L W S A I A K Y C P H A L V N M I S N ntO: AAG AAC CCT TGG TCT GCC ATC GCC AAG TAC TGC CCT CAT GCA CTT GTT AAT ATG ATT AGC AAC RtO: 003 001 012 223 212 311 021 311 003 201 231 112 102 310 122 322 002 023 022 031 001 RtG: 003 001 012 223 212 311 021 311 003 201 231 112 102 310 122 322 002 023 022 031 001 ntG: AAG AAC CCT TGG TCT GCC ATC GCC AAG TAC TGC CCT CAT GCA CTT GTT AAT ATG ATT AGC AAC aaG: K N L W S A I A K Y C P H A L V N M I S N | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: P V N S T V P I A A E I F K K A G M Y D E ntO: CCG GTC AAC TCT ACT GTT CCG ATT GCA GCA GAG ATT TTC AAG AAG GCT GGT ATG TAC GAT GAA RtO: 113 321 001 212 012 322 113 022 310 310 303 022 221 003 003 312 332 023 201 302 300 RtG: 113 321 001 212 012 322 113 022 310 310 303 022 221 003 003 312 332 023 201 302 300 ntG: CCG GTC AAC TCT ACT GTT CCG ATT GCA GCA GAG ATT TTC AAG AAG GCT GGT ATG TAC GAT GAA aaG: P V N S T V P I A A E I F K K A G M Y D E | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
| aaO: K K L F G V T T L D V V R V K T S Y A G K ntO: AAG AAG CTG TTT GGT GTT ACC ACT CTT GAC GTT GTT AGG GTC AAG ACT TCC TAT GCT GGA AAG RtO: 003 003 123 222 332 322 011 012 122 301 322 322 033 321 003 012 211 202 312 330 003 RtG: 003 003 123 222 332 322 011 012 122 301 322 322 033 321 003 012 211 202 312 330 003 ntG: AAG AAG CTG TTT GGT GTT ACC ACT CTT GAC GTT GTT AGG GTC AAG ACT TCC TAT GCT GGA AAG aaG: K K L F G V T T L D V V R V K T S Y A G K | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | N | V | P | V | A | E | V | N | V | P | A | I | V | G | H | A | G | V | T | I | |
| ntO: | GCT | AAT | GTC | CCT | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GCG | ATT | GTT | GGT | CAT | GCT | GGG | GTT | ACA | ATC | |
| RtO: | 312 | 002 | 321 | 112 | 322 | 310 | 300 | 322 | 002 | 322 | 113 | 313 | 022 | 322 | 332 | 102 | 312 | 333 | 322 | 010 | 021 | |
| RtG: | 312 | 002 | 321 | 112 | 322 | 310 | 300 | 322 | 002 | 322 | 113 | 313 | 022 | 322 | 332 | 102 | 312 | 333 | 322 | 010 | 021 | |
| ntG: | GCT | AAT | GTC | CCT | GTT | GCA | GAA | GTT | AAT | GTT | CCG | GCG | ATT | GTT | GGT | CAT | GCT | GGG | GTT | ACA | ATC | |
| aaG: | A | N | V | P | V | A | E | V | N | V | P | A | I | V | G | H | A | G | V | T | I | |
| aaO: | L | P | L | F | S | Q | A | T | P | Q | A | I | L | S | G | D | A | L | T | V | T | |
| ntO: | CTC | CCA | CTC | TTC | TCT | CAG | GCC | ACT | CCC | CAA | GCT | ATC | TTG | TCA | GGT | GAC | GCG | CTC | ACT | GTC | ACC | |
| RtO: | 121 | 110 | 121 | 221 | 212 | 103 | 311 | 012 | 111 | 100 | 312 | 021 | 223 | 210 | 332 | 301 | 313 | 121 | 012 | 321 | 011 | |
| RtG: | 121 | 110 | 121 | 221 | 212 | 103 | 311 | 012 | 111 | 100 | 312 | 021 | 223 | 210 | 332 | 301 | 313 | 121 | 012 | 321 | 011 | |
| ntG: | CTC | CCA | CTC | TTC | TCT | CAG | GCC | ACT | CCC | CAA | GCT | ATC | TTG | TCA | GGT | GAC | GCG | CTC | ACT | GTC | ACC | |
| aaG: | L | P | L | F | S | Q | A | T | P | Q | A | I | L | S | G | D | A | L | T | V | T | |
| aaO: | T | K | R | T | Q | D | G | G | T | E | V | E | A | K | A | G | K | G | S | A | | |
| ntO: | ACC | AAG | CGT | ACT | CAA | GAT | GGA | GGT | ACA | GAA | GTT | GAA | GAG | GCT | AAG | GCA | GGG | AAA | GGT | TCA | GCT | |
| RtO: | 011 | 003 | 132 | 012 | 100 | 302 | 330 | 332 | 010 | 300 | 322 | 300 | 303 | 312 | 003 | 310 | 333 | 000 | 332 | 210 | 312 | |
| RtG: | 011 | 003 | 132 | 012 | 100 | 302 | 330 | 332 | 010 | 300 | 322 | 300 | 303 | 312 | 003 | 310 | 333 | 000 | 332 | 210 | 312 | |
| ntG: | ACC | AAG | CGT | ACT | CAA | GAT | GGA | GGT | ACA | GAA | GTT | GAA | GAG | GCT | AAG | GCA | GGG | AAA | GGT | TCA | GCT | |
| aaG: | T | K | R | T | Q | D | G | G | T | E | V | E | E | A | K | A | G | K | G | S | A | |
| aaO: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G | |
| ntO: | ACA | TTG | TCA | ATG | GCG | TAT | GCT | GGG | GCA | CTA | TTG | GCT | GAC | GCA | TGC | TTG | AAG | GGA | CTC | AAT | GGT | |
| RtO: | 010 | 223 | 210 | 023 | 313 | 202 | 312 | 330 | 332 | 310 | 120 | 221 | 312 | 301 | 310 | 231 | 223 | 003 | 330 | 121 | 002 | 332 |
| RtG: | 010 | 223 | 210 | 023 | 313 | 202 | 312 | 330 | 310 | 120 | 221 | 312 | 301 | 310 | 231 | 223 | 003 | 330 | 121 | 002 | 332 | |
| ntG: | ACA | TTG | TCA | ATG | GCG | TAT | GCT | GGG | GCA | CTA | TTG | GCT | GAC | GCA | TGC | TTG | AAG | GGA | CTC | AAT | GGT | |
| aaG: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G | |
| aaO: | V | P | D | V | V | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A | |
| ntO: | GTC | CCA | GAT | GTA | GTC | GAA | TGC | TCA | TAC | GTC | CAG | TCC | ACA | ATC | ACT | GAG | CTT | CCT | TTC | TTT | GCC | |
| RtO: | 321 | 110 | 302 | 320 | 321 | 300 | 231 | 210 | 201 | 323 | 103 | 211 | 010 | 021 | 012 | 303 | 122 | 112 | 221 | 222 | 311 | |
| RtG: | 321 | 110 | 302 | 320 | 321 | 300 | 231 | 210 | 201 | 323 | 103 | 211 | 010 | 021 | 012 | 303 | 122 | 112 | 221 | 222 | 311 | |
| ntG: | GTC | CCA | GAT | GTA | GTC | GAA | TGC | TCA | TAC | GTC | CAG | TCC | ACA | ATC | ACT | GAG | CTT | CCT | TTC | TTT | GCC | |
| aaG: | V | P | D | V | V | E | C | S | Y | V | Q | S | T | I | T | E | L | P | F | F | A | |
| aaO: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D | |
| ntO: | TCA | AAG | GTG | AGG | TTG | GGG | AAG | AAC | GGT | GTT | GAG | GAG | GTT | CTT | GAT | TTG | GGG | CCA | CTC | TCT | GAC | |
| RtO: | 210 | 003 | 323 | 033 | 223 | 333 | 003 | 001 | 332 | 322 | 303 | 303 | 322 | 122 | 302 | 223 | 333 | 110 | 121 | 212 | 301 | |
| RtG: | 210 | 003 | 323 | 033 | 223 | 333 | 003 | 001 | 332 | 322 | 303 | 303 | 322 | 122 | 302 | 223 | 333 | 110 | 121 | 212 | 301 | |
| ntG: | TCA | AAG | GTG | AGG | TTG | GGG | AAG | AAC | GGT | GTT | GAG | GAG | GTT | CTT | GAT | TTG | GGG | CCA | CTC | TCT | GAC | |
| aaG: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D | |
| aaO: | F | E | K | E | G | L | E | A | L | R | P | G | I | K | S | T | I | E | K | G | V | |
| ntO: | TTC | GAG | AAG | GAA | GGC | TTG | GAA | GCA | TTG | AGA | CCG | GGA | ATC | AAG | TCT | ACC | ATT | GAG | AAG | GGA | GTT | |
| RtO: | 221 | 303 | 003 | 300 | 331 | 223 | 300 | 310 | 310 | 223 | 030 | 113 | 330 | 021 | 003 | 212 | 011 | 022 | 303 | 003 | 330 | 322 |
| RtG: | 221 | 303 | 003 | 300 | 331 | 123 | 300 | 310 | 223 | 030 | 113 | 330 | 021 | 003 | 212 | 011 | 022 | 303 | 003 | 330 | 322 | |
| ntG: | TTC | GAG | AAG | GAA | GGC | CTG | GAA | GCA | TTG | AGA | CCG | GGA | ATC | AAG | TCT | ACC | ATT | GAG | AAG | GGA | GTT | |
| aaG: | F | E | K | E | G | L | E | A | L | R | P | G | I | K | S | T | I | E | K | G | V | |
| aaO: | K | F | A | N | Q | | | | | | | | | | | | | | | | | |
| ntO: | AAG | TTT | GCC | AAC | CAA | | | | | | | | | | | | | | | | | |
| RtO: | 003 | 222 | 311 | 001 | 100 | | | | | | | | | | | | | | | | | |
| RtG: | 003 | 222 | 311 | 001 | 100 | | | | | | | | | | | | | | | | | |
| ntG: | AAG | TTT | GCC | AAC | CAA | | | | | | | | | | | | | | | | | |
| aaG: | K | F | A | N | Q | | | | | | | | | | | | | | | | | |

 Figura 5.17: Proteína reproduzida em $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$.

Para os demais casos de polinômios analisados na Tabela 5.15, ou o rotulamento reproduzido não é o mesmo em ambas as sequências (SD e proteína), ou a parte da SD na proteína gerada não foi reproduzida sem alteração de aminoácido, assim todas essas possibilidades são descartadas como candidatas ao código concatenado.

Outro exemplo de sequências de DNA reproduzidas pelos códigos parcialmente concatenados é mostrado através das observações realizadas na geração e reprodução da sequência de direcionamento Seq.25 e a proteína Seq.52. Note na Tabela 5.16 que a SD não foi gerada com $D(\mathbf{a}, \mathbf{b}) = 1$ em $GR(4, 6)$, assim como a proteína também não foi gerada com $D(\mathbf{a}, \mathbf{b}) = 1$

em $GR(4, 10)$. No caso de $D(\mathbf{a}, \mathbf{b}) = 2$, somente o polinômio primitivo $p_{01}(x) = x^6 + x + 1$ foi capaz de reproduzir a SD sem alteração de aminoácido (Figura 5.18).

Seq.25| *A. thaliana* – Mitocôndria - Malate dedhydrogenase 2 – GI número 145338521

| $p_{01}(x) = x^6 + x + 1 - g_{01}(x) = x^6 + 2x^3 + 3x + 1 - Rotulamento A: (0, 1, 3, 2) - (A, C, G, T)$ | | | | | | | | | | | | | | | | | | | | | | | | |
|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|--|
| aaO: F R S M I V R S A S P V K Q G L L R R G F | | | | | | | | | | | | | | | | | | | | | | | | |
| ntO: TTC CGA TCA ATG ATT GTT CGA TCT GCT TCC CCA GTG AAG CAG GGT CTT CTC CGC AGA GGA TTC | | | | | | | | | | | | | | | | | | | | | | | | |
| RtO: 221 130 210 023 022 322 130 212 312 211 110 323 003 103 332 122 121 131 030 330 221 | | | | | | | | | | | | | | | | | | | | | | | | |
| RtG: 221 130 210 023 022 322 130 212 312 211 110 321 003 103 332 122 121 130 030 330 221 | | | | | | | | | | | | | | | | | | | | | | | | |
| ntG: TTC CGA TCA ATG ATT GTT CGA TCT GCT TCC CCA GTC AAG CAG GGT CTT CTC CGA AGA GGA TTC | | | | | | | | | | | | | | | | | | | | | | | | |
| aaG: F R S M I V R S A S P V K Q G L L R R G F | | | | | | | | | | | | | | | | | | | | | | | | |

Figura 5.18: Sequência de direcionamento reproduzida em $GR(4, 6)$ com $D(\mathbf{a}, \mathbf{b}) = 2$.

Veja, na Tabela 5.16, que os monômios do polinômio $p_{01}(x)$ estão contidos nos polinômios primitivos da extensão $r = 10$, porém somente o polinômio $p_{14}(x)$ foi capaz de reproduzir a proteína no mesmo rotulamento da SD e sem alteração de aminoácido, como mostra a Figura 5.19.

| Sistema parcialmente concatenado: Seq.25 e Seq.52 - BCH sobre Anéis | | | |
|---|---|-----|---|
| Diferença em nucleotídeos | Polinômios primitivos da extensão de Galois $r = 6$ responsáveis pela geração da Seq.25 | Rot | Polinômios primitivos da extensão de Galois $r = 10$ responsáveis pela geração da Seq.52 |
| $D(\mathbf{a}, \mathbf{b}) = 1$ | não foi gerado nos 6 polinômios | - | não foi gerado nos 60 polinômios |
| $D(\mathbf{a}, \mathbf{b}) = 2$ | $p_{01}(x) = x^6 + x + 1$ | A | $ \begin{aligned} p_{04}(x) &= x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1 \\ p_{06}(x) &= x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 \\ p_{11}(x) &= x^{10} + x^9 + x^8 + x^6 + x^5 + x + 1 \\ \textcolor{red}{p_{14}(x)} &= \textcolor{red}{x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1} - \text{Rot.A} \\ p_{18}(x) &= x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x + 1 \\ p_{19}(x) &= x^{10} + x^9 + x^7 + x^6 + x^4 + x + 1 \\ p_{20}(x) &= x^{10} + x^9 + x^8 + x^6 + x^2 + x + 1 \\ p_{21}(x) &= x^{10} + x^7 + x^6 + x^5 + x^2 + x + 1 \\ p_{23}(x) &= x^{10} + x^8 + x^7 + x^6 + x^2 + x + 1 \\ p_{24}(x) &= x^{10} + x^7 + x^6 + x^4 + x^2 + x + 1 \\ p_{31}(x) &= x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1 \\ p_{32}(x) &= x^{10} + x^9 + x^6 + x^4 + x^3 + x + 1 \\ p_{34}(x) &= x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ p_{37}(x) &= x^{10} + x^8 + x^6 + x^5 + x^3 + x + 1 \\ p_{38}(x) &= x^{10} + x^7 + x^6 + x^5 + x^4 + x + 1 \\ p_{39}(x) &= x^{10} + x^9 + x^7 + x^6 + x^4 + x^3 + x^2 + x + 1 \\ p_{42}(x) &= x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 \\ p_{48}(x) &= x^{10} + x^9 + x^6 + x^3 + x^2 + x + 1 \\ p_{49}(x) &= x^{10} + x^8 + x^6 + x^4 + x^2 + x + 1 \\ \textcolor{red}{p_{51}(x)} &= \textcolor{red}{x^{10} + x^9 + x^6 + x + 1} - \text{Rot.B e C} \\ p_{54}(x) &= x^{10} + x^8 + x^6 + x + 1 \end{aligned} $ |

Tabela 5.16: Interpretando o código concatenado - *A. thaliana*.

Vale lembrar que os demais polinômios da extensão $r = 10$ reproduziram a proteína,

contudo, não são indicados aos códigos concatenados pois, ou os rotulamentos entre SD e proteína não são os mesmos ou a parte da SD gerada pela proteína apresenta uma alteração de aminoácido.

Seq.52 | A. thaliana – Mitocôndria - Malate dedhydrogenase 2 – GI número 145338521
Código Z₄-linearidade ((1023, 1013, 3) BCH primitivo sobre GR(4,10), rotulamento A)

$$p_{14}(x) = x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 - g_{14}(x) = x^{10} + 3x^9 + 2x^8 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1$$

Caso 2 – (A,C,G,T) = (0,1,3,2)

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aao: | M | F | R | S | M | I | V | R | S | A | S | P | V | K | Q | G | L | L | R | R | G |
| nto: | ATG | TTC | CGA | TCA | ATG | ATT | GTT | CGA | TCT | GCT | TCC | CCA | GTG | AAG | CAG | GCT | CTT | CTC | CGC | AGA | GGG |
| Rto: | 023 | 221 | 130 | 210 | 023 | 022 | 322 | 130 | 212 | 312 | 211 | 110 | 323 | 003 | 103 | 332 | 122 | 121 | 131 | 030 | 330 |
| Rtg: | 023 | 221 | 130 | 210 | 023 | 022 | 322 | 130 | 212 | 312 | 211 | 110 | 323 | 003 | 103 | 332 | 122 | 121 | 131 | 030 | 330 |
| ntg: | ATG | TTC | CGA | TCA | ATG | ATT | GTT | CGA | TCT | GCT | TCC | CCA | GTG | AAG | CAG | GCT | CTT | CTC | CGC | AGA | GGG |
| aag: | M | F | R | S | M | I | V | R | S | A | S | P | V | K | Q | G | L | L | R | R | G |
| aao: | F | A | S | E | S | V | P | D | R | K | V | V | I | L | G | A | A | G | G | I | G |
| nto: | TTC | GCC | TCT | GAA | TCT | GTT | CCC | GAT | CGC | AAA | GTC | GTC | ATC | CTC | GGT | GCC | GCC | GGT | GGG | ATC | GGC |
| Rto: | 221 | 311 | 212 | 300 | 212 | 322 | 111 | 302 | 131 | 000 | 321 | 321 | 021 | 121 | 332 | 311 | 311 | 332 | 333 | 021 | 331 |
| Rtg: | 221 | 311 | 212 | 300 | 212 | 322 | 111 | 302 | 131 | 000 | 321 | 321 | 021 | 121 | 332 | 311 | 311 | 332 | 333 | 021 | 331 |
| ntg: | TTC | GCC | TCT | GAA | TCT | GTT | CCC | GAT | CGC | AAA | GTC | GTC | ATC | CTC | GGT | GCC | GCC | GGT | GGG | ATC | GGC |
| aag: | F | A | S | E | S | V | P | D | R | K | V | V | I | L | G | A | A | G | G | I | G |
| aao: | Q | P | L | S | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I |
| nto: | CAG | CCA | CTT | TCT | CTT | CTC | ATG | AAG | CTT | AAC | CCT | CTC | GTC | TCT | TCT | CTC | TCC | CTC | TAT | GAT | ATC |
| Rto: | 103 | 110 | 122 | 212 | 122 | 121 | 023 | 003 | 122 | 001 | 112 | 121 | 321 | 212 | 212 | 121 | 211 | 121 | 202 | 302 | 021 |
| Rtg: | 103 | 110 | 122 | 212 | 122 | 121 | 023 | 003 | 122 | 001 | 112 | 121 | 321 | 212 | 212 | 121 | 211 | 121 | 202 | 302 | 021 |
| ntg: | CAG | CCA | CTT | TCT | CTT | CTC | ATG | AAG | CTT | AAC | CCT | CTC | GTC | TCT | TCT | CTC | TCC | CTC | TAT | GAT | ATC |
| aag: | Q | P | L | S | L | L | M | K | L | N | P | L | V | S | S | L | S | L | Y | D | I |
| aao: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | Q | V | S | G |
| nto: | GCC | AAC | ACT | CCC | GCG | GTT | GCT | GCT | GAC | GTC | GGT | CAC | ATC | AAC | ACC | CGA | TCT | CAG | GTT | TCT | GGG |
| Rto: | 311 | 001 | 012 | 111 | 331 | 322 | 312 | 312 | 301 | 321 | 332 | 101 | 021 | 001 | 011 | 130 | 212 | 103 | 322 | 212 | 333 |
| Rtg: | 311 | 001 | 012 | 111 | 331 | 322 | 312 | 312 | 301 | 321 | 332 | 101 | 021 | 001 | 011 | 130 | 212 | 103 | 322 | 212 | 333 |
| ntg: | GCC | AAC | ACT | CCC | GCG | GTT | GCT | GCT | GAC | GTC | GGT | CAC | ATC | AAC | ACC | CGA | TCT | CAG | GTT | TCT | GGG |
| aag: | A | N | T | P | G | V | A | A | D | V | G | H | I | N | T | R | S | Q | V | S | G |
| aao: | Y | M | G | D | D | L | G | K | A | L | E | G | A | D | D | L | V | I | I | P | A |
| nto: | TAC | ATG | GGT | GAT | GAT | TTG | GGG | AAA | GCT | CTA | GAA | GGC | GCT | GAC | CTC | GTC | ATT | ATT | CCA | GCT | |
| Rto: | 201 | 023 | 332 | 302 | 302 | 302 | 223 | 333 | 000 | 312 | 120 | 300 | 331 | 312 | 301 | 121 | 321 | 022 | 022 | 110 | 312 |
| Rtg: | 201 | 023 | 332 | 302 | 302 | 302 | 223 | 333 | 000 | 312 | 120 | 300 | 331 | 312 | 301 | 121 | 321 | 022 | 022 | 110 | 312 |
| ntg: | TAC | ATG | GGT | GAT | GAT | TTG | GGG | AAA | GCT | CTA | GAA | GGC | GCT | GAC | CTC | GTC | ATT | ATT | CCA | GCT | |
| aag: | Y | M | G | D | D | L | G | K | A | L | E | G | A | D | D | L | V | I | I | P | A |
| aao: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V |
| nto: | GGT | GTC | CCA | AGG | AAG | CCT | GGC | ATG | ACC | CGT | GAT | GAT | CTT | TTC | AAC | ATC | AAT | GCT | GGC | ATT | GTC |
| Rto: | 332 | 321 | 110 | 033 | 003 | 112 | 331 | 023 | 011 | 132 | 302 | 302 | 122 | 221 | 001 | 021 | 002 | 312 | 331 | 022 | 321 |
| Rtg: | 332 | 321 | 110 | 033 | 003 | 112 | 331 | 023 | 011 | 132 | 302 | 302 | 122 | 221 | 001 | 021 | 002 | 312 | 331 | 022 | 321 |
| ntg: | GGT | GTC | CCA | AGG | AAG | CCT | GGC | ATG | ACC | CGT | GAT | GAT | CTT | TTC | AAC | ATC | AAT | GCT | GGC | ATT | GTC |
| aag: | G | V | P | R | K | P | G | M | T | R | D | D | L | F | N | I | N | A | G | I | V |
| aao: | K | N | L | S | I | A | I | A | K | Y | C | P | Q | A | L | V | N | M | I | S | N |
| nto: | AAG | AAC | CTC | AGC | ATC | GCC | ATC | GCC | AAG | TAT | TGC | CCA | CAA | GCA | CTT | GTC | AAC | ATG | ATC | AGC | AAC |
| Rto: | 003 | 001 | 121 | 031 | 021 | 311 | 021 | 311 | 003 | 202 | 231 | 110 | 100 | 310 | 122 | 321 | 001 | 023 | 021 | 031 | 001 |
| Rtg: | 003 | 001 | 121 | 031 | 021 | 311 | 021 | 311 | 003 | 202 | 231 | 110 | 100 | 310 | 122 | 321 | 001 | 023 | 021 | 031 | 001 |
| ntg: | AAG | AAC | CTC | AGC | ATC | GCC | ATC | GCC | AAG | TAT | TGC | CCA | CAA | GCA | CTT | GTC | AAC | ATG | ATC | AGC | AAC |
| aag: | K | N | L | S | I | A | I | A | K | Y | C | P | Q | A | L | V | N | M | I | S | N |
| aao: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | T | Y | D | E |
| nto: | CCT | GTG | AAC | TCC | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATT | TTT | AAG | AAG | GCT | GGT | ACC | TAT | GAT | GAG |
| Rto: | 112 | 323 | 001 | 211 | 012 | 322 | 110 | 022 | 310 | 312 | 303 | 022 | 222 | 003 | 003 | 312 | 332 | 011 | 202 | 302 | 303 |
| Rtg: | 112 | 323 | 001 | 211 | 012 | 322 | 110 | 022 | 310 | 312 | 303 | 022 | 222 | 003 | 003 | 312 | 332 | 011 | 202 | 302 | 303 |
| ntg: | CCT | GTG | AAC | TCC | ACT | GTT | CCA | ATT | GCA | GCT | GAG | ATT | TTT | AAG | AAG | GCT | GGT | ACC | TAT | GAT | GAG |
| aag: | P | V | N | S | T | V | P | I | A | A | E | I | F | K | K | A | G | T | Y | D | E |
| aao: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |
| nto: | AAG | AAA | TTA | TTC | GGT | GTC | ACC | ACT | CIT | GAT | GTT | GTC | AGG | GCT | AGG | ACT | TTC | TAT | GCT | GGA | AAA |
| Rto: | 003 | 000 | 220 | 221 | 332 | 321 | 011 | 012 | 122 | 302 | 322 | 321 | 033 | 312 | 033 | 012 | 221 | 202 | 312 | 330 | 000 |
| Rtg: | 003 | 000 | 220 | 221 | 333 | 321 | 011 | 012 | 122 | 302 | 322 | 321 | 033 | 312 | 033 | 012 | 221 | 202 | 312 | 330 | 000 |
| ntg: | AAG | AAA | TTA | TTC | GGG | GTC | ACC | ACT | CIT | GAT | GTT | GTC | AGG | GCT | AGG | ACT | TTC | TAT | GCT | GGA | AAA |
| aag: | K | K | L | F | G | V | T | T | L | D | V | V | R | A | R | T | F | Y | A | G | K |

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | S | D | V | N | V | A | E | V | N | V | P | V | V | G | G | H | A | G | I | T | I |
| ntO: | TCG | GAT | GTC | AAT | GTT | GCA | GAG | GTT | AAT | GTT | CCA | GTT | GTT | GGT | GGT | CAT | GCT | GGC | ATC | ACG | ATT |
| RtO: | 213 | 302 | 321 | 002 | 322 | 310 | 303 | 322 | 002 | 322 | 110 | 322 | 322 | 332 | 332 | 102 | 312 | 331 | 021 | 013 | 022 |
| RtG: | 213 | 302 | 321 | 002 | 322 | 310 | 303 | 322 | 002 | 322 | 110 | 322 | 322 | 332 | 332 | 102 | 312 | 331 | 021 | 013 | 022 |
| ntG: | TCG | GAT | GTC | AAT | GTT | GCA | GAG | GTT | AAT | GTT | CCA | GTT | GTG | GGT | GGT | CAT | GCT | GGC | ATC | ACG | ATT |
| aaG: | S | D | V | N | V | A | E | V | N | V | P | V | V | G | G | H | A | G | I | T | I |
| aaO: | L | P | L | F | S | Q | A | S | P | Q | A | N | L | S | D | D | L | I | R | A | L |
| ntO: | CTT | CCT | CTC | TTT | TCT | CAG | GCT | AGC | CCT | CAA | GCC | AAC | TTG | TCG | GAT | GAT | TTA | ATC | AGG | GCC | CTC |
| RtO: | 122 | 112 | 121 | 222 | 212 | 103 | 312 | 031 | 112 | 100 | 311 | 001 | 223 | 213 | 302 | 302 | 220 | 021 | 033 | 311 | 121 |
| RtG: | 122 | 112 | 121 | 222 | 212 | 103 | 312 | 031 | 112 | 100 | 311 | 001 | 223 | 213 | 302 | 302 | 220 | 021 | 033 | 311 | 121 |
| ntG: | CTT | CCT | CTC | TTT | TCT | CAG | GCT | AGC | CCT | CAA | GCC | AAC | TTG | TCG | GAT | GAT | TTA | ATC | AGG | GCC | CTC |
| aaG: | L | P | L | F | S | Q | A | S | P | Q | A | N | L | S | D | D | L | I | R | A | L |
| aaO: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| ntO: | ACA | AAG | CGT | ACC | CAG | GAC | GGA | GGG | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCT | GGA | AAG | GGT | TCA | GCT |
| RtO: | 010 | 003 | 132 | 011 | 103 | 301 | 330 | 333 | 010 | 300 | 321 | 323 | 303 | 310 | 000 | 312 | 330 | 003 | 332 | 210 | 312 |
| RtG: | 010 | 003 | 132 | 011 | 103 | 301 | 330 | 333 | 010 | 300 | 321 | 323 | 303 | 310 | 000 | 312 | 330 | 003 | 332 | 210 | 312 |
| ntG: | ACA | AAG | CGT | ACC | CAG | GAC | GGA | GGG | ACA | GAA | GTC | GTG | GAG | GCA | AAA | GCT | GGA | AAG | GGT | TCA | GCT |
| aaG: | T | K | R | T | Q | D | G | G | T | E | V | V | E | A | K | A | G | K | G | S | A |
| aaO: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| ntO: | ACA | TTG | TCA | ATG | GCC | TAT | GCG | GGA | GCA | CTC | TTT | GCT | GAT | GCA | TGC | TTA | AAG | GGA | CTC | AAC | GGT |
| RtO: | 010 | 223 | 210 | 023 | 311 | 202 | 313 | 330 | 310 | 121 | 222 | 312 | 302 | 310 | 231 | 220 | 003 | 330 | 121 | 001 | 332 |
| RtG: | 010 | 223 | 210 | 023 | 311 | 202 | 313 | 330 | 310 | 121 | 222 | 312 | 302 | 310 | 231 | 220 | 003 | 330 | 121 | 001 | 332 |
| ntG: | ACA | TTG | TCA | ATG | GCC | TAT | GCG | GGA | GCA | CTC | TTT | GCT | GAT | GCA | TGC | TTA | AAG | GGA | CTC | AAC | GGT |
| aaG: | T | L | S | M | A | Y | A | G | A | L | F | A | D | A | C | L | K | G | L | N | G |
| aaO: | V | P | N | V | V | E | C | S | F | V | Q | S | T | I | T | E | L | P | F | F | A |
| ntO: | GTT | CCA | AAT | GTG | GTG | GAA | TGC | TCA | TTC | GTC | CAA | TCT | ACC | ATC | ACC | GAG | CTT | CCT | TTC | TTC | GCC |
| RtO: | 322 | 110 | 002 | 323 | 323 | 300 | 231 | 210 | 221 | 321 | 100 | 212 | 011 | 021 | 011 | 303 | 122 | 112 | 221 | 221 | 311 |
| RtG: | 322 | 110 | 002 | 323 | 323 | 300 | 231 | 210 | 221 | 321 | 100 | 212 | 011 | 021 | 011 | 303 | 122 | 112 | 221 | 221 | 311 |
| ntG: | GTT | CCA | AAT | GTG | GTG | GAA | TGC | TCA | TTC | GTC | CAA | TCT | ACC | ATC | ACC | GAG | CTT | CCT | TTC | TTC | GCC |
| aaG: | V | P | N | V | V | E | C | S | F | V | Q | S | T | I | T | E | L | P | F | F | A |
| aaO: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| ntO: | TCA | AAG | GTA | AGA | CTG | GGA | AAG | AAC | GGA | GTG | GAG | GAA | GTG | CTA | GAT | CTA | GGG | CCA | CTC | TCA | GAC |
| RtO: | 210 | 003 | 320 | 030 | 123 | 330 | 003 | 001 | 330 | 323 | 303 | 300 | 323 | 120 | 302 | 120 | 333 | 110 | 121 | 210 | 301 |
| RtG: | 210 | 003 | 320 | 030 | 123 | 330 | 003 | 001 | 330 | 323 | 303 | 300 | 323 | 120 | 302 | 120 | 333 | 110 | 121 | 210 | 301 |
| ntG: | TCA | AAG | GTA | AGA | CTG | GGA | AAG | AAC | GGA | GTG | GAG | GAA | GTG | CTA | GAT | CTA | GGG | CCA | CTC | TCA | GAC |
| aaG: | S | K | V | R | L | G | K | N | G | V | E | E | V | L | D | L | G | P | L | S | D |
| aaO: | F | E | K | E | G | L | E | A | L | K | A | E | L | K | S | S | I | E | K | G | I |
| ntO: | TTT | GAA | AAG | GAA | GCC | TTA | GAA | GCC | CTC | AAG | GCA | GAA | CTC | AAA | TCC | TCT | ATC | GAG | AAG | GGC | ATC |
| RtO: | 222 | 300 | 003 | 300 | 331 | 220 | 300 | 311 | 121 | 003 | 310 | 300 | 121 | 000 | 211 | 212 | 021 | 303 | 003 | 331 | 021 |
| RtG: | 222 | 300 | 003 | 300 | 331 | 220 | 300 | 311 | 121 | 003 | 310 | 300 | 121 | 000 | 211 | 212 | 021 | 303 | 003 | 331 | 021 |
| ntG: | TTT | GAA | AAG | GAA | GCC | TTA | GAA | GCC | CTC | AAG | GCA | GAA | CTC | AAA | TCC | TCT | ATC | GAG | AAG | GGC | ATC |
| aaG: | F | E | K | E | G | L | E | A | L | K | A | E | L | K | S | S | I | E | K | G | I |
| aaO: | K | F | A | N | Q | | | | | | | | | | | | | | | | |
| ntO: | AAA | TTT | GCC | AAC | CAA | | | | | | | | | | | | | | | | |
| RtO: | 000 | 222 | 311 | 001 | 100 | | | | | | | | | | | | | | | | |
| RtG: | 000 | 222 | 311 | 001 | 100 | | | | | | | | | | | | | | | | |
| ntG: | AAA | TTT | GCC | AAC | CAA | | | | | | | | | | | | | | | | |
| aaG: | K | F | A | N | Q | | | | | | | | | | | | | | | | |

 Figura 5.19: Proteína reproduzida em $GR(4, 10)$ com $D(\mathbf{a}, \mathbf{b}) = 2$

No casos das proteínas Seq.50 e Seq.52 (Figuras 5.17 e 5.19), note que, nos dois casos, as proteínas foram reproduzidas inteiramente sem alteração de aminoácido. Observe ainda que os polinômios responsáveis pela reprodução das SD's sem alteração de aminoácidos não são extamente os monônimos finais dos polinômios responsáveis pela reprodução das proteínas. Logo, inferimos que essas proteínas são reproduzidas por códigos entrelaçados e portanto, possuem a propriedade de códigos parcialmente concatenados.

