UNICAMP

JULIO CESAR ARAUJO DA SILVA

BASES MOLECULARES DA DIMINUIÇÃO DA CAPACIDADE FUNCIONAL DO RECEPTOR DE ANDROGÊNIO MUTADO ESTUDADAS POR SIMULAÇÕES DE DINÂMICA MOLECULAR

CAMPINAS
2012

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA

JULIO CESAR ARAUJO DA SILVA

BASES MOLECULARES DA DIMINUIÇÃO DA CAPACIDADE FUNCIONAL DO RECEPTOR DE ANDROGÊNIO MUTADO ESTUDADAS POR SIMULAÇÕES DE DINÂMICA MOLECULAR

ORIENTADOR: PROF. DR. MUNIR SALOMÃO SKAF

TESE DE DOUTORADO APRESENTADA AO INSTITUTO DE QUÍMICA DA UNICAMP PARA OBTENÇÃO DO TÍTULO DE DOUTOR EM CIÊNCIAS

ESTE EXEMPLAR CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA POR JULIO CESAR ARAUJO DA SILVA E ORIENTADA PELO PROF. DR. MUNIR SALOMÃO SKAF.

_____
Assinatura do Orientador

CAMPINAS
2012

FICHA CATALOGRÁFICA ELABORADA POR SIMONE LUCAS - CRB8/8144 - BIBLIOTECA DO INSTITUTO DE QUÍMICA DA UNICAMP

Si38b

da Silva, Julio Cesar Araujo (1974-).
Bases moleculares da diminuição da capacidade funcional do receptor de androgênio mutado estudadas por simulações de dinâmica molecular / Julio Cesar Araujo da Silva. – Campinas, SP: [s.n.], 2012.

Orientador: Munir Salomão Skaf.

Tese (doutorado) - Universidade Estadual de Campinas, Instituto de Química.

1. Receptor nuclear. 2. Receptor de androgênio. 3. Dinâmica molecular-Simulação. I. Skaf, Munir Salomão. II. Universidade Estadual de Campinas. Instituto de Química. III. Título.

Informações para Biblioteca Digital

Título em inglês: Molecular basis of functional impairment of androgen receptor mutants studied by molecular dynamics simulations

Palavras-chave em inglês:
Nuclear receptor
Androgen receptor
Molecular dynamics simulations

Área de concentração: Físico-Química

Titulação: Doutor em Ciências

Banca examinadora:
Munir Salomão Skaf [Orientador]
Alexandre Suman de Araujo
Mario de Oliveira Neto
Leandro Martinez
Rogério Custodio

Data de defesa: 14/12/2012

Programa de pós-graduação: Química

“E eu, tantas vezes reles, tantas vezes vil.”
(Álvaro de Campos)

Dedico esta tese ao meu tio João Bosco
“O dotô da família chegou!”

Agradecimentos

- Aos colegas de laboratório. Os que já estavam quando eu cheguei e aos que chegaram quando eu já estava.
- Ao IQ e à Unicamp por me acolherem em uma instituição de tamanha excelência acadêmica.
- À PRPG-Unicamp pela oportunidade de vivenciar a experiência docente em todas as suas árduas, porém edificantes, atribuições e pela oportuna bolsa para minha manutenção no programa de pós-graduação.
- Aos “caras da química”, de todos os anos e sexos, e aos “lamacentos” que mais uma vez tornaram meus anos na Unicamp muito mais divertidos, leves e, portanto, profícuos. Porque “imagine o poeta inspirado em coca-cola. Que poesia mais estranha ele iria expressar”.
- Em especial gostaria de agradecer ao Paulo, meu coorientador informal, quem compartilhou comigo sua experiência em simulações de receptores nucleares e sempre esteve pronto a discutir ciência mesmo quando minhas colocações ficavam aquém de seus conhecimentos.
- Também agradeço em especial ao Rodrigo pela extrema paciência com algumas de minhas limitações conceituais e por achar um tempinho entre suas atividades no PIQ para discutir ciência e política educacional comigo.
- À toda equipe da CPG-IQ, especialmente à Bel uma pessoa que em todos os meus anos de IQ-Unicamp sempre, friso, sempre esteve disposta a resolver as demandas levadas a ela e, mesmo quando não era possível, sua calma e profissionalismo me convencia da sua impossibilidade naquele momento.
- Ao professor Munir Skaf pela orientação, acolhimento em seu grupo de pesquisa e paciência ao longo destes anos. Presente nas horas fundamentais em que precisei, mostrando minhas limitações de forma precisa, sem rodeios, mas de forma a estimular minha autonomia. Pelo apoio e orientação nas decisões mais importantes de minha carreira, sou-lhe grato.
- Quero lembrar de duas pessoas que em diferentes fases de minha vida me suportaram, em todas as acepções deste verbo, e tornaram possível eu chegar onde cheguei. São elas: Teresa, minha base e meu norte na graduação e Cecília, meu contraponto intelectual durante o mestrado.
- Agradeço aos meus pais, Waldir e Severina, que, mesmo estranhos aos bancos da academia, sempre estimularam minha curiosidade infantil, minha capacidade de me maravilhar com os fenômenos da natureza e minha busca pelo saber como e porquê. Sou muito grato a vocês.
- Por fim, agradeço à minha esposa Renata e ao meu filho Dimitri, que está por vir. Ela representa a maturidade desafiadora que minha vida alcançou e ele, as possibilidades do devir.

Curriculum Vitae

Julio Cesar Araujo da Silva
julioc.araujo@gmail.com

Formação acadêmica/titulação

2008-2012 Doutorado em Química.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil.
Título: Bases Moleculares da Diminuição da Capacidade Funcional do Receptor de Androgênio Mutado Estudadas por Simulações de Dinâmica Molecular.
Orientador: prof. Dr. Munir Salomão Skaf.
Bolsista do: Programa de Estágio Docente – PRPG/Unicamp.

2004 - 2008 Mestrado em Química.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil.
Título: Estudos Estruturais de Agonistas da Acetilcolina pela Espectroscopia de RMN e Cálculos Teóricos, Ano de obtenção: 2008
Orientador: prof. Dr. Roberto Rittner.
Bolsista da: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior.

2003 - 2005 Graduação em Bacharelado em Química.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil.

1995 - 2000 Graduação em Licenciatura Em Química.
Universidade Estadual de Campinas, UNICAMP, Campinas, Brasil.

1991 - 1994 Ensino Profissional de nível técnico.
Escola Técnica Federal de Química, ETFQ-RJ, Rio de Janeiro, Brasil.

Atuação profissional

1. Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina - IF/SC
   Professor, Regime: Dedicação exclusiva.

02/2011 – 07/2011 Ensino Médio *Disciplina Ministrada: Química Geral*

2011 - Atual *Linhas de pesquisa: Desenvolvimento Agrícola e Agroindustrial da Região Serrana Catarinense.*

2. Fundação de Amparo à Pesquisa e Inovação do Estado de Santa Catarina - FAPESC
2011 – Atual Consultor *ad hoc*, Regime: Parcial

3. Universidade Estadual de Campinas - UNICAMP
Professor Estagiário, Carga horária: 12, Regime: Parcial

03/2011 - 07/2011 Graduação. D*isciplina ministrada: Química I (Biologia).*

03/2011 - 07/2011 Graduação. D*isciplina ministrada: Química.*

08/2010 - 12/2010 Graduação. *Disciplina ministrada: Química Experimental I.*

03/2010 - 07/2010 Graduação. *Disciplina ministrada: Química Inorgânica Experimental*

08/2009 - 12/2009 Graduação. *Disciplina ministrada: Química.*

08/2009 - 12/2009 Graduação. *Disciplina ministrada: Química Geral Experimental*

03/2009 - 07/2009 Graduação. *Disciplina ministrada: Química.*

08/2008 - 12/2008 Graduação. *Disciplina ministrada: Física Aplicada II.*

02/2008 - 07/2008 Graduação. *Disciplina ministrada: Fundamentos de Química Aplicada à Construção Civil.*

4. Integrando Escola de Educação Infantil LTDA – ME, Bragança Paulista – SP.
2007 – 2008 Professor de Ensino Médio.

5. Sociedade Educacional Cidade de Itu S/C LTDA, ITU – SP.
2006 – 2008 Professor de Ensino Médio.

6. Instituto Educacional São João da Escócia, Poços de Caldas – MG.
2002 – 2003 Professor de Ensino Médio.

7. Associação Cultural e Educacional Atibaiense S/C LTDA, Atibaia – SP.
2001 – 2002 Professor de Ensino Médio.

Prêmios e títulos

2012 2º Colocado no Concurso Público para o cargo de Professor Universitário do Quadro de Pessoal Permanente - Edital 03/2011 (Portaria 607 de 27/04/2012), Universidade do Estado de Santa Catarina – Udesc.

Produção bibliográfica

Artigos completos publicados em periódicos:

1. da Silva, Julio Cesar A., Ducati, Lucas C., Rittner, Roberto
Conformational and stereoeletronic investigations of muscarinic agonists of acetylcholine by NMR and theoretical calculations. Journal of Molecular Structure (Print), v.1015, p.33 - 40, 2012.

Trabalhos publicados em anais de eventos (resumo)

1. da Silva, Julio Cesar A., Souza, Paulo C. T., Skaf, M. S.
Molecular Basis of Functional Impairment of Androgen Receptor Mutants Studied by Molecular Dynamics Simulations In: XXXVIII Congress of Theoretical Chemists os Latin Expression, 2012, Natal. 2012 Abstract book, 2012.

2. da Silva, Julio Cesar A., Souza, Paulo C. T., Skaf, M. S.
Molecular Dynamics Simulations of the Androgen Receptor LBD: Molecular Basis of Structure-Function Relationship In: 7th International Conference of the Brazilian Association for Bioinformatics and Computational Biology (AB3C) and 3rd International Conference of the IberoAmerican Society for Bioinformatics (SolBio), 2011, Florianópolis. 2011 Abstract book, 2011.

3. da Silva, J. C. A., Rittner, Roberto
Estudo do Efeito de Solvente sobre o Equilíbrio Conformacional dos Agonistas Muscarínicos da Acetilcolina por RMN 1H e 13C. In: 33a. Reunião Anual da Sociedade Brasileira de Química, 2010, Águas de Lindóia. 33a. Reunião Anual da Sociedade Brasileira de Química, 2010.

4. da Silva, J. C. A., Rittner, Roberto
Investigação dos Fatores de Estabilização dos Confôrmeros dos Agonistas Muscarínicos da Acetilcolina por Cálculos ab initio. In: 33a. Reunião Anual da Sociedade Brasileira de Química, 2010, Águas de Lindóia. 33a. Reunião Anual da Sociedade Brasileira de Química, 2010.

5. da Silva, Julio Cesar A., Rittner, Roberto
Estudos dos Fatores de Estabilização dos Confôrmeros de Agonistas da Acetilcolina pela Espectroscopia de RMN e Cálculos Teóricos In: XV Simpósio Brasileiro de Química Teórica, 2009, Poços de Caldas. Livro de Resumo 2009, 2009.

Resumo

BASES MOLECULARES DA DIMINUIÇÃO DA CAPACIDADE FUNCIONAL DO RECEPTOR DE ANDROGÊNIO MUTADO ESTUDADAS POR SIMULAÇÕES DE DINÂMICA MOLECULAR.

Receptores de androgênio (AR) são membros da superfamília de receptores nucleares que incluem os receptores de esteroides, entre outros. O AR liga os esteroides sexuais endógenos diidrotestosterona e testosterona. O desenvolvimento normal do fenótipo masculino e do sistema reprodutivo necessita de ações pré- e pós-natais promovidas pela interação do AR com esses hormônios. Mutações no gene do receptor de androgênio podem levar a várias doenças como o câncer de próstata e a síndrome de insensibilidade ao androgênio (AIS). Substituições diferentes no mesmo resíduo de aminoácido podem resultar em impactos variáveis na atividade do receptor levando a diferentes graus de AIS. Um grande número de mutações tem sido reportado para o AR envolvendo AIS e células tumorais de câncer de próstata e sua localização e função podem ajudar a entender como essas doenças devem ser tratadas. Entretanto, pouco se sabe sobre como as mutações mudam a estrutura e a dinâmica do AR, uma vez que apenas poucas estruturas cristalográficas de mutantes foram obtidas.

Neste trabalho, apresentamos estudos de simulação de dinâmica molecular de algumas estruturas do AR humano com mutações localizadas no domínio de ligação do ligante (LBD) em comparação com a estrutura nativa complexadas com o ligante sintético metiltrienolona (R1881). Nosso objetivo é investigar as bases moleculares das mudanças sutis no receptor causadas pelas mutações que afetam sua afinidade pelo ligante R1881. Embora nenhum dos resíduos mutados deste estudo interajam diretamente com o ligante, os resultados das simulações indicaram que as mutações causam mudanças estruturais e dinâmicas no AR-LBD na região onde se localiza a mutação e na cavidade de ligação do ligante (LBP). A principal mudança observada foi o deslocamento do resíduo Arg752, facilitando a entrada de moléculas de água no LBP e o reposicionamento de cadeias laterais dos resíduos do domínio F, uma importante região que contribui para a estabilidade da estrutura do AR-LBD e da conformação ativa da hélice 12.

Os resultados obtidos mostram que essas mutações, que ocorrem naturalmente, são exemplos de resíduos que não estão em contato direto com o ligante e não pertencem à região de recrutamento do coativador, mas que possuem um importante papel na ligação do ligante e na ativação do receptor por estabilizar a hélice 12 e o domínio F na conformação ativa.

Abstract

MOLECULAR BASIS OF FUNCTIONAL IMPAIRMENT OF ANDROGEN RECEPTOR MUTANTS STUDIED BY MOLECULAR DYNAMICS.

Androgen receptors (AR) are members of the superfamily of nuclear receptors that includes the steroid receptors, among others. AR binds the male endogenous sex steroids, dihydrotestosterone and testosterone. Normal development of the male phenotype and reproductive system requires pre- and postnatal actions promoted by AR interaction with these hormones. Mutations in the androgen receptor gene may lead to several diseases like prostate cancer (PCa) and the androgen insensitivity syndrome (AIS). Different substitutions at the same amino acid residue may result in variable impact on the activity of the receptor leading to different degrees of AIS. A number of mutations have been reported for the AR in AIS and PCa tumor cells and their location and function may help us to understand how these diseases should be treated. Nevertheless, not much is known about how the mutations change the structure and dynamics of the AR since only a few crystallographic structures of mutants were obtained.

In this work we present molecular dynamics simulation (MD) studies of some human AR mutations located in the ligand binding domain (LBD) in comparison with wild type (WT) structure in complex with the synthetic ligand methyltrienolone (R1881). Our goal is to investigate the molecular basis of subtle changes in the receptor caused by mutations in the AR-LBD/R1881 affinity. Although the mutated residues do not interact with the ligand, the simulations results indicated that the mutants cause structural and dynamical changes in the AR-LBD in the region in which the mutation is placed and in the binding pocket (LBP). The principal change observed was the displacement of residue Arg752, facilitating water penetration in LBP, and the repositioning of F-domain side chains, which makes important contributions to the stability of the AR-LBD structure and helix 12 active conformation.

The results obtained show that these naturally occurring mutations are examples of residues that are not in contact with the ligand and do not belong to coactivator recruitment region, but which do have an important role in ligand binding and receptor activation by stabilizing the helix H12 and the F-domain in the active conformation.

Índice

Lista de Abreviaturas xxi
Lista de Tabelas xxii
Lista de Figuras xxiv
1. Introdução 1
1.1. Os Receptores Nucleares 2
1.1.1. Os Domínios Estruturais 4
1.1.2. Estrutura e Função das Proteínas 5
1.1.3. O Papel da Hélice 12 em NRs 6
1.1.4. As Mutações 8
1.2. O Receptor de Androgênio 8
1.2.1. Ligantes do Receptor de Androgênio 14
1.2.2. As mutações no AR e seus fenótipos. 15
1.3. Dinâmica Molecular 17
1.3.1. Fundamentos teóricos 19
1.3.2. Etapas de uma simulação de Dinâmica Molecular 23
1.3.3. Geração da configuração inicial 25
1.3.4. Condições periódicas de contorno 26
1.3.5. Convenção da imagem mínima e raio de corte 27
1.3.6. Controle da temperatura e da pressão 28
1.3.7. Vínculos e restrições 29
1.3.8. Tratamento do Solvente 29
1.3.9. Velocidades iniciais 30
1.3.10. Movimentação das partículas 31
1.3.11. Parametrização do Ligante R1881 32
2. Objetivo da Tese 34
2.1. Objetivo geral 34
2.2. Objetivos específicos 34
3. Materiais e Métodos 35
4. Resultados e Discussão 38
4.1. As Mutações do AR 38

4.2. Sistema AR-LBD-cofator/R1881 40
4.2.1 Estabilidade da simulação 40
4.2.2. Perfil das Interações dos Ligantes Hidrofílicos do LBP 42
4.2.3. Perfil de Solvatação do LBP 45
4.2.4. Perfil dos Contatos das Regiões Chaves do AR-LBD 49
4.2.4.1 Perfil da Dinâmica do Ligante R1881 50
4.2.4.2 Perfil da Dinâmica dos Resíduos Mutados 50
4.2.4.3 Perfil da Dinâmica do Domínio F 61
4.2.4.4. Perfil em torno do AF2 64
5. Conclusões 65
6. Bibliografia 67
Anexo I: Percentuais de tempo relativo de contatos 77
Anexo II: Parâmetros do Ligante R1881 84
II.1. Arquivo de Parâmetros 85
II. 2. Arquivo de Topologia 87
Anexo III: Lista dos Aminoácidos Padrões 94

Lista de Abreviaturas

AF1/2: ActivationFunction 1/2 – Função de Ativação 1/2.
AIS: Androgen Insensitivity Syndrome – Síndrome de Insensibilidade ao Androgênio.
(h) AR: (human) Androgen Receptor – Receptor de Androgênio (humano).
BPTI: Bovine pancreatic trypsin inhibitor – Inibidor da Tripsina Pancreática Bovina.
CAIS: Complete Androgen Insensitivity Syndrome – Síndrome Completa de Insensibilidade ao Androgênio.
CTD:Carboxi-terminal Domain – Domínio Carbóxi-terminal.
CTE: Carboxi-terminal Extension – Extensão C-terminal.
DBD: DNA Binding Domain – Domínio de Ligação do DNA.
DHT: 5α-Dihydrotestosterone – 5α-Diidrotestosterona
ER: Estrogen Receptor – Receptor de Estrogênio.
FRET: Förster Resonance Energy Transfer – Transferência de Energia Por Ressonância de Förster.
GR: Glucocorticoid Receptor – Receptor de Glucocorticóide.
HGMD: Human Gene Mutation Database – Base de Dados de Mutação Genética Humana.
HSP: Heat-shock Proteins – Proteínas de Choque Térmico.
LBP: Ligand BindingPocket – Bolsão de Ligação do Ligante.
LDB: Ligand Binding Domain – Domínio de Ligação do Ligante.
MAIS: AIS: Mild Androgen Insensitivity Syndrome – Síndrome Leve de Insensibilidade ao Androgênio.
MC: Monte Carlo Methods – Métodos de Monte Carlo.
MD: Molecular Dynamics – Dinâmica Molecular.
MR: Mineralocorticoid Receptor – Receptor de Mineralocorticóide.
NMR: Nuclear Magnetic Resonance – Ressonância Magnética Nuclear.
NR: Nuclear Receptor – Receptor Nuclear.
NTD: N-terminal Domain – Domínio N-terminal.
PAIS: AIS: Partial Androgen Insensitivity Syndrome – Síndrome Parcial de Insensibilidade ao Androgênio.
PCa: Prostate Cancer – Câncer de Próstata.
PPAR: Peroxisome Proliferator Activated Receptor – Receptor Ativado de Proliferador de Peroxissomos.
PR: Progesterone Receptor – Receptor de Progesterona.
R1881: Metiltrienolona; Metribolona.
R752: Arginina 752.
RAR: Retinoic Acid Receptor – Receptor de Ácido Retinóico
SRC: Steroid Receptor Coativator – Coativador do Receptor de Esteróide
TR: Thyroid Receptor - Receptor do hormônio Tireoidiano.
VDR: D Vitamin Receptor – Receptor de vitamina D.
XRD: X-ray Diffraction – Difração de Raios-X.

Lista de Tabelas

Tabela 1: mutações selecionadas para o estudo. As cores fazem correspondência com as esferas que indicam a localização das mutações na Figura 14A. 17

Tabela 2: Porcentagem de tempo de contato entre o resíduo 916 e resíduos até 4 Å dele para as estruturas nativa e mutada. 52

Tabela 3: Porcentagem de tempo de contato entre o resíduo 817 e resíduos até 4 Å dele para as estruturas nativa e mutada. 54

Tabela 4: Porcentagem de tempo de contato entre o resíduo 840 e resíduos até 4 Å dele para as estruturas nativa e mutada. 55

Tabela 5: Porcentagem de tempo de contato entre o resíduo 743 e resíduos até 4 Å dele para as estruturas nativa e mutada. 56

Tabela 6: Porcentagem de tempo de contato entre o resíduo 871 e resíduos até 4 Å dele para as estruturas nativa e mutada. 57

Tabela 7: Porcentagem de tempo de contato entre o resíduo 909 e resíduos até 4 Å dele para as estruturas nativa e mutada. 59

Tabela 8: Porcentagem de tempo de contato entre o resíduo 919 e resíduos até 4 Å dele para as estruturas nativa e mutada. 60

Tabela 9: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação F916A. 77

Tabela 10: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação F916A. 77

Tabela 11: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação P817A. 78

Tabela 12: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação P817A. 78

Tabela 13: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação R840C. 79

Tabela 14: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação R840C. 79

Tabela 15: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação G743V. 80

Tabela 16: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação G743V. 80

Tabela 17: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação R871G. 81

Tabela 18: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação R871G..... 81

Tabela 19: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação G909E..... 82

Tabela 20: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação G909E..... 82

Tabela 21: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação Q919R..... 83

Tabela 22: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação Q919R..... 83

Tabela 23: Tipos atômicos atribuídos e cargas parciais calculadas para todos os átomos do ligante R1881. 84

Lista de Figuras

Figura 1: Representação da estrutura comum dos domínios de um receptor nuclear. Em cima: Esquema 1D. As regiões N- e C-terminal estão indicadas. AF1, função de ativação 1; AF2, função de ativação 2; DBD, domínio de ligação do DNA; LBD, domínio de ligação do ligante. Em baixo: Estrutura experimental 3D do receptor de estrogênio. As estruturas não determinadas NTD, dobradiça e domínio F estão indicadas pelas linhas pontilhadas (vermelho, rosa e laranja, respectivamente) (adaptado da ref. 9)..... 5

Figura 2: Representação do mecanismo clássico de ativação/inibição dos receptores nucleares pela hélice 12. Agonistas e antagonistas promovem diferentes conformações no receptor. A ligação do agonista resulta no reposicionamento da H12 próximo aos resíduos das hélices 3, 4, 5 e 6 criando uma superfície para a ligação do coativador. A ligação do antagonista inibe a capacidade da H12 adotar a conformação ativa. Ao invés disso, a hélice 12 ocupa uma posição que impede a ligação do coativador (retirado da ref. 47). 7

Figura 3: Representação do mecanismo da ação do androgênio na célula. Da ligação à ativação do gene para a expressão da proteína. T, testosterona; DHT, diidrotestosterona, N-term, região N-terminal; hsps, proteínas de choque térmico; P, sítio de fosforilação, RNA Pol II, RNA polimerase II. (retirado da ref. 55) 10

Figura 4: (A) Representação esquemática da organização estrutural do AR humano. (retirada da ref. 58); (B) Representação da estrutura do AR-LBD. Em violeta: alfas-hélices; em azul: hélices 3-10; em amarelo: fitas beta. O ligante R1881 está representado no modelo “esferas e varetas”. O peptídeo mimetizando o coativador, CoA, está em vermelho. 11

Figura 5: Representações da estrutura do AR-LBP. (A) Resíduos que fazem parte do sítio de ligação do R1881 coloridos segundo o tipo (apolar: branco, polar: verde e básico: azul); (B) Superfície de van der Waals mostrando a cavidade hidrofóbica do LBP; (C) Resíduos hidrofílicos que ancoram o ligante R1881 através de ligações de hidrogênio (os resíduos N705 e T877 ligam-se ao grupo OH-17 e os resíduos Q711 e R752 ao O-3 do grupo ceto). Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 12

Figura 6: Representação da superfície do AF2 do AR-LBD. Os resíduos pertencentes ao grampo de cargas que flanqueiam o AF2 estão em vermelho (negativo) e azul (positivo). Os demais resíduos que compõem o AF2 estão em branco (apolar) e verde (polar). 13

Figura 7: Representação das interações específicas entre os resíduos da fita β S4 pertencente à extensão C-terminal e outras partes do AR-LBD (S3, H9 e H10). Os resíduos incluídos fazem ou ligação de hidrogênio ou contatos de van der Waals. Os átomos estão coloridos segundo o tipo (carbono: azul; oxigênio: vermelho; nitrogênio:

azul). Para maior clareza, os átomos de hidrogênio foram omitidos. A esfera vermelha representa uma molécula de água. 14

Figura 8: Estruturas moleculares dos principais ligantes do AR. Naturais: (a) testosterona e (b) diidrotestosterona. Sintético: (c) metiltrienolona (R1881). A posição dos grupos polares está indicada na estrutura “a”..... 15

Figura 9: Faixa de escala temporal para dinâmicas de sistemas biomoleculares. Embora os passos temporais individuais de dinâmicas moleculares seja de 1 a 2 fs, a computação paralela permite simular em uma escala de microssegundos e técnicas de computação distribuída podem amostrar mesmo processos mais lentos, quase atingindo milissegundos. 21

Figura 10: Fluxograma simplificado das etapas de uma simulação de dinâmica molecular típica. A ideia básica é gerar estruturas a partir de um conjunto natural calculando as funções potenciais e integrando as equações de movimento de Newton; essas estruturas são então usadas para estimar propriedades de equilíbrio do sistema. Um passo temporal típico é da ordem de 1 a 2 fs. 23

Figura 11: Resumo dos passos em uma simulação típica de MD de proteína. O primeiro passo é importar uma estrutura inicial do PDB. Em seguida, solvente é adicionado e a estrutura é relaxada via um algoritmo de minimização para remover maus contatos. Então o sistema é equilibrado: a estrutura experimental minimizada é submetida à simulação de MD para trazê-la para o equilíbrio com as condições de simulação. Apenas após esse estágio de equilibração podem as simulações de produção serem realizadas para fornecerem resultados realísticos (adaptada da ref. 99). 24

