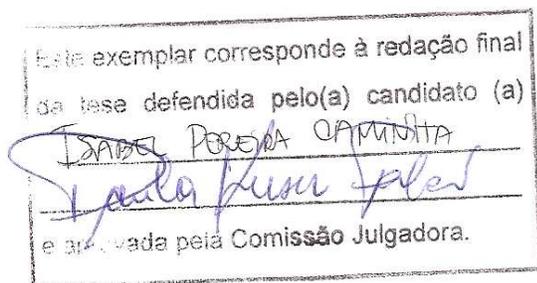


**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA**



ISABEL PEREIRA CAMINHA

**ESTUDOS ESTRUTURAIS DAS ENZIMAS ENVOLVIDAS COM O MECANISMO DE
PRODUÇÃO DE ÁCIDOS ORGÂNICOS DA BACTÉRIA *GLUCONACETOBACTER
DIAZOTROPHICUS* UTILIZANDO MÉTODOS COMPUTACIONAIS**



Dissertação apresentada ao Instituto de Biologia para obtenção do Título de Mestre em Genética e Biologia Molecular na área de Bioinformática

Orientador: Prof^ª. Dr^ª. Paula Regina Kuser Falcão

Campinas, 2011

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE BIOLOGIA – UNICAMP

C146e

Caminha, Isabel Pereira

Estudos estruturais das enzimas envolvidas com o mecanismo de produção de ácidos orgânicos da bactéria *Gluconacetobacter diazotrophicus* utilizando métodos computacionais / Isabel Pereira Caminha. – Campinas, SP: [s.n.], 2011.

Orientadora: Paula Regina Kuser Falcão.
Dissertação (mestrado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Modelagem molecular. 2. Enzimas. 3. *Gluconacetobacter diazotrophicus*. 4. Ácidos orgânicos.
I. Falcão, Paula Regina Kuser. II. Universidade Estadual de Campinas. Instituto de Biologia. III. Título.

(rcdt/ib)

Título em inglês: Structural studies on the enzymes related to the mechanism of organic acid production in the *Gluconacetobacter diazotrophicus* bacteria using computational methods.

Palavras-chave em inglês: Molecular modelling; Enzymes; *Gluconacetobacter diazotrophicus*; Organic acids.

Área de concentração: Bioinformática.

Titulação: Mestre em Genética e Biologia Molecular.

Banca examinadora: Paula Regina Kuser Falcão, Marcelo Mendes Brandão, Michel Eduardo Beleza Yamagishi.

Data da defesa: 23/02/2011.

Programa de Pós-Graduação: Genética e Biologia Molecular.

Campinas, 23 de fevereiro de 2011

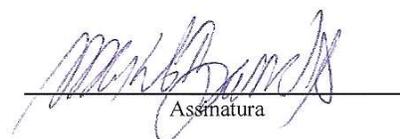
BANCA EXAMINADORA

Prof(a). Dr(a). Paula Regina Kuser Falcão (Orientador(a))



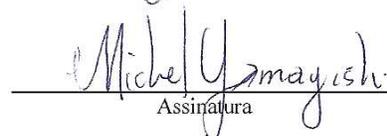
Assinatura

Prof. Dr. Marcelo Mendes Brandão



Assinatura

Prof(a). Dr(a). Michel Eduardo Beleza Yamagishi



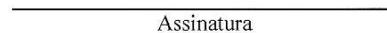
Assinatura

Prof(a). Dr(a). Poliana Fernanda Giachetto



Assinatura

Prof(a). Dr(a). Renato Vicentin



Assinatura

***Bela, seu sorriso é a força que não me
deixa desistir, você é a inspiração que
me incentiva a ser uma pessoa melhor.***

***Pai, pelos longos meses de estudo que me
proporcionaram as melhores discussões
a respeito deste trabalho, e por ter
possibilitado essa mudança tão
significante em minha vida.***

***Mãe, pelo apoio nos momentos em
que eu mais precisei, e pela confiança
depositada em mim hoje e sempre.***

***Rodrigo, por me levantar todas as vezes
em que eu caí, e sempre ter acreditado
que eu conseguiria.***

AGRADECIMENTOS

À minha filha Isabela, por estar comigo todas as noites e me fazer acreditar que todo nosso esforço está sendo recompensado.

À minha querida família, minha inspiração, essencial para que eu me tornasse o que sou hoje.

Ao meu marido Rodrigo, pessoa da qual sinto orgulho todos os dias, e que me incentiva a sempre buscar o meu melhor.

À minha orientadora Paula, que me acolheu como uma filha no momento mais difícil, e que me ofereceu todas as oportunidades para crescer como profissional.

À melhor amiga do mundo, Bruna, que deixa meus dias mais felizes e sempre está com o ombro disponível para uma tristeza básica.

Aos colegas do LBA, em especial ao Fábio, que me deu suporte todas as infinitas vezes em que eu pedi.

Ao amigo Roberto Herai, que me possibilitou chegar até aqui, e me ajudou com todos os problemas acadêmicos possíveis.

Aos amigos de Campinas e Itajubá, que tornam nossas vidas mais divertidas.

À empresa Embrapa Informática Agropecuária, pela infra-estrutura fornecida.

Ao CNPq, pelo suporte financeiro.

RESUMO

A capacidade de *Gluconacetobacter diazotrophicus* produzir ácidos orgânicos representa uma das suas potenciais aplicações biotecnológicas. Entretanto, a via metabólica do gluconato é bastante complexa e, muitas vezes, inviável economicamente. Este estudo tem como objetivo buscar uma maneira de otimizar a síntese do produto de interesse. A opção sugerida é alterar as enzimas envolvidas na conversão de gluconato em cetogluconatos, visando aumentar a concentração de ácido glicônico. A mutagênese dirigida é uma técnica utilizada neste tipo de estudo, entretanto, realizar e avaliar diversos tipos de mutações exige grande demanda de tempo e recursos financeiros. A biologia computacional pode ajudar a encurtar o caminho dos processos experimentais, por isto foi utilizada neste estudo. Após uma busca no genoma de *G. diazotrophicus*, foram selecionadas oito ORFs relacionadas com a produção de ácido glicônico. Como nenhuma delas possui uma estrutura resolvida experimentalmente, foi utilizada uma técnica de predição, a modelagem por homologia, com o objetivo de inferir a respeito do funcionamento de cada proteína. Uma estratégia de substituição de resíduos selecionados por alaninas aliada à modelagem molecular foi utilizada para avaliar o impacto dessa mutação. Um estudo da via metabólica do gluconato em diversos organismos forneceu dados suficientes para eleger os dois melhores alvos para a mutagênese, uma gluconato 5-desidrogenase e uma putativa gluconato quinase, ambas relacionadas com a conversão de ácido glicônico em cetogluconatos. A partir de um estudo do mecanismo catalítico das enzimas, foi possível selecionar os possíveis resíduos críticos para a atividade enzimática. Estes foram substituídos, *in silico*, por alaninas, e as consequências das alterações foram analisadas. As mutações que acusaram um impacto no funcionamento da enzima foram Arginina 106 e Serina 152 na gluconato 5-desidrogenase, e Histidina 133 na gluconato quinase. Espera-se que, ao substituímos estes resíduos, o potencial catalítico seja minimizado ou até mesmo inativado, diminuindo a quantidade de gluconato convertido em outros produtos. Estes resultados serão testados experimentalmente por colaboradores.

Palavras-chave: *Gluconacetobacter diazotrophicus*, ácido glicônico, modelagem molecular, mutagênese sítio-dirigida.

ABSTRACT

The ability of *Gluconacetobacter diazotrophicus* to produce organic acids represents one of its potential biotechnological applications. However, the metabolic pathway of gluconate is quite complex and often impracticable economically. This study aims to find a way to optimize the synthesis product of interest. The suggested option is to alter the enzymes involved in the conversion of gluconate into cetogluconates, to increase the concentration of gluconic acid. Site directed mutagenesis is a technique used in this type of study, however, implement and evaluate various types of mutations requires a great deal of time and financial resources. The computational biology can help shorten the path of experimental processes, so it was used in this study. After a search in the genome of *G. diazotrophicus*, eight ORFs were selected related to the production of gluconic acid. As there is little physical or structural information available for this enzyme, we used a technique of prediction, homology modeling, in order to infer about the function of each protein. A molecular modeling approach with an amino acid replacement by alanines strategy was employed to assess the impact of this mutation. A study of the metabolic pathway of gluconate in several organisms has provided sufficient data to choose the two best targets for mutagenesis, a gluconate 5-dehydrogenase and a putative gluconate kinase, both related to the conversion of gluconic acid in cetogluconates. From a study of the catalytic mechanism of enzymes, it was possible to select any residues critical for enzymatic activity. These were replaced *in silico* by alanines, and the consequences of changes were analyzed. Detailed analysis of the modeled structures reveals that the mutations that may cause an impact on the functioning of the enzyme were Arginine 106 and Serine 152 in gluconate 5-dehydrogenase, and Histidine 133 in gluconate kinase. It is expected that, in replacing these residues, the catalytic potential is minimized or even inactivated, reducing the amount of gluconate converted into other products. These results will be experimentally tested by collaborators, and once proved right our findings may be applicable in the use of this bacteria in biotechnological processes.

Keywords: *Gluconacetobacter diazotrophicus*, gluconic acid, molecular modeling, site directed mutagenesis.

LISTA DE FIGURAS

Figura 1.1: Via metabólica geral do Gluconato	05
Figura 1.2: Oxidação da glicose por <i>Aspergillus niger</i>	06
Figura 1.3: Via metabólica para oxidação da glicose por <i>Gluconobacter</i>	07
Figura 1.4: Esquema das vias metabólicas envolvidas na biossíntese de ácido glicônico e ceto-derivados por <i>G. oxydans</i>	08
Figura 1.5: Aplicações dos modelos construídos por meio de modelagem	13
Figura 3.1: Pipeline referente à metodologia utilizada	19
Figura 3.2: Crescimento do banco de dados Genbank	20
Figura 3.3: Crescimento do banco de dados de estruturas de proteínas	21
Figura 3.4: Etapas da modelagem comparativa	24
Figura 3.5: Gráfico de Ramachandran	29
Figura 3.6: Representação dos ângulos phi e psi de um aminoácido	29
Figura 3.7: Predição de estruturas secundárias do modelo construído	30
Figura 3.8: Etapas do processo de dinâmica molecular	32
Figura 3.9: Interações frequentes em complexos proteína-ligante	33
Figura 4.1: Via de metabolismo de cetogluconatos em <i>E. coli</i>	36
Figura 4.2: Alinhamento das sequências da ORF GDI0691 e 3CXR	38
Figura 4.3: Representação esquemática da relação proteína-ligante (3CXT) .	39
Figura 4.4: Alinhamento estrutural do molde/modelo construído (GDI0691) ...	41
Figura 4.5: Gráfico de Ramachandran do modelo construído/ORF GDI0691 .	42

Figura 4.6: Predição de estruturas secundárias do modelo/ORF GDI0691.....	43
Figura 4.7: Predição de estruturas secundárias do molde 3CXR	43
Figura 4.8: Via da biossíntese do corismato em organismos diversos	46
Figura 4.9: Via de metabolismo de cetogluconatos em <i>E. coli</i>	47
Figura 4.10: Alinhamento das sequências da ORF GDI0293/ 1KNQ	49
Figura 4.11: Representação da relação proteína-ligante do PDB 1KO5	49
Figura 4.12: Representação da relação proteína-ligante do PDB 1KO8	50
Figura 4.13: Representação da relação proteína-ligante do PDB 1KOF	50
Figura 4.14: Alinhamento estrutural do molde/modelo (GDI0293)	52
Figura 4.15: Gráfico de Ramachandran do modelo construído/ORF GDI0293	53
Figura 4.16: Predição de estruturas secundárias do modelo/ORF GDI0293 ...	54
Figura 4.17: Predição de estruturas secundárias do molde 1KNQ	54
Figura 5.1: Sugestão da via metabólica de ácido glicônico e cetoderivados em <i>G. diazotrophicus</i>	57
Figura 5.2: Objetivo do estudo	57
Figura 5.3: Representação da interação modelo-ligante (GDI 0691)	59
Figura 5.4: Sobreposição da Ga5DH antes e após a substituição da Arginina 106 por uma alanina	61
Figura 5.5: Representação esquemática da interação do modelo com os ligantes após a substituição da Arginina 106 por uma alanina	62
Figura 5.6: Sobreposição da Ga5DH antes e após a substituição da Serina 152 por uma alanina	63

Figura 5.7: Representação esquemática da interação do modelo com os ligantes após a substituição da Serina 152 por uma alanina	64
Figura 5.8: Representação da interação modelo-ligante (ATP)	65
Figura 5.9: Representação da interação modelo-ligante (6PG)	66
Figura 5.10: Sobreposição da Gluconato quinase antes e após a substituição da Histidina 133 por uma alanina	68
Figura 5.11: Representação esquemática da interação do modelo com o ligante ATP após a substituição da Histidina 133 por uma alanina	68
Figura 5.12: Representação esquemática da interação do modelo com o ligante 6PG após a substituição da Histidina 133 por uma alanina	69

LISTA DE TABELAS

Tabela 4.1: ORFs de <i>G. diazotrophicus</i> possivelmente relacionadas com a produção de Ácidos Orgânicos	35
Tabela 4.2: Características analisadas para escolha do molde adequado da ORF GDI0691	37
Tabela 4.3: Resíduos que interagem com os ligantes no PDB 3CXT	40
Tabela 4.4: Características analisadas para escolha do molde adequado da ORF GDI0293	48
Tabela 4.5: Resíduos que interagem com os ligantes no PDB 3CXT	51
Tabela 5.1: Resíduos que interagem com os ligantes na estrutura 3CXT	58
Tabela 5.2: Comparação das interações dos ligantes com o molde e o modelo (ORF GDI 0691)	60
Tabela 5.3: Resíduos que interagem com os ligantes em 1KO5 e 1KO8	65
Tabela 5.4: Comparação das interações dos ligantes com o molde e o modelo (ORF GDI0293)	67

LISTA DE SIGLAS

2,5 KGA – Ácido 2,5 dicetoglicônico

2KGA – Ácido 2-cetoglicônico

5KGA – Ácido 5-ceto-D-gluconato

ACP – Ácido Fosfo Metilfosfônico

ADP – Adenosina Difosfato

ATP – Adenosina Trifosfato

BLAST – Basic Local Alignment Search Tool

CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico

EMBRAPA – Empresa Brasileira de Pesquisa Agropecuária

FAD – Flavina Adenina Dinucleotídeo

FAPERJ – Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro

FBN – Fixação Biológica de Nitrogênio

Ga5DH – Gluconato 5-desidrogenase

GADH – Gluconato desidrogenase

GCO – Gluconato

GDH – Glicose desidrogenase

GKR – Glucarato

GNO – Gluconato 5-oxidoredutase

Gntk – Gluconato Quinase

GP6 – Gluconato 6-fosfato

KGDH – α -cetogluturato desidrogenase

NAD – Nicotinamida Adenina Dinucleotídeo

NADP – Nicotinamida Adenina Dinucleotídeo Fosfato

NK – Nucleotídeo/Nucleosídeo Quinases

NMP – Nucleosídeo Monofosfato

ORF – Open Reading Frame

PDB – Protein Data Bank

PQQ – Pirroloquinolina quinona

PUC-Rio – Pontifícia Universidade Católica do Rio de Janeiro

RMN – Ressonância Magnética Nuclear

SDR – Desidrogenases/Redutases de cadeia curta

UENF – Universidade Estadual do Norte Fluminense Darcy Ribeiro

UERJ – Universidade do Estado do Rio de Janeiro

UFRJ – Universidade Federal do Rio de Janeiro

UFRRJ – Universidade Federal Rural do Rio de Janeiro

VMD – Visual Molecular Dynamics

ÍNDICE

RESUMO	vi
ABSTRACT	vii
LISTA DE FIGURAS.....	viii
LISTA DE TABELAS.....	xi
LISTA DE SIGLAS.....	xii
CAPÍTULO 1 - INTRODUÇÃO	1
1.1. <i>Gluconacetobacter diazotrophicus</i>	1
1.2. Ácido Glicônico	3
1.2.1. Produção e Vias Metabólicas	4
1.3. Bioinformática estrutural aplicada ao estudo de mutações	9
1.4. Aplicações da Modelagem Molecular Comparativa	12
1.5. Possíveis erros da modelagem molecular comparativa	14
CAPÍTULO 2 – OBJETIVOS	16
2.1. Justificativa	16
2.2. Objetivos Específicos.....	17
2.3. Atividades	17
CAPÍTULO 3 - MATERIAIS E MÉTODOS	19
3.1. Modelagem Molecular	19
3.1.1. Busca por sequências relacionadas à sequência alvo	24
3.1.2. Definição de moldes.....	25
3.1.3. Alinhamento de sequências	26
3.1.4. Construção dos modelos.....	27
3.1.5. Validação do modelo.....	28
3.2. Dinâmica Molecular	31
3.3. Interações proteína-ligante.....	32
3.4. Mutagênese sítio-dirigida	34
CAPÍTULO 4 - DESENVOLVIMENTO	35
4.1. ORF GDI0691	36
4.1.1. Procura e seleção do molde	37

4.1.2. Alinhamento de sequências e interação com os ligantes	38
4.1.3. Modelagem comparativa e dinâmica molecular.....	41
4.1.4. Validação do modelo.....	42
4.1.5. Mecanismo catalítico	44
4.2. ORF GDIO293	45
4.2.1. Procura e seleção do molde	46
4.2.2. Alinhamento de sequências e interação com os ligantes	48
4.2.3. Modelagem comparativa e dinâmica molecular.....	52
4.2.4. Validação do modelo.....	52
4.2.5. Mecanismo catalítico	55
CAPÍTULO 5 - RESULTADOS	56
5.1. Análise da via metabólica.....	56
5.2. Mutagênese sítio-dirigida	58
5.2.1. ORF GDI0691	58
5.2.2. ORF GDIO293	64
CAPÍTULO 6 - CONCLUSÕES E DESENVOLVIMENTOS FUTUROS	70
REFERÊNCIAS BIBLIOGRÁFICAS	73

CAPÍTULO 1 - INTRODUÇÃO

1.1. *Gluconacetobacter diazotrophicus*

As bactérias ácido-acéticas são aeróbicas obrigatórias capazes de oxidar diversos açúcares e polióis, como D-glicose, glicerol, D-sorbitol e etanol. Este tipo de reação é denominado fermentação oxidativa, uma vez que envolve oxidações incompletas acompanhadas pelo acúmulo dos produtos da oxidação no meio de cultura (Matsushita *et al.*, 2003).

Gluconacetobacter diazotrophicus é uma bactéria diazotrófica associativa, endofítica, isolada primeiramente em cana de açúcar (*Saccharum spp.*) (Gillis, 1989) e posteriormente em outras plantas (James & Olivares, 1998). Além da cana de açúcar, pode associar-se a capim-elefante (*Pennisetum purpureum*), batata doce (*Ipomoea batatas*), abacaxi (*Ananas comosus*) e café (*Coffea arabica*), demonstrando potencial para beneficiar diversas plantas hospedeiras (Reis *et al.*, 1994; Dobereiner *et al.*, 1988; Tapia-Hernandez *et al.*, 2000; Weber *et al.*, 1999)

A simbiose entre a bactéria em questão e a cana de açúcar possui grande relevância devido à importância econômica da colheita. O Brasil é responsável por uma grande parte da cultura de cana no mundo, sendo portanto um dos maiores produtores de açúcar (Zanin *et al.*, 2000; Orellana & Neto, 2006). Além disso, há no país um programa de incentivo à produção de combustíveis alternativos, visando a preservação do meio ambiente.

G. diazotrophicus exerce um importante papel na agricultura, pois assimila nitrogênio para as plantas infectadas por meio da Fixação Biológica de Nitrogênio (FBN), utilizando a glicose como fonte de carbono (Luna, 2002). Embora muitos organismos sejam capazes de contribuir com a FBN, esta bactéria destaca-se por fixar N_2 sem causar doenças (James *et al.*, 1994). Atualmente, mais de metade do nitrogênio necessário para o crescimento das plantas resulta deste processo (Saravanan *et al.*, 2008). Isto traz como consequência uma diminuição no número de fertilizantes utilizados, diminuindo a poluição proveniente da agricultura. O uso de bactérias

promotoras de crescimento de plantas como biofertilizantes está sendo considerado como meio alternativo ou suplementar de reduzir o uso de produtos químicos nas plantações. O estímulo do crescimento por microorganismos pode ser uma consequência da fixação de nitrogênio, produção de fitormônios, biocontrole de fitopatógenos na raiz ou aumento na disponibilidade de minerais (Luna, 2010).

Muitas propriedades de *G. diazotrophicus* exaltam sua importância em várias áreas da biotecnologia. Um exemplo é a capacidade deste organismo em suportar altas concentrações de açúcar, característica que demonstra sua adaptação às plantações de cana de açúcar, onde o conteúdo de sucrose é frequentemente alto. Além disso, possui tolerância a baixo pH e ácidos orgânicos, sendo capaz de fixar nitrogênio a pHs próximos de 2.5 (Tejera *et al.*, 2003).

Esta bactéria também produz hormônios relacionados ao crescimento de plantas, como auxinas e giberelinas, como demonstrado em vários experimentos (Sevilla *et al.*, 2001). Além disso, tem sido relatado um possível papel de defesa, devido a sua atividade antagonista contra *Xanthomonas albilineans* e *Colletotrichum falcatum* por meio da produção de bacteriocina (Blanco *et al.*, 2005) e de sua habilidade de fermentar açúcares em meio com pH abaixo de 3.0 (Muthukumarasamy, 2000), respectivamente.

Podemos notar, portanto, que o organismo em questão possui múltiplas qualidades que podem ser aplicadas na indústria biotecnológica. Finalmente, não podemos deixar de ressaltar sua capacidade em produzir ácidos orgânicos por meio da fermentação oxidativa de diversas fontes de carbono. A biossíntese desses ácidos envolve vias um tanto complexas, muitas vezes até inviáveis economicamente (Magnuson & Lasure, 2004). A peculiaridade desta espécie em crescer a condições extremas poderia auxiliar no desenvolvimento de um processo mais estável e econômico de biossíntese de ácido glicônico.

O sequenciamento do genoma da *Gluconacetobacter diazotrophicus* foi finalizado em 2007, pelo projeto RioGene (Programa Genoma do Estado do Rio de Janeiro). Esse projeto reuniu laboratórios fluminenses de diversas instituições, entre elas a Universidade Federal do Rio de Janeiro (UFRJ), Universidade do Estado do Rio

de Janeiro (UERJ), Universidade Estadual do Norte Fluminense Darcy Ribeiro (UENF), Universidade Federal Rural do Rio de Janeiro (UFRRJ), Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), além da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) Agrobiologia, recebendo o apoio da Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro (FAPERJ) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Seus resultados relataram um genoma composto por um cromossomo de 3,9 Mb e dois plasmídios com 16,6 e 38,8 Kb. (Bertalan *et al.*, 2009).

Os dados do genoma representam uma importante fonte de informação que pode ser usada para estudar mais profundamente a importância biotecnológica desta bactéria, visando compreender melhor suas vias metabólicas e, posteriormente, utilizá-la em outras plantas para otimizar a síntese de produtos de interesse.

1.2. Ácido Glicônico

O ácido glicônico é um ácido orgânico simples, não corrosivo, volátil nem tóxico, que foi descoberto em meados de 1870. Dez anos após, foram encontradas pela primeira vez bactérias ácido acéticas, capazes de oxidar diversos tipos de açúcares (Rohr *et al.*, 1983). Desde então, a produção deste ácido foi demonstrada em diversas espécies de bactérias, como *Pseudomonas*, *Gluconobacter* e *Acetobacter*, além de vários tipos de fungos. Também pode ser encontrado em plantas, frutas, arroz, carnes e outros produtos alimentícios (Ramachandran *et al.*, 2006).

A produção de ácido glicônico e seus cetoderivados possui importante papel na indústria alimentícia, pois é muito utilizado para conservar, fermentar e conferir sabor aos alimentos. São de grande importância econômica também para as indústrias farmacêuticas, como fonte de cálcio, nutrição animal, precursor da vitamina C, na produção de agentes antimicrobianos, além de ser útil na indústria têxtil e em outras inúmeras aplicações (Das & Kundu, 1987; Milson & Meers, 1985; Elfari *et al.*, 2005).

O ácido glicônico pode ser convertido em outros ácidos, denominados cetogluconatos. Um exemplo é o ácido 5-ceto-D-Gluconato (5KGA), que pode produzir

ácido xilárico, diversos condimentos e ser efetivamente convertido em ácido tartárico (Matzerath et *al.*, 1995). Este último é de interesse considerável, pois pode ser utilizado como antioxidante na indústria alimentícia, como um agente redutor de acidez na indústria têxtil e como um reagente quiral em síntese orgânica. Os métodos para preparação de 5KGA são baseados na fermentação da glicose. Esses processos levam à co-produção do ácido isomérico 2-cetoglicônico (2KGA) e posteriormente também ácido 2,5 dicetoglicônico (2,5 DKGA) (Shinagawa et *al.*, 1983; Weenk et *al.*, 1984), precursores de ácido ascórbico (Salusjarvi et *al.*, 2004). É importante ressaltar que a maioria das bactérias ácido acéticas produz tanto 2KGA quanto 5KGA via ácido glicônico, e os acumula simultaneamente no meio. As taxas de produção variam de acordo com as condições do meio e características de cada organismo.

1.2.1. Produção e Vias Metabólicas

O ácido glicônico é produzido a partir da glicose por meio de uma simples reação de desidrogenação catalisada por uma glicose oxidase. A oxidação da glicose dá origem à uma glucona- δ -lactona que, espontaneamente ou por meio de uma lactonase, é convertida a ácido glicônico. O ciclo do Gluconato encontra-se detalhado na Figura 1.1.

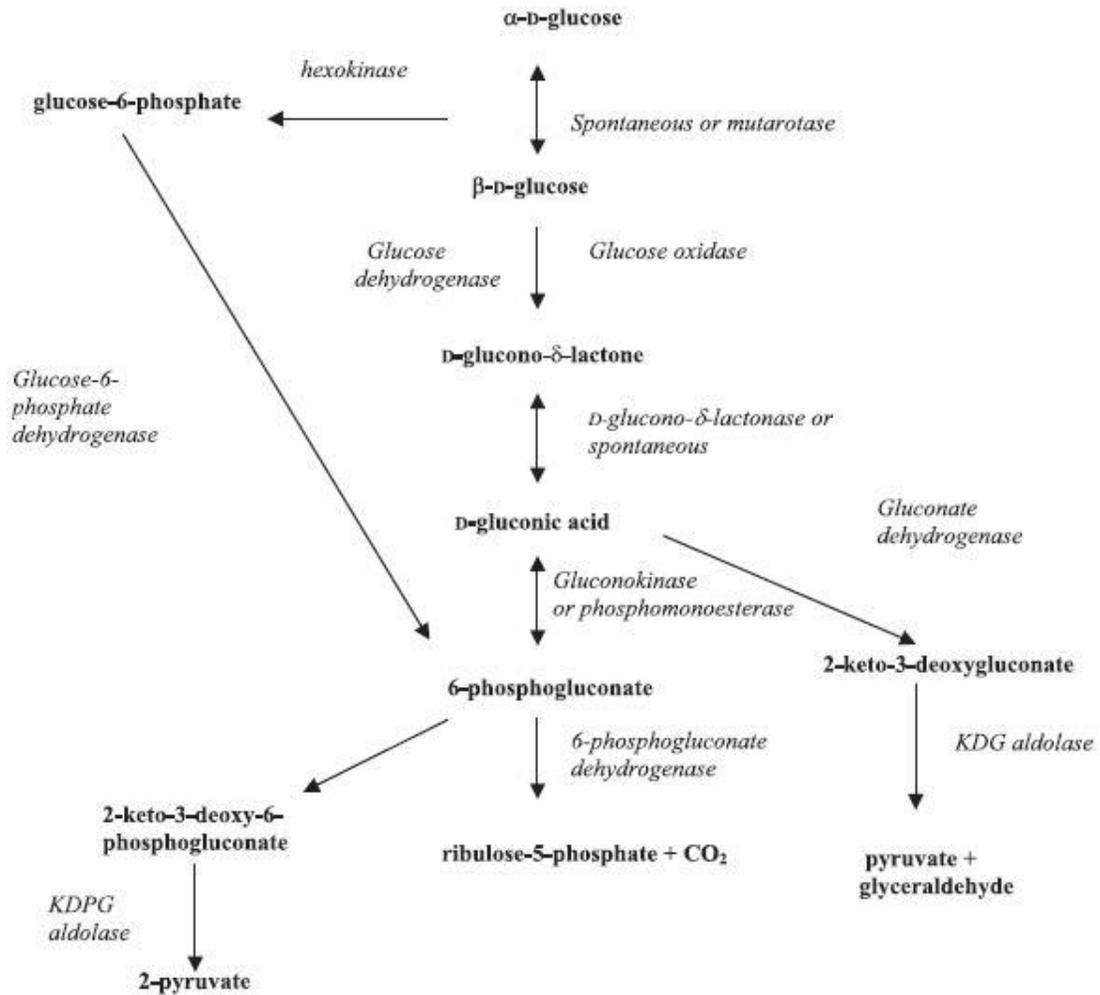


Figura 1.1. Via metabólica geral do Gluconato (Ramachandran *et al.*, 2006)

Embora existam várias abordagens diferentes para a produção de ácido glicônico (químicas, eletroquímicas, etc.) e vários agentes de oxidação disponíveis, estes parecem ser mais custosos e menos eficientes do que o processo de fermentação. A produção por meio de enzimas é vantajosa por não exigir processos de purificação (Godjevargova *et al.*, 2004). Portanto, esta última tem sido a técnica mais utilizada na manufatura destes ácidos. Processos utilizando bactérias como *Gluconacetobacter oxydans* têm ganhado cada vez mais importância, embora o método com o fungo *Aspergillus niger* seja ainda um dos mais amplamente empregados.