Sistema totalmente concatenado

Aqui, conjecturamos que um sistema é totalmente *nested* quando o polinômio primitivo da SD é exatamente os monômios finais do polinômio primitivo da proteína, como mostra sequência de direcionamento Seq.24 e a proteína Seq.51. Veja na Figura 5.20 que a SD (1^{a} sequência) foi reproduzida somente no polinômio primitivo $p_{04}(x)$ com $D(\mathbf{a}, \mathbf{b}) = 1$, porém com alteração de aminoácido, logo essa SD está descartada das análises. Nas análises com $D(\mathbf{a}, \mathbf{b}) = 2$, note que a SD foi reproduzida sem alteração de aminoácido nos polinômios $p_{01}(x)$ (rotulamentos A e B) e $p_{02}(x)$ (rotulamento C), cujos monômios destes dois polinômios devem estar contidos nos polinômios da $r = 10$.

Seq.24 | *A. thaliana* – Mitocôndria – Malate dehydrogenase 1 – GI número 30695458

| | | | |
|-----------------------|--|---|---|
| 1^a) | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | $- g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1$ | - Rot.B: (0, 1, 2, 3) – (A, C, G, T) |
| aaO: | F R S M L V R S S A S A K Q A V I R R S F | | |
| ntO: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC TTC | | |
| RtO: | 331 020 313 032 131 231 120 313 313 211 311 212 002 102 212 233 031 121 123 021 331 | | |
| RtG: | 331 020 313 032 131 231 120 313 313 211 311 212 003 102 212 233 031 121 123 021 331 | | |
| ntG: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAAT CAG GCG GTT ATC CGC CGT AGC TTC | | |
| aAG: | F R S M L V R S S A S A N Q A V I R R S F | | |
| 2^a) | $p_{01}(x) = x^6 + x + 1$ | $- g_{01}(x) = x^6 + 2x^3 + 3x + 1$ | - Rot.A: (0, 1, 3, 2) – (A, C, G, T) |
| aaO: | F R S M L V R S S A S A K Q A V I R R S F | | |
| ntO: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC TTC | | |
| RtO: | 221 030 212 023 121 321 130 212 212 311 211 313 003 103 313 322 021 131 132 031 221 | | |
| RtG: | 221 030 212 023 120 321 130 212 212 311 211 313 003 103 313 322 022 131 132 031 221 | | |
| ntG: | TTC AGA TCT ATG CTA GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATT CGC CGT AGC TTC | | |
| aAG: | F R S M L V R S S A S A K Q A V I R R S F | | |
| 3^a) | $p_{01}(x) = x^6 + x + 1$ | $- g_{01}(x) = x^6 + 2x^3 + 3x + 1$ | - Rot.B: (0, 1, 2, 3) – (A, C, G, T) |
| aaO: | F R S M L V R S S A S A K Q A V I R R S F | | |
| ntO: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC TTC | | |
| RtO: | 331 020 313 032 131 231 120 313 313 211 311 212 002 102 212 233 031 121 123 021 331 | | |
| RtG: | 331 020 313 032 131 231 120 313 313 211 311 212 002 102 212 233 030 121 121 021 331 | | |
| ntG: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATA CGC CGC AGC TTC | | |
| aAG: | F R S M L V R S S A S A K Q A V I R R S F | | |
| 4^a) | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ | $- g_{01}(x) = x^6 + 2x^3 + 3x + 1$ | - Rot.C: (0, 2, 1, 3) – (A, C, G, T) |
| aaO: | F R S M L V R S S A S A K Q A V I R R S F | | |
| ntO: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCC TCC GCG AAG CAG GCG GTT ATC CGC CGT AGC TTC | | |
| RtO: | 332 010 323 031 232 132 210 323 323 122 322 121 001 201 121 133 032 212 213 012 332 | | |
| RtG: | 332 010 323 031 232 132 210 323 323 121 322 121 001 201 121 131 032 212 213 012 332 | | |
| ntG: | TTC AGA TCT ATG CTC GTC CGA TCT TCT GCG TCC GCG AAG CAG GCG GTG ATC CGC CGT AGC TTC | | |
| aAG: | F R S M L V R S S A S A K Q A V I R R S F | | |

Figura 5.20: Sequência de DNA n^o 24 gerada pelo código BCH (63, 57, 3).

Na Tabela 5.17, indicamos os possíveis polinômios primitivos da extensão $r = 10$ que contêm os monômios dos polinômios $p_{01}(x)$ e $p_{02}(x)$ candidatos ao código concatenado. Fi-

zemos as simulações em todos esses polinômios com a proteína Seq.51, e consideramos apenas os casos reproduzidos no mesmo rotulamento da SD e sem alteração de aminoácido na parte da SD. Veja na tabela que o polinômio $p_{42}(x)$ satisfaz essas condições tanto para $D(\mathbf{a}, \mathbf{b}) = 1$ quanto para $D(\mathbf{a}, \mathbf{b}) = 2$, ou seja, reproduz a proteína no mesmo rotulamento da SD (rotulamento C) e sem alteração de aminoácido na parte da SD. Note que o polinômio $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ está contido na parte final do polinômio $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$.

Sistema totalmente concatenado: Seq.24 e Seq.51 - BCH sobre Anéis

| Diferença em nucleotídeos | Polinômios primitivos da extensão de Galois $r = 6$ responsáveis pela geração da Seq.24 | Rot | Polinômios primitivos da extensão de Galois $r = 10$ responsáveis pela geração da Seq.51 |
|---------------------------------|---|-------|---|
| $D(\mathbf{a}, \mathbf{b}) = 1$ | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | B | $p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1$ - Rot.C |
| $D(\mathbf{a}, \mathbf{b}) = 2$ | $p_{01}(x) = x^6 + x + 1$ | A e B | $ \begin{aligned} p_{04}(x) &= x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1 \\ p_{06}(x) &= x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 \\ p_{11}(x) &= x^{10} + x^9 + x^8 + x^6 + x^5 + x + 1 \\ p_{14}(x) &= x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ p_{18}(x) &= x^{10} + x^9 + x^8 + x^7 + x^6 + x^5 + x^4 + x + 1 \\ p_{19}(x) &= x^{10} + x^9 + x^7 + x^6 + x^4 + x + 1 \\ p_{20}(x) &= x^{10} + x^9 + x^8 + x^6 + x^2 + x + 1 \\ p_{21}(x) &= x^{10} + x^7 + x^6 + x^5 + x^2 + x + 1 \\ p_{23}(x) &= x^{10} + x^8 + x^7 + x^6 + x^2 + x + 1 \\ p_{24}(x) &= x^{10} + x^7 + x^6 + x^4 + x^2 + x + 1 \\ p_{31}(x) &= x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^2 + x + 1 \\ p_{32}(x) &= x^{10} + x^9 + x^6 + x^4 + x^3 + x + 1 \\ p_{34}(x) &= x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ p_{37}(x) &= x^{10} + x^8 + x^6 + x^5 + x^3 + x + 1 \\ p_{38}(x) &= x^{10} + x^7 + x^6 + x^5 + x^4 + x + 1 \\ p_{39}(x) &= x^{10} + x^9 + x^7 + x^6 + x^4 + x^3 + x^2 + x + 1 \\ p_{42}(x) &= x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 \\ p_{48}(x) &= x^{10} + x^9 + x^6 + x^3 + x^2 + x + 1 \\ p_{49}(x) &= x^{10} + x^8 + x^6 + x^4 + x^2 + x + 1 \\ p_{51}(x) &= x^{10} + x^9 + x^6 + x + 1 \\ p_{54}(x) &= x^{10} + x^8 + x^6 + x + 1 \end{aligned} $ |
| | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ | C | $ \begin{aligned} p_{06}(x) &= x^{10} + x^8 + x^7 + x^6 + x^5 + x^4 + x^3 + x + 1 \\ p_{14}(x) &= x^{10} + x^9 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ p_{32}(x) &= x^{10} + x^9 + x^6 + x^4 + x^3 + x + 1 \\ p_{34}(x) &= x^{10} + x^7 + x^6 + x^5 + x^4 + x^3 + x^2 + x + 1 \\ p_{39}(x) &= x^{10} + x^9 + x^7 + x^6 + x^4 + x^3 + x^2 + x + 1 \\ p_{42}(x) &= x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 - \text{Rot.C} \end{aligned} $ |

Tabela 5.17: Interpretando o código concatenado - *A.thaliana*.

No Capítulo 4, a proteína Seq.51 foi reproduzida com $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ como mostradas nas Figuras 4.6 e 4.9. Nas duas sequências, note que a parte da SD foi reproduzida sem alteração de aminoácido e, na parte restante da proteína apresenta uma alteração de aminoácido. Assim, a reprodução da SD e sua respectiva proteína *Malate dehydrogenase 1* (GI número 30695458) satisfazem a conjectura e portanto, inferimos que essa proteína é reproduzida por códigos totalmente concatenados.

No sistema concatenado, Battail sugere que as informações mais antigas são mais protegidas que as informações mais recentes. O fato de mostrarmos que existe a presença de um sistema parcialmente ou totalmente concatenados em algumas sequências de DNA através dos polinômios primitivos, ou seja, que os códigos das sequências de direcionamento e suas respectivas proteínas precursoras estão de alguma maneira concatenadas, não quer dizer que as informações contidas nessas SD's são mais antigas que as informações contidas nas proteínas. Sem contar que as sequências de DNA, de maneira geral, podem ter sofrido pressões evolutivas diferentes ao longo do tempo e, isso deve ser estudado caso-a-caso. É necessário uma análise evolutiva dessas sequências de DNA ao longo do tempo para que seja confirmada a hipótese apresentada por Battail com mais propriedade.

5.7 Relação entre o Corpo $GF(4^r)$ e o Anel $GR(4, r)$

Nos processos de geração dos códigos G -linearidade (em termos dos códigos BCH primitivos cujo grau da extensão de Galois é par), identificação e reprodução das sequências de DNA para os comprimentos n iguais a 63, 255 e 1023, observamos que:

- 1º) O grau da extensão de Galois no corpo é a metade do grau da extensão de Galois no anel;
- 2º) A quantidade de polinômios primitivos a serem analisados no corpo é o dobro da quantidade de polinômios primitivos a serem analisados no anel;
- 3º) A cada dois polinômios primitivos no corpo temos um polinômio gerador correspondente, cujo polinômio gerador é igual a um dos polinômios primitivos na extensão do anel correspondente.

Na Tabela 5.18 vimos os exemplos da primeira e da segunda observação. O exemplo da terceira observação é mostrado na Tabela 5.19.

O critério mais simples para determinarmos, quais, entre os polinômios primitivos no corpo possuem o mesmo polinômio gerador, pode ser determinado através da comparação entre os termos dos polinômios primitivos, da seguinte maneira:

- a) Os termos com coeficientes 1 devem ser iguais;
- b) Os termos com coeficientes diferentes de 1, tais coeficientes devem ser recíprocos.

Por exemplo, os polinômios $p_{01}(x) = x^3 + x^2 + ax + b$ e $p_{02}(x) = x^3 + x^2 + bx + a$ sobre $GF(4^3)$ mencionados na Tabela 5.19. Note que os termos com coeficientes 1 (x^3 e x^2) são

| BCH sobre $GF(4^r)$ e sobre $GR(4, r)$ | | | |
|--|--------------------|---------------------------|--------------------------|
| | Extensão de Galois | Comprimento n | Quantidade de polinômios |
| | r | $n = (2^r - 1)$ | $p(x)$ |
| Anéis de Galois | $r = 6$ | $n = (2^6 - 1) = 63$ | 6 |
| | $r = 8$ | $n = (2^8 - 1) = 255$ | 16 |
| | $r = 10$ | $n = (2^{10} - 1) = 1023$ | 60 |
| | $r = 12$ | $n = (2^{12} - 1) = 4095$ | 144 |
| Corpos de Galois | r | $n = (4^r - 1)$ | $p(x)$ |
| | $r = 3$ | $n = (4^3 - 1) = 63$ | 12 |
| | $r = 4$ | $n = (4^4 - 1) = 255$ | 32 |
| | $r = 5$ | $n = (4^5 - 1) = 1023$ | 120 |
| | $r = 6$ | $n = (4^6 - 1) = 4095$ | 288 |

Tabela 5.18: Relação entre os códigos BCH sobre $GF(4^r)$ e sobre $GR(4, r)$.

iguais nos dois polinômios $p_{01}(x)$ e $p_{02}(x)$; e os termos com coeficientes diferentes de 1 ($ax+b$ e $bx+a$) os coeficientes são recíprocos. Logo, esses dois polinômios primitivos possuem o mesmo polinômio gerador. Esse critério de comparação entre os termos dos polinômios primitivos pode ser verificado para todos os polinômios primitivos nos códigos BCH primitivos sobre $GF(4^r)$, como mostra a Tabela 5.19 para os polinômios primitivos de grau $r = 3$ e $r = 4$.

A estrutura de corpo não foi capaz de reproduzir as sequências de DNA com comprimentos n iguais a 21, 51 e 93 nucleotídeos com o padrão de erro $D(\mathbf{a}, \mathbf{b}) = 1$ e, as simulações para o padrão de erros $D(\mathbf{a}, \mathbf{b}) = 2$ não foram feitas em virtude da complexidade computacional, por isso o processo de constução e geração dos códigos BCH não primitivos sobre $GF(4^r)$ não foram mostrados neste trabalho. Porém, durante a construção desses códigos, notamos que no caso do código BCH não primitivo $(21, 15, 3)$ os 12 polinômios primitivos da extensão $r = 3$ tinham como resultado 2 polinômios geradores; no caso do código BCH não primitivo $(51, 57, 3)$ os 32 polinômios primitivos da extensão $r = 4$ tinham como resultado 4 polinômios geradores; e no caso do código BCH não primitivo $(93, 83, 3)$ os 120 polinômios primitivos da extensão $r = 5$ tinham como resultado 6 polinômios geradores. Portanto, fazemos aqui mais uma observação, a quantidade de polinômios geradores 2, 4 e 6 são iguais a quantidade de classes de polinômios primitivos nos códigos BCH não primitivos sobre $GR(4, r)$ já mostradas nas Tabelas 5.9, 5.10 e 5.11, respectivamente.

A Tabela 5.20⁵ mostra um resumo das análises e simulações que foram feitas neste trabalho, em termos dos códigos BCH primitivos e não primitivos sobre o anel \mathbb{Z}_4 e o corpo $GF(4)$ e suas extensões de Galois com os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$.

⁵A notação $p(x)/g(x)$ na tabela, indica a quantidade de polinômios primitivos/classes de polinômios geradores em cada extensão; * corresponde as 10 sequências de direcionamento identificadas e reproduzidas em [20].

| Relação entre os polinômios no Corpo $GF(4^r)$ e no Anel $GR(4, r)$ | |
|--|---|
| $GF(4^3)$ | $GR(4, 6)$ |
| $p_{01}(x) = x^3 + x^2 + ax + b$ e $p_{02}(x) = x^3 + x^2 + bx + a$ $g_{01}(x) = g_{02}(x) = x^6 + x + 1$ | $p_{01}(x) = x^6 + x + 1$ |
| $p_{06}(x) = x^3 + x^2 + x + b$ e $p_{09}(x) = x^3 + x^2 + x + a$ $g_{06}(x) = g_{09}(x) = x^6 + x^4 + x^3 + x + 1$ | $p_{02}(x) = x^6 + x^4 + x^3 + x + 1$ |
| $p_{10}(x) = x^3 + ax^2 + bx + b$ e $p_{12}(x) = x^3 + bx^2 + ax + a$ $g_{10}(x) = g_{12}(x) = x^6 + x^5 + 1$ | $p_{03}(x) = x^6 + x^5 + 1$ |
| $p_{03}(x) = x^3 + ax^2 + bx + a$ e $p_{04}(x) = x^3 + bx^2 + ax + b$ $g_{03}(x) = g_{04}(x) = x^6 + x^5 + x^2 + x + 1$ | $p_{04}(x) = x^6 + x^5 + x^2 + x + 1$ |
| $p_{05}(x) = x^3 + ax^2 + ax + a$ e $p_{07}(x) = x^3 + bx^2 + bx + b$ $g_{05}(x) = g_{07}(x) = x^6 + x^5 + x^3 + x^2 + 1$ | $p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$ |
| $p_{08}(x) = x^3 + bx^2 + x + a$ e $p_{11}(x) = x^3 + ax^2 + x + b$ $g_{08}(x) = g_{11}(x) = x^6 + x^5 + x^4 + x + 1$ | $p_{06}(x) = x^6 + x^5 + x^4 + x + 1$ |
| $GF(4^4)$ | $GR(4, 8)$ |
| $p_{01}(x) = x^4 + x^3 + ax^2 + ax + a$ e $p_{02}(x) = x^4 + x^3 + bx^2 + b$ $g_{01}(x) = g_{02}(x) = x^8 + x^4 + x^3 + x^2 + 1$ | $p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$ |
| $p_{03}(x) = x^4 + bx^2 + ax + a$ e $p_{06}(x) = x^4 + ax^2 + bx + b$ $g_{03}(x) = g_{06}(x) = x^8 + x^4 + x^6 + x^5 + x^3 + 1$ | $p_{02}(x) = x^8 + x^4 + x^6 + x^5 + x^3 + 1$ |
| $p_{04}(x) = x^4 + bx^3 + ax + b$ e $p_{08}(x) = x^4 + ax^3 + bx + a$ $g_{04}(x) = g_{08}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ | $p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1$ |
| $p_{05}(x) = x^4 + x^2 + bx + a$ e $p_{10}(x) = x^4 + x^2 + ax + b$ $g_{05}(x) = g_{10}(x) = x^8 + x^5 + x^3 + x + 1$ | $p_{04}(x) = x^8 + x^5 + x^3 + x + 1$ |
| $p_{07}(x) = x^4 + ax^2 + ax + a$ e $p_{14}(x) = x^4 + bx^2 + bx + b$ $g_{07}(x) = g_{14}(x) = x^8 + x^6 + x^5 + x^2 + 1$ | $p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1$ |
| $p_{09}(x) = x^4 + ax^2 + ax + b$ e $p_{17}(x) = x^4 + bx^2 + bx + a$ $g_{09}(x) = g_{17}(x) = x^8 + x^6 + x^5 + x + 1$ | $p_{06}(x) = x^8 + x^6 + x^5 + x + 1$ |
| $p_{11}(x) = x^4 + ax^3 + ax^2 + b$ e $p_{20}(x) = x^4 + bx^3 + bx^2 + a$ $g_{11}(x) = g_{20}(x) = x^8 + x^7 + x^3 + x^2 + 1$ | $p_{07}(x) = x^8 + x^7 + x^3 + x^2 + 1$ |
| $p_{12}(x) = x^4 + x^3 + bx^2 + a$ e $p_{23}(x) = x^4 + x^3 + ax^2 + b$ $g_{12}(x) = g_{23}(x) = x^8 + x^5 + x^3 + x^2 + 1$ | $p_{08}(x) = x^8 + x^5 + x^3 + x^2 + 1$ |
| $p_{13}(x) = x^4 + x^3 + x + a$ e $p_{15}(x) = x^4 + x^3 + x + b$ $g_{13}(x) = g_{15}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ | $p_{09}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$ |
| $p_{16}(x) = x^4 + bx^3 + x^2 + bx + a$ e $p_{24}(x) = x^4 + ax^3 + x^2 + ax + b$ $g_{16}(x) = g_{24}(x) = x^8 + x^7 + x^6 + x + 1$ | $p_{10}(x) = x^8 + x^7 + x^6 + x + 1$ |
| $p_{18}(x) = x^4 + ax^3 + bx^2 + b$ e $p_{26}(x) = x^4 + bx^3 + ax^2 + a$ $g_{18}(x) = g_{26}(x) = x^8 + x^7 + x^5 + x^3 + 1$ | $p_{11}(x) = x^8 + x^7 + x^5 + x^3 + 1$ |
| $p_{19}(x) = x^4 + ax^3 + bx^2 + ax + b$ e $p_{28}(x) = x^4 + bx^3 + ax^2 + bx + a$ $g_{19}(x) = g_{28}(x) = x^8 + x^7 + x^2 + x + 1$ | $p_{12}(x) = x^8 + x^7 + x^2 + x + 1$ |
| $p_{21}(x) = x^4 + x^3 + x^2 + b$ e $p_{27}(x) = x^4 + x^3 + x^2 + a$ $g_{21}(x) = g_{27}(x) = x^8 + x^6 + x^3 + x^2 + 1$ | $p_{13}(x) = x^8 + x^6 + x^3 + x^2 + 1$ |
| $p_{22}(x) = x^4 + bx^3 + x + a$ e $p_{30}(x) = x^4 + ax^3 + x + b$ $g_{22}(x) = g_{30}(x) = x^8 + x^6 + x^3 + x^2 + 1$ | $p_{14}(x) = x^8 + x^6 + x^3 + x^2 + 1$ |
| $p_{25}(x) = x^4 + ax^3 + ax + a$ e $p_{29}(x) = x^4 + bx^3 + bx + b$ $g_{25}(x) = g_{29}(x) = x^8 + x^7 + x^6 + x^4 + x^2 + 1$ | $p_{15}(x) = x^8 + x^7 + x^6 + x^4 + x^2 + 1$ |
| $p_{31}(x) = x^4 + x^3 + x^2 + ax + a$ e $p_{32}(x) = x^4 + x^3 + x^2 + bx + b$ $g_{08}(x) = g_{11}(x) = x^8 + x^6 + x^5 + x^4 + 1$ | $p_{16}(x) = x^8 + x^6 + x^5 + x^4 + 1$ |

Tabela 5.19: Relação entre os polinômios geradores do corpo com os polinômios primitivos do anel.

Das análises feitas com as sequências de DNA no corpo $GF(4^r)$ e no anel $GR(4, 6)$, apresentadas resumidamente nas Tabelas 5.1 e 5.20, observamos que a maioria das sequências possuem apenas a estrutura de anel, sendo menos da metade das sequências que apresentaram as estruturas de corpo e de anel. Com isso, conluímos que toda sequência reproduzida no corpo também será reproduzida no anel, mas nem toda sequência reproduzida no anel será reproduzida no corpo. Um fato esperado.

| Estrutura Algébrica | | $C(n, k, d_H)$ | $p(x)/g(x)$ | Total de Seq. | $D(\mathbf{a}, \mathbf{b}) = 1$ | $D(\mathbf{a}, \mathbf{b}) = 2$ |
|-----------------------|---------------------|---|----------------------------------|-------------------|---------------------------------|---------------------------------|
| A E N E L | $r = 6$ | (21, 15, 3) (63, 57, 3) | 6/2 6/6 | 6 27 | 0 $10^* + 4$ | 2 27 |
| | $r = 8$ | (51, 43, 3) (255, 57, 3) | 16/4 16/16 | 13 10 | 0 5 | 12 10 |
| | $r = 9$ | (511, 502, 3) | 48/48 | 1 | 1 | 1 |
| | $r = 10$ | (93, 83, 3) (1023, 1013, 3) | 60/6 60/60 | 5 10 | 0 6 | 2 10 |
| | $r = 11$ | (2047, 2036, 3) | 176/176 | 1 | 1 | 1 |
| | $r = 12$ | (39, 27, 3) (45, 33, 3) (105, 93, 3) (195, 183, 3) | 144/- 144/- 144/- 144/- | 5 11 2 1 | 0 0 0 0 | 0 0 0 0 |
| Total das análises: | | | | 92 | $10^* + 17$ | 65 |
| C O R P O | $r = 3$ | (21, 15, 3) (63, 57, 3) | 12/2 12/2 | 6 27 | 0 6 | - - |
| | $r = 4$ | (51, 43, 3) (255, 57, 3) | 32/4 32/4 | 13 10 | 0 2 | - - |
| | $r = 10$ | (93, 83, 3) (1023, 1013, 3) | 120/6 120/6 | 5 10 | 0 2 | - - |
| | Total das análises: | | | 71 | 10 | - |

Tabela 5.20: Quantidade das sequências de DNA analisadas em anéis e corpos com os padrões de erro $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$.

Do ponto de vista de CCE é conhecido na literatura que a estrutura de corpo é mais inflexível que a estrutura de anel. Com isso, inferimos que as sequências de DNA reproduzidas sobre corpos contêm propriedades matemáticas na sua estrutura genômica mais rígidas que as demais sequências reproduzidas sobre anéis.

Novamente inferimos que essa propriedade de inflexibilidade nestas sequências pode estar relacionada à importância da sua estrutura topológica (função biológica da sequência), ou seja, a sua estrutura deve ser mais protegida para que a sequência de DNA não perca a sua função; ou pode estar relacionado também ao índice mutacional das sequências de DNA genômicas, ou seja, as sequências reproduzidas sobre corpos foram originadas primeiro que as sequências reproduzidas em anéis. Para comprovação dessas observações, é necessário um estudo laboratorial detalhado.