Figura 12: Representação de uma caixa de simulação bidimensional (em cinza) e suas réplicas periódicas (em branco). As condições periódicas de contorno correspondem à criação de réplicas idênticas da caixa de simulação colocadas lado a lado, de modo que qualquer movimento que ocorra na caixa de simulação ocorra também nas réplicas... 27

Figura 13: Alternativas a um corte brusco para interações coulômbicas não ligadas. Em cima: desligando a interação (tracejada) antes do corte, a força será a derivada exata do potencial, mas a derivada (e, portanto, a força) crescerá extraordinariamente logo após o corte. Em baixo: PME é um algoritmo no qual a interação de Coulomb (sólida) é dividida em um termo de curto alcance que é resolvido dentro de um corte (tracejada) e um termo de longo alcance que pode ser resolvido exatamente em um espaço recíproco com transformadas de Fourier (ponto e traço). (retirado da ref. 95)..... 28

Figura 14: (A) Localização no hAR-LBD das mutações selecionadas para o estudo. As cores das esferas fazem correspondência com as cores do texto na Tabela 1; (B) Ordem crescente do aumento da perda de afinidade relativa do ligantes. 39

Figura 15: Representação esquemática do alinhamento de corpo rígido, necessário para computar o RMSD. 40

Figura 16: Média da evolução temporal do RMSD dos carbonos do backbone do AR-LBD da estrutura nativa em relação à estrutura cristalográfica da própria nativa..... 41

Figura 17: Média da evolução temporal dos RMSD's dos carbonos do backbone do AR-LBD em relação à estrutura cristalográfica: (A) F916A, (B) P817A, (C) R840C, (D) G743V, (E) R871G, (F) G909E e (G) Q919R. 42

Figura 18: Instantâneos da simulação do F916A mostrando as estruturas do AR-LBD com o ligante R1881 e o resíduo R752: (A) 0,0 ns; (B) 2,6 ns; e (C) 3,1 ns. O mesmo comportamento é observado nas outras estruturas em que ocorre a perda de contato com o ligante. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 43

Figura 19: Perfil da variação da distância entre o resíduo R752 e o ligante R1881 em função do tempo de simulação: (A) nativa; (B) F916A; (C) P817A; (D) R840C; (E) G743V; (F) R871G; (G) G909E; (H) Q919R. Em cada gráfico, as curvas mostram as 10 simulações de 8 ns feitas para cada um dos sistemas..... 44

Figura 20: Distribuição relativa média das moléculas de água posicionadas até 4 Å do ligante R1881 no AR-LBD das estruturas mutadas em relação à estrutura nativa: (A) F916A; (B) P817A; (C) R840C; (D) G743V; (E) R871G; (F) G909E; (G) Q919R. 46

Figura 21: Instantâneo da simulação mostrando estruturas representativas do AR-LBD: (A) Na estrutura nativa, a molécula de R1881 interage diretamente com resíduo R752 e com uma molécula de água apenas; (B) Na variante F916A, o ligante R1881 perde sua interação com R752, mas passa a interagir com maior frequência com um número maior de moléculas de água. Comportamento semelhante ocorre nas demais mutações em que há aumento da frequência de entrada de moléculas de água devido à perda da interação R752-R1881. 47

Figura 22: Relação entre a entrada de moléculas de água no AR-LBP e a perda de interação entre o ligante R1881 e o resíduo R752. (A) nativa; (B) F916A; (C) P817A; (D) R840C; (E) G743V; (F) R871G; (G) G909E; (H) Q919R. 49

Figura 23: Representação da região em torno da posição 916. (A) estrutura nativa: F916; (B) estrutura mutada: A916. Estão incluídos os principais resíduos que fazem ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 2 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 52

Figura 24: Representação da região em torno do resíduo A817. (A) estrutura nativa: P817; (B) estrutura mutada: A817. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 11 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 54

Figura 25: Representação da região em torno da posição 743. (A) estrutura nativa: G743; (B) estrutura mutada: V743. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 5 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 56

Figura 26: Representação da região em torno da posição 871. (A) estrutura nativa: R871; (B) estrutura mutada: C871. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 6 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 58

Figura 27: Representação da região em torno da posição 909. (A) estrutura nativa: G909; (B) estrutura mutada: E909. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 7 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 59

Figura 28: Representação da região em torno da posição 919. (A) estrutura nativa: Q919; (B) estrutura mutada: R919. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 8 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul). 61

Figura 29: Representação das estruturas médias do AR-LBD mostrando os resíduos que diminuem (em vermelho) e os que aumentam (em preto) o número de contatos com o domínio F (em verde) ao longo da simulação para cada mutação em relação à estrutura nativa: (A) F916A; (B) P817A; (C) R840C; (D) G743V; (E)R871G; (F) G909E; (G) Q919R. Apenas diferenças de porcentagem maiores que 5% são apresentadas. . 63

Figura 31: Aminoácidos essenciais estão indicados com círculos amarelos. A polaridade, basicidade e acidez refere-se à cadeia lateral. (retirado de http://www.unc.edu/~bzafer/aminoacids). Acessado em 17 de outubro de 2012. 94

1. Introdução

Uma das linhas de pesquisa de grande interesse do nosso grupo é a determinação das bases moleculares relacionadas à função dos receptores nucleares (NR, na sigla em inglês), fatores de transcrição intracelulares que regulam a atividade da complexa maquinaria dos genes. Estas proteínas definem uma superfamília responsável pelos principais aspectos do desenvolvimento dos eucariotos: diferenciação, reprodução e homeostase metabólica.[1–4]

Devido à natureza dinâmica de muitos dos mecanismos moleculares dos NRs, os estudos baseados em estruturas cristalográficas de raios-X ou em estruturas de ressonância magnética nuclear nem sempre são capazes de explicar o grande número de doenças associadas às mutações destes receptores. Dentre eles, o receptor de androgênio (AR) é um dos maiores alvos para o desenvolvimento de fármacos devido a seu importante papel na diferenciação e manutenção das características masculinas do indivíduo e no tratamento do câncer de próstata. Os ligantes de AR de diferentes estruturas químicas e/ou propriedades farmacológicas são largamente usados para estas aplicações terapêuticas e os atualmente disponíveis para terapia modulam a função do AR via ligação direta com o LBD do receptor.[5]

A grande maioria das anomalias do AR está associada a mutações pontuais em que um único nucleotídeo é trocado. Por sua vez, a maioria das mutações identificadas são mutações germinais associadas com a síndrome de insensibilidade ao androgênio (AIS) e/ou ao câncer de próstata (PCa).[5] Este defeito em um único resíduo fornece a base para uma abordagem estrutural, em complemento a estudos das habilidades funcionais do AR *in vitro*.[6]

Neste sentido, o propósito desta Tese é estudar como algumas das mutações associadas a AIS e a PCa afetam a função do AR através de mudanças na estrutura do AR-LBD, empregando para este fim simulações de dinâmica molecular (MD). Mais especificamente, estamos interessados nas bases moleculares das mudanças estruturais que possam estar associadas a perda ou diminuição de algumas funções biológicas por parte do receptor de androgênio.

1.1. Os Receptores Nucleares

O primeiro receptor nuclear foi identificado há cerca de 45 anos. Era um receptor intracelular que ligava algumas moléculas de esteroides.[7,8] Contudo, mais de 20 anos se passaram antes que ficasse claro que aqueles receptores de esteroides fazem parte de uma superfamília de fatores de transcrição de metazoários que evoluiu antes da separação dos vertebrados e invertebrados. Os 48 membros conhecidos dos receptores nucleares humanos incluem tanto receptores para os quais ligantes são conhecidos como "receptores órfãos" para os quais não há ainda ligantes descobertos.[9]

Receptores nucleares são fatores de transcrição essenciais para o desenvolvimento embrionário, manutenção de fenótipos celulares diferenciados, metabolismo e morte celular.[10] A superfamília dos NRs é tipicamente subdividida em três subfamílias, ou classes. A subfamília dos receptores de esteroides (classe I) inclui o receptor de progesterona (PR), receptor de estrogênio (ER), receptor de glicocorticóide (GR), receptor de androgênio (AR) e receptor de mineralocorticóide (MR). A subfamília dos tireóides/retinóides (classe II) inclui o receptor de hormônios tireoideanos (TR), receptor de vitamina D (VDR), receptor de ácido retinóico (RAR) e receptor ativado de proliferador de peroxissomos (PPAR). A terceira classe de receptores foi nomeada família dos receptores órfãos que define um conjunto de proteínas identificadas pela análise comparativa da sequência como pertencendo à superfamília dos receptores nucleares, mas para os quais o ligante cognato é desconhecido. Embora todos os receptores nucleares regulem a expressão gênica, entre as três classes existem diferenças sutis nos mecanismos bioquímicos pelos quais os receptores desempenham suas funções. Por exemplo, enquanto os receptores da classe I, normalmente, formam homodímeros, os da classe II, tipicamente, funcionam como heterodímeros.[1–4]

À medida que o conhecimento sobre NRs foi crescendo, tornou-se evidente que os receptores nucleares formam uma família muito diversa em termos tanto de papel fisiológico como o de ação molecular. A visão clássica derivada dos receptores endócrinos é que cada um deles é ativado por um único ligante de alta afinidade com uma constante de dissociação na faixa de nanomol por litro. Os melhores exemplos

desses receptores são os membros da subfamília dos receptores de hormônio esteróide como o receptor de estrogênio (ER) e o receptor de androgênio.[9]

Mais recentemente, identificou-se que os chamados receptores metabólicos são ativados por diversos ligantes, mas de baixa afinidade, com constantes de dissociação na faixa de micromol por litro. Por exemplo, PPARs são ativados por vários ácidos graxos, eicosanóides e prostanóides e receptores X do fígado (LXRs) são ativados por oxiesteróis[11]. Hoje sabe-se que receptores nucleares podem tanto ativar como reprimir a transcrição e que, portanto, o papel dos ligantes pode ser ativar, inativar ou reprimir receptores para desativar ou superativar receptores constitutivamente ativos ou, em alguns poucos casos, para servir mesmo como correguladores estruturais não regulatórios. Somado a isso, descobriu-se que alguns receptores aparentemente carecem de uma cavidade de ligação do ligante e, portanto, podem ser regulados por outros meios.[9]

Devido à diversidade e dos importantes papéis biológicos, os receptores nucleares são alvos primários na descoberta de fármacos. A disfunção da sinalização do NR leva a doenças proliferativas, reprodutivas e metabólicas tais como câncer, infertilidade, obesidade e diabetes. Agonistas ou antagonistas farmacêuticos dos receptores nucleares, tais como tamoxifeno para receptores de estrogênio (alvo para câncer de mama)[12], tiazolidinedionas para PPAR-γ (alvo para diabetes tipo II)[13] ou dexametasona para receptores de glicocorticóides (alvo para doenças inflamatórias)[14], estão entre as drogas mais usadas.[10] Portanto, é evidente que a elucidação dos mecanismos de ativação dos receptores nucleares possuem uma grande gama de implicações para o entendimento da regulação biológica, bem como para os esforços de explorar receptores nucleares como alvos de fármacos.[9]

1.1.1. Os Domínios Estruturais

NRs são proteínas modulares com domínios de ligação ao DNA (DBD, do inglês *DNA binding domain*) e de ligação ao ligante evolucionariamente conservados. Um receptor nuclear típico possui ainda uma região N-terminal pouco conservada, uma região dobradiça entre o DBD e o LBD que permite a rotação de um em relação ao outro e, em alguns casos, uma região C-terminal chamada de domínio F. Sendo fatores de transcrição, receptores nucleares também contém regiões necessárias para a ativação transcricional. A função de ativação 1 (AF1) localizada na região N-terminal é independente do ligante, já a função de ativação 2 (AF2), localizada no LBD, é essencial para a transativação dependente do ligante e é conservada ao longo dos membros dos NRs (Figura 1).

O LBD pode ser visto como um interruptor molecular que, uma vez acionado pelo ligante, habilita o receptor a ativar a transcrição dos genes alvo e a interferir com outros sinais de caminhos de transdução. Além disso, a ligação do ligante é necessária, em muitos casos, para a interação do receptor com seus elementos responsivos de DNA cognatos, homo e heterodimerização e, mais importante, posicionar o AF2. Dentro do AF2 de vários NRs, um fraco domínio autônomo de ativação, nomeado de AF2 AD, foi identificado. A integridade da α-hélice anfipática conservada que constitui o núcleo do AF2 AD mostrou-se necessária à transativação e à interação dependente do ligante entre o LBD e os fatores intermediários transcricionais (também chamados de coativadores ou mediadores) que medeiam a atividade do AF2 na maquinaria de transcrição. Por outro lado, também já foi mostrado que LBDs não ligados do TR e do RAR podem interagir com fatores que aparentemente medeiam o silenciamento da atividade transcricional desses NRs em particular, os correpressores.[15]

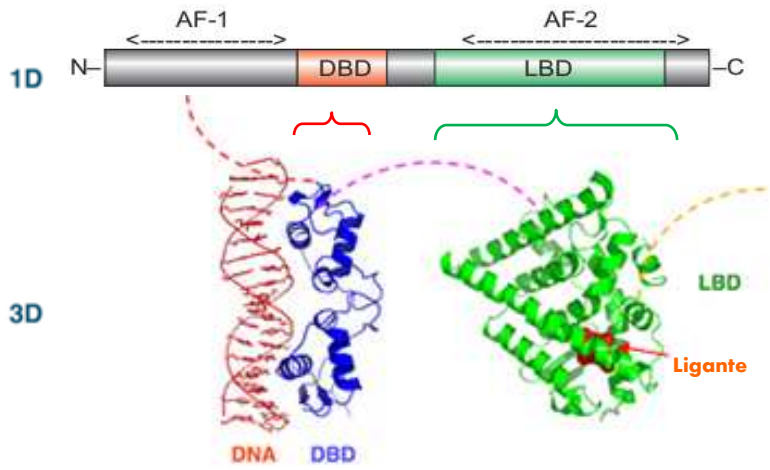


Figura 1: Representação da estrutura comum dos domínios de um receptor nuclear. Em cima: Esquema 1D. As regiões N- e C-terminal estão indicadas. AF1, função de ativação 1; AF2, função de ativação 2; DBD, domínio de ligação do DNA; LBD, domínio de ligação do ligante. Em baixo: Estrutura experimental 3D do receptor de estrogênio. As estruturas não determinadas NTD, dobradiça e domínio F estão indicadas pelas linhas pontilhadas (vermelho, rosa e laranja, respectivamente) (adaptado da ref. 9).

1.1.2. Estrutura e Função das Proteínas

A diferença estrutural entre as famílias de proteínas é formada pela divergência sequencial acumulada como consequência da evolução de novas espécies, da duplicação de genes e de eventos de anulação, bem como pela pressão seletiva evolucionária exercida em cada proteína de acordo com a estrutura tridimensional correspondente e a função específica desempenhada.[16,17] O balanço entre rearranjos genômicos e pressão seletiva para aumentar o repertório funcional disponível para os organismos leva ao surgimento de novas subfamílias na escala evolucionária.[18] Existem muitos aspectos da função de proteínas que contribuem para a evolução da organização da família. Estes incluem a conservação global dos mecanismos catalíticos (no caso de enzimas), ligação específica a substratos e cofatores, bem como a interação com outras proteínas em processos tais como sinalização celular, regulação de reações e formação de complexos macromoleculares. Uma hipótese comumente aceita é que enquanto posições completamente conservadas estão relacionadas aos aspectos funcionais comuns a todos os membros da família, outros resíduos estão

relacionados à especificidade funcional, por exemplo, diferença na ligação de cofatores.[19]

Chothia, Lesk e Gerstein foram os pioneiros nos estudos de flexibilidade de proteínas de larga escala por comparações de dois ou mais estados estruturais da mesma proteína.[20–22] Seus estudos levaram à criação de uma importante base de dados de movimentos de proteína.[23] Eles sugeriram uma classificação dos principais tipos de grandes movimentos proteicos caracterizados por três amplitudes (sem movimento, movimentos menores e movimentos maiores), três tamanhos (fragmento, domínio e subunidade) e três mecanismos (dobradiça, tesoura e outros).[24]

As mudanças conformacionais em resposta à ligação de outra molécula são de extrema importância para a função da proteína. Uma mudança pode tanto envolver apenas uma simples cadeia ou pode ser tão drástica como um reposicionamento (movimento) de longo alcance de fragmentos maiores ou domínios e até mesmo enovelamento e/ou reenovelamento parcial.[24]

A função da proteína depende em muito da sinergia entre estrutura e dinâmica. A dinâmica envolve a interconversão de estados conformacionais em uma complexa superfície de energia livre multidimensional. Torna-se muito difícil estudar tais transições conformacionais experimentalmente em um nível de detalhes atômicos, embora técnicas como a de transferência de energia por ressonância de Förster (single-molecule FRET),[25] cristalografia de raios-X e ressonância magnética nuclear (NMR) deem informações úteis, porém limitadas. Por isso é que simulações de dinâmica molecular estão desempenhando um papel crescente na determinação de superfícies de energia livre, como um suplemento aos estudos experimentais.[26]

1.1.3. O Papel da Hélice 12 em NRs

A visão corrente da farmacologia de receptores nucleares estabelece que quando um ligante liga-se ao LBD, ele causa um rearranjo estrutural no qual a hélice 12 (H12) C-terminal adota uma das duas posições possíveis dependendo se o ligante é um agonista ou antagonista (Figura 2). Neste modelo, que ficou conhecido como “ratoeira”, o reposicionamento da H12 abre a rota de entrada e saída para o bolsão de ligação do

ligante.[27] Baseado na posição da H12, proteínas coativadoras que induzem a transcrição gênica ou correpressores que induzem a inibição são recrutados.[28]

A descrição estática e descrita acima é, em muitos casos, suficiente para explicar resultados experimentais, mas para entender o processo em detalhe é preciso apreciar a natureza dinâmica das moléculas de proteínas e ligantes. Vários estudos envolvendo estruturas cristalinas dos LBDs dos receptores nucleares têm mostrado que na ausência do ligante ou na presença de antagonistas, a hélice 12 adota uma variedade de posições, incluindo posições similares muito próximas da conformação ativa.[29–41] Entretanto, análises bioquímicas e estruturais mostram que a hélice 12 deve ser deslocada da posição ativa para permitir a ligação de correpressor.[39,42–45] Tomados juntos, esses estudos sugerem que a hélice 12 é crucial para regular o recrutamento de cofator, mas o papel exato dela no receptor livre do ligante continua incerto.[46]

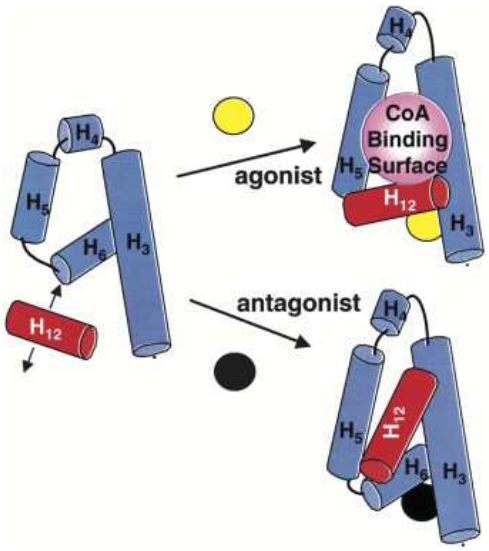


Figura 2: Representação do mecanismo clássico de ativação/inibição dos receptores nucleares pela hélice 12. Agonistas e antagonistas promovem diferentes conformações no receptor. A ligação do agonista resulta no reposicionamento da H12 próximo aos resíduos das hélices 3, 4, 5 e 6 criando uma superfície para a ligação do coativador. A ligação do antagonista inibe a capacidade da H12 adotar a conformação ativa. Ao invés disso, a hélice 12 ocupa uma posição que impede a ligação do coativador (retirado da ref. 47).

1.1.4. As Mutações

Mais de três mil diferentes genes humanos foram identificados onde uma ou mais modificações da sequência são causas diretas de doença.[48] Essas mutações podem afetar a função da proteína através de diversos mecanismos, tais como mudanças na transcrição, processamento de RNA, expressão da proteína, enovelamento da cadeia polipeptídica, estabilidade do estado enovelado, modificação pós-translacional, interações com ligantes parceiros e alterações na catálise. Uma análise do banco de dados de mutações genéticas humanas (HGMD)[48] mostra que a grande maioria de casos conhecidos agem através de mudanças na sequência codificante, com mutações pontuais (uma única base muda resultando em uma mudança de um único aminoácido da proteína). Este é, de longe, o efeito mais comum, contabilizando mais de 60% de todas as doenças de mutações monogênicas. Segundo um trabalho de Yue e colaboradores o mecanismo mais comum (mais de 80% dos casos) pelo qual uma mudança pontual na base resulta em doença é a desestabilização na estrutura da proteína relativa ao estado desenovelado. Observou-se que outros 10% de mutações pontuais afetam algum aspecto conhecido da função molecular e os 10% restantes operam através de outros mecanismos não identificados.[49]

1.2. O Receptor de Androgênio

Os efeitos dos androgênios são mediados por um simples receptor nuclear/intracelular, o receptor de androgênio (NR3C4, www.nursa.org),[50] uma proteína de 110 kDa pertencente à subfamília dos receptores esteroides, cujos membros compartilham homologia básica funcional e estrutural.[51,52] Ele age como um fator de transcrição controlado por hormônio que converte a mensagem tanto de androgênios naturais como sintéticos ao nível dos genes.

Embora uma isoforma do AR (ARα) tenha sido identificada como sendo uma proteína de 87 kDa truncada a partir do AR de 110 kDa (ARβ),[53] ainda é desconhecido se essas isoformas possuem capacidades diferenciadas de sinalização de transdução biológica em humanos. Contudo, sabe-se que ARα não é capaz de formar complexos

de alta afinidade com o DNA para modular os genes responsivos do AR.[54] Como o papel funcional da isoforma alfa continua pouco esclarecido, normalmente ao se referir ao receptor de androgênio, faz-se apenas em relação à isoforma beta, nomeada simplesmente como receptor de androgênio.

No citoplasma celular o AR não ligado esta associado a uma série de proteínas de choque térmico (HSPs, em inglês). O AR é ativado pela ligação de uma molécula de androgênio, normalmente 5α-diidrotestosterona, e a liberação das HSPs. A ligação do ligante também promove a hiperfosforilação do AR. O AR alterado assume uma conformação na qual os dois dedos de zinco no DBD ficam expostos. Ele é então translocado para o núcleo, onde ocorre a homodimerização e o homodímero liga-se aos elementos responsivos específicos de androgênio em certos genes, adquirindo a capacidade de regular a taxa de transcrição desses genes (Figura 3). Para exercer tal regulação, o complexo de um hormônio andrógeno e um AR deve interagir também com proteínas transcricionalmente ativas que se ligam ao elemento responsivo do DNA. Essas interações com uma série de correguladores, incluindo fatores de transcrição basais e proteínas promotoras centrais de elemento de ligação, determinam o controle vetorial sobre a expressão transcricional de uma dado gene alvo de androgênio.[55]

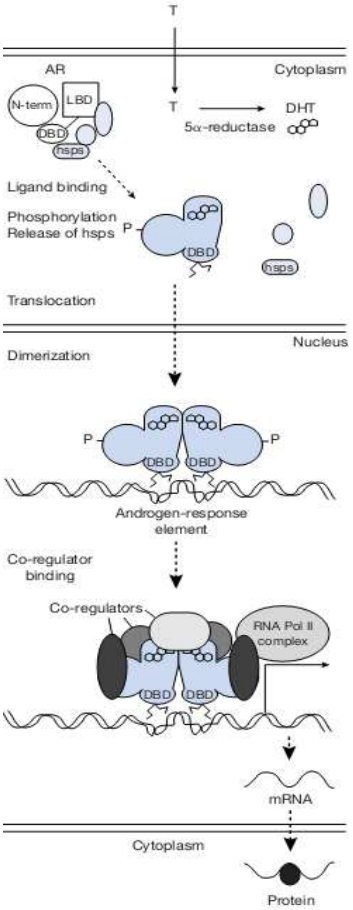
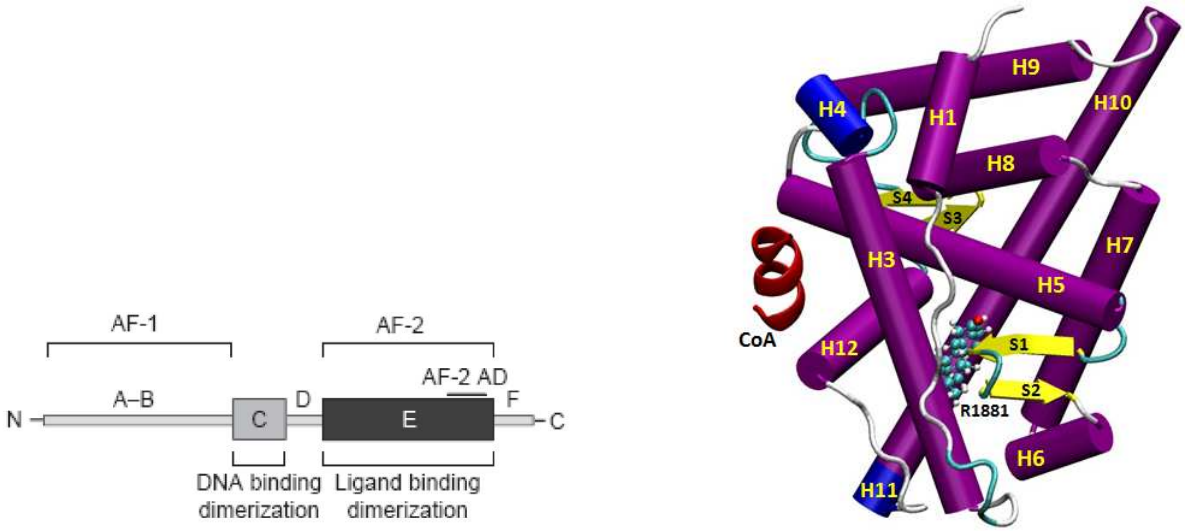


Figura 3: Representação do mecanismo da ação do androgênio na célula. Da ligação à ativação do gene para a expressão da proteína. T, testosterona; DHT, diidrotestosterona, N-term, região N-terminal; hsps, proteínas de choque térmico; P, sítio de fosforilação, RNA Pol II, RNA polimerase II. (retirado da ref. 55)

Como um membro da superfamília dos NRs, o receptor de androgênio é uma proteína modular organizada em domínios funcionais (Figura 4A).[56] A estrutura tridimensional do domínio N-terminal (NTD) ainda é imprecisa e este é o domínio menos conservado da família dos receptores esteroides; por exemplo, a similaridade de sequência entre o NTD dos receptores de androgênio e de progesterona é de apenas 20%. Ao contrário, a conservação do DBD é bastante alta (~80%) entre todos os receptores esteroides, exceto para o DBD do ER, cuja sequência é apenas 59% (ERα) similar ao do AR.[57]

A estrutura cristalográfica do AR-LBD revelou que ele possui um domínio globular composto por α-hélices dispostas em uma espécie de sanduíche de folhas helicoidais antiparalelas (Fig. 4B). O AR-LBD difere de outros NRs por possuir 11 α-

hélices e não 12, como os demais. Contudo, estas hélices são numeradas de 1 a 12, com a hélice 2 sendo omitida para manter a identidade posicional do AR em relação aos outros membros da subfamília dos receptores esteroides.