A reação envolvendo a conversão de glicose a ácido glicônico por fungos filamentosos é catalisada pela enzima glicose oxidase (E.C. 1.1.3.4). Ela catalisa a reação onde a glicose é desidratada a glucono- δ -lactona, enquanto o hidrogênio é transferido para seu co-fator, FAD. A enzima é induzida na presença de altos níveis de glicose e oxigênio no meio e pH próximo a 5.5 (Liu *et al.*, 2001). A conversão de glucono- δ -lactona a ácido glicônico pode ocorrer espontaneamente ou por meio de uma lactonase. Um esquema da oxidação da glicose pelo fungo *Aspergillus niger* é apresentado na Figura 1.2.

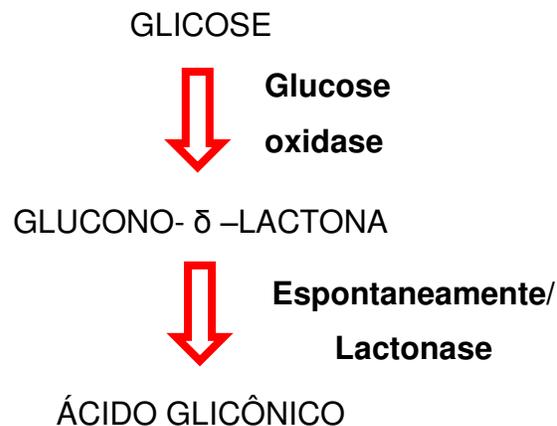


Figura 1.2. Oxidação da glicose por *Aspergillus niger*

A produção de ácidos orgânicos em bactérias ácido acéticas é realizada por uma glicose desidrogenase (GDH, E.C. 1.1.99.17), que oxida glicose a ácido glicônico. Este, por sua vez, é oxidado por uma gluconato desidrogenase (GADH), transformando-se em 2-cetogluconato. O último passo da oxidação, ou seja, a oxidação para 2,5 dicetoglicônico, é realizado por meio de uma α -cetoglutarato desidrogenase (KGDH). Os passos da reação em questão são demonstrados na Figura 1.3. Todas estas enzimas estão situadas na membrana dos organismos e são induzidas pela alta concentração de glicose no meio (Velizarov & Beschkov, 1994).

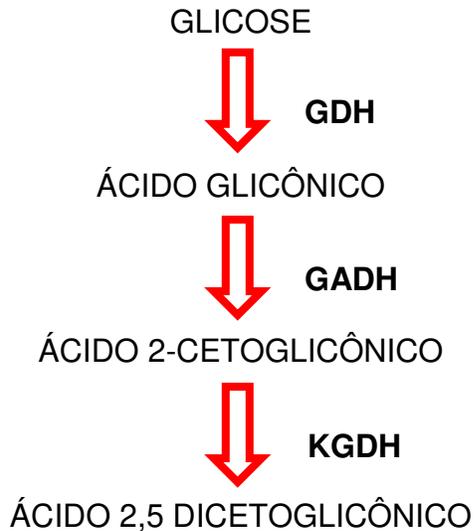


Figura 1.3. Via metabólica para oxidação da glicose por *Gluconobacter* (Ramachandran *et al.*, 2006)

GDH é uma proteína extracelular que possui Pirroloquinolina quinona (PQQ) como coenzima. Existe também uma enzima intracelular, dependente de NADP^+ , chamada glicose desidrogenase, envolvida na produção de ácidos orgânicos. Estes são exportados para a célula e catabolizados por meio de reações na via pentose fosfato, que é uma das vias essenciais do metabolismo de muitos organismos. Quando a concentração de glicose é maior do que 15 mM, esta via é reprimida, dando lugar ao acúmulo de ácido glicônico.

Gluconacetobacter oxydans é uma bactéria aeróbica obrigatória que oxida glicose por meio de duas vias metabólicas alternativas. A primeira requer uma fosforilação inicial seguida pela oxidação via pentose fosfato. A segunda é a via de oxidação direta da glicose, que resulta na formação dos ácidos glicônico e dicetoglicônico (Stadler-Szoke, 1980). Este organismo converte D-glicose em ácido 2,5 dicetoglicônico pela ação das três enzimas descritas na Figura 1.3.

Existem dois sistemas enzimáticos separados em *Gluconacetobacter oxydans* para redução de glicose a 5KGA: o primeiro oxida glicose a glucono- δ -lactona por meio de uma glicose desidrogenase situada na membrana. Este produto intermediário é convertido a ácido glicônico por meio de uma lactonase. Posteriormente, a oxidação do

ácido glicônico é realizada pelas enzimas Gluconato desidrogenase e α -cetoglutarato desidrogenase, resultando na formação de 2-KGA e 2,5-DKGA (Shinagawa *et al.*, 1984). A proteína de membrana Gluconato-oxidorreductase converte ácido glicônico a 5KGA (Matsushita *et al.*, 2003). O segundo sistema enzimático está localizado no citoplasma, onde a glicose desidrogenase solúvel (dependente de NADP^+) catalisa a formação de glucono- δ -lactona, que é convertida a ácido glicônico por uma lactonase. A Gluconato: NADP -5 oxidorreductase citoplasmática converte ácido glicônico a 5KGA. Uma representação esquemática destas reações é apresentada na Figura 1.4.

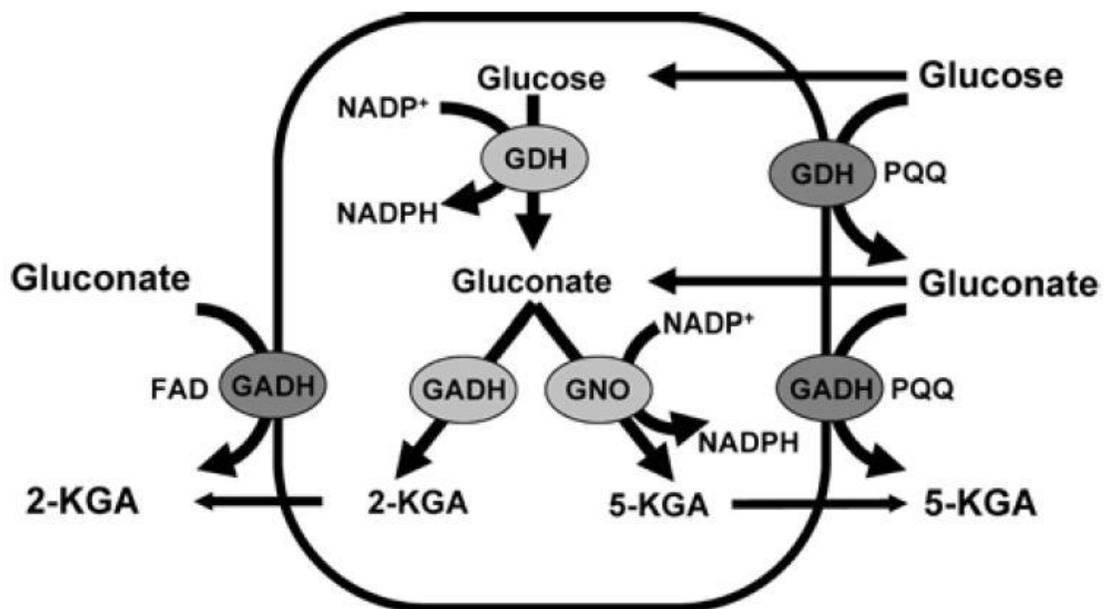


Figura 1.4: Esquema das vias metabólicas envolvidas na biossíntese de ácido glicônico e cetoderivados por *G. oxydans* (Herrmann *et al.*, 2004; Elfari *et al.*, 2005)

Legenda: 2-KGA: ácido 2-cetoglicônico; 5-KGA: ácido 5-cetoglicônico; GADH: Gluconato desidrogenase; GDH: Glicose desidrogenase; GNO: Gluconato 5-oxidorreductase; PQQ: Pirroquinolina quinona; FAD: Flavina Adenina Dinucleotídeo; NADP^+ : Nicotinamida adenina dinucleotídeo fosfato oxidado; NADPH: Nicotinamida adenina dinucleotídeo fosfato reduzida.

Como a reação leva à formação de um produto ácido, é necessária a utilização de agentes neutralizantes. Caso contrário, a acidez irá inativar as enzimas, resultando na interrupção da produção. Analisando este fato, podemos sugerir que a tolerância de

Gluconacetobacter diazotrophicus a ambientes ácidos seja uma característica importante a ser considerada na elaboração de novos processos de produção.

A bactéria ácido acética *Gluconacetobacter diazotrophicus*, exibiu altas taxas de produção de ácido glicônico. A oxidação da glicose era menos sensível a baixos valores de pH do que em *oxydans*. A produção de ácido glicônico está diretamente associada com o início do crescimento exponencial. Ao longo deste processo, um metabolismo aeróbico intenso ocorre, com alto consumo de glicose e uma redução de pH concomitante (Intorne *et al.*, 2009). Ácido glicônico, então, representa uma importante fonte de carbono para a bactéria (Stephan *et al.*, 1991). Além de uma glicose desidrogenase dependente de NAD^+ , *diazotrophicus* contém uma glicose desidrogenase dependente de PQQ, que era inicialmente responsável pela produção de ácido glicônico. A produção de ácido glicônico por bactérias possui sucesso limitado em escala industrial, já que a oxidação prossegue com as reações secundárias, levando a ácidos oxoglicônicos (Ramachandran, 2006).

Embora a produção de ácido glicônico seja um simples processo de oxidação que pode ser realizado por meio de métodos eletroquímicos ou bioquímicos, a produção através de processos de fermentação envolvendo fungos e bactérias está bem estabelecida comercialmente. Progressos consideráveis têm sido feitos no entendimento do mecanismo do processo de fermentação por meio de diferentes microorganismos, e um processo de produção altamente eficiente foi desenvolvido ao longo dos anos (Ramachandran, 2006). Entretanto, o desenvolvimento de um processo mais novo e econômico para a conversão de glicose a ácido glicônico seria promissor.

1.3. Bioinformática estrutural aplicada ao estudo de mutações

As proteínas são macromoléculas envolvidas na maioria dos processos biológicos. Seu papel em uma determinada via metabólica constitui uma das informações mais importantes que pode ser obtida nesta era da biologia de sistemas. Tradicionalmente, a função de uma proteína é estudada por meio de investigações diretas na proteína nativa. Uma abordagem diferente de obter estas informações é

manipular racionalmente uma proteína, para se obter novas propriedades funcionais e estruturais, importantes na investigação de interações biológicas. Esta técnica pode ser denominada engenharia de proteínas, e tem como objetivo gerar moléculas desconhecidas, muitas vezes com base em outras anteriormente estudadas (Alberghina, 2005).

As enzimas são importantes alvos na engenharia de proteínas, e muitas delas já foram aplicadas em processos industriais. Métodos de desenho racional, mutagênese aleatória e evolução molecular tem sido aplicados na intenção de controlar a especificidade enzimática (Shao & Arnold, 1996; Vita, 1997). Uma abordagem comum é utilizar a base de uma proteína conhecida para introduzir novas propriedades à outra. Estudos demonstraram que uma única alteração de aminoácido é suficiente para, por exemplo, converter uma protease papaína a uma nitrila hidratase (Dufour *et al.*, 1995).

Outra alternativa é construir proteínas híbridas nas quais diferentes módulos catalíticos e/ou regulatórios são incorporados (Nixon *et al.*, 1998). Sítios de ligação para metais muitas vezes são úteis para regular a atividade enzimática. Por exemplo, a introdução de um sítio de ligação de Cu^{2+} é suficiente para permitir a regulação da atividade da tripsina (Halfon & Craik, 1996). Também existem estudos onde houve alteração do cofator utilizado e que comprovam a eficiência do método na mudança da especificidade do substrato (Alberghina, 2005).

Na maioria das vezes, pouquíssimos resíduos são essenciais para uma interação (Bromberg & Burkhard, 2008), e o grande número de aminoácidos existentes no interior de uma proteína dificulta a análise de sua função. Substituir cada resíduo de uma proteína pelos outros 19 e estudar as variantes resultantes é impossível. Se os dados da estrutura tridimensional da proteína estiverem disponíveis, a escolha da substituição pode ser facilitada. Entretanto, na ausência desta informação, os experimentos se tornam mais difíceis, exigindo outros dados para limitar o alvo de mutagênese para as substituições que provavelmente possuirão resultados interessantes.

Uma maneira experimental de investigar a funcionalidade de uma proteína é alterar os resíduos que estão hipoteticamente envolvidos na função, pois isso causaria um impacto significativo na reação (Bogan & Thorn, 1998). Uma técnica muito utilizada

para identificar sítios funcionalmente importantes é a mutagênese por varredura (*scanning mutagenesis*), que consiste na substituição individual de resíduos por um tipo de aminoácido, que pode ser alanina, glicina, prolina ou cisteína (Clackson & Wells, 1995; Gardsvoll *et al.*, 2006; Konishi *et al.*, 1999; Kouadio *et al.*, 2005; Qin *et al.*, 2003). Raramente mutações múltiplas são testadas para o mesmo resíduo (Xiang *et al.*, 2006; Yang *et al.*, 2003), e a alanina é o aminoácido mais utilizado para substituições. Isto acontece porque, diferente da glicina, a alanina retém a estereoquímica da molécula. Além disso, a ausência de cargas na cadeia lateral (que não ocorre em nenhum dos outros 18 aminoácidos) torna improvável que sejam introduzidos conflitos estéricos ou eletrônicos após a substituição (Coombs & Corey, 1998; Bromberg & Burkhard, 2008).

Se uma substituição não resultar em uma significativa mudança de função, o resíduo original provavelmente não era crítico para a atividade no tipo selvagem. A maioria dos resíduos se encontra nesta categoria. Geralmente, o interesse encontra-se naqueles aminoácidos importantes funcionalmente, e sua substituição causará uma mudança considerável na estrutura ou função da proteína. Entretanto, uma pequena alteração na proteína pode afetar múltiplas propriedades, como estrutura, enovelamento, flexibilidade, estabilidade ou mesmo enzimáticas e de interação. Isto dificulta a interpretação do conhecimento adquirido a partir de proteínas mutadas, pois uma alteração na atividade da proteína, ocasionada por uma substituição, não indica que o resíduo original estava diretamente envolvido em uma importante interação funcional. A inativação pode ser resultado, por exemplo, de um impedimento do correto enovelamento da proteína.

A principal desvantagem da mutagênese por varredura na investigação da função de macromoléculas são os recursos exigidos na fabricação, purificação e análise de um grande número de mutantes. Por isso é necessário utilizar toda a informação disponível para limitar o número de resíduos candidatos à alteração. Por exemplo, quando realizamos uma busca por resíduos críticos, é sensato assumir que eles estarão conservados em homólogos relacionados. Portanto, limitar a mutagênese a resíduos conservados é uma abordagem inicial para o emprego das substituições. No caso das enzimas, existem propriedades que implicam em certos aminoácidos

funcionalmente importantes, aos quais pode-se restringir a mutagênese (Coombs & Corey, 1998).

Ainda analisando as desvantagens, estudos experimentais de mutações em proteínas são caros e consomem um grande período de tempo, tornando o número de mutantes inclusos limitado. A partir de uma única molécula de proteína, a mutagênese permitirá a geração de inúmeros mutantes. Entretanto, investigações computacionais de mutagênese podem ser realizadas em um grande número de mutantes, e com algoritmos e softwares eficientes, todos os possíveis resultados de uma dada proteína podem ser analisados (Masso *et al.*, 2006). Muitos estudos de mutagênese computacional foram descritos recentemente (Chasman & Adams, 2001; Karchin *et al.*, 2005; Krishnan & Westhead, 2003; Ng & Henikoff, 2003; Verzilli *et al.*, 2005; Wang & Moulton, 2001). O desafio é utilizar o conhecimento disponível para definir as alterações que resultem em novos conhecimentos a respeito da função da proteína, ou até mesmo informações potenciais para aplicações laboratoriais, médicas ou industriais.

1.4. Aplicações da Modelagem Molecular Comparativa

A modelagem por homologia é um meio eficiente de se obter informações úteis a respeito da proteína de interesse. A aplicação dos modelos construídos por meio desta técnica está diretamente relacionada com a sua precisão. Podemos classificar os modelos em três tipos: de baixa, média e alta precisão. Cada um deles possui qualidade suficiente para auxiliar em determinadas investigações, conforme representado na Figura 1.5.

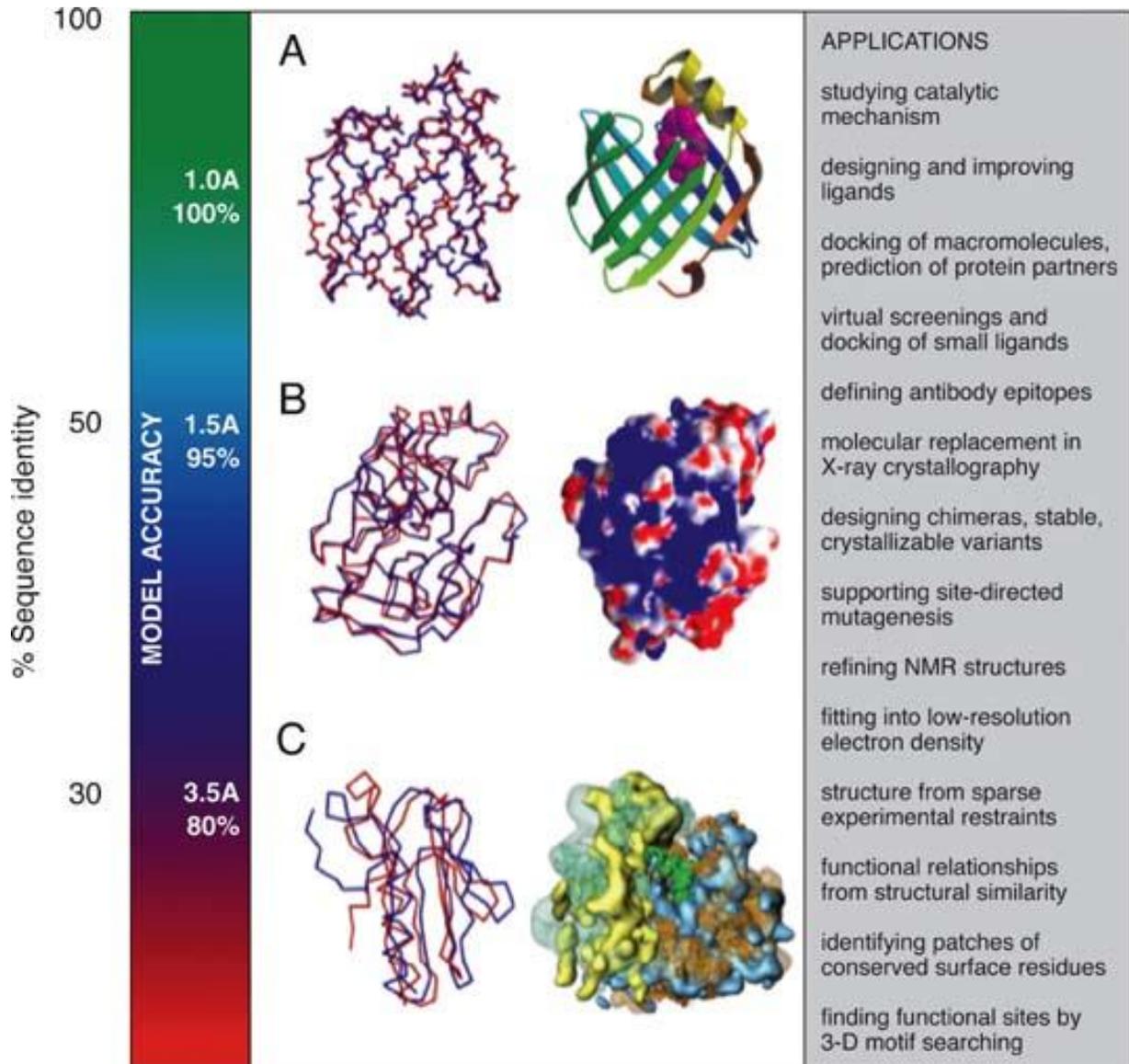


Figura 1.5: Aplicações dos modelos construídos por meio de modelagem por homologia. Esta figura ilustra a relação entre a precisão do modelo e suas potenciais aplicações. É representado o grau de informação estrutural disponível em uma variação de 80 a 100% (que corresponderiam a modelos de resoluções de 3,5 a 1Å, respectivamente). Nos melhores índices de precisão, é possível analisar átomos individuais. No outro extremo, apenas o padrão de enovelamento pode ser predito. (Eswar et al., 2007)

Os modelos de baixa precisão são baseados em uma identidade de sequências em torno de 30%. Apesar deste baixo valor, esses modelos muitas vezes estão enovelados corretamente, e esta informação pode ser muito útil na predição da função de uma proteína.

Os modelos baseados em uma identidade em torno de 35-40% podem ser considerados de média precisão. Como os sítios ativos e os de ligação são geralmente mais conservados do que o restante da molécula, a modelagem destas áreas é muitas vezes mais precisa (Sanchez & Sali, 1998). A interação com o ligante é mais diretamente determinada pela estrutura do sítio de ligação do que por sua sequência, portanto esses modelos permitem um refinamento da predição funcional. Além disso, a partir deles podem ser obtidas informações da estrutura alvo que não estão presentes no molde, como a predição da localização do sítio ativo (em grupos de resíduos carregados) ou do tamanho do ligante (a partir do volume da cavidade do sítio ativo) (Xu *et al.*, 1996). A aplicação à qual se destina este trabalho também é direcionada a modelos de média precisão. Estes podem ser utilizados para construir mutantes sítio-dirigidos com capacidade de ligação alterada ou exterminada, nos quais depois serão testadas hipóteses a respeito da relação sequência-estrutura-função. Existem muitos exemplos em que modelos *in silico* foram utilizados para inferir funções de proteínas, ajudando a sugerir localização de sítios ativos, parceiros de interação e inclusive possibilitando a criação de mutantes para alterar fenótipos distintos (Kryshtafovych, 2008).

Os modelos mais precisos, por sua vez, são os baseados em uma identidade maior ou igual a 50%. Estes quase não contêm erros e são comparáveis a modelos de alta resolução ($\sim 1,5\text{\AA}$) resolvidos por cristalografia de raio-x. Devido a isso, podem ser utilizados no estudo do mecanismo catalítico e aprimoramento de interações (Eswar *et al.*, 2007).

1.5. Possíveis erros da modelagem molecular comparativa

Os modelos construídos por meio de modelagem por homologia muitas vezes podem possuir erros. Felizmente, dependendo do que está sendo analisado, muitos destes erros não afetarão as investigações biológicas. Portanto, conhecer as limitações dos modelos construídos é um importante aspecto a ser considerado na modelagem.

Os tipos de erros mais encontrados estão, na maioria das vezes, relacionados à implementação incorreta da metodologia da modelagem. São exemplos, erros na

escolha do molde, na elaboração do alinhamento e no posicionamento das estruturas secundárias. Outros tipos de erros podem ocorrer devido à baixa identidade entre as sequências alvo e molde, como enovelamento incorreto de *loops* ou erros nas conformações das cadeias laterais (Bujnicki, 2009).

CAPÍTULO 2 – OBJETIVOS

2.1. Justificativa

A produção de ácido glicônico e cetoderivados por microorganismos possui grande impacto e aplicação comercial nas indústrias farmacêutica, alimentícia, têxtil, entre outras. Apesar de existirem várias abordagens para a produção deste ácido, quimicamente, por exemplo, estas parecem ser menos eficientes e mais custosas do que o processo de fermentação realizado por fungos e bactérias. Mesmo assim, a biossíntese destes ácidos envolve vias muito complexas, muitas vezes inviáveis economicamente.

Métodos de produção utilizando fungos são amplamente empregados, e a utilização de bactérias neste processo tem sido cada vez mais eficiente. Um exemplo é a bactéria *Gluconobacter oxydans*, que converte glicose a ácido glicônico e, posteriormente a cetogluconatos como 2KGA e 5KGA. Uma das dificuldades nesta via metabólica é que, como o produto da reação é um ácido, se não houver a ação de neutralizantes no meio, essa acidez irá inativar as enzimas correspondentes, interrompendo a produção. Este fato nos motiva a buscar um microorganismo resistente a tais ambientes inóspitos.

Gluconacetobacter diazotrophicus é uma bactéria diazotrófica, associativa e endofítica, que é encontrada em diversas espécies de plantas, como a cana de açúcar. Possui duas características de alta importância para o presente estudo, a tolerância a ambientes ácidos e a capacidade de produzir ácidos orgânicos. Além do ácido glicônico, este organismo produz 2KGA, 5KGA e 2,5DKGA.

Os dados relacionados anteriormente, somados ao recente sequenciamento do genoma de *G. diazotrophicus* fornecem informações essenciais para um estudo mais aprofundado da produção de ácido glicônico e seus cetoderivados por esta bactéria.

A função de uma proteína em determinada via metabólica geralmente é estudada por meio de investigações diretas na própria macromolécula. Atualmente existem maneiras de alterar uma proteína em um local determinado com o objetivo de inferir a

respeito de seu funcionamento. Este método é denominado mutagênese sítio dirigida, e possui diversas abordagens, como a mutagênese por varredura.

Estudos experimentais de mutações em proteínas, todavia, consomem grandes períodos de tempo e muitos recursos financeiros. As investigações realizadas através de métodos computacionais podem ser utilizadas como meio de encurtar o caminho dos processos experimentais e propor soluções de uma maneira mais dirigida.

2.2. Objetivos Específicos

O objetivo desse estudo é gerar informações que possam auxiliar no desenvolvimento de um processo mais simples e eficiente de produção de ácido glicônico em *Gluconacetobacter diazotrophicus*. Este projeto está vinculado a uma linha de pesquisa desenvolvida na Embrapa Agrobiologia e na Embrapa Informática Agropecuária.

2.3. Atividades

As atividades desenvolvidas durante este trabalho para atingir o objetivo proposto foram:

1. Levantar referencial teórico através de:
 - a. Revisão da literatura;
 - b. Disciplinas de pós-graduação.
2. Analisar dados fornecidos pela EMBRAPA Agrobiologia e definir plano de trabalho;
3. Levantamento de sequências de aminoácidos e estruturas homólogas em bancos de dados de estruturas, buscando encontrar determinadas proteínas com sequências similares às sequências alvo;
4. Levantamento e aprendizagem no uso de softwares utilizados para visualização, edição, modelagem, docagem e dinâmica molecular;
5. Modelagem comparativa das sequências alvo;
6. Validação e análise estrutural dos modelos obtidos;

7. Modelagem das estruturas com substituição dos aminoácidos escolhidos por alaninas;
8. Refinamento dos modelos através de dinâmica molecular, após substituição;
9. Validação e análise estrutural dos modelos obtidos com as alterações;
10. Sugestão de mutações sítio-dirigidas baseada nas informações obtidas durante o estudo;
11. Preparação de artigo e tese.

CAPÍTULO 3 - MATERIAIS E MÉTODOS

A metodologia aplicada neste trabalho pode ser resumida na figura 3.1, na forma de um pipeline, o qual será detalhado ao longo do capítulo em questão.