5.7.1 Grau das extensões de Galois r e os comprimentos das sequências de DNA.

Em [26], é conhecido que as sequências de direcionamento possuem comprimentos entre 13 e 85 aminoácidos. Aqui, determinamos em quais extensões de Galois os códigos G -linearidade (BCH primitivos e BCH não primitivos) deverão ser construídos obedecendo os comprimentos das sequências de DNA. A Tabela 5.21, ilustra alguns comprimentos de sequências, as respectivas extensões de Galois de grau r e, alguns valores de d para a construção destes códigos.

| Comprimento em aminoácidos | Comprimento em nucleotídeos | Grau da extensão de Galois (anel/corpo) | Valor de d para anel/corpo |
|----------------------------|-----------------------------|---|------------------------------|
| 7 | 21 | $r = 06/03$ | $d = 6/3$ |
| 13 | 39 | $r = 12/06$ | $d = 210/105$ |
| 15 | 45 | $r = 12/06$ | $d = 182/91$ |
| 17 | 51 | $r = 08/04$ | $d = 10/5$ |
| 19 | 57 | $r = 18/09$ | — |
| 21 | 63 | $r = 06/03$ | $d = 2/1$ |
| 23 | 69 | $r = 22/11$ | — |
| 25 | 75 | $r = 20/10$ | — |
| 27 | 81 | $r = 54/27$ | — |
| 29 | 87 | $r = 28/14$ | — |
| 31 | 93 | $r = 10/05$ | $d = 22/11$ |
| 33 | 99 | $r = 30/15$ | — |
| 35 | 105 | $r = 12/06$ | $d = 78/39$ |
| 37 | 111 | $r = 36/18$ | — |
| 39 | 117 | $r = 12/06$ | $d = 70/35$ |
| 41 | 123 | $r = 20/10$ | — |
| 43 | 129 | $r = 14/07$ | $d = 254/127$ |
| 45 | 135 | $r = 36/18$ | — |
| 47 | 141 | $r = 46/23$ | — |
| 49 | 147 | $r = 42/21$ | — |
| 51 | 153 | $r = 24/22$ | — |
| 53 | 159 | $r = 52/26$ | — |
| 55 | 165 | $r = 20/10$ | — |
| 57 | 171 | $r = 18/09$ | — |
| 59 | 177 | $r = 58/29$ | — |
| 61 | 183 | $r = 60/30$ | — |
| 63 | 189 | $r = 18/09$ | — |
| 65 | 195 | $r = 12/06$ | $d = 42/21$ |
| 67 | 201 | $r = 66/33$ | — |
| 69 | 207 | $r = 66/33$ | — |
| 71 | 213 | $r = 70/35$ | — |
| 73 | 219 | $r = 18/09$ | — |
| 75 | 225 | $r = 60/30$ | — |
| 77 | 231 | $r = 30/15$ | — |
| 79 | 237 | $r = 78/39$ | — |
| 81 | 243 | $r = 88/44$ | — |
| 83 | 249 | $r = 82/41$ | — |
| 85 | 255 | $r = 08/04$ | $d = 2/1$ |
| 341 | 1023 | $r = 10/05$ | $d = 2/1$ |
| 1365 | 4095 | $r = 12/06$ | $d = 2/1$ |

Tabela 5.21: Grau das extensões de Galois r e os comprimentos das sequências de DNA.

As sequências com comprimentos $n = 21$, $n = 93$, $n = 1023$ e $n = 4095$ nucleotídeos não

são sequências de direcionamento, apenas estão ilustradas na Tabela 5.21 para identificar o grau r da extensão de Galois e o valor de d . Os comprimentos que estão em vermelho foram analisados através dos códigos BCH não primitivos; os comprimentos em azul foram analisados através dos códigos BCH primitivos, e, as demais sequências, não foram analisadas neste trabalho em virtude da dificuldade em encontrar sequências de DNA no *NCBI* com estes comprimentos, sem contar com a complexidade computacional envolvida nos cálculos das extensões de Galois de grau maiores.

Capítulo 6

Conclusões e Perspectivas Futuras

Tanto as teorias da informação, comunicação e codificação quanto a genética preocu-pam-se com a transferência da informação. Essa transferência da informação nessas teorias são realizadas pelo homem consistindo de um poderoso ferramental conceitual, enquanto que, a transferência da informação genética é realizada por um processo natural de maneira excepcional e intrigante garantindo a evolução e perpetuação das espécies. Por exemplo, a teoria da comunicação estabelece o envio de mensagens de um lugar para outro, já a genética estabelece o envio de mensagens hereditárias no tempo.

No mundo moderno, o desenvolvimento constante de tecnologias de comunicação nos mais diversificados ramos de aplicação deve-se a consistência matemática garantida por essas teorias. Por outro lado, a evolução da biologia molecular deve-se, principalmente, a descoberta e elucidação da dupla hélice do DNA onde as bases moleculares da genética foram estabelecidas proporcionando o melhor entendimento da ciência da vida e o desenvolvimento de novas áreas relacionadas a biotecnologia. A descoberta do DNA deixou claro que a informação genética composta por quatro letras está armazenada na dupla hélice do DNA e, é transmitida de geração em geração.

No presente trabalho, apontamos várias vezes que os benefícios entre o casamento dessas teorias com a ciência da vida são enormes. O ferramental conceitual em teoria de informação, comunicação e codificação aplicados em sistemas biológicos já é praticado por cientistas há várias décadas. Um dos grandes desafios da comunidade científica em teorias de informação genética, comunicação genética e codificação genética é determinar uma estrutura matemática relacionada a estrutura do DNA. A utilização adequada do ferramental, bem como, o esta-belecimento e a caracterização matemática de uma linguagem acessível para ambas as áreas não são tarefas simples de serem implementadas de modo que a estrutura matemática do DNA seja determinada.

Este trabalho, além de mostrar que as semelhanças entre os sistemas de comunicação e

os sistemas biológicos são enormes, aponta para a existência de uma estrutura matemática nas fitas simples do DNA (informação genética) e na dupla hélice do DNA (informação genômica) associada aos CCEs e, portanto é possível uma caracterização matemática desses sistemas. Com base nessas semelhanças propomos o modelo de codificação genética e o modelo de codificação genômica. O primeiro modelo identifica a sequência de nucleotídeos de uma fita simples do DNA como palavra-código de um código G -linearidade e, o segundo identifica a sequência das bases complementares da fita dupla do DNA como palavra-código de um código G -linearidade. Ambos os modelos são capazes de identificar, reproduzir e classificar matematicamente diferentes sequências de DNA. A caracterização desses modelos e a reprodução dessas sequências possibilitam divisar metodologias que poderá ser aplicada em análises mutacionais e de polimorfismos, produção de novos fármacos, melhoramento genético, entre outros, reduzindo tempo e custos laboratoriais, como proposto no processo descrito em [21].

O grande diferencial deste trabalho está na possibilidade de identificar uma estrutura matemática em diferentes sequências de DNA. Até onde é de nosso conhecimento, pela primeira vez sequências de DNA com características e funções biológicas distintas, incluindo proteínas, gene e genoma em termos das fitas simples do DNA e da dupla hélice do DNA são reproduzidas e classificadas matematicamente. Dentre as sequências de DNA reproduzidas e classificadas matematicamente, temos sequências relacionadas a vírus e bactérias causadoras de patologias clínicas, tais como: sequência de direcionamento do vírus *Influenza A H1N1* e sequências referente as proteínas p53 e BCRA1 relacionadas ao câncer. Outra novidade é com relação a existência de códigos concatenados entre algumas sequências de direcionamento e suas respectivas proteínas organelares apresentada neste trabalho. Além disso, foi possível a constatação da arquitetura biológica (*Biological frame*) do genoma do plasmídeo *Lactococcus lactis*.

6.1 Desenvolvimento do Trabalho

O Capítulo 2 é introdutório, nele apresentamos alguns conceitos da biologia celular e molecular com ênfase no papel do DNA e a síntese de proteínas e, os principais conceitos e definições da teoria de CCEs utilizados no trabalho.

Nos Capítulos 3,4 e 5 encontram-se as contribuições deste trabalho. No Capítulo 3 propomos um modelo de um sistema de comunicação de informação genética e um modelo de um sistema de comunicação de informação genômica, ambos os modelos análogos ao modelo de um sistema de comunicação digital. Nestes modelos o bloco transmissor do sistema de comunicação é caracterizado matematicamente, onde o codificador genético e o codificador

genômico são estabelecidos. O diferencial deste modelo é a utilização de conceitos atuais empregados na teoria da comunicação, tais como: códigos geometricamente uniformes, códigos G -linearidade, mapeamento casado, códigos sobre anéis e sobre corpos na identificação, reprodução e classificação matemática de diferentes sequências de DNA.

Ainda neste capítulo, mostramos as semelhanças entre os sistemas biológicos e os procedimentos usados na comunicação em redes locais de computadores, dentre eles, os protocolos, a arquitetura, a topologia das redes e a estruturação dos quadros. Com isso, duas conjecturas são feitas: a) A informação genômica pode ser organizada e armazenada de maneira similar a informação armazenada em CD's; b) Os procedimentos de armazenamento e de transmissão da informação genômica seguem similarmente aos procedimentos da transmissão de dados tornando possível a montagem de um quadro da informação genômica semelhante a montagem do quadro *ethernet* em redes locais de computadores onde os protocolos são estabelecidos de acordo com o interesse.

No Capítulo 4 apresentamos o desenvolvimento de algoritmos computacionais responsáveis pela identificação de sequências de DNA através da codificação. Os passos dos algoritmos descrevem a construção dos códigos G -linearidade, tais como: os códigos BCH sobre o corpo $GF(4^r)$ e, os códigos BCH sobre o anel \mathbb{Z}_4 através das suas extensões de Galois, respectivamente. Na construção dos códigos G -linearidade, abordamos as extensões de Galois de grau $r = 3$ para corpo e as extensões de Galois de grau $r = 6, r = 8$ e $r = 10$ para anel, assim como, as possibilidades dos códigos BCH primitivos e dos códigos BCH não primitivos com os padrões de erros $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$ (1 e 2 nucleotídeos diferindo da sequência do NCBI).

Os algoritmos além de identificarem uma estrutura de CCEs nas sequências de DNA, permitem uma nova abordagem para a classificação destas sequências sob um ponto de vista matemático. As sequências de DNA que foram reproduzidas pelo código \mathbb{Z}_4 -linearidade (código BCH sobre anel e rotulamento A) são classificadas como sequências não-lineares. Enquanto, que as sequências reproduzidas pelos códigos $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade (códigos BCH sobre anel rotulamento B e rotulamento C, respectivamente) são classificadas como sequências lineares.

No Capítulo 5 fazemos uma análise dos resultados e das observações relevantes decorrentes da caracterização matemática dos modelos de codificação genética, codificação genômica, da construção dos códigos G -linearidade e, dos processos de geração destes códigos, bem como, dos processos de identificação e reprodução das sequências de DNA. De maneira resumida, mostramos todas as sequências que foram geradas e reproduzidas através dos códigos G -linearidades: códigos BCH sobre o corpo $GF(4^r)$ e códigos BCH sobre o anel $GR(4, r)$ e suas extensões de Galois, respectivamente.

6.2 Contribuições do Trabalho

O estudo e a aplicação do ferramental conceitual em teorias da codificação e da comunicação proporcionou, no presente trabalho, apresentarmos contribuições nas áreas da teoria da comunicação (genética e genômica) e da codificação (genética e genômica), bem como na área da bioinformática, da seguinte maneira:

- A caracterização matemática do modelo de codificação genética para a identificação de sequências da fita simples do DNA;
- A caracterização matemática do modelo de codificação genômica para a identificação de sequências da dupla hélice do DNA;
- A construção e análises de diferentes CCEs, sendo códigos G -linearidade: BCH primitivos sobre corpos, códigos BCH primitivos e não primitivos sobre anéis;
- O desenvolvimento dos algoritmos de identificação das sequências de DNA sobre as estruturas algébricas de corpo e anel;
- Os algoritmos permitem a identificação dessas sequências com até dois nucleotídeos diferindo das sequências postadas no *NCBI*;
- A reprodução de diferentes sequências de DNA com funcionalidades biológicas diversas, tais como: sequências de direcionamento, sequência de direcionamento ambígua, enzima, sinal interno, hormônio, ítron, DNA repetitivo, proteínas, micro RNA (*miRNA*), gene e genoma.
- A identificação da característica de ciclicidade nas sequências de DNA reproduzidas;
- A representação algébrica da dupla hélice do DNA;
- A representação algébrica entre as fitas codantes e não codantes através dos seus polinômios primitivos/geradores e seus recíprocos;
- A relação entre polinômios primitivos e polinômios geradores do corpo e do anel;
- A determinação de classes de polinômios primitivos/geradores tanto em corpo quanto em anel;
- A classificação das sequências de DNA, sob o ponto de vista matemático: sequências não-lineares (código \mathbb{Z}_4 -linearidade), sequências lineares (códigos $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade);

- A identificação da existência de códigos concatenados na estrutura do DNA, em algumas sequências;
- A constatação da arquitetura biológica do genoma do plasmídeo *Lactococcus lactis*.
- A exploração do modelo de codificação genética em análises mutacionais e polimorfismos em sequências de DNA;

Parte desse trabalho, referente às sequências de DNA identificadas e reproduzidas através do código \mathbb{Z}_4 -lineariade, foi apresentado no congresso internacional em [64]. Outra parte do trabalho, relacionada à aplicações do processo de identificação e reprodução das sequências de DNA foi apresentado em [65]. Os resultados da identificação e reprodução das sequências de DNA sobre $GF(4)$ foi publicado em [66], que foi selecionado o melhor artigo do mês de fevereiro de 2010 pela revista inglesa. O desenvolvimento do algoritmo para geração de sequências de DNA, cujos resultados podem ser direcionados em metodologias voltadas às análises mutacionais e de polimorfismos foi descrito em [21].

6.3 Sugestões para Trabalhos Futuros

A fim de dar prosseguimento a este trabalho, apresentamos algumas sugestões:

- O estudo da proteção desigual em relação aos nucleotídeos associados às sequências de DNA.
- A capacidade de correção dos códigos que reproduziram as sequências de DNA é igual a $D_H \geq 3$. Neste trabalho consideramos os padrões de erro iguais a $D(\mathbf{a}, \mathbf{b}) = 1$ e $D(\mathbf{a}, \mathbf{b}) = 2$. Para $D(\mathbf{a}, \mathbf{b}) = 1$ está dentro da capacidade de correção do código, porém, no caso de $D(\mathbf{a}, \mathbf{b}) = 2$ consideramos em virtude do arranjo padrão do código ser capaz de corrigir alguns erros iguais a $D(\mathbf{a}, \mathbf{b}) = 2$. É de suma importância a determinação e caracterização destes arranjos padrões e uma análise detalhada dos resultados da flexibilização dos códigos descritos neste trabalho;
- Do ponto de vista de CCEs é conhecido que a estrutura de corpo é mais inflexível que a estrutura de anel. Uma análise laboratorial detalhada das sequências de DNA reproduzidas sobre as estrutura de anéis e corpos ao mesmo tempo deve ser estudada. Os resultados dessa análise poderão ser úteis na determinação da importância da estrutura topológica (função biológica) dessas sequências, ou seja, a sua estrutura deve ser mais protegida para que a sequência de DNA não perca a sua função; ou pode estar relacionado ao índice mutacional das sequências de DNA genômicas;

- A compreensão porque determinados polinômios primitivos são responsáveis pela reprodução de algumas sequências de DNA é fundamental. Esse entendimento levará a uma melhor caracterização da relação entre esses polinômios e a reprodução dessas sequências e, possibilitará estudar uma possível elucidação da estrutura topológica das proteínas;
- No caso dos códigos BCH não primitivos deve ser estudado uma generalização para determinar a classe dos polinômios geradores;
- É necessário uma análise evolutiva das sequências de DNA reproduzidas neste trabalho pelos códigos *nested*. O entendimento e uma melhor caracterização desses códigos;
- Embora este trabalho tenha abordado diferentes sequências de DNA com diferentes comprimentos, foi necessário considerarmos a restrição de comprimento: $n = (4^r - 1)$ para o corpo, e, $n = (2^r - 1)$ para o anel. Este fator deve-se à complexidade do assunto. Salientamos que é de grande interesse que os resultados apresentados neste trabalho sejam generalizados para qualquer comprimento;
- Observamos que a característica de ciclicidade presente nas matrizes geradoras dos códigos BCH sobre $GF(4^r)$ e BCH sobre $GR(4, r)$ construídos neste trabalho, são as mesmas presentes nas matrizes geradoras dos códigos convolucionais. Outro ponto importante é que os códigos convolucionais podem ser construídos de acordo com o comprimento desejado. Portanto, uma nova abordagem utilizando códigos convolucionais, poderá contribuir para o desenvolvimento nesta área;
- As sequências de DNA que não foram reproduzidas pelos códigos *G*-linearidade neste trabalho, provavelmente, essas sequências não apresentam a característica de ciclicidade na sua formação. O estudo de funções algébricas conhecidas como Funções *Booleanas* pode ser útil no entendimento do processo de geração e reprodução dessas sequências ampliando consideravelmente a pesquisa;
- O Estudo sobre as semelhanças entre a topologia das estruturas primárias das sequências reproduzidas pelos códigos \mathbb{Z}_4 -linearidade, $\mathbb{Z}_2 \times \mathbb{Z}_2$ -linearidade e Klein-linearidade e as estruturas da *alpha*-hélice, folha *beta* paralela e folha *beta* anti-paralela deve ser abordado.
- Uma abordagem entre a topologia da estrutura primária e a topologia da estrutura secundária das sequências de DNA fica como um forte apelo para uma proposta futura por ser um assunto de extrema importância, tanto no âmbito da biologia molecular quanto na teoria de codificação genética e genômica. Os benefícios que os laboratórios

e as indústrias nas áreas farmacêutica e biotecnologia podem galgar com os resultados dessa abordagem são enormes;

- O sistema de transmissão de informação têm muito a contribuir com o entendimento do sistema biológico e vice-versa. Os conceitos e as propriedades presentes nesses sistemas devem ser explorados e melhor caracterizados, tornando possível uma modelagem matemática apropriada. Por exemplo: a comunicação célula-célula ou múlti-células, os processos de comunicação intercelular e intracelular, as estratégias de comunicação, bem como, os fatores em relação ao distanciamento entre as células, todos eles guardam semelhanças com as rede de computadores: os protocolos de comunicação, a arquitetura da rede, o quadro/pacote e a própria topologia da rede. Estudar o genoma como um conjunto de protocolos pode contribuir para novas analogias entre o sistema de transmissão e o genoma.

6.4 Comentário Final

A pesquisa em áreas de fronteira da ciência produz conhecimento, tecnologia e crescimento para um país. A exemplo do que ocorre em outros países com estudos mais avançados em biologia celular, molecular e biotecnologia, a parceria entre as áreas da matemática, da biologia, da física, da química e da teoria de informação e codificação, podem render bons resultados no mundo acadêmico e no âmbito industrial e comercial. Com a necessidade de redução de tempo e custos nos experimentos laboratoriais, o modelo apresentado neste trabalho propõe uma abordagem metemática capaz de identificar, reproduzir e classificar sequências de DNA matematicamente, possibilitando o desenvolvimento de uma metodologia que poderá ser aplicada em análises mutacionais e de polimorfismos nas sequências de DNA.

Ressaltamos que o modelo pode ser ajustado na análise de qualquer sequência de DNA produtora de proteínas dentro da célula. Ao conhecermos a estrutura matemática da proteína é possível alterar a ordem das bases e também corrigir as mutações ou erros que possam acontecer para voltar à condição normal de uma proteína. Outra possibilidade seria fabricar proteínas a partir do código matemático ou ainda encontrar proteínas não conhecidas existentes nas células. Hoje se faz uma alteração na sequência de DNA que codifica uma proteína e depois são feitos os testes em laboratório para verificar a eficácia da reação num experimento de tentativa e erro. Com o modelo matemático será possível testar a afinidade e a estabilidade da proteína em um trabalho preliminar de forma a verificar mutações e, posteriormente, testá-las a partir de experimentos laboratoriais para confirmar se a mutação na sequência de DNA deu o resultado esperado.

Com um aprimoramento do modelo e o desenvolvimento de um software, teremos um leque de aplicações de grande interesse industrial e comercial bastante abrangente. Pode ser usado para a produção de novos fármacos, criação de novos produtos, aplicação em empresas de agronegócios, melhoramento genético, indústrias têxteis, e, de biotecnologia de um modo geral, [21].

Apêndice A

Códigos BCH $C(n, k, d_H)$ sobre $GF(4^r)$ e sobre $GR(4, r)$

| t | $d_H \geq 2t + 1$ | $\alpha^1, \dots, \alpha^{2t}$ | $g(x) = mmc(M_1, M_2), \dots, M_{n-1}$ | $C(n, k, d_H)$ |
|-----|-------------------|----------------------------------|--|-----------------|
| 31 | $d_H \geq 63$ | $(\alpha^1, \dots, \alpha^{62})$ | $g(x) = mmc(M_1, \dots, M_{62})$ | $C(63, 1, 63)$ |
| 30 | $d_H \geq 61$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{60})$ | $C(63, 1, 61)$ |
| 29 | $d_H \geq 59$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{58})$ | $C(63, 5, 59)$ |
| 28 | $d_H \geq 57$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{56})$ | $C(63, 7, 57)$ |
| 27 | $d_H \geq 55$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{54})$ | $C(63, 9, 55)$ |
| 26 | $d_H \geq 53$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{52})$ | $C(63, 11, 53)$ |
| 25 | $d_H \geq 51$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{50})$ | $C(63, 13, 51)$ |
| 24 | $d_H \geq 49$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{48})$ | $C(63, 15, 49)$ |
| 23 | $d_H \geq 47$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{46})$ | $C(63, 17, 47)$ |
| 22 | $d_H \geq 45$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{44})$ | $C(63, 15, 45)$ |
| 21 | $d_H \geq 43$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{42})$ | $C(63, 17, 43)$ |
| 20 | $d_H \geq 41$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{40})$ | $C(63, 19, 41)$ |
| 19 | $d_H \geq 39$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{38})$ | $C(63, 21, 39)$ |
| 18 | $d_H \geq 37$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{36})$ | $C(63, 23, 37)$ |
| 17 | $d_H \geq 35$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{34})$ | $C(63, 25, 35)$ |
| 16 | $d_H \geq 33$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{32})$ | $C(63, 27, 33)$ |
| 15 | $d_H \geq 31$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{30})$ | $C(63, 29, 31)$ |
| 14 | $d_H \geq 29$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{28})$ | $C(63, 31, 29)$ |
| 13 | $d_H \geq 27$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{26})$ | $C(63, 33, 27)$ |
| 12 | $d_H \geq 25$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{24})$ | $C(63, 35, 25)$ |
| 11 | $d_H \geq 23$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{22})$ | $C(63, 37, 23)$ |
| 10 | $d_H \geq 21$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{20})$ | $C(63, 39, 21)$ |
| 09 | $d_H \geq 19$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{18})$ | $C(63, 41, 19)$ |
| 08 | $d_H \geq 17$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{16})$ | $C(63, 43, 17)$ |
| 07 | $d_H \geq 15$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{14})$ | $C(63, 45, 15)$ |
| 06 | $d_H \geq 13$ | $(\alpha^1, \dots, \alpha^{60})$ | $g(x) = mmc(M_1, \dots, M_{12})$ | $C(63, 47, 13)$ |
| 05 | $d_H \geq 11$ | $(\alpha^1, \dots, \alpha^{10})$ | $g(x) = mmc(M_1, \dots, M_{10})$ | $C(63, 49, 11)$ |
| 04 | $d_H \geq 09$ | $(\alpha^1, \dots, \alpha^8)$ | $g(x) = mmc(M_1, \dots, M_8)$ | $C(63, 51, 9)$ |
| 03 | $d_H \geq 07$ | $(\alpha^1, \dots, \alpha^6)$ | $g(x) = mmc(M_1, \dots, M_6)$ | $C(63, 53, 7)$ |
| 02 | $d_H \geq 05$ | $(\alpha^1, \dots, \alpha^4)$ | $g(x) = mmc(M_1, \dots, M_4)$ | $C(63, 55, 5)$ |
| 01 | $d_H \geq 03$ | (α^1, α^2) | $g(x) = mmc(M_1, M_2)$ | $C(63, 57, 3)$ |

Tabela A.1: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(63, k, d_H)$ BCH primitivos sobre $GF(64)$ - extensão de Galois $r = 3$.

| t | $d_H \geq 2t + 1$ | $\alpha^1, \dots, \alpha^{2^t}$ | $g(x) = mmc\{M_1(x), M_2(x), \dots, M_{n-1}(x)\}$ | $C(n, k, d_H)$ |
|----------|-------------------|-----------------------------------|---|------------------|
| 127 | $d_H \leq 255$ | $(\alpha^1, \dots, \alpha^{254})$ | $g(x) = mmc\{M_1(x), \dots, M_{254}(x)\}$ | $C(255, 1, 255)$ |
| 126 | $d_H \geq 253$ | $(\alpha^1, \dots, \alpha^{252})$ | $g(x) = mmc\{M_1(x), \dots, M_{252}(x)\}$ | $C(255, k, 253)$ |
| 125 | $d_H \geq 251$ | $(\alpha^1, \dots, \alpha^{250})$ | $g(x) = mmc\{M_1(x), \dots, M_{250}(x)\}$ | $C(255, k, 251)$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| 04 | $d_H \geq 09$ | $(\alpha^1, \dots, \alpha^8)$ | $g(x) = mmc\{M_1(x), M_3(x), M_5(x), M_7(x)\}$ | $C(255, k, 9)$ |
| 03 | $d_H \geq 07$ | $(\alpha^1, \dots, \alpha^6)$ | $g(x) = mmc\{M_1(x), M_3(x), M_5(x)\}$ | $C(255, k, 7)$ |
| 02 | $d_H \geq 05$ | $(\alpha^1, \dots, \alpha^4)$ | $g(x) = mmc\{M_1(x), M_3(x)\}$ | $C(255, k, 5)$ |
| 01 | $d_H \geq 03$ | (α^1, α^2) | $g(x) = \{M_1(x)\}$ | $C(255, 247, 3)$ |

Tabela A.2: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos BCH primitivos em $GF(256)$ - extensão de Galois $r = 4$.

| t | $d_H \geq 2t + 1$ | $\beta^1, \dots, \beta^{2^t}$ | $g(x) = mmc(M_1(x), M_2(x), \dots, M_{n-1}(x))$ | $C(n, k, d_H)$ |
|-----|-------------------|--------------------------------|---|-----------------|
| 31 | $d_H \geq 63$ | $(\beta^1, \dots, \beta^{62})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15}, M_{21}, M_{23}, M_{27}, M_{31})$ | $C(63, 01, 63)$ |
| 30 | $d_H \geq 61$ | $(\beta^1, \dots, \beta^{60})$ | " | $C(63, 01, 61)$ |
| 29 | $d_H \geq 59$ | $(\beta^1, \dots, \beta^{58})$ | " | $C(63, 01, 59)$ |
| 28 | $d_H \geq 57$ | $(\beta^1, \dots, \beta^{56})$ | " | $C(63, 01, 57)$ |
| 27 | $d_H \geq 55$ | $(\beta^1, \dots, \beta^{54})$ | " | $C(63, 01, 55)$ |
| 26 | $d_H \geq 53$ | $(\beta^1, \dots, \beta^{52})$ | " | $C(63, 01, 53)$ |
| 25 | $d_H \geq 51$ | $(\beta^1, \dots, \beta^{50})$ | " | $C(63, 01, 51)$ |
| 24 | $d_H \geq 49$ | $(\beta^1, \dots, \beta^{48})$ | " | $C(63, 01, 49)$ |
| 23 | $d_H \geq 47$ | $(\beta^1, \dots, \beta^{46})$ | " | $C(63, 01, 47)$ |
| 22 | $d_H \geq 45$ | $(\beta^1, \dots, \beta^{44})$ | " | $C(63, 01, 45)$ |
| 21 | $d_H \geq 43$ | $(\beta^1, \dots, \beta^{42})$ | " | $C(63, 01, 43)$ |
| 20 | $d_H \geq 41$ | $(\beta^1, \dots, \beta^{40})$ | " | $C(63, 01, 41)$ |
| 19 | $d_H \geq 39$ | $(\beta^1, \dots, \beta^{38})$ | " | $C(63, 01, 39)$ |
| 18 | $d_H \geq 37$ | $(\beta^1, \dots, \beta^{36})$ | " | $C(63, 01, 37)$ |
| 17 | $d_H \geq 35$ | $(\beta^1, \dots, \beta^{34})$ | " | $C(63, 01, 35)$ |
| 16 | $d_H \geq 33$ | $(\beta^1, \dots, \beta^{32})$ | " | $C(63, 01, 33)$ |
| 15 | $d_H \geq 31$ | $(\beta^1, \dots, \beta^{30})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15}, M_{21}, M_{23}, M_{27})$ | $C(63, 07, 31)$ |
| 14 | $d_H \geq 29$ | $(\beta^1, \dots, \beta^{28})$ | " | $C(63, 07, 29)$ |
| 13 | $d_H \geq 27$ | $(\beta^1, \dots, \beta^{26})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15}, M_{21}, M_{23})$ | $C(63, 10, 27)$ |
| 12 | $d_H \geq 25$ | $(\beta^1, \dots, \beta^{24})$ | " | $C(63, 10, 25)$ |
| 11 | $d_H \geq 23$ | $(\beta^1, \dots, \beta^{22})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15}, M_{21})$ | $C(63, 16, 23)$ |
| 10 | $d_H \geq 21$ | $(\beta^1, \dots, \beta^{20})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15})$ | $C(63, 18, 21)$ |
| 09 | $d_H \geq 19$ | $(\beta^1, \dots, \beta^{18})$ | " | $C(63, 18, 19)$ |
| 08 | $d_H \geq 17$ | $(\beta^1, \dots, \beta^{16})$ | " | $C(63, 18, 17)$ |
| 07 | $d_H \geq 15$ | $(\beta^1, \dots, \beta^{14})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13})$ | $C(63, 24, 15)$ |
| 06 | $d_H \geq 13$ | $(\beta^1, \dots, \beta^{12})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11})$ | $C(63, 30, 13)$ |
| 05 | $d_H \geq 11$ | $(\beta^1, \dots, \beta^{10})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9)$ | $C(63, 36, 11)$ |
| 04 | $d_H \geq 09$ | $(\beta^1, \dots, \beta^8)$ | $g(x) = mmc(M_1, M_3, M_5, M_7)$ | $C(63, 39, 09)$ |
| 03 | $d_H \geq 07$ | $(\beta^1, \dots, \beta^6)$ | $g(x) = mmc(M_1, M_3, M_5)$ | $C(63, 45, 07)$ |
| 02 | $d_H \geq 05$ | $(\beta^1, \dots, \beta^4)$ | $g(x) = mmc(M_1, M_3)$ | $C(63, 51, 05)$ |
| 01 | $d_H \geq 03$ | (β^1, β^2) | $g(x) = mmc(M_1)$ | $C(63, 57, 03)$ |