A B

Figura 4: (A) Representação esquemática da organização estrutural do AR humano. (retirada da ref. 58); (B) Representação da estrutura do AR-LBD. Em violeta: alfas-hélices; em azul: hélices 3-10; em amarelo: fitas beta. O ligante R1881 está representado no modelo “esferas e varetas”. O peptídeo mimetizando o coativador, CoA, está em vermelho.

O domínio de ligação com ligantes do AR (AR-LBD), importante para a ligação de esteroides, dimerização e interações proteína-proteína com proteínas corregulatórias,[59,60] possui diferentes regiões as quais desempenham um papel crítico no processo de transcrição. Alterações estruturais nestas regiões afetam de maneira significativa a função do receptor. A cavidade de ligação do ligante (LBP), a função de ativação 2 (AF2) e a extensão carbóxi-terminal, também conhecida como domínio F, são três destas regiões.

LBP: Um total de 22 resíduos de aminoácidos, situados dentro do LBD na chamada cavidade de ligação do ligante são responsáveis pela interação direta do receptor com o ligante (Figura 5A). Embora a maioria destas interações seja de natureza hidrofóbica (Figura 5B), o sítio de ligação também compreende, em ambas as extremidades do ligante, alguns resíduos polares que prendem firmemente a molécula do esteróide via

uma rede de ligações de hidrogênio (Figura 5C). Os resíduos que delineiam o LBP pertencem a cinco hélices (H3, H5, H7, H10, e H12), a uma fita β localizada entre H5 e H6 e a uma alça entre H11 e H12 (Figura 5B).[61,62]

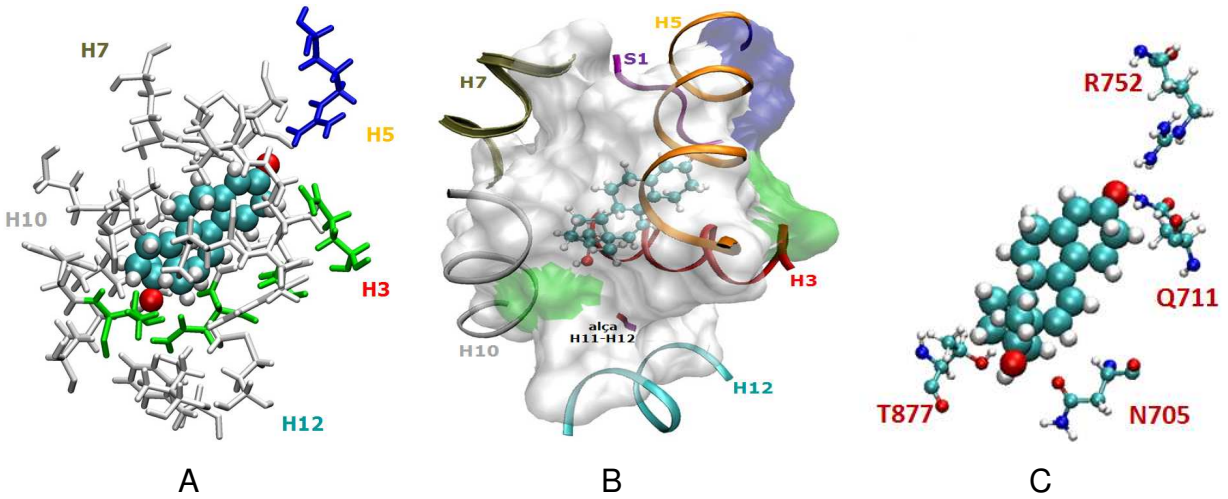


Figura 5: Representações da estrutura do AR-LBP. (A) Resíduos que fazem parte do sítio de ligação do R1881 coloridos segundo o tipo (apolar: branco, polar: verde e básico: azul); (B) Superfície de van der Waals mostrando a cavidade hidrofóbica do LBP; (C) Resíduos hidrofílicos que ancoram o ligante R1881 através de ligações de hidrogênio (os resíduos N705 e T877 ligam-se ao grupo OH-17 e os resíduos Q711 e R752 ao O-3 do grupo ceto). Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

AF2: Os receptores esteroides possuem duas regiões predominantes de ativação. A função de ativação 1 na região N-terminal e a função de ativação 2 que faz parte do LBD e do domínio de transativação amino-terminal e dependente do ligante. O AF2 é um sulco hidrofóbico altamente conservado contíguo ao LBP formado por resíduos pertencentes às hélices H3, H4, H5 e H12. Ele é flanqueado por resíduos de cargas opostas (os grampos de carga) que ligam motivos LXXLL do coativador do receptor de esteróide (SRC) da família de coativadores (Figura 6), onde X pode ser qualquer resíduo de aminoácido. O domínio AF2 do AR, entretanto, ao contrário dos demais receptores esteroides, demonstra uma afinidade maior pelos peptídeos contendo motivos FXXLF derivados do NTD e das proteínas correguladoras do AR do que pelos peptídeos contendo LXXLL derivados do coativador. A Figura 4B mostra um peptídeo

de um cofator ligado ao AR-LBD na região do AF2. A conservação de sequência do AF2 dos NR's reflete uma função comum da ligação do coativador.[63]

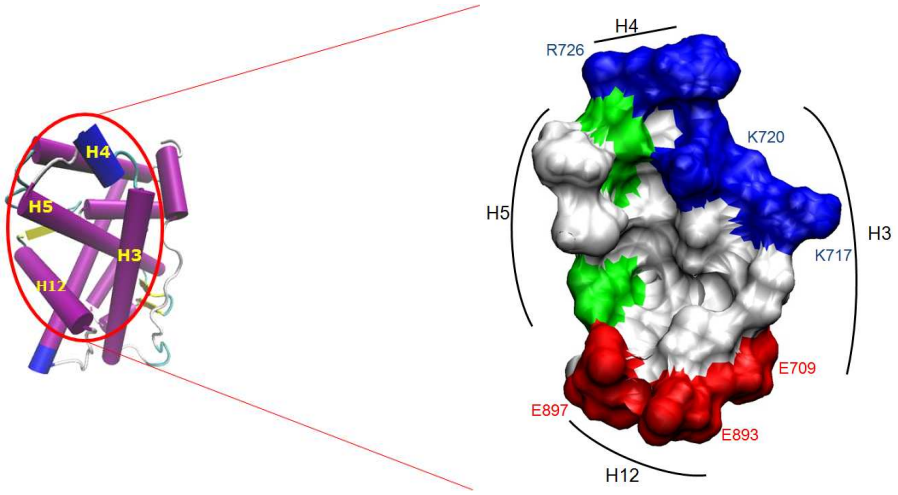


Figura 6: Representação da superfície do AF2 do AR-LBD. Os resíduos pertencentes ao grampo de cargas que flanqueiam o AF2 estão em vermelho (negativo) e azul (positivo). Os demais resíduos que compõem o AF2 estão em branco (apolar) e verde (polar).

Domínio F: Em particular, no caso da subfamília dos receptores esteroides existe ainda uma sequência curta carbóxi-terminal bem no fim do LBD. Esta extensão C-terminal, também conhecida como domínio F, desempenha um importante papel na ligação e estabilidade do ligante na cavidade de ligação e também na ativação do AR e dos outros receptores esteroides.[64–68] Embora não esteja diretamente associado à cavidade hidrofóbica, sua deleção completa ou parcial impede a ligação do agonista. No caso do receptor de androgênio humano (hAR) já foi demonstrado que a truncagem dos últimos 13 resíduos de aminoácidos do domínio F (resíduos 907 a 919) suprime a ligação do hormônio tanto com agonistas como com vários antagonistas[68] alterando a ativação do receptor.

As estruturas cristalográficas do AR, PR e GR mostraram que estes resíduos enovelam-se como uma fita β (S4) na conformação ativa e formam uma folha β conservada antiparalela com a fita β S3 entre H8 e H9 (Figura 4B).[60,69,70] No AR, o domínio F é mantido firmemente no lugar não apenas pelas interações de ligação de hidrogênio entre os grupos NH da cadeia principal e os átomos de CO de ambas as fitas adjacentes S3 (815-817) e S4 (911-913), mas também pelas interações de

algumas cadeias laterais dos seus resíduos com a parte interna do LBD (Figura 7). Em particular, os resíduos V911, I914 e F916 estão orientados de tal forma que eles estão envolvidos em fortes interações de van der Waals com resíduos das hélices 9, 10 e 5. Muitos dos resíduos que participam destas interações estabilizantes são especificamente conservados na subfamília dos receptores esteroides. Ademais, ensaios de mutação sítio-dirigidas do hAR-LBD mostraram que, dos 11 resíduos que formam o domínio F, as substituições do resíduos F916, I914 e V911 induziram as mudanças mais drásticas nas afinidades de ligação do ligante e na ativação do receptor sugerindo, portanto, a relevância fisiológica do arranjo observado. Parece, portanto, que a posição particular e a conformação adotada pela extensão C-terminal são necessárias para a ativação dos receptores esteroides e análises estruturais sugerem que a desestabilização da extensão C-terminal altera a funcionalidade do hAR.[6,68]

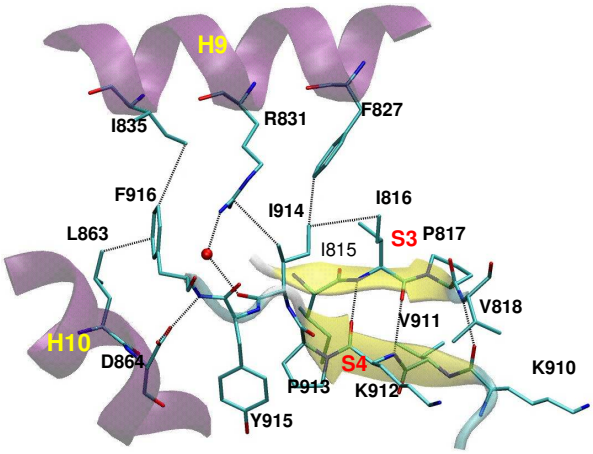
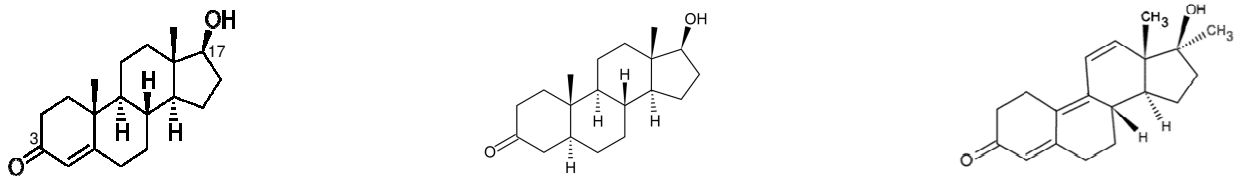


Figura 7: Representação das interações específicas entre os resíduos da fita β S4 pertencente à extensão C-terminal e outras partes do AR-LBD (S3, H9 e H10). Os resíduos incluídos fazem ou ligação de hidrogênio ou contatos de van der Waals. Os átomos estão coloridos segundo o tipo (carbono: azul; oxigênio: vermelho; nitrogênio: azul). Para maior clareza, os átomos de hidrogênio foram omitidos. A esfera vermelha representa uma molécula de água.

1.2.1. Ligantes do Receptor de Androgênio

Androgênio é qualquer composto natural ou sintético, geralmente um hormônio esteróide, que estimula ou controla o desenvolvimento e manutenção das características masculinas em vertebrados ao ligar-se ao receptor de androgênio.[51,52]

Androgênios são cruciais para o desenvolvimento, diferenciação e função dos órgãos reprodutivos masculinos. Eles também regulam características sexualmente dismórficas e processos em tecidos não genitais como crescimento de pelos faciais, púbicos e em outras partes do corpo, alargamento das cordas vocais, desenvolvimento de força muscular e comportamento masculino.[71] Testosterona (Figura 8a) e 5α-diidrotestosterona (DHT) (Figura 8b) são os dois mais potentes e importantes hormônios androgênios naturais, com a testosterona sendo produzida principalmente nas células de Leydig nos testículos e o DHT localmente a partir deste em tecidos alvo, tais como a próstata, pela enzima 5α-redutase. Androgênios também têm efeitos "não genômicos", i. e., alguns dos seus efeitos, por exemplo, em células neuroendócrinas, ocorrem tão rápido para ser mediado através da regulação de efeitos genômicos que ocorrem na presença de inibidores de transcrição.



(a) (b) (c)

Figura 8: Estruturas moleculares dos principais ligantes do AR. Naturais: (a) testosterona e (b) diidrotestosterona. Sintético: (c) metiltrienolona (R1881). A posição dos grupos polares está indicada na estrutura "a".

Em fêmeas, os efeitos da ação excessiva dos androgênios são amplamente conhecidos, mas o papel normal dos androgênios na fisiologia feminina é, entretanto, menos entendido. A glândula adrenal, tecidos periféricos e ovário são os principais tecidos produtores de androgênios em fêmeas.[57]

1.2.2. As mutações no AR e seus fenótipos.

Algumas mutações no AR estão associadas ao câncer de próstata humano e faz do AR um alvo bem estabelecido para o tratamento deste tipo de câncer,[60,61,72] enquanto outras resultam em graus variáveis de insensibilidade aos androgênios em indivíduos XY. A título de exemplo, o número total de mutações reportadas cresceu de

cerca de 600 em 2004 para mais de 1020 mutações, segundo a última atualização de 2012 do *The Androgen Receptor Gene Mutations Database* (ARDB).[55,73]

A síndrome de insensibilidade ao androgênio varia da forma completa (CAIS) à suave (MAIS), passando pela forma parcial (PAIS). Esta, que já foi conhecida como síndrome da feminilização testicular[74], é uma desordem hereditária que possui dentre outras características, o desenvolvimento de seios e outras características sexuais secundárias na puberdade, ausência ou escassez de pelos pubianos e nas axilas, genitália feminina externa e ausência de genitália feminina interna. No LBD existe uma surpreendente preponderância de mutações pontuais, com um número significativamente maior de casos CAIS[75,76] do que PAIS[6]. Diferentes substituições no mesmo aminoácido podem resultar em receptores com diferentes graus de AIS. Atualmente, são conhecidas mais de 400 diferentes mutações no AR que levam à AIS.[57,77,78]

O AR possui um papel central no crescimento e desenvolvimento normal da próstata, bem como na carcinogênese da próstata e na progressão da doença, tanto na fase dependente do androgênio como na fase não dependente.[79] A dependência do câncer de próstata pela sinalização do androgênio é conhecida a cerca de 70 anos e ainda assim o PCa continua sendo uma das principais causas de morte de homens no mundo ocidental. Avanços na metodologia e no tempo de diagnóstico têm permitido a detecção de tumores ainda no estágio inicial, quando a doença pode ser tratada cirurgicamente ou por radioterapia. Entretanto, apesar dos resultados positivos do tratamento, estágios avançados da doença apresentam um prognóstico muito mais difícil.[80] A biossíntese anormal de androgênio ou mutações que inativem o AR estão associadas com o desenvolvimento incompleto ou mesmo o não desenvolvimento da próstata em indivíduos XY. Portanto, uma via de sinalização normal é um requisito necessário não apenas para o desenvolvimento normal da próstata, mas também para o desenvolvimento do câncer de próstata.[81]

As mutações selecionadas para este estudo (Tabela 1) ocorrem naturalmente em indivíduos XY e foram todas reportadas em estudos clínicos. Além de estarem envolvidos nas doenças AIS e PCa, também estão associados a: (i) alterações significativas na constante de afinidade ligante-receptor; (ii) alterações na estrutura do

sítio de ligação, direta ou indiretamente; e (iii) alteração dos mecanismos de estabilização do domínio F. Além disso, escolheu-se estudar mutações que transformassem resíduos carregados em neutros e resíduos volumosos em resíduos menores e vice e versa. Isso nem sempre leva a modificações importantes na estrutura da proteína, mas como critério inicial se mostra válido, inclusive para o mapeamento do comportamento deste tipo de receptor.

Tabela 1: mutações selecionadas para o estudo. As cores fazem correspondência com as esferas que indicam a localização das mutações na Figura 14A.

Mutação	**Localização no LBD**	**Fenótipo**	**Afinidade relativa R1881 (%)**
F916A	F	PCa	1,3[68]
P817A	S3	PAIS	12[6]
R840C	H9	PAIS	36[82]
G743V	H5	CAIS	40[82]
R871G	H10	MAIS	51[83]
G909E	F	PCa	59[68]
Q919R	F	PCa	100[68]

1.3. Dinâmica Molecular

Não existe técnica experimental capaz de fornecer informações dinâmicas com o grau de completeza e precisão de processos moleculares que ocorrem na resolução temporal da ordem de femtossegundos e com uma resolução espacial atômica. Por exemplo, informações de como as moléculas de um determinado sistema químico se movem, quais são as energias de interação entre essas moléculas, quais são as correlações estruturais e dinâmicas e como essas correlações afetam a atividade bioquímica destas moléculas. Embora algumas técnicas modernas de espectroscopia[84] e cristalografia[85] resolvidas no tempo tenham obtido resultados importantes e promissores, as simulações de dinâmica molecular constituem uma ferramenta muito valiosa para a compreensão de fenômenos dinâmicos e de flutuações estruturais em escala atômica para sistemas biomoleculares.

Por isso, a simulação computacional tornou-se uma ferramenta científica chave no estudo de sistemas químicos e bioquímicos. O uso de técnicas de modelagem molecular já é bastante comum no estudo de uma grande variedade de sistemas químicos, tais como proteínas e DNA, e no estudo de suas interações com candidatos a agentes farmacêuticos. O uso cada vez maior de simulações para sistemas moleculares tem sido possível devido ao avanço na química teórica e computacional e ao rápido desenvolvimento de recursos computacionais mais acessíveis.[86,87]

Embora a primeira simulação de dinâmica molecular tenha sido feita em 1957,[88] apenas na década de 1970 é que foi possível simular o comportamento da água[89] e de biomoléculas[90]. A primeira simulação de dinâmica molecular de uma macromolécula de interesse biológico foi publicada em 1977 por McCammon e Karplus para o inibidor da tripsina pancreática bovina (BPTI). Embora essa simulação tenha sido feita na ausência de solvente, com um potencial de interação grosseiro comparado com os disponíveis atualmente e durado apenas 9,2 ps, os resultados foram fundamentais para substituir a visão das proteínas como estruturas relativamente rígidas pela percepção de que elas são sistemas dinâmicos, cujos movimentos internos desempenham um papel funcional. Contudo, deve-se registrar que dados experimentais, como os de experimentos de troca de hidrogênio feitos por Linderstrom-Lang e colaboradores[91] já existiam e apontavam nessa direção.

Dois anos após a simulação do BPTI, foi reconhecido[92] que o fator B de temperatura calculados durante o refinamento da estrutura cristalográfica obtida por raios-X podiam ser usados para inferir movimentos internos de proteínas. Durante os 10 anos seguintes, uma grande quantidade de fenômenos de movimento foi investigada por simulações de dinâmica molecular de proteínas e de ácidos nucleicos. A maioria desses estudos focou nos aspectos físicos dos movimentos internos e na interpretação dos experimentos.[93]

A simulação computacional é uma importante ferramenta também para o entendimento da ligação de ligantes a proteínas. Enquanto o poder da biologia estrutural está em traduzir o conhecimento das estruturas proteicas em percepções sobre suas forças, ligações e mecanismos, o poder complementar da simulação computacional está em mostrar “o resto da história”, isto é, como movimentos em

conjuntos de confôrmeros alternativos e as entropias e forças que não podem ser vistas em estruturas moleculares únicas também contribuem para as afinidade de ligação. Na ligação a uma proteína, um ligante pode ligar-se em múltiplas orientações; a proteína ou o ligante pode ser deformado pelo evento da ligação; moléculas de água, íons ou cofatores podem ter envolvimento inesperado; e entropias conformacionais ou de solvatação podem às vezes desempenhar papéis maiores e outras vezes imprevisíveis. E a simulação computacional está ajudando a elucidar esses fatores.[94]

Em resumo, simulações computacionais podem fornecer o detalhe fundamental que permite a compreensão de questões específicas sobre as propriedades de um sistema modelo, frequentemente mais facilmente do que experimentos com um sistema real. Para muitos aspectos da função biomolecular, são esses detalhes que despertam interesse.[93]

1.3.1. Fundamentos teóricos

Idealmente, a equação de Schrödinger dependente do tempo deveria ser capaz de prever todas as propriedades de qualquer sistema com precisão. Entretanto, quando uma grande quantidade de partículas está envolvida, como no caso de simulações de biomoléculas, é necessário introduzir aproximações para que haja um ganho no tempo de cálculo da propriedade desejada. Simulações computacionais dependem de um modelo capaz de descrever as interações atômicas em um sistema químico. Para a maioria de sistemas biomoleculares, portanto, costuma-se escolher trabalhar com parametrizações empíricas dos modelos; por exemplo, interações coulômbicas clássicas entre cargas atômicas pontuais ao invés da descrição quântica dos elétrons são expressas em funções de energias potenciais. Estas formas funcionais analíticas, juntamente com seus parâmetros, relacionam a estrutura química e a conformação à energia no chamado campo de força empírico.[86,95]

Na prática, campos de força clássicos geralmente incluem um termo de energia cinética e termos para estiramento de ligação, deformação angular, torções diedrais e deformação fora do plano, bem como interações não covalentes: de van der Waals e eletrostáticas. Para a maioria dos propósitos, essas aproximações funcionam bem, mas

elas não podem reproduzir efeitos quânticos tais como formação ou quebra de ligações.[95] Uma forma funcional típica para uma Hamiltoniana contendo estes termos é:[86]

$$H\left(r^N, p^N\right) = \frac{1}{2}\sum_{i}^{N}\frac{p_i^2}{m_i} + \sum_{n=1}^{N_b}\frac{1}{2}C_{b,n}(b_n - b_{0,n})^2 + \sum_{n=1}^{N_\theta}\frac{1}{2}C_{\theta,n}(\theta_n - \theta_{0,n})^2 + \sum_{n=1}^{N_\xi}\frac{1}{2}C_{\xi,n}(\xi_n - \xi_{0,n})^2 +$$

$$+\sum_{n=1}^{N_\varphi}\frac{1}{2}C_{\varphi,n}[1+\cos(m_n\varphi_n - \delta_n)] + \sum_{i}^{N-1}\sum_{j=i+1}^{N}\left[\frac{C_{12,ij}}{r_{ij}^{12}} - \frac{C_{6,ij}}{r_{ij}^{6}}\right] + \sum_{i}^{N-1}\sum_{j=i+1}^{N}\left[\frac{q_i q_j}{4\pi\varepsilon_r\varepsilon_0 r_{ij}}\right] \quad \textbf{(1)}$$

$H(r^N, p^N)$ é a hamiltoniana clássica que descreve as interações no sistema em termos das coordenadas **r** e momentos **p** de todas as partículas *i* e *j*; C_b e b_0 são a constante de estiramento e a distância de equilíbrio da ligação; C_θ e θ_0 são a constante de deformação angular e o ângulo de equilíbrio da ligação; C_ξ e ξ_0 são a constante de deformação angular impróprio e o ângulo impróprio de equilíbrio da ligação; C_φ, φ e δ são a constante de força torsional, a periodicidade do termo torsional e a fase do ângulo; $C_{12,ij}$ e $C_{6,ij}$ são os coeficientes de repulsão e de atração não ligados (Lennard-Jones); r_{ij} é a distância interatômica entre os átomos i e j; q_i e q_j são as cargas atômicas parciais dos átomos i e j; e ε_0 e ε_r são a constante dielétrica efetiva é a constante dielétrica relativa do meio.

Campos de força mais elaborados podem incluir termos cruzados entre interações covalentes, termos representando polarização eletrônica ou termos para a interação com campos externos.

Dinâmicas biomoleculares ocorrem em uma larga faixa de escalas espacial e temporal e a escolha da abordagem para estudá-las depende da questão a ser respondida (Figura 9).

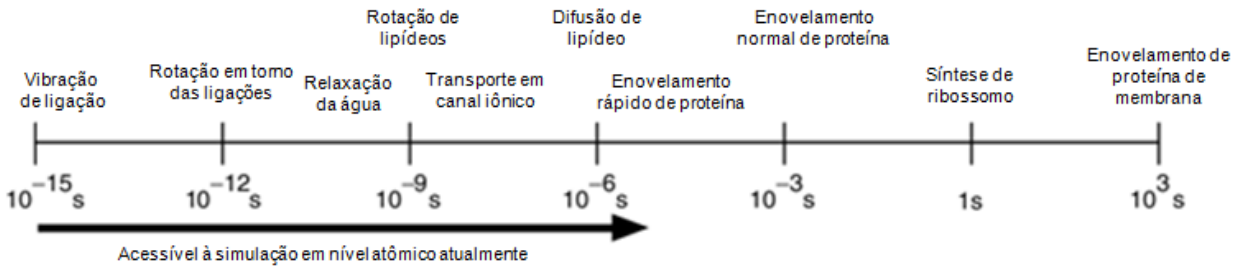


Figura 9: Faixa de escala temporal para dinâmicas de sistemas biomoleculares. Embora os passos temporais individuais de dinâmicas moleculares seja de 1 a 2 fs, a computação paralela permite simular em uma escala de microssegundos e técnicas de computação distribuída podem amostrar mesmo processos mais lentos, quase atingindo milissegundos.

Para obter propriedades a partir das simulações moleculares uma coleção de configurações do sistema de estudo precisa ser gerada (*ensemble*) de tal forma que ao menos as regiões mais populosas do espaço de fase sejam suficientemente representadas. Dois métodos largamente utilizados para gerar um *ensemble* termodinamicamente relevante a partir de uma Hamiltoniana como a mostrada na equação **1** são o método Monte Carlo (MC) e o método de dinâmica molecular (MD). Sendo que este último tem a vantagem de reproduzir propriedades dinâmicas.[86,95]

Nas simulações de Monte Carlo, configurações escolhidas aleatoriamente são aceitas ou rejeitadas com base na probabilidade de Boltzmann. Estas configurações são subsequentemente usadas para obter médias de *ensemble* para as propriedades de interesse.[86]

Em simulações de dinâmica molecular a derivada do gradiente do potencial é usada para calcular as forças em cada partícula do sistema. Usando-se um algoritmo de integração adequado, essas forças são usadas para obter a evolução temporal do sistema. As propriedades são obtidas como médias temporais:[86]

$$\overline{\Psi} = \frac{1}{\tau}\int_0^{\tau}\Psi(t)\,dt \approx \frac{1}{N_{md}}\sum_i^{N_{md}}\Psi(t_i) \qquad \textbf{(2)}$$

Ψ é a propriedade desejada calculada ao longo do tempo t e N_{md} é o número total de passos de dinâmica molecular.