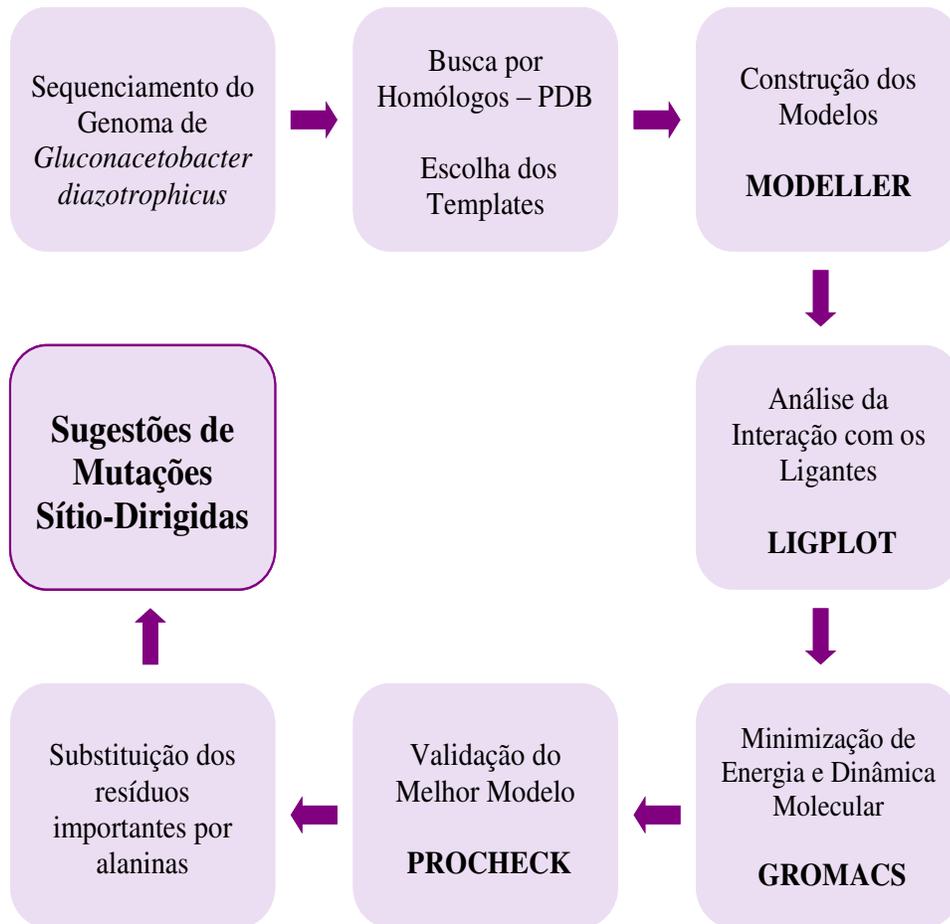


Figura 3.1: Pipeline referente à metodologia utilizada no estudo.

3.1. Modelagem Molecular

A grande maioria dos processos biológicos existentes requer o envolvimento de uma ou mais moléculas de proteína. Na era da biologia de sistemas, é cada vez mais importante entender a respeito do funcionamento do proteoma dos organismos. O papel

dos complexos protéicos e suas redes de interação na regulação e função celular ganharam destaque nas recentes pesquisas biológicas.

Conhecer a estrutura tridimensional de proteínas é essencial para estudar as interações bioquímicas nas quais estas estão inseridas. O aumento exponencial do número de genes sequenciados (Figura 3.2) evidencia a pequena quantidade de estruturas resolvidas mundialmente (Figura 3.3) (Venselaar *et al.*, 2010). Isto se deve ao fato de que existem apenas duas técnicas espectroscópicas capazes de produzir coordenadas tridimensionais de macromoléculas a resolução atômica, a ressonância magnética nuclear (RMN) e a cristalografia de raio-x.

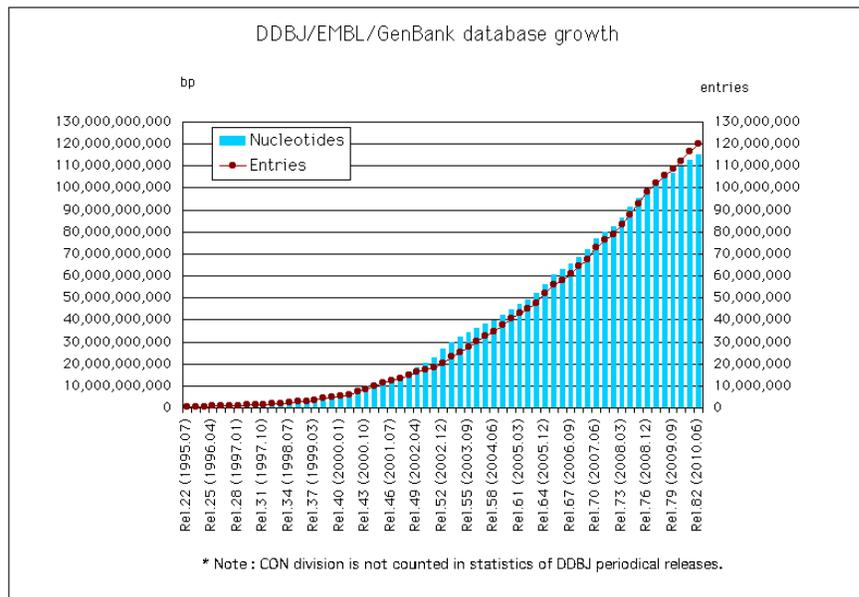


Figura 3.2: Crescimento do banco de dados Genbank. A última atualização, em junho de 2010, relata 120.034.097 entradas e 115.169.689.543 pares de bases depositados.

Fonte: http://www.ddbj.nig.ac.jp/images/breakdown_stats/DBGrowth-e.gif - 27/08/2010

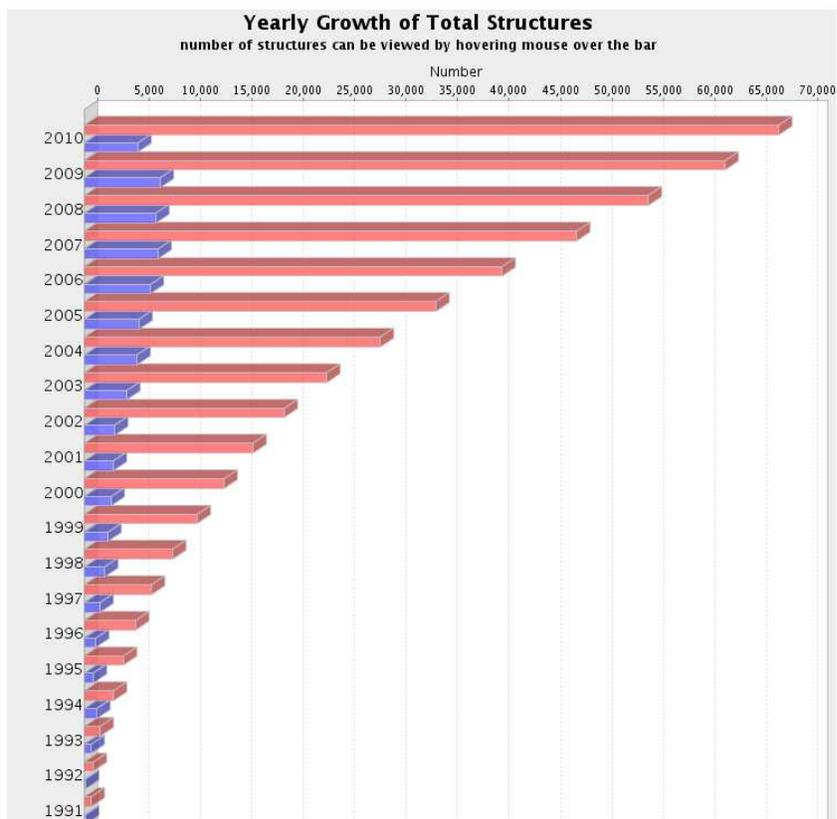


Figura 3.3: Crescimento do banco de dados de estruturas de proteínas, o PDB. As barras azuis indicam o número de estruturas depositadas em cada ano, e as vermelhas representam o total de estruturas depositadas até o momento. A última atualização (24/08/2010) indicava 5.243 estruturas depositadas em 2010, e 67.529 no total.

Fonte: <http://www.pdb.org/pdb/statistics/contentGrowthChart.do?content=total&seqid=100> – 27/08/2010

Os métodos experimentais para determinação de estruturas de proteínas são bastante complexos e exigem grande demanda de tempo e recursos financeiros. Além disso, embora sejam bastante confiáveis, ainda é um desafio obter informações a respeito dos aspectos dinâmicos das subunidades e das consequências das muitas modificações as quais as moléculas são submetidas. Se a maioria das interações ocorre nestas áreas dinâmicas, pode-se considerar esta dificuldade de predição como um problema considerável. Seria importante, então, o desenvolvimento de métodos que, além da informação estrutural, nos permitisse inferir a respeito da dinâmica e regulação funcional das proteínas (Zhou & Robinson, 2010).

O constante aumento da necessidade desta informação tridimensional faz da predição de estruturas de proteínas um dos maiores desafios da biologia estrutural computacional. Nos últimos anos, diversos métodos computacionais foram desenvolvidos, visando validação, modelagem, análise de interações, entre outros. Estes vem se estabelecendo, em muitas áreas, como alternativas viáveis para encurtar o caminho dos processos experimentais.

Muitos projetos de genômica estrutural foram iniciados nos últimos anos com o objetivo de resolver experimentalmente um grupo cuidadosamente selecionado de proteínas. Estas poderiam ser utilizadas como moldes para modelagem computacional de aproximadamente 100 vezes mais proteínas relacionadas (Burley *et al.*, 1999). A maioria das estruturas depositadas recentemente são advindas destes projetos (Burley *et al.*, 2008). Apesar de esta abordagem evidenciar a importância dos métodos experimentais, acredita-se que, no futuro, uma grande parte dos modelos tridimensionais serão obtidos a partir de predição (Manjasetty *et al.*, 2007).

A sugestão de Anfinsen e colaboradores, no final dos anos 50, de que a informação necessária para guiar o enovelamento de uma proteína estaria em sua própria sequência de aminoácidos, criou a possibilidade de criar novas metodologias para a predição de estruturas (Sela *et al.*, 1957). Sabe-se que a estrutura de uma proteína pode fornecer informações importantes para a compreensão de sua função. Considerando este conceito, podemos sugerir que é possível formular uma hipótese a respeito da função de uma proteína apenas a partir de sua sequência de aminoácidos.

A técnica de predição de estrutura de proteínas mais usada e geralmente mais precisa é a modelagem comparativa (Marti-Renom, *et al.*, 2000; Tramontano *et al.*, 2001; Eswar *et al.*, 2007). Na ausência de uma estrutura determinada experimentalmente, tais modelos podem ser usados para verificação de hipóteses ou para complementar experimentos. Quando a identidade entre as sequências de aminoácidos do alvo e do modelo é alta, os modelos determinados por modelagem comparativa podem até mesmo ser usados para avaliar possíveis ligantes ou para inferir sobre o mecanismo catalítico de uma enzima (Bjelic & Aqvist, 2004; Chmiel *et al.*, 2005; Caffrey *et al.*, 2005; Xu *et al.*, 2005, Eramian *et al.*, 2008). Embora informações valiosas possam ser obtidas

a partir da sequência, sua alta maleabilidade muitas vezes impossibilita o encontro de resíduos relevantes. No caso das enzimas, pode-se assumir uma função similar entre duas proteínas se a identidade de sequência entre elas for maior que 40%. Estima-se, no entanto, que 75% das enzimas homólogas compartilhem menos de 30% de posições idênticas (Todd *et al.*, 2001).

A modelagem comparativa baseia-se no conceito de que a estrutura é mais conservada do que a sequência, e que uma pequena mudança na sequência geralmente resulta em uma pequena mudança na estrutura tridimensional. Sendo assim, se as proteínas forem estreitamente relacionadas, a modelagem comparativa permite obter um modelo tridimensional para uma proteína de estrutura desconhecida (alvo) baseada em uma ou mais proteínas de estruturas conhecidas (molde/*template*) (Johnson *et al.*, 1994; Sali & Blundel, 1993; Sanchez & Sali, 1997; Marti-Renom *et al.*, 2000; Fiser *et al.*, 2002; Fiser & Sali, 2002; Ginalski, 2006; Petrey and Honig, 2005).

Todos os atuais métodos de modelagem baseada em moldes consistem de 5 passos principais (Figura 3.4). O primeiro é a busca por proteínas com estruturas conhecidas que são relacionadas à sequência alvo. O segundo consiste em definir as estruturas que serão usadas como molde. O terceiro passo é alinhar as sequências escolhidas com a sequência alvo. O quarto é construir um modelo para o alvo, baseado em seu alinhamento com os moldes. O último passo trata-se da validação dos modelos construídos, por meio da utilização de inúmeros critérios. Uma relação detalhada de cada etapa será descrita nos tópicos 3.1.1 a 3.1.5.

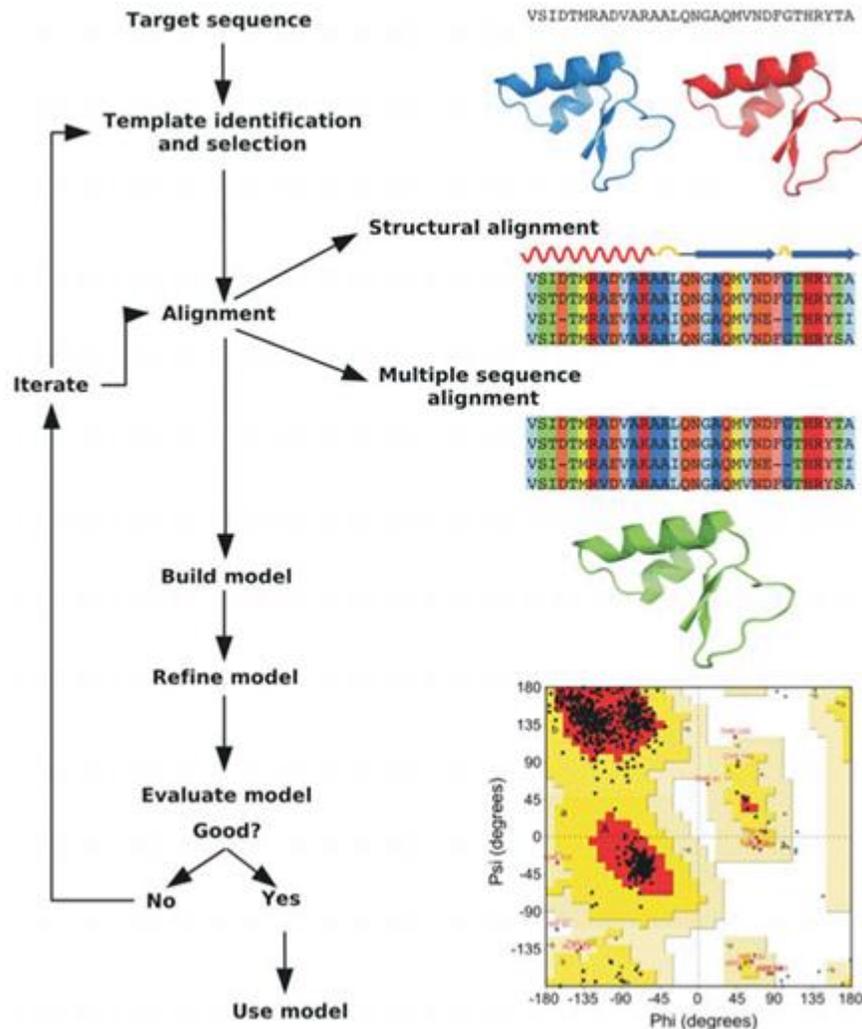


Figura 3.4: Etapas da modelagem comparativa. Representa a metodologia frequentemente utilizada para obtenção de modelos adequados para determinado estudo. O primeiro passo consiste na busca de seqüências homólogas à seqüência alvo. Em seguida são determinados os moldes para modelagem. O alinhamento das seqüências é feito e utilizado como arquivo de entrada para o quarto passo, a modelagem em si. A última etapa consiste na avaliação dos modelos construídos, para verificar o mais adequado para o estudo em questão.

Fonte: http://biskit.pasteur.fr/use/workflows/img/homology_modelling.png – 27/08/2010

A busca pelo molde adequado tradicionalmente é feita a partir da execução de uma busca no BLAST (Altschul, 1990) contra o banco de dados de proteínas (PDB) (Berman et al., 2007). Geralmente esta é feita comparando a seqüência alvo com a seqüência de cada uma das estruturas do banco de dados. Esta abordagem é frequentemente bem sucedida quando o alvo é similar a alguma estrutura do banco.

Entretanto, as sequências que estão próximas ao limiar da modelagem por homologia muitas vezes não são detectadas (Sander e Schneider, 1991).

O desenvolvimento do PSI_BLAST (Altschul *et al.*, 1997) e de métodos de reconhecimento de enovelamento (Jones, 1999) aprimoraram a busca por utilizarem um perfil ao invés de uma única sequência na pesquisa no banco de dados. Além disso, o aumento no número de estruturas depositadas aumenta as chances de se encontrar estruturas homólogas.

3.1.2. Definição de moldes

Existem muitas abordagens diferentes para definir o molde mais adequado para a modelagem. A forma mais simples é selecionar a estrutura com a maior identidade de sequência em relação à sequência alvo. Entretanto, muitas vezes são encontradas várias sequências com identidade similar, sendo mais preciso utilizar outros critérios de seleção para melhorar a qualidade da modelagem.

Juntamente com a identidade, a similaridade entre sequências deve ser observada. A substituição de um resíduo por outro não implica, necessariamente, na alteração da estrutura. No caso de uma substituição conservativa, na qual o aminoácido original possui a mesma característica que o atual (polar, apolar, etc.), muito provavelmente haverá uma conservação da estrutura da molécula.

O tamanho das sequências também deve ser analisado, pois uma alta identidade pode significar um falso resultado se a área de cobertura na sequência alvo for pequena. Uma cobertura de 80 – 90% da sequência, aliada a uma alta identidade, nos permite supor que as proteínas possuem uma similaridade na estrutura como um todo. Por sua vez, uma baixa cobertura aliada a uma alta identidade pode indicar, por exemplo, que as proteínas possuem apenas um domínio em comum. No caso das proteínas que apresentaram baixa similaridade de sequência com outros membros de sua possível família, foram realizadas buscas baseadas na similaridade estrutural, utilizando o programa 3D-BLAST (Tung *et al.*, 2007; Yang & Tung, 2006).

A existência de motivos ou domínios conservados não deixa de ser importante, já que estes estão relacionados a funções específicas. Um domínio conhecido conservado pode auxiliar na inferência do funcionamento da proteína.

Após análise dos aspectos relacionados a sequências, devemos considerar a qualidade do modelo experimental. Além da resolução da estrutura, que é uma medida de avaliação muito utilizada, dados como *R-factor*, *R-free* e muitos outros que auxiliam na definição da credibilidade das estruturas disponíveis, podem ser analisados na escolha do molde mais adequado. Fatores que impliquem na ausência de informação, como grandes partes da sequência que não são alinhados com algum modelo estrutural (*gaps*) ou muitos resíduos faltantes no arquivo PDB, devem ser identificados para que sejam realizados procedimentos de otimização, como uma modelagem específica no local.

A escolha dos critérios utilizados para selecionar o molde mais adequado depende, principalmente, do propósito da modelagem. No presente estudo, por exemplo, é importante avaliar a interação estrutura-ligante. Portanto, deve haver um equilíbrio entre os parâmetros que definem um bom molde. Os moldes que possuem estruturas resolvidas livres e complexadas com o substrato tiveram preferência neste trabalho, pois facilitam a compreensão do funcionamento da enzima. A maioria das proteínas candidatas a molde possui mais de uma estrutura resolvida. Foi escolhida, sempre que existente, a proteína livre de substrato para realizar a modelagem. As estruturas complexadas foram comparadas a estas para verificar mudanças conformacionais ocorridas devido à interação com os ligantes.

3.1.3. Alinhamento de sequências

Em geral, programas de modelagem por homologia exigem informações a respeito da equivalência entre as sequências para a construção de um modelo adequado. Muitas delas são definidas a partir de um correto alinhamento entre a sequência alvo e a sequência molde. Portanto, uma vez que o molde é escolhido, é necessário utilizar um método de alinhamento eficiente. Este é obtido com certa

facilidade se a identidade entre as sequências for igual ou maior que 40%. Entretanto, se esta for menor que 30%, muitas vezes o primeiro alinhamento não é o mais adequado. Nestes casos, é importante fazer a correção manualmente, pois este é um fator que afeta a qualidade do modelo construído (Fiser, 2010).

Existem vários programas disponíveis para alinhar um determinado número de sequências relacionadas, e o utilizado neste estudo foi o CLUSTAL W (Thompson et al., 1994). Além do alinhamento entre as sequências alvo e molde, foram realizados alinhamentos múltiplos com proteínas da mesma família, para que fossem obtidas informações adicionais como resíduos conservados, áreas de superfície, inserções e deleções.

A existência de um modelo estrutural como molde facilita na construção do alinhamento, que pode ser melhorado através da inclusão de informações estruturais. Por exemplo, deve-se evitar *gaps* em estruturas secundárias ou regiões enterradas. Podemos concluir, portanto, que a qualidade do alinhamento está diretamente relacionada com a qualidade do modelo a ser construído.

3.1.4. Construção dos modelos

O programa escolhido para a construção dos modelos foi o MODELLER 9v6 (Fiser & Sali, 2003; Sali & Blundell, 1993), por ser uma ferramenta amplamente utilizada que produz resultados satisfatórios. Neste programa é implementada a modelagem por satisfação de restrições espaciais, na qual são geradas restrições ou limitações na estrutura da sequência alvo, baseadas no alinhamento com as estruturas relacionadas. Para obtenção do modelo final, estas restrições são aliadas a uma função objetivo que é otimizada por meio de uma combinação de gradientes conjugados e dinâmica molecular (Petrey et al., 2003).

Para construção do modelo mais adequado, foram realizadas modelagens com um molde, vários moldes e refinamento de *loops*, de acordo com a necessidade de cada proteína. Neste também foram realizados alinhamentos, os quais foram comparados aos gerados pela ferramenta CLUSTAL W (Thompson et al., 1994) e

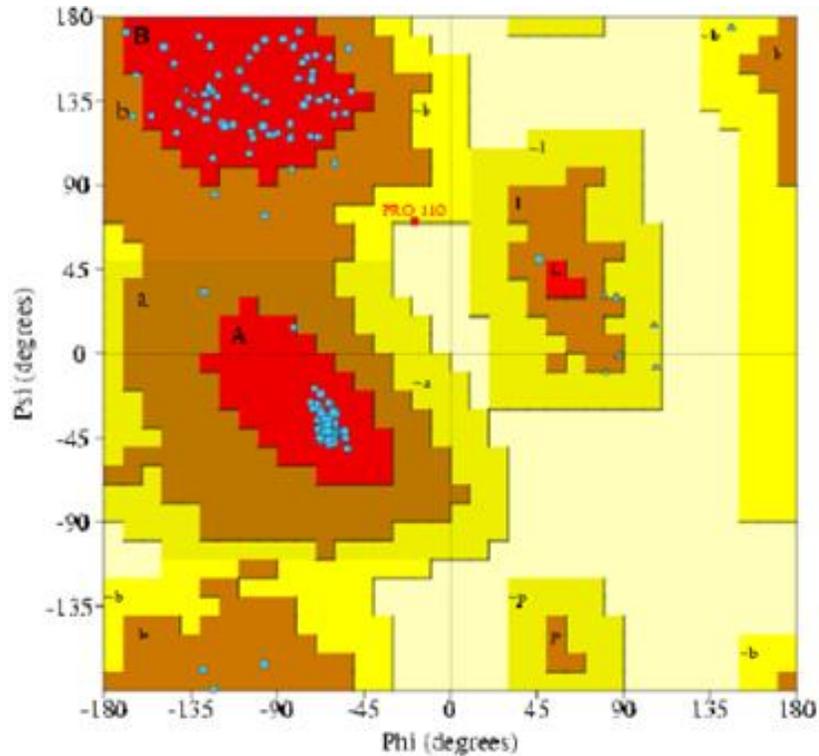
modificados manualmente sempre que julgado necessário. Estes casos possuem como objetivo manter as estruturas secundárias íntegras, ou manter conservados os resíduos importantes para a interação com o substrato.

Foram construídos dez modelos para cada sequência alvo, e a escolha do mais indicado foi feita levando-se em conta a menor função objetivo encontrada. A função objetivo do MODELLER é uma medida de quão bem o modelo satisfaz as restrições espaciais inseridas. Menores valores da função objetivo indicam um melhor ajuste aos dados inseridos e, portanto, estes modelos provavelmente são mais acurados (Sali & Blundell, 1993). Estes são otimizados por meio da dinâmica molecular, que será discutida no item 3.4.

3.1.5. Validação do modelo

Após a construção dos modelos, é necessário investigar a existência de erros. A qualidade de um modelo pode ser predita inicialmente a partir da similaridade entre as estruturas do alvo e do molde. A avaliação da estereoquímica de um modelo (ângulos, ligações e distâncias entre átomos) pode ser feita por meio de diversas ferramentas. Neste estudo foi utilizado o programa PROCHECK (Laskowski et al., 1993), que avalia a condição do modelo gerado por meio da qualidade da estereoquímica geral e individual de cada resíduo.

Uma das visualizações feitas é do gráfico de Ramachandran (Figura 3.5), que representa os ângulos diédricos ϕ e ψ dos resíduos da estrutura correspondente. Estes ângulos determinam o grau de liberdade de uma ligação peptídica (Figura 3.6). Sua análise permite estimar quais resíduos estão em posições menos favoráveis na estrutura, muitas vezes sendo resultado de erros na modelagem. É necessário enfatizar, no entanto, que muitos aminoácidos estão situados nessas regiões devido a uma funcionalidade específica na molécula de proteína, não devendo ser considerado um erro.



		Número de resíduos	%
Regiões mais favoráveis	[A, B, L]	139	92,1 %
Regiões permitidas	[a, b, l, p]	12	7,9%
Regiões generosamente permitidas	[~a, ~b, ~l, ~p]	0	0,0 %
Regiões não permitidas		0	0,0 %
Resíduos não prolina e não glicinas		151	100%

Figura 3.5: Gráfico de Ramachandran. Nele estão representados os ângulos phi e psi de cada resíduo, indicando se as conformações adotadas por ele são favoráveis ou não à estrutura. As regiões em vermelho são as mais favoráveis, onde se encontram 92,1% dos resíduos. O restante, 7,9%, está localizado em regiões ainda permitidas, representadas em marrom claro. Os espaços em amarelo e branco representam regiões generosamente permitidas e não favoráveis, respectivamente. Nota-se que os resíduos de prolina e glicina não são levados em conta nos dados. Estes devem ser analisados separadamente, por apresentarem características estereoquímicas próprias.