Tabela A.3: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(63, k, d_H)$ BCH primitivos sobre $GR^*(4, 6)$ - extensão de Galois $r = 6$.

| t | $d_H \geq 2t + 1$ | $\beta^1, \dots, \beta^{2t}$ | $g(x) = mmc\{M_1(x), M_2(x), \dots, M_{n-1}(x)\}$ | $C(n, k, d_H)$ |
|----------|-------------------|---------------------------------|---|------------------|
| 127 | $d_H \leq 255$ | $(\beta^1, \dots, \beta^{254})$ | $g(x) = mmc\{M_1(x), \dots, M_{254}(x)\}$ | $C(255, 1, 255)$ |
| 126 | $d_H \geq 253$ | $(\beta^1, \dots, \beta^{252})$ | $g(x) = mmc\{M_1(x), \dots, M_{252}(x)\}$ | $C(255, k, 253)$ |
| 125 | $d_H \geq 251$ | $(\beta^1, \dots, \beta^{250})$ | $g(x) = mmc\{M_1(x), \dots, M_{250}(x)\}$ | $C(255, k, 251)$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| 04 | $d_H \geq 09$ | $(\beta^1, \dots, \beta^8)$ | $g(x) = mmc\{M_1(x), M_3(x), M_5(x), M_7(x)\}$ | $C(255, k, 9)$ |
| 03 | $d_H \geq 07$ | $(\beta^1, \dots, \beta^6)$ | $g(x) = mmc\{M_1(x), M_3(x), M_5(x)\}$ | $C(255, k, 7)$ |
| 02 | $d_H \geq 05$ | $(\beta^1, \dots, \beta^4)$ | $g(x) = mmc\{M_1(x), M_3(x)\}$ | $C(255, k, 5)$ |
| 01 | $d_H \geq 03$ | (β^1, β^2) | $g(x) = \{M_1(x)\}$ | $C(255, 247, 3)$ |

Tabela A.4: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(255, k, d_H)$ BCH primitivos em $GR^*(4, 8)$ - extensão de Galois $r = 8$.

| t | $d_H \geq 2t + 1$ | $\beta^1, \dots, \beta^{2t}$ | $g(x) = mmc(M_1(x), M_2(x), \dots, M_{n-1}(x))$ | $C(n, k, d_H)$ |
|-----|-------------------|--------------------------------|---|-----------------|
| 10 | $d_H \geq 21$ | $(\beta^1, \dots, \beta^{20})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{13}, M_{15})$ | $C(21, 01, 21)$ |
| 09 | $d_H \geq 19$ | $(\beta^1, \dots, \beta^{18})$ | " | $C(21, 01, 19)$ |
| 08 | $d_H \geq 17$ | $(\beta^1, \dots, \beta^{16})$ | " | $C(21, 01, 17)$ |
| 07 | $d_H \geq 15$ | $(\beta^1, \dots, \beta^{14})$ | " | $C(21, 01, 15)$ |
| 06 | $d_H \geq 13$ | $(\beta^1, \dots, \beta^{12})$ | " | $C(21, 01, 13)$ |
| 05 | $d_H \geq 11$ | $(\beta^1, \dots, \beta^{10})$ | " | $C(21, 01, 11)$ |
| 04 | $d_H \geq 09$ | $(\beta^1, \dots, \beta^8)$ | $g(x) = mmc(M_1, M_3, M_5, M_{07})$ | $C(21, 04, 09)$ |
| 03 | $d_H \geq 07$ | $(\beta^1, \dots, \beta^6)$ | $g(x) = mmc(M_1, M_3, M_5)$ | $C(21, 06, 07)$ |
| 02 | $d_H \geq 05$ | $(\beta^1, \dots, \beta^4)$ | $g(x) = mmc(M_1, M_3)$ | $C(21, 09, 05)$ |
| 01 | $d_H \geq 03$ | (β^1, β^2) | $g(x) = mmc(M_1)$ | $C(21, 15, 3)$ |

Tabela A.5: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(21, k, d_H)$ BCH não primitivos sobre $GR^*(4, 6)$ - extensão de Galois $r = 6$.

| t | $d_H \geq 2t + 1$ | $\beta^1, \dots, \beta^{2t}$ | $g(x) = mmc(M_1(x), M_2(x), \dots, M_{n-1}(x))$ | $C(n, k, d_H)$ |
|-----|-------------------|--------------------------------|--|-----------------|
| 25 | $d_H \geq 51$ | $(\beta^1, \dots, \beta^{50})$ | $g(x) = mmc(M_1, M_3, M_5, M_9, M_{11}, M_{17}, M_{19})$ | $C(51, 01, 51)$ |
| 24 | $d_H \geq 49$ | $(\beta^1, \dots, \beta^{48})$ | " | $C(51, 01, 49)$ |
| 23 | $d_H \geq 47$ | $(\beta^1, \dots, \beta^{46})$ | " | $C(51, 01, 47)$ |
| 22 | $d_H \geq 45$ | $(\beta^1, \dots, \beta^{44})$ | " | $C(51, 01, 45)$ |
| 21 | $d_H \geq 43$ | $(\beta^1, \dots, \beta^{42})$ | " | $C(51, 01, 43)$ |
| 20 | $d_H \geq 41$ | $(\beta^1, \dots, \beta^{40})$ | " | $C(51, 01, 41)$ |
| 19 | $d_H \geq 39$ | $(\beta^1, \dots, \beta^{38})$ | " | $C(51, 01, 39)$ |
| 18 | $d_H \geq 37$ | $(\beta^1, \dots, \beta^{36})$ | " | $C(51, 01, 37)$ |
| 17 | $d_H \geq 35$ | $(\beta^1, \dots, \beta^{34})$ | " | $C(51, 01, 35)$ |
| 16 | $d_H \geq 33$ | $(\beta^1, \dots, \beta^{32})$ | " | $C(51, 01, 33)$ |
| 15 | $d_H \geq 31$ | $(\beta^1, \dots, \beta^{30})$ | " | $C(51, 01, 31)$ |
| 14 | $d_H \geq 29$ | $(\beta^1, \dots, \beta^{28})$ | " | $C(51, 01, 29)$ |
| 13 | $d_H \geq 27$ | $(\beta^1, \dots, \beta^{26})$ | " | $C(51, 01, 27)$ |
| 12 | $d_H \geq 25$ | $(\beta^1, \dots, \beta^{24})$ | " | $C(51, 01, 25)$ |
| 11 | $d_H \geq 23$ | $(\beta^1, \dots, \beta^{22})$ | " | $C(51, 01, 23)$ |
| 10 | $d_H \geq 21$ | $(\beta^1, \dots, \beta^{20})$ | " | $C(51, 01, 21)$ |
| 09 | $d_H \geq 19$ | $(\beta^1, \dots, \beta^{18})$ | $g(x) = mmc(M_1, M_3, M_5, M_9, M_{11}, M_{17})$ | $C(51, 07, 19)$ |
| 08 | $d_H \geq 17$ | $(\beta^1, \dots, \beta^{16})$ | $g(x) = mmc(M_1, M_3, M_5, M_9, M_{11})$ | $C(51, 11, 17)$ |
| 07 | $d_H \geq 15$ | $(\beta^1, \dots, \beta^{14})$ | " | $C(51, 11, 15)$ |
| 06 | $d_H \geq 13$ | $(\beta^1, \dots, \beta^{12})$ | " | $C(51, 11, 13)$ |
| 05 | $d_H \geq 11$ | $(\beta^1, \dots, \beta^{10})$ | $g(x) = mmc(M_1, M_3, M_5, M_9)$ | $C(51, 19, 11)$ |
| 04 | $d_H \geq 09$ | $(\beta^1, \dots, \beta^8)$ | $g(x) = mmc(M_1, M_3, M_5)$ | $C(51, 27, 09)$ |
| 03 | $d_H \geq 07$ | $(\beta^1, \dots, \beta^6)$ | " | $C(51, 27, 07)$ |
| 02 | $d_H \geq 05$ | $(\beta^1, \dots, \beta^4)$ | $g(x) = mmc(M_1, M_3)$ | $C(51, 35, 05)$ |
| 01 | $d_H \geq 03$ | (β^1, β^2) | $g(x) = mmc(M_1)$ | $C(51, 43, 3)$ |

Tabela A.6: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(51, k, d_H)$ BCH não primitivos sobre $GR^*(4, 8)$ - extensão de Galois $r = 8$.

| t | $d_H \geq 2t + 1$ | $\beta^1, \dots, \beta^{2t}$ | $g(x) = mmc(M_1(x), M_2(x), \dots, M_{n-1}(x))$ | $C(n, k, d_H)$ |
|----------|-------------------|--------------------------------|---|-----------------|
| 46 | $d_H \geq 93$ | $(\beta^1, \dots, \beta^{92})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17}, M_{21}, M_{23}, M_{31}, M_{33}, M_{45})$ | $C(93, 01, 93)$ |
| 45 | $d_H \geq 91$ | $(\beta^1, \dots, \beta^{90})$ | " | $C(93, 01, 91)$ |
| \vdots | \vdots | \vdots | \vdots | \vdots |
| 23 | $d_H \geq 47$ | $(\beta^1, \dots, \beta^{46})$ | " | $C(93, 01, 47)$ |
| 22 | $d_H \geq 45$ | $(\beta^1, \dots, \beta^{44})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17}, M_{21}, M_{23}, M_{31}, M_{33})$ | $C(93, 06, 45)$ |
| 21 | $d_H \geq 43$ | $(\beta^1, \dots, \beta^{42})$ | " | $C(93, 06, 43)$ |
| 20 | $d_H \geq 41$ | $(\beta^1, \dots, \beta^{40})$ | " | $C(93, 06, 41)$ |
| 19 | $d_H \geq 39$ | $(\beta^1, \dots, \beta^{38})$ | " | $C(93, 06, 39)$ |
| 18 | $d_H \geq 37$ | $(\beta^1, \dots, \beta^{36})$ | " | $C(93, 06, 37)$ |
| 17 | $d_H \geq 35$ | $(\beta^1, \dots, \beta^{34})$ | " | $C(93, 06, 35)$ |
| 16 | $d_H \geq 33$ | $(\beta^1, \dots, \beta^{32})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17}, M_{21}, M_{23}, M_{31})$ | $C(93, 11, 33)$ |
| 15 | $d_H \geq 31$ | $(\beta^1, \dots, \beta^{30})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17}, M_{21}, M_{23})$ | $C(93, 13, 31)$ |
| 14 | $d_H \geq 29$ | $(\beta^1, \dots, \beta^{28})$ | " | $C(93, 13, 29)$ |
| 13 | $d_H \geq 27$ | $(\beta^1, \dots, \beta^{26})$ | " | $C(93, 13, 27)$ |
| 12 | $d_H \geq 25$ | $(\beta^1, \dots, \beta^{24})$ | " | $C(93, 13, 25)$ |
| 11 | $d_H \geq 23$ | $(\beta^1, \dots, \beta^{22})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17}, M_{21})$ | $C(93, 23, 23)$ |
| 10 | $d_H \geq 21$ | $(\beta^1, \dots, \beta^{20})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15}, M_{17})$ | $C(93, 28, 21)$ |
| 09 | $d_H \geq 19$ | $(\beta^1, \dots, \beta^{18})$ | " | $C(93, 28, 19)$ |
| 08 | $d_H \geq 17$ | $(\beta^1, \dots, \beta^{16})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11}, M_{15})$ | $C(93, 38, 17)$ |
| 07 | $d_H \geq 15$ | $(\beta^1, \dots, \beta^{14})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9, M_{11})$ | $C(93, 43, 15)$ |
| 06 | $d_H \geq 13$ | $(\beta^1, \dots, \beta^{12})$ | " | $C(93, 43, 13)$ |
| 05 | $d_H \geq 11$ | $(\beta^1, \dots, \beta^{10})$ | $g(x) = mmc(M_1, M_3, M_5, M_7, M_9)$ | $C(93, 53, 11)$ |
| 04 | $d_H \geq 09$ | $(\beta^1, \dots, \beta^8)$ | $g(x) = mmc(M_1, M_3, M_5, M_7)$ | $C(93, 58, 09)$ |
| 03 | $d_H \geq 07$ | $(\beta^1, \dots, \beta^6)$ | $g(x) = mmc(M_1, M_3, M_5)$ | $C(93, 68, 07)$ |
| 02 | $d_H \geq 05$ | $(\beta^1, \dots, \beta^4)$ | $g(x) = mmc(M_1, M_3)$ | $C(93, 78, 05)$ |
| 01 | $d_H \geq 03$ | (β^1, β^2) | $g(x) = mmc(M_1)$ | $C(93, 83, 3)$ |

Tabela A.7: Polinômios geradores $g(x)$ relativos às distâncias mínimas d_H dos códigos $(93, k, d_H)$ BCH não primitivos sobre $GR^*(4, 10)$ - extensão de Galois $r = 10$.

| Comprimento | $GR^*(4, r)$ | $d \leq n$ | $d \leq 2t + 1$ | t |
|-------------|---------------|-------------|-------------------------|--------------------|
| $n = 21$ | $GR^*(4, 6)$ | $d \leq 21$ | $d \leq 2 \cdot 10 + 1$ | $1 \leq t \leq 10$ |
| $n = 39$ | $GR^*(4, 12)$ | $d \leq 39$ | $d \leq 2 \cdot 19 + 1$ | $1 \leq t \leq 19$ |
| $n = 45$ | $GR^*(4, 12)$ | $d \leq 45$ | $d \leq 2 \cdot 22 + 1$ | $1 \leq t \leq 22$ |
| $n = 51$ | $GR^*(4, 8)$ | $d \leq 51$ | $d \leq 2 \cdot 25 + 1$ | $1 \leq t \leq 25$ |
| $n = 93$ | $GR^*(4, 10)$ | $d \leq 93$ | $d \leq 2 \cdot 46 + 1$ | $1 \leq t \leq 46$ |

Tabela A.8: Distâncias de Hamming de códigos BCH não primitivos relativos aos comprimentos n .

Apêndice B

Sequências de DNA Reproduzidas através dos Códigos *G*-linearidade

B.1 BCH Primitivo sobre $GF(4)$

Seq.17-SD | *T. sativum* – Retículo Endoplasmático – wPR4g gene for putative vacuolar defense protein – GI número 78096542

Código *G*-linearidade ((63,57,3) BCH primitivo sobre $GF(4)$, rotulamento S_4)

$$p_{08}(x) = x^3 + bx^2 + x + a - g_{08}(x) = x^6 + x^5 + x^4 + x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | A | A | R | L | A | L | V | A | A | L | L | C | A | G | A | T | A | A | A | A |
| ntO: | ATG | GCC | GCA | CGC | CTC | CGG | CTG | GTG | GCG | GCG | CTC | CTG | TGC | GCC | GGT | GCC | ACG | GCC | GCC | GCG | GCG |
| RtO: | Oba | a11 | a10 | a11 | 1b1 | ala | 1ba | aba | ala | ala | 1b1 | 1ba | bal | a11 | aab | a11 | 01a | a11 | a11 | a11 | ala |
| RtG: | Oba | a11 | a10 | a11 | 1b1 | ala | 1ba | ala | ala | 1b1 | 1ba | bal | b11 | aab | a11 | 01a | a11 | a11 | a11 | ala | ala |
| ntG: | ATG | GCC | GCA | CGC | CTC | CGG | CTG | GTG | GCG | GCG | CTC | CTG | TGC | TCC | GGT | GCC | ACG | GCC | GCG | GCG | GCG |
| aaG: | M | A | A | R | L | A | L | V | A | A | L | L | C | S | G | A | T | A | A | A | A |

Seq.19-SD | *S. cerevisiae* – Mitocôndria – 54S ribosomal protein – GI número 45269853

Código *G*-linearidade ((63,57,3) BCH primitivo sobre $GF(4)$, rotulamento S_4)

$$p_{10}(x) = x^3 + ax^2 + bx + b - g_{10}(x) = x^6 + x^5 + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | Q | K | I | F | R | P | F | Q | L | T | R | G | F | T | S | S | V | K | N | F |
| ntO: | ATG | CAA | AAA | ATT | TTC | AGA | CCA | TTC | CAA | TTA | ACG | AGA | GCG | TTT | ACC | TCT | TCC | GTA | AAA | AAC | TTC |
| RtO: | Oba | 100 | 000 | 0bb | bb1 | 0a0 | 110 | bb1 | 100 | bb0 | 01a | 0a0 | a11 | bbb | 011 | b1b | b11 | ab0 | 000 | 001 | bb1 |
| RtG: | Oba | 100 | 000 | 0 1b | bb1 | 0a0 | 110 | bb1 | 100 | bb0 | 01a | 0a0 | a11 | bbb | 011 | b1b | b11 | ab0 | 000 | 001 | bb1 |
| ntG: | ATG | CAA | AAA | AC T | TTC | AGA | CCA | TTC | CAA | TTA | ACG | AGA | GCG | TTT | ACC | TCT | TCC | GTA | AAA | AAC | TTC |
| aaG: | M | Q | K | T | F | R | P | F | Q | L | T | R | G | F | T | S | S | V | K | N | F |

Seq.31-SD | *P. dominulus* – Retículo endoplasmático – Allergen Pol d 5 – GI número 51093376

Código *G*-linearidade ((63,57,3) BCH primitivo sobre $GF(4)$, rotulamento S_4)

$$p_{10}(x) = x^3 + ax^2 + bx + b - g_{10}(x) = x^6 + x^5 + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | K | I | S | C | L | I | C | L | V | I | V | L | T | I | I | H | L | S | Q | A |
| ntO: | ATG | AAA | ATT | AGT | TGC | TTA | ATT | TGT | CTC | GTA | ATT | GTT | CTT | ACG | ATC | ATT | CAT | TTG | TCT | CAA | GCT |
| RtO: | Oba | 000 | 0bb | 0ab | bab | bb0 | bab | 1b1 | ab0 | 0bb | abb | 1bb | 01a | Ob1 | 0bb | 10b | bba | b1b | 100 | alb | |
| RtG: | Oba | 000 | 0bb | 0a 0 | bab | bb0 | bab | 1b1 | ab0 | 0bb | abb | 1bb | 01a | Ob1 | 0bb | 10b | bba | b1b | 100 | alb | |
| ntG: | ATG | AAA | ATT | AG A | TGC | TTA | ATT | TGT | CTC | GTA | ATT | GTT | CTT | ACG | ATC | ATT | CAT | TTG | TCT | CAA | GCT |
| aaG: | M | K | I | R | C | L | I | C | L | V | I | V | L | T | I | I | H | L | S | Q | A |

Figura B.1: Sequências de direcionamento (SD) com 63 nucleotídeos e $D(a, b) = 1$.

Seq.31-SD | *P.dominulus* – Retículo endoplasmático – Allergen Pol d 5 – GI número 51093376

Código G-linearidade ((63,57,3) BCH primitivo sobre GF(4), rotulamento S_4)

$$p_{08}(x) = x^3 + bx^2 + x + a - g_{08}(x) = x^6 + x^5 + x^4 + x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | K | I | S | C | L | I | C | L | V | I | V | L | T | I | I | H | L | S | Q | A |
| ntO: | ATG | AAA | ATT | AGT | TGC | TTA | ATT | TGT | CTC | GTA | ATT | GTT | CTT | ACG | ATC | ATT | CAT | TTG | TCT | CAA | GCT |
| RtO: | 0ba | 000 | 0bb | 0ab | ba1 | bb0 | 0bb | bab | 1b1 | ab0 | 0bb | abb | 1bb | 01a | 0b1 | 0bb | 10b | bba | b1b | 100 | a1b |
| RtG: | 0ba | 000 | 0bb | 0ab | ba1 | bb0 | 0bb | bab | 1b1 | ab0 | 0bb | abb | 1bb | 01a | 0b1 | 0bb | 10b | bba | b1b | b00 | a1b |
| ntG: | ATG | AAA | ATT | AGT | TGC | TTA | ATT | TGT | CTC | GTA | ATT | GTT | CTT | ACG | ATC | ATT | CAT | TTG | TCT | TAA | GCT |
| aaG: | M | K | I | S | C | L | I | C | L | V | I | V | L | T | I | I | H | L | S | sto | A |

Seq.36-SI | *S. cerevisiae* – OXA 1 – Sinal interno – GI número 832917

Código G-linearidade ((63,57,3) BCH primitivo sobre GF(4), rotulamento S_4)

$$p_{01}(x) = x^3 + x^2 + ax + b - g_{01}(x) = x^6 + x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |
| ntO: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| RtO: | a11 | abb | 10b | abb | b01 | b1b | aaa | bb1 | 11b | baa | baa | aa0 | 01b | 0b1 | a1a | a11 | 011 | 011 | 0b1 | 1b1 | 0bb |
| RtG: | a11 | abb | 10b | abb | b11 | b1b | aaa | bb1 | 11b | baa | baa | aa0 | 01b | 0b1 | a1a | a11 | 011 | 011 | 0b1 | 1b1 | 0bb |
| ntG: | GCC | GTT | CAT | GTT | TCC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| aaG: | A | V | H | V | S | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |

$$p_{03}(x) = x^3 + ax^2 + bx + a - g_{03}(x) = x^6 + x^5 + x^2 + x + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|-----|-----|-----|-----|
| aaO: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |
| ntO: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| RtO: | a11 | abb | 10b | abb | b01 | b1b | aaa | bb1 | 11b | baa | baa | aa0 | 01b | 0b1 | a1a | a11 | 011 | 011 | 0b1 | 1b1 | 0bb |
| RtG: | a11 | abb | 10b | abb | b01 | b1b | aaa | bb1 | 11b | baa | baa | aa0 | 01b | 0b1 | a1a | a11 | b11 | 011 | 0b1 | 1b1 | 0bb |
| ntG: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | TCC | ACC | ATC | CTC | ATT |
| aaG: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | S | T | I | L | I |

Seq.38-I | *R. norvegicus* - NADH ubiquinone oxidoreductase subunit (IP13) gene – GI número 600528

Código G-linearidade ((63,57,3) BCH primitivo sobre GF(4), rotulamento S_4)

$$p_{05}(x) = x^3 + ax^2 + bx + a - g_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ntO: | GTA | ACG | AGG | TAA | CGG | CCG | TAA | TGC | CTG | GAA | CCC | GAG | ACT | GAC | GGT | AGC | AGG | GAG | CGT | GGC | AAG |
| RtO: | ab0 | 01a | 0aa | b00 | 1aa | 11a | b00 | ba1 | 1ba | a00 | 111 | a0a | 01b | a01 | aab | 0a1 | 0aa | a0a | 1ab | a11 | 00a |
| RtG: | ab0 | 01a | 0aa | b00 | 1aa | 11a | b00 | ba1 | 1ba | a00 | 111 | a0b | 01b | a01 | aab | 0a1 | 0aa | a0a | 1ab | a11 | 00a |
| aaG: | GTA | ACG | AGG | TAA | CGG | CCG | TAA | TGC | CTG | GAA | CCC | GAT | ACT | GAC | GGT | AGC | AGG | GAG | CGT | GGC | AAG |

Figura B.2: Sequências de Direcionamento (SD), sinal interno (SI) e íntron (I) com 63 nucleotídeos e $D(a, b) = 1$.

Seq.41-DNARep | *O. sativae* - Japonica Group DNA, transposon like sequence – GI número 21734424

Código G-linearidade (255,247,3) BCH primitivo sobre GF(4), rotulamento S_4)

$$p_{13}(x) = x^4 + x^3 + x + a - g_{13}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1$$

```

ntO: GGG CTG GTT TGG TTT GAG ACC TAA ATT AGA CTT ACC AAT ATT TGA TAA TTT AAA TAG TGT TTA
RtO: aaa 1ba abb baa bbb a0a 011 b00 0bb 0a0 1bb 011 00b 0bb ba0 b00 bbb000 b0a bab bb0
RtG: aaa 1ba abb baa a0a 011 b00 0bb 0a0 1bb 011 00b 0bb ba0 b00 bbb 000 b0a bab bb0
ntG: GGG CTG GTT TGG TTT GAG ACC TAA ATT AGA CTT ACC AAT ATT TGA TAA TTT AAA TAG TGT TTA

ntO: GTG TCT ATT TGG GAA GCC AAA TTT TGA CAT GCC TAA GAA AAT ATG TCA TTT CAA TAG TGA
RtO: aba b1b Obb baa bbb a00 a11 000 bbb ba0 10b 11b 000 0bb Oba b10 bbb 100 b0a ba0
RtG: aba b1b Obb baa bbb a00 a11 000 bbb ba0 10b 11b 000 0bb Oba b10 bbb 100 b0a ba0
ntG: GTG TCT ATT TGG GAA GCC AAA TTT TGA CAT GCC TAA GAA AAT ATG TCA TTT CAA TAG TGA

ntO: ACT TAG GCT ATT TTG ACT TCA ATC CAA ACA CAA CTT TAC CTT ACC AAA ATT AGT CAT GCC AAA
RtO: 01b b0a alb Obb bba 01b b10 0b1 100 010 100 1bb b01 1bb 011 000 0bb Oab 10b a11 000
RtG: 01b b0a alb Obb bba 01b b10 0b1 100 010 100 1bb b01 1bb 011 000 0bb Oab 10b a11 000
ntG: ACT TAG GCT ATT TTG ACT TCA ATC CAA ACA CAA CTT TAC CTT ACC AAA ATT AGT CAT GCC AAA

ntO: ACT TGC CAA AAT TTG GCA CTG ACA AAA TTT GGT AAG GCC TAT TTA GGC CAC AAA CCA AAC CAA
RtO: 01b ba1 100 00b bba a10 1ba 010 000 bbb aab 00a a11 b0b bb0 aal 101 000 110 001 100
RtG: 01b ba1 100 00b bba a10 1ba 010 000 bab aab 00a a11 b0b bb0 aal 101 000 110 001 100
ntG: ACT TGC CAA AAT TTG GCA CTG ACA AAA TG TGT GGT AAG GCC TAT TTA GGC CAC AAA CCA AAC CAA

ntO: CCC
RtO: 111
RtG: 111
ntG: CCC

```

Seq.46-P | *A. denitrificans* - EST4002 plasmid pEST4011 - hypothetical protein – GI número 45368559

Código G-linearidade (255,247,3) BCH primitivo sobre GF(4), rotulamento S_4)

$$p_{01}(x) = x^4 + x^3 + ax^2 + ax + a - g_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1$$

```

aaO: M S E A N I R L E C L R P A N D G W E Q P
ntO: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CGT CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG
RtO: Oba b10 a0a a10 00b 0b1 0aa 1bb a00 bal 1ba laa 111 a10 001 a01 a11 baa a0a 10a 11a
RtG: Oba b10 a0a a10 00b 0b1 0aa 1bb a00 bal 1ba laa 111 a10 001 a01 a11 baa a0a 10a 11a
ntG: ATG TCA GAG GCA AAT ATC AGG CTT GAA TGC CGT CGG CCC GCA AAC GAC GGC TGG GAG CAG CCG
aaG: M S E A N I R L E C L R P A N D G W E Q P

aaO: T G E E V R E A L K A A G F T G G Q A A K
ntO: ACC GGC GAA GAG GTG CGC GAG GCG CTG AAA GCG GCG GGC TTC ACG GGA GGC CAA GCC GCG AAA
RtO: 011 a11 a00 a0a aba 1a1 1a0 ala 1ba 000 ala ala aal bb1 01a aa0 a11 100 a11 a1a 000
RtG: 011 a11 a00 a0a aba bal 1a0 ala 1ba 000 ala ala aal bb1 01a aa0 a11 100 a11 a1a 000
ntG: ACC GGC GAA GAG GTG TG GAG GCG CTG AAA GCG GCG GGC TTC ACG GGA GGC CAA GCC GCG AAA
aaG: T G E E V C E A L K A A G F T G G Q A A K

aaO: A L G L G A K G D R T V R R W I G G D S A
ntO: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC
RtO: a11 1b1 aaa 1ba aaa a10 00a a10 00b 0a0 011 a0b 1a1 1a1 baa 0b1 a11 aaa a0b b1a a11
RtG: a11 1b1 aaa 1ba aaa a10 00a a10 00b 0a0 011 a0b 1a1 1a1 baa 0b1 a11 aaa a0b b1a a11
ntG: GCC CTC GGG CTG GGG GCA AAG GGC GAT AGA ACC GTA CGC CGC TGG ATC GGC GGG GAT TCG GCC
aaG: A L G L G A K G D R T V R R W I G G D S A

aaO: I P Y A A W A L L C D F G N L G Q I W K K
ntO: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CAA ATC TGG AAG AAG
RtO: 0b1 111 b0b a11 a10 baa a11 bba 1ba bal a01 bb1 a11 00b bba a11 100 0b1 baa 00a 00a
RtG: 0b1 111 b0b a11 a10 baa a11 bba 1ba bal a01 bb1 a11 00b bba a11 100 0b1 baa 00a 00a
ntG: ATC CCC TAT GCC GCA TGG GCC TTG CTG TGC GAC TTC GGC AAT TTG GGC CAA ATC TGG AAG AAG
aaG: I P Y A A W A L L C D F G N L G Q I W K K

aaO: D
ntO: GAC
RtO: a01
RtG: a01
ntG: GAC
aaG: D

```

Figura B.3: DNA repetitivo (DNARep) e proteína (P) com 255 nucleotídeos e $D(a, b) = 1$.