Valores esperados são substituídos por médias do *ensemble* das simulações de Monte Carlo ou médias temporais das simulações de dinâmica molecular, assumindo-se que o espaço de fase foi adequadamente amostrado no *ensemble* simulado ou na trajetória dinâmica. Propriedades dinâmicas e estruturais podem ser prontamente obtidas como médias de *ensemble* ou de médias temporais se um valor para esta propriedade puder ser calculado a partir de uma configuração simples. Energia potencial, energia cinética, energias de ligação, temperatura e pressão são exemplos de tais propriedades. Uma vez que simulações de dinâmica molecular seguem a evolução temporal de um sistema, muitas outras propriedades também estão disponíveis, como coeficientes de difusão e tempos de relaxação. Diferentemente, energias livres e entropias dependem da extensão total do espaço de fase e não podem ser facilmente derivadas das médias estatísticas que as simulações fornecem.[86,87]

Na prática, sistemas fisicamente interessantes possuem tantos graus de liberdade que não é possível uma completa amostragem do espaço de fase. A principal dificuldade prática enfrentada nos cálculos surge da complexidade e irregularidade subjacente do panorama energético das flutuações conformacionais dos sistemas de macromoléculas biológicas.[86,87]

O algoritmo básico de simulação biomolecular não mudou substancialmente desde o primeiro estudo de MD de uma proteína há mais de 30 anos.[96] Embora simulações rotineiras de MD com solvente explícito estejam atualmente em 4 ou 5 ordens de magnitude maiores (100-1000 ns geralmente), estudos modernos de MD ainda parecem serem curtos frente ao que é necessário para simulações de equilíbrio estatisticamente válidas. Grosso modo, seria mais interessante correr simulações pelo menos 10 vezes maiores do que a escala de tempo importante mais lenta em um sistema. Infelizmente, muitas escalas de tempo biomoleculares excedem o 1 ms e, em alguns casos, por ordens de magnitudes.

1.3.2. Etapas de uma simulação de Dinâmica Molecular

Algumas técnicas são necessárias para simular um sistema de partículas usando dinâmica molecular para cada etapa do processo de descrição das posições e velocidades das partículas a cada instante. De uma maneira geral as etapas incluem a geração da configuração inicial do sistema; o cálculo das forças exercidas sobre cada partícula por conta das interações intermoleculares; movimentação das partículas; os *ensembles* e controle da simulação; e a análise ou armazenamento de configurações (trajetórias)[97] (Figura 10).

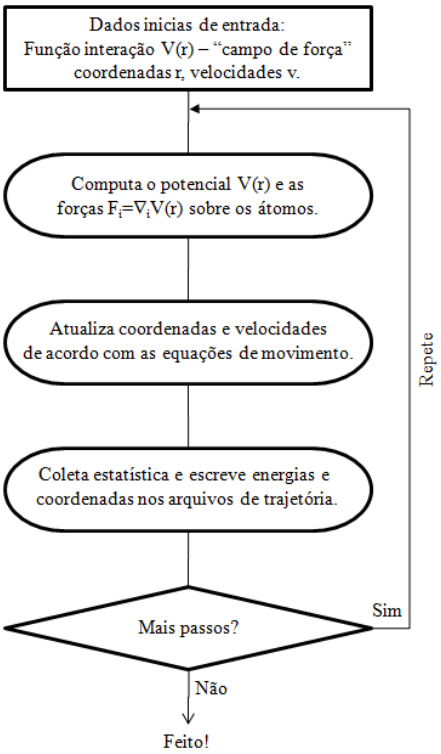


Figura 10: Fluxograma simplificado das etapas de uma simulação de dinâmica molecular típica. A ideia básica é gerar estruturas a partir de um conjunto natural calculando as funções potenciais e integrando as equações de movimento de Newton; essas estruturas são então usadas para estimar propriedades de equilíbrio do sistema. Um passo temporal típico é da ordem de 1 a 2 fs.

Entretanto, simulações utilizando estruturas experimentais como partida sempre precisarão de um chamado estágio de equilibração antes que trajetórias utilizáveis possam ser coletadas. É entendido que o propósito da equilibração é imprescindível para eliminação dos artefatos dinâmicos e estruturais das condições iniciais de não

equilíbrio. Portanto, embora esse estágio da simulação não tenha qualquer pretensão quanto à exploração de todo espaço de fase de equilíbrio, ele é necessário para identificar o ponto de partida da tal exploração. Deve-se notar que qualquer método para a equilibração que começa com uma estrutura determinada experimentalmente possui um ponto fraco principal: o sistema pode ficar preso em um mínimo local de energia que não é o poço nativo da proteína, dependendo das propriedades físicas da proteína e da qualidade da estrutura experimental.[98]

A Figura 11 resume os passos de uma simulação típica de dinâmica molecular de uma proteína partindo de uma estrutura obtida do *Protein Data Bank* (PDB) passando pela etapas de minimização, equilibração e produção e posterior análise da trajetória de simulação.

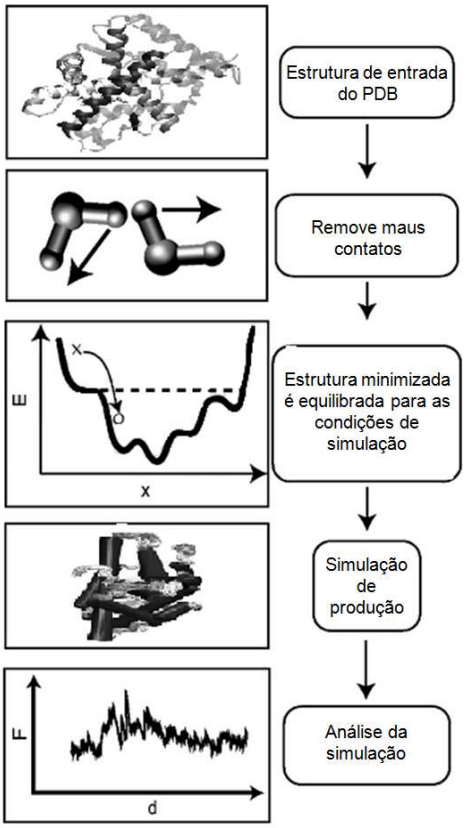


Figura 11: Resumo dos passos em uma simulação típica de MD de proteína. O primeiro passo é importar uma estrutura inicial do PDB. Em seguida, solvente é adicionado e a estrutura é relaxada via um algoritmo de minimização para remover maus contatos. Então o sistema é equilibrado: a estrutura experimental minimizada é submetida à simulação de MD para trazê-la para o equilíbrio com as condições de simulação. Apenas após esse estágio de equilibração podem as simulações de produção serem realizadas para fornecerem resultados realísticos (adaptada da ref. 98).

Várias técnicas para a definição deste equilíbrio existem e já foram amplamente discutidas na literatura. Entretanto, uma técnica muito comum é a análise do RMSD (root mean square deviation, em inglês). Quando o RMSD alcança um valor de platô, o sistema atingiu um poço na superfície de energia potencial. Supõe-se que esse poço energético corresponde ao espaço de fase no equilíbrio e o sistema é considerado apropriado para a etapa de produção da simulação de dinâmica molecular.[98]

1.3.3. Geração da configuração inicial

No caso do sistema ser isotrópico homogêneo (um fluído, por exemplo), o procedimento mais usual é criar uma caixa de simulação cúbica, podendo ser também de outra geometria, onde as N moléculas do sistema são colocadas. As dimensões da caixa são escolhidas de tal forma que a densidade numérica, N/V, corresponda àquela do sistema real nas condições termodinâmicas desejadas.[97]

A construção de caixas de simulação de sistemas mais complexos, como soluções de proteínas, não é possível sem o conhecimento prévio da estrutura da macromolécula. Por conta disso, as estruturas terciárias de uma proteína podem ser obtidas a partir de diferentes técnicas experimentais, como cristalografia por difração de raios-X, espectroscopia de RMN e, mais recentemente, de técnicas integradas de proteômica com espectrometria de massas.[99]

Normalmente, como configuração inicial para estudos de dinâmica molecular de estruturas de proteínas inteiras ou de domínios grandes, tais como o LBD dos NRs, utilizam-se as estruturas obtidas por cristalografia por apresentarem melhor resolução. O dado colhido a partir de um experimento de difração de raios-X não é a estrutura de equilíbrio da proteína; mais precisamente, é o mapa da densidade de probabilidade eletrônica, produzido sobre muitas células unitárias, a partir das quais uma estrutura média pode ser determinada. Isso fornece um “chute” inicial para um conjunto de posições atômicas. Estruturas obtidas a partir de ressonância magnética nuclear possuem limitações semelhantes.[97,98]

Quando uma configuração de partida está muito longe da de equilíbrio, forças grandes, devido à sobreposição de átomos, por exemplo, podem causar a queda da

simulação ou a distorção do sistema e, nesse caso, é necessário realizar uma minimização de energia do sistema antes da simulação de dinâmica molecular. A minimização de energia é uma técnica que visa a encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema. O procedimento básico consiste em "caminhar" sobre a superfície de potencial na direção em que a energia decresce até o sistema atingir um mínimo local de energia. Minimizações de energia também são normalmente usadas para refinar estruturas experimentais de baixa resolução.[95,100]

Os algoritmos de minimização mais conhecidos são: o método steepest descent, o método dos gradientes conjugados e o método de Newton-Raphson.[87]

1.3.4. Condições periódicas de contorno

Em simulações de MD, os sistemas possuem normalmente 10^2-10^6 partículas, sendo muitas ordens de grandeza menores em comparação com o número de átomos de um sistema macroscópico. Assim, o número relativo de partículas na superfície do sistema é muito maior em sistemas de MD do que para um sistema macroscópico, podendo causar efeitos de superfície indesejados.[100]

Tipicamente, simulações de biomoléculas usam condições periódicas de contorno para evitar efeitos de superfície, ou efeitos de fronteira, de tal forma que uma molécula de água que sai pela direita reaparece na esquerda (Figura 12). Se a caixa é suficientemente grande, as moléculas não irão interagir significativamente com suas cópias periódicas. Réplicas idênticas da caixa de simulação são consideradas como estando dispostas ao redor da célula principal, formando um sistema que procure reproduzir de forma mais relista o sistema termodinâmico (com N e V tendendo ao infinito, mas N/V uma constante). Isso está intimamente relacionado às interações não ligadas, que idealmente deveriam ser somadas sobre todos os vizinhos no sistema periódico infinito resultante.[95,97]

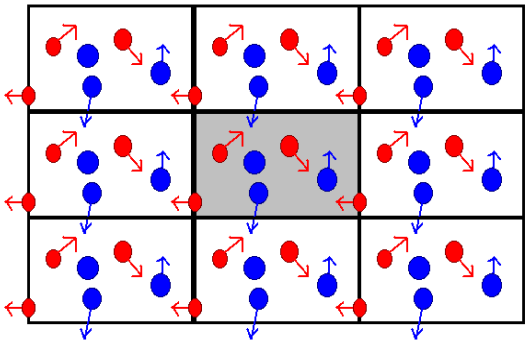


Figura 12: Representação de uma caixa de simulação bidimensional (em cinza) e suas réplicas periódicas (em branco). As condições periódicas de contorno correspondem à criação de réplicas idênticas da caixa de simulação colocadas lado a lado, de modo que qualquer movimento que ocorra na caixa de simulação ocorra também nas réplicas (retirado de http://kalilbn.wordpress.com, acessado em 20/10/2012).

1.3.5. Convenção da imagem mínima e raio de corte

A implementação das condições periódicas de contorno requer a avaliação das forças exercidas sobre cada molécula por todas as demais do sistema. Portanto, esse tratamento deve ser acompanhado por uma indicação de como as interações entre as imagens de diferentes réplicas periódicas devem ser consideradas. Para o tratamento das interações de curto alcance, onde $V \sim r^{-6}$ (interações de van der Waals), emprega-se o procedimento de convenção da imagem mínima.[87] Nesse caso, é empregada a truncagem do potencial em um raio de corte esférico. Cortes simples podem funcionar para interações de Lennard-Jones que decaem muito rapidamente, mas, para interações de longo alcance (interações de Coulomb), um corte súbito pode levar a grandes erros. Uma alternativa é “desligar” a interação antes do corte, como mostrado na Figura 13, mas uma melhor opção é usar a soma de Ewald para calcular as interações eletrostáticas infinitas pela separação da soma em partes de curto e de longo alcance. Para PME, o corte determina apenas o balanço entre as duas partes e a parte de longo alcance é tratada pela fixação de cargas em uma grade que é resolvido em um espaço recíproco através de transformadas de Fourier.[87,95,97,100]

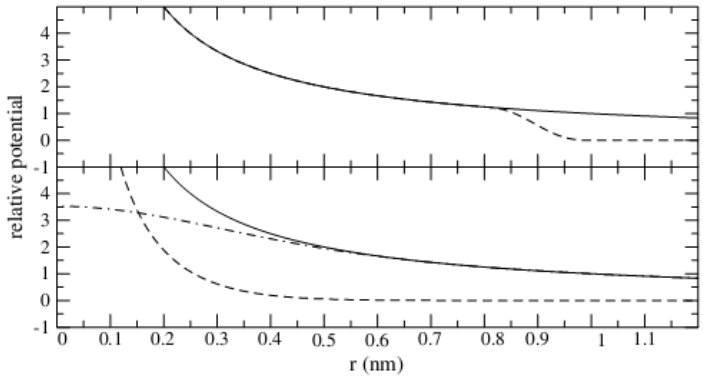


Figura 13: Alternativas a um corte brusco para interações coulômbicas não ligadas. Em cima: desligando a interação (tracejada) antes do corte, a força será a derivada exata do potencial, mas a derivada (e, portanto, a força) crescerá extraordinariamente logo após o corte. Em baixo: PME é um algoritmo no qual a interação de Coulomb (sólida) é dividida em um termo de curto alcance que é resolvido dentro de um corte (tracejada) e um termo de longo alcance que pode ser resolvido exatamente em um espaço recíproco com transformadas de Fourier (ponto e traço). (retirado da ref. 95)

1.3.6. Controle da temperatura e da pressão

Cortes e arredondamento de erros podem levar a desvios de energia, que fazem o sistema aquecer durante a simulação. Para controlar isso, o sistema é normalmente acoplado a um termostato que escalona a velocidade durante a integração para manter a temperatura ambiente. De forma parecida, a pressão total no sistema pode ser ajustada através do escalonamento do tamanho da caixa de simulação, isotropicamente ou separadamente nas dimensões x, y e z.[95]

Para gerar a distribuição correta do *ensemble* o sistema pode ser acoplado a um reservatório com um banho determinístico ou estocástico. Acoplamentos determinísticos geralmente possuem alguma quantidade que é conservada (similar a energia total).[101]

Durante as simulações de dinâmica molecular o controle de temperatura pode ser obtido através de um processo conhecido como termostato de Berendsen,[102] que mantém a temperatura acoplando o sistema é a um banho térmico com temperatura fixa. Durante as simulações as velocidades são reescalonadas a cada passo de

integração, a fim de ajustar a energia cinética do sistema até a temperatura selecionada.[102]

Outros métodos mais sofisticados como os termostatos de Nosé-Hoover[103,104] e Nosé-Poincaré[104] também são utilizados. Para o controle da pressão, as variáveis de temperatura são substituídas pelas de pressão, e as velocidades dos átomos pelas coordenadas atômicas. Os barostatos mais utilizados são os de Berendsen[102] e Parrinello-Rahaman.[105] O programa NAMD[101], contudo, usa uma abordagem de banhos estocásticos devido à sua facilidade de implementação e também pelo fato de que os termos de fricção tendem a melhorar a estabilidade dinâmica. Para simulações NpT, como a deste trabalho, Phillips e colaboradores propuseram um novo conjunto de equações de movimento e implementaram um integrador numérico no NAMD.[101]

1.3.7. Vínculos e restrições

A parte de maior demanda das simulações é o computo das interações não covalentes, devido aos milhões de pares que têm de ser avaliados a cada passo de tempo. Estender o passo de tempo é, portanto, um importante modo de melhorar performance da simulação, mas, infelizmente, erros são introduzidos nas vibrações de ligação já em 1 fs. Entretanto, na maioria das simulações, as vibrações de ligação não são de interesse por si e podem ser removidas inteiramente pela introdução de algoritmos de restrição de ligação como o SHAKE,[106] o RATTLE[107] ou LINCS.[108] Tais restrições tornam possível estender os passos de tempo para 2 fs. [95]

1.3.8. Tratamento do Solvente

Dois são os modelos de solvatação utilizados em simulações de MD: o modelo de solvatação implícita e o de solvatação explícita. Neste último, milhares de moléculas de solvente são incorporadas ao sistema como um componente adicional. Dentre os vários potenciais desenvolvidos e parametrizados para descrever moléculas de água explicitamente, os mais utilizados são os da família TIP (Transferable Intermolecular Potentials), desenvolvidos por Jorgensen e colaboradores:[109,110] TIP3P, TIP4P e TIP5P,

os modelos de carga pontual, SPC e SPC/E, desenvolvidos por Berendsen e colaboradores[111] e o modelo ST2 de Stillinger e Rahman[112].

Em casos de sistemas muito grandes, envolvendo complexos de proteínas com DNA e cofatores, por exemplo, o uso de modelos implícitos de solvatação podem ser mais interessantes do ponto de vista do custo computacional da simulação, dependendo da propriedade a ser estudada. Em geral, um tratamento implícito considera a influência média do solvente através da estimativa direta da energia livre de solvatação, mas não considera os graus de liberdade do solvente explicitamente e sim como um meio contínuo. Os modelos de solvatação implícita mais comumente usados são o de área superficial acessível ao solvente,[113,114] o de Poisson-Boltzmann[115,116] e o modelo generalizado de Born.[117,118] A combinação do modelo generalizado de Born com o de Poisson-Boltzmann tem sido reconhecida como uma boa escolha para o tratamento de implícito do solvente em sistemas biomoleculares.

1.3.9. Velocidades iniciais

Além das configurações iniciais, o método MD requer especificação das velocidades iniciais de todos os átomos do sistema, para que as equações de movimento possam ser resolvidas. Dentre as maneiras pelas quais as velocidades iniciais podem ser estabelecidas, uma possibilidade é a partir da distribuição de Maxwell-Boltzmann na temperatura de interesse:[87]

$$p(v_{ix}) = \left(\frac{m_i}{2\pi K_b T}\right)^{1/2} exp\left[-\frac{1}{2}\frac{m_i v_{ix}^2}{K_b T}\right] \quad \textbf{(3)}$$

A equação de Maxwell-Boltzmann fornece a probabilidade p que um átomo i de massa m_i tenha uma velocidade v_{ix} na direção x na temperatura T. Sendo k_b a constante de Boltzmann.

Desta forma, a energia cinética do sistema é determinada pela temperatura especificada por:

$$\frac{3}{2}Nk_BT = \frac{1}{2}\sum_{i=1}^{N} m_i v_i^2 \quad \textbf{(4)}$$

em que N é o número de partículas e m_i e v_i são a massa e a velocidade da partícula i, respectivamente. Seja qual for a escolha, é importante que a soma dos momentos sobre todas as partículas do sistema seja um vetor nulo, evitando o deslocamento da caixa como um todo.

1.3.10. Movimentação das partículas

Em dinâmica molecular, configurações sucessivas do sistema são geradas através da integração das leis de movimento de Newton. O resultado é uma trajetória que especifica como as posições e velocidades das partículas no sistema variam em função do tempo. A trajetória é obtida resolvendo equações diferenciais incorporadas na segunda lei de Newton.

Existem muitos algorítimos para a integração das equações de movimento usando métodos de diferenças finitas, contudo, o algoritmo de Verlet[119] é provavelmente o método mais usado para integrar as equações de movimento em uma simulação de dinâmica molecular. O algoritmo de Verlet usa as posições e acelerações no tempo *t* e as posições do passo anterior, r(t - δt), para calcular as novas posições em t + δt, r(t + δt). As relações entre essas quantidades e as velocidades no tempo *t* podem ser escritas como:[87]

$$\boldsymbol{r}(t+\delta t) = \boldsymbol{r}(t) + \boldsymbol{v}(t)\delta t + \frac{1}{2}\boldsymbol{a}(t)\delta t^2 + \cdots \quad \textbf{(5)}$$

$$\boldsymbol{r}(t-\delta t) = \boldsymbol{r}(t) - \boldsymbol{v}(t)\delta t + \frac{1}{2}\boldsymbol{a}(t)\delta t^2 - \cdots \quad \textbf{(6)}$$

Onde **r** é a posição da partícula, **v** sua velocidade e **a** representa a aceleração experimentada por ela.

Somando-se as equações (5) e (6), tem-se:

$$\boldsymbol{r}(t+\delta t) = 2\boldsymbol{r}(t) - \boldsymbol{r}(t-\delta t) + \boldsymbol{a}(t)\delta t^2 \quad \textbf{(7)}$$

As velocidades não aparecem explicitamente no algoritmo de integração de Verlet, mas podem ser calculadas de muitas formas. Uma forma simples é dividir a diferença nas posições em t+δt e t-δt por 2δt:

$$\boldsymbol{v}(t) = [\boldsymbol{r}(t + \delta t) - \boldsymbol{r}(t - \delta t)]/2\delta t \qquad \textbf{(8)}$$

A implementação do algoritmo de Verlet é direto e os requisitos de armazenagem são modestos. O método é temporalmente reversível, conserva o momento angular e requer apenas uma avaliação das forças para cada passo temporal. Uma desvantagem do método é que, para um período fixo de tempo, o método exibe um erro global proporcional a δt^2 o que pode levar a uma perda de precisão.[87,101]

Embora já existam métodos mais exatos (de maior ordem) para a integração de equações de movimento, eles não são convenientes para sistemas biomoleculares porque requerem várias avaliações das forças para cada passo temporal. Desta forma, o algoritimo de Verlet e suas variações (leap-frog[120] e Verlet-velocidade,[121] por exemplo) ainda continuam sendo a opção mais prática e acurada para simulações biomoleculares.

1.3.11. Parametrização do Ligante R1881

Foi necessário realizar a parametrização do ligante R1881 uma vez que faltavam alguns parâmetros no campo de força CHARMM[122] correspondentes a cargas parciais atômicas, comprimentos e ângulos de ligação e ângulos diedros deste ligante.

Parametrizar o ligante consiste basicamente em ajustar ou otimizar um conjunto de parâmetros empíricos que dependem da molécula à forma funcional do campo de força (Equação 1). De um modo geral, os campos de força têm seus parâmetros ajustados de modo a reproduzir algumas propriedades estruturais ou termodinâmicas obtidas experimentalmente. Nos casos em que dados termodinâmicos são escassos outra forma de determinar parâmetros são os cálculos de estrutura eletrônica *ab initio*. O próprio campo de força CHARMM, por exemplo, foi desenvolvido tanto a partir de dados experimentais como de cálculos quânticos.

Uma importante característica dos campos de força é a transferabilidade, isto é, um mesmo conjunto de parâmetros pode ser utilizado para modelar moléculas relacionadas, ao invés de ter que definir um novo conjunto de parâmetros para cada molécula individualmente.[87] Parâmetros relacionados a graus de liberdade "rígidos", como estiramento de ligação e deformação angular, influenciam pouco o desempenho de um campo de força e são especialmente transferíveis. Também costuma-se assumir que um mesmo conjunto de parâmetros de van der Waals pode ser usado para a maioria, senão todos, os tipos atômicos de um mesmo elemento. As carga atômicas, por sua vez, podem ser determinadas por ajuste do potencial eletrostático e os parâmetros de torção são comumente obtidos através de cálculos de mecânica quântica.

2. Objetivo da Tese

2.1. Objetivo geral

Através de simulações por dinâmica molecular, estudar a estrutura e dinâmica do domínio de ligação do ligante do receptor de androgênio, AR-LBD, em suas formas nativa e mutadas, complexadas com o ligante sintético metiltrienolona (R1881).

2.2. Objetivos específicos

- determinar as bases moleculares da diminuição da capacidade funcional do receptor de AR decorrentes das seguintes mutações: F916A, P817A, R840C, G743V, R871G, G909E e Q919R.
- buscar por alterações estruturais na cavidade de ligação do ligante (LBP), na superfície de ligação do cofator (AF2); e na extensão C-terminal (domínio F).
- estudar como alterações na posição da extensão C-terminal (domínio F) estão relacionadas à diminuição da capacidade funcional do AR.

3. Materiais e Métodos

A preparação do sistema para a simulação por dinâmica molecular consiste basicamente em escolher uma estrutura de partida, no caso, estruturas cristalográficas disponíveis no *Protein Data Bank* (http://www.rcsb.org/pdb). Para este estudo, a estrutura cristalográfica de partida principal foi a de PDB ID 1XOW[63] que consiste na estrutura do LBD do receptor humano de androgênio ligado ao agonista artificial R1881 e também a um peptídeo NH_2-terminal (peptídeo AR 20-30) que mimetiza parte de um coativador. Esta estrutura será denominada AR-LBD-Cofator/R1881. A estrutura de PDB ID 1E3G[60] foi usada para modelar as partes faltantes das estruturas de partida: o loop anterior à hélice 1(resíduos 669 e 670) e parte do loop entre as hélice 8 e 9 (resíduos 844 e 849). O programa psfgen implementado no pacote NAMD[101] foi usado para adicionar os átomos de hidrogênio faltantes.

As mutações foram construídas trocando-se computacionalmente os resíduos de aminoácidos desejados, já que estruturas cristalográficas do AR-LBD contendo as mutações escolhidas para este trabalho não estão disponíveis até o momento. É importante ressaltar que todas as mutações escolhidas ocorrem naturalmente em indivíduos XY e fazem parte de estudos clínicos reportados (Tabela 1). O programa psfgen foi novamente usado para reconstruir as cadeias laterais trocadas computacionalmente.

Todos os modelos foram imersos em uma caixa cúbica com moléculas de água utilizando o *plugin solvate* implementado no programa VMD[123]; íons Na^+ e Cl^- foram adicionados ao sistema para manter a neutralidade e mimetizar as condições fisiológicas, na concentração de 0,15 $mol.L^{-1}$, usando-se o *plugin autoionize* também implementado no VMD.[123] Os sistemas assim construídos totalizaram aproximadamente 55 mil átomos cada um.