Fonte: <http://www.ebi.ac.uk/pdbsum/> – 27/08/2010

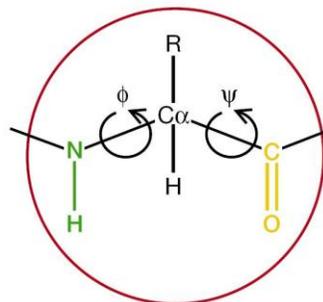


Figura 3.6: Representação dos ângulos phi e psi de um aminoácido. Cada aminoácido possui dois graus de liberdade, ou seja, podem girar em duas ligações. O ângulo de rotação phi (ϕ) situa-se em torno da ligação N-C α , e o psi (ψ) em torno de C α -C (Mount, 2001)

Além deste gráfico, existem outras ferramentas de avaliação, como a predição de estrutura secundária (Figura 3.7). Estas são comparadas às estruturas secundárias do molde, disponíveis no PDB, para verificar a similaridade correspondente

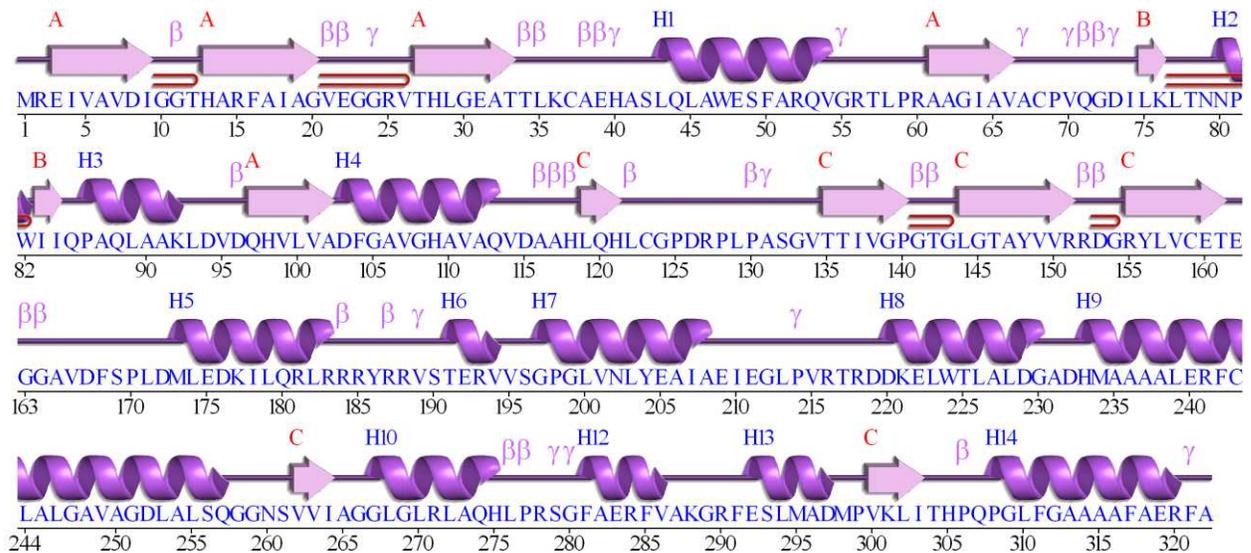


Figura 3.7: Representação da predição de estruturas secundárias do modelo construído. Em roxo estão representadas as α -hlices, em rosa as fitas- β . As letras A, B e C em vermelho determinam as folhas- β nas quais as fitas se situam e motivos específicos são representados por β e γ .

Fonte: <http://www.ebi.ac.uk/pdbsum/> - 27/08/2010

Muitas das análises foram realizadas a partir da observação das estruturas em ferramentas de visualização. As mais utilizadas foram Pymol (DeLano, 2002), Swiss Pdb Viewer (Guex & Peitsch, 1997) e Visual Molecular Dynamics (VMD) (Humphrey *et al.*, 1996). Este último também foi utilizado para gerar modelos estruturais com mais de uma cadeia, como por exemplo o tetrâmero da Gluconato 5-desidrogenase. Também foram feitas modelagens a partir destas estruturas e comparadas com as elaboradas a partir de apenas um monômero.

3.2. Dinâmica Molecular

Proteínas são moléculas com grande flexibilidade conformacional, a qual é muito importante para sua funcionalidade. É possível obter informações sobre mudanças conformacionais das macromoléculas a partir de simulações da dinâmica molecular.

A dinâmica molecular é, provavelmente, a técnica mais poderosa para simulação de movimentos de proteínas dentre os métodos teóricos existentes. Apesar dos avanços substanciais de algoritmos, a teoria básica por trás das simulações de dinâmica molecular é bastante simples. As posições e os momentos dos átomos são especificados, e estes interagem entre si, ao longo do tempo, segundo as leis da mecânica clássica. Os átomos “movimentam-se” através de potenciais de interação, chocando-se uns contra os outros (em um sistema fluido), de um modo bastante similar ao que acontece na realidade.

Na maioria das interações, a água é o principal ambiente das proteínas. Para que a simulação se aproxime ao máximo do real, é necessário adicionar moléculas de água e íons em concentração fisiológica ao sistema, visando a solvatação da proteína. Estes elementos são inseridos no interior de uma caixa, construída com base na estrutura da proteína, para que seja ajustada à sua forma.

As equações são resolvidas simultaneamente em pequenos espaços de tempo, tomando-se o devido cuidado para que a pressão e temperatura do sistema mantenham os valores desejados. A Figura 3.8 representa os passos de uma dinâmica molecular. A ferramenta utilizada para a execução desta etapa do trabalho foi o pacote GROMACS (Van der Spoel et *al.*, 2005).

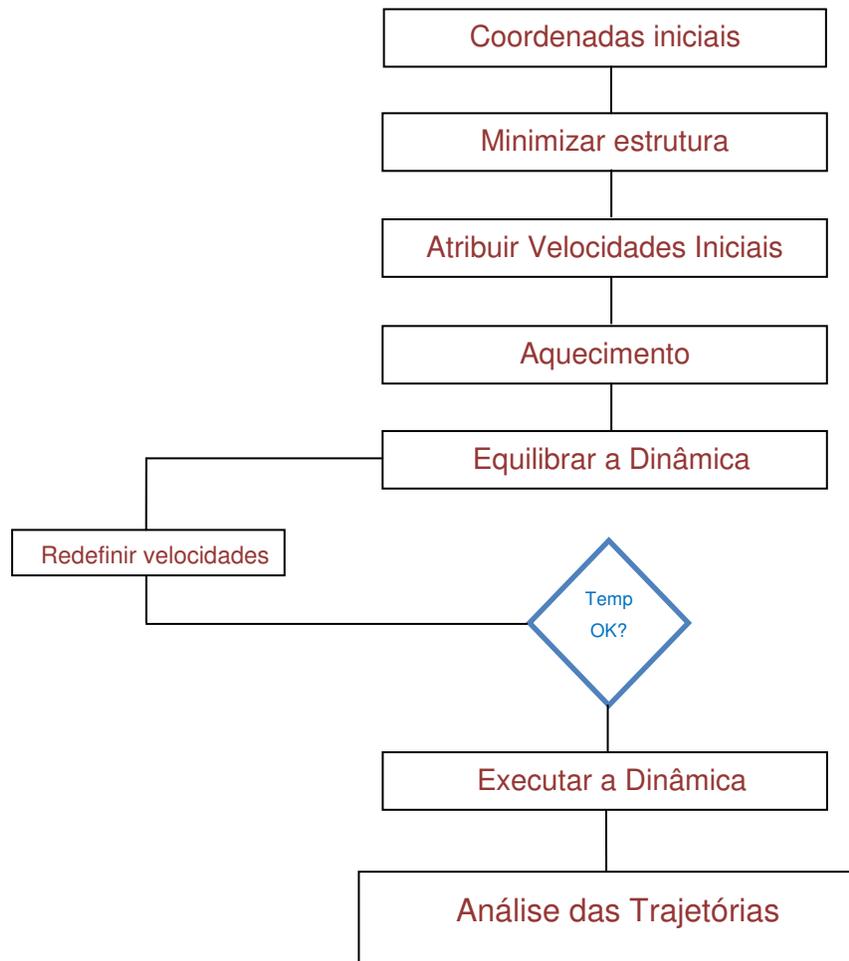


Figura 3.8: Etapas da Dinâmica molecular (Van der Spoel, 2005)

3.3. Interações proteína-ligante

As proteínas estão envolvidas em diversos sistemas biológicos, e todas elas possuem uma função específica. Uma das mais importantes funções relacionadas à superfície de uma proteína são suas interações com outras moléculas. A ligação seletiva de um ligante a uma proteína específica é determinada pelo reconhecimento estrutural e energético entre ambos, e as ligações diretas entre elas (Figura 3.9) são extremamente importantes para a interação (Böhm, 2003).

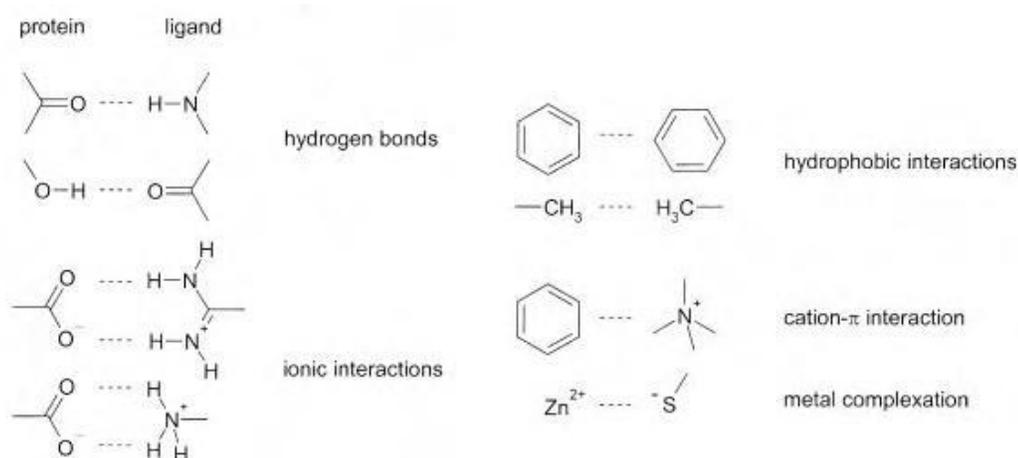


Figura 3.9: Interações frequentemente encontradas em complexos proteína-ligante. (Böhm, 2003)

A busca por sítios de ligação em estruturas de macromoléculas podem ser classificados em três tipos, por docagem, cavidades ou similaridade (Kinoshita et *al.*, 2009). A abordagem utilizando docagem é feita otimizando a avaliação da adaptação do ligante nas estruturas das proteínas.

Outra alternativa é realizar uma busca por cavidades na estrutura da proteína (Laurie & Jackson, 2005). Este método tem como objetivo identificar potenciais sítios de ligação, mas não fornece informações a respeito dos possíveis ligantes. Portanto, para otimizar a predição destes sítios, seria promissor que esta técnica fosse combinada à outra, como a de busca por similaridade.

A informação mais confiável para inferir sobre potenciais sítios de ligação, na abordagem de busca por similaridade, é a homologia. Se proteínas complexadas possuírem uma alta identidade entre si, provavelmente o modo de ligação será similar.

A atual existência de dados experimentais disponíveis a respeito de estruturas tridimensionais de complexos proteína-ligante permitem sua utilização na identificação de características comuns entre os sítios de ligação.

Este tipo de análise foi utilizado neste trabalho. Estudos aprofundados a respeito das interações proteína-ligante dos moldes foram realizados por meio da ferramenta LIGPLOT (Wallace et *al.*, 1995). Seus resultados foram extensivamente estudados e as

informações obtidas a partir deles foram utilizadas na predição de mutações sítio-dirigidas de interesse.

3.4. Mutagênese sítio-dirigida

A mutagênese *in silico* do modelo foi realizada com o objetivo de verificar as possíveis consequências de uma substituição de aminoácidos considerados importantes. Estes foram substituídos, inicialmente, todos de uma vez, e posteriormente, um a um, na sequência FASTA. Esta sequência alterada foi utilizada para realização de um novo processo de modelagem, e as mudanças conformacionais ocorridas devido às substituições foram determinantes na escolha dos resíduos potenciais para a mutagênese experimental.

O estudo do mecanismo catalítico dos moldes é determinante na escolha dos resíduos selecionados, pois a partir disso é possível entender o funcionamento da molécula, facilitando a identificação dos aminoácidos essenciais. A docagem dos prováveis ligantes no modelo construídos foi realizada com o objetivo de verificar se houve alteração nas interações após as mudanças.

CAPÍTULO 4 - DESENVOLVIMENTO

O capítulo em questão aborda, detalhadamente, a aplicação da metodologia descrita no capítulo 3, para cada sequência alvo. É importante salientar que este trabalho é realizado em colaboração com a EMBRAPA Agrobiologia, que participou do projeto de sequenciamento da *Gluconacetobacter diazotrophicus*. Esta bactéria vem sendo estudada por seus pesquisadores por muitos anos, fato que possibilitou uma busca minuciosa por ORFs importantes na produção de ácido glicônico e seus cetoderivados. A partir deste estudo, foram selecionadas oito ORFs possivelmente relacionadas ao produto de interesse. Estas representam o ponto de partida do presente estudo, e estão relacionadas na tabela 4.1:

Tabela 4.1: ORFs de *G. diazotrophicus* possivelmente relacionadas com a produção de Ácidos Orgânicos.

ORF	Produto
GDI0293	Putativa Chiquimato Quinase
GDI0691	Gluconato 5-desidrogenase
GDI1760	Glucoquinase
GDI1785	Putativa Isoquinolina 1-oxidorreductase subunidade alfa
GDI1864	Glucoquinase
GDI2456	Putativa precursora de L-aminoácido oxidase
GDI2625	Glicose 1-desidrogenase
GDI3432	2 ketogluconate reductase

A metodologia de pesquisa foi aplicada para todas as ORFs selecionadas, pois todas apresentaram resultados satisfatórios para a execução da modelagem. Entretanto, após um estudo mais detalhado da via de produção de ácido glicônico, o trabalho foi direcionado à análise das ORFs mais diretamente relacionadas à síntese dos produtos de interesse. Estas representam três das oito sequências selecionadas, e a aplicação da metodologia nestas está descrita nos tópicos 4.1 a 4.3.

4.1. ORF GDI0691

A ORF GDI0691 é uma sequência de 261 aminoácidos que origina uma gluconato 5-desidrogenase (Ga5DH - E.C. 1.1.1.69), uma enzima com capacidade de catalisar uma oxidorredução reversível entre ácido glicônico e 5-ceto-D-gluconato. Quando a energia se esgota, Ga5DH pode converter 5-ceto-D-gluconato em ácido glicônico, fornecendo uma fonte de carbono e NADP^+ para a bactéria (Sweeney et. *al.*, 1996). Por outro lado, sob condições de excesso de energia, a reação oposta pode ocorrer como um meio de armazenamento de energia (Klasen et. *al.*, 1995). Uma busca em bancos de dados de vias metabólicas resultou na identificação da participação desta enzima no metabolismo de cetogluconatos em *Escherichia coli*, representada na Figura 4.1.

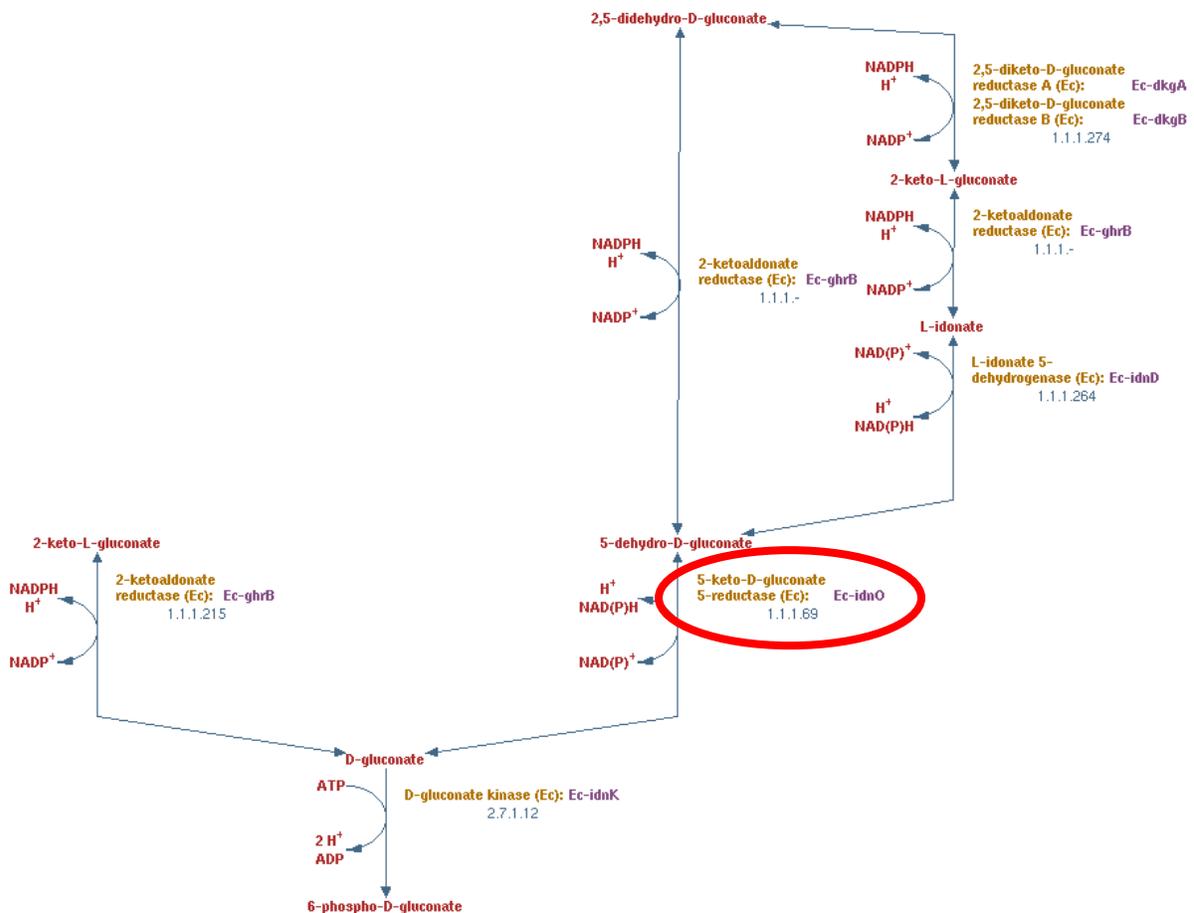


Figura 4.1: Via de metabolismo de cetogluconatos em *E. coli*. A enzima de interesse, circulado em vermelho, catalisa a conversão reversível de D-gluconato a 5-deidro-D-Gluconato. Seus produtos, por sua vez, podem ser transformados em 2-ceto-L-Gluconato, 2,5-dideidro-D-gluconato e 6-fosfo-D-gluconato.

Fonte: <http://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=KETOGLUCONMET-PWY>

Esta enzima pertence à família das desidrogenases/reduases de cadeia curta (SDR), que geralmente catalisam reações dependentes de NAD(P)(H) com uma grande variedade de especificidade de substrato (Bray et. al., 2009). Seus membros compartilham motivos sequenciais de sítios de ligação ao co-fator (TGxxxGxG) e da tétrede catalítica (N-S-Y-K), que algumas vezes podem apresentar variações (Filling et. al., 2002). As estruturas das SDR são nitidamente homólogas, com um padrão comum de enovelamento α/β caracterizado por uma folha- β central com hélices em ambos os lados, típico domínio de enovelamento Rossmann (Persson et al., 2009).

4.1.1. Procura e seleção do molde

A busca por estruturas homólogas à ORF GDI0691 no BLAST contra o PDB resultou em diversas sequências com um bom alinhamento com a sequência alvo. Duas gluconato 5-desidrogenases possuíam boas pontuações, entretanto, devido a melhores valores de identidade e resolução, por exemplo, foi escolhida a estrutura de código PDB 3CXR. Este representa uma enzima da espécie *Streptococcus suis*, e possui uma estrutura relacionada, de código 3CXT, que retrata a mesma enzima, do mesmo organismo, entretanto, complexada com o substrato. Os aspectos analisados para sua escolha estão descritos na tabela 4.2.

Tabela 4.2: Características analisadas para escolha do molde adequado da ORF GDI0691

Molde 3CXR/3CXT	
Produto	Gluconato 5-desidrogenase de <i>S. suis</i>
Tamanho da sequência	182 aa
Identidade/Similaridade	42% / 58%
Gaps	11 (4%)
Domínios Conservados	Superfamília NADB_Rossmann
Resolução	2.0Å
Resíduos faltantes	1-3, 208-213, 270, 271
Estrutura sem substrato	Sim – 3CXR
Estrutura com substrato	Sim – 3CXT
Ligantes	CA (Íon cálcio), NAP (NADP Nicotinamida adenina dinucleotídeo fosfato), GCO (ácido glicônico)

4.1.2. Alinhamento de sequências e interação com os ligantes

Após análise dos alinhamentos construídos com as ferramentas CLUSTAL W e MODELLER, o resultado mais adequado foi gerado, e está demonstrado na Figura 4.2.



Figura 4.2: Alinhamento das sequências da ORF GDI0691 e o molde escolhido, de código 3CXR. As cores representam as características dos resíduos. Os aminoácidos destacados com setas foram identificados como participantes na interação com o substrato. As setas vermelhas indicam interação com NAP e as azuis com GCO.

No alinhamento foram observados os resíduos conservados e, no caso de substituição, se esta foi conservativa, ou seja, não teve alteração das características do resíduo alterado. Além disso, foi feita uma análise para verificar se o alinhamento estava adequado, como por exemplo, sem gaps interrompendo estruturas secundárias.

A análise dos resíduos importantes para a enzima foi feita por meio do estudo de sua interação com os ligantes. As ferramentas utilizadas nesta etapa foram o CLUSTAL W, para construção de alinhamentos múltiplos com membros da mesma família, e o LIGPLOT, que gera representações esquemáticas da interação proteína/ligante. Foram analisadas as interações com os ligantes NAP e GCO. A figura 4.3 representa as interações dos ligantes com a proteína.

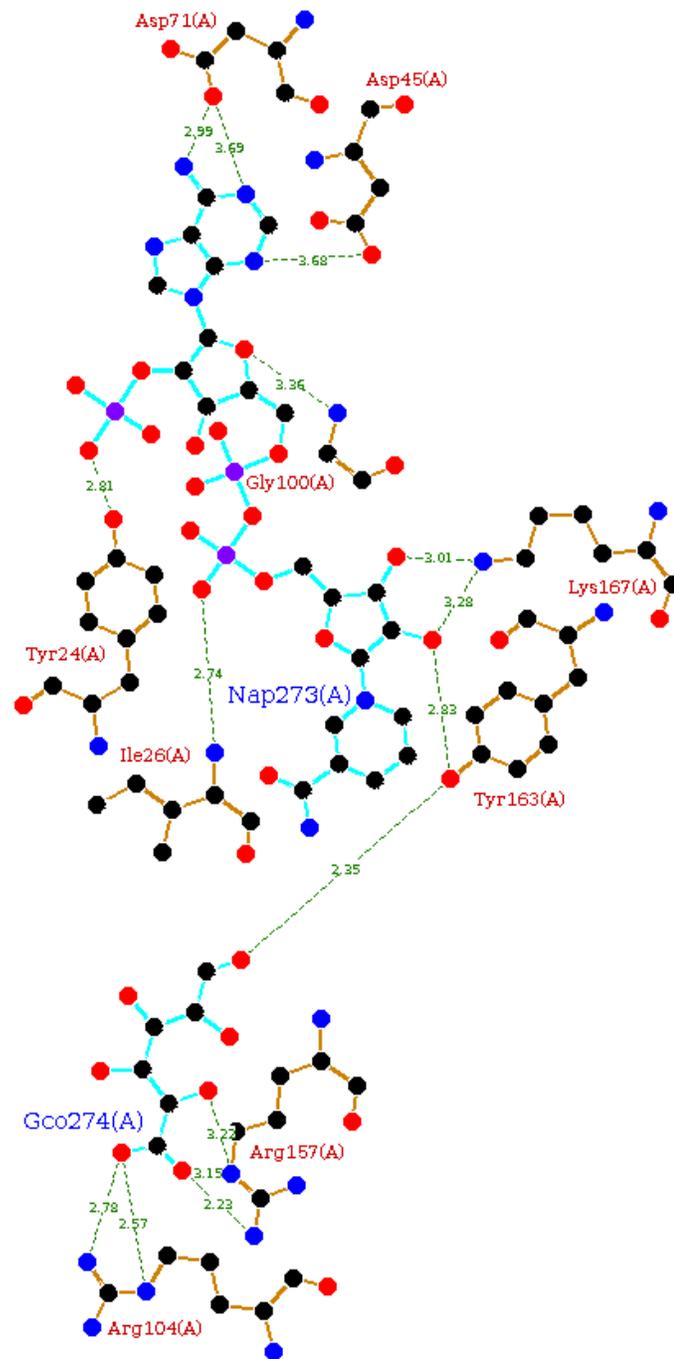


Figura 4.3: Representação esquemática da relação proteína-ligante. Os ligantes Nap e Gkr são representados em azul, e os aminoácidos que interagem com eles, em marrom. As distâncias entre os átomos são demonstradas pelos tracejados verdes, e as esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

Vale ressaltar que no banco de dados de estruturas, o PDB, constava a presença de glucarato (GKR) e não gluconato (GCO) como ligante da Ga5DH de *S. suis*. Com a análise da sua estrutura realizada e os estudos feitos a respeito da enzima no decorrer desse trabalho, foi possível detectar o erro de anotação e notificar os responsáveis pela deposição da estrutura. Esses confirmaram o erro na anotação do PDB e fizeram a devida correção.

Analisando a Figura 4.3, é possível identificar quais resíduos interagem com a enzima em questão. Estes foram marcados no alinhamento (Figura 4.2), com o objetivo de verificar a conservação na sequência-alvo. Os resultados desta análise estão representados na tabela 4.3, onde os asteriscos (*) representam a conservação do resíduo.

Tabela 4.3: Resíduos que interagem com os ligantes no PDB 3CXT.

NAP		GKR	
3CXR	GDI0691	3CXR	GDI0691
Tyr 24	Gly		
Ile 26	*		
Asp 45	Gly		
Asp 71	*		
Gly 100	*		
		Arg 104	*
		Arg 157	*
Tyr 163	*	Tyr 163	*
Lys 167	*		

Pode-se observar, neste caso, que a grande maioria dos resíduos importantes na interação com o substrato estão conservados na sequência alvo. Um mecanismo catalítico é proposto por Zhang e colaboradores (2009), e seu estudo é importante para análise dos resíduos interessantes para mutagênese. Este será discutido mais adiante.

4.1.3. Modelagem comparativa e dinâmica molecular

Após verificar a semelhança entre as sequências alvo e molde, foi realizada a modelagem molecular, que foi realizada utilizando o template 3CXR, que não está complexado com os ligantes. Foram construídos dez modelos e, como explicado no Capítulo 3, o melhor foi escolhido de acordo com a menor função objetivo. A estrutura selecionada foi o modelo 5, e este foi submetido à minimização de energia e dinâmica molecular, ambas realizadas na ferramenta GROMACS. O resultado desta etapa pode ser conferido por meio de uma sobreposição da estrutura molde e do modelo construído, representada na Figura 4.4.

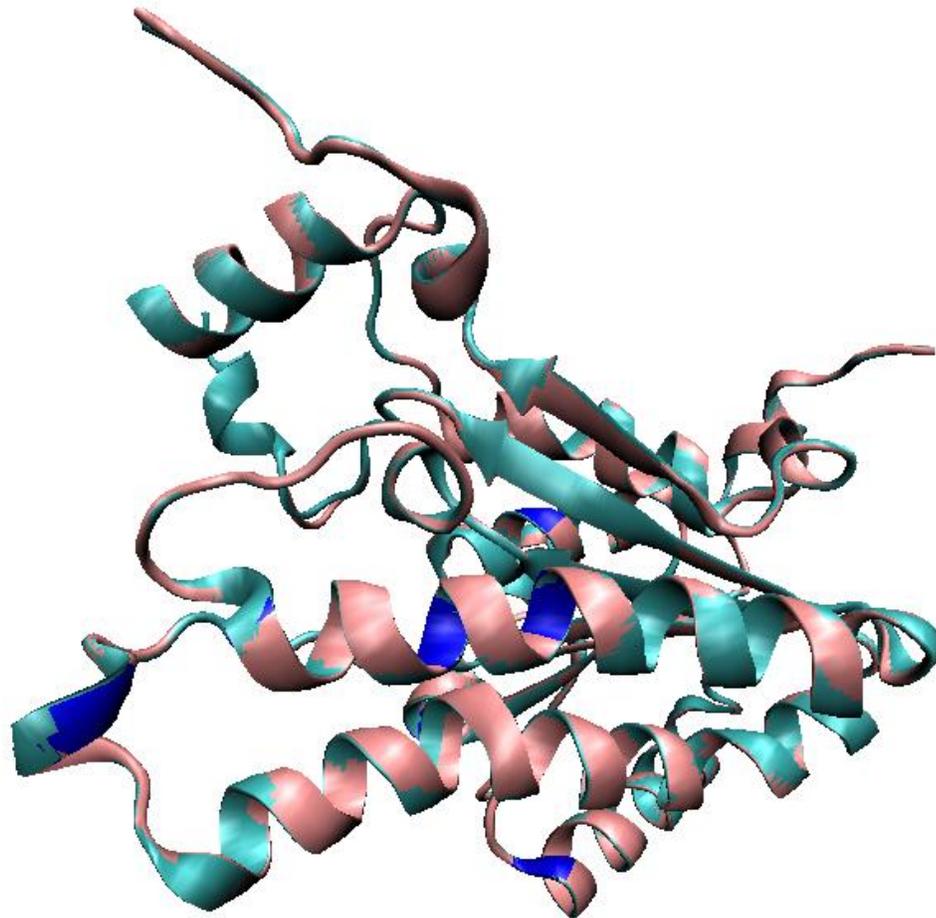
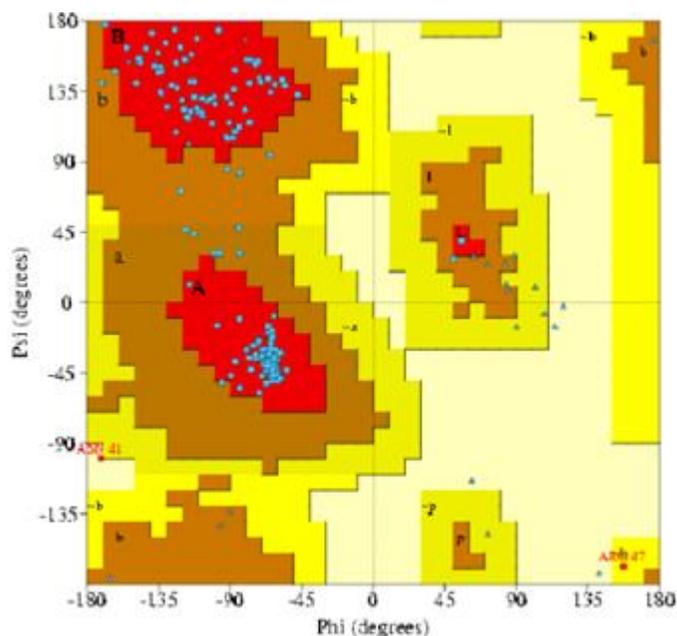


Figura 4.4: Alinhamento estrutural do molde (em azul) e melhor modelo construído (em rosa), após dinâmica. Observando o modelo final antes e após a dinâmica molecular, nota-se que as diferenças entre molde e alvo ocorreram após sua execução. Os resíduos envolvidos na interação com os ligantes estão destacados, e é visível uma conservação também estrutural destes.