B.2 BCH Primitivo sobre $GR(4, r)$

Seq.36-SI | *S. aeruginosa* - OXA 1 – Sinal interno – GI número 832917

Código Z_2xZ_2 -linearidade (63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento B)

$$p_{02}(x) = x^6 + x^4 + x^3 + x + 1 \quad - \quad g_{02}(x) = x^6 + 2x^5 + x^4 + x^3 + 3x + 1 \quad - \quad \text{Caso } 01 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|
| aaO: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |
| ntO: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| RtO: | 211 | 233 | 103 | 233 | 301 | 313 | 222 | 332 | 113 | 322 | 322 | 220 | 013 | 031 | 212 | 211 | 011 | 011 | 031 | 131 | 033 |
| RtG: | 211 | 233 | 103 | 233 | 301 | 313 | 222 | 332 | 113 | 322 | 322 | 220 | 013 | 031 | 212 | 311 | 011 | 011 | 031 | 131 | 033 |
| ntG: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | TCC | ACC | ACC | ATC | CTC | ATT |
| aaG: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | S | T | T | I | L | I |

Código Z_4 -linearidade (63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento A)

$$p_{04}(x) = x^6 + x^5 + x^2 + x + 1 \quad - \quad g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1 \quad - \quad \text{Caso } 02 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 1, 3, 2)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | V | H | V | Y | S | G | L | P | W | W | G | T | I | A | A | T | T | I | L | I |
| ntO: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GGA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| RtO: | 311 | 322 | 102 | 322 | 201 | 212 | 333 | 223 | 112 | 233 | 233 | 330 | 012 | 021 | 313 | 311 | 011 | 011 | 021 | 121 | 022 |
| RtG: | 311 | 322 | 102 | 322 | 201 | 212 | 333 | 223 | 112 | 233 | 233 | 310 | 012 | 021 | 313 | 311 | 011 | 011 | 021 | 121 | 022 |
| ntG: | GCC | GTT | CAT | GTT | TAC | TCT | GGG | TTG | CCT | TGG | TGG | GCA | ACT | ATC | GCG | GCC | ACC | ACC | ATC | CTC | ATT |
| aaG: | A | V | H | V | Y | S | G | L | P | W | W | A | T | I | A | A | T | T | I | L | I |

Seq.38-I | *R. norvegicus* - NADH ubiquinone oxidoreductase subunit (IP13) gene – GI número 600528

Código Z_2xZ_2 -linearidade (63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento B)

$$p_{04}(x) = x^6 + x^5 + x^2 + x + 1 \quad - \quad g_{04}(x) = x^6 + 3x^5 + 2x^4 + x^2 + x + 1 \quad - \quad \text{Caso } 01 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| ntO: | CAT | TGC | TCC | ATT | GCC | GGC | ATT | ACG | GAC | CTT | GGG | CTC | TGA | CTG | CCA | TCG | TCC | CTC | GCA | CCG | TTC |
| RtO: | 103 | 321 | 311 | 033 | 211 | 221 | 033 | 012 | 201 | 133 | 222 | 131 | 320 | 132 | 110 | 312 | 311 | 131 | 210 | 112 | 331 |
| RtG: | 103 | 321 | 211 | 033 | 211 | 221 | 033 | 012 | 201 | 133 | 222 | 131 | 320 | 132 | 110 | 312 | 311 | 131 | 210 | 112 | 331 |
| ntG: | CAT | TGC | GCC | ATT | GCC | GGC | ATT | ACG | GAC | CTT | GGG | CTC | TGA | CTG | CCA | TCG | TCC | CTC | GCA | CCG | TTC |

Código Klein-linearidade (63,57,3) BCH primitivo sobre $GR(4,6)$, rotulamento C)

$$p_{05}(x) = x^6 + x^5 + x^3 + x^2 + 1 \quad - \quad g_{05}(x) = x^6 + 3x^5 + x^3 + x^2 + 2x + 1 \quad - \quad \text{Caso } 03 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| ntO: | GTA | ACG | AGG | TAA | CGG | CCG | TAA | TGC | CTG | GAA | CCC | GAG | ACT | GAC | GGT | AGC | AGG | GAG | CGT | GGC | AAG |
| RtO: | 130 | 021 | 011 | 300 | 211 | 221 | 300 | 312 | 231 | 100 | 222 | 101 | 023 | 102 | 113 | 012 | 011 | 101 | 213 | 112 | 001 |
| RtG: | 130 | 021 | 011 | 300 | 211 | 221 | 300 | 312 | 231 | 100 | 222 | 101 | 023 | 102 | 113 | 012 | 011 | 101 | 213 | 112 | 201 |
| ntG: | GTA | ACG | AGG | TAA | CGG | CCG | TAA | TGC | CTG | GAA | CCC | GAG | ACT | GAC | GGT | AGC | AGG | GAG | CGT | GGC | CAG |

Figura B.4: Sinal interno (SI) e Ítron (I) com 63 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.41-DNARep | O. sativae - Japonica Group DNA, transposon like sequence – GI número 21734424

Código Klein-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento C)

$$P_{09}(x) = x^8 + x^6 + x^4 + x^3 + x^2 + x + 1 - g_{09}(x) = x^8 + 2x^7 + 3x^6 + x^4 + 3x^3 + x^2 + x + 1 - \text{Caso 03- (A, C, G, T)} = (0, 2, 1, 3)$$

```

ntO: GGG CTG GTT TGG TTT GAG ACC TAA ATT AGA CTT ACC AAT ATT TGA TAA TTT YAA TAG TGT TTA
RtO: 111 231 133 311 333 101 022 300 033 010 233 022 003 033 310 300 333 000 301 313 330
RtG: 111 231 133 311 333 101 022 300 033 010 233 022 003 033 310 300 333 000 301 313 330
ntG: GGG CTG GTT TGG TTT GAG ACC TAA ATT AGA CTT ACC AAT ATT TGA TAA TTT AAA TAG TGT TTA

ntO: GTG TCT ATT TGG TTT GAA GCC AAA TTT TGA CAT GCC TAA GAA AAT ATG TCA TTT CAA TAG TGA
RtO: 131 323 033 311 333 100 122 000 333 310 203 122 300 100 003 031 320 333 200 301 310
RtG: 131 323 033 311 333 100 122 000 333 310 203 122 300 100 003 031 320 333 200 301 310
ntG: GTG TCT ATT TGG TTT GAA GCC AAA TTT TGA CAT GCC TAA GAA AAT ATG TCA TTT CAA TAG TGA

ntO: ACT TAG GCT ATT TTG ACT TCA ATC CAA ACA CAA CTT TAC CTT ACC AAA ATT AGT CAT GCC AAA
RtO: 023 301 123 033 331 023 320 032 200 020 200 233 302 233 022 000 033 013 203 122 000
RtG: 023 301 123 033 331 023 320 032 200 220 200 233 302 233 022 000 033 013 203 122 000
ntG: ACT TAG GCT ATT TTG ACT TCA ATC CAA CCA CTT TAC CTT ACC AAA ATT AGT CAT GCC AAA

ntO: ACT TGC CAA AAT TTG GCA CTG ACA AAA TTT GGT AAG GCC TAT TTA GGC CAC AAA CCA AAC CAA
RtO: 023 312 200 003 331 120 231 020 000 333 113 001 122 303 330 112 202 000 220 002 200
RtG: 023 312 200 003 331 120 231 020 000 333 113 001 122 303 330 112 202 000 220 002 200
ntG: ACT TGC CAA AAT TTG GCA CTG ACA AAA TTT GGT AAG GCC TAT TTA GGC CAC AAA CCA AAC CAA

ntO: CCC
RtO: 222
RtG: 222
ntG: CCC

```

Seq.42-P | A. thaliana - proteína – GI número 26556996

Código Z₂xZ₂-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento B)

$$P_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1 - g_{02}(x) = x^8 + 2x^7 + x^6 + 3x^5 + x^3 + 1 - \text{Caso 01- (A, C, G, T)} = (0, 1, 2, 3)$$

```

aaO: M T K R E Y N S Q P E M L E G A K S I G A
ntO: ATG ACA AAG CGT GAG TAT AAT TCT CAA CCC GAG ATG TTA GAA GGT GCA AAA TCA ATA GGT GCC
RtO: 032 010 002 123 202 303 003 313 100 111 202 032 330 200 223 210 000 310 030 223 211
RtG: 032 010 002 123 202 303 003 313 100 111 202 032 330 200 223 210 000 310 030 223 211
ntG: ATG ACA AAG CGT GAG TAT AAT TCT CAA CCC GAG ATG TTA GAA GGT GCA AAA TCA ATA GGT GCC
aaG: M T K R E Y N S Q P E M L E G A K S I G A

aaO: G A A T I A S A G A A I G I G N V F S S L
ntO: GGA GCT GCT ACA ATT GCT TCA GCG GGA GCT GCT ATC GGT ATT GGA AAC GTA TTC AGT TCT TTG
RtO: 220 213 213 010 033 213 310 212 220 213 213 031 223 033 220 001 230 331 023 313 332
RtG: 220 213 213 010 033 213 310 212 220 213 213 031 223 033 220 001 230 331 023 313 332
ntG: GGA GCT GCT ACA ATT GCT TCA GCG GGA GCT GCT ATC GGT ATT GGA AAC GTA TTC AGT TCT TTG
aaG: G A A T I A S A G A A I G I G N V F S S L

aaO: I H S V A R N P S L A K Q S F G Y A I L G
ntO: ATT CAT TCT GTG GCG CGA AAT CCA TCA TTG GCT AAA CAA TCA TTT GGT TAT GCC ATT TTG GGC
RtO: 033 103 313 232 212 120 003 110 310 332 213 000 100 310 333 223 303 211 033 332 221
RtG: 033 103 313 232 212 120 003 110 310 332 213 000 100 310 333 223 303 211 033 332 221
ntG: ATT CAT TCT GTG GCG CGA AAT CCA TCA TTG GCT AAA CAA TCA TTT GGT TAT GCC ATT TTG GGC
aaG: I H S V A R N P S L A K Q S F G Y A I L G

aaO: F A L T E A I A L F A P M M A F L I L F V
ntO: TTT GCT CTA ACC GAA GCT ATT GCA TTG TTT GCC CCA ATG ATG GCC TTT TTG ATC TTA TTC GTC
RtO: 333 213 130 011 200 213 033 210 332 333 211 110 032 032 211 333 332 031 330 331 230
RtG: 333 203 130 011 200 213 033 210 332 333 211 110 032 032 211 333 332 031 330 331 230
ntG: TTT GAT CTA ACC GAA GCT ATT GCA TTG TTT GCC CCA ATG ATG GCC TTT TTG ATC TTA TTC GTC
aaG: F D L T E A I A L F A P M M A F L I L F V

aaO: F
ntO: TTC
RtO: 331
RtG: 331
ntG: TTC
aaG: F

```

Figura B.5: DNA repetitivo (DNARep) e Proteína (P) com 255 nucleotídeos e $D(a, b) = 1$.

Seq.42-P | *A. thaliana* - proteína – GI número 26556996

Código Klein-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento C)

$$P_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1 - g_{02}(x) = x^8 + 2x^7 + x^6 + 3x^5 + x^3 + 1 - \text{Caso 03-}(\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 2, 1, 3)$$

```

aaO: M T K R E Y N S Q P E M L E G A K S I G A
ntO: ATG ACA AAG CGT GAG TAT AAT TCT CAA CCC GAG ATG TTA GAA GGT GCA AAA TCA ATA GGT GCC
RtO: 031 020 001 213 101 303 003 323 200 222 101 031 330 100 113 120 000 320 030 113 122
RtG: 031 020 001 213 101 303 003 323 200 222 101 031 330 100 113 120 000 320 030 113 122
ntG: ATG ACA AAG CGT GAG TAT AAT TCT CAA CCC GAG ATG TTA GAA GGT GCA AAA TCA ATA GGT GCC
aaO: M T K R E Y N S Q P E M L E G A K S I G A

aaO: G A A T I A S A G A A I G I G N V F S S L
ntO: GGA GCT GCT ACA ATT GCT TCA GCG GGA GCT GCT ATC GGT ATT GGA AAC GTA TTC AGT TCT TTG
RtO: 110 123 123 020 033 123 320 121 110 123 123 032 113 033 110 002 130 332 013 323 331
RtG: 110 123 123 020 033 123 320 121 110 123 123 032 113 033 110 002 130 332 013 323 331
ntG: GGA GCT GCT ACA ATT GCT TCA GCG GGA GCT GCT ATC GGT ATT GGA AAC GTA TTC AGT TCT TTG
aaO: G A A T I A S A G A A I G I G N V F S S L

aaO: I H S V A R N P S L A K Q S F G Y A I L G
ntO: ATT CAT TCT GTG GCG CGA AAT CCA TCA TTG GCT AAA CAA TCA TTT GGT TAT GCC ATT TTG GGC
RtO: 033 203 323 131 121 210 003 220 320 331 123 000 200 320 333 113 303 122 033 331 112
RtG: 033 203 323 131 121 210 003 220 320 331 123 000 200 320 333 113 303 122 033 331 113
ntG: ATT CAT TCT GTG GCG CGA AAT CCA TCA TTG GCT AAA CAA TCA TTT GGT TAT GCC ATT TTG GGT
aaO: I H S V A R N P S L A K Q S F G Y A I L G

aaO: F A L T E A I A L F A P M M A F L I L F V F
ntO: TTT GCT CTA ACC GAA GCT ATT GCA TTG TTT GCC CCA ATG ATG GCC TTT TTG ATC TTA TTC GTC TTC
RtO: 333 123 230 022 100 123 033 120 331 333 122 220 031 031 122 333 331 032 330 332 130 332
RtG: 333 123 230 022 100 123 033 120 331 333 122 220 031 031 122 333 331 032 330 332 130 332
ntG: TTT GCT CTA ACC GAA GCT ATT GCA TTG TTT GCC CCA ATG ATG GCC TTT TTG ATC TTA TTC GTC TTC
aaO: F A L T E A I A L F A P M M A F L I L F V F

```

Seq.44-P | *A. marina* - MBIC11017 plasmid pREB2, complete sequence – Bacteria – GI número 158339871

Código Klein-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento C)

$$P_{12}(x) = x^8 + x^7 + x^4 + x^2 + x + 1 - g_{12}(x) = x^8 + 3x^7 + 2x^5 + x^4 + x^2 + x + 1 - \text{Caso 03-}(\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 2, 1, 3)$$

```

aaO: M M T T L L S F L A D L S N H L G L A W W
ntO: ATG ATG ACT ACT CTA TTA AGT TTC CTG GCT GAT CTA TCG AAT CAC CTG GGT CTA GCC TGG TGG
RtO: 031 031 023 230 330 013 332 231 123 103 230 321 003 202 231 113 230 122 311 311
RtG: 031 031 023 230 330 013 332 231 123 103 230 321 003 202 231 113 230 122 311 311
ntG: ATG ATG ACT ACT CTA TTA AGT TTC CTG GCT GAT CTA TCG AAT CAC CTG GGT CTA GCC TGG TGG
aaO: M M T T L L S F L A D L S N H L G L A W W

Oaa: V E I K T L S P I C T Y F F G P F L I R K
ntO: GTT GAA ATC AAA ACG TTA TCC CCG ATA TGT ACT TAT TTT TTT GGT CCC TTT CTG ATT CGC AAG
RtO: 133 100 032 000 021 330 322 221 030 313 023 303 333 333 113 222 333 231 033 212 001
RtG: 133 100 032 000 021 330 322 221 030 313 023 303 333 333 113 222 333 231 033 212 001
ntG: GTT GAA ATC AAA ACG TTA TCC CCG ATA TGT ACT TAT TTT TTT GGT CCC TTT CTG ATT CGC AAG
aaO: V E I K T L S P I C T Y F F G P F L I R K

aaO: E A E A A L F G Y V E D L E A E Q A Q N I
ntO: GAA GCA GAA GCA GCA TTG TTT GGG TAT GTA GAA GAT TTA GAG GCA GAA CAG GCA CAG AAC ATA
RtO: 100 120 100 120 120 331 333 111 303 130 100 103 330 101 120 100 201 120 201 002 030
RtG: 100 120 100 120 120 331 333 111 303 130 100 103 330 101 120 100 201 120 201 002 030
ntG: GAA GCA GAA GCA GA TTG TTT GGG TAT GTA GAA GAT TTA GAG GCA GAA CAG GCA CAG AAC ATA
aaO: E A E A G L F G Y V E D L E A E Q A Q N I

aaO: V A H I Q R C H P P H L T I C N E M E Y S C
ntO: GTT GCC CAT ATT CAG CGT TGT CAT CCA CCC CAT TTG ACC ATT TGC AAT GAG ATG GAG TAT TCG TGC
RtO: 133 122 203 033 201 213 313 203 220 222 203 331 022 033 312 003 101 031 101 303 321 312
RtG: 133 122 203 033 201 213 313 203 220 222 203 331 022 033 312 003 101 031 101 303 321 312
ntG: GTT GCC CAT ATT CAG CGT TGT CAT CCA CCC CAT TTG ACC ATT TGC AAT GAG ATG GAG TAT TCG TGC
aaO: V A H I Q R C H P P H L T I C N E M E Y S C

```

Figura B.6: Proteína codificada por 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.46-P | *A. denitrificans* - EST4002 plasmid pEST4011, complete – Bacteria — GI número 45368559

Código Z₄-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento A)

$$p_{11}(x) = x^8 + x^7 + x^5 + x^3 + 1 - g_{11}(x) = x^8 + 3x^7 + 2x^6 + x^5 + 3x^3 + 1 - \text{ Caso } 02 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 1, 3, 2)$$

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | |
| ntO: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCA | AAC | GAC | GGC | TGG | GAG | CAG | CCG | |
| RtO: | 023 | 210 | 303 | 310 | 002 | 021 | 033 | 122 | 300 | 231 | 123 | 133 | 111 | 310 | 001 | 301 | 331 | 233 | 303 | 103 | 113 | |
| RtG: | 023 | 210 | 303 | 310 | 002 | 021 | 033 | 122 | 300 | 231 | 123 | 133 | 111 | 313 | 001 | 301 | 331 | 233 | 303 | 103 | 113 | |
| ntG: | ATG | TCA | GAG | GCA | AAT | ATC | AGG | CTT | GAA | TGC | CTG | CGG | CCC | GCG | AAC | GAC | GGC | TGG | GAG | CAG | CCG | |
| aaO: | M | S | E | A | N | I | R | L | E | C | L | R | P | A | N | D | G | W | E | Q | P | |
| aaO: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | |
| ntO: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | CTG | AAA | GCG | GCG | GCG | TTC | ACG | GGA | GGC | CAA | GCC | GCG | AAA | |
| RtO: | 011 | 331 | 300 | 303 | 323 | 131 | 303 | 313 | 123 | 000 | 313 | 313 | 331 | 221 | 013 | 330 | 331 | 100 | 311 | 313 | 000 | |
| RtG: | 011 | 331 | 300 | 303 | 323 | 131 | 303 | 313 | 123 | 000 | 313 | 313 | 331 | 221 | 013 | 330 | 331 | 100 | 311 | 313 | 000 | |
| ntG: | ACC | GGC | GAA | GAG | GTG | CGC | GAG | GCG | CTG | AAA | GCG | GCG | GCG | TTC | ACG | GGA | GGC | CAA | GCC | GCG | AAA | |
| aaO: | T | G | E | E | V | R | E | A | L | K | A | A | G | F | T | G | G | Q | A | A | K | |
| aaO: | A | L | G | L | G | A | K | G | D | R | T | V | R | R | W | I | G | G | D | S | A | |
| ntO: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GGC | GAT | AGA | ACC | GTA | CGC | CGC | TGG | ATC | GGA | GGG | GAT | TCG | GCC | |
| RtO: | 311 | 121 | 333 | 123 | 333 | 310 | 003 | 331 | 302 | 030 | 011 | 320 | 131 | 131 | 233 | 021 | 331 | 333 | 302 | 213 | 311 | |
| RtG: | 311 | 121 | 333 | 123 | 333 | 310 | 003 | 331 | 302 | 030 | 011 | 320 | 131 | 131 | 233 | 021 | 331 | 333 | 302 | 213 | 311 | |
| ntG: | GCC | CTC | GGG | CTG | GGG | GCA | AAG | GGC | GAT | AGA | ACC | GTA | CGC | CGC | TGG | ATC | GGA | GGG | GAT | TCG | GCC | |
| aaO: | A | L | G | L | G | A | K | G | D | R | T | V | R | R | W | I | G | G | D | S | A | |
| aaO: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | D |
| ntO: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GCG | AAT | TAT | GGC | CAA | ATC | TGG | AAG | AAG | GAC |
| RtO: | 021 | 111 | 202 | 311 | 310 | 233 | 311 | 223 | 123 | 231 | 301 | 221 | 331 | 002 | 223 | 331 | 100 | 021 | 233 | 003 | 003 | 301 |
| RtG: | 021 | 111 | 202 | 311 | 310 | 233 | 311 | 223 | 123 | 231 | 301 | 221 | 331 | 002 | 223 | 331 | 100 | 021 | 233 | 003 | 003 | 301 |
| ntG: | ATC | CCC | TAT | GCC | GCA | TGG | GCC | TTG | CTG | TGC | GAC | TTC | GCG | AAT | TAT | GGC | CAA | ATC | TGG | AAG | AAG | GAC |
| aaO: | I | P | Y | A | A | W | A | L | L | C | D | F | G | N | L | G | Q | I | W | K | K | D |

Seq.47-P | *A. hospitalis* - plasmid pAH1, complete sequence – Archaea – GI número 207258119

Código Z₄-linearidade (255,247,3) BCH primitivo sobre GR(4,8), rotulamento A)

$$p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1 - g_{12}(x) = x^8 + 2x^7 + x^6 + x^5 + 2x^3 + x^2 + 2x + 1 - \text{ Caso } 02 - (\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}) = (0, 1, 3, 2)$$

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | S | L | V | Q | L | V | E | K | V | A | K | K | Y | N | N | I | K | V | N | S | L |
| ntO: | ATG | AGC | TTA | GTT | CAA | CTC | GTT | GAA | AAA | GTT | GCG | AAG | AAG | TAT | ATT | ATC | AAA | GTA | AAT | TCT | CTC | |
| RtO: | 023 | 031 | 220 | 322 | 100 | 121 | 322 | 300 | 000 | 322 | 313 | 003 | 003 | 202 | 002 | 021 | 000 | 320 | 002 | 212 | 121 | |
| RtG: | 023 | 031 | 220 | 322 | 100 | 121 | 322 | 300 | 000 | 322 | 313 | 003 | 003 | 202 | 002 | 021 | 000 | 320 | 002 | 212 | 121 | |
| ntG: | ATG | AGC | TTA | GTT | CAA | CTC | GTT | GAA | AAA | GTT | GCG | AAG | AAG | TAT | ATT | ATC | AAA | GTA | AAT | TCT | CTC | |
| aaO: | M | S | L | V | Q | L | V | E | K | V | A | K | K | Y | N | N | I | K | V | N | S | L |
| aaO: | P | N | G | V | I | I | L | V | K | N | D | I | G | Y | V | Q | I | A | A | V | R | |
| ntO: | CCT | AAT | GGT | GTG | ATA | ATT | CTT | GTA | AAA | AAT | GAC | ATA | GGC | TAT | GTG | CAA | ATA | GCT | GCA | GTT | AGA | |
| RtO: | 112 | 002 | 332 | 323 | 020 | 022 | 122 | 320 | 000 | 002 | 301 | 020 | 331 | 202 | 323 | 100 | 020 | 312 | 310 | 322 | 030 | |
| RtG: | 112 | 002 | 332 | 323 | 020 | 022 | 122 | 320 | 000 | 002 | 301 | 020 | 331 | 202 | 323 | 100 | 020 | 312 | 310 | 322 | 030 | |
| ntG: | CCT | AAT | GGT | GTG | ATA | ATT | CTT | GTA | AAA | AAT | GAC | ATA | GGC | TAT | GTG | CAA | ATA | GCT | GCA | GTT | AGA | |
| aaO: | P | N | G | V | I | I | L | V | K | N | D | I | G | Y | V | Q | I | A | A | V | R | |
| aaO: | N | V | Y | Y | V | R | Y | L | T | K | N | E | A | Y | I | I | H | K | L | N | E | |
| ntO: | AAT | GTT | TAC | TAT | GTC | AGA | TAC | TTA | ACG | AAA | AAT | GAG | GCA | TAT | ATT | ATA | CAT | AAA | CCT | AAC | GAA | |
| RtO: | 002 | 322 | 201 | 202 | 321 | 030 | 201 | 220 | 013 | 000 | 002 | 303 | 310 | 202 | 022 | 020 | 102 | 000 | 122 | 001 | 300 | |
| RtG: | 002 | 322 | 201 | 202 | 321 | 030 | 201 | 220 | 013 | 000 | 002 | 303 | 310 | 202 | 022 | 020 | 102 | 000 | 122 | 001 | 300 | |
| ntG: | AAT | GTT | TAC | TAT | GTC | AGA | TAC | TTA | ACG | AAA | AAT | GAG | GCA | TAT | ATT | ATA | CAT | AAA | CTT | AAC | GAA | |
| aaO: | N | V | Y | Y | V | R | Y | L | T | K | N | E | A | Y | I | I | H | K | L | N | E | |
| aaO: | E | V | I | E | W | I | L | E | E | K | L | D | E | T | K | A | L | K | I | P | D | V |
| ntO: | GAG | GTC | ATA | GAA | TGG | ATT | TTA | GAA | GAA | AAG | TTA | GAT | GAA | ACT | AAA | GCC | CTC | AAA | ATC | CCT | GAT | GTT |
| RtO: | 303 | 321 | 020 | 300 | 233 | 022 | 220 | 300 | 300 | 003 | 220 | 302 | 300 | 012 | 000 | 311 | 121 | 000 | 021 | 112 | 302 | 322 |
| RtG: | 303 | 321 | 020 | 300 | 233 | 022 | 220 | 300 | 300 | 003 | 220 | 302 | 300 | 012 | 000 | 311 | 121 | 000 | 021 | 112 | 302 | 322 |
| ntG: | GAG | GTC | ATA | GAA | TGG | CTT | TTA | GAA | GAA | AAG | TTA | GAT | GAA | ACT | AAA | GCC | CTC | AAA | ATC | CCT | GAT | GTT |
| aaO: | E | V | I | E | W | L | E | E | K | L | D | E | T | K | A | L | K | I | P | D | V | |

Figura B.7: Proteína codificada por 255 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.53-P | *B. subtilis* - subsp. *subtilis* str. 168 – Bacteria - GI número: 50812173

Código Z₂Z₂-linearidade (1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento B)