Todas as simulações de MD foram feitas com o programa NAMD 2.6[101] utilizando-se o algorítimo *SHAKE*[107] para restringir os movimentos dos átomos de hidrogênio e o campo de força de átomos explícitos CHARMM22.[122] O passo de

integração das equações de movimento nas simulações foi de 2 fs com as configurações sendo gravadas a cada 2 ps.

Condições periódicas de contorno foram aplicadas. O método de *Particle Mesh Ewald* (PME) foi empregado para calcular as interações eletrostáticas de longo alcance. Todas as simulações foram feitas no *ensemble* NpT, usando-se o termostato de Langevin onde a temperatura foi mantida em 298,15 K. Tanto a estrutura da proteína nativa como as estruturas das proteínas mutadas foram submetidas à mesma rotina de simulação.

Para cada configuração do sistema hAR-LBD-Cofator/R1881, isto é, para a nativa e para cada uma das sete estruturas mutadas, foram feitas 10 simulações. Para cada uma dessas simulações, o seguinte roteiro foi seguido:

1) uma primeira minimização de 1000 passos, seguida de uma dinâmica de 200 ps, nesta etapa todos os átomos do receptor são mantidos fixos, exceto as partes modeladas, o solvente e os íons adicionados;

2) uma segunda minimização também de 1000 passos e com uma nova dinâmica de 200 ps, nesta etapa apenas o *backbone* da proteína é mantido fixo;

3) seguiu-se então para a etapa de produção num total de 8,6 ns para cada uma das simulações com o objetivo de melhorar a amostragem do espaço de fase e, com isso, tornar mais confiável a posterior análise das propriedades estudadas.

As análises de todos os resultados foram feitas utilizando-se programas desenvolvidos pelo nosso próprio grupo de pesquisa de acordo com as propriedades a serem analisadas e as necessidades específicas de cada sistema estudado. Os programas de análise estão disponíveis em http://lm-mdanalysis.googlecode.com.

A parametrização do ligante foi realizada em conjunto com os pesquisadores Paulo Cesar Telles de Souza e Karen Weber (trabalho não publicado). Os parâmetros faltantes no campo de força CHARMM[122] correspondentes a cargas parciais atômicas, comprimentos e ângulos de ligação e ângulos diedros do ligante R1881 foram obtidos empregando cálculos quânticos em nível HF/6-31(d,p) no vácuo, utilizando o programa Gaussian 03.[124] A geometria do ligante cristalizado com o AR-LBD (PDB ID 1XOW)[63] foi usada como estrutura de partida para a otimização da estrutura do ligante com o

objetivo de determinar os parâmetros geométricos de equilíbrio. Na sequência, cálculos das cargas atômicas foram realizados usando o método Merz-Singh-Kollman.[125] Cálculos de frequências vibracionais foram efetuados para obter as constantes de força de estiramento, deformação angular e torção. Os parâmetros de van der Waals foram obtidos dos arquivos de parâmetros e topologias do campo de força CHARMM.

Para testar o conjunto de parâmetros obtidos para o ligante, foi feita uma minimização de energia do ligante no vácuo, com 2000 passos num gradiente conjugado. Em seguida, uma etapa de equilibração de 100 ps foi realizada, seguida por uma simulação de produção de 2 ns nas mesmas condições.

Para a parametrização do ligante R1881 não foi necessário definir tipos atômicos diferentes daqueles existentes no campo de força CHARMM,[122] de modo que os parâmetros de van der Waals puderam ser transferidos diretamente para o nosso sistema. O esquema de numeração do ligante R1881, os tipos atômicos atribuídos a cada átomo e suas respectivas cargas parciais calculadas estão mostrados no anexo II.

4. Resultados e Discussão

Ao se fazer o estudo das bases moleculares da diminuição da capacidade funcional através de mutações pontuais no AR-LBD, o interesse reside nas possíveis diferenças das propriedades estruturais e dinâmicas entre as configurações nativa e mutadas. Variações na posição relativa de estruturas secundárias, nos contatos entre os resíduos e na solvatação de uma dada região são alguns dos fatores que podem dar pistas dos mecanismos que levam à perda da afinidade pelo ligante e da função do receptor.

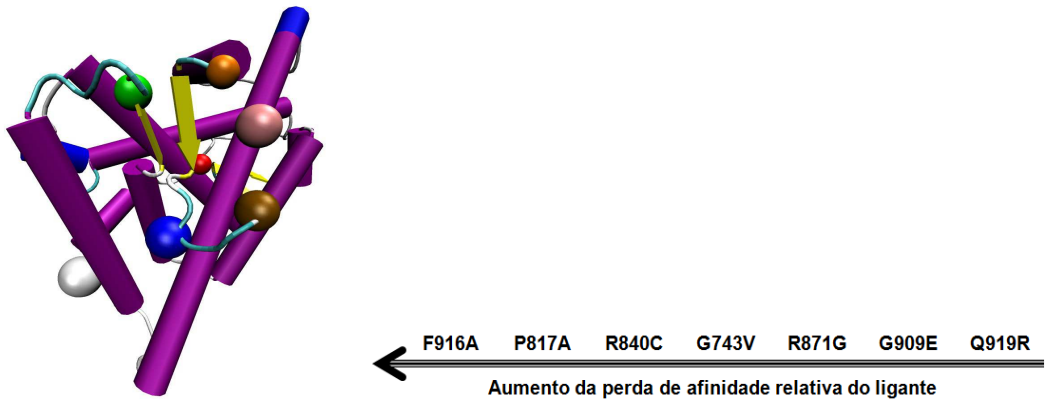
A escolha por se fazer múltiplas simulações relativamente menores de 8 ns ao invés de uma única mais longa equivalente à soma dos tempos das simulações menores deve-se ao fato de que uma amostragem mais efetiva do espaço de fase conformacional das proteínas pode ser obtida de múltiplas trajetórias curtas melhor do que uma única trajetória longa como mostrado em um estudo de Caves e colaboradores.[126] Contudo, esta escolha depende da propriedade a ser estudada ou do fenômeno que se pretende estudar. É obvio que as simulações não podem ser menores do que a escala temporal do fenômeno a ser observado. Além disso, para a tentativa de uma melhor amostragem do espaço de fase, as múltiplas trajetórias devem partir de diferentes condições inicias.

Para o melhor entendimento deste trabalho no anexo III é possível encontrar a nomenclatura, as estruturas dos aminoácidos padrões e os códigos de 3 letras e de 1 letra normalmente usados ao se re referir à resíduos de aminoácidos.

4.1. As Mutações do AR

A Figura 14A mostra a localização no LBD do AR dos resíduos mutados selecionadas para este estudo. Mutações que não afetam diretamente a afinidade do ligante de forma significativa, mas que, por outro lado, estão envolvidas diretamente na atividade de ligação do ligante foram escolhidas para servirem como controle para as análises. Os dados de afinidade reportados na Tabela 1 foram obtidos de fontes diferentes dependendo da mutação estudada. Por este motivo, estes dados foram

normalizados em relação à afinidade do ligante na configuração nativa, tomada como 100%, para cada mutante (Figura 14B). Vale ressaltar que a constante de afinidade ou a constante de dissociação aparente (K_d) determinada experimentalmente deve ser usada com cautela pois seus valores podem variar bastante entre estudos diferentes para o mesmo mutante. Em alguns casos, os erros relativos das medidas são quase da mesma ordem de grandeza do próprio valor estimado para K_d. Um exemplo é a mutação G743V que em três estudos diferentes apresentam três conjuntos diferentes de valores de K_d para os complexos ligante-nativa e ligante-mutante: nativa=0,22/G743V=0,54;[82] nativa=0,65/G743V=1,30;[127] e nativa=0,40 ±0,20/G743V=0,16 ±0,06.[128] Todos os valores expressos em $nmol.L^{-1}$. Mesmo considerando-se a afinidade relativa, os dois primeiros valores apontam para uma perda de afinidade da ordem de 41% e 46 %, respectivamente, enquanto o terceiro valor aponta para um ganho de afinidade de 60%. Por isso, estes valores serão usados mais como uma referência relativa no sentido de apontar quais mutações levam a perdas expressivas de afinidade do ligante e quais se mantêm na mesma ordem de grandeza da afinidade apresentada pela estrutura nativa.



A B

Figura 14: (A) Localização no hAR-LBD das mutações selecionadas para o estudo. As cores das esferas fazem correspondência com as cores do texto na Tabela 1; (B) Ordem crescente do aumento da perda de afinidade relativa do ligantes.

4.2. SISTEMA AR-LBD-COFATOR/R1881

4.2.1 Estabilidade da simulação

A análise de RMSD (root mean square deviation) dá informações sobre variações conformacionais da proteína com relação a uma estrutura de referência que pode ser a estrutura nativa ou uma estrutura média obtida ao longo da simulação:

$$\langle \boldsymbol{RMSD} \rangle(\boldsymbol{t}) = \sqrt{\frac{\sum_{j}^{N}\left[\boldsymbol{r_j}(\boldsymbol{t}) - \boldsymbol{r_j}(\mathbf{0})\right]^2}{\boldsymbol{N}}} \qquad \textbf{(9)}$$

onde N é o número de átomos considerados da proteína inteira ou de cada resíduo; $r_j(t)$ a posição do átomo j no tempo t e $r_j(0)$ sua posição de referência. Geralmente, se está interessado em observar o comportamento estrutural da proteína em relação à estrutura cristalográfica, mais precisamente dos átomos de carbono α do *backbone* da proteína.

Um aspecto importante do cálculo do RMSD é a necessidade do alinhamento da estrutura proteica, de maneira a retirar desvios da posição atômica relativos a translação e rotação da proteína como um todo. Neste trabalho foi usado o alinhamento descrito por Kearsley[129] representado na Figura 15.

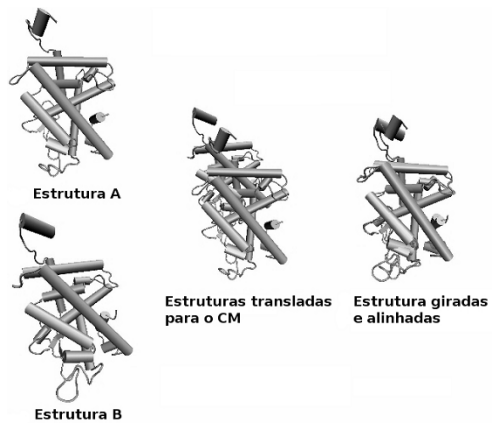


Figura 15: Representação esquemática do alinhamento de corpo rígido, necessário para computar o RMSD.

A Figura 16 mostra as médias das evoluções temporais dos RMSDs das 10 simulações feitas para a estrutura nativa tomando-se como referência a estrutura cristalográfica da proteína nativa sem as alças. Essa análise é usada para verificar a estabilidade e a relaxação da proteína em relação a referência.

A variação de ±0,3 Å em torno da média no perfil do RMSD (Figura 16) mostra a manutenção da estrutura terciária do AR-LBD da estrutura nativa. A ausência de mudanças bruscas na configuração nativa indica a robustez dos parâmetros de simulação escolhidos e a confiabilidade no que diz respeito à rotina de simulação adotada, uma vez que não era esperada nenhuma modificação no sentido de perda de estrutura terciária para este sistema nas condições adotadas para a simulação.

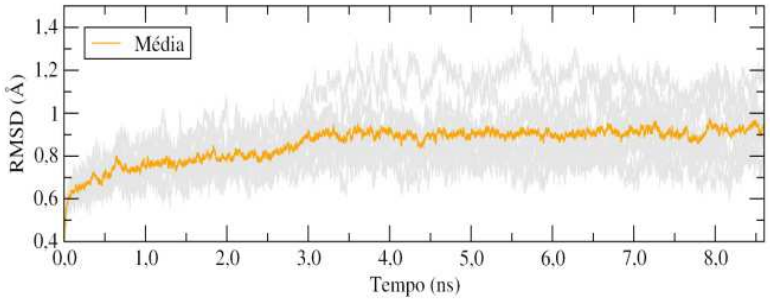


Figura 16: Média da evolução temporal do RMSD dos carbonos do backbone do AR-LBD da estrutura nativa em relação à estrutura cristalográfica da própria nativa.

Toda a discussão em torno das configurações nativa e mutadas será feita usando-se os últimos 8,0 ns de cada simulação, pois, embora não seja possível afirmar que esse sistema esteja equilibrado a partir de 0,6 ns de simulação, a principal mudança observada ocorre ainda no primeiro nanossegundo, que é a abertura do resíduo de arginina na posição 752. As propriedades analisadas associadas a esse fenômeno não dependem da equilibração do sistema e serão discutidos nos itens a seguir.

Comportamento semelhante ao da estrutura nativa com relação à estabilidade e confiabilidade das simulações pode ser observado nos gráficos de RMSD em função do tempo para cada uma das estruturas mutadas (Figura 17).

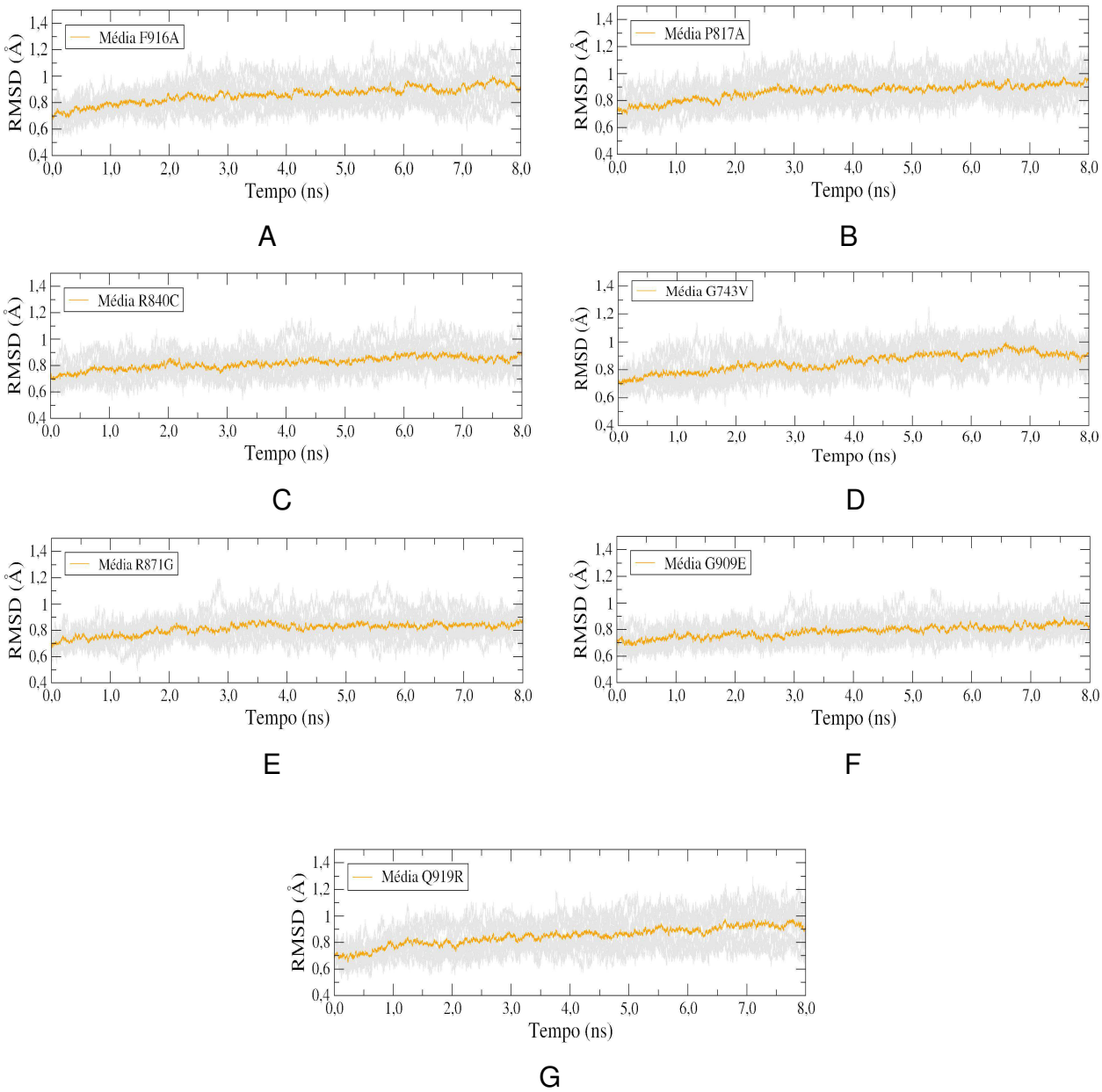


Figura 17: Média da evolução temporal dos RMSD's dos carbonos do backbone do AR-LBD em relação à estrutura cristalográfica: (A) F916A, (B) P817A, (C) R840C, (D) G743V, (E) R871G, (F) G909E e (G) Q919R.

4.2.2. Perfil das Interações dos Ligantes Hidrofílicos do LBP

Outra forma de analisar as variações na estrutura do receptor é estudar o arquivo da trajetória, isto é, o "filme da simulação" buscando diferenças de mobilidade em sua estrutura tridimensional ao longo do tempo de simulação. Justamente durante essas

análises observou-se uma variação brusca na posição do resíduo arginina 752 (R752) perdendo sua interação com o átomo de oxigênio carbonílico do ligante R1881 (Figura 18). Embora os dois resíduos R752 e Q711 localizados nas vizinhanças do ciclo A do ligante esteroide interajam com o átomo O3, é provável que o R752, devido à sua carga positiva, seja mais capaz de estabelecer interações mais fortes com o grupo ceto. Isto está em acordo com o fato de que uma mutação nesta posição R752Q afeta a atividade funcional do AR causando AIS.[130]

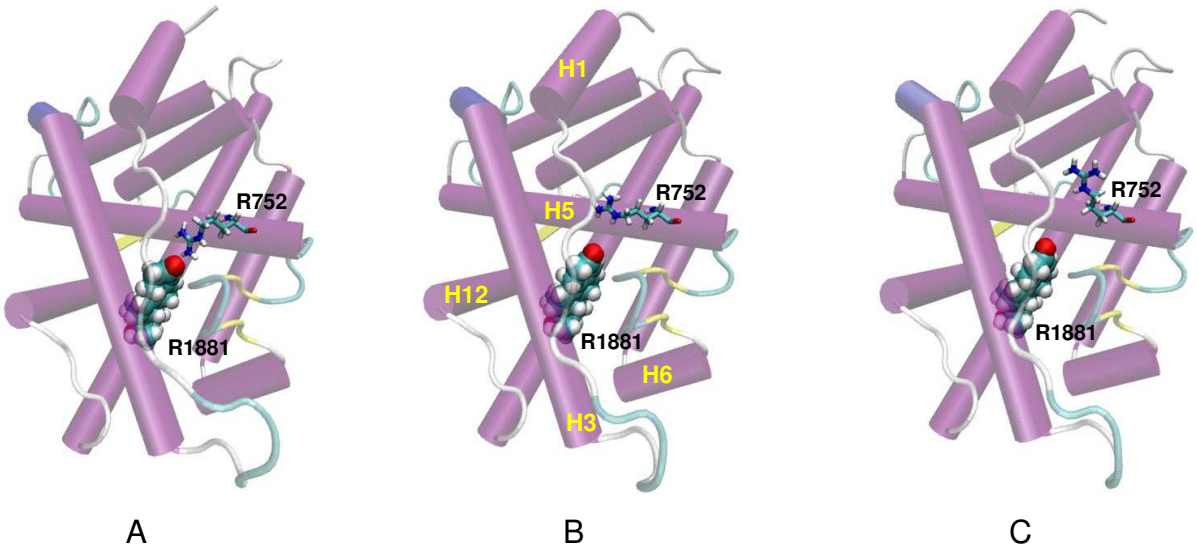


Figura 18: Instantâneos da simulação do F916A mostrando as estruturas do AR-LBD com o ligante R1881 e o resíduo R752: (A) 0,0 ns; (B) 2,6 ns; e (C) 3,1 ns. O mesmo comportamento é observado nas outras estruturas em que ocorre a perda de contato com o ligante. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

A perda de interação R752-R1881 ocorreu com mais frequência nas estruturas mutadas. Das dez simulações realizadas para cada sistema esta alteração ocorreu, por exemplo, uma única vez na estrutura nativa (Figura 19A), em cinco simulações para a mutação F916A (Figura 19B) e em duas para a P817A (Figura 19C), indicando que o fenômeno observado possui relevância no conjunto de simulações realizado. Para as demais mutações, o comportamento do resíduo R752 está indicado na Figura 19. Uma mudança conformacional semelhante envolvendo o resíduo R752 foi observada em um trabalho recente de simulação de Xu e colaboradores[131], entretanto, tal fenômeno não é relacionado ao perfil de solvatação do ligante. Possivelmente, a análise mais detalhada

nas estruturas das simulações onde tal fenômeno ocorre pode lançar luz sobre as principais diferenças capazes de explicar as bases da diminuição da capacidade funcional do AR-LBD quando ocorrem as mutações. Assim sendo, decidiu-se analisar com maior atenção as simulações em que ocorre este segundo modo de ligação nas estruturas mutadas.

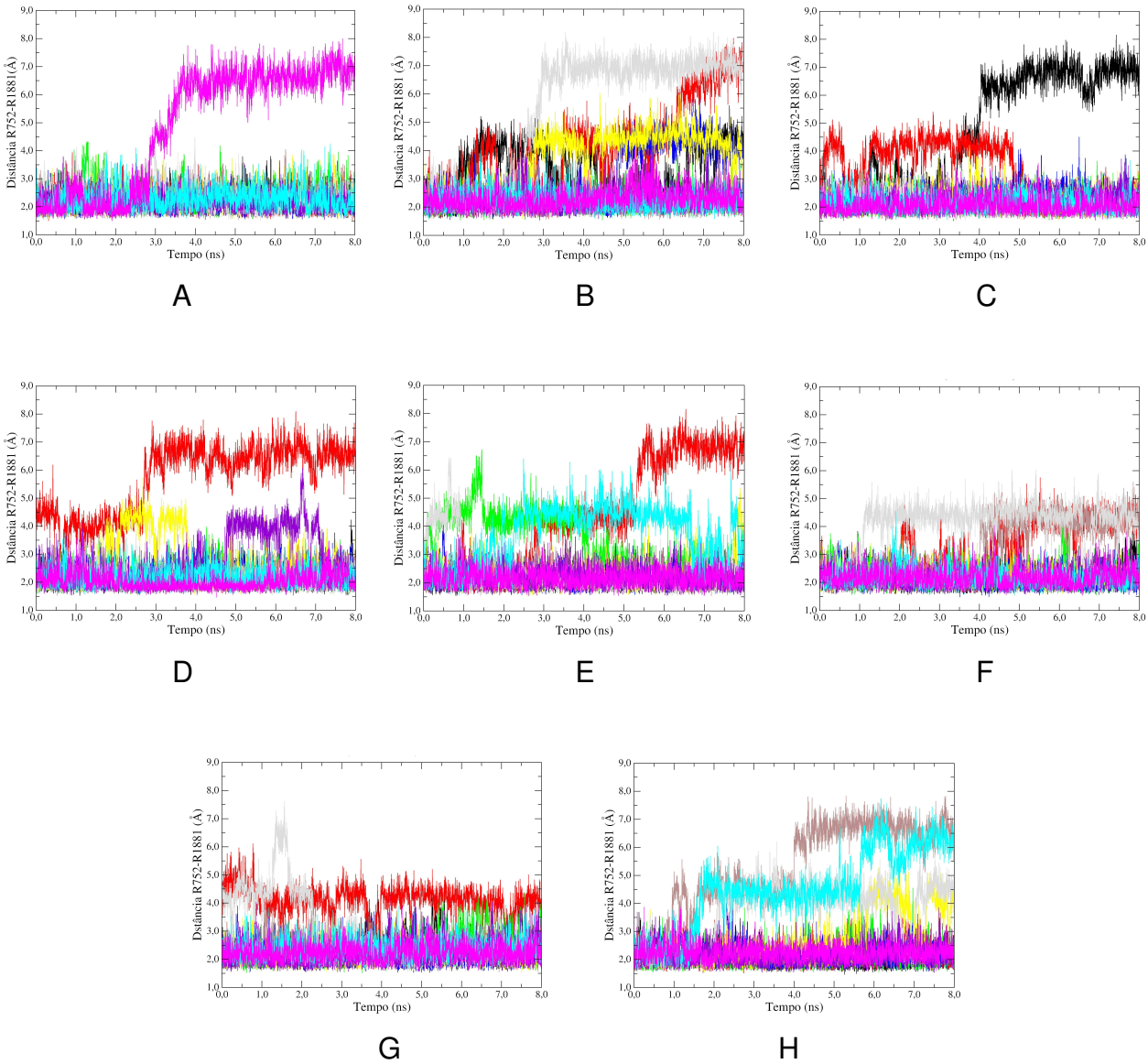
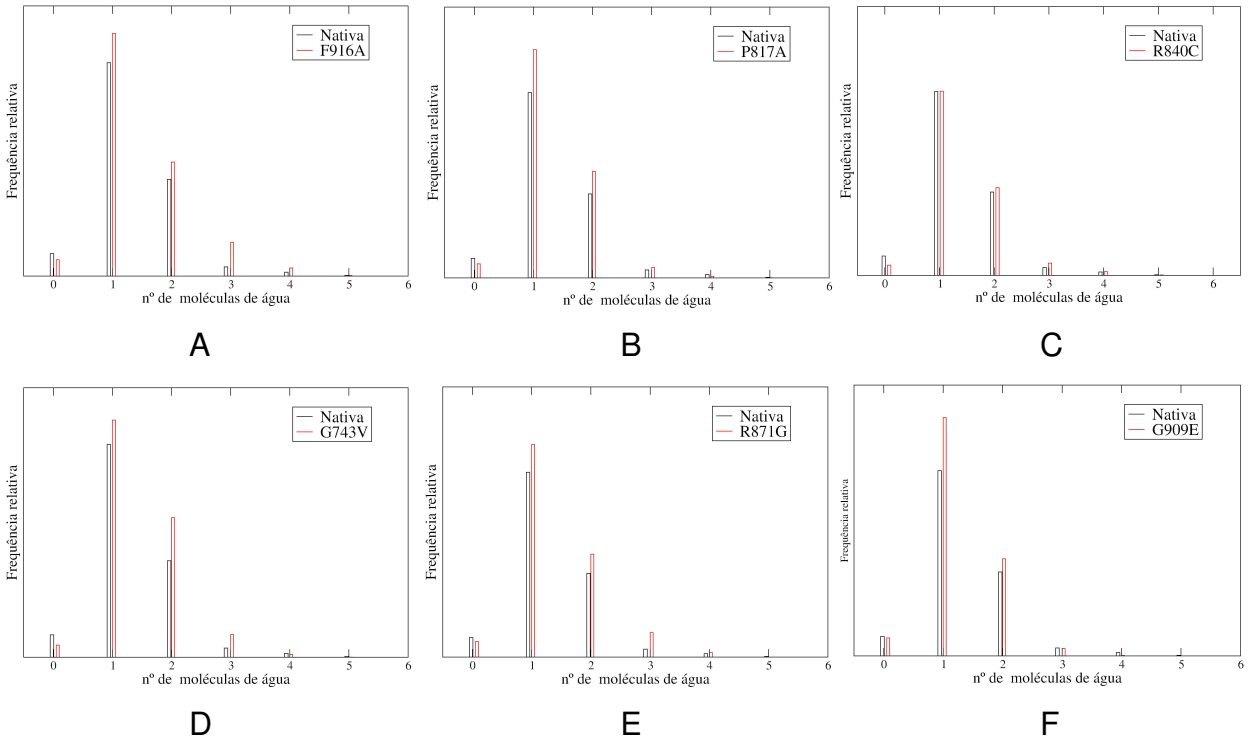


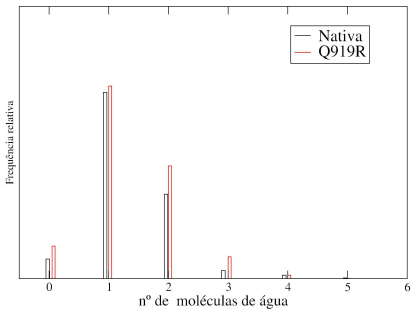
Figura 19: Perfil da variação da distância entre o resíduo R752 e o ligante R1881 em função do tempo de simulação: (A) nativa; (B) F916A; (C) P817A; (D) R840C; (E) G743V; (F) R871G; (G) G909E; (H) Q919R. Em cada gráfico, as curvas mostram as 10 simulações de 8 ns feitas para cada um dos sistemas.