Figura preparada com a ferramenta Swiss-pdb Viewer.

4.1.4. Validação do modelo

O modelo final construído foi submetido a uma avaliação estereoquímica geral por meio da ferramenta PROCHECK. Seu resultado é disponibilizado na forma de diversas representações esquemáticas, entre elas o gráfico de Ramachandran, representado na Figura 4.5, e a predição de estruturas secundárias (Figura 4.6), que são comparadas às estruturas do molde (Figura 4.7) para verificar se estas foram mantidas íntegras.



		Número de resíduos	%
Regiões mais favoráveis	[A, B, L]	202	92,7 %
Regiões permitidas	[a, b, l, p]	14	6,4 %
Regiões generosamente permitidas	[-a, ~b, ~l, ~p]	1	0,5 %
Regiões não permitidas		1	0,5 %

Resíduos não prolina e não glicinas		218	100%
Resíduos finais (exc Gly e Pro)		2	
Resíduos Glicina		28	
Resíduos Prolina		13	

Total de Resíduos		261	

Figura 4.5: Gráfico de Ramachandran do modelo construído para a ORF GDI0691. As regiões em vermelho são mais favoráveis, e um modelo com mais de 90% dos resíduos nesta região é considerado adequado. O resíduo situado na região “não permitida” deve ser analisado para verificar as consequências deste posicionamento

Fonte: PdbSum/ PROCHECK

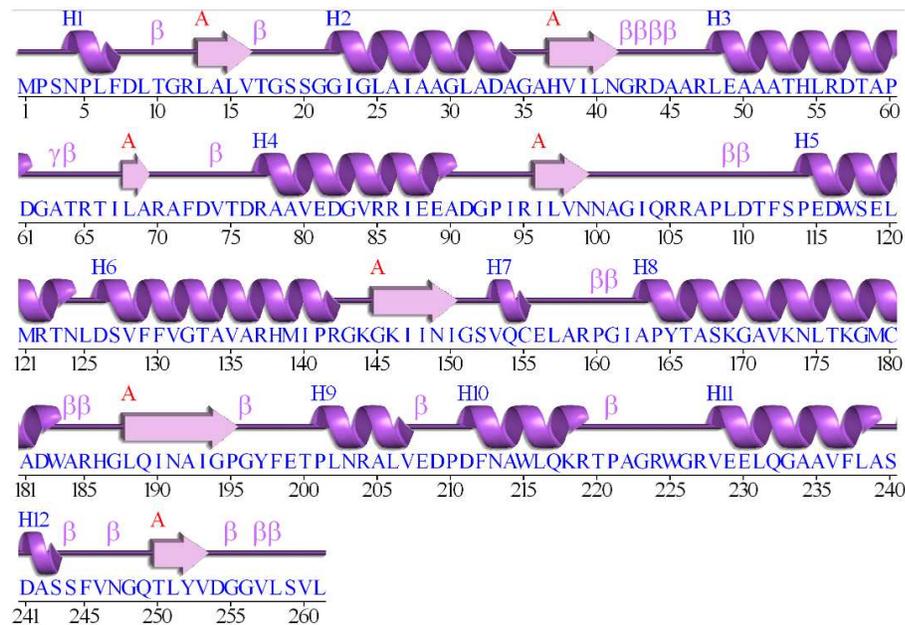


Figura 4.6: Predição de estruturas secundárias do modelo construído para a ORF GDI0691. As regiões em roxo representam α -hélices e as setas na cor rosa indicam fitas- β .

Fonte: PdbSum/ PROCHECK

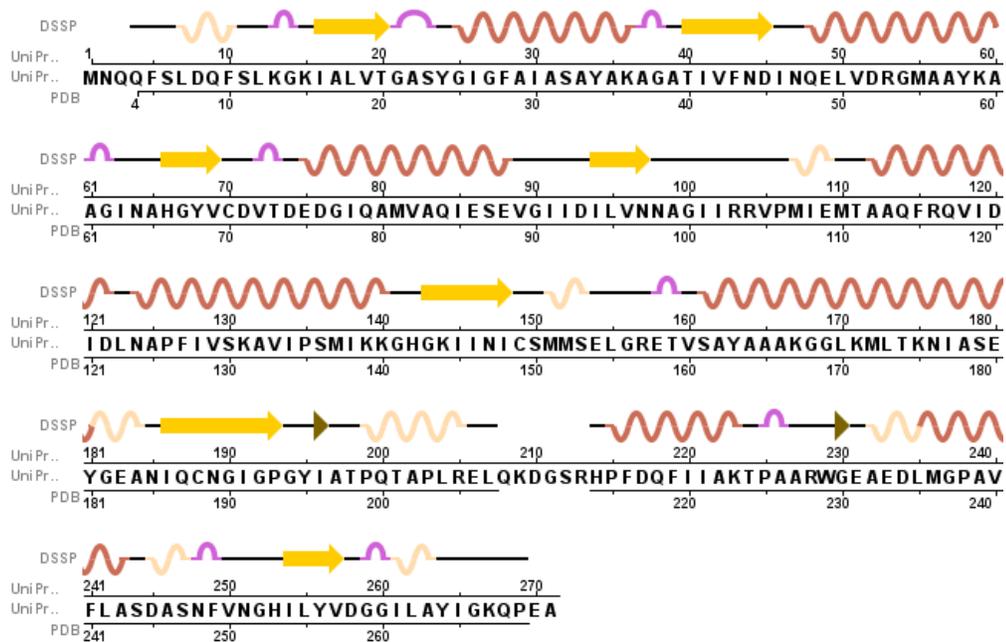


Figura 4.7: Predição de estruturas secundárias molde 3CXR. As α -hélices estão representadas em rosa e vermelho escuro, e as fitas- β em setas amarelas.

Fonte: PDB

Os resultados apresentados neste item nos permitem considerar o modelo final construído adequado para a continuação dos experimentos. Este apresenta uma estrutura similar ao molde, mantendo os resíduos considerados importantes conservados, assim como suas estruturas secundárias. Além disso, sua qualidade estereoquímica e a identidade e resolução do molde proporcionam dados suficientes para a realização da etapa final, a sugestão de locais para mutagênese sítio-dirigida.

4.1.5. Mecanismo catalítico

Para auxiliar na definição das mutações mais indicadas é importante estudar o mecanismo catalítico de cada enzima. No caso da gluconato 5-desidrogenase, Zhang e colaboradores sugeriram um mecanismo um pouco diferente dos propostos anteriormente (2009). Este propõe a existência de uma tétrade catalítica (Arg104-Ser150-Tyr163-Lys167), adicionando a Arginina 104 à tríade (Ser-Tyr-Lys) normalmente funcional em outras enzimas SDR. Uma descrição do mecanismo catalítico é feita a seguir.

A Arginina 104 e a Serina 150 conservadas interagem com o gluconato e exercem um papel importante no reconhecimento do substrato. A Arginina 157 interage com o substrato, e com a ajuda do íon metal e de moléculas de água, este é posicionado em uma posição favorável para que ocorra o mecanismo catalítico. A Lisina 167 interage com a Tirosina 163, e esta por sua vez forma uma ponte de hidrogênio com o gluconato. O processo termina com a liberação do 5-ceto-D-gluconato do sítio ativo.

A partir dessa análise, podemos obter mais informações a respeito da função dos resíduos catalíticos. Uma substituição da serina por alanina resulta em uma perda quase completa de todas as atividades enzimáticas (menos de 0,04% de atividade residual comparada à atividade do tipo selvagem). Entretanto, a substituição desta por uma treonina dá origem a uma enzima mutante, totalmente ativa. Isso indica que uma cadeia lateral hidroxila nesta posição é essencial para a atividade catalítica (Oppermann, *et. al.*, 1997).

Uma proteína mutante obtida por mutagênese sítio dirigida (R104A) possui atividade 20% menor do que a Ga5DH nativa, sugerindo que a Arginina carregada positivamente pode facilitar o acesso e a orientação do substrato (carregado negativamente) no interior do sítio ativo. Isto sustenta a ideia de uma tétrede catalítica em contraste à tríade já estabelecida.

4.2. ORF GDI0293

A ORF GDI0293 possui 182 aminoácidos, e suas anotações no genoma sequenciado pelo projeto RioGene indicam que seu produto é uma chiquimato quinase (E.C. 2.7.1.71). Essa ORF está envolvida na via metabólica do chiquimato (também conhecido como corismato), onde catalisa sua fosforilação usando ATP como substrato, resultando em chiquimato 3-fosfato e ADP (Figura 4.8). A enzima em questão pertence à família estrutural das nucleosídeo monofosfato quinases (NMP), que se destacam por sofrerem grandes mudanças conformacionais durante a catálise (Vonrhein *et. al.*, 1995).

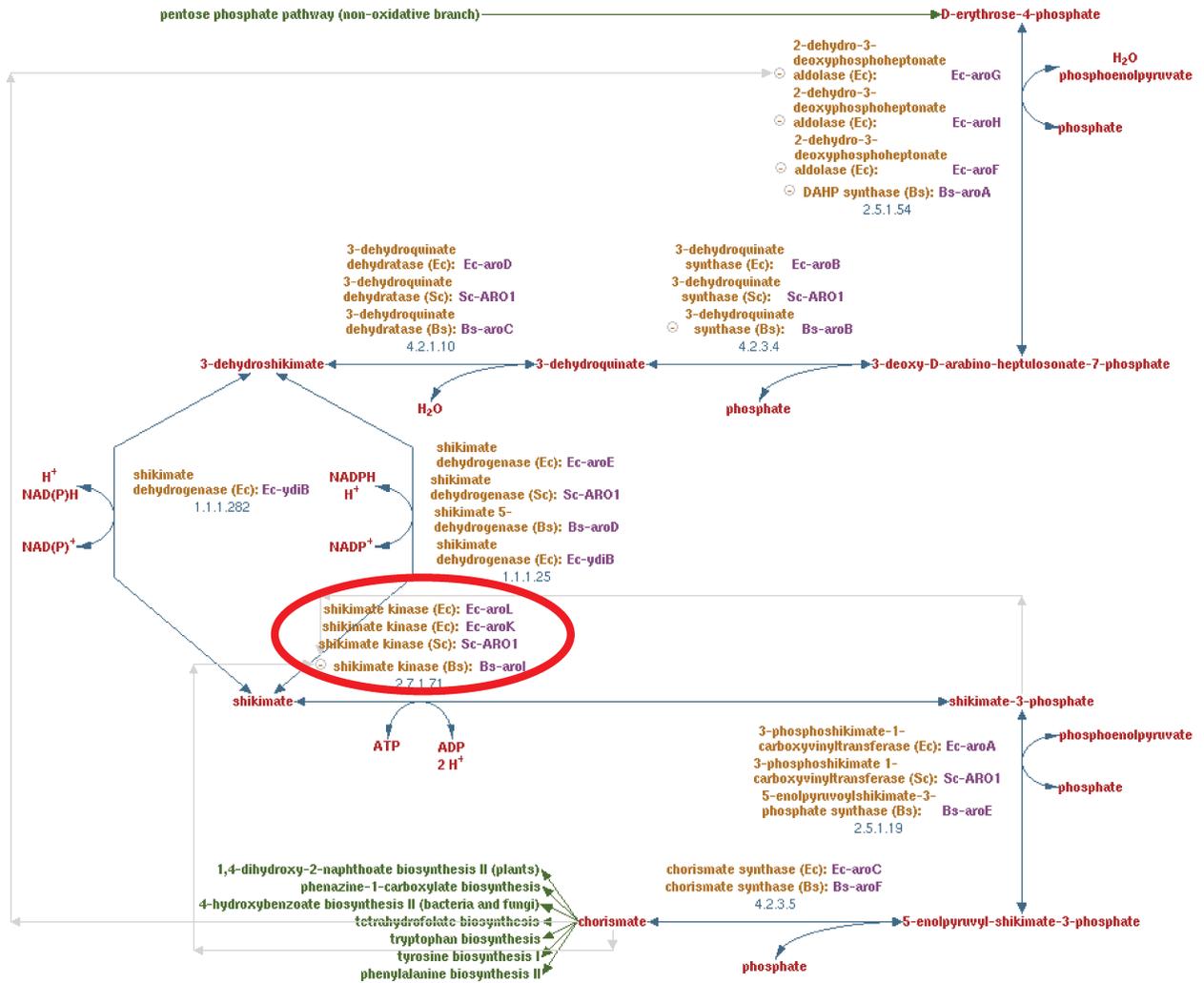


Figura 4.8: Via metabólica da biossíntese do corismato em organismos diversos. A enzima de interesse, em destaque, catalisa a conversão reversível de chiquimato a chiquimato 3-fosfato, uma das etapas da síntese do corismato.

Fonte: <http://biocyc.org/META/NEW-IMAGE?type=REACTION&object=SHIKIMATE-KINASE-RXN&redirect=T>

4.2.1. Procura e seleção do molde

A busca no BLAST contra o PDB por sequências homólogas à ORF GDI0293 resultou em um “hit” específico GntK, referente às gluconato quinases, e a uma super família conservada, das nucleotídeo/nucleosídeo quinases (NK). Essa última consiste de múltiplas famílias de enzimas que compartilham similaridade estrutural e estão funcionalmente relacionadas à catálise de transferências reversíveis de grupos fosfato

de nucleosídeos trifosfatos para nucleosídeos monofosfatos. (Marchler-Bauer *et. al.*, 2009).

Apenas dois alinhamentos significativos foram identificados, os quais representavam a mesma sequência de nucleotídeos, de códigos 1KNQ e 1KO4. Ambos são referentes à enzima Gluconato quinase de *Escherichia coli*, que possui três estruturas relacionadas, 1KO5, 1KO8 e 1KOF. Estas representam a enzima complexada com três diferentes ligantes, ATP (Adenosina 5-trifosfato), GP6 (gluconato 6-fosfato) e AMPPCP (Ácido fosfo metilfosfônico), respectivamente.

Gluconato quinase (2.7.1.12) é uma enzima envolvida na via de degradação de ácido orgânico (D-gluconato), e possui topologia similar às NMP, como a chiquimato quinase. Essa enzima é responsável por catalisar a conversão de D-gluconato a 6-fosfo-D-gluconato na via de metabolismo de cetogluconatos, como mostrado na Figura 4.9.

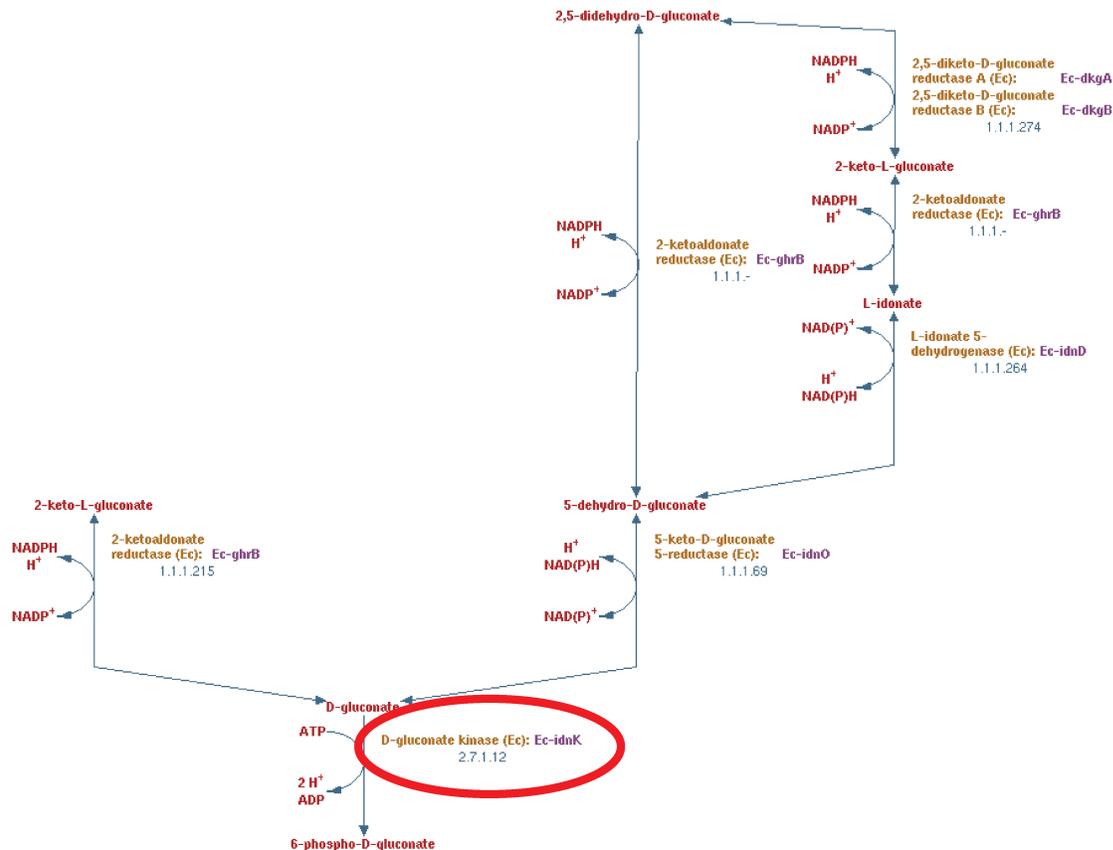


Figura 4.9: Via de metabolismo de cetogluconatos em *E. coli*. A enzima de interesse, circulado, catalisa a conversão de D-gluconato a 6-fosfo-D-gluconato.

Fonte: <http://biocyc.org/META/NEW-IMAGE?type=PATHWAY&object=KETOGLUCONMET-PWY>

A semelhança entre a ORF estudada e a gluconato quinase nos incentivou a buscar maiores informações a respeito da anotação desta sequência. No segundo genoma sequenciado de *G. diazotrophicus*, consta que a sequência em questão refere-se a uma carboidrato quinase, grupo no qual se insere a gluconato quinase. Foi sugerido aos pesquisadores da EMBRAPA Agrobiologia que fosse realizada uma análise da anotação deste gene, pois a gluconato quinase possui papel fundamental no metabolismo de ácidos orgânicos e sua identificação resultaria em uma otimização do projeto. As características do molde escolhido estão descritas na tabela 4.4.

Tabela 4.4: Características analisadas para escolha do molde adequado da ORF GDI0293

Molde 1KNQ/1KO5/1KO8/1KOF	
Produto	Gluconato quinase
Tamanho da sequência	165 aa
Identidade/Similaridade	42% / 61%
Gaps	2 (1%)
Domínios Conservados	Superfamília NK / <i>Hit</i> específico: GntK
Resolução	2.0Å
Resíduos faltantes	1, 2, 174, 175
Estrutura sem substrato	Sim – 1KNQ
Estrutura com substrato	Sim – 1KO5, 1KO8, 1KOF
Ligantes	ATP (1KO5), 6PG (Gluconato 6-fosfato – 1KO8), ACP (Ácido fosfo metilfosfônico – 1KOF)

4.2.2. Alinhamento de sequências e interação com os ligantes

Foi realizada uma análise dos alinhamentos gerados, e o melhor resultado encontra-se demonstrado na Figura 4.10. As considerações aplicadas na primeira ORF foram aplicadas em todas as outras. Estas se encontram na página 43. As interações com os ligantes são demonstradas nas Figuras 4.11, 4.12 e 4.13.



Figura 4.10: Alinhamento das seqüências da ORF GDI0293 e o molde escolhido, 1KNQ. As cores representam as características dos resíduos. Os aminoácidos destacados com setas foram identificados como participantes na interação com o substrato. As setas vermelhas indicam interação com ATP (adenosina trifosfato), as verdes com 6PG (Gluconato 6-fosfato) e as azuis com ACP (ácido fosfo metilfosfônico).

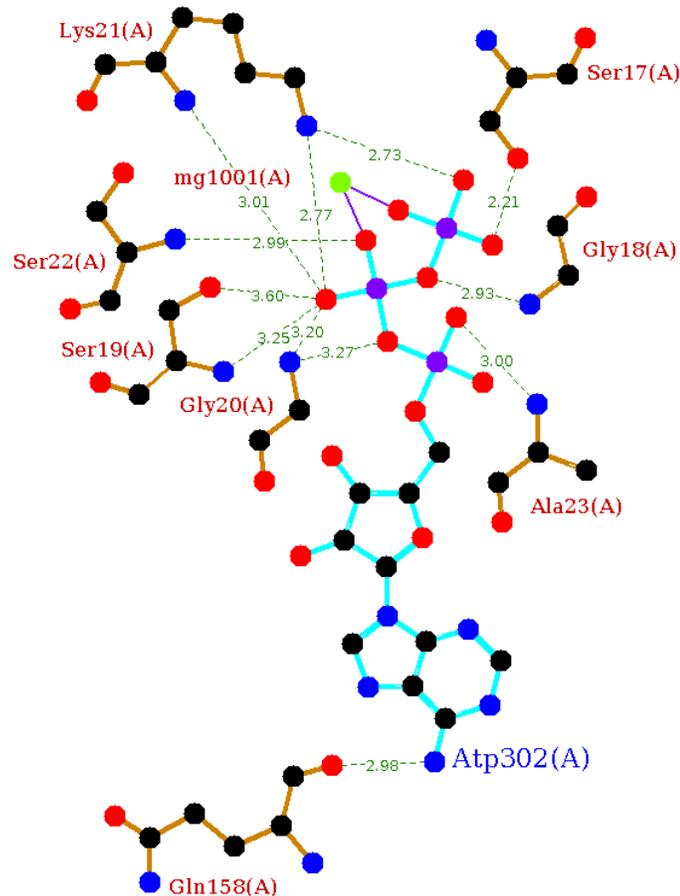


Figura 4.11: Representação esquemática da relação proteína-ligante do PDB 1KO5. O ligante ATP está representado em azul, e os aminoácidos que interagem com ele, em marrom. As distâncias entre os átomos são demonstradas nos tracejados verdes, e as esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

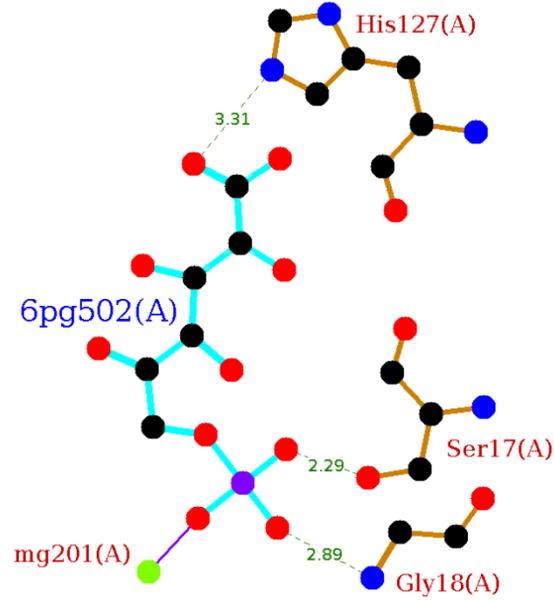


Figura 4.12: Representação esquemática da relação proteína-ligante do PDB 1KO8. O ligante 6PG (Gluconato 6-fosfato) está representado em azul, e os aminoácidos que interagem com ele, em marrom. Distâncias entre átomos em verde, e as esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

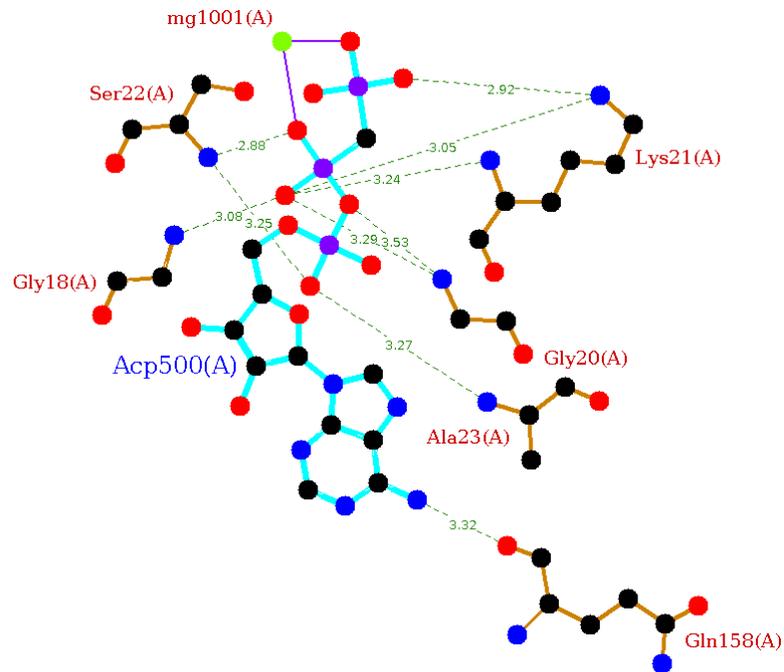


Figura 4.13: Representação esquemática da relação proteína-ligante do PDB 1KOF. O ligante ACP (Ácido fosfo metilfosfônico) está representado em azul, e os aminoácidos que interagem com ele, em marrom. Distâncias entre átomos em verde, e as esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

Analisando as Figuras 4.11 a 4.13, podemos identificar os resíduos que interagem com a enzima em questão. Estes foram marcados no alinhamento (Figura 4.10), com o objetivo de verificar a conservação na sequência-alvo. Os resultados estão representados na tabela 4.5, onde os asteriscos (*) representam a conservação do resíduo.

Tabela 4.5: Resíduos que interagem com os ligantes no PDB 3CXT.

ATP		6PG		ACP	
1KO5	GDI0293	1KO8	GDI0293	1KOF	GDI0293
Ser 17	*	Ser 17	*		
Gly 18	*	Gly 18	*	Gly 18	*
Ser 19	(C)				
Gly 20	*			Gly 20	*
Lys 21	*			Lys 21	*
Ser 22	(T)			Ser 22	(T)
Ala 23	(T)			Ala 23	(N)
		His 127	*		
Gln 158	(N)			Gln 158	(N)

Analisando a tabela e as propriedades dos aminoácidos em questão, podemos presumir que a proteína alvo possui estrutura similar à proteína molde. Isso ocorre devido ao fato de que muitos aminoácidos estão conservados nas sequências, e as mutações ocorridas são conservativas, ou seja, mantém as características principais dos resíduos envolvidos. Foi realizada uma análise mais detalhada do papel de cada um deles na interação com o substrato, fundamental para a indicação de mutações significantes. Esta será abordada mais adiante.

4.2.3. Modelagem comparativa e dinâmica molecular

O resultado desta etapa está representado na Figura 4.14, que retrata a sobreposição da estrutura molde e do modelo construído.