$$p_{42}(x) = x^{10} + x^9 + x^8 + x^7 + x^6 + x^4 + x^3 + x + 1 - g_{42}(x) = x^{10} + x^9 + x^8 + 3x^7 + x^6 + x^4 + x^3 + 3x + 1 - \text{Caso 01- } (A, C, G, T) = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Oaa: | M | I | I | S | Y | K | C | P | N | C | G | S | D | M | A | F | D | S | E | T | G |
| Ont: | ATG | ATA | ATA | TCT | TAT | AAG | TGT | CCG | AAC | TGC | GGC | AGT | GAT | ATG | GCA | TTT | GAC | AGT | GAA | ACC | GGC |
| Olb: | 032 | 030 | 030 | 313 | 303 | 002 | 323 | 112 | 001 | 321 | 221 | 023 | 203 | 032 | 210 | 333 | 201 | 023 | 200 | 011 | 221 |
| Glb: | 032 | 030 | 030 | 313 | 303 | 002 | 323 | 112 | 001 | 321 | 221 | 023 | 203 | 032 | 210 | 333 | 201 | 023 | 200 | 011 | 221 |
| Gnt: | ATG | ATA | ATA | TCT | TAT | AAG | TGT | CCG | AAC | TGC | GGC | AGT | GAT | ATG | GCA | TTT | GAC | AGT | GAA | ACC | GGC |
| Gaa: | M | I | I | S | Y | K | C | P | N | C | G | S | D | M | A | F | D | S | E | T | G |
| Oaa: | S | L | S | C | S | S | C | G | R | Q | D | N | I | E | S | L | P | K | E | N | I |
| Ont: | TCG | TTA | TCC | TGC | AGC | AGC | TGC | GGG | AGA | CAG | GAC | AAT | ATT | GAA | AGC | CTT | CCG | AAA | GAA | AAC | ATT |
| Olb: | 312 | 330 | 311 | 321 | 021 | 021 | 321 | 220 | 020 | 102 | 201 | 003 | 033 | 200 | 021 | 133 | 112 | 000 | 200 | 001 | 033 |
| Glb: | 312 | 330 | 311 | 321 | 021 | 021 | 321 | 220 | 020 | 102 | 201 | 003 | 033 | 200 | 021 | 133 | 112 | 000 | 200 | 001 | 033 |
| Gnt: | TCG | TTA | TCC | TGC | AGC | AGC | TGC | GGG | AGA | CAG | GAC | AAT | ATT | GAA | AGC | CTT | CCG | AAA | GAA | AAC | ATT |
| Gaa: | S | L | S | C | S | S | C | G | R | Q | D | N | I | E | S | L | P | K | E | N | I |
| Oaa: | A | A | R | F | S | D | D | E | A | K | E | Y | Q | C | K | N | C | G | A | V | L |
| Ont: | GCG | GCA | CGG | TTT | TCT | GAT | GAT | GAA | GCA | AAA | GAA | TAT | CAA | TGT | AAA | AAC | TGC | GGT | GCC | GTT | TTG |
| Olb: | 212 | 210 | 122 | 333 | 313 | 203 | 203 | 200 | 210 | 000 | 200 | 303 | 100 | 323 | 000 | 001 | 321 | 223 | 211 | 233 | 332 |
| Glb: | 212 | 210 | 122 | 333 | 313 | 203 | 203 | 200 | 210 | 000 | 200 | 303 | 100 | 323 | 000 | 001 | 321 | 223 | 211 | 233 | 332 |
| Gnt: | GCG | GCA | CGG | TTT | TCT | GAT | GAT | GAA | GCA | AAA | GAA | TAT | CAA | TGT | AAA | AAC | TGC | GGT | GCC | GTT | TTG |
| Gaa: | A | A | R | F | S | D | D | E | A | K | E | Y | Q | C | K | N | C | G | A | V | L |
| Oaa: | I | T | E | A | E | T | T | A | T | T | C | S | F | C | G | G | A | A | I | L | A |
| Ont: | ATC | ACG | GAA | GCT | GAA | ACG | ACA | GCA | ACG | ACG | TGC | AGC | TTC | TGC | GGG | GGT | GCT | GCA | ATA | CTT | GCC |
| Olb: | 031 | 012 | 200 | 213 | 200 | 012 | 010 | 210 | 012 | 012 | 321 | 021 | 331 | 321 | 220 | 223 | 213 | 210 | 030 | 133 | 211 |
| Glb: | 031 | 012 | 200 | 213 | 200 | 012 | 010 | 210 | 012 | 012 | 321 | 021 | 331 | 321 | 220 | 223 | 213 | 210 | 030 | 133 | 211 |
| Gnt: | ATC | ACG | GAA | GCT | GAA | ACG | ACA | GCA | ACG | ACG | TGC | AGC | TTC | TGC | GGG | GGT | GCT | GCA | ATA | CTT | GCC |
| Gaa: | I | T | E | A | E | T | T | A | T | T | C | S | F | C | G | G | A | A | I | L | A |
| Oaa: | D | R | L | S | G | H | L | A | P | A | K | V | I | P | F | T | I | S | K | Q | E |
| Ont: | GAT | CGT | TTA | TCA | GGG | CAT | TTG | GCG | CCG | GCG | AAG | GTC | ATT | CCA | TTT | ACA | ATC | AGC | AAG | CAA | GAA |
| Olb: | 203 | 123 | 330 | 310 | 220 | 103 | 332 | 212 | 112 | 212 | 002 | 231 | 033 | 110 | 333 | 010 | 031 | 021 | 002 | 100 | 200 |
| Glb: | 203 | 123 | 330 | 310 | 220 | 103 | 332 | 212 | 112 | 212 | 002 | 231 | 033 | 110 | 333 | 010 | 031 | 021 | 002 | 100 | 200 |
| Gnt: | GAT | CGT | TTA | TCA | GGG | CAT | TTG | GCG | CCG | GCG | AAG | GTC | ATT | CCA | TTT | ACA | ATC | AGC | AAG | CAA | GAA |
| Gaa: | D | R | L | S | G | H | L | A | P | A | K | V | I | P | F | T | I | S | K | Q | E |
| Oaa: | A | E | Q | A | F | R | K | W | C | K | K | G | L | L | T | P | R | G | F | M | S |
| Ont: | GCG | GAG | CAG | GCA | TTT | CGA | AAG | TGG | TGC | AAA | AAA | GGC | CTT | CTG | ACA | CCA | AGA | GTT | TTC | ATG | TCT |
| Olb: | 212 | 202 | 102 | 210 | 333 | 120 | 002 | 322 | 321 | 000 | 000 | 221 | 133 | 132 | 010 | 110 | 020 | 223 | 331 | 032 | 313 |
| Glb: | 212 | 202 | 102 | 210 | 333 | 120 | 002 | 322 | 321 | 000 | 000 | 221 | 133 | 132 | 010 | 110 | 020 | 223 | 331 | 032 | 313 |
| Gnt: | GCG | GAG | CAG | GCA | TTT | CGA | AAG | TGG | TGC | AAA | AAA | GGC | CTT | CTG | ACA | CCA | AGA | GTT | TTC | ATG | TCT |
| Gaa: | A | E | Q | A | F | R | K | W | C | K | K | G | L | L | T | P | R | G | F | M | S |
| Oaa: | A | D | R | I | K | S | I | T | G | M | Y | I | P | F | W | M | F | D | L | N | S |
| Ont: | GCT | GAT | CGT | ATC | AAA | AGC | ATC | ACC | GGC | ATG | TAT | ATT | CCA | TTT | TGG | ATG | TTT | GAT | TTA | AAT | AGT |
| Olb: | 213 | 203 | 123 | 031 | 000 | 021 | 031 | 011 | 221 | 032 | 303 | 033 | 110 | 333 | 322 | 032 | 333 | 203 | 330 | 003 | 023 |
| Glb: | 213 | 203 | 123 | 031 | 000 | 021 | 031 | 011 | 221 | 032 | 303 | 033 | 110 | 333 | 322 | 032 | 333 | 203 | 330 | 003 | 023 |
| Gnt: | GCT | GAT | CGT | ATC | AAA | AGC | ATC | ACC | GGC | ATG | TAT | ATT | CCA | TTT | TGG | ATG | TTT | GAT | TTA | AAT | AGT |
| Gaa: | A | D | R | I | K | S | I | T | G | M | Y | I | P | F | W | M | F | D | L | N | S |
| Oaa: | E | V | Q | V | R | A | N | C | T | R | V | H | Q | Y | E | E | G | D | Y | I | C |
| Ont: | GAA | GTA | CAG | GTC | AGA | GCA | AAC | TGT | ACC | AGA | GTC | CAT | CAA | TAT | GAA | GAA | GGG | GAT | TAT | ATT | TGC |
| Olb: | 200 | 230 | 102 | 232 | 020 | 210 | 001 | 323 | 011 | 020 | 231 | 103 | 100 | 303 | 200 | 200 | 222 | 203 | 303 | 033 | 321 |
| Glb: | 200 | 230 | 102 | 232 | 020 | 210 | 001 | 323 | 011 | 020 | 231 | 103 | 100 | 303 | 200 | 200 | 222 | 203 | 303 | 033 | 321 |
| Gnt: | GAA | GTA | CAG | GTC | AGA | GCA | AAC | TGT | ACC | AGA | GTC | CAT | CAA | TAT | GAA | GAA | GGG | GAT | TAT | ATT | TGC |
| Gaa: | E | V | Q | V | R | A | N | C | T | R | V | H | Q | Y | E | E | G | D | Y | I | C |
| Oaa: | T | E | T | E | H | F | E | A | F | R | D | I | N | L | D | Y | L | K | I | P | V |
| Ont: | ACG | GAA | ACA | GAG | CAC | TTT | GAA | GCG | TTT | CGT | GAT | ATC | AAT | CTC | GAT | TAT | TTG | AAA | ATC | CCT | GTC |
| Olb: | 012 | 200 | 010 | 202 | 101 | 333 | 200 | 212 | 333 | 123 | 203 | 031 | 003 | 131 | 203 | 303 | 332 | 000 | 031 | 113 | 231 |
| Glb: | 012 | 200 | 010 | 202 | 101 | 333 | 200 | 212 | 333 | 123 | 203 | 031 | 003 | 131 | 203 | 303 | 332 | 000 | 031 | 113 | 231 |
| Gnt: | ACG | GAA | ACA | GAG | CAC | TTT | GAA | GCG | TTT | CGT | GAT | ATC | AAT | CTC | GAT | TAT | TTG | AAA | ATC | CCT | GTC |
| Gaa: | T | E | T | E | H | F | E | A | F | R | D | I | N | L | D | Y | L | K | I | P | V |

```

Oaa: D A S E K M K D E L M D K L E P Y S Y E E
Ont: GAT GCC TCT GAA AAA ATG AAA GAC GAA TTA ATG GAC AAA TTG GAG CCT TAT TCA TAC GAA GAG
Olb: 203 211 313 200 000 032 000 201 200 330 032 201 000 332 202 113 303 310 301 200 202
Glb: 203 211 313 200 000 032 000 201 203 330 032 201 000 332 202 113 303 310 301 200 202
Gnt: GAT GCC TCT GAA AAA ATG AAA GAC GAT TTA ATG GAC AAA TTG GAG CCT TAT TCA TAC GAA GAG
Gaa: D A S E K M K D D L M D K L E P Y S Y E E

Oaa: L K D F Q T A Y L A G Y I A E K Y N Y T D
Ont: CTG AAG GAC TTT CAA ACG GCA TAT TTG GCC GGT TAT ATT GCG GAA AAG TAC AAT TAT ACC GAT
Olb: 132 002 201 333 100 012 210 303 332 211 223 303 033 212 200 002 301 003 303 011 203
Glb: 132 002 201 333 100 012 210 303 332 211 223 303 033 212 200 002 301 003 303 011 203
Gnt: CTG AAG GAC TTT CAA ACG GCA TAT TTG GCC GGT TAT ATT GCG GAA AAG TAC AAT TAT ACC GAT
Gaa: L K D F Q T A Y L A G Y I A E K Y N Y T D

Oaa: E E L F P R A K E K I S S Y I D S Y I H S
Ont: GAG GAG CTT TTT CCG AGG GCA AAA GAG AAA ATC AGC AGT TAT ATA GAT TCA TAC ATC CAT TCT
Olb: 202 202 133 333 112 022 210 000 202 000 031 021 023 303 030 203 310 301 031 103 313
Glb: 202 202 133 333 112 022 210 000 202 000 031 021 023 303 030 203 310 301 031 103 313
Gnt: GAG GAG CTT TTT CCG AGG GCA AAA GAG AAA ATC AGC AGT TAT ATA GAT TCA TAC ATC CAT TCT
Gaa: E E L F P R A K E K I S S Y I D S Y I H S

Oaa: T F S G Y T S V N V R D K H I H T K N V N
Ont: ACT TTT TCC GGA TAT ACG TCA GTC AAT GTG AGG GAC AAA CAT ATT CAC ACG AAA AAC GTG AAC
Olb: 013 333 311 220 303 012 310 231 003 232 022 201 000 103 033 101 012 000 001 232 001
Glb: 013 333 311 220 303 012 310 231 003 232 022 201 000 103 033 101 012 000 001 232 001
Gnt: ACT TTT TCC GGA TAT ACG TCA GTC AAT GTG AGG GAC AAA CAT ATT CAC ACG AAA AAC GTG AAC
Gaa: T F S G Y T S V N V R D K H I H T K N V N

Oaa: S F Y V L L P V W M V S Y D Y E R A E H I
Ont: AGC TTT TAC GTT TTG CTT CCT GTT TGG ATG GTC AGT TAC GAT TAT GAA AGA GCG GAG CAT ATC
Olb: 021 333 301 233 332 133 113 233 322 032 231 023 301 203 303 200 020 212 202 103 031
Glb: 021 333 301 233 332 133 113 233 322 032 231 023 301 203 303 200 020 212 202 103 031
Gnt: AGC TTT TAC GTT TTG CTT CCT GTT TGG ATG GTC AGT TAC GAT TAT GAA AGA GCG GAG CAT ATC
Gaa: S F Y V L L P V W M V S Y D Y E R A E H I

Oaa: F A M N G Q T G K V V G K P P I S R G K V
Ont: TTT GCG ATG AAC GGG CAA ACA GGA AAG GTC GTT GGA AAG CCG CCG ATC AGT CGA GGA AAA GTG
Olb: 333 212 032 001 222 100 010 220 002 231 233 220 002 112 112 031 023 120 220 000 232
Glb: 333 212 032 001 222 100 010 220 002 231 233 220 002 112 112 031 023 120 220 000 232
Gnt: TTT GCG ATG AAC GGG CAA ACA GGA AAG GTC GTT GGA AAG CCG CCG ATC AGT CGA GGA AAA GTG
Gaa: F A M N G Q T G K V V G K P P I S R G K V

Oaa: A A W F S G I A G G T F L A L K L V S L M
Ont: GCT GCA TGG TTT AGC GGA ATA GCA GGC GGG ACA TTT CTT GCG TTG AAG CTC GTC TCA TTG ATG
Olb: 213 210 322 333 021 220 030 210 221 222 010 333 133 212 332 002 131 231 310 332 032
Glb: 213 210 322 333 021 220 030 210 221 222 010 333 133 212 332 002 131 231 310 332 032
Gnt: GCT GCA TGG TTT AGC GGA ATA GCA GGC GGG ACA TTT CTT GCG TTG AAG CTC GTC TCA TTG ATG
Gaa: A A W F S G I A G G T F L A L K L V S L M

Oaa: M G G G F
Ont: ATG GGA GGC GGA TTT
Olb: 032 220 221 220 333
Glb: 032 220 221 220 333
Gnt: ATG GGA GGC GGA TTT
Gaa: M G G G F

```

Figura B.8: Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.54-P | A. pasteurianus - plasmid pAP12875, Bacteria – GI número: 32455378

Código Z₄-linearidade (1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento A)

$$p_{04}(x) = x^{10} + x^6 + x^5 + x^3 + x^2 + x + 1 - g_{04}(x) = x^{10} + 2x^8 + 3x^6 + x^5 + 3x^3 + 3x^2 + x + 1 - \text{Caso 02- (A,C,G,T)} = (0,1,3,2)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Oaa: | V | R | P | R | V | P | L | S | D | H | S | G | V | R | N | A | L | G | H | V | Q |
| Ont: | GTG | CGG | CCA | AGG | GTT | CCC | CTT | AGC | GAT | CAT | TCA | GGC | GTC | AGA | AAT | GCA | CTC | GGC | CAC | GTC | CAA |
| Olb: | 323 | 133 | 110 | 033 | 322 | 111 | 122 | 031 | 302 | 102 | 210 | 331 | 321 | 030 | 002 | 310 | 121 | 331 | 101 | 321 | 100 |
| Glb: | 323 | 133 | 110 | 033 | 322 | 111 | 122 | 031 | 302 | 102 | 210 | 331 | 321 | 030 | 002 | 310 | 121 | 331 | 101 | 321 | 100 |
| Gnt: | GTG | CGG | CCA | AGG | GTT | CCC | CTT | AGC | GAT | CAT | TCA | GGC | GTC | AGA | AAT | GCA | CTC | GGC | CAC | GTC | CAA |
| Gaa: | V | R | P | R | V | P | L | S | D | H | S | G | V | R | N | A | L | G | H | V | Q |
| Oaa: | T | S | I | P | E | V | L | N | H | A | P | N | Y | S | R | R | C | P | T | R | H |
| Ont: | ACC | TCT | ATC | CCA | GAG | GTT | CTG | AAC | CAT | GCC | CCG | AAT | TAC | AGC | CGC | CGG | TGC | CCA | ACG | AGA | CAT |
| Olb: | 011 | 212 | 021 | 110 | 303 | 322 | 123 | 001 | 102 | 311 | 113 | 002 | 201 | 031 | 131 | 133 | 231 | 110 | 013 | 030 | 102 |
| Glb: | 011 | 212 | 021 | 110 | 303 | 322 | 123 | 001 | 102 | 311 | 113 | 002 | 201 | 031 | 131 | 133 | 231 | 110 | 013 | 030 | 102 |
| Gnt: | ACC | TCT | ATC | CCA | GAG | GTT | CTG | AAC | CAT | GCC | CCG | AAT | TAC | AGC | CGC | CGG | TGC | CCA | ACG | AGA | CAT |
| Gaa: | T | S | I | P | E | V | L | N | H | A | P | N | Y | S | R | R | C | P | T | R | H |
| Oaa: | A | P | E | S | R | N | L | G | V | W | T | L | Q | H | A | P | R | V | R | D | A |
| Ont: | GCC | CCG | GAA | AGC | CGA | AAT | CTT | GGC | GTC | TGG | ACA | CTT | CAA | CAT | GCG | CCG | CGT | GTA | AGA | GAC | GCC |
| Olb: | 311 | 113 | 300 | 031 | 130 | 002 | 122 | 331 | 321 | 233 | 010 | 122 | 100 | 102 | 313 | 113 | 132 | 320 | 030 | 301 | 311 |
| Glb: | 311 | 113 | 300 | 031 | 130 | 002 | 122 | 331 | 321 | 233 | 010 | 122 | 100 | 102 | 313 | 113 | 132 | 320 | 030 | 301 | 311 |
| Gnt: | GCC | CCG | GAA | AGC | CGA | AAT | CTT | GGC | GTC | TGG | ACA | CTT | CAA | CAT | GCG | CCG | CGT | GTA | AGA | GAC | GCC |
| Gaa: | A | P | E | S | R | N | L | G | V | W | T | L | Q | H | A | P | R | V | R | D | A |
| Oaa: | P | G | G | R | S | D | P | R | D | S | L | W | L | A | Y | G | A | P | A | L | R |
| Ont: | CCG | GGC | GGC | AGA | TCG | GAC | CCG | AGA | GAC | AGC | CTT | TGG | CTG | GCC | TAC | GCG | GCT | CCC | GCT | CTG | CGC |
| Olb: | 113 | 331 | 331 | 030 | 213 | 301 | 113 | 030 | 301 | 031 | 122 | 233 | 123 | 311 | 201 | 331 | 312 | 111 | 312 | 123 | 131 |
| Glb: | 113 | 331 | 331 | 030 | 213 | 301 | 113 | 030 | 301 | 031 | 122 | 233 | 123 | 311 | 201 | 331 | 312 | 111 | 312 | 123 | 131 |
| Gnt: | CCG | GGC | GGC | AGA | TCG | GAC | CCG | AGA | GAC | AGC | CTT | TGG | CTG | GCC | TAC | GCG | GCT | CCC | GCT | CTG | CGC |
| Gaa: | P | G | G | R | S | D | P | R | D | S | L | W | L | A | Y | G | A | P | A | L | R |
| Oaa: | S | L | S | H | R | H | R | L | A | G | G | V | G | T | A | Q | G | E | G | V | C |
| Ont: | TCG | CTC | TCT | CAC | CGG | CAT | CGT | CTT | GCG | GCG | GGG | GTT | GGG | ACT | GCT | CAG | GGG | GAG | GGA | GTG | TGT |
| Olb: | 213 | 121 | 212 | 101 | 133 | 102 | 132 | 122 | 313 | 331 | 333 | 322 | 333 | 012 | 312 | 103 | 333 | 303 | 330 | 323 | 232 |
| Glb: | 213 | 121 | 212 | 101 | 133 | 102 | 132 | 122 | 313 | 331 | 333 | 322 | 333 | 012 | 312 | 103 | 333 | 303 | 330 | 323 | 232 |
| Gnt: | TCG | CTC | TCT | CAC | CGG | CAT | CGT | CTT | GCG | GCG | GGG | GTT | GGG | ACT | GCT | CAG | GGG | GAG | GGA | GTG | TGT |
| Gaa: | S | L | S | H | R | H | R | L | A | G | G | V | G | T | A | Q | G | E | G | V | C |
| Oaa: | R | S | S | S | P | P | A | R | L | S | L | C | L | D | E | R | R | A | R | G | V |
| Ont: | AGG | TCG | AGC | AGT | CCC | CCC | GCC | CGC | CTG | TCC | TTG | TGC | CTG | GAC | GAG | CGC | AGA | GCG | CGC | GGC | GTT |
| Olb: | 033 | 213 | 031 | 032 | 111 | 111 | 311 | 131 | 123 | 211 | 223 | 231 | 123 | 301 | 303 | 131 | 030 | 313 | 131 | 331 | 322 |
| Glb: | 033 | 213 | 031 | 032 | 111 | 111 | 311 | 131 | 123 | 211 | 223 | 231 | 123 | 301 | 303 | 131 | 030 | 313 | 131 | 331 | 322 |
| Gnt: | AGG | TCG | AGC | AGT | CCC | CCC | GCC | CGC | CTG | TCC | TTG | TGC | CTG | GAC | GAG | CGC | AGA | GCG | CGC | GGC | GTT |
| Gaa: | R | S | S | S | P | P | A | R | L | S | L | C | L | D | E | R | R | A | R | G | V |
| Oaa: | H | L | N | T | R | L | S | L | G | I | R | P | H | Q | T | P | E | S | R | I | R |
| Ont: | CAT | CTC | AAC | ACG | CGC | TTG | AGC | CTT | GGA | ATA | CGA | CCG | CAT | CAG | ACG | CCG | GAA | AGC | CGC | ATA | CGC |
| Olb: | 102 | 121 | 001 | 013 | 131 | 223 | 031 | 122 | 330 | 020 | 130 | 113 | 102 | 103 | 013 | 113 | 300 | 031 | 131 | 020 | 131 |
| Glb: | 102 | 121 | 001 | 013 | 131 | 223 | 031 | 122 | 330 | 020 | 130 | 113 | 102 | 103 | 013 | 113 | 300 | 031 | 131 | 020 | 131 |
| Gnt: | CAT | CTC | AAC | ACG | CGC | TTG | AGC | CTT | GGA | ATA | CGA | CCG | CAT | CAG | ACG | CCG | GAA | AGC | CGC | ATA | CGC |
| Gaa: | H | L | N | T | R | L | S | L | G | I | R | P | H | Q | T | P | E | S | R | I | R |
| Oaa: | R | H | I | A | E | L | K | L | E | S | L | S | D | R | P | D | A | R | L | V | P |
| Ont: | CGC | CAT | ATT | GCC | GAA | CTC | AAA | CTT | GAG | AGC | CTC | AGC | GAC | CGG | CCA | GAC | GCC | CGC | ATA | CGC | |
| Olb: | 131 | 102 | 022 | 311 | 300 | 121 | 000 | 122 | 303 | 031 | 121 | 031 | 301 | 133 | 110 | 301 | 311 | 130 | 122 | 321 | 110 |
| Glb: | 131 | 102 | 022 | 311 | 300 | 121 | 000 | 122 | 303 | 031 | 121 | 031 | 301 | 133 | 110 | 301 | 311 | 130 | 122 | 321 | 110 |
| Gnt: | CGC | CAT | ATT | GCC | GAA | CTC | AAA | CTT | GAG | AGC | CTC | AGC | GAC | CGG | CCA | GAC | GCC | CGA | CTT | GTC | CCA |
| Gaa: | R | H | I | A | E | L | K | L | E | S | L | S | D | R | P | D | A | R | L | V | P |
| Oaa: | H | A | A | S | L | P | L | P | A | R | R | Q | T | M | G | G | T | L | N | M | P |
| Ont: | CAT | GCG | GCC | AGT | CTG | CCC | CTG | CCA | GCC | AGA | CGG | CAG | ACA | ATG | GGA | GGA | ACG | CTG | AAT | ATG | CCG |
| Olb: | 102 | 313 | 311 | 032 | 123 | 111 | 123 | 110 | 311 | 030 | 133 | 103 | 010 | 023 | 330 | 330 | 013 | 123 | 002 | 023 | 113 |
| Glb: | 102 | 313 | 311 | 032 | 123 | 111 | 123 | 110 | 311 | 030 | 133 | 103 | 010 | 023 | 330 | 330 | 013 | 123 | 002 | 023 | 113 |
| Gnt: | CAT | GCG | GCC | AGT | CTG | CCC | CTG | CCA | GCC | AGA | CGG | CAG | ACA | ATG | GGA | GGA | ACG | CTG | AAT | ATG | CCG |
| Gaa: | H | A | A | S | L | P | L | P | A | R | R | Q | T | M | G | G | T | L | N | M | P |

```

Oaa: S A T R R V L R Y I L R P A Q S V I D V L
Ont: AGC GCC ACG CGC CGC GTG CTT CGC TAC ATA CTG AGA CCA GCC CAG AGC GTC ATA GAT GTG CTT
Olb: 031 311 013 131 131 323 122 131 201 020 123 030 110 311 103 031 321 020 302 323 122
Glb: 031 311 013 131 131 323 122 131 201 020 123 030 110 311 103 031 321 020 302 323 122
Gnt: AGC GCC ACG CGC CGC GTG CTT CGC TAC ATA CTG AGA CCA GCC CAG AGC GTC ATA GAT GTG CTT
Gaa: S A T R R V L R Y I L R P A Q S V I D V L

Oaa: D R L R P L L R P V G G R Q M R P D G Y K
Ont: GAC CGT CTG AGA CCA CTT CTG AGA CCC GTA GGA GGC CGC CAG ATG CGT CCA GAC GGC TAC AAG
Olb: 301 132 123 030 110 122 123 030 111 320 330 331 131 103 023 132 110 301 331 201 003
Glb: 301 132 123 030 110 122 123 030 111 320 330 331 131 103 023 332 110 301 331 201 003
Gnt: GAC CGT CTG AGA CCA CTT CTG AGA CCC GTA GGA GGC CGC CAG ATG GGT CCA GAC GGC TAC AAG
Gaa: D R L R P L L R P V G G R Q M G P D G Y K

Oaa: P M F H S R R P A S S L Y G F L S E I H G
Ont: CCT ATG TTC CAT AGC AGA AGG CCC GCC AGC AGC CTC TAC GGC TTC CTG AGC GAA ATA CAC GGC
Olb: 112 023 221 102 031 030 033 111 311 031 031 121 201 331 221 123 031 300 020 101 331
Glb: 112 023 221 102 031 030 033 111 311 031 031 121 201 331 221 123 031 300 020 101 331
Gnt: CCT ATG TTC CAT AGC AGA AGG CCC GCC AGC AGC CTC TAC GGC TTC CTG AGC GAA ATA CAC GGC
Gaa: P M F H S R R P A S S L Y G F L S E I H G

Oaa: P M Q V R N A A L L P L S D P V Q A H G R
Ont: CCC ATG CAG GTG AGG AAC GCC GCG CTT TTG CCA CTC AGT GAC CCA GTG CAA GCG CAT GGC CGA
Olb: 111 023 103 323 033 001 311 313 122 223 110 121 032 301 110 323 100 313 102 331 130
Glb: 111 023 103 323 033 001 311 313 122 223 110 121 032 301 110 323 100 313 102 331 130
Gnt: CCC ATG CAG GTG AGG AAC GCC GCG CTT TTG CCA CTC AGT GAC CCA GTG CAA GCG CAT GGC CGA
Gaa: P M Q V R N A A L L P L S D P V Q A H G R

Oaa: R G N L R S S H E S N R K T I N Q P R P V
Ont: AGG GGA AAC CTT CGT TCC TCT CAT GAG AGC AAC AGA AAG ACG ATC AAC CAG CCG AGA CCA GTC
Olb: 033 330 001 122 132 211 212 102 303 031 001 030 003 013 021 001 103 113 030 110 321
Glb: 033 330 001 122 132 211 212 102 303 031 001 030 003 013 021 001 103 113 030 110 321
Gnt: AGG GGA AAC CTT CGT TCC TCT CAT GAG AGC AAC AGA AAG ACG ATC AAC CAG CCG AGA CCA GTC
Gaa: R G N L R S S H E S N R K T I N Q P R P V

Oaa: L S S L R A C L N G E G Q R E P Y A V K F
Ont: CTT AGC AGT CTC AGG GCA TGT CTG AAC GGT GAA GGT CAA AGA GAA CCC TAC GCC GTC AAG TTT
Olb: 122 031 032 121 033 310 232 123 001 332 300 332 100 030 300 111 201 311 321 003 222
Glb: 122 031 032 121 033 310 232 123 001 332 300 332 100 030 300 111 201 311 321 003 222
Gnt: CTT AGC AGT CTC AGG GCA TGT CTG AAC GGT GAA GGT CAA AGA GAA CCC TAC GCC GTC AAG TTT
Gaa: L S S L R A C L N G E G Q R E P Y A V K F

Oaa: V V F D R P E E T H I S T S N S R R P A A
Ont: GTC GTT TTC GAC AGA CCA GAG GAA ACG CAT ATT TCG ACG AGC AAC AGC CGG AGA CCA GCC GCT
Olb: 321 322 221 301 030 110 303 300 013 102 022 213 013 031 001 031 133 030 110 311 312
Glb: 321 322 221 301 030 110 303 300 013 102 022 213 013 031 001 031 133 030 110 311 312
Gnt: GTC GTT TTC GAC AGA CCA GAG GAA ACG CAT ATT TCG ACG AGC AAC AGC CGG AGA CCA GCC GCT
Gaa: V V F D R P E E T H I S T S N S R R P A A

Oaa: T R T T F
Ont: ACA AGA ACC ACG TTT
Olb: 010 030 011 013 222
Glb: 010 030 011 013 222
Gnt: ACA AGA ACC ACG TTT
Gaa: T R T T F

```

Figura B.9: Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.55-P | A. marina - MBIC11017 plasmid pREB8 - Bacteria – GI número: 158341503