4.2.3. Perfil de Solvatação do LBP

A perda de um importante contato como o do resíduo R752, um dos quatros resíduos que mantém interações hidrofílicas com o ligante, normalmente está associada a mudanças no perfil de solvatação principalmente em torno do ligante. Nos histogramas mostrados na Figura 20 pode-se observar uma maior frequência na entrada de moléculas de água no LBP ao longo da simulação para quase todas as estruturas mutadas, exceto para o R840C que apresenta perfil de solvatação muito próximo ao da estrutura nativa. Contudo, essa discrepância é aparente, pois analisando o arquivo da trajetória vê-se claramente um maior número de moléculas de água entrando no LBP quando da abertura do R752 (Figura 22D), assim como o observado nas trajetórias dos demais mutantes. Porém, na simulação em que ocorre a abertura do R752 está entrada de moléculas de água ocorre com uma frequência mais baixa do que nos demais mutantes, fazendo com que na média seu perfil de solvatação seja muito semelhante ao da nativa.



…→



G

Figura 20: Distribuição relativa média das moléculas de água posicionadas até 4 Å do ligante R1881 no AR-LBD das estruturas mutadas em relação à estrutura nativa: (A) F916A; (B) P817A; (C) R840C; (D) G743V; (E) R871G; (F) G909E; (G) Q919R.

Os dados mostrados nos histogramas da Figura 20 confirmam de forma quantitativa o comportamento observado ao longo da simulação. A perda da ligação de hidrogênio entre o átomo de oxigênio O3 do R1881 e o hidrogênio do grupo guanidina do resíduo R752 passa a ser compensada pelas novas ligações de hidrogênio com as moléculas de água que entram em maior quantidade nesta região (Figura 21). As moléculas de água entram no LBP através de uma conhecida região acessível ao solvente entre as hélices 3 e 5 e posicionam-se entre o grupo 3-ceto do ligante e o grupo guanidina do R752 e o grupo amida do Q711, agindo como mediador de ligação de hidrogênio entre estes grupos.[132]

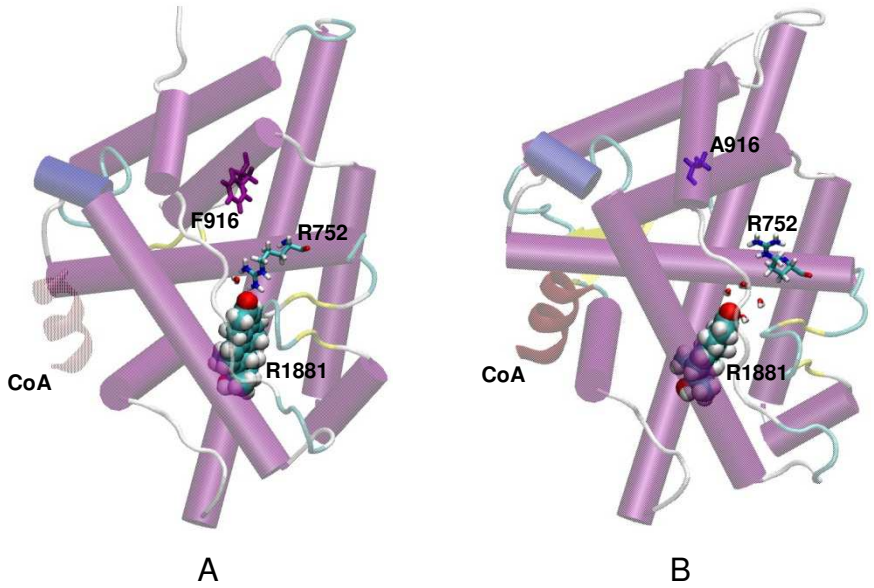
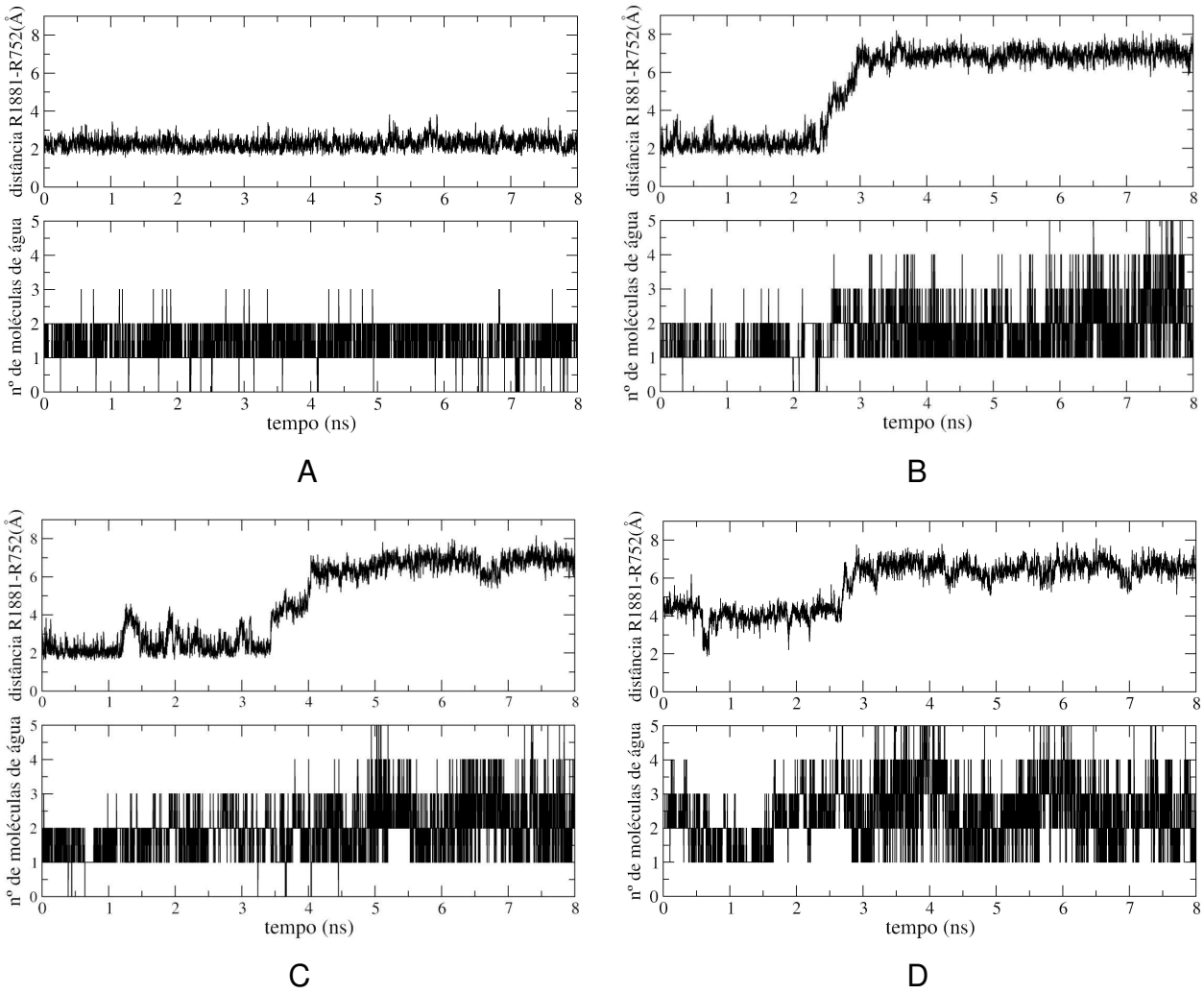


Figura 21: Instantâneo da simulação mostrando estruturas representativas do AR-LBD: (A) Na estrutura nativa, a molécula de R1881 interage diretamente com resíduo R752 e com uma molécula de água apenas; (B) Na variante F916A, o ligante R1881 perde sua interação com R752, mas passa a interagir com maior frequência com um número maior de moléculas de água. Comportamento semelhante ocorre nas demais mutações em que há aumento da frequência de entrada de moléculas de água devido à perda da interação R752-R1881.

A entrada de moléculas de água desestabiliza a região do LBP, uma vez que este é essencialmente hidrofóbico (Figura 5B). Esta é uma evidência que, de uma forma geral, corrobora para as perdas de afinidade observadas experimentalmente e, uma vez que a região de entrada das moléculas de água é exatamente onde se localiza o resíduo R752, esses dados se mostram bastante coerentes. Contudo, o perfil demonstrado não permite estabelecer uma relação unívoca entre a extensão da perda de afinidade do ligante e o aumento da frequência de solvatação no LBP das estruturas mutadas, pois os valores extremos de perda de afinidade não correspondem aos perfis extremos de solvatação observados nas simulações. Por outro lado, a maior frequência de entrada de moléculas de água no LBP das estruturas mutadas acontece de maneira quase que concertada com a mudança brusca de posição do resíduo R752 (Figura 22B a 22H) não permitindo estabelecer uma relação de causa e consequência entre estes dois eventos. Como contraponto, o perfil de entrada de moléculas de água em uma simulação em que não ocorre a abertura da R752 é mostrado para a estrutura nativa (Figura 22A). Observa-se que não ocorre acréscimo do número de moléculas de água

em torno do ligante ao mesmo tempo que a interação R752-R1881 se mantém ao longo da simulação. Tal comportamento demonstra o papel das moléculas de água na estabilização das interações hidrofílicas do grupo ceto do ligante R1881 levando, contudo, à já citada desestabilização do bolsão hidrofóbico do ligante. Os perfis mostrados na Figura 22 se repetem para os respectivos casos de abertura e de não abertura do R752 observado nas demais simulações.



…→

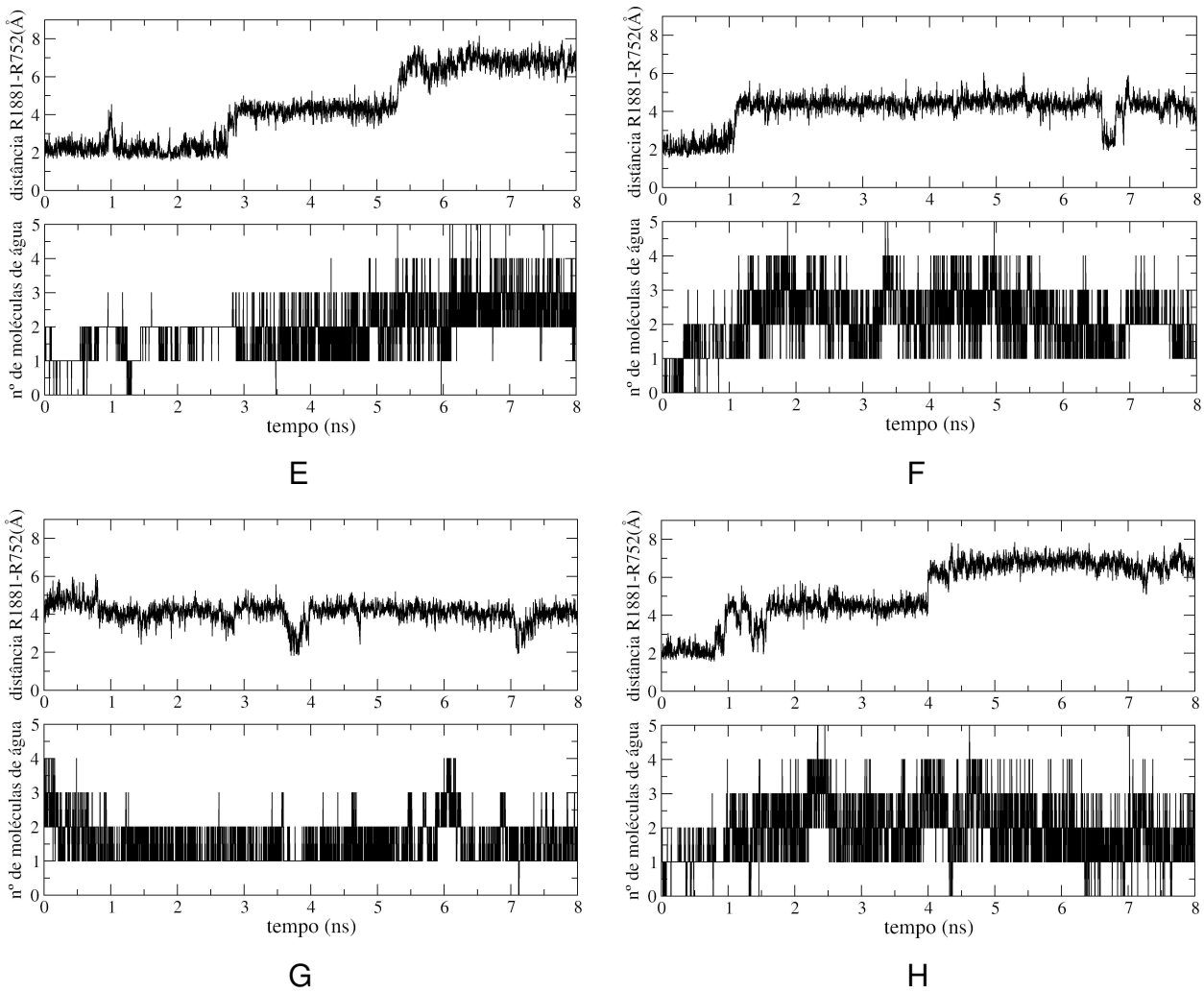


Figura 22: Relação entre a entrada de moléculas de água no AR-LBP e a perda de interação entre o ligante R1881 e o resíduo R752. (A) nativa; (B) F916A; (C) P817A; (D) R840C; (E) G743V; (F) R871G; (G) G909E; (H) Q919R.

4.2.4. Perfil dos Contatos das Regiões Chaves do AR-LBD

Buscando uma visão estrutural mais geral de como as mudanças observadas no LBP se relacionam com as mutações estudadas e como estas, por sua vez, alteram localmente a mobilidade dos resíduos envolvidos em cada um dos processos, decidiu-se investigar mais de perto as regiões chaves para a função do receptor: em torno do ligante R1881, em torno do resíduo mutado, na região do domínio F e na superfície do AF2.

4.2.4.1 Perfil da Dinâmica do Ligante R1881

Foi contabilizado o número de contatos ao longo da simulação entre o ligante R1881 e os resíduos localizados até 4 Å dele. Em todas as estruturas mutadas o único resíduo que apresentou mudança significativa na porcentagem de contatos foi o R752, como já havia sido observado na análise da trajetória da simulação e quantificado na variação da distância R752-R1881 (Figura 19). Dessa forma, baseando-se nas evidências das simulações, a perda de afinidade do ligante, no que concerne diretamente ao LBP, estaria associada apenas ao aumento da frequência de entrada de moléculas de água na região entre o loop H1/H3 e fitas S1/S2 e à perda de contato com o resíduo R752.

4.2.4.2 Perfil da Dinâmica dos Resíduos Mutados

Neste item serão mostradas as porcentagens de tempo ao longo das simulações em que cada resíduo estudado fica em contato com os demais resíduos distantes até 4 Å dele. Os valores são mostrados como porcentagem relativa ao tempo total no qual dois resíduos quaisquer podem permanecer em contato durante toda a simulação. Estão listados os resíduos a n ±3 do resíduo estudado e apenas aqueles que ficam em contato superior a 10% do tempo total de contatos na média das 10 simulações.

F916A

Uma condição importante para proporcionar a conformação adequada para a ligação do ligante é a hidrofobicidade principalmente envolvendo os resíduos L907, V911 e I914. Estudos mostraram que mutações na posição 907 apresentam um efeito drástico na ligação do ligante quando este é substituído por outros resíduos de menor hidrofobicidade e/ou menor volume estéreo.[68,133] Além disso, embora não conservadas, as posições correspondentes V911 e I914 em outros receptores esteroides são todas ocupadas por resíduos hidrofóbicos.[68]

Neste sentido, o aspecto mais crítico é a presença da aromática fenilalanina na posição 916. Dados experimentais mostram que sob proteólise, o mutante F916A

degrada-se,[68] indicando que a compactação devida do LBD não pôde ocorrer após a ligação do agonista. O que demonstra o envolvimento do resíduo F916 no processo de promover a conformação ativa do receptor.

Dois resíduos próximos ao resíduo F916 são conhecidamente importantes para a manutenção da estabilidade da extensão C-terminal, o R831 e o D864[68] (Figura 23). Estes resíduos, assim como o próprio F916, também são conservados através da família dos receptores esteroides. Além disso, algumas mutações no R831 e no D864 estão associadas à CAIS, demonstrando, portanto, sua importância para a função do receptor.

Neste sentido, enquanto que a porcentagem do tempo de contatos entre os resíduos 916 e D864 se mantêm praticamente inalterada ao longo dos 8,0 ns de simulação na mutante em relação à nativa, a porcentagem de tempo de contato com o resíduo R831 apresenta uma queda significativa: 85% na nativa e 20% na mutante (Tabela 2), devido à perda da interação amino-aromática. Além deste resíduo, ocorre diminuições bastante significativas de tempo de contato entre outros resíduos localizados na hélice 8 (L810), alça H8/S3 (S814) e hélice 9 (I835). É possível que como decorrência dessa perda de interação a extensão C-terminal, onde se encontra o resíduo 916, diminua sua compactação com a parte interna do sanduíche de α-hélices comprometendo a conformação ativa do receptor como proposto por Tahiri.[68]

Tabela 2: Porcentagem de tempo de contato entre o resíduo 916 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	F916A
H8	L810*	92	0
H8/S3	S814	15	0
H9	R831*	85	20
H9	I835*	99	0
H10	F856*	66	0
H10	T860	100	97
H10	K861	3	12
H10	L863*	91	0
H10	D864*	98	95
Domínio F	Q919	10	9

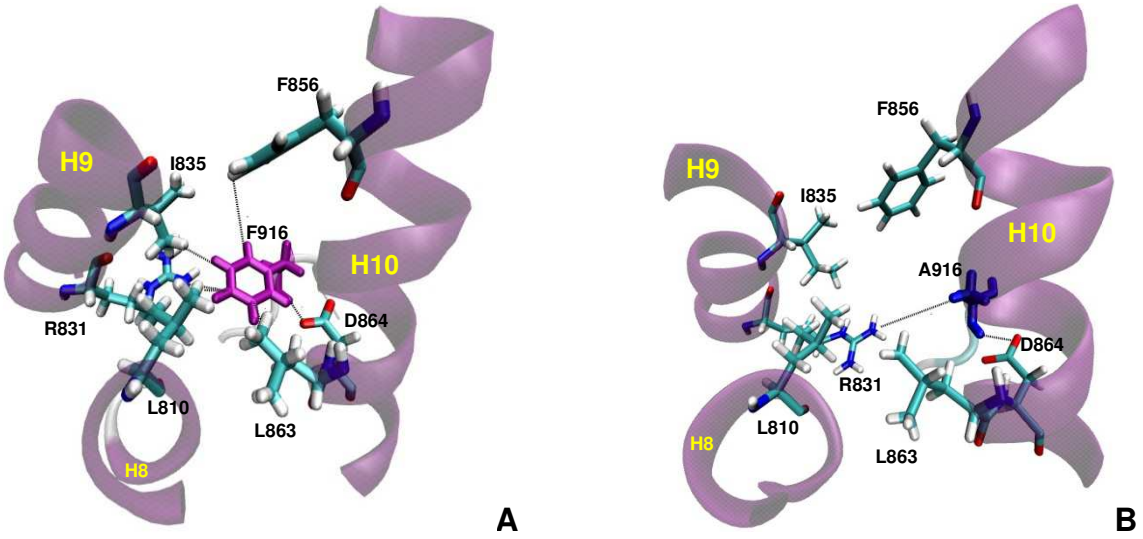


Figura 23: Representação da região em torno da posição 916. (A) estrutura nativa: F916; (B) estrutura mutada: A916. Estão incluídos os principais resíduos que fazem ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 2 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

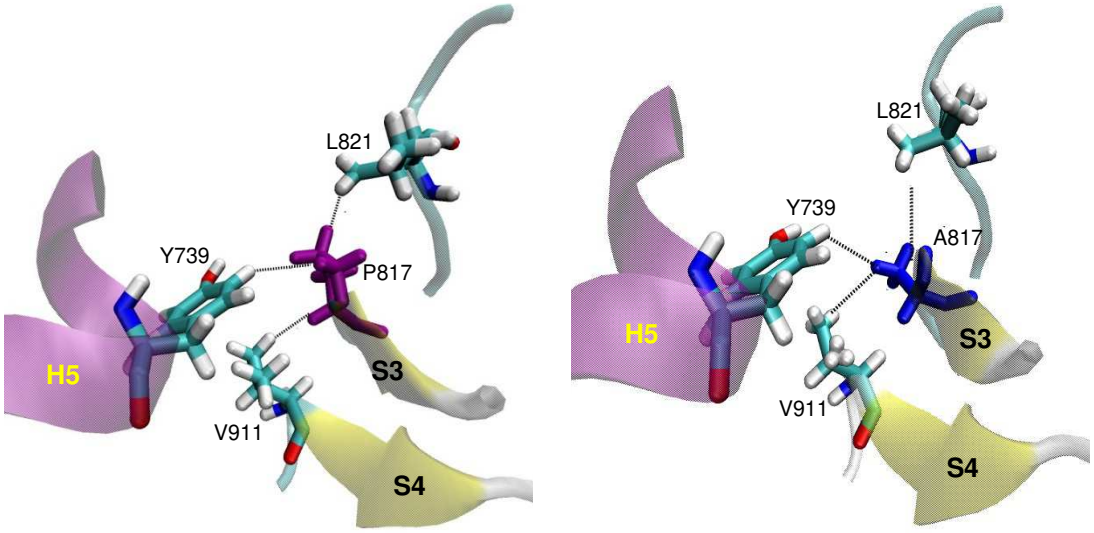
P817A

A observação de que o resíduo P817 é conservado em todos os receptores esteroides, exceto o ER, sugere um papel estrutural específico para este resíduo na subfamília. Embora não esteja localizado no domínio F, ao contrário da mutação F916A, substituições na prolina 817 também impactam consideravelmente na atividade e afinidade do hAR (Tabela 1). Experimentos *in vitro* confirmaram que a substituição do resíduo P817 altera a função do AR e análises estruturais sugerem que tal alteração é devida à desestabilização da extensão C-terminal,[6] uma vez que este resíduo faz interações diretas com os resíduos do domínio F. Assim sendo, o estudo da evolução dos contatos envolvendo o resíduo 817 também pode lançar luz sobre o papel do domínio F no AR-LBD.

Dos três resíduos localizados na hélice 5 que fazem contato direto com o resíduo 817 (A735, V736 e Y739) apenas a tirosina 739 está em contato ao longo de praticamente toda a simulação na estrutura nativa (Tabela 3). Na estrutura mutada, o tempo de contato desta interação apresenta uma queda significativa deixando a interação da fita S3 com a hélice 5 menos efetiva (Figura 24). Com relação às interações do resíduo 817 com os resíduos do domínio F, mais especificamente os da fita S4, estas se mantêm inalteradas a despeito da mutação operada. Portanto, segundo as simulações, a liberação das restrições rotacionais impostas pelo resíduo de prolina, ao contrário do proposto por Lumbroso e colaboradores,[6] não leva ao rompimento da fita β S3 e nem da interação estabilizante com a fita β S4. O que, ainda segundo Lumbroso e colaboradores, poderia explicar a redução da estabilidade do AR-LBD advinda da mutação P817A. Segundo os resultados de nossas simulações, tal redução da estabilidade observada quando desta mutação pode estar associada à mudanças no perfil dos contatos do resíduos do domínio F. Esta discussão será feita no próximo item.

Tabela 3: Porcentagem de tempo de contato entre o resíduo 817 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	P817A
H5	A735	15	6
H5	V736	13	2
H5	Y739*	99	76
S3/H9	G820	99	93
S3/H9	L821*	100	68
Domínio F	K910	99	96
Domínio F	V911*	100	100
Domínio F	K912	99	97



A B

Figura 24: Representação da região em torno do resíduo A817. (A) estrutura nativa: P817; (B) estrutura mutada: A817. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 11 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

R840C

Os contatos mantidos pelo resíduo 840 e seus vizinhos permanecem praticamente inalterados (Tabela 4) e nada no que diz respeito a estas interações fornece evidências para a perda de afinidade do ligante quando desta mutação.