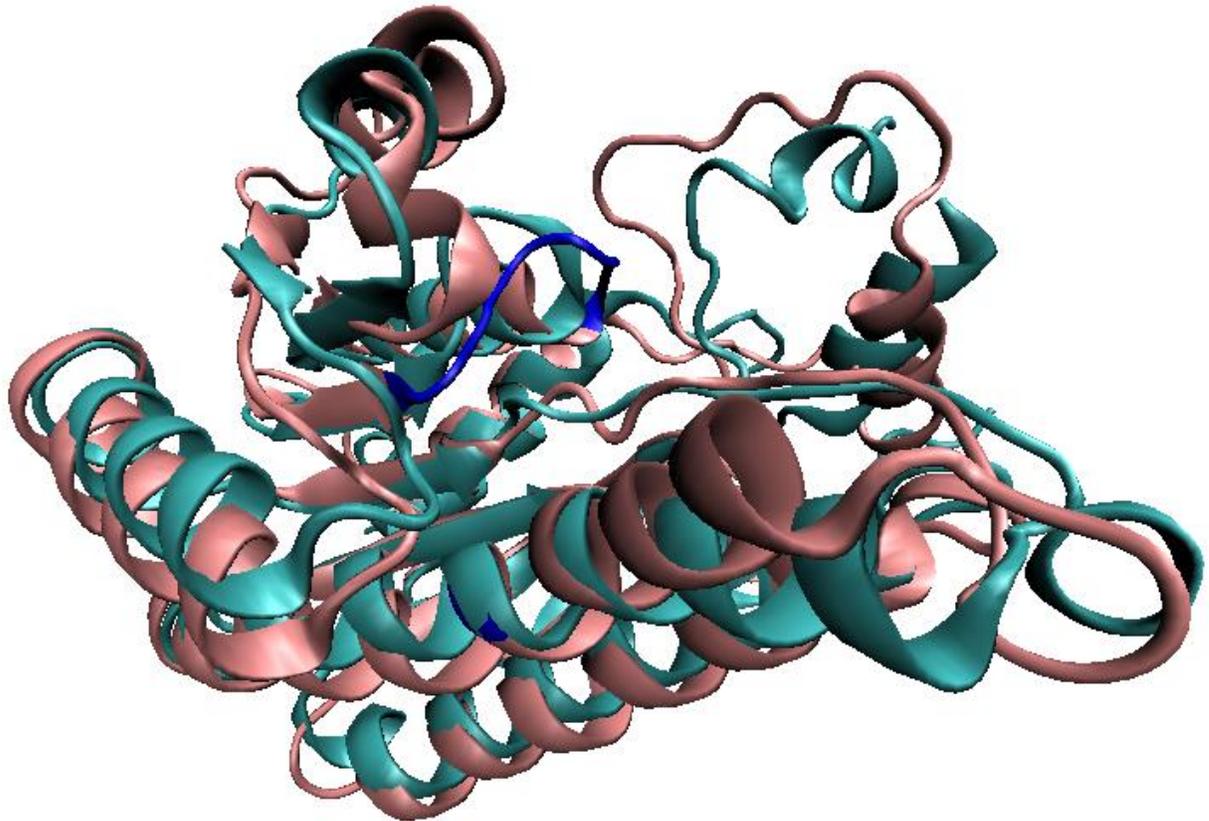
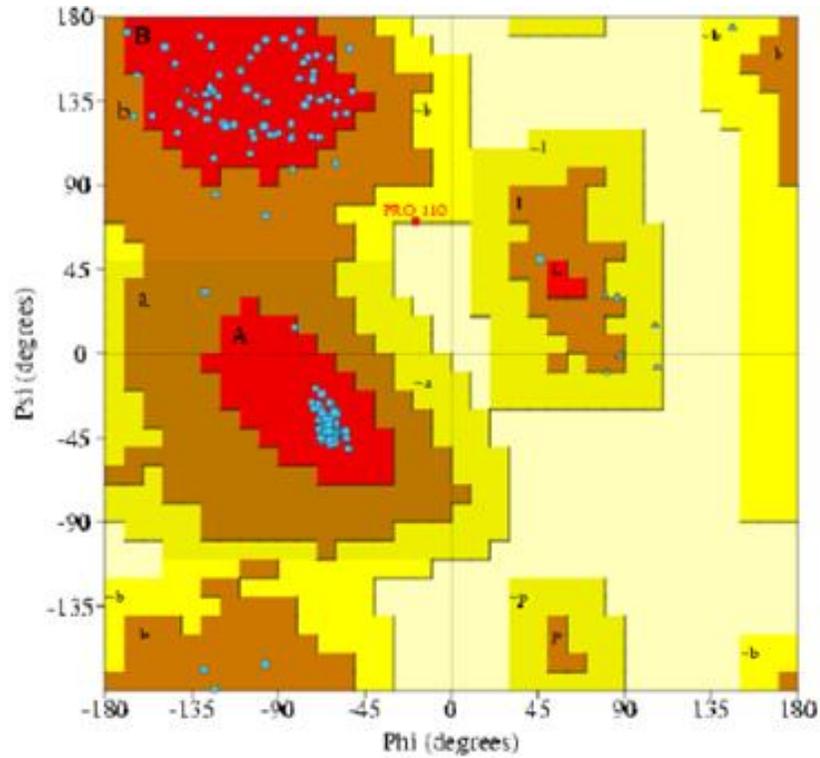


Figura 4.14: Alinhamento estrutural do molde (em azul) e do melhor modelo (em rosa), após dinâmica molecular. Os resíduos envolvidos na interação com os ligantes estão destacados.

Figura preparada com a ferramenta Swiss-pdb Viewer.

4.2.4. Validação do modelo

As figuras 4.15, 4.16 e 4.17 demonstram a qualidade do modelo obtido para a ORF GDI0293.



		Número de resíduos	%
Regiões mais favoráveis	[A, B, L]	139	92,1 %
Regiões permitidas	[a, b, l, p]	12	7,9 %
Regiões generosamente permitidas	[~a, ~b, ~l, ~p]	0	0,0 %
Regiões não permitidas		0	0,0 %
		-----	-----
Resíduos não prolina e não glicinas		151	100%
Resíduos finais (exc Gly e Pro)		2	
Resíduos Glicina		15	
Resíduos Prolina		14	

Total de Resíduos		182	

Figura 4.15: Gráfico de Ramachandran do modelo construído para a ORF GDI0293. As regiões em vermelho são mais favoráveis, e um modelo com mais de 90% dos resíduos nesta região é considerado adequado.

Fonte: PdbSum/PROCHECK

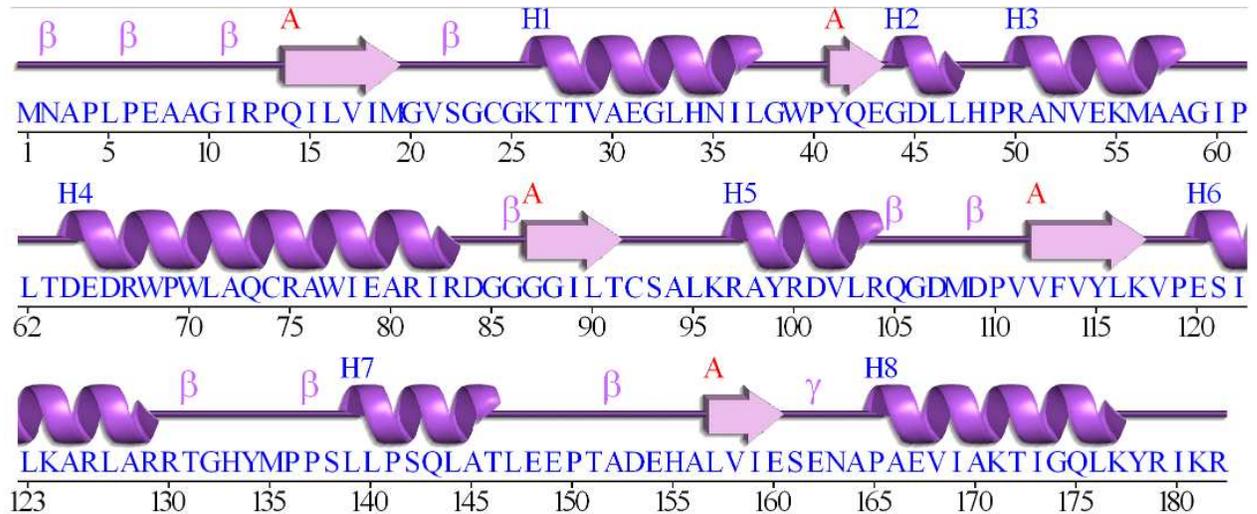


Figura 4.16: Predição de estruturas secundárias do modelo construído para a ORF GDI0293. As regiões em roxo representam α -hélices e as setas na cor rosa indicam fitas- β .

Fonte: PdbSum/PROCHECK

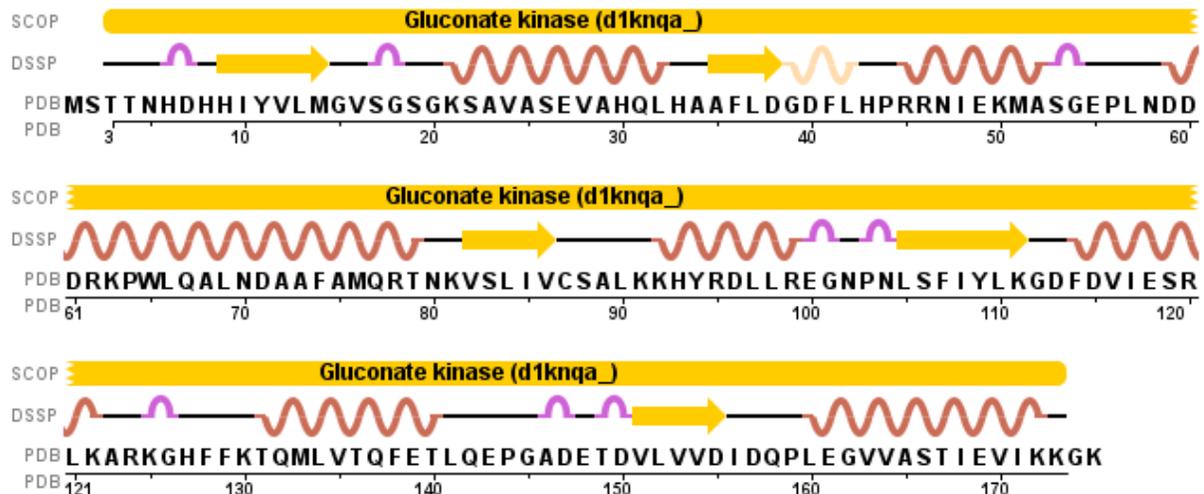


Figura 4.17: Predição de estruturas secundárias do molde 1KNQ. As α -hélices são representadas em rosa e vermelho escuro, e as fitas- β em setas amarelas.

Fonte: PDB

Após as diversas análises apresentadas neste item, o modelo final construído pode ser considerado adequado para a continuação dos experimentos. Esse apresenta uma estrutura similar ao molde, mantendo os resíduos importantes conservados, assim

como suas estruturas secundárias. Os dados obtidos a respeito da estrutura são suficientes para a realização da etapa final do estudo, a sugestão de sítios de mutação.

4.2.5. Mecanismo catalítico

A análise do mecanismo catalítico da enzima de interesse é essencial para a definição de mutações sítio dirigidas com um objetivo específico. O modelo estudado foi proposto por Kraft e colaboradores, e é relativo ao mecanismo catalítico de uma gluconato quinase.

Os substratos possuem diferentes sítios de ligação, e a interação desses parece ocorrer de forma ordenada. O sítio de ligação do ATP é acessível sem a ligação do gluconato. Esse, por sua vez, parece ser incapaz de se ligar a seu sítio ativo antes da ligação com o ATP. Após a ligação do ATP há uma mudança conformacional considerável, que resulta em uma ampliação do sítio ativo. Esta conformação mais aberta possibilita o acesso do gluconato a seu sítio ativo.

A gluconato quinase catalisa a reação posicionando o ATP e o gluconato em uma posição favorável para a reação. O íon magnésio auxilia na orientação do grupo γ -fosforil alinhado ao gluconato, criando a geometria correta para a transferência do fosforil (Walker *et. al.*, 1982). A cadeia lateral da lisina 21 conservada é essencial para a transferência do fosforil, estabilizando sua carga durante a catálise (Kraft *et. al.*, 2002).

CAPÍTULO 5 - RESULTADOS

Diversos estudos foram realizados com o objetivo de sugerir substituições interessantes nas sequências alvo envolvidas com a via metabólica de síntese de ácido glicônico da bactéria *Gluconacetobacter diazotrophicus*. Primeiramente, foram feitos estudos a respeito da via metabólica de síntese de ácido glicônico e seus cetoderivados. Este passo foi determinante na escolha das enzimas que seriam utilizadas para otimizar sua produção. Posteriormente, foi feita uma análise do mecanismo catalítico das enzimas alvo, para identificação dos resíduos essenciais para seu funcionamento.

Depois de identificar os principais alvos de mutação, estes foram substituídos, um a um na sequência, para verificar, *in silico*, as consequências das alterações. Para finalizar, foram feitas predições de ancoragens dos ligantes nas estruturas modificadas, para analisar possíveis efeitos que possam ocorrer *in vivo*, otimizando a síntese do produto de interesse. A análise do mecanismo catalítico foi descrita no capítulo 4, itens 4.1.5 e 4.2.5, e os outros itens serão descritos a seguir, nos itens 5.1 e 5.2.

5.1. Análise da via metabólica

A via de produção de ácido glicônico e seus cetoderivados não está disponível em bancos de dados, quando se trata da bactéria *Gluconacetobacter diazotrophicus*. Entretanto, essa via já foi estabelecida em outros organismos, podendo ser utilizadas para inferir sobre a síntese na bactéria de interesse. Foram feitas adaptações às vias já depositadas, juntamente com pesquisadores colaboradores da Embrapa Agrobiologia, agregando a elas as informações disponíveis a respeito do microorganismo em questão. Estas análises originaram uma hipótese da via metabólica de ácido glicônico em *G. diazotrophicus*, a qual está demonstrada na Figura 5.1.

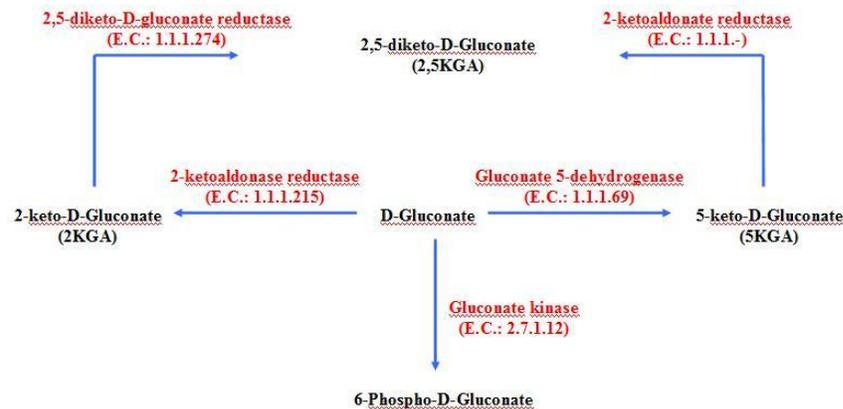


Figura 5.1: Sugestão da via metabólica de ácido glicônico e cetoderivados em *G. diazotrophicus*. O Gluconato é convertido em diversos produtos, como 2KGA e 5KGA. Estes, por sua vez, se convertem em outro cetogluconato.

Nesta hipótese, podemos notar que o ácido glicônico, representado como *D-gluconate*, passa por diversas reações para ser convertido em outros produtos. Isto faz com que a concentração de ácido glicônico no organismo de interesse seja pequena. A partir da análise desta via, foram determinadas quais enzimas seriam mais efetivas na síntese do ácido em questão. Estas já foram citadas anteriormente, e consistem de gluconato quinase e gluconato 5-desidrogenase. O objetivo é impedir a conversão de gluconato em 6-fosfo-D-gluconato e 5-ceto-D-gluconato, consequentemente aumentando a concentração de ácido glicônico na bactéria. A Figura 5.2 representa o objetivo deste estudo.

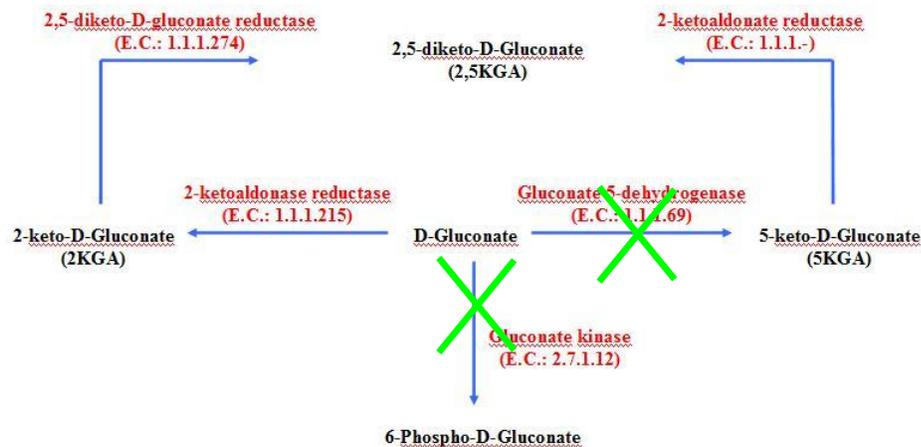


Figura 5.2: Objetivo do estudo. Inativando as enzimas que realizam a conversão de ácido glicônico em outros ácidos, é possível aumentar a concentração deste ácido na bactéria. As marcações em verde indicam os locais escolhidos para mutagênese sítio dirigida.

5.2. Mutagênese sítio-dirigida

Após definir as enzimas mais indicadas para otimizar a produção de ácidos orgânicos, analisar suas estruturas e estudar seus mecanismos catalíticos, foi possível sugerir algumas mutações sítio-dirigidas. Com o objetivo de otimizar estes resultados, foram feitas algumas simulações computacionais visando a obtenção de hipóteses a respeito das mudanças que irão ocorrer *in vivo*. Primeiramente, foi construída uma estrutura do melhor modelo com os ligantes do molde. Depois foram feitas as alterações dos resíduos importantes, um a um, por alaninas. As mudanças conformacionais visualizadas serviram de base para definir os melhores alvos para mutação.

5.2.1. ORF GDI0691

O molde utilizado para a construção dos modelos da ORF em questão, a estrutura cristalográfica da gluconato 5-desidrogenase de *Streptococcus suis* (PDBID 3CXR), possui uma estrutura relacionada depositada no PDB, complexada com dois ligantes: NAP (NADP Nicotinamida adenina dinucleotídeo fosfato) e GCO (ácido glicônico) (PDBID: 3CXT). A interação dos ligantes com a enzima é feita por meio dos resíduos representados na tabela 5.1.

Tabela 5.1: Resíduos que interagem com os ligantes na estrutura 3CXT.

NAP	GKR
Tyr 24	
Ile 26	
Asp 45	
Asp 71	
Gly 100	
	Arg 104
	Arg 157
Tyr 163	Tyr 163
Lys 167	

Utilizando a ferramenta Swiss-Pdb Viewer, foi construída uma estrutura na qual se encontra o modelo construído por modelagem e os ligantes de seu respectivo molde. O objetivo era analisar se os ligantes se posicionariam na mesma posição, e quais as interações seriam modificadas. O resultado está representado na Figura 5.3.

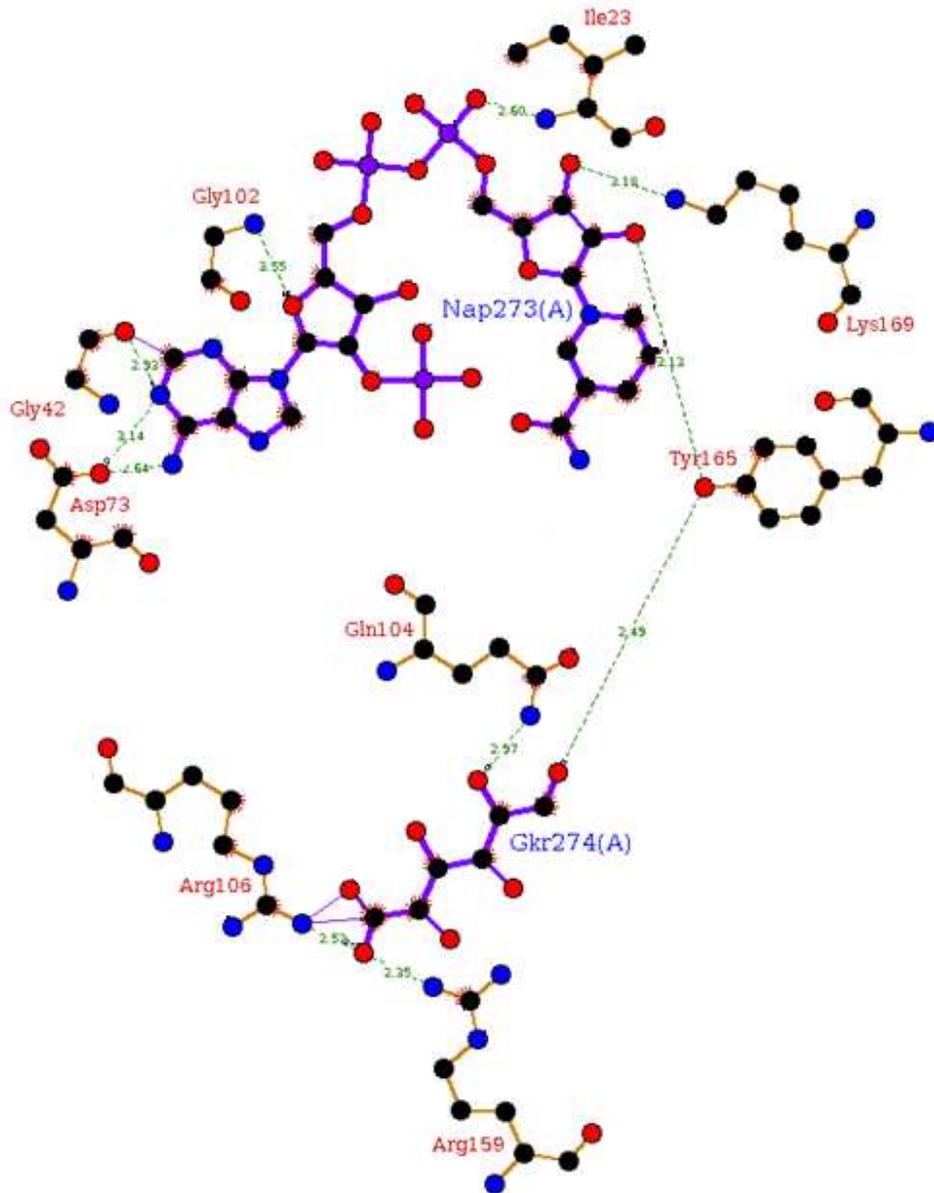


Figura 5.3: Representação esquemática da interação modelo-ligante. Os ligantes Nap e Gkr são representados em roxo, e os aminoácidos que interagem com eles, em marrom. As distâncias entre os átomos são demonstradas pelos tracejados verdes, e as esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

Analisando a figura, e considerando o alinhamento das duas sequências, podemos notar que a maioria das interações foi mantida após a ancoragem dos ligantes no modelo construído por modelagem. A tabela 5.2 demonstra a conservação das interações.

Tabela 5.2: Comparação da interações dos ligantes com o molde e o modelo.

NAP		GKR	
3CXT	Modelo	3CXT	Modelo
Tyr 24	--		
Ile 26	Ile 23		
Asp 45	--		
Asp 71	Asp 73		
Gly 100	Gly 102		
		Arg 104	Arg 106
		Arg 157	Arg 159
Tyr 163	Tyr 165	Tyr 163	Tyr 165
Lys 167	Lys 169		

Estes resíduos são considerados importantes para o funcionamento da molécula, e são fortes candidatos à mutagênese. Entretanto, para refinar as mutações sugeridas, foi feito um estudo do mecanismo catalítico, relatado no item 4.1.5. Os resíduos da téttrade catalítica encontrada no molde foram selecionados, sendo eles: Arginina 104, Serina 152, Tirosina 165 e Lisina 167. A serina, embora não tenha interação direta, possui um papel importante no reconhecimento do substrato. Estes foram modificados um a um na sequência, e foram feitas análises para verificar o impacto destas alterações na estrutura ou funcionamento da molécula. Os resíduos sugeridos para realização da mutagênese sítio dirigida são a Arginina 106 e a Serina 152. Suas análises serão demonstradas nos itens 5.2.1.1 e 5.2.1.2.

5.2.1.1. Arginina 106

A mutação da Arginina 106 por uma alanina resultou em uma perda das interações da tétrade catalítica com os ligantes. A função deste resíduo está relacionada com o reconhecimento do substrato, sendo que, provavelmente, sua carga positiva facilita a orientação e o acesso do substrato, carregado negativamente. Além disso, um estudo comprovou uma atividade catalítica 20% menor do que a da enzima nativa quando ocorre esta mutação (Zhang *e. al.*, 2009). As mudanças na estrutura e na interação com o substrato estão representadas nas Figuras 5.4 e 5.5.

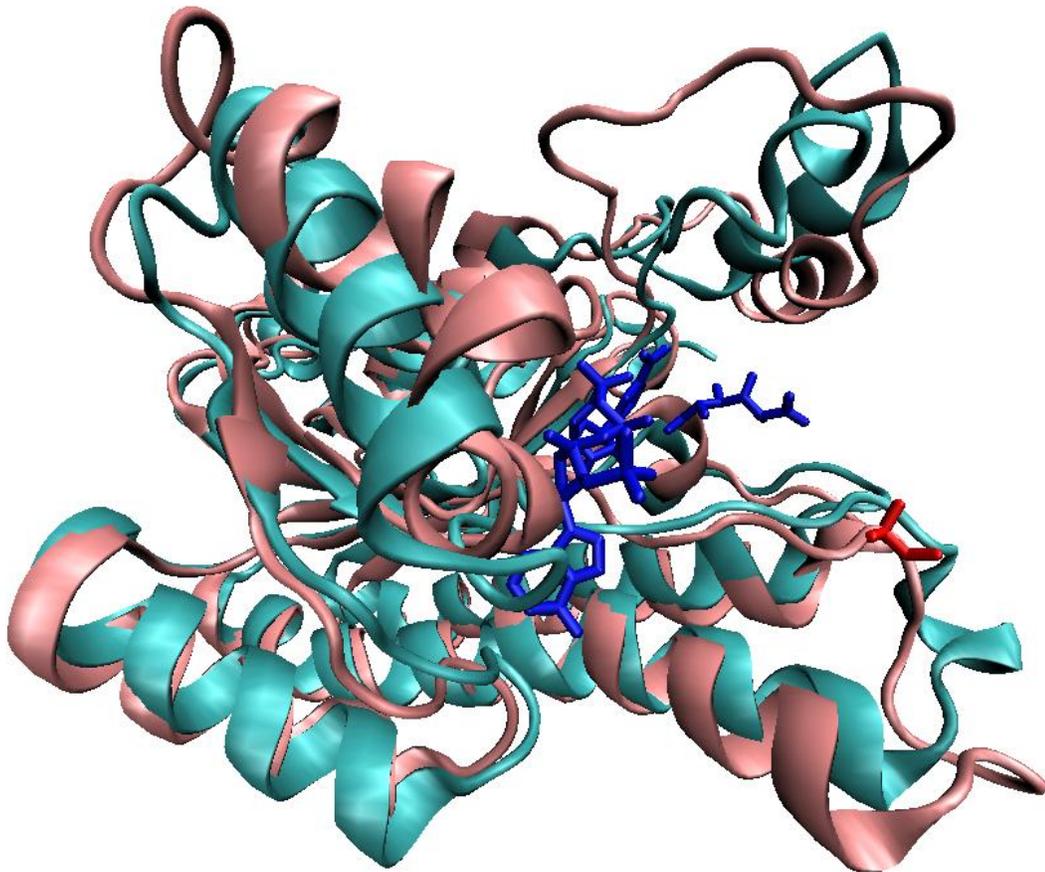


Figura 5.4: Sobreposição da Ga5DH antes e após a substituição da Arginina 106 por uma alanina. O modelo em azul representa a enzima antes da substituição, e o modelo em rosa representa o modelo após a modificação. Os ligantes estão representados em azul escuro, e a alanina alterada em vermelho.