Código Klein-linearidade (1023,1013,3) BCH primitivo sobre GR(4,10), rotulamento C)

$$p(x)=x^{10}+x^7+1 \quad - \quad g(x)=x^{10}+3x^7+2x^5+1 \quad - \quad \text{Caso 03 : } (A, C, G, T) = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Oaa: | M | E | T | E | S | I | E | A | I | V | I | D | L | G | T | S | L | T | K | V | L |
| Ont: | ATG | GAA | ACT | GAA | TCA | ATC | GAA | GCG | ATC | GTG | ATC | GAC | CTG | GGG | ACA | TCC | CTC | ACG | AAA | GTT | CTG |
| Olb: | 031 | 100 | 023 | 100 | 320 | 032 | 100 | 121 | 032 | 131 | 032 | 102 | 231 | 111 | 020 | 322 | 232 | 021 | 000 | 133 | 231 |
| Glb: | 031 | 100 | 023 | 100 | 320 | 032 | 100 | 121 | 032 | 131 | 032 | 102 | 231 | 111 | 020 | 322 | 232 | 021 | 000 | 133 | 231 |
| Gnt: | ATG | GAA | ACT | GAA | TCA | ATC | GAA | GCG | ATC | GTG | ATC | GAC | CTG | GGG | ACA | TCC | CTC | ACG | AAA | GTT | CTG |
| Gaa: | M | E | T | E | S | I | E | A | I | V | I | D | L | G | T | S | L | T | K | V | L |
| Oaa: | Y | Y | Q | Q | G | Q | Y | H | F | F | A | Q | S | S | S | V | L | E | L | D | H |
| Ont: | TAC | TAC | CAA | CAG | GGG | CAA | TAT | CAC | TTC | TTT | GCT | CAA | TCG | AGT | TCC | GTC | CTA | GAG | CTA | GAC | CAT |
| Olb: | 302 | 302 | 200 | 201 | 111 | 200 | 303 | 202 | 332 | 333 | 123 | 200 | 321 | 013 | 322 | 132 | 230 | 101 | 230 | 102 | 203 |
| Glb: | 302 | 302 | 200 | 201 | 111 | 200 | 303 | 202 | 332 | 333 | 123 | 200 | 321 | 013 | 322 | 132 | 230 | 101 | 230 | 102 | 203 |
| Gnt: | TAC | TAC | CAA | CAG | GGG | CAA | TAT | CAC | TTC | TTT | GCT | CAA | TCG | AGT | TCC | GTC | CTA | GAG | CTA | GAC | CAT |
| Gaa: | Y | Y | Q | Q | G | Q | Y | H | F | F | A | Q | S | S | S | V | L | E | L | D | H |
| Oaa: | R | K | Y | Q | Q | H | A | A | H | P | Q | T | K | T | H | A | I | W | T | E | G |
| Ont: | CGC | AAG | TAT | CAA | CAA | CAT | GCA | GCT | CAC | CCC | CAG | ACC | AAG | ACC | CAT | GCC | ATC | TGG | ACT | GAA | GGG |
| Olb: | 212 | 001 | 303 | 200 | 200 | 203 | 120 | 123 | 202 | 222 | 201 | 022 | 001 | 022 | 203 | 122 | 032 | 311 | 023 | 100 | 111 |
| Glb: | 212 | 001 | 303 | 200 | 200 | 203 | 120 | 123 | 202 | 222 | 201 | 022 | 001 | 022 | 203 | 122 | 032 | 311 | 023 | 100 | 111 |
| Gnt: | CGC | AAG | TAT | CAA | CAA | CAT | GCA | GCT | CAC | CCC | CAG | ACC | AAG | ACC | CAT | GCC | ATC | TGG | ACT | GAA | GGG |
| Gaa: | R | K | Y | Q | Q | H | A | A | H | P | Q | T | K | T | H | A | I | W | T | E | G |
| Oaa: | A | G | Y | L | V | G | R | G | A | S | Q | P | Y | E | E | V | D | L | S | Q | T |
| Ont: | GCG | GGC | TAT | CTC | GTG | GGG | CGA | GGG | GCC | TCT | CAA | CCC | TAT | GAA | GAA | GTG | GAC | CTC | TCT | CAA | ACC |
| Olb: | 121 | 112 | 303 | 232 | 131 | 111 | 210 | 111 | 122 | 323 | 200 | 222 | 303 | 100 | 100 | 131 | 102 | 232 | 323 | 200 | 022 |
| Glb: | 121 | 112 | 303 | 232 | 131 | 111 | 210 | 111 | 122 | 323 | 200 | 222 | 303 | 100 | 100 | 131 | 102 | 232 | 323 | 200 | 022 |
| Gnt: | GCG | GGC | TAT | CTC | GTG | GGG | CGA | GGG | GCC | TCT | CAA | CCC | TAT | GAA | GAA | GTG | GAC | CTC | TCT | CAA | ACC |
| Gaa: | A | G | Y | L | V | G | R | G | A | S | Q | P | Y | E | E | V | D | L | S | Q | T |
| Oaa: | K | A | T | S | A | I | A | K | V | L | A | A | I | G | E | I | K | D | R | V | N |
| Ont: | AAA | GCC | ACT | TCA | GCC | ATC | GCT | AAG | GTA | CTC | GCC | GCA | ATC | GGA | GAA | ATC | AAA | GAC | CGG | GTG | AAC |
| Olb: | 000 | 122 | 023 | 320 | 122 | 032 | 123 | 001 | 130 | 232 | 122 | 120 | 032 | 110 | 100 | 032 | 000 | 102 | 211 | 131 | 002 |
| Glb: | 000 | 122 | 023 | 320 | 122 | 032 | 123 | 001 | 130 | 232 | 122 | 120 | 032 | 110 | 100 | 032 | 000 | 102 | 211 | 131 | 002 |
| Gnt: | AAA | GCC | ACT | TCA | GCC | ATC | GCT | AAG | GTA | CTC | GCC | GCA | ATC | GGA | GAA | ATC | AAA | GAC | CGG | GTG | AAC |
| Gaa: | K | A | T | S | A | I | A | K | V | L | A | A | I | G | E | I | K | D | R | V | N |
| Oaa: | S | D | K | A | I | K | L | E | N | V | I | L | L | P | L | K | E | R | K | E | |
| Ont: | AGT | GAC | AAG | GCC | ATC | AAA | CTG | GAG | AAC | GTC | ATC | CTG | TTG | CTT | CCC | CTC | AAA | GAA | CGC | AAA | GAA |
| Olb: | 013 | 102 | 001 | 122 | 032 | 000 | 231 | 101 | 002 | 132 | 032 | 231 | 331 | 233 | 222 | 232 | 000 | 100 | 212 | 000 | 100 |
| Glb: | 013 | 102 | 001 | 122 | 032 | 000 | 231 | 101 | 002 | 132 | 032 | 231 | 331 | 233 | 222 | 232 | 000 | 100 | 212 | 000 | 100 |
| Gnt: | AGT | GAC | AAG | GCC | ATC | AAA | CTG | GAG | AAC | GTC | ATC | CTG | TTG | CTT | CCC | CTC | AAA | GAA | CGC | AAA | GAA |
| Gaa: | S | D | K | A | I | K | L | E | N | V | I | L | L | P | L | K | E | R | K | E | |
| Oaa: | Y | E | T | L | K | T | A | I | V | K | A | L | Y | G | F | G | F | N | G | K | T |
| Ont: | TAT | GAA | ACC | CTA | AAG | ACT | GCC | ATC | GTT | AAA | GCA | CTC | TAT | GGA | TTT | GGG | TTC | AAC | GGC | AAG | ACC |
| Olb: | 303 | 100 | 022 | 230 | 001 | 023 | 122 | 032 | 133 | 000 | 120 | 232 | 303 | 110 | 333 | 111 | 332 | 002 | 112 | 001 | 022 |
| Glb: | 303 | 100 | 022 | 230 | 001 | 023 | 122 | 032 | 133 | 000 | 120 | 232 | 303 | 110 | 333 | 111 | 332 | 002 | 112 | 001 | 022 |
| Gnt: | TAT | GAA | ACC | CTA | AAG | ACT | GCC | ATC | GTT | AAA | GCA | CTC | TAT | GGA | TTT | GGG | TTC | AAC | GGC | AAG | ACC |
| Gaa: | Y | E | T | L | K | T | A | I | V | K | A | L | Y | G | F | G | F | N | G | K | T |
| Oaa: | T | K | F | W | V | G | R | F | N | I | Q | Y | E | G | Y | G | F | N | G | K | T |
| Ont: | ACA | AAG | TTT | TGG | GTC | GGA | CGG | TTC | AAT | ATC | CAA | TAT | GAG | GGG | TAT | GGC | ATC | ACC | ACG | ATT | TCA |
| Olb: | 020 | 001 | 333 | 311 | 132 | 110 | 211 | 332 | 003 | 032 | 200 | 303 | 101 | 111 | 303 | 112 | 032 | 022 | 021 | 033 | 320 |
| Glb: | 020 | 001 | 333 | 311 | 132 | 110 | 211 | 332 | 003 | 032 | 200 | 303 | 101 | 111 | 303 | 112 | 032 | 022 | 021 | 033 | 320 |
| Gnt: | ACA | AAG | TTT | TGG | GTC | GGA | CGG | TTC | AAT | ATC | CAA | TAT | GAG | GGG | TAT | GGC | ATC | ACC | ACG | ATT | TCA |
| Gaa: | T | K | F | W | V | G | R | F | N | I | Q | Y | E | G | Y | G | I | T | T | I | S |
| Oaa: | T | E | E | R | A | A | F | A | V | F | G | Q | K | D | L | T | I | G | A | M | N |
| Ont: | ACG | GAA | GAA | CGA | GCA | GCC | TTC | GCA | GTC | TTT | GGT | CAG | AAA | GAT | TTG | ACC | ATT | GGG | GCC | ATG | AAT |
| Olb: | 021 | 100 | 100 | 210 | 120 | 122 | 332 | 120 | 132 | 333 | 113 | 201 | 000 | 103 | 331 | 022 | 033 | 111 | 122 | 031 | 003 |
| Glb: | 021 | 100 | 100 | 210 | 120 | 122 | 332 | 120 | 132 | 333 | 113 | 201 | 000 | 103 | 331 | 022 | 033 | 111 | 122 | 031 | 003 |
| Gnt: | ACG | GAA | GAA | CGA | GCA | GCC | TTC | GCA | GTC | TTT | GGT | CAG | AAA | GAT | TTG | ACC | ATT | GGG | GCC | ATG | AAT |
| Gaa: | T | E | E | R | A | A | F | A | V | F | G | Q | K | D | L | T | I | G | A | M | N |

```

Oaa: Q G D L T P N M S Q T L V G W G M S K V L
Ont: CAG GGA GAC TTA ACC CCA AAT ATG AGT CAA ACC CTC GTG GGT TGG GGC ATG TCC AAA GTC CTT
Olb: 201 110 102 330 022 220 003 031 013 200 022 232 131 113 311 112 031 322 000 132 233
Glb: 201 110 102 330 022 220 003 031 013 200 022 232 131 113 311 112 031 322 000 132 233
Gnt: CAG GGA GAC TTA ACC CCA AAT ATG AGT CAA ACC CTC GTG GGT TGG GGC ATG TCC AAA GTC CTT
Gaa: Q G D L T P N M S Q T L V G W G M S K V L

Oaa: E E V Q Y T F E S D V Y G A K C V Y S Y N
Ont: GAA GAA GTG CAA TAC ACC TTT GAG TCG GAT GTG TAT GGG GCC AAG TGC GTC TAC TCC TAT AAC
Olb: 100 100 131 200 302 022 333 101 321 103 131 303 111 122 001 312 132 302 322 303 002
Glb: 100 100 131 200 302 022 333 101 321 103 131 303 111 122 001 312 132 302 322 303 002
Gnt: GAA GAA GTG CAA TAC ACC TTT GAG TCG GAT GTG TAT GGG GCC AAG TGC GTC TAC TCC TAT AAC
Gaa: E E V Q Y T F E S D V Y G A K C V Y S Y N

Oaa: Q R R T N L K K L I P A E K V E F V S A S
Ont: CAA CGC CGC ACC AAC CTC AAG AAA CTG ATT CCC GCC GAG AAA GTT GAA TTT GTC TCC GCT TCC
Olb: 200 212 212 022 002 232 001 000 231 033 222 122 101 000 133 100 333 132 322 123 322
Glb: 200 212 212 022 002 232 001 000 231 033 222 122 101 000 133 100 333 132 322 123 322
Gnt: CAA CGC CGC ACC AAC CTC AAG AAA CTG ATT CCC GCC GAG AAA GTT GAA TTT GTC TCC GCT TCC
Gaa: Q R R T N L K K L I P A E K V E F V S A S

Oaa: I D R A L S S S W Y Q I E R K L D S S Q A
Ont: ATA GAC CGG GCG TTG TCT TCG TCC TGG TAT CAA ATT GAG CGT AAA TTG GAT TCC TCC CAA GCC
Olb: 030 102 211 121 331 323 321 322 311 303 200 033 101 213 000 331 103 322 322 200 122
Glb: 030 102 211 121 331 323 321 322 311 303 200 033 101 213 000 332 103 322 322 200 122
Gnt: ATA GAC CGG GCG TTG TCT TCG TCC TGG TAT CAA ATT GAG CGT AAA TTC GAT TCC TCC CAA GCC
Gaa: I D R A L S S S W Y Q I E R K F D S S Q A

Oaa: L M D A V T I Y V T G G S S Q L W R K Q L
Ont: CTG ATG GAT GCC GTC ACC ATC TAC GTC ACC GGG GGC AGT TCT CAG CTC TGG CGC AAA CAG CTC
Olb: 231 031 103 122 132 022 032 302 132 022 111 112 013 323 201 232 311 212 000 201 232
Glb: 231 031 103 122 132 022 032 302 132 022 111 112 013 323 201 232 311 212 000 201 232
Gnt: CTG ATG GAT GCC GTC ACC ATC TAC GTC ACC GGG GGC AGT TCT CAG CTC TGG CGC AAA CAG CTC
Gaa: L M D A V T I Y V T G G S S Q L W R K Q L

Oaa: K G V Y G R R L N C L D G I A R E I A E T
Ont: AAA GGG GTC TAT GGT CGG CGG CTG AAT TGC CTA GAT GGC ATC GCG AGA GAA ATC GCA GAG ACA
Olb: 000 111 132 303 113 211 211 231 003 312 230 103 112 032 121 010 100 032 120 101 020
Glb: 000 111 132 303 113 211 211 231 003 312 230 103 112 032 121 010 100 032 120 101 020
Gnt: AAA GGG GTC TAT GGT CGG CGG CTG AAT TGC CTA GAT GGC ATC GCG AGA GAA ATC GCA GAG ACA
Gaa: K G V Y G R R L N C L D G I A R E I A E T

Oaa: F P Q L D K D P M I L R M A D A Y L T I R
Ont: TTC CCC CAA TTG GAC AAA GAC CCC ATG ATT CTG AGA ATG GCT GAT GCC TAT CTG ACG ATT AGA
Olb: 332 222 200 331 102 000 102 222 031 033 231 010 031 123 103 122 303 231 021 033 010
Glb: 332 222 200 331 102 000 102 222 031 033 231 010 031 123 103 122 303 231 021 033 010
Gnt: TTC CCC CAA TTG GAC AAA GAC CCC ATG ATT CTG AGA ATG GCT GAT GCC TAT CTG ACG ATT AGA
Gaa: F P Q L D K D P M I L R M A D A Y L T I R

Oaa: G M V A A
Ont: GGA ATG GTG GCC GCA
Olb: 110 031 131 122 120
Glb: 110 031 131 122 120
Gnt: GGA ATG GTG GCC GCA
Gaa: G M V A A

```

Figura B.10: Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

Seq.58-P | *B.marismortui* - ATCC 43049 plasmid pNG100 - Archea – GI número: 55376107

Código Z₂x Z₂-linearidade (1023,1013,3) BCH primitivo sobre GR(4,10), rotulado B)

$$p_{05}(x) = x^{10} + x^9 + x^8 + x^6 + x^3 + x^2 + 1 \quad g_{05}(x) = x^{10} + x^9 + 3x^8 + 2x^7 + x^6 + x^3 + x^2 + 2x + 1 \quad \text{Caso 01: } (A, C, G, T) = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| Oaa: | M | D | R | L | P | P | A | Y | D | A | T | P | G | D | G | A | L | E | A | A | I | |
| Ont: | ATG | GAT | CGG | CTT | CCG | CCA | GCA | TAC | GAC | GCT | ACT | CCC | GGG | GAT | GGC | GCC | CTC | GAA | GCA | GCA | GCC | ATC |
| Olb: | 032 | 203 | 122 | 133 | 112 | 110 | 210 | 301 | 201 | 213 | 013 | 111 | 222 | 203 | 221 | 211 | 131 | 200 | 210 | 211 | 031 | |
| Glb: | 032 | 203 | 122 | 133 | 112 | 110 | 210 | 301 | 201 | 213 | 013 | 111 | 222 | 203 | 221 | 211 | 131 | 200 | 210 | 211 | 031 | |
| Gnt: | ATG | GAT | CGG | CTT | CCG | CCA | GCA | TAC | GAC | GCT | ACT | CCC | GGG | GAT | GGC | GCC | CTC | GAA | GCA | GCA | GCC | ATC |
| Gaa: | M | D | R | L | P | P | A | Y | D | A | T | P | G | D | G | A | L | E | A | A | I | |
| Oaa: | E | E | R | L | V | D | I | D | S | G | R | Y | R | T | N | V | A | S | V | L | R | |
| Ont: | GAG | GAA | CGG | CTG | GTA | GAT | ATC | GAC | TCT | GGA | CGC | TAC | CGA | ACA | AAC | GTC | GCG | AGC | GTG | CTG | CGA | |
| Olb: | 202 | 200 | 122 | 132 | 230 | 203 | 031 | 201 | 313 | 220 | 121 | 301 | 120 | 010 | 001 | 231 | 212 | 021 | 232 | 132 | 120 | |
| Glb: | 202 | 200 | 122 | 132 | 230 | 203 | 031 | 201 | 313 | 220 | 121 | 301 | 120 | 010 | 001 | 231 | 212 | 021 | 232 | 132 | 120 | |
| Gnt: | GAG | GAA | CGG | CTG | GTA | GAT | ATC | GAC | TCT | GGA | CGC | TAC | CGA | ACA | AAC | GTC | GCG | AGC | GTG | CTG | CGA | |
| Gaa: | E | E | R | L | V | D | I | D | S | G | R | Y | R | T | N | V | A | S | V | L | R | |
| Oaa: | K | F | A | T | W | A | R | D | Q | H | G | I | S | S | P | E | D | I | D | D | D | |
| Ont: | AAG | TTC | GCA | ACG | TGG | GCT | CGG | GAC | CAG | CAC | GGT | ATC | TCC | AGT | CCT | GAA | GAC | ATC | GAC | GAC | GAT | |
| Olb: | 002 | 331 | 210 | 012 | 322 | 213 | 122 | 201 | 102 | 101 | 223 | 031 | 311 | 023 | 113 | 200 | 201 | 031 | 201 | 201 | 203 | |
| Glb: | 002 | 331 | 210 | 012 | 322 | 213 | 122 | 201 | 102 | 101 | 223 | 031 | 311 | 023 | 113 | 200 | 201 | 031 | 201 | 201 | 203 | |
| Gnt: | AAG | TTC | GCA | ACG | TGG | GCT | CGG | GAC | CAG | CAC | GGT | ATC | TCC | AGT | CCT | GAA | GAC | ATC | GAC | GAC | GAT | |
| Gaa: | K | F | A | T | W | A | R | D | Q | H | G | I | S | S | P | E | D | I | D | D | D | |
| Oaa: | L | C | R | Q | Y | A | R | D | L | A | R | A | D | D | R | D | D | I | S | P | E | |
| Ont: | CTC | TGC | CGA | CAG | TAT | GCT | CGC | GAC | CTT | GCT | CGA | GCT | GAC | GAC | CGG | GAC | GAC | ATC | TCA | CCG | GAG | |
| Olb: | 131 | 321 | 120 | 102 | 303 | 213 | 121 | 201 | 133 | 213 | 120 | 213 | 201 | 201 | 122 | 201 | 201 | 031 | 310 | 112 | 202 | |
| Glb: | 131 | 321 | 120 | 102 | 303 | 213 | 121 | 201 | 133 | 213 | 120 | 213 | 201 | 203 | 122 | 201 | 201 | 031 | 310 | 112 | 202 | |
| Gnt: | CTC | TGC | CGA | CAG | TAT | GCT | CGC | GAC | CTT | GCT | CGA | GCT | GAC | GAT | CGG | GAC | GAC | ATC | TCA | CCG | GAG | |
| Gaa: | L | C | R | Q | Y | A | R | D | L | A | R | A | D | D | R | D | D | I | S | P | E | |
| Oaa: | T | A | R | R | Y | F | A | Y | V | R | S | F | L | T | W | A | V | Y | E | G | L | |
| Ont: | ACC | GCC | CGC | CGC | TAC | TTC | GCG | TAT | GTC | CGC | TCG | TTT | CTC | ACA | TGG | GCT | GTC | TAC | GAG | GGC | CTG | |
| Olb: | 011 | 211 | 121 | 121 | 301 | 331 | 212 | 303 | 231 | 121 | 312 | 333 | 131 | 010 | 322 | 213 | 231 | 301 | 202 | 221 | 132 | |
| Glb: | 011 | 211 | 121 | 121 | 301 | 331 | 212 | 303 | 231 | 121 | 312 | 333 | 131 | 010 | 322 | 213 | 231 | 301 | 202 | 221 | 132 | |
| Gnt: | ACC | GCC | CGC | CGC | TAC | TTC | GCG | TAT | GTC | CGC | TCG | TTT | CTC | ACA | TGG | GCT | GTC | TAC | GAG | GGC | CTG | |
| Gaa: | T | A | R | R | Y | F | A | Y | V | R | S | F | L | T | W | A | V | Y | E | G | L | |
| Oaa: | I | P | T | N | P | A | K | T | N | H | A | E | G | P | L | P | T | D | E | T | E | |
| Ont: | ATT | CCG | ACG | AAT | CCG | GCC | AAG | ACC | AAC | CAC | GCC | GAA | GGG | CCG | CTC | CCG | ACC | GAT | GAG | ACT | GAA | |
| Olb: | 033 | 112 | 012 | 003 | 112 | 211 | 002 | 011 | 001 | 101 | 211 | 200 | 222 | 112 | 131 | 112 | 011 | 203 | 202 | 013 | 200 | |
| Glb: | 033 | 112 | 012 | 003 | 112 | 211 | 002 | 011 | 001 | 101 | 211 | 200 | 222 | 112 | 131 | 112 | 011 | 203 | 202 | 013 | 200 | |
| Gnt: | ATT | CCG | ACG | AAT | CCG | GCC | AAG | ACC | AAC | CAC | GCC | GAA | GGG | CCG | CTC | CCG | ACC | GAT | GAG | ACT | GAA | |
| Gaa: | I | P | T | N | P | A | K | T | N | H | A | E | G | P | L | P | T | D | E | T | E | |
| Oaa: | T | D | Q | Q | Y | W | T | T | R | D | R | D | A | I | C | A | T | A | T | A | R | |
| Ont: | ACA | GAC | CAG | CAG | TAC | TGG | ACG | ACA | CGC | GAC | CGC | GAC | GCT | ATC | TGC | GCT | ACC | GCG | ACG | GCT | CGC | |
| Olb: | 010 | 201 | 102 | 102 | 301 | 322 | 012 | 010 | 121 | 201 | 121 | 201 | 213 | 031 | 321 | 213 | 011 | 212 | 012 | 213 | 121 | |
| Glb: | 010 | 201 | 102 | 102 | 301 | 322 | 012 | 010 | 121 | 201 | 121 | 201 | 213 | 031 | 321 | 213 | 011 | 212 | 012 | 213 | 121 | |
| Gnt: | ACA | GAC | CAG | CAG | TAC | TGG | ACG | ACA | CGC | GAC | CGC | GAC | GCT | ATC | TGC | GCT | ACC | GCG | ACG | GCT | CGC | |
| Gaa: | T | D | Q | Q | Y | W | T | T | R | D | R | D | A | I | C | A | T | A | T | A | R | |
| Oaa: | V | D | K | A | G | E | S | D | D | I | D | R | T | A | A | Y | R | D | Q | A | L | |
| Ont: | GTC | GAC | AAG | GCC | GGC | GAG | AGC | GAC | GAC | GCT | CGC | ACG | GCG | GCC | TAC | GCT | CGC | GAC | CAA | GCA | CTC | |
| Olb: | 231 | 201 | 002 | 211 | 221 | 202 | 021 | 201 | 201 | 031 | 201 | 121 | 012 | 212 | 211 | 301 | 121 | 201 | 100 | 210 | 131 | |
| Glb: | 231 | 201 | 002 | 211 | 221 | 202 | 021 | 201 | 201 | 031 | 201 | 121 | 012 | 212 | 211 | 301 | 121 | 201 | 100 | 210 | 131 | |
| Gnt: | GTC | GAC | AAG | GCC | GGC | GAG | AGC | GAC | GAC | GCT | CGC | ACG | GCG | GCC | TAC | GCT | CGC | GAC | CAA | GCA | CTC | |
| Gaa: | V | D | K | A | G | E | S | D | D | I | D | R | T | A | A | Y | R | D | Q | A | L | |
| Oaa: | V | F | L | L | A | Y | S | G | A | R | S | A | E | L | V | A | V | S | D | D | E | |
| Ont: | GTC | TTC | TTG | CTC | CGC | TAC | TCA | GGA | GCC | CGC | AGC | GCG | GAA | CTG | GTT | GCT | GTC | TCT | GAT | GAT | GAA | |
| Olb: | 231 | 331 | 332 | 131 | 212 | 301 | 310 | 220 | 211 | 121 | 021 | 212 | 200 | 132 | 233 | 213 | 313 | 203 | 203 | 200 | | |
| Glb: | 231 | 331 | 332 | 131 | 212 | 301 | 310 | 220 | 211 | 121 | 021 | 212 | 200 | 132 | 233 | 213 | 313 | 203 | 203 | 200 | | |
| Gnt: | GTC | TTC | TTG | CTC | CGC | TAC | TCA | GGA | GCC | CGC | AGC | GCG | GAA | CTG | GTT | GCT | GTC | TCT | GAT | GAT | GAA | |
| Gaa: | V | F | L | L | A | Y | S | G | A | R | S | A | E | L | V | A | V | S | D | D | E | |

```

Oaa: E R N G L R W R H V N L E A G T M Q V F G
Ont: GAA CGG AAT GGT CTG CGG TGG CGA CAC GTC AAC CTC GAA GCC GGC ACA ATG CAG GTG TTC GGC
Olb: 200 122 003 223 132 122 322 120 101 231 001 131 200 211 221 010 032 102 232 331 221
Glb: 200 122 003 223 132 122 322 120 101 231 001 131 200 211 221 010 032 102 232 331 221
Gnt: GAA CGG AAT GGT CTG CGG TGG CGA CAC GTC AAC CTC GAA GCC GGC ACA ATG CAG GTG TTC GGC
Gaa: E R N G L R W R H V N L E A G T M Q V F G

Oaa: K N R T R E S A P I L D D A L R P L R R W
Ont: AAG AAC CGT ACC CGA GAG TCT GCG CCG ATC CTC GAC GAT GCA CTT CGC CCG CTT CGG CGC TGG
Olb: 002 001 123 011 120 202 313 212 112 031 131 201 203 210 133 121 112 133 122 121 322
Glb: 002 001 123 011 120 202 313 212 112 031 131 201 203 210 133 121 112 133 122 121 322
Gnt: AAG AAC CGT ACC CGA GAG TCT GCG CCG ATC CTC GAC GAT GCA CTT CGC CCG CTT CGG CGC TGG
Gaa: K N R T R E S A P I L D D A L R P L R R W

Oaa: K Q L R E P D E N E A V F P R L D N A A K
Ont: AAG CAG CTT CGA GAA CCC GAC GAG AAT GAG GCC GTG TTC CCC AGA CTC GAC AAC GCT GCG AAA
Olb: 002 102 133 120 200 111 201 202 003 202 211 232 331 111 020 131 201 001 213 212 000
Glb: 002 102 133 120 200 111 201 202 003 202 211 232 331 111 020 131 201 001 213 212 000
Gnt: AAG CAG CTT CGA GAA CCC GAC GAG AAT GAG GCC GTG TTC CCC AGA CTC GAC AAC GCT GCG AAA
Gaa: K Q L R E P D E N E A V F P R L D N A A K

Oaa: A L D P T P S I T T Q S A R N I L A D L C
Ont: GCG CTT GAT CCC ACG CCG TCG ATC ACC ACG CAG TCG GCC CGG AAC ATC TTG GCA GAC CTG TGT
Olb: 212 133 203 111 012 112 312 031 011 012 102 312 211 122 001 031 332 210 201 132 323
Glb: 212 133 203 111 012 112 312 031 011 012 102 312 211 122 001 031 332 210 201 132 323
Gnt: GCG CTT GAT CCC ACG CCG TCG ATC ACC ACG CAG TCG GCC CGG AAC ATC TTG GCA GAC CTG TGT
Gaa: A L D P T P S I T T Q S A R N I L A D L C

Oaa: E W S E Y E F E D P L K P H G A R R G L G
Ont: GAG TGG TCT GAG TAC GAG TTC GAA GAC CCG CTG AAG CCA CAC GGT GCT CGC CGT GGA TTG GGA
Olb: 202 322 313 202 301 202 331 200 201 112 132 002 110 101 223 213 121 123 220 332 220
Glb: 202 322 313 202 301 202 331 200 201 112 132 002 110 101 223 213 121 123 220 332 220
Gnt: GAG TGG TCT GAG TAC GAG TTC GAA GAC CCG CTG AAG CCA CAC GGT GCT CGC CGT GGA TTG GGA
Gaa: E W S E Y E F E D P L K P H G A R R G L G

Oaa: R E I Y R E N P Q L A Q D V L R H K S I E
Ont: CGG GAG ATC TAT CGC GAG AAC CCG CAA CTG GCA CAG GAC GTG CTC CGG CAC AAA TCC ATC GAG
Olb: 122 202 031 303 121 202 001 112 100 132 210 102 201 232 131 122 101 000 311 031 202
Glb: 122 202 031 303 121 202 001 112 100 132 210 102 201 232 131 122 101 000 311 031 202
Gnt: CGG GAG ATC TAT CGC GAG AAC CCG CAA CTG GCA CAG GAC GTG CTC CGG CAC AAA TCC ATC GAG
Gaa: R E I Y R E N P Q L A Q D V L R H K S I E

Oaa: T T H E G Y A Q E A A K R T R D E A N E I
Ont: ACA ACC CAC GAA GGC TAC GCT CAG GAG GCC GCG AAA CGG ACT CGT GAC GAG GCG AAC GAG ATT
Olb: 010 011 101 200 221 301 213 102 202 211 212 000 122 013 123 201 202 212 001 202 033
Glb: 010 011 101 200 221 301 213 102 202 211 212 000 122 013 123 201 202 212 001 202 033
Gnt: ACA ACC CAC GAA GGC TAC GCT CAG GAG GCC GCG AAA CGG ACT CGT GAC GAG GCG AAC GAG ATT
Gaa: T T H E G Y A Q E A A K R T R D E A N E I

Oaa: I S D T E
Ont: ATC TCA GAC ACA GAA
Olb: 031 310 201 010 200
Glb: 031 310 201 010 200
Gnt: ATC TCA GAC ACA GAA
Gaa: I S D T E

```

Figura B.11: Proteína codificada por 1023 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 1$.

B.3 BCH não Primitivo sobre $GR(4, r)$

Seq.60- | *H. sapiens* - Sequence 19 from Patent WO2008043521 – GI número: 240153946

Código Klein-linearidade (21,15,3) BCH primitivo sobre $GR(4,6)$, rotulamento C)

$$p_{04}(x) = x^6 + x^9 + x^5 + x^2 + x + 1 \quad - \quad g_{04}(x) = x^6 + 3x^3 + 2x^4 + x^2 + x + 1 \quad - \quad \text{Caso 03: } (A, C, G, T) = (0, 2, 1, 3)$$

| | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|
| ntO: | TAA | AGT | GCT | GAC | AGT | GCA | GAT |
| RtO: | 300 | 013 | 123 | 102 | 013 | 120 | 103 |
| RtG: | 300 | 203 | 123 | 102 | 013 | 120 | 103 |
| ntG: | TAA | CAT | GCT | GAC | AGT | GCA | GAT |

Figura B.12: MicroRNA com 21 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$.