Tabela 4: Porcentagem de tempo de contato entre o resíduo 840 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	R840C
Coil-H1	C669	36	19
H9	K836	100	100
H9	E837	100	100
H9	A843	93	100
Loop H9/H10	C844	82	67

G743V

As simulações mostraram que dentre as interações reconhecidamente mais importantes para o resíduo 743,[128] duas delas sofrem aumentos significativos em sua porcentagem de tempo de contato: as com o resíduos I815, localizado na fita β S3, e, principalmente, com o resíduo Q867, localizado na hélice 10 (Tabela 5). Essas mudanças podem contribuir para a desestabilização da hélice 5 através da perturbação de interações chaves para o empacotamento envolvendo as hélices 10 e 11 (Figura 25), como proposto por Poujol e colaboradores em um trabalho de biologia molecular de 2002[127] e por He e colaboradores em um trabalho de cristalografia de 2006.[128]

Tabela 5: Porcentagem de tempo de contato entre o resíduo 743 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	G743V
H5	Y739	2	16
H5	S740	100	100
H5	V746	100	100
H5	F747	100	100
H8	L811	100	100
S3	I815*	0	25
H10	V866	100	100
H10	Q867*	9	100
H10	A870	100	100
H10	H874	15	11

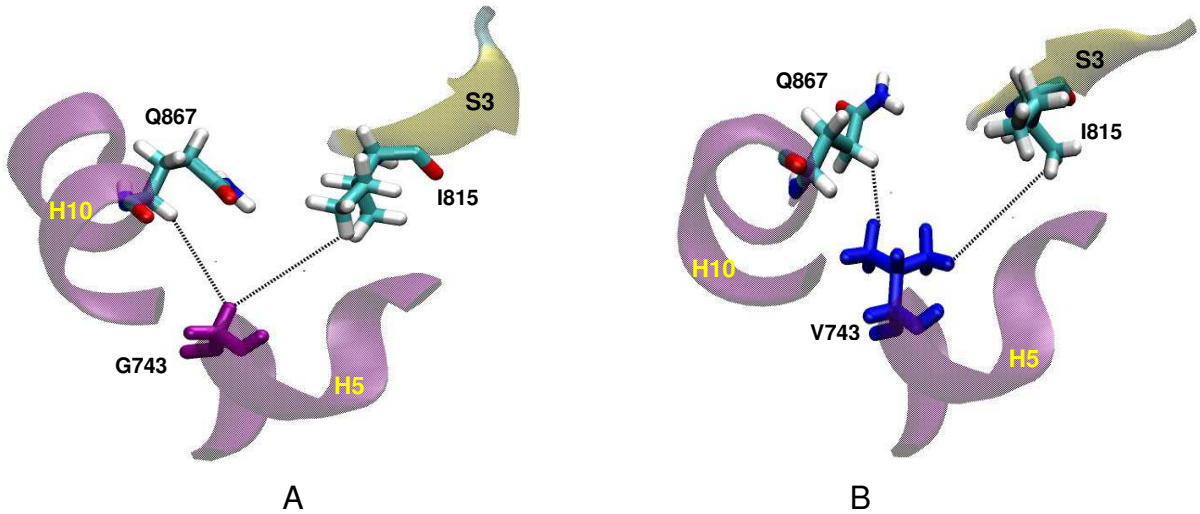


Figura 25: Representação da região em torno da posição 743. (A) estrutura nativa: G743; (B) estrutura mutada: V743. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 5 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

R871G

O resíduo R871 está na extremidade final da hélice 10 e supõem-se que devido à sua posição, mutações nesse resíduo possam interferir com a capacidade da hélice 12 mover-se apropriadamente afetando a cinética dos ligantes de AR.[83] De fato, as simulações mostram que ocorre uma diminuição grande no contato desse resíduo com o resíduo I906 situado na hélice 12 e uma diminuição menor mas também significativa em relação ao resíduo 907 (Tabela 6). Com isso a interação da hélice 12 com a região C-terminal da hélice 10 fica comprometida. Contudo, afirmar que tal perda de interação leva à perda de capacidade da hélice 12 mover-se dependeria de estudos com maiores tempos de simulação e ainda assim talvez não seja possível observar tal movimento nos tempos de simulação disponíveis. De qualquer forma, como nos demais casos, essas perdas de interações são um importante indicativo no que diz respeito às contribuições sutis para a manutenção da estrutura da proteína, pois, como discutido para a mutação F916A, a manutenção das interações hidrofóbicas em torno da região da extensão C-terminal são fundamentais para promover a conformação ativa do receptor.

Tabela 6: Porcentagem de tempo de contato entre o resíduo 871 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	R871G
H10	Q867	100	100
H10	P868	100	100
H10	H874	100	100
H10	Q875	100	100
H12	I906*	79	10
H12	L907*	100	83

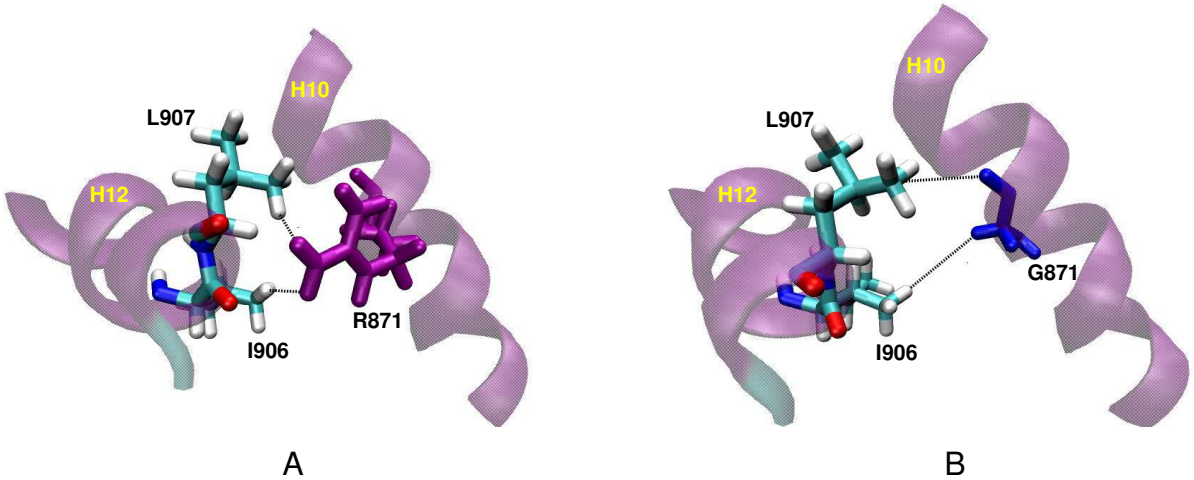


Figura 26: Representação da região em torno da posição 871. (A) estrutura nativa: R871; (B) estrutura mutada: C871. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 6 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

G909E

Embora esta mutação seja uma das que têm o menor impacto no valor de constante de afinidade do R1881 dentre as estudadas neste trabalho, sabe-se que o papel do resíduo de glicina na posição 909 é o de promover a correta orientação da extensão C-terminal, conforme indicam resultados experimentais de biologia molecular obtidos por Tahiri e colaboradores.[68] Observou-se que, devido à flexibilidade conformacional característica do resíduo de glicina, qualquer mutação nesta posição modifica as propriedades do receptor. O que confirma a importância deste resíduo para a orientação ativa correta da extensão C-terminal.

Para essa mutação, as simulações mostraram um aumento significativo do tempo de interação com o resíduo R871 da hélice 10 e também com o resíduo I906 localizado no fim da hélice 12 (Tabela 7). Essas interações, juntamente com a da V818 localizada no fim da fita S3, parecem deixar a região em torno da mutação com menor liberdade conformacional quando comparada à estrutura nativa, por conta de uma tripla “ancoragem” (Figura 27). Esta menor flexibilidade mostrada pelas simulações corroboram para a hipótese defendida por Tahiri e colaboradores.[68]

Tabela 7: Porcentagem de tempo de contato entre o resíduo 909 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	G909E
S3/H9	V818*	90	90
H10	R871*	7	51
H12	K905	99	99
H12	I906*	84	100
Domínio F	K912	63	86

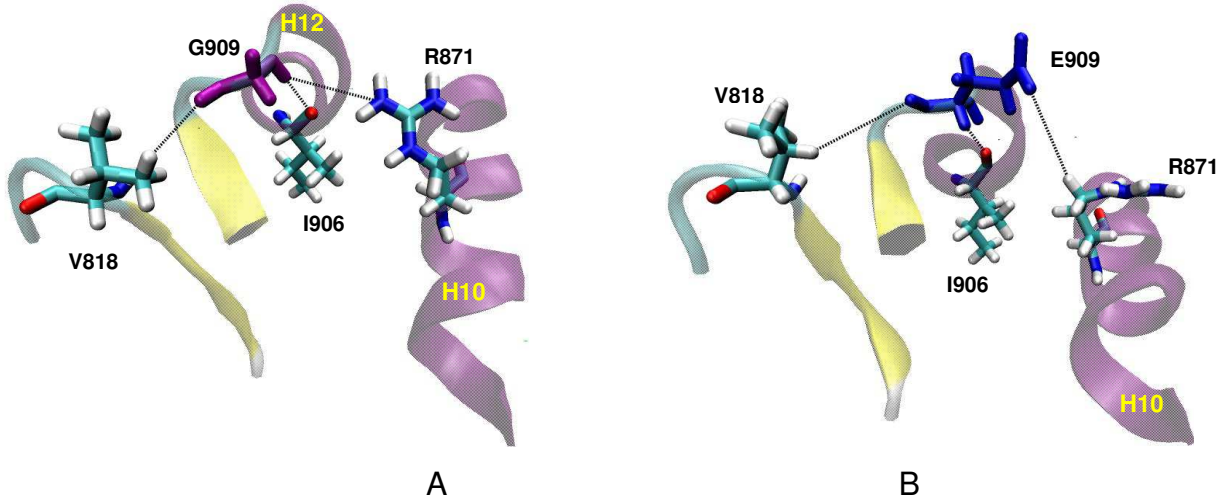


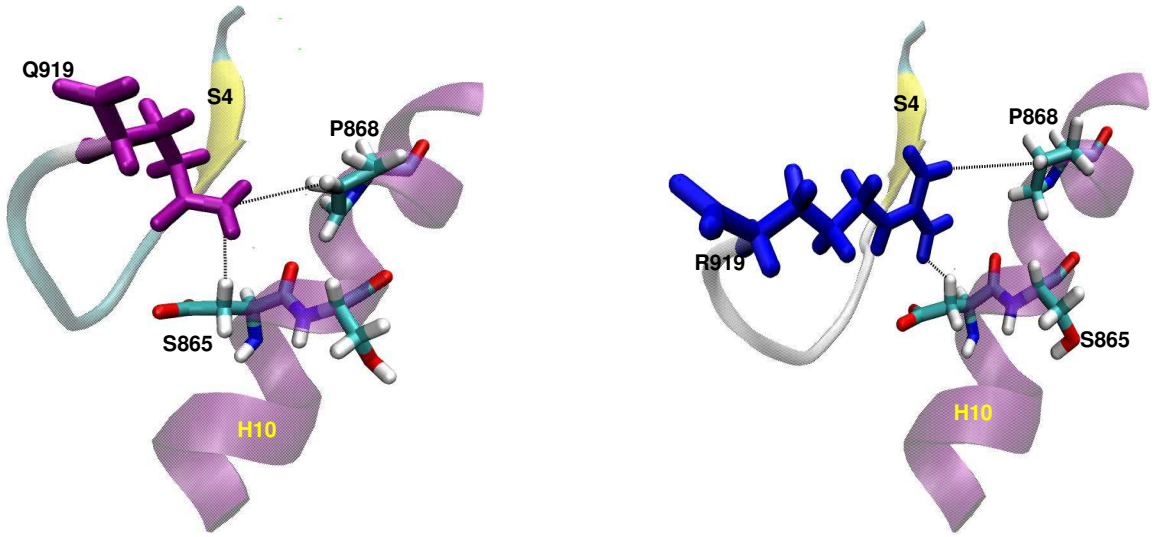
Figura 27: Representação da região em torno da posição 909. (A) estrutura nativa: G909; (B) estrutura mutada: E909. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 7 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

Q919R

Nos outros receptores esteroides a posição 919 é ocupada por um resíduo básico de lisina. Provavelmente por isso a substituição do resíduo de glutamina por um outro resíduo básico de arginina não tenha grande influência na afinidade do ligante. Por outro lado, esta substituição apresenta uma alteração na função da proteína associada ao câncer de próstata. Sendo um resíduo pertencente ao domínio F, o Q919 torna-se importante para a manutenção da orientação correta da extensão C-terminal e consequentemente do papel funcional do AR. A tabela 8 mostra que há um aumento nas interações com resíduos da hélice 10, mais especificamente os resíduos S865 e P868 (Figura 28), além de um pequeno aumento das interações com o já citado resíduo D864. Este aumento de interações com a H10 pode levar a um posicionamento não adequado para a extensão C-terminal no que diz respeito à orientação mais efetiva para a ativação do receptor.

Tabela 8: Porcentagem de tempo de contato entre o resíduo 919 e resíduos até 4 Å dele para as estruturas nativa e mutada.

Localização	Resíduo	% de contato	
		Nativa	Q919R
H10	K861	26	20
H10	D864	35	41
H10	S865*	1	20
H10	P868*	7	44
Domínio F	Y915	67	72



A B

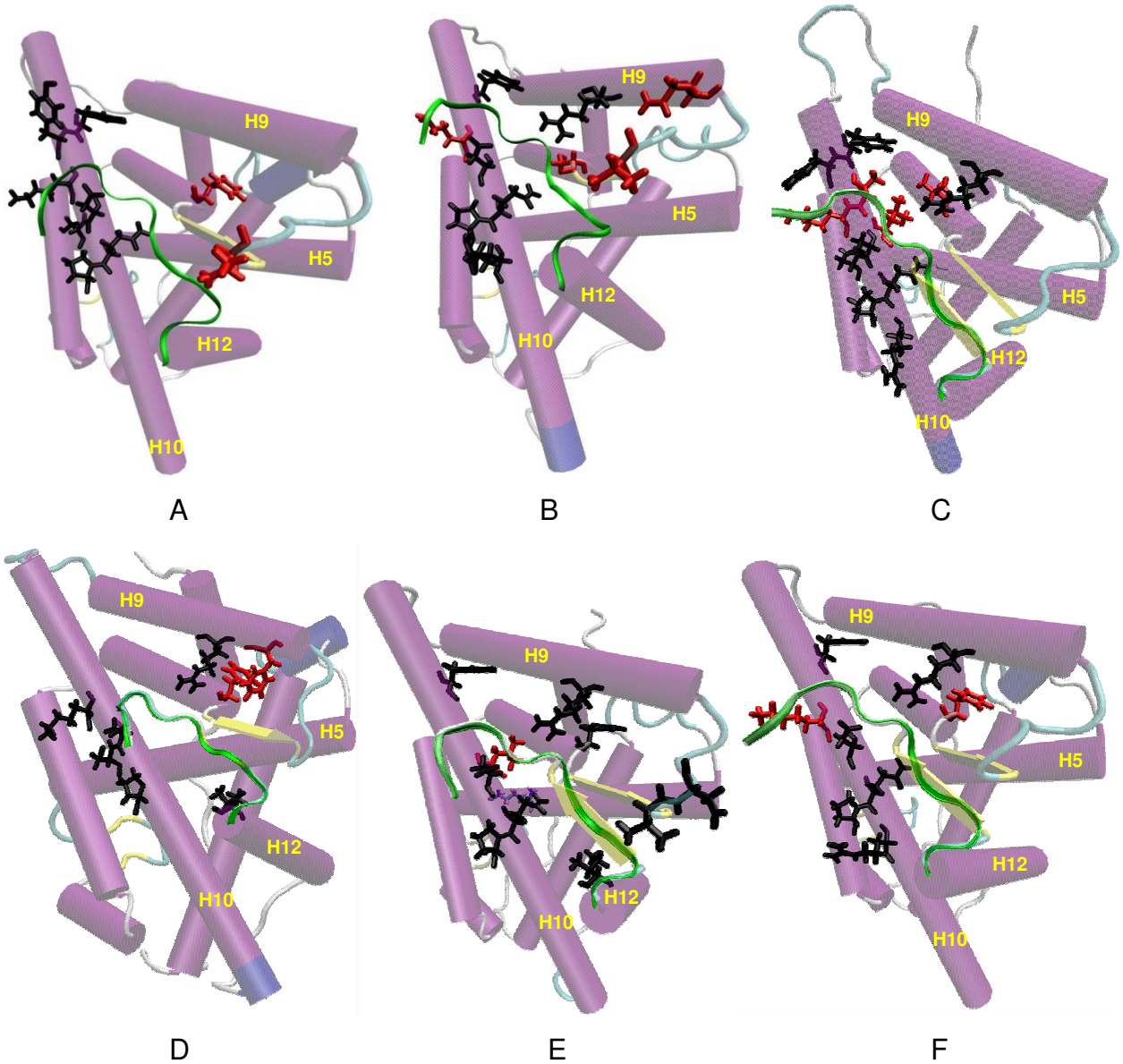
Figura 28: Representação da região em torno da posição 919. (A) estrutura nativa: Q919; (B) estrutura mutada: R919. Estão incluídos os principais resíduos que fazem ou ligação de hidrogênio ou contatos de van der Waals, indicados na Tabela 8 com o asterisco. Os átomos estão coloridos segundo o tipo (carbono: azul; hidrogênio: branco; oxigênio: vermelho; nitrogênio: azul).

4.2.4.3 Perfil da Dinâmica do Domínio F

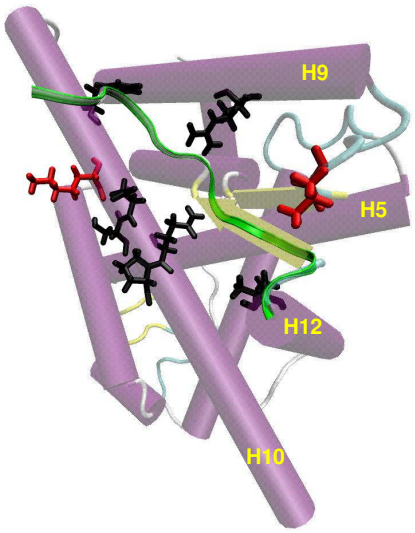
Por conta do conhecido papel da extensão C-terminal, característica dos receptores esteroides, e da importância da necessidade de seu posicionamento correto para o desempenho funcional do AR, decidiu-se monitorar também a evolução dos contatos dos outros resíduos que formam o domínio F, além daqueles já envolvidos nas mutações F916A, G909E e Q919R, a fim de se estudar eventuais mudanças relevantes nas interações para esta região. A Figura 29 mostra uma visão geral para cada estrutura mutada dos contatos que foram perdidos e ganhos.

De uma forma geral, pode-se observar que existe uma clara tendência dos resíduos do domínio F diminuírem o tempo em que fazem interações com a região do AR-LBD onde se localizam as hélices 8 e 9 e a fita S3, ao passo que aumentam as interações com a região da hélice 10, principalmente com os dois resíduos reconhecidamente importantes para o papel funcional do AR, os já citados R831 e D864.[68] Desta forma, as evidências sugerem que os resíduos do domínio F passam a interagir mais com a fatia externa do sanduíche de alfa-hélices em detrimento das

interações com a fatia interna (Figura 29). Observa-se ainda que, embora a fita β S4 seja preservada nas estruturas mutadas, há mudanças estruturais no domínio F como um todo. Assim, supomos que o balanço entre pequenas perdas de contatos da extensão C-terminal com um lado da proteína e pequenos ganhos de contatos com o outro pode levar ao posicionamento incorreto, ou menos eficiente, desta região do AR-LBD fundamental na ativação transcricional. As tabelas com os valores dos percentuais de tempo de contato dos resíduos do domínio F para cada mutante estão no anexo I.



…→



G

Figura 29: Representação das estruturas médias do AR-LBD mostrando os resíduos que diminuem (em vermelho) e os que aumentam (em preto) o número de contatos com o domínio F (em verde) ao longo da simulação para cada mutação em relação à estrutura nativa: (A) F916A; (B) P817A; (C) R840C; (D) G743V; (E)R871G; (F) G909E; (G) Q919R. Apenas diferenças de porcentagem maiores que 5% são apresentadas.

Uma observação complementar a cerca dos contatos envolvendo os resíduos do domínio F diz respeito a proposição feita por Lumbroso e colaboradores de que por conta da mutação P817A ocorreria um rompimento da fita β S3 e da interação estabilizante com a fita β S4 e que isso poderia explicar a redução da estabilidade do AR-LBD.[6] Tal proposição não foi sustentada pelos resultados observados nas simulações (item 4.2.4.2). Contudo, a desestabilização do domínio F decorrente da mutação pode estar associada à mudança no perfil das interações de outros resíduos do domínio F que não aqueles interagindo diretamente com o resíduo 817. De forma parecida como que ocorre na F916A, os resíduos finais do domínio F diminuem o tempo em que fazem interações com alguns resíduos da região da fita β S3 e das hélices 8 e 9 quando da mutação P817A (Tabela 11 – Anexo I) e aumentam o tempo com que fazem interações com alguns resíduos da hélice 10 (Tabela 12 – Anexo I), embora de forma mais tímida do que no caso da F916A.

4.2.4.4. Perfil em torno do AF2

Mesmo tendo partido da estrutura do AR-LBD já com o peptídeo representando o cofator ligado, e, portanto, com a região do AF2 já estabilizada, a análise desta região poderia mostrar alguma mudança na direção de perdas de ancoramento do peptídeo ligado ou na mobilidade do próprio peptídeo.

Contudo, análises de variação de distância entre os pares de resíduos do grampo de carga (K720/E897, K717/E893 e R726/E709) não mostraram nenhuma mudança significativa. Análises de mobilidade do peptídeo, bem como de contatos entre seus resíduos e os resíduos do AF2 também não apresentaram mudanças.

Duas são as possibilidades para este comportamento, ou uma vez o cofator ligado, a região do AF2 não sofre grandes variações por conta das mutações, ou eventuais mudanças na mobilidade do cofator ocorrem em um tempo de simulação maior. Na verdade, como a mutação ocorre no gene, uma eventual mudança na posição dos resíduos do AF2, incluído os do grampo de carga, ocorreria durante o enovelamento e afetariam o próprio recrutamento do cofator. Porém, não foi possível realizar simulações desta ordem temporal.

5. Conclusões

As análises estruturais do AR-LBD como um todo, mostraram a manutenção da estrutura terciária da proteína para todos os sistemas simulados. De fato, devido às condições e ao tempo das simulações não esperava-se que alterações drásticas ocorressem na estrutura da proteína. Portanto, as simulações são coerentes com o comportamento das proteínas em condições experimentais semelhantes.

As simulações corroboraram para algumas hipóteses da literatura que apontam para a importância do papel do domínio F na estabilização de uma conformação ativa no AR-LDB. Ao que parece, a manutenção de alguns contatos hidrofóbicos são de extrema importância para a estabilidade da conformação ativa do receptor. Como aqueles envolvendo o resíduo F916. Por outro lado, embora não leve ao rompimento da fita β S3 e nem de sua interação com a fita β S4, como supõe Lumbroso e colaboradores, a mutação P817A perturba as interações com a fatia interna do sanduíche de α-hélices, alterando, como consequência, as interações do domínio F.

As demais mutações de uma maneira geral também acabam por aumentar as interações do domínio F com a hélice 10 (fatia externa de α-hélices) e, portanto, mudam o posicionamento do que se acredita ser a conformação ideal do domínio F que contribui para o papel funcional do AR.

As simulações sugerem que a perda de afinidade do ligante, no que concerne ao LBP, está associada apenas ao aumento da frequência de entrada de moléculas de água na região do loop H1/H3 e fitas S1/S2 e à perda de contato com o resíduo R752. Contudo, não foi possível estabelecer como as perturbações locais na estrutura da proteína decorrentes das mutações se propagam até o sítio de ligação a ponto de serem diretamente relacionadas às mudanças no perfil de solvatação da região do LBP e nas interações hidrofílicas do ligante R1881. Ficou evidente que mesmo sem provocar alterações no posicionamento do ligante estas perturbações contribuem para o rearranjo das interações do AR-LBD, alterando a compactação de algumas regiões e é possível que estas mudanças contribuam de forma expressiva para a diminuição da capacidade funcional do AR.

O que mais chama a atenção é que nossos resultados mostram que as mutações naturais estudadas são todas exemplos de resíduos que não estão em contato com o ligante e não pertencem ao AF2, mas que possuem um importante papel na afinidade ligante-proteína e na ativação do receptor através da estabilização da conformação ativa, ou melhor dizendo, de um estado conformacional ativo. Outros fatores podem contribuir para esta perda de afinidade como mudanças entrópicas ou fatores entálpicos presentes em outras partes do sistema. Investigar como as mudanças estruturais em torno das mutações se propagam até o sítio de ligação alterando a mobilidade de importantes resíduos desta região, como a entrada das águas está associada à perda de interação do resíduo R752 e, por fim, como estas mudanças afetam outras regiões do AR-LBD importantes para a dimerização e ligação do cofator continuarão sendo objetos de investigação.

6. Bibliografia

1. Bain, D. L., Heneghan, A. F., Connaghan-Jones, K. D. & Miura, M. T. Nuclear receptor structure: implications for function. *Annu. Rev. Physiol.* **69**, 201–220 (2007).

2. Bisson, W. H., Abagyan, R. & Cavasotto, C. N. Molecular basis of agonicity and antagonicity in the androgen receptor studied by molecular dynamics simulations. *J. Mol. Graph. Model.* **27**, 452–458 (2008).

3. Forster, M. J. Molecular modelling in structural biology. *Micron* **33**, 365–384 (2002).

4. Jääskeläinen, J. Molecular biology of androgen insensitivity. *Mol. Cel. Endocrinol.* **352**, 4–12 (2012).

5. Gao, W. Androgen receptor as a therapeutic target. *Adv. Drug. Deliver. Rev.* **62**, 1277–1284 (2010).

6. Lumbroso, S. *et al.* A new mutation of the androgen receptor, P817A, causing partial androgen insensitivity syndrome: in vitro and structural analysis. *J. Mol. Endocrinol.* **32**, 679–687 (2004).

7. Toft, D., Shyamala, G. & Gorski, J. A receptor molecule for estrogens: studies using a cell-free system. *Proc. Natl. Acad. Sci.* **57**, 1740–1743 (1967).

8. Jensen, E. V. & DeSombre, E. R. Mechanism of action of the female sex hormones. *Annu. Rev. Biochem.* **41**, 203–230 (1972).

9. Nagy, L. & Schwabe, J. W. R. Mechanism of the nuclear receptor molecular switch. *Trends Biochem. Sci.* **29**, 317–324 (2004).

10. Gronemeyer, H., Gustafsson, J.-A. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nat. Rev. Drug Discov.* **3**, 950–964 (2004).

11. Chawla, A., Repa, J. J., Evans, R. M. & Mangelsdorf, D. J. Nuclear receptors and lipid physiology: opening the x-files. *Science* **30**, 1866–1870 (2001).