Figura preparada com a ferramenta Swiss-Pdb Viewer

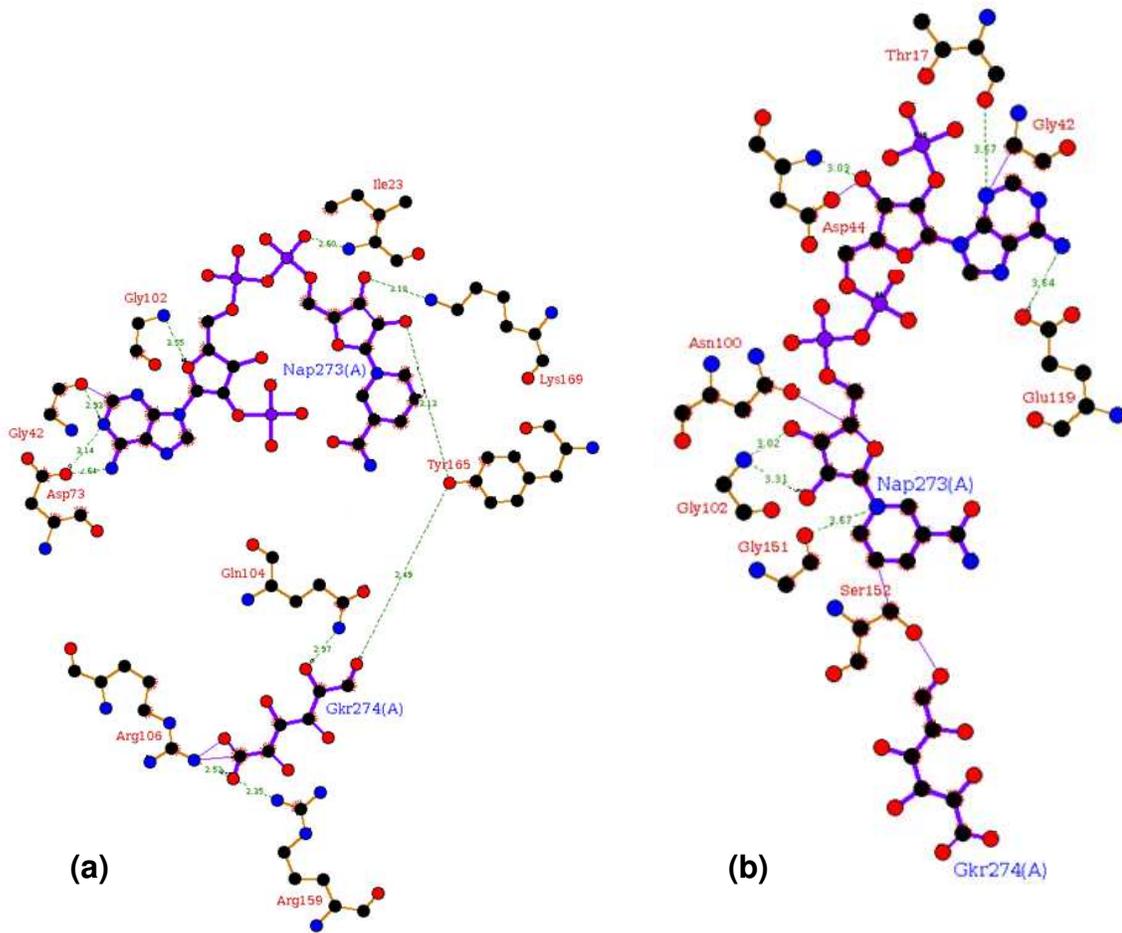


Figura 5.5: Representação esquemática antes (a) e depois (b) da interação do modelo com os ligantes após a substituição da Arginina 106 por uma alanina. As interações com os resíduos da tétrede catalítica foram totalmente perdidas. Os ligantes estão representados em azul, e os aminoácidos que interagem com ele, em marrom. Distância entre os átomos representadas em verde, e esferas azuis, vermelhas e pretas representam nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

5.2.1.2. Serina 152

A serina da tétrede catalítica das enzimas SDR, embora não interaja diretamente com os ligantes, também possui um papel importante no reconhecimento do substrato. Estudos de Opperman e colaboradores, em 2003, demonstraram que uma substituição deste resíduo por uma alanina resulta em uma perda quase total das atividades

enzimáticas. Esta substituição, quando realizada *in silico*, resultou em uma mudança considerável na estrutura (Figura 5.6), e em uma grande perda na interação com os substratos, principalmente com o gluconato (Figura 5.7), sustentando a ideia de que este resíduo é essencial para a atividade enzimática em questão.

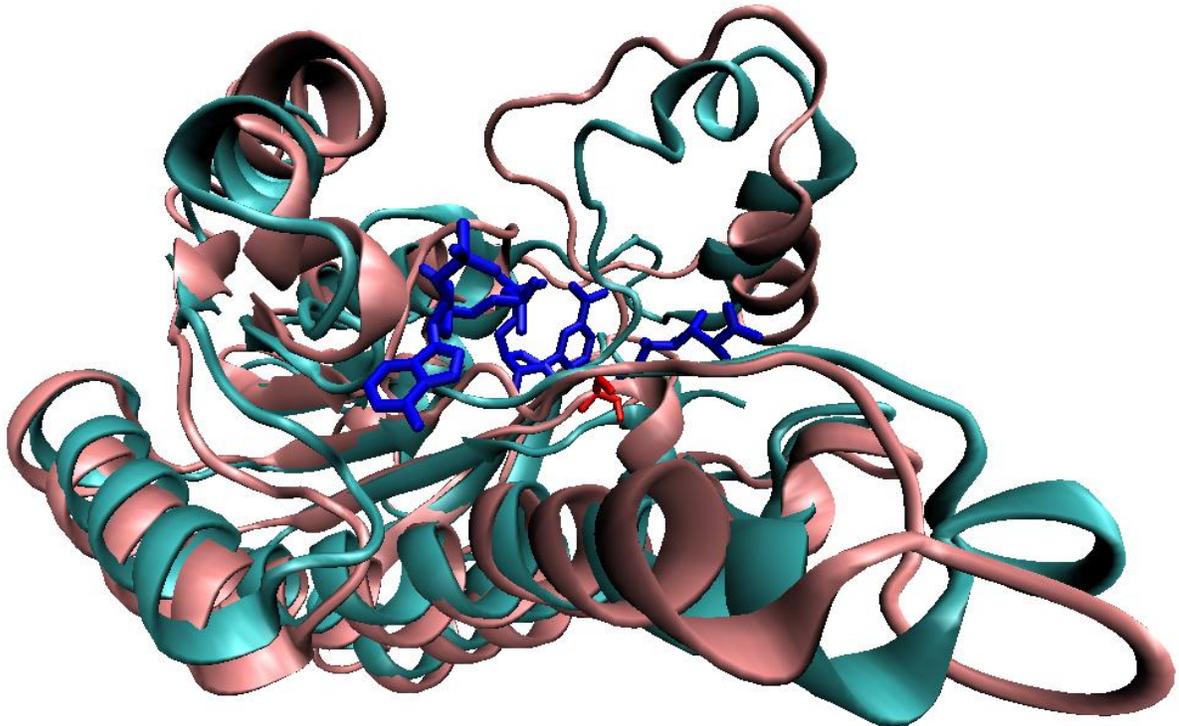


Figura 5.6: Sobreposição da Ga5DH antes e após a substituição da Serina 152 por uma alanina. O modelo em azul representa a enzima antes da substituição, e o modelo rosa representa o modelo após a modificação. Os ligantes estão representados em azul escuro, e a alanina alterada em vermelho.

Figura preparada com a ferramenta Swiss-Pdb Viewer

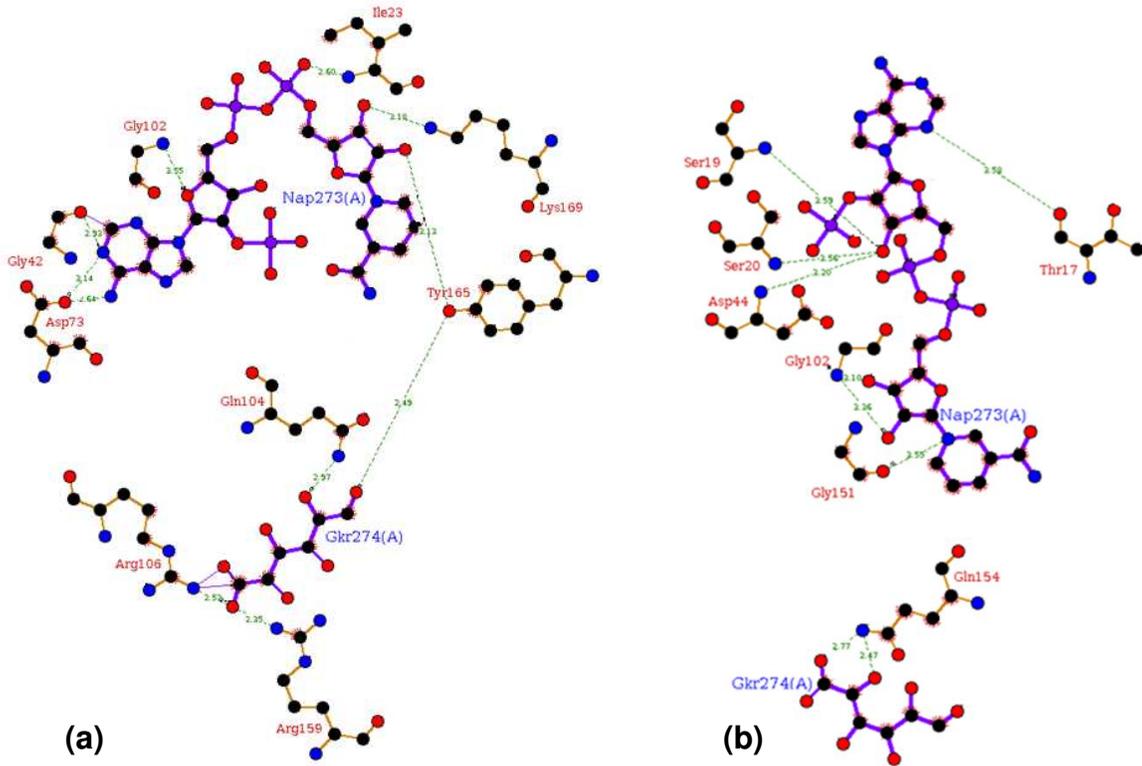


Figura 5.7: Representação esquemática da interação do modelo com os ligantes antes (a) e após (b) a substituição da Serina 152 por uma alanina. Os ligantes estão representados em roxo, e os aminoácidos que interagem com ele, em marrom. Distância entre os átomos representadas em verde, e esferas azuis, vermelhas e pretas representando nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

5.2.2. ORF GDI0293

O molde utilizado para a construção dos modelos desta ORF (PDBID 1KNQ) possui três estruturas relacionadas, complexadas com diferentes ligantes: 1KO5 (ATP), 1KO8 (6PG – Gluconato 6-fosfato) e 1KOF (ACP – Ácido fosfo metilfosfônico). Os substratos ACP e ATP foram considerados muito similares, e por isso somente um deles foi relatado no artigo destas estruturas (Kraft *et. al.*, 2002). Estes substratos interagem com a enzima por meio dos resíduos representados na tabela 5.3.

Tabela 5.3: Resíduos que interagem com os ligantes nas estruturas 1KO5 e 1KO8.

ATP	6PG
Ser 17	Ser 17
Gly 18	Gly 18
Ser 19	
Gly 20	
Lys 21	
Ser 22	
Ala 23	
	His 127
Gln 158	

Como no item anterior, foi construído um modelo com os ligantes citados acima, com o objetivo de analisar as mudanças nas interações entre eles. O resultado está representado nas Figuras 5.8 e 5.9.

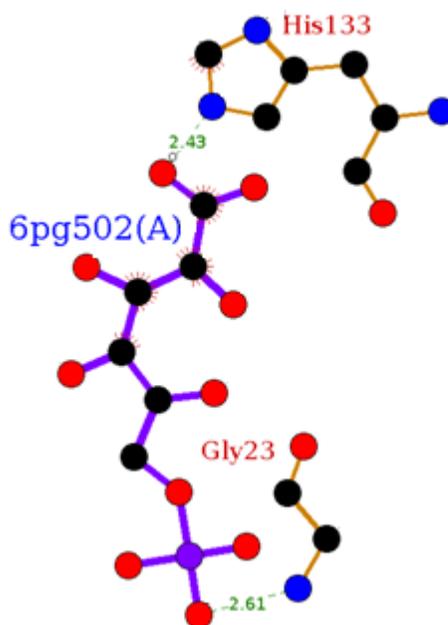


Figura 5.8: Representação esquemática da interação modelo-ligante. O ligante 6PG está representado em roxo, e os aminoácidos que interagem com ele, em marrom. As distâncias entre os átomos são demonstradas pelos tracejados verdes, e esferas azuis, vermelhas e pretas representando nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

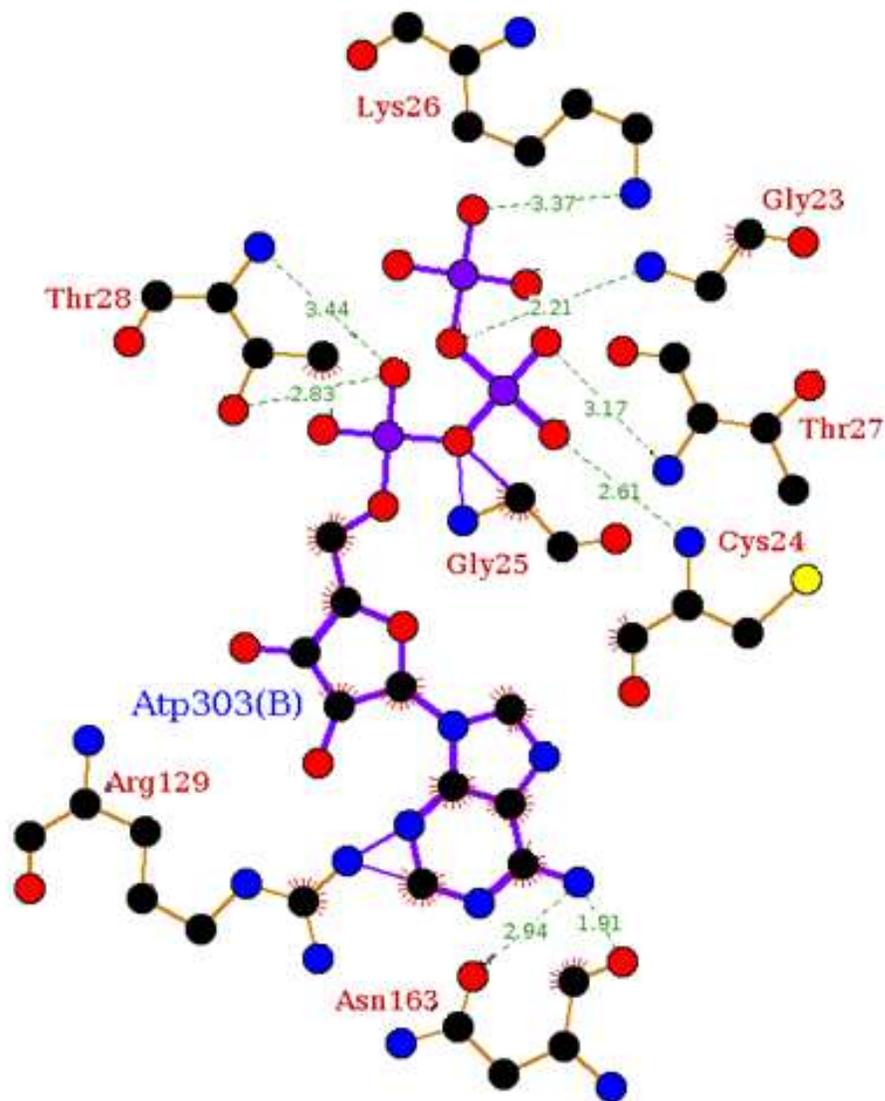


Figura 5.9: Representação esquemática da interação modelo-ligante. O ligante ATP está representado em roxo, e os aminoácidos que interagem com ele, em marrom. As distâncias entre os átomos são demonstradas pelos tracejados verdes, e esferas azuis, vermelhas e pretas representando nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

Analisando as figuras e considerando o alinhamento das sequências, podemos notar que a maioria das interações foi mantida após a ancoragem dos ligantes no modelo construído por modelagem. Na Tabela 5.4 estão listados os resíduos que interagem com os substratos nas estruturas molde, e a conservação destas interações no modelo construído.

Tabela 5.4: Comparação da interações dos ligantes com o molde e o modelo.

ATP		6PG	
1KO5	Modelo	1KO8	Modelo
Ser 17	--	Ser 17	
Gly 18	Gly 23	Gly 18	Gly 23
Ser 19	Cys 24		
Gly 20	Gly 25		
Lys 21	Lys 26		
Ser 22	Thr 27		
Ala 23	Thr 28		
		His 127	His 133
Gln 158	Asn 163		

Estes resíduos são considerados importantes para o funcionamento da molécula, e são fortes candidatos a mutagênese. Para refinar as sugestões de mutações, foi feito um estudo do mecanismo catalítico, relatado no item 4.1.5. A mutagênese foi aplicada em todos os resíduos, e o resultado mais significativo foi com a Histidina 133. Esta é considerada essencial para a transferência do substrato, e suas análises serão demonstradas no item 5.2.2.1.

5.2.2.1. Histidina 133

A substituição da Histidina 133 por uma alanina não alterou consideravelmente a interação da enzima com o ATP, a maioria das interações foram mantidas. Entretanto, a mutação altera a principal interação da enzima com o gluconato, podendo resultar na diminuição ou até mesmo perda da atividade catalítica. Nota-se também que a mutação gera uma grande mudança estrutural no loop no qual a histidina se encontra, dificultando ainda mais a entrada do gluconato. As mudanças na estrutura e na interação com o substrato estão representadas nas Figuras 5.10, 5.11 e 5.12.

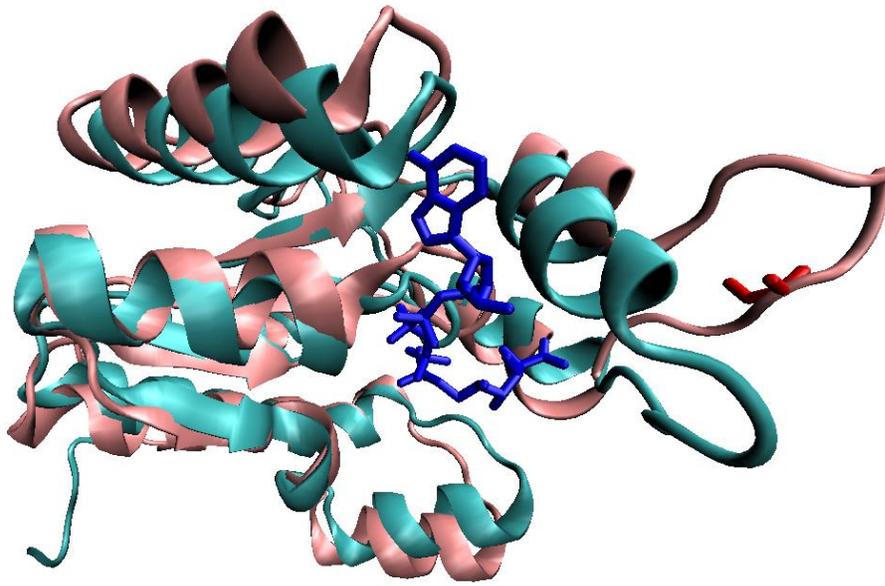


Figura 5.10: Sobreposição da Gluconato quinase antes e após a substituição da Histidina 133 por uma alanina. O modelo em azul escuro representa a enzima antes da substituição, e o modelo colorido por RMS representa o modelo após a modificação. A área de superfície e cadeias principal e lateral dos ligantes estão em rosa, e a alanina alterada em preto.

Fonte: Swiss-Pdb Viewer

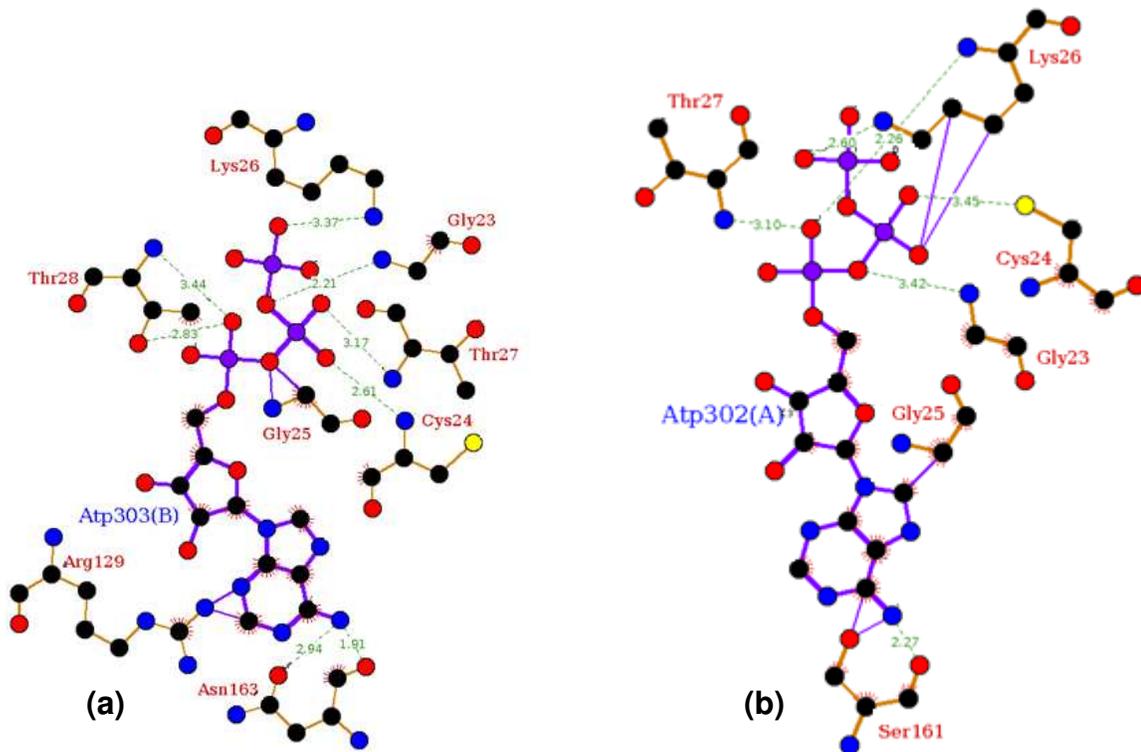


Figura 5.12: Representação esquemática da interação do modelo com o ligante 6PG antes (a) e após (b) a substituição da Histidina 133 por uma alanina. O ligante está representado em roxo, e os aminoácidos que interagem com ele, em marrom. Distância entre os átomos representadas em verde, e esferas azuis, vermelhas e pretas representando nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

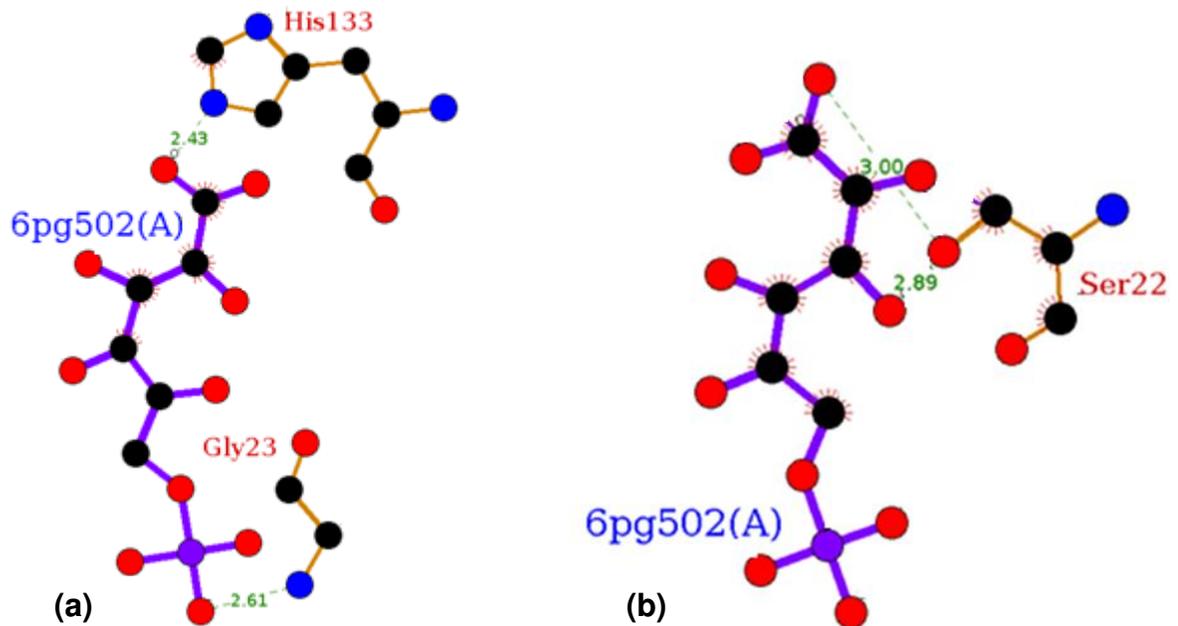


Figura 5.12: Representação esquemática da interação do modelo com o ligante 6PG antes (a) e após (b) a substituição da Histidina 133 por uma alanina. O ligante está representado em roxo, e os aminoácidos que interagem com ele, em marrom. Distância entre os átomos representadas em verde, e esferas azuis, vermelhas e pretas representando nitrogênio, oxigênio e carbono, respectivamente.

Figura preparada com a ferramenta LIGPLOT

CAPÍTULO 6 - CONCLUSÕES E DESENVOLVIMENTOS FUTUROS

A produção de ácidos orgânicos por microorganismos é um processo de grande impacto e aplicação comercial para a produção e conservação de alimentos, fármacos e síntese de produtos biodegradáveis. Dentre eles, destaca-se o ácido glicônico, que possui papel importante nas indústrias alimentícia, farmacêutica, têxtil, entre outras. Este é encontrado em diversas espécies de bactérias, assim como em plantas, frutas e outros produtos alimentícios.

Um exemplo de microorganismo produtor de ácido glicônico é a bactéria *Gluconacetobacter diazotrophicus*, que foi isolada inicialmente em cana-de-açúcar e posteriormente em outras plantas. Trata-se de uma diazotrófica associativa e endofítica, que exerce um papel importante na agricultura por assimilar nitrogênio para as plantas por meio da Fixação Biológica de Nitrogênio.

A capacidade de *Gluconacetobacter diazotrophicus* produzir ácidos orgânicos derivados da oxidação incompleta da glicose inclusive em condições de fixação de nitrogênio representa mais uma das suas potenciais aplicações biotecnológicas. No entanto, para garantir a eficiência desses processos biológicos é necessário entender os mecanismos envolvidos. A via metabólica do ácido glicônico, entretanto, é bastante complexa e, muitas vezes, inviável economicamente. Diante disso esse projeto foi realizado para buscar uma maneira de otimizar a produção destes ácidos nessa determinada bactéria. A opção sugerida é alterar as enzimas envolvidas na conversão de ácido glicônico em cetoderivados, para que aumente a concentração de gluconato nas reações. Esta alteração nas proteínas pode ser feita por meio de mutações sítio-dirigidas, entretanto, os processos experimentais exigem grande demanda de tempo e recursos financeiros.

Utilizamos então ferramentas de biologia computacional para realização de mutagênese *in silico* visando obter informações a respeito das consequências que essas causarão a uma proteína específica. No presente estudo, as enzimas de interesse não possuíam estrutura tridimensional resolvida, sendo necessário recorrer a uma técnica de predição estrutural, a modelagem por homologia. A partir desta, foram

construídos modelos confiáveis das sequências-alvo, e estes foram utilizados para inferir a respeito do funcionamento de cada enzima.

Após análise estrutural dos modelos construídos, foi realizado um estudo da via metabólica do ácido glicônico e de cada enzima envolvida. As proteínas consideradas melhores alvos para a mutagênese foram uma gluconato 5-desidrogenase e uma gluconato quinase, enzimas diretamente relacionadas com a conversão de ácido glicônico em cetogluconatos.

Uma vez definidas as enzimas potenciais, foi realizada a análise do mecanismo catalítico de cada uma, visando encontrar os resíduos catalíticos essenciais para seu funcionamento. Para entender a contribuição desses resíduos no mecanismo da enzima, foram feitas mutações *in silico* para alaninas e dinâmica molecular foi realizada nos mutantes. Os resultados da dinâmica revelam a variabilidade conformacional e os possíveis movimentos para os resíduos da enzima mutada, para que estes fossem comparados aos modelos sem alterações.

Uma das maneiras de se determinar a importância de um resíduo na proteína é substituí-lo e verificar as consequências dessa alteração. Se, após a alteração, houver uma grande mudança estrutural ou na interação com o substrato, o resíduo substituído deve ter grande importância no funcionamento da enzima. Estas análises também foram feitas *in silico*, e a partir delas foram obtidos os resultados do presente estudo.