Seq.01-SD | *A. thaliana* - Mitocôndria – GI número: 30686795

Código Z_4 -linearidade (51,43,3) BCH não primitivo sobre $GR(4,8)$, rotulamento A)

$$p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 3x^3 + 2x^2 + 3x + 1$$

Caso 02– (A, C, G, T) = (0, 1, 3, 2)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | S | A | L | A | L | K | R | L | L | S | S | S | I | A | P | R |
| ntO: | GCA | TCT | GCT | CTC | GCT | CTT | AAG | AGA | CTC | CTA | TCA | TCC | TCC | ATC | GCT | CCA | CGT |
| RtO: | 310 | 212 | 312 | 121 | 312 | 122 | 003 | 030 | 121 | 120 | 210 | 211 | 211 | 021 | 312 | 110 | 132 |
| RtG: | 310 | 212 | 312 | 121 | 312 | 122 | 003 | 030 | 121 | 120 | 213 | 211 | 211 | 021 | 312 | 113 | 132 |
| ntG: | GCA | TCT | GCT | CTC | GCT | CTT | AAG | AGA | CTC | CTA | TCG | TCC | TCC | ATC | GCT | CCG | CGT |
| aaG: | A | S | A | L | A | L | K | R | L | L | S | S | S | I | A | P | R |

Código Klein-linearidade (255,247,3) BCH não primitivo sobre $GR(4,8)$, rotulamento C)

$$p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1 - g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$$

Caso 02– (A, C, G, T) = (0, 2, 1, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | A | S | A | L | A | L | K | R | L | L | S | S | S | I | A | P | R |
| ntO: | GCA | TCT | GCT | CTC | GCT | CTT | AAG | AGA | CTC | CTA | TCA | TCC | TCC | ATC | GCT | CCA | CGT |
| RtO: | 120 | 323 | 123 | 232 | 123 | 233 | 001 | 010 | 232 | 230 | 320 | 322 | 322 | 032 | 123 | 220 | 213 |
| RtG: | 120 | 223 | 123 | 232 | 123 | 233 | 001 | 010 | 332 | 230 | 320 | 322 | 322 | 032 | 123 | 220 | 213 |
| ntG: | GCA | CCT | GCT | CTC | GCT | CTT | AAG | AGA | TTG | CTA | TCA | TCC | TCC | ATC | GCT | CCA | CGT |
| aaG: | A | P | A | L | A | L | K | R | F | L | S | S | S | I | A | P | R |

Figura B.13: Sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$.

Seq.02-SD | S.tuberosum - Precursor of the 59kDa subunit of the mitochondrial, Matriz – GI número: 438130

Código Z₄-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento A

$$P_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 3x^3 + 2x^2 + 3x + 1$$

$$\text{Caso 02- (A, C, G, T)} = (0, 1, 3, 2)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | W | R | V | A | R | S | A | A | S | T | F | R | R | T | R | R | L |
| ntO: | TGG | AGA | GTG | GCT | CGA | TCT | GCG | GCG | TCG | ACT | TTC | CGC | CGT | ACG | CGG | CGG | TTA |
| RtO: | 233 | 030 | 323 | 312 | 130 | 212 | 313 | 313 | 213 | 012 | 221 | 131 | 132 | 013 | 133 | 133 | 220 |
| RtG: | 233 | 030 | 323 | 312 | 130 | 212 | 323 | 313 | 213 | 012 | 222 | 131 | 132 | 013 | 133 | 133 | 220 |
| ntG: | TGG | AGA | GTG | GCT | CGA | TCT | G TG | GCG | TCG | ACT | TTT | CGC | CGT | ACG | CGG | CGG | TTA |
| aaG: | W | R | V | A | R | S | V | A | S | T | F | R | R | T | R | R | L |

Código Z₂x Z₂ (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento B

$$P_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 3x^3 + 2x^2 + 3x + 1$$

$$\text{Caso 01- (A, C, G, T)} = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|------------|-----|-----|-----|-----|-----|-----|-------------|-----|-----|-----|-----|-----|-----|
| aaO: | W | R | V | A | R | S | A | A | S | T | F | R | R | T | R | R | L |
| ntO: | TGG | AGA | GTG | GCT | CGA | TCT | GCG | GCG | TCG | ACT | TTC | CGC | CGT | ACG | CGG | CGG | TTA |
| RtO: | 322 | 020 | 232 | 213 | 120 | 313 | 212 | 212 | 312 | 013 | 331 | 121 | 123 | 012 | 122 | 122 | 330 |
| RtG: | 322 | 020 | 232 | 212 | 120 | 313 | 212 | 212 | 312 | 013 | 321 | 121 | 123 | 012 | 122 | 122 | 330 |
| ntG: | TGG | AGA | GTG | GCG | CGA | TCT | GCG | GCG | TCG | ACT | T GC | CGC | CGT | ACG | CGG | CGG | TTA |
| aaG: | W | R | V | A | R | S | A | A | S | T | C | R | R | T | R | R | L |

Código Klein-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento C)

$$P_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 3x^3 + 2x^2 + 3x + 1$$

$$\text{Caso 03- (A, C, G, T)} = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | W | R | V | A | R | S | A | A | S | T | F | R | R | T | R | R | L |
| ntO: | TGG | AGA | GTG | GCT | CGA | TCT | GCG | GCG | TCG | ACT | TTC | CGC | CGT | ACG | CGG | CGG | TTA |
| RtO: | 311 | 010 | 131 | 123 | 210 | 323 | 121 | 121 | 321 | 023 | 332 | 212 | 213 | 021 | 211 | 211 | 330 |
| RtG: | 311 | 010 | 131 | 121 | 210 | 320 | 121 | 121 | 321 | 023 | 332 | 212 | 213 | 021 | 211 | 211 | 330 |
| ntG: | TGG | AGA | GTG | GCG | CGA | TCA | GCG | GCG | TCG | ACT | TTC | CGC | CGT | ACG | CGG | CGG | TTA |
| aaG: | W | R | V | A | R | S | A | A | S | T | F | R | R | T | R | R | L |

Figura B.14: Sequência de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$.

Seq.03-SD | H. sapiens – Retículo endoplasmático – GI número: 298590

Código Z₄-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento A)

$$p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1 - g_{02}(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$$

$$\text{Caso 02- (A, C, G, T)} = (0, 1, 3, 2)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | D | Y | L | L | M | I | F | S | L | L | F | V | A | C | Q | G |
| ntO: | ATG | GAT | TAT | TTG | CTC | ATG | ATT | TTC | TCT | CTG | CTG | TTT | GTG | GCT | TGC | CAA | GGA |
| RtO: | 023 | 302 | 202 | 223 | 121 | 023 | 022 | 221 | 212 | 123 | 123 | 222 | 323 | 312 | 231 | 100 | 330 |
| RtG: | 023 | 302 | 201 | 223 | 121 | 023 | 022 | 221 | 212 | 123 | 123 | 222 | 323 | 312 | 233 | 100 | 330 |
| ntG: | ATG | GAT | TAC | TTG | CTC | ATG | ATT | TTC | TCT | CTG | CTG | TTT | GTG | GCT | TGG | CAA | GGA |
| aaG: | M | D | Y | L | L | M | I | F | S | L | L | F | V | A | W | Q | G |

Código Klein-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento C)

$$p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1 - g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$$

$$\text{Caso 03- (A, C, G, T)} = (0, 2, 1, 3)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | D | Y | L | L | M | I | F | S | L | L | F | V | A | C | Q | G |
| ntO: | ATG | GAT | TAT | TTG | CTC | ATG | ATT | TTC | TCT | CTG | CTG | TTT | GTG | GCT | TGC | CAA | GGA |
| RtO: | 031 | 103 | 303 | 331 | 232 | 031 | 033 | 332 | 323 | 231 | 231 | 333 | 131 | 123 | 312 | 200 | 110 |
| RtG: | 031 | 103 | 303 | 331 | 232 | 031 | 033 | 332 | 313 | 221 | 231 | 333 | 131 | 123 | 312 | 200 | 110 |
| ntG: | ATG | GAT | TAT | TTG | CTC | ATG | ATT | TTC | TGT | CCG | CTG | TTT | GTG | GCT | TGC | CAA | GGA |
| aaG: | M | D | Y | L | L | M | I | F | C | P | L | F | V | A | C | Q | G |

Seq.04-SD | S. cerevisiae – ATP syntase, Membrana Interna Mitochondrial 8 – GI número: 114685

Código Z_{2x} Z₂ (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento B)

$$p_{02}(x) = x^8 + x^6 + x^5 + x^3 + 1 - g_{02}(x) = x^8 + 2x^6 + 3x^4 + 3x^3 + 3x + 1$$

$$\text{Caso 01- (A, C, G, T)} = (0, 1, 2, 3)$$

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | F | N | R | V | F | T | R | S | F | A | S | S | L | R | A | A |
| ntO: | ATG | TTT | AAT | AGA | GTC | TTT | ACC | AGG | TCA | TTT | GCA | TCA | AGC | TTA | AGA | GCT | GCT |
| RtO: | 032 | 333 | 003 | 020 | 231 | 333 | 011 | 022 | 310 | 333 | 210 | 310 | 021 | 330 | 020 | 213 | 213 |
| RtG: | 032 | 333 | 003 | 020 | 232 | 333 | 011 | 032 | 310 | 333 | 210 | 310 | 021 | 330 | 020 | 213 | 213 |
| ntG: | ATG | TTT | AAT | AGA | GTC | TTT | ACC | ATG | TCA | TTT | GCA | TCA | AGC | TTA | AGA | GCT | GCT |
| aaG: | M | F | N | R | V | F | T | M | S | F | A | S | S | L | R | A | A |

Figura B.15: Sequências de direcionamento com 51 nucleotídeos e $D(\mathbf{a}, \mathbf{b}) = 2$.

Seq.05-SD | S. cerevisiae – Phenylalanyl-tRNA synthetase, mitochondrial precursor (Phenylalanine-tRNA ligase) (PheRS) – GI número: 114152885

Código Z₂x Z₂ (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento B)

$$p_{01}(x) = x^8 + x^4 + x^3 + x^2 + 1 - g_{01}(x) = x^8 + x^7 + x^6 + x^5 + x^4 + 3x^3 + 2x^2 + 3x + 1$$

Caso 01– (A, C, G, T) = (0, 1, 2, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | F | L | N | R | M | M | K | T | R | T | G | L | Y | R | L | Y |
| ntO: | ATG | TTT | CTC | AAT | AGA | ATG | ATG | AAG | ACC | AGG | ACT | GGT | CTT | TAT | CGC | TTA | TAT |
| RtO: | 032 | 333 | 131 | 003 | 020 | 032 | 032 | 002 | 011 | 022 | 013 | 223 | 133 | 303 | 121 | 330 | 303 |
| RtG: | 032 | 333 | 131 | 003 | 020 | 033 | 032 | 002 | 011 | 022 | 013 | 123 | 133 | 303 | 121 | 330 | 303 |
| ntG: | ATG | TTT | CTC | AAT | AGA | ATT | ATG | AAG | ACC | AGG | ACT | CGT | CTT | TAT | CGC | TTA | TAT |
| aaG: | M | F | L | N | R | I | M | K | T | R | T | R | L | Y | R | L | Y |

Código Z₂x Z₂ (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento B)

$$p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1 - g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$$

Caso 01– (A, C, G, T) = (0, 1, 2, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | F | L | N | R | M | M | K | T | R | T | G | L | Y | R | L | Y |
| ntO: | ATG | TTT | CTC | AAT | AGA | ATG | ATG | AAG | ACC | AGG | ACT | GGT | CTT | TAT | CGC | TTA | TAT |
| RtG: | 032 | 333 | 131 | 003 | 020 | 032 | 032 | 002 | 011 | 022 | 013 | 223 | 133 | 303 | 121 | 330 | 303 |
| RtG: | 032 | 333 | 131 | 003 | 020 | 032 | 032 | 002 | 111 | 022 | 013 | 223 | 133 | 333 | 121 | 330 | 303 |
| ntG: | ATG | TTT | CTC | AAT | AGA | ATG | ATG | AAG | CCC | AGG | ACT | GGT | CTT | TTT | CGC | TTA | TAT |
| aaG: | M | F | L | N | R | M | M | K | P | R | T | G | L | F | R | L | Y |

Seq.06-SD | X. laevis – Single-stranded DNA-binding protein R, mitochondrial precursor, (Mt-SSB-R) (Mt-SSB 2) – GI número: 2815500

Código Klein-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento C)

$$p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1 - g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$$

Caso 03– (A, C, G, T) = (0, 2, 1, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| aaO: | M | F | H | R | P | V | L | Q | V | F | R | Q | F | A | R | C | Q |
| ntO: | ATG | TTT | CAT | CGG | CCG | GTC | CTT | CAG | GTG | TTT | CGC | CAG | TTT | GCA | AGA | TGT | CAG |
| RtO: | 031 | 333 | 203 | 211 | 221 | 132 | 233 | 201 | 131 | 333 | 212 | 201 | 333 | 120 | 010 | 313 | 201 |
| RtO: | 031 | 333 | 203 | 211 | 221 | 132 | 233 | 201 | 131 | 330 | 212 | 201 | 333 | 120 | 013 | 313 | 201 |
| ntG: | ATG | TTT | CAT | CGG | CCG | GTC | CTT | CAG | GTG | TTA | CGC | CAG | TTT | GCA | AGT | TGT | CAG |
| aaG: | M | F | H | R | P | V | L | Q | V | L | R | Q | F | A | N | C | Q |

Figura B.16: Sequências de direcionamento com 51 nucleotídeos e $D(a, b) = 2$.

Seq.06-SD | X. laevis – Single-stranded DNA-binding protein R, mitochondrial precursor, (Mt-SSB-R) (Mt-SSB-2)– GI número: 2815500

Código Z₂x Z₂ (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento B)

$$p_{05}(x) = x^8 + x^6 + x^5 + x^2 + 1 - g_{05}(x) = x^8 + 3x^7 + 2x^6 + x^4 + x^3 + x^2 + x + 1$$

Caso 01– (A, C, G, T) = (0, 1, 2, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|
| aaO: | M | F | H | R | P | V | L | Q | V | F | R | Q | F | A | R | C | Q |
| ntO: | ATG | TTT | CAT | CGG | CCG | GTC | CTT | CAG | GTG | TTT | CGC | CAG | TTT | GCA | AGA | TGT | CAG |
| RtO: | 032 | 333 | 103 | 122 | 112 | 231 | 133 | 102 | 232 | 333 | 121 | 102 | 333 | 210 | 020 | 323 | 102 |
| RtG: | 032 | 303 | 103 | 122 | 112 | 231 | 133 | 102 | 232 | 333 | 121 | 102 | 333 | 210 | 020 | 323 | 302 |
| ntG: | ATG | TAT | CAT | CGG | CCG | GTC | CTT | CAG | GTG | TTT | CGC | CAG | TTT | GCA | AGA | TGT | TAG |
| aaG: | M | Y | H | R | P | V | L | Q | V | F | R | Q | F | A | R | C | sto |

Seq.07-SD | S. cerevisiae – 54S ribosomal protein L20, mitochondrial precursor – GI número: 548791

Código Z₄-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento A)

$$p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1 - g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$$

Caso 02– (A, C, G, T) = (0, 1, 3, 2)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----------|-----|
| aaO: | I | G | R | G | V | C | C | R | S | F | H | T | A | G | S | A | W |
| ntO: | ATT | GGC | AGA | GGT | GTG | TGC | TGT | AGA | TCG | TTC | CAC | ACT | GCT | GGA | TCT | GCC | TGG |
| RtO: | 022 | 331 | 030 | 332 | 323 | 231 | 232 | 030 | 213 | 221 | 101 | 012 | 312 | 330 | 212 | 311 | 233 |
| RtG: | 022 | 331 | 030 | 332 | 313 | 231 | 232 | 030 | 213 | 221 | 101 | 012 | 312 | 330 | 212 | 301 | 233 |
| ntG: | ATT | GGC | AGA | GGT | CG | TGC | TGT | AGA | TCG | TTC | CAC | ACT | GCT | GGA | TCT | AC | TGG |
| aaG: | I | G | R | G | A | C | C | R | S | F | H | T | A | G | S | D | W |

Seq.08-SD | S. cerevisiae – Malate dehydrogenase, mitochondrial precursor, matrix – GI número: 547901

Código Klein-linearidade (51,43,3) BCH não primitivo sobre GR(4,8), rotulamento C)

$$p_{03}(x) = x^8 + x^7 + x^6 + x^5 + x^2 + x + 1 - g_{03}(x) = x^8 + 3x^7 + 3x^5 + 3x^4 + 2x^2 + 1$$

Caso 03– (A, C, G, T) = (0, 2, 1, 3)

| | | | | | | | | | | | | | | | | | |
|------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|------------|------------|-----|
| aaO: | M | L | S | R | V | A | K | R | A | F | S | S | T | V | A | N | P |
| ntO: | ATG | TTG | TCA | AGA | GTA | GCT | AAA | CGT | GCG | TTT | TCC | TCT | ACA | GTT | GCC | AAC | CCT |
| RtO: | 031 | 331 | 320 | 010 | 130 | 123 | 000 | 213 | 121 | 333 | 322 | 323 | 020 | 133 | 122 | 002 | 223 |
| RtG: | 031 | 331 | 320 | 010 | 130 | 123 | 000 | 213 | 121 | 333 | 322 | 323 | 020 | 133 | 022 | 022 | 223 |
| ntG: | ATG | TTG | TCA | AGA | GTA | GCT | AAA | CGT | GCG | TTT | TCC | TCT | ACA | GTT | AC | AC | CCT |
| aaG: | M | L | S | R | V | A | K | R | A | F | S | S | T | V | T | T | P |

Figura B.17: Sequências de direcionamento com 51 nucleotídeos e $D(a, b) = 2$.

Referências Bibliográficas

- [1] G. Battail, *An Outline of Informational Genetics*, Morgan & Claypool Publishers, (2008).
- [2] C.E. Shannon, “A mathematical theory of communication,” *Bell Syst. Tech. J.*, **27**, pp. 397-423 and 623-656, (1948). Reprinted in: C.E.Shannon and W.Weaver, eds., *A Mathematical Theory of Communication*, (Univ. of Illinois Press, Urbana, Illinois, (1963)).
- [3] E. Mayr, *The Growth of Biological Thought*, Harvard University Press, (1982).
- [4] “Convenção sobre Diversidade Biológica (Artigo 2. Utilização de Termos).” *Nações Unidas*, 1992. Recuperado em 27 de março de 2008.
- [5] T.D. Schneider, G.D. Stormo, L. Gold, and A. Dhrenfeucht, “Information content of binding sites on nucleotide sequences,” *Journal of Molecular Biology*, **188**, pp. 415-431, (1986).
- [6] T.D. Schneider and R.M. Stephens, “Sequence Logos: a New Way to Display Consensus Sequences,” *Nucleic Acids Research*, **18** (20), pp. 6097-6100, September (1990).
- [7] T.D. Schneider, “Information content of individual genetic sequences,” *Journal of Theoretical Biology*, **189**, pp.427-441, (1997).
- [8] H. Yockey, *Information Theory and Molecular Biology*, Cambridge University Press: Cambridge, (1992).
- [9] D.R. Forsdyke, “Are introns in-series error detecting sequences?,” *Intel. J. Theor. Biol.*, vol.**93**, pp.861-866, (1981).
- [10] D.R. Forsdyke, “Conservation of stem-loop potential in introns of snake venom phospholipase A2 genes. An application of FORS-D analysis,” *Mol. Biol. and Evol.*, vol.**12**, pp.1157-1165, (1995).

- [11] J. Rzeszowska-Wolny, "Is genetic code error-correcting?", *J. Theor. Biol.*, vol.**104**, pp.701-702, (1983).
- [12] L.S. Liebovitch, Y.Tao, A.T. Todorov, and L. Levine, "Is there an error correcting code in the base sequence in DNA?", *Biophysical Journal*, vol.**71**, pp.1539-1544, (1996).
- [13] G.L.Rosen, "Examining coding structure and redundancy in DNA," *IEEE Engineering in Medicine and Biology*, vol.**25**, pp.62-68, (2006).
- [14] G. Battail, *Information Theory and Error Correcting Codes in Genetics and Biological Evolution, Introduction to Biosemiotics*, Springer: New York, USA, (2006).
- [15] E.E. May, M. Vouk, D. Bitzer and D. Rosnick, "An error-correcting code framework for genetic sequence analysis," *Journal of the Franklin Institute*, vol.**34**, pp.89-109, (2004).
- [16] Mac Donaill DA, "Why nature chose A, C, G and U/T: an error-coding perspective of nucleotide alphabet composition," *Origins of life and evolution of the Biosphere*, **33**, pp.433-455, (2003).
- [17] R. Sánchez, L.A. Perfetti, R. Grau, E. Morgado, "A new DNA sequences vector space on a genetic code Galois field," *MATCH Commun. Math. Comput. Chem.*, vol.**54**, (2005).
- [18] Z. Dawy, P. Hanus, J. Weindl, J. Dingel and F. Morcos, "On genomic coding theory," *European Transactions on Telecommunications*, **18**, pp.873-879, (2007).
- [19] P. Hanus, B. Goebel, J. Dingel, J. Weindl, J. Zech, Z. Dawy, J. Hagenauer, J.C. Meller "Information and Communication Theory in Molecular Biology," *AdET*, March (2007).
- [20] A.S.L. Rocha, R. Palazzo Jr. e M.C. Silva-Filho. *Modelo de Sistema de Comunicações Digital para o Mecanismo de Importação de Proteínas Mitochondriais através de Códigos Corretores de Erros*, Tese de Doutorado, DT-FEEC-UNICAMP, (2010).
- [21] A.S.L. Rocha, L.C.B. Faria, R. Palazzo Jr., M.C. Silva-Filho and J.H. Kleinschmidt, US Patent: *Generation and Reproduction of DNA Sequences and Analysis of Polymorphisms and Mutations by Using Error Correcting Codes*, International Application n°. PCT/IB2010/002299 (19/08/2010), Pub.n°. WO/2011/021105 (24/02/2011).
- [22] M. Lonquety, Z. Lacroix, N. Papandreou and J. Chomilier, "SPROUTS: a database fot the evaluation of protein stability upon point mutation," *Nucleic Acids Research*, vol.**37**, pp.D374-D379, (2008).

- [23] N.M. Silva *Investigaçāo da resistencia a inseticidas na mosca-da-bicheira Cochliomyia hominivorax (Diptera; Calliphoridae)*, Tese de Doutorado, IB-UNICAMP, (2009).
- [24] A.S.L. Rocha, *Modelo Matemático para a Previsão de Recombinação Sítio-Específica do DNA*, Dissertação de Mestrado, DT-FEEC-UNICAMP, (2004).
- [25] L.C.B. Faria, *Caracterizações Topológica, Geométrica e Algébrica dos Produtos da Recombinação do DNA através dos Modelos Tangle e Frações Contínuas*, Dissertação de Mestrado, DT-FEEC-UNICAMP, (2004).
- [26] Lodish, Berk, Matsudaira, Kaiser, Krieger, Scott, Zipursky, Darnell, *Molecular Cell Biology*, Editora Artmed, 5th Edition, (2005).
- [27] Alberts, Bruce; Johnson, Alexander; Lewis, Julian; Raff, Martin; Roberts, Keith; Walter, Peter; *Molecular Biology of the Cell*, Editora Artmed, 4th Edition, (2005).
- [28] E. Freese, “The difference between spontaneous and base-analogue induced mutations of phage T4,” *Proc. of PNAS*, vol.**45**, pp.622-633, (1959).
- [29] E. Freese, “The specific mutagenic effect of base analogues on phage T4”, *Journal Molecular Biology*, vol.**1**, pp.87-105, (1959).
- [30] S.A. Sawyer, J.Parsch, Z.Zhang, D.L.Hartl, “Prevalence of positive selection among nearly neutral amino acid replacements in *Drosophila*,” *Proceedings of the National Academy of Sciences U.S.A*, vol.**104**, pp.6504-6510, (2007).
- [31] Y. Ionov, M.A. Peinado, S. Malkhosyan, D. Shibata, M. Perucho “Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis,” *Nature*, vol.**363**, pp.558-561, (1993).
- [32] A. Galvani, M.S latkin, “Evaluating plague and smallpox as historical selective pressures for the *CCR5 – Δ32* HIV-resistance allele,” *Proceedings of the National Academy of Sciences U.S.A*, vol.**100**, pp.15276-15279, (2003).
- [33] I.N. Herstein, *Topics in Algebra*, John Wiley and Sons, New York, (1975).
- [34] J.B. Fraleigh, *A First Course in Asbtract Algebra*, Addison-Welwy Publishing Co, (1982).
- [35] M.J.E. Golay, “Notes on digital conding,” *Proc. IEEE*, **37**, pp.657, (1949).
- [36] R.W. Hamming, “Error detecting and error correcting codes”, *Bell Syst. Tech. J.*, **29**, pp.147-160, (1950).

- [37] W.W.Peterson and E.J.Weldon Jr., *Error-Correcting Codes*, 2nd ed. MIT Press, (1972).
- [38] F.J. McWilliams and N.J.A. Sloane, *The Theory of Error Correcting Codes*, North-Holland Publishing Company, (1977).
- [39] A. Viterbi and J.K. Omura, *Principles of digital Communication and Coding*, New York: McGraw-Hil, (1979).
- [40] Shu Lin and D.J. Costello Jr. *Error Control Coding: Fundamentals and Applications*. Prentice-Hall, Inc., Englewood Clis, NJ, (1983).
- [41] G. Ungerboeck, “Channel coding with multilevel/phase signals,” *IEEE Trans. Inform. Theory*, vol.IT-**28**, pp.56-67, (1982).
- [42] G.D. Forney. “Geometrically uniform codes”, *IEEE Trans. Inform. Theory*, vol.IT-**37**, pp.1241-1260, September (1991).
- [43] H.A.Loeliger, “Signal sets matedched to groups,” *IEEE Trans. Inform. Theory*, vol.IT-**37**, pp.1675-1682, November 1991.
- [44] P.R. Barbosa, *Construção de Códigos \mathbb{Z}_{2^k} -pseudolineares através de Aplicações Isométricas e Extensões de Galois sobre Anéis Locais*, Dissertação de Mestrado, DT-FEEC-UNICAMP, (2000).
- [45] B.R. McDonald, *Finite Rings with Identity*, Marcel Dekker, New York, (1974).
- [46] J.C. Interlando, *Uma Contribuição à Construção e Decodificação de Códigos Lineares sobre Grupos Abelianos via Concatenação de Códigos sobre Anéis de Inteiros Residuais*, Tese de Doutorado, DT-FEEC-UNICAMP, (1994).
- [47] J.C. Interlando, R. Palazzo, Jr., J.R. Gerônimo, A.A. Andrade, O.M. Favareto, e T.P. da Nóbrega Neto, *Códigos Corretores de Erros sobre Estruturas de Corpos, Anéis e Grupos*, DT-FEEC-UNICAMP, (1998).
- [48] P. Shankar, “On BCH codes over arbitrary integer rings”, *IEEE Trans. Inform. Theory*, vol.IT-**25**, pp.480-483, July (1979).
- [49] J.R. Gerônimo, *Extensões da \mathbb{Z}_4 -linearidade via Grupos de Simetrias*, Tese de Doutorado, DT-FEEC-UNICAMP, (1997).
- [50] A.R. Hammons Jr., A.R. Calderbank, P.V. Kumar, N.J.A. Sloane and P. Solé, “The \mathbb{Z}_4 -linearity of Kerdock, Preparata, Goethals, and related codes”, *IEEE Trans. Inform. Theory*, vol.IT-**40**, p.301-319, março (1994).

- [51] F.J. McWillians and N.J.A. Sloane, *The Theory of Error Correcting Codes*, North-Holland Publishing Company, (1977).
- [52] J.C. Interlando and R. Palazzo, Jr., “Modified Berlekamp-Massey algorithm for decoding BCH codes defined over rings,” *Proceedings of the IEEE International Symposium on Inform. Theory (ISIT)* - Trondheim, Norway, **27**, pp.94, June - 1 July (1994).
- [53] R.G. Cavalcante, R. Palazzo Jr., “Performance analysis of M-PSK signal constellations in Riemannian varieties,” *AAECC'03 Proceedings of the 15th international conference on Applied algebra, algebraic algorithms and error-correcting codes*, Springer-Verlag Berlin, Heidelberg, (2003).
- [54] <http://adventista.forumbrasil.net/t1388-a-flor-do-genoma/artigoscientificos>.
- [55] <http://www1.folha.uol.com.br/colunas/marceloleite/831492-a-flor-do-genoma.shtml>.
- [56] Jan A. Witkowski, J.D. Watson, Myers, Richard M, A.A. Caudy, *Dna Recombinante - Genes e Genomas*, Editora Artmed, Quarta Edição, (2010).
- [57] P.C. Naquim, *Em Nome do DNA*, Editora LMP, Livraria Médica Paulista, (2010).
- [58] *How the CD Was Developed*, BBC News, (2007-08-17).
- [59] A.S. Tanenbaum, *Computer Networks*, Prentice-Hall, Inc., New Jersey, (1981).
- [60] R. Sá, *Sistemas e Redes de Telecomunicações*, Editora FCA, (2007).
- [61] W.F. Giozza, J.F.M. Araújo, J.A.B. Moura e J.P. Sauvé, *Redes Locais de Computadores: Tecnologia e Aplicações*, São Paulo, Editora McGraw-Hill, (1986).
- [62] W. Stallings, *Data & Computer Communications*, Prince Hall, Upper Saddle River, New Jersey, Sixth Edition, (2000).
- [63] W. Tavares, “Resistência Bacteriana”, *Manual de antibióticos e quimioterápicos Antiinfecciosos*, Rio de Janeiro, Atheneu, cap.5, p.50-109, (1990).
- [64] A.S.L. Rocha, L C.B. Faria, J.H. Kleinschmidt, R. Palazzo Jr. and M.C. Silva-Filho, “DNA Sequences Generated by \mathbb{Z}_4 -linear Codes”, in *Procedings of the IEEE International Symposium o Inform. Theory (ISIT)* - Austin, Texas, U.S.A., vol.1, pp-98, June 13-18, (2010).

- [65] L.C.B. Faria, A.S.L. Rocha, J.H. Kleinschmidt, M.C. Silva-Filho and R. Palazzo Jr., “Analyses of amino acid substitutions by use of error-correcting codes,” *6th International Conference of the Brazilian Association for Bioinformatics and Computacional Biology* (X-meeting) - Ouro Preto, November 15 to 18 (2010).
- [66] L.C.B. Faria, A.S.L. Rocha, J.H. Kleinschmidt, R. Palazzo Jr. and M.C. Silva-Filho, “DNA sequences generated by BCH codes over $GF(4)$,” *IEE Electronics Letters*, Vol.**46**, nº 03, pp-202-203, 4th February (2010).