A ORF GDI0691, referente a uma gluconato 5 desidrogenase, foi modelada com base em uma Ga5DH de outra bactéria, *Streptococcus suis*. Esta última possui dois ligantes, NAP e Gluconato. Relata-se que a interação com estes ligantes depende de uma téttrade catalítica, principal alvo, portanto, para a mutagênese sítio dirigida. Dentre os quatro resíduos substituídos, duas alterações produziram resultados mais significativos: Arginina 106 e Serina 152, ambos relacionados com o reconhecimento do substrato. A troca da arginina por alanina resultou em uma alteração na interação com o substrato, e estudos concluíram que esta alteração diminui a atividade catalítica da enzima. A substituição da serina, por outro lado, teve como consequência uma significativa mudança estrutural, e uma perda na interação com os substratos, principalmente o gluconato.

Em relação à ORF GDI0293, indica-se apenas um resíduo para mutagênese, a Histidina 133. Sua substituição não alterou consideravelmente a interação da enzima com o ligante ATP, entretanto a principal interação com o gluconato foi perdida, o que pode resultar na diminuição ou até mesmo perda da atividade catalítica. Além disso, a mutação gera uma grande mudança estrutural no loop no qual a histidina se encontra, dificultando ainda mais a entrada do gluconato.

Os resultados apresentados neste trabalho serão transmitidos aos pesquisadores da Embrapa Agrobiologia, que pretendem realizar estudos experimentais para comprovar a eficácia dos métodos computacionais. A Bioinformática sugere soluções, mas a comprovação experimental é necessária.

REFERÊNCIAS BIBLIOGRÁFICAS

- Alberghina, L. **Systems Biology: Definitions and Perspectives**. Topics in Current Genetics, 13: Springer: 2005;
- Altschul, S.; Gish, W.; Miller, W.; Myers, E.; Lipman, D. J. **Basic Local Alignment Search Tool**. Journal of Molecular Biology, 215: 403-410, 1990;
- Altschul, S.; Madden, T.; Schäffer, A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. **Gapped blast and psi-blast: a new generation of protein database search programs**. Nucleic Acids Research, 25(17): 3389-3402, 1997;
- Berman, H.; Henrick, K.; Nakamura, H.; Markley, J.L. **The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data**. Nucleic Acids Research, 35(Database issue):D301-3, 2007;
- Bertalan et al. **Complete genome sequence of the sugarcane nitrogen-fixing endophyte *Gluconacetobacter diazotrophicus* Pal5**. BMC Genomics, 10:450, 2009;
- Bjelic, S., Aqvist, J. **Computational prediction of structure, substrate binding mode, mechanism, and rate for a malaria protease with a novel type of active site**. Biochemistry, 43:14521–14528, 2004;
- Blanco, Y.; Blanch, M.; Pinon, D.; Legaz, M.E.; Vicente, C. **Antagonism of *Gluconacetobacter diazotrophicus* (a sugarcane endosymbiont) against *Xanthomonas albilineans* (pathogen) studied in alginate-immobilized sugarcane stalk tissues**. Journal of Bioscience and Bioengineering, 99:366–71, 2005;
- Bogan, A.A.; Thorn, K.S. **Anatomy of Hot Spots in Protein Interfaces**. Journal of Molecular Biology, 280: 1-9, 1998;
- Böhm, H.J.; Schneider, G. **Protein-Ligand Interaction: From Molecular Recognition to Drug Design**. Wiley-VCH, Weinheim, New York, 2003;
- Bromberg, Y.; Burkhard, R. **Comprehensive in silico mutagenesis highlights functionally important residues in proteins**. Bioinformatics. 24 (ECCB Proceedings): i207-i212, 2008;
- Bujnicki, J.; Nabieva; Singh. **Protein function prediction via analysis of interactomes**. Prediction of Protein Structures, Functions, and Interactions (ed), 2009;
- Burley, S.K.; Almo, S.C.; Bonanno, J.B.; Capel, M.; Chance, M.R.; Gaasterland, T.; Lin, D.; Sali, A.; Studier, F.W.; Swaminathan, S. **Structural genomics: Beyond the human genome project**. Nature Genetics, 23:151–157, 1999;
- Caffrey, C.R.; Placha, L.; Barinka, C.; Hradilek, M.; Dostal, J.; Sajid, M.; McKerrow, J.H.; Majer, P.; Konvalinka, J.; Vondrasek, J. **Homology modeling and SAR analysis of *Schistosoma japonicum* cathepsin D (SjCD) with statin inhibitors identify a unique active site steric barrier with potential for the design of specific inhibitors**. Biological Chemistry, 386:339–349, 2005;
- Chasman, D.; Adams, R.M. **Predicting the functional consequences of non-synonymous single nucleotide polymorphisms: structure-based assessment of amino acid variation**. Journal of Molecular Biology, 307: 683–706, 2001;

- Chmiel, A.A.; Radlinska, M.; Pawlak, S.D.; Krowarsch, D.; Bujnicki, J.M.; Skowronek, K.J. **A theoretical model of restriction endonuclease NlaIV in complex with DNA, predicted by fold recognition and validated by site-directed mutagenesis and circular dichroism spectroscopy.** Protein Engineering, Design and Selection, 18:181–189, 2005;
- Clackson, T.; Wells, J.A. **A hot spot of binding energy in a hormone-receptor interface.** Science, 267: 383–386, 1995;
- Coombs, G.S.; Corey, D.R. **Site-directed mutagenesis and protein engineering**, in Proteins, (Angeletti RH, ed.), Academic Press, San Diego, CA, pp. 259-311, 1998;
- Das, A.; Kundu, P.N. **Microbial production of gluconic acid.** Journal of Scientific & Industrial Research, 46:307–31, 1987;
- DeLano, W.L. **The PyMOL Molecular Graphics System.** California, USA: DeLano Scientific, San Carlos, 2002;
- Döbereiner, J.; Reis, V.; Lazarini, A.C. In: Bothe, H.; Briyn, F.J.; Newton, W.E.; editors. **Nitrogen Fixation.** Stuttgart, Cologne: Gustav Fisher, p. 717–22, 1988;
- Dufour, E.; Storer, A.C.; Ménard, R. **Engineering nitrile hydratase activity into a cysteine protease by a single mutation.** Biochemistry, 34:16382-16388, 1995;
- Elfari, M.; Ha, S.W.; Bremus, C.; Merfort, M.; Khodaverdi, V.; Herrmann, U.; Gorish, H. **A *Gluconobacter oxydans* mutant converting glucose almost quantitatively to 5-keto-D-gluconic acid.** Applied Microbiology and Biotechnology, Vol. 66, no. 6:668-674, 2005;
- Eramian, D.; Shen, M.Y.; Devos, D.; Melo, F.; Sali, A.; Marti-Renom, M.A. **A composite score for predicting errors in protein structure models.** Protein Science, 15:1653–1666, 2006;
- Eswar, N.; Webb, B.M.; Marti-Renom, M.; Madhusudhan, M.S.; Eramian, D.; Shen, M.Y.; Pieper, U.; Sali, A. **Comparative protein structure modeling using MODELLER.** Current Protocols in Protein Science, Chapter 2: Unit 2.9, 2007;
- Filling, C.; Berndt, K.D.; Benach, J.; Knapp, S.; Prozorovski, T.; Nordling, E.; Ladenstein, R.; Jörnvall, H.; Oppermann, U. **Critical residues for structure and catalysis in short-chain dehydrogenases/reductases.** Journal of Biological Chemistry, 277: 25677–25684, 2002;
- Fiser, A. **Template-based protein structure modeling.** Methods in Molecular Biology, 673:73-94, 2010;
- Fiser, A.; Sali, A. **Comparative protein structure modeling with Modeller: A practical approach.** Methods in Enzymology in press, 2002;
- Fiser, A.; Sali, A. **Modeller: generation and refinement of homology-based protein structure models.** Methods in Enzymology, 374:461-91, 2003;
- Gardsvoll, H. *et al.* **Characterization of the functional epitope on the urokinase receptor. Complete alanine scanning mutagenesis supplemented by chemical crosslinking.** Journal of Biological Chemistry, 281: 19260–19272, 2006;
- Gillis, M.; Kersters, K.; Hoste, B.; Janssens, D.; Kroppeenstedt, R.M.; Stephan, M.P.; Teixeira, K.R.; Do, S.; Döbereiner, J.; De Ley, J. ***Acetobacter diazotrophicus* sp. nov., a nitrogen-fixing acetic acid bacterium associated with sugar cane.** International Journal of Systematic Bacteriology, 39:361-364, 1989;

- Ginalski, K. **Comparative modeling for protein structure prediction.** Current Opinion in Structural Biology, 16:172–177, 2006;
- Godjevargova, T.; Dayal, R.; Turmanova, S. **Gluconic acid production in bioreactor with immobilized glucose oxidase plus catalase on polymer membrane adjacent to anion-exchange membrane.** Macromolecular Bioscience, 4: 950–956, 2004;
- Guex, N.; Peitsch, M.C. **SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modelling.** Electrophoresis, 18:2714-2723, 1997;
- Halfon, S.; Craik, C.S. **Regulation of proteolytic activity by engineered metal binding loops.** Journal of the American Chemical Society, 118: 1227–1228, 1996;
- Herrmann, U.; Merfort, M.; Jeude, M.; Bringer-Meyer, S.; Sahm, H. **Biotransformation of glucose to 5-keto-D-gluconic acid by recombinant *Gluconobacter oxydans* DSM 2343.** Applied Microbiology and Biotechnology, 64: 86-90, 2004;
- Humphrey, W. F.; Dalke, A.; Schulten, K. **VMD - Visual Molecular Dynamics.** Journal of Molecular Graphics, 14: 33-38, 1996;
- Intorne, A.C.; de Oliveira, M.V.V.; Lima, M.L.; da Silva, J.F.; Olivares, F.L.; Filho, G.A.S. **Identification and characterization of *Gluconacetobacter diazotrophicus* mutants defective in the solubilization of phosphorus and zinc.** Archives of Microbiology, 191:477–483, 2009;
- James, E.; Bray, A.; Brian, D.; Marsden, A.; Oppermann, U. **The human short-chain dehydrogenase/reductase (SDR) superfamily: A bioinformatics summary.** Chemo-Biological Interactions, Issues 1-3, 178: 99-109, 2009;
- James, E.K.; Olivares, F.L. **Infection and colonization of sugar cane and other graminaceous plants by endophytic diazotrophs.** Critical Reviews in Plant Sciences, 17: 77-119, 1998;
- James, E.K.; Reis, V.M.; Olivares, F.L.; Baldani, J.I.; Döbereiner, J. **Infection of sugarcane by the nitrogen-fixing bacterium *Acetobacter diazotrophicus*.** Journal of Experimental Botany, 45:757-766, 1994;
- Johnson M. S., Srinivasan N., Sowdhamini R., and Blundell T.L. 1994. Knowledge-based protein modeling. *Crit Rev Biochem Mol Biol* 29:1-68;
- Jones, D.T. **Protein secondary structure prediction based on position-specific scoring matrices.** Journal of Molecular Biology, 292: 195-202, 1999;
- Karchin, R.; Kelly, L.; Sali, A. **Improving functional annotation of non-synonymous SNPs with information theory.** Pacific Symposium on Biocomputing, 397-408, 2005;
- Kinoshita, E.; Kinoshita-Kikuta, E.; Ujihara, H.; Koike, T. **Mobility shift detection of phosphorylation on large proteins using a Phos-tag SDS-PAGE gel strengthened with agarose.** Proteomics, 9: 4098- 4101, 2009;
- Klasen, R.; Bringer-Meyer, S.; Sahm, H. **Biochemical characterization and sequence analysis of the gluconate: NADP 5- oxidoreductase gene from *Gluconobacter oxydans*.** The Journal of Bacteriology, 177:2637–2643, 1995;
- Konishi, S. *et al.* **Cysteine-scanning mutagenesis around transmembrane segment VI of Tn10-encoded metal-tetracycline/H(+) antiporter.** FEBS Letters, 461: 315–318, 1999;

Kouadio, J.L. *et al.* **Shotgun alanine scanning shows that growth hormone can bind productively to its receptor through a drastically minimized interface.** *Journal of Biological Chemistry*, 280: 25524–25532, 2005;

Kraft, L.; Sprenger, G.A.; Linqvist, Y. **Conformational changes during the catalytic cycle of gluconate kinase as revealed by x-ray crystallography.** *Journal of Molecular Biology*, 318(4): 1057-1069, 2002;

Krishnan,V.G.; Westhead, D.R. **A comparative study of machinelearning methods to predict the effects of single nucleotide polymorphisms on protein function.** *Bioinformatics*, 19: 2199–2209, 2003;

Kryshtafovych, A. **Basics of CASP8 evaluation measures and automatic data assessment.** CASP 8 meeting. Cagliari, Italy, 2008;

Laskowski, R.A.; MacArthur, M.W.; Moss, D.S.; Thornton, J.M. **PROCHECK: A program to check the stereochemical quality of protein structures.** *Journal of Applied Crystallography*. 26: 283-291, 1993;

Laurie, A.T.; Jackson, R.M. **Q-SiteFinder: an energy-based method for the prediction of protein–ligand binding sites.** *Bioinformatics*, 21:1908-1916, 2005;

Liu, J.Z.; Huang, Y.Y.; Liu, J.; Weng, L.P.; Ji, L.N. **Effects on simultaneous production of glucose oxidase and catalase by *Aspergillus niger*.** *Letters in Applied Microbiology*, 32: 16–19, 2001;

Luna MF, Bernardelli CE, Mignone CF, Boiardi JL. **Energy Generation by Extracellular Aldose Oxidation in N2-Fixing *Gluconacetobacter diazotrophicus*.** *Applied And Environmental Microbiology*, 68 (4): 2054–2056, 2002;

Luna, M.F.; Galar, M.L.; Aprea, J.; Molinari, M.L.; Boiardi, J.L. **Colonization of sorghum and wheat by seed inoculation with *Gluconacetobacter diazotrophicus*.** *Biotechnology Letters*, 32:1071–1076, 2010;

Magnuson, J.K.; Lasure, L.L. **Organic acid production by filamentous fungi.** In: TCAKZ, J.; Lange, L. (Ed.). *Advances in fungal bio/technology for industry, agriculture and medicine*, Berlin: Springer, p. 307-340, 2004;

Manjasetty, B. A.; Wuxian, S; Chenyang, Z.; András, F.; Chance, M.R. **A high-throughput approach to protein structure analysis.** *Genetic engineering*, 28:105-28, 2007;

Marchler-Bauer A. *et al.* **CDD: specific functional annotation with the Conserved Domain Database.** *Nucleic Acids Res.* 37(D)205-210, 2009;

Marti-Renom, M. A.; Stuart, A.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. **Comparative protein structure modeling of genes and genomes.** *Annual Review of Biophysics and Biomolecular Structure*, 29:291-325, 2000;

Marti-Renom, M.A.; Stuart, A.C.; Fiser, A.; Sanchez, R.; Melo, F.; Sali, A. **Comparative protein structure modeling of genes and genomes.** *Annual Review of Biophysics and Biomolecular Structure* 29: 291–325, 2000;

Masso, M.; Lu, Z.; Vaisman, I.I. **Computational mutagenesis studies of protein structure-function correlations Proteins.** *Structure, Function, and Bioinformatics*, Issue 1, 64: 234–245, 2006;

Matsushita, K.; Fujii, Y.; Ano, Y.; Toyama, H.; Shinjoh, M.; Tomiyama, N.; Miyazaki, T.; Sugisawa, T.; Hoshino, T.; Adachi O. **1 5-Keto-D-Gluconate Production Is Catalyzed by a Quinoprotein Glycerol Dehydrogenase, Major Polyol Dehydrogenase, in Gluconobacter Species.** Applied And Environmental Microbiology, 1959–1966, 2003;

Matzerath, I.; Kläui, W.; Klasen, R.; Sahm, H. **Vanadate catalysed oxidation of 5-keto-D-gluconic acid to tartaric acid: the unexpected effect of phosphate and carbonate on rate and selectivity.** Inorganica Chimica Acta, 237:203–205, 1995;

Milsom, P.E.; Meers, J.L. **Gluconic and itaconic acids.** In: Moo-Young M, editor. Comprehensive biotechnol, vol. 3. Oxford: Pergamon Press; p. 681–700, 1985;

Mount, D.W. Bioinformatics, Cold Spring Harbor Lab Press, pp.382-478, 2001;

Muthukumarasamy, R.; Revathi, G.; Vadivelu, M. **Antagonistic potential of N₂-fixing *Acetobacter diazotrophicus* against *Colletotrichum falcatum* Went.: a causal organism of red-rot of sugarcane.** Current Science, 78:1063–1065, 2000;

Ng, P.C.; Henikoff, S. **Predicting deleterious amino acid substitutions.** Genome Research, 11: 863–874, 2001;

Nixon, A. E.; Ostermeier, M.; Benkovic, S. J. **Hybrid enzymes: manipulating enzyme design.** Trends Biotechnology, 16: 258-264, 1998;

Oppermann, U.; Filling, C.; Hult, M.; Shafqat, N.; Wu, X.; Lindh, M.; Shafqat, J.; Nordling, E.; Kallberg, Y.; Persson, B.; Jörnvall, H. **Short-chain dehydrogenases/reductases (SDR): the 2002 update.** Chemico-biological interactions, 143-144:247-53, 2003;

Oppermann, U.C.; Filling, C.; Berndt, K.D.; Persson, B.; Benach, J.; Ladenstein, R.; Jörnvall, H. **Active site directed mutagenesis of 3 beta/17 beta-hydroxysteroid dehydrogenase establishes differential effects on short-chain dehydrogenase/reductase reactions.** Biochemistry, 36(1):34-40, 1997;

Orellana, C.; Neto, R.B. **Brazil and Japan give fuel to ethanol market.** Nature Biotechnology, 24:232, 2006;

Persson, B.; Kallberg, Y.; Bray, J. E.; Bruford, E.; Dellaporta, S.L.; Favia, A. D.; Duarte, R. G.; Jörnvall, H.; Kavanagh, K. L.; Kedishvili, N.; Kisiela, M.; Maser, E.; Mindnich, R.; Orchard, S.; Penning, T. M.; Thornton, J. M.; Adamski, J.; Oppermann, U. **The SDR (short-chain dehydrogenase/reductase and related enzymes) nomenclature initiative.** Chemico-biological Interactions, 178:94-98, 2009;

Petrey D, Honig B. 2005. Protein structure prediction: Inroads to biology . Mol Cell 20:811–81;

Petrey, D.; Xiang, Z.; Tang, C.L.; Xie, L.; Gimpelev, M.; Mitros, T.; Soto, C.S.; Goldsmith-Fischman, S.; Kernysky, A.; Schlessinger, A.; Koh, I.Y.; Alexov, E.; Honig, B. **Using multiple structure alignments, fast model building, and energetic analysis in fold recognition and homology modeling.** Proteins, 53:430- 435, 2003;

Qin, L. *et al.* **Cysteine-scanning analysis of the dimerization domain of EnvZ, an osmosensing histidine kinase.** Journal of. Bacteriology, 185: 3429–3435, 2003;

Ramachandran, S.; Fontanille, P.; Pandey, A.; Larroche, C. **Microbial Production.** Food Technology. Biotechnology, 44 (2): 185–195, 2006;

- Reis, V.M.; Döbereiner, J. **Effect of high sugar concentration on nitrogenase activity of *Acetobacter diazotrophicus***. *Arch Microbiol*, 171:13-18, 1998;
- Reis, V.M.; Olivares, F.L.; Döbereiner, J. **Improved methodology for isolation and identification of *Acetobacter diazotrophicus* and confirmation of its endophytic habitat**. *World Journal of Microbiology and Biotechnology*, 10:401–405, 1994;
- Röhr, M.; Kubicek, C.P.; Kominek, J. **Gluconic Acid**. *Biotechnology*, Vol. 3, Rehm, H.J.; Reed, G. (Eds.), Verlag Chemie, Weinheim, Germany. pp. 455–465, 1983;
- Sali, A.; Blundell, T.L. **Comparative protein modelling by satisfaction of spatial restraints**. *Journal of Molecular Biology*, 234: 779-815, 1993;
- Sali, A.; Blundell, T.L. **Comparative protein modelling by satisfaction of spatial restraints**. *Journal of Molecular Biology*, 234:779-815, 1993;
- Salusjärvi, T.; Povelainen, M.; Hvorslev, N.; Eneyskaya, E.V.; Kulminskaya, A.A.; Shabalin, K.A.; Neustroev, K.N.; Kalkkinen, N.; Miasnikov, A.N. **Cloning of a gluconate/polyol dehydrogenase gene from *Gluconobacter suboxydans* IFO 12528, characterization of the enzyme and its use for the production of 5-ketogluconate in a recombinant *Escherichia coli* strain**. *Applied Microbiology and Biotechnology*, 65: 306–314, 2004;
- Sanchez, R.; Sali, A. **Advances in comparative protein-structure modelling**. *Current Opinion on Structural Biology*, 7:206-214, 1997;
- Sánchez, R.; Sali, A. **Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome**. *Proceedings of the National Academy of Science*, 95, 13597-13602, 1998;
- Sander, C.; Schneider, R. **Database of homology-derived protein structures and the structural meaning of sequence alignment**. *Proteins*, 9 (1): 56-68, 1991;
- Saravanan, V.S.; Madhaiyan, M.; Osborne, J.; Thangaraju, M.; Sa, T.M. **Ecological occurrence of *Gluconacetobacter diazotrophicus* and nitrogen-fixing *Acetobacteraceae* members: their possible role in plant growth promotion**. *Microbial Ecology*, 55:130–40, 2008;
- Sela, M.; White, F.H.; Anfinsen, C.B. **Reductive cleavage of disulfide bridges in ribonuclease**. *Science*, 125 (3250): 691–692, 1957;
- Sevilla, M.; Burris, R.; Gunapala, N.; Kennedy, C.; **Comparison of benefit to sugarcane plant growth and 15 N_2 incorporation following inoculation of sterile plants with *Acetobacter diazotrophicus* wild-type and *nif*- mutant strains**. *Molecular Plant-Microbe Interactions*, 14: 358–66, 2001;
- Shao, Z.; Arnold, F.H. **Engineering new functions and altering existing functions**. *Current Opinion in Structural Biology*, 6: 513–517, 1996;
- Shinagawa, E.; Matsushita, K.; Ameyama, M.; Adachi, O. **D-gluconate dehydrogenase, 2-keto-D-gluconate yielding from *Gluconobacter dioxyaceticus*: purification and characterization**. *Agricultural and Biological Chemistry*, 48:1517–1522, 1984;
- Shinagawa, E.; Matsushita, K.; Toyama, H.; Adachi, O. **Selective production of 5-keto-D-gluconate by *gluconobacter* strains**. *Journal of Fermentation Technology*, 61:359–363, 1983;
- Stadler-Szok, A.; Nyeste, L.; Hollo, J. **Studies on the factors affecting gluconic acid and 5-keto gluconic acid formation by *Acetobacter***. *Acta Alimentaria*, 9: 155–172, 1980;

Stephan, M.P.; Oliveira, M.; Teixeira, K.R.S.; Martínez-Drets, G.; Döbereiner, J. **Physiology and dinitrogen fixation of *Acetobacter diazotrophicus***. FEMS Microbiology Letters, Amsterdam, 77: 67-72, 1991;

Stephen K.B.; Andrzej, J.; Gaetano, T.M.; Wilson, I.A. **Contributions to the NIH-NIGMS Protein Structure Initiative from the PSI Production Centers**. Structure, 16(1): 5–11, 2008;

Sweeney, N.J.; Laux, D.C.; Cohen, P.S. ***Escherichia coli* F-18 and *E. coli* K-12 eda mutants do not colonize the streptomycin-treated mouse large intestine**. Infection and Immunity, 64:3504–3511, 1996;

Tapia-Hernandez, A.; Bustillos-Cristales, M.R.; Jimenez-Salgado, T.; Caballero-Mellado, J.; Fuentes-Ramirez, L.E. **Natural endophytic occurrence of *Acetobacter diazotrophicus* in pineapple plants**. Microbial Ecology, 39:49–55, 2000;

Tejera, N.A.; Ortega, E.; González-López, J.; Lluch, C. **Effect of some abiotic factors on the biological activity of *Gluconacetobacter diazotrophicus***. Journal of Applied Microbiology, 95:528-535, 2003;

Thompson, J.D.; Higgins, D.G.; Gibson, T.J. **CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice**. Nucleic Acids Research, 11; 22(22): 4673–4680, 1994;

Todd, A.E.; Orengo, C.A.; Thornton, J.M. **Evolution of Function in Protein Superfamilies, from a Structural Perspective**. Journal of Molecular Biology, 307: 1113-1143, 2001;

Tramontano, A.; Leplae, R.; Morea, V. **Analysis and assessment of comparative modeling predictions in CASP4**. Proteins Suppl 5: 22–38, 2001;

Tung, C.H.; Huang, J.W.; Yang, J.M. **Kappa-alpha plot derived structural alphabet and BLOSUM-like substitution matrix for fast protein structure database search**. Genome Biology, 8: R31.1-R31.16, 2007;

Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A.E.; Berendsen, H.J. **GROMACS: fast, flexible, and free**. Journal of Computational Chemistry, 26 (16): 1701–1718, 2005;

Velizarov, S.; Beschkov, V. **Production of free gluconic acid by cells of *Gluconobacter oxydans***. Biotechnology Letters, 16: 715–720, 1994;

Venselaar, H.; Te Beek, T.A.H.; Kuipers, R.K.P.; Hekkelman, M.L.; Vriend, G. **Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces**. BMC Bioinformatics, 11:548, 2010;

Vita, C. **Engineering novel proteins by transfer of active sites to natural scaffolds**. Current Opinion in Biotechnology, 8: 429-434, 1997;

Vonrhein, C.; Schlauderer, G.J.; Schulz, G.F. **Movie of the structural changes during a catalytic cycle of nucleoside monophosphate kinases**. Structure, 3: 483–490, 1995;

Walker, J.E.; Saraste, M.; Runswick, M.J.; GayDistantly, N.J. **Related sequences in the a- and f-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold**. The EMBO Journal Vol. 1, 8:945 - 951, 1982;

Wallace, A.C.; Laskowski, R.A.; Thornton, J.M. **LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions.** Protein Engineering, 8: 127–134, 1995;

Wang, Z.; Moulton, J. **SNPs, protein structure, and disease.** Human Mutation, 17: 263–270, 2001;

Weber, O.B.; Baldani, V.L.D.; Teixeira, K.R.S.; Kirchhof, G.; Baldaniand, J.I.; Dobereiner, J. **Isolation and characterization of diazotrophic bacteria from banana and pineapple plants.** Plant Soil, 210:103–113, 1999;

Weenk, G.; Olijve, W.; Harder, W. **Ketogluconate formation by Gluconobacter species.** Applied Microbiology and Biotechnology, 20:400–405, 1984;

Xiang, Z. *et al.* **Pharmacological characterization of 40 human melanocortin-4 receptor polymorphisms with the endogenous proopiomelanocortin-derived agonists and the agouti-related protein (AGRP) antagonist.** Biochemistry, 45: 7277–7288, 2006;

Xu, W.; Yuan, X.; Xiang, Z.; Mimnaugh, E.; Marcu, M.; Neckers, L. **Surface charge and hydrophobicity determine ErbB2 binding to the Hsp90 chaperone complex.** Nature Structure & Molecular Biology, 12:120–126, 2005;

Xu, X.; Burke, S.P. **Roles of active site residues and the NH2-terminal domain in the catalysis and substrate binding of human Cdc25.** Journal of Biological Chemistry, 271 (9): 5118–5124, 1996;

Yang, J.M.; Tung, C.H. **Protein structure databases search and evolutionary classification.** Nucleic Acids Research, 34:3646-3659, 2006;

Yang, Y. *et al.* **Molecular determination of agouti-related protein binding to human melanocortin-4 receptor.** Molecular Pharmacology, 64: 94–103, 2003;

Zanin, G.M. *et al.* **Brazilian bioethanol program.** Applied Biochemistry and Biotechnology, 84–86:1147–1161, 2000;

Zhang, Q.; Peng, H.; Gao, F.; Liu, Y.; Cheng, H.; Thompson, J.; Gao, G.F. **Structural insight into the catalytic mechanism of gluconate 5-dehydrogenase from *Streptococcus suis*: Crystal structures of the substrate-free and quaternary complex enzymes.** Protein Science, 18: 294-303, 2009;

Zhou, M.; Robinson C.V. **When proteomics meets structural biology.** Trends Biochemistry Science, 35: 522-529, 2010;