12. Shao, W. & Brown, M. Advances in estrogen receptor biology: prospects for improvements in targeted breast cancer therapy. *Breast. Cancer Res.* **6**, 39–52 (2004).

13. Kersten, S., Desvergne, B. & Wahli, W. Roles of PPARs in health and disease. *Nature* **405**, 421–424 (2000).

14. Laue, L. *et al.* Receptor-mediated effects of glucocorticoids on inflammation: enhancement of the inflammatory response with a glucocorticoid antagonist. *J. Steroid Biochem.* **29**, 591–598 (1988).

15. Wurtz, J.-M. *et al.* A canonical structure for the ligand-binding domain of nuclear receptors. *Nat. Struct. Biol.* **3**, 87–94 (1996).

16. Koehl, P. & Levitt, M. Improved recognition of native-like structures using a family of designed sequences. *Proc. Natl. Acad. Sci.* **99**, 691–696 (2002).

17. Orengo, C. A. & Thornton, J. M. Protein families and their evolution-a structural perspective. *Annu. Rev. Biochem.* **74**, 867–900 (2005).

18. Koonin, E. V., Wolf, Y. I. & Karev, G. P. The structure of the protein universe and genome evolution. *Nature* **420**, 218–223 (2002).

19. Rausell, A., Juan, D., Pazos, F. & Valencia, A. Protein interactions and ligand binding: from protein subfamilies to functional specificity. *Proc. Natl. Acad. Sci.* **107**, 1995–2000 (2010).

20. Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, C. Transmission of conformational change in insulin. *Nature* **302**, 500–505 (1983).

21. Lesk, A. M. & Chothia, C. Mechanisms of domain closure in proteins. *J. Mol. Biol.* **174**, 175–191 (1984).

22. Gerstein, M. & Chothia, C. Analysis of protein loop closure. Two types of hinges produce one motion in lactate dehydrogenase. *J. Mol. Biol.* **220**, 133–149 (1991).

23. Krebs, W. G. *et al.* Studying protein flexibility in a statistical framework: Tools and databases for analyzing structures and approaches for mapping this onto sequences. *Method. Enzymol.* **374**, 544–584 (2003).

24. Rashin, A. a, Rashin, A. H. L. & Jernigan, R. L. Diversity of function-related conformational changes in proteins: coordinate uncertainty, fragment rigidity, and stability. *Biochemistry-US* **49**, 5683–5704 (2010).

25. Schuler, B. & Eaton, W. A. Protein folding studied by single-molecule FRET. *Curr. Opin. Struct. Biol.* **18**, 16–26 (2008).

26. Rao, F. & Karplus, M. Protein dynamics investigated by inherent structure analysis. *Proc. Natl. Acad. Sci.* **107**, 9152–9157 (2010).

27. Moras, D. & Gronemeyer, H. The nuclear receptor ligand-binding domain: structure and function. *Curr. Opin. Cell Biol.* **10**, 384–391 (1998).

28. Carlsson, P., Koehler, K. F. & Nilsson, L. Glucocorticoid receptor point mutation V571M facilitates coactivator and ligand binding by structural rearrangement and stabilization. *Mol. Endocrinol.* **19**, 1960–1977 (2005).

29. Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and the antagonism of this interaction by tamoxifen. *Cell* **95**, 927–937 (1998).

30. Nolte, R. T. *et al.* Ligand binding and co-activator assembly of the peroxisome proliferator-activated receptor-g. *Nature* **395**, 137–143 (1998).

31. Bourguet, W., Ruff, M., Chambon, P., Gronemeyer, H. & Moras, D. Crystal structure of the ligand-binding domain of the human nuclear receptor RXR- α. *Nature* **375**, 377–382 (1995).

32. Brzozowski, A. M. *et al.* Molecular basis of agonism and antagonism in the oestrogen receptor. *Nature* **389**, 753–758 (1997).

33. Pike, A. C. *et al.* Structure of the ligand-binding domain of oestrogen receptor beta in the presence of a partial agonist and a full antagonist. *Embo J.* **18**, 4608–4618 (1999).

34. Gampe, R. T. J. *et al.* Structural basis for autorepression of retinoid X receptor by tetramer formation and the AF-2 helix. *Gene. Dev.* **14**, 2229–2241 (2000).

35. Bourguet, W. *et al.* Crystal structure of a heterodimeric complex of RAR and RXR ligand-binding domains. *Mol. Cell.* **5**, 289–298 (2000).

36. Pike, A. C. *et al.* Structural insights into the mode of action of a pure antiestrogen. *Structure* **9**, 145–153 (2001).

37. Watkins, R. E. *et al.* The human nuclear xenobiotic receptor PXR: structural determinants of directed promiscuity. *Science* **292**, 2329–2333 (2001).

38. Clayton, G. M., Peak-Chew, S. Y., Evans, R. M. & Schwabe, J. W. The structure of the ultraspiracle ligand-binding domain reveals a nuclear receptor locked in an inactive conformation. *Proc. Natl. Acad. Sci.* **98**, 1549–1554 (2001).

39. Xu, H. *et al.* Structural basis for antagonist-mediated recruitment of nuclear co-repressors by PPARalpha. *Nature* **415**, 813–817 (2002).

40. Sonoda, M. T., Martínez, L., Webb, P., Skaf, M. S. & Polikarpov, I. Ligand dissociation from estrogen receptor is mediated by receptor dimerization: evidence from molecular dynamics simulations. *Mol. Endocrinol.* **22**, 1565–1578 (2008).

41. Martínez, L. *et al.* Molecular Dynamics Simulations Reveal Multiple Pathways of Ligand Dissociation from Thyroid Hormone Receptors. *Biophys. J.* **89**, 2011–2023 (2005).

42. Hu, X. & Lazar, M. A. The CoRNR motif controls the recruitment of corepressors by nuclear hormone receptors. *Nature* **402**, 93–96 (1999).

43. Perissi, V. *et al.* Molecular determinants of nuclear receptor-corepressor interaction. *Gene. Dev.* **13**, 3198–31208 (1999).

44. Nagy, L. *et al.* Mechanism of corepressor binding and release from nuclear hormone receptors. *Gene. Dev.* **13**, 3209–3216 (1999).

45. Webb, P. *et al.* The nuclear receptor corepressor (N-CoR) contains three isoleucine motifs (I/LXXII) that serve as receptor interaction domains (IDs). *Mol. Endocrinol.* **14**, 1976–1985 (2000).

46. Kallenberger, B. C., Love, J. D., Chatterjee, V. K. K. & Schwabe, J. W. R. A dynamic mechanism of nuclear receptor activation and its perturbation in a human disease. *Nat. Struct. Biol.* **10**, 136–140 (2003).

47. Schulman, I. G. & Heyman, R. A. The flip side: identifying small molecule regulators of nuclear receptors. *Chem. Biol.* **11**, 639–646 (2004).

48. Stenson, P. D. *et al.* The human gene mutation database: 2008 update. *Genome Med.* **1**, 13 (2009).

49. Yue, P., Li, Z. & Moult, J. Loss of protein structure stability as a major causative factor in monogenic disease. *J. Mol. Biol.* **353**, 459–473 (2005).

50. Nuclear Receptor Signaling Atlas (NURSA). (2012).at <www.nursa.org>

51. Estébanez-Perpiñá, E. *et al.* A surface on the androgen receptor that allosterically regulates coactivator binding. *Proc. Natl. Acad. Sci.* **104**, 16074–16079 (2007).

52. Gobinet, J., Poujol, N. & Sultan, C. Molecular action of androgens. *Mol. Cel. Endocrinol.* **198**, 15–24 (2002).

53. Wilson, C. M. & McPhaul, M. J. A and B forms of the androgen receptor are expressed in a variety of human tissues. *Mol. Cell. Endocrinol.* **120**, 51–57 (1996).

54. Liegibel, U. M. *et al.* Androgen receptor isoforms AR-A and AR-B display functional differences in cultured human bone cells and genital skin fibroblasts. *Steroids* **68**, 1179–1187 (2003).

55. Gottlieb, B., Beitel, L. K., Wu, J. H. & Trifiro, M. The androgen receptor gene mutations database (ARDB): 2004 update. *Hum. Mutat.* **23**, 527–533 (2004).

56. Centenera, M. M., Harris, J. M., Tilley, W. D. & Butler, L. M. The contribution of different androgen receptor domains to receptor dimerization and signaling. *Mol. Endocrinol.* **22**, 2373–2382 (2008).

57. Palvimo, J. J. The androgen receptor. *Mol. Cel. Endocrinol.* **352**, 1–3 (2012).

58. Bourguet, W., Germain, P. & Gronemeyer, H. Nuclear receptor ligand-binding domains: three-dimensional structures, molecular interactions and pharmacological implications. *Trends Pharmacol. Sci.* **21**, 381–388 (2000).

59. Duff, J., Davies, P., Watt, K. & McEwan, I. J. Structural dynamics of the human androgen receptor: implications for prostate cancer and neurodegenerative disease. *Biochem. Soc. T.* **34**, 1098–1102 (2006).

60. Matias, P. M. *et al.* Structural evidence for ligand specificity in the binding domain of the human androgen receptor. Implications for pathogenic gene mutations. *J. Biol. Chem.* **275**, 26164–26171 (2000).

61. Dehm, S. M. & Tindall, D. J. Androgen receptor structural and functional elements: role and regulation in prostate cancer. *Mol. Endocrinol.* **21**, 2855–2863 (2007).

62. Cantin, L. *et al.* Structural characterization of the human androgen receptor ligand-binding domain complexed with EM5744, a rationally designed steroidal ligand bearing a bulky chain directed toward helix 12. *J. Biol. Chem.* **282**, 30910–30919 (2007).

63. He, B. *et al.* Structural basis for androgen receptor interdomain and coactivator interactions suggests a transition in nuclear receptor activation function dominance. *Mol. Cell.* **16**, 425–438 (2004).

64. Jenster, G. *et al.* Domains of the Human Androgen Receptor Involved in Steroid Binding, Transcriptional Activation, and Subcellular Localization. *Mol. Endocrinol.* **5**, 1396–1404 (1991).

65. Xu, J., Nawaz, Z., Tsai, S. Y., Tsai, M. J. & O'Malley, B. W. The extreme C terminus of progesterone receptor contains a transcriptional repressor domain that functions through a putative corepressor. *Proc. Natl. Acad. Sci.* **93**, 12195–12199 (1996).

66. Zhang, S., Liang, X. & Danielsen, M. Role of the C terminus of the glucocorticoid receptor in hormone binding and agonist/antagonist discrimination. *Mol. Endocrinol.* **10**, 24–34 (1996).

67. Couette, B. *et al.* Folding requirements of the ligand-binding domain of the human mineralocorticoid receptor. *Mol. Endocrinol.* **12**, 855–863 (1998).

68. Tahiri, B., Auzou, G., Nicolas, J., Sultan, C. & Lupo, B. Participation of critical residues from the extreme C-terminal end of the human androgen receptor in the ligand binding function. *Biochemistry-US* **40**, 8431–8437 (2001).

69. Sack, J. S. *et al.* Crystallographic structures of the ligand-binding domains of the androgen receptor and its T877A mutant complexed with the natural agonist dihydrotestosterone. *Proc. Natl. Acad. Sci.* **98**, 4904–4909 (2001).

70. Bledsoe, R. K. *et al.* Crystal structure of the glucocorticoid receptor ligand binding domain reveals a novel mode of receptor dimerization and coactivator recognition. *Cell* **110**, 93–105 (2002).

71. Bardin, C. W. & Catterall, J. F. Testosterone: a major determinant of extragenital sexual dimorphism. *Science* **211**, 1285–1284 (1981).

72. Mao, H.-L., Zhu, Z.-Q. & Chen, C. D. The androgen receptor in hormone-refractory prostate cancer. *Asian J. Androl.* **11**, 69–73 (2009).

73. Gottlieb, B., Beitel, L. K., Nadarajah, A., Paliouras, M. & Trifiro, M. The androgen receptor gene mutations database: 2012 update. *Hum. Mutat.* **33**, 887–894 (2012).

74. Morris, J. M. The Syndrome of Testicular Feminization in Male Pseudohermaphrodites. *Am. J. Obstet. Gynecol.* **65**, 1192–1211 (1953).

75. Elhaji, Y. a *et al.* Impaired helix 12 dynamics due to proline 892 substitutions in the androgen receptor are associated with complete androgen insensitivity. *Hum. Mol. Genet.* **15**, 921–931 (2006).

76. Wu, J. H. *et al.* Bridging structural biology and genetics by computational methods: an investigation into how the R774C mutation in the AR gene can result in complete androgen insensitivity syndrome. *Hum. Mutat.* **22**, 465–475 (2003).

77. Langley, E., Kemppainen, J. a & Wilson, E. M. Intermolecular NH2-/carboxyl-terminal interactions in androgen receptor dimerization revealed by mutations that cause androgen insensitivity. *J. Biol. Chem.* **273**, 92–101 (1998).

78. Elhaji, Y. a *et al.* An examination of how different mutations at arginine 855 of the androgen receptor result in different androgen insensitivity phenotypes. *Mol. Endocrinol.* **18**, 1876–1886 (2004).

79. Koochekpour, S. Androgen receptor signaling and mutations in prostate cancer. *Asian J. Androl.* **12**, 639–657 (2010).

80. Vasaitis, T. S., Bruno, R. D. & Njar, V. C. O. CYP17 inhibitors for prostate cancer therapy. *J. Steroid Biochem.* **125**, 23–31 (2011).

81. Marcelli, M. *et al.* Androgen receptor mutations in prostate cancer. *Cancer Res.* **60**, 944–949 (2000).

82. Georget, V., Térouanne, B., Lumbroso, S., Nicolas, J. C. & Sultan, C. Trafficking of androgen receptor mutants fused to green fluorescent protein: a new investigation of partial androgen insensitivity syndrome. *J. Clin. Endocr. Metab.* **83**, 3597–3603 (1998).

83. Shkolny, D. L. *et al.* Discordant measures of androgen-binding kinetics in two mutant androgen receptors causing mild or partial androgen insensitivity, respectively. *J. Clin. Endocr. Metab.* **84**, 805–810 (1999).

84. Gaiduk, A., Yorulmaz, M., Ruijgrok, P. V. & Orrit, M. Room temperature detection of a single molecule's absorption by photothermal contrast. *Science* **330**, 353–356 (2010).

85. Zewail, A. H. 4D ultrafast electron diffraction crystallography and microscopy. *Annu. Rev. Phys. Chem.* **57**, 65–103 (2006).

86. *Encyclopedia of Computational Chemistry*. 1036 (John Wiley and Sons, Inc.: London, 1998).

87. Leach, A. R. *Molecular Modelling: principles and applications*. 563 (Pearson Prentice Hall: England, 2001).

88. Alder, B. J. & Wainwright, T. E. Phase transition for a hard sphere system. *J. Chem. Phys.* **27**, 1208–1209 (1957).

89. Rahman, A. & Stillinger, F. H. Molecular dynamics study of liquid water. *J. Chem. Phys.* **55**, 3336–3359 (1971).

90. McCammon, J. A. & Karplus, M. Internal motions of antibody molecules. *Nature* **268**, 765–766 (1977).

91. Linderstrom-Lang, K. Deuterium exchange between paptides and water. *Chem. Soc. (London)* **2**, 1–20 (1955).

92. Frauenfelder, H., Petsko, G. A. & Tsernoglou, D. Temperature-dependent X-ray diffraction as a probe of protein structural dynamics. *Nature* **280**, 558–563 (1979).

93. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **9**, 646–652 (2002).

94. Mobley, D. L. & Dill, K. a Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure* **17**, 489–498 (2009).

95. Kukol, A. *Molecular Modeling of Proteins. Gene Therapy* 390 (Humana Press: Hertfordshire, 2008).

96. McCammon, J. A., Gelin, B. R. & Karplus, M. Dynamics of folded proteins. *Nature* **267**, 585–590 (1977).

97. Morgon, N. H. & Coutinho, K. *Métodos de Química Teórica E Modelagem Molecular.* 533 (Livraria da Física Editora: São Paulo, 2007).

98. Walton, E. & VanVliet, K. Equilibration of experimentally determined protein structures for molecular dynamics simulation. *Phys. Rev. E* **74**, 1–8 (2006).

99. Zhou, M. & Robinson, C. V. When proteomics meets structural biology. *Trends Biochem. Sci.* **35**, 522–529 (2010).

100. Namba, A. M., Da Silva, V. B. & Da Silva, C. H. T. P. Dinâmica molecular : teoria e aplicações em planejamento de fármacos. *Eclet. Quim.* **33**, 13–23 (2008).

101. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **26**, 1781–1802 (2005).

102. Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. van, DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690 (1984).

103. Nosé, S. A molecular dynamics method for simulations in the canonical ensemble. *Mol. Phys.* **52**, 255–268 (1984).

104. Hoover, W. G. Canonical dynamics: Equilibrium phase-space distributions. *Phys. Rev. A* **31**, 1695–1697 (1985).

105. Parrinello, M. & Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **52**, 7182–7190 (1981).

106. Ryckaert, J. ., Ciccotti, G. & Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of n-alkanes. *J. Comp. Phys.* **23**, 327–341 (1977).

107. Andersen, H. C. Rattle: A “Velocity” Version of the Shake Algorithm for Molecular Dynamics Calculations. *J. Comp. Phys.* **52**, 24–34 (1983).

108. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulation. *J. Comput. Chem.* **18**, 1463–1472 (1998).

109. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).

110. Mahoney, M. W. & Jorgensen, W. L. A five-site model liquid water and the reproduction of the density anomaly by rigid, non-polarizable models. *J. Chem. Phys.* **112**, 8910–8922 (2000).

111. Berendsen, H. J. C., Grigera, J. R. & Straatsma, T. P. The Missing Term in Effective Pair Potentials. *J. Phys. Chem.* **91**, 6269–6271 (1987).

112. Stillinger, F. H. & Rahman, A. Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* **60**, 1545–1557 (1974).

113. Roux, B. & Simonson, T. Implicit solvent models. *Biophys. Chem.* **78**, 1–20 (1999).

114. Ferrara, P., Apostolakis, J. & Caflisch, A. Evaluation of a fast implicit solvent model for molecular dynamics simulations. *Proteins Struct. Funct. Genet.* **46**, 24–33 (2002).

115. Gilson, M. K., Davis, M. E., Luty, B. A. & McCammon, J. A. Computation of electrostatic forces on solvated molecules using the Poisson-Boltzmann equation. *J. Phys. Chem.* **97**, 3591–3600 (1993).

116. Baker, N. A. Improving implicit solvent simulations: a Poisson-centric view. *Curr. Opin. Struct. Biol.* **15**, 137–143 (2005).

117. Still, W. C., Tempczyk, A., Hawley, R. C. & Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**, 6127–6129 (1990).

118. Bashford, D. & Case, D. A. Generalized Born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.* **51**, 129–152 (2000).

119. Verlet, L. Computer “Experiments” on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98–103 (1967).

120. Hockney, R. W. The potential calculation and some applications. *Method. Comp. Phys.* **9**, 136–211 (1970).

121. Swope, W. C., Andersen, H. C., Berens, P. H. & Wilson, K. R. A computer simulation method for the calculation of equilibrium constants for the formation of

physical clusters of molecules: application to small water clusters. *J. Chem. Phys.* **76**, 637–649 (1982).

122. Mackerell, Jr., A. D. *et al.* All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **102**, 3586–3616 (1998).

123. Humphrey, W., Dalke, A. & Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **14**, 33–38 (1996).

124. Frisch, M. J. *et al.* Gaussian 03. (2004).

125. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).

126. Caves, L. S., Evanseck, J. D. & Karplus, M. Locally accessible conformations of proteins: multiple molecular dynamics simulations of crambin. *Protein Sci.* **7**, 649–666 (1998).

127. Poujol, N. *et al.* Pathophysiology of Androgen Insensitivity Syndromes: Molecular and Structural Approaches of Natural and Engineered Androgen Receptor Mutations at Amino Acid 743. *J. Clin. Endocr. Metab.* **87**, 5793–5800 (2002).

128. He, B. *et al.* Probing the functional link between androgen receptor coactivator and ligand-binding sites in prostate cancer and androgen insensitivity. *J. Biol. Chem.* **281**, 6648–6663 (2006).

129. Kearsley, S. K. On the orthogonal transformation used for structural comparisons. *Acta Cryst. A* **45**, 208–210 (1989).

130. Sakai, N. *et al.* Bilateral testicular tumors in androgen insensitivity syndrome. *Int. J. Urol.* **7**, 390–392 (2000).

131. Xu, X. *et al.* Dynamic communication between androgen and coactivator: mutually induced conformational perturbations in androgen receptor ligand-binding domain. *Proteins* **79**, 1154–1171 (2011).

132. Marhefka, C. *et al.* Homology modeling using multiple molecular dynamics simulations and docking studies of the human androgen receptor ligand binding domain bound to testosterone and nonsteroidal ligands. *J. Med. Chem.* **44**, 1729–1740 (2001).

133. Bevan, C. L., Hughes, I. A. & Patterson, M. N. Wide variation in androgen receptor dysfunction in complete androgen insensitivity syndrome. *J. Steroid Biochem. Mol. Biol.* **61**, 19–26 (1997).

Anexo I: Percentuais de tempo relativo de contatos

Neste anexo, são apresentadas as porcentagens de tempo de contato entre os resíduos do domínio F com os demais resíduos até 4 Å deles para cada uma das mutações deste estudo. Apenas os contatos que apresentaram diferenças maiores do que 5% entre a estrutura nativa e a mutada são apresentados.

Tabela 9: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação F916A.

Resíduo		% de contato	
F	Contato	Nativa	F916A
G909	V818 (S3/H9)	90	83
K912	V818 (S3/H9)	98	89
I914	F813 (H8)	85	78
I914	Q867 (H10)	21	14
H917	K861 (H10)	88	80

Tabela 10: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação F916A.

Resíduo		% de contato	
F	Contato	Nativa	F916A
I914	D864 (H10)	9	29
Y915	Q867 (H10)	81	92
Y915	P868 (H10)	76	85
H917	F856 (H10)	17	34
T918	Y857 (H10)	4	15
T918	K861 (H10)	6	22
Q919	K861 (H10)	26	46
Q919	D864 (H10)	35	70
Q919	S865 (H10)	1	31
Q919	Q867 (H10)	1	9
Q919	P868 (H10)	7	32

P817A

Tabela 11: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação P817A.

Resíduo		% de contato	
F	Contato	Nativa	P817A
G909	V818 (S3/H9)	90	58
V911	V818 (S3/H9)	89	65
K912	V818 (S3/H9)	98	80
I914	Q824 (H9)	11	5
F916	S814 (H8/S3)	15	4
Q919	K861 (H10)	26	14

Tabela 12: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação P817A.

Resíduo		% de contato	
F	Contato	Nativa	P817A
Y915	R831 (H9)	11	17
Y915	Q867 (H10)	81	88
Y915	P868 (H10)	76	84
Y915	R871 (H10)	0	18
F916	R831 (H9)	85	98
F916	F856 (H10)	66	72
Q919	D864 (H10)	35	47
Q919	P868 (H10)	7	13

R840C

Tabela 13: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação R840C.

Resíduo		% de tempo de contato	
F	Contato	Nativa	R840C
G909	K912 (S4)	63	55
K912	Y915 (F)	31	24
Y915	T918 (F)	20	9
F916	L810 (H8)	92	87
F916	S814 (H8/S3)	15	4
F916	L863 (H10)	91	79
F916	Q919 (F)	10	4
H917	F856 (H10)	17	7
H917	T860 (H10)	98	91
Q919	K861(H10)	26	16

Tabela 14: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação R840C.

Resíduo		% de tempo de contato	
F	Contato	Nativa	R840C
P913	R871(H10)	3	12
P913	I906 (H12)	50	61
I914	Q867 (H10)	21	28
Y915	Q867 (H10)	81	92
Y915	P868 (H10)	76	87
Y915	Q919 (F)	67	81
F916	R831 (H9)	85	99
F916	F856 (H10)	66	74
F916	Y857 (H10)	0	6
H917	D864 (H10)	87	93
Q919	D864 (H10)	35	47
Q919	S865 (H10)	1	7
Q919	P868 (H10)	7	17

G743V

Tabela 15: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação G743V.

Resíduo		% de contato	
F	Contato	Nativa	G743V
K912	Y915 (F)	31	21
I914	F813 (H8)	85	77
Y915	T918 (F)	20	6
H917	F856 (H10)	17	9

Tabela 16: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação G743V.

Resíduo		% de contato	
F	Contato	Nativa	G743V
G909	K912 (F)	63	78
K912	I906 (H12)	15	28
P913	I906 (H12)	50	62
Y915	Q867 (H10)	81	93
Y915	P868 (H10)	76	90
Y915	Q919 (F)	67	92
F916	R831 (H9)	85	97
H917	D864 (H10)	87	94
Q919	D864 (H10)	35	62
Q919	S865 (H10)	1	16
Q919	P868 (H10)	7	26
Q919	Y915 (F)	67	92

R871G

Tabela 17: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação R871G.

Resíduo		% de tempo de contato	
F	Contato	Nativa	R871G
F916	L863 (H10)	91	84
H917	F856 (H10)	17	12

Tabela 18: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação R871G.

Resíduo		% de tempo de contato	
F	Contato	Nativa	R871G
G909	V818 (S3/H9)	90	96
K910	D819 (S3/H9)	13	23
K912	I906 (H12)	15	26
P913	I906 (H12)	50	75
I914	F813 (H8)	85	92
Y915	Q867 (H10)	81	91
Y915	P868 (H10)	76	88
Y915	Q919 (F)	67	83
F916	R831 (H9)	85	98
F916	F856 (H10)	66	73
H917	D864 (H10)	87	94
Q919	D864 (H10)	35	53
Q919	S865 (H10)	1	9
Q919	P868 (H10)	7	23

G909E

Tabela 19: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação G909E.

Resíduo		% de contato	
F	Contato	Nativa	G909E
909E	P904 (H12)	10	4
K912	Y915 (F)	31	20
I914	F813 (H8)	85	78
Y915	T918 (F)	20	11
F916	S814 (H8/S3)	15	5
Q919	K861 (H10)	26	16

Tabela 20: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação G909E.

Resíduo		% de contato	
F	Contato	Nativa	G909E
909E	R871 (H10)	7	51
909E	I906 (H12)	84	100
909E	K912 (S4)	63	86
K910	I906 (H12)	31	48
Y915	Q867 (H10)	81	92
Y915	P868 (H10)	76	87
Y915	Q919 (F)	67	74
F916	R831 (H9)	85	98
F916	F856 (H10)	66	74
H917	D864 (H10)	87	92
Q919	D864 (H10)	35	48
Q919	S865 (H10)	1	8
Q919	P868 (H10)	7	18

Q919R

Tabela 21: Resíduos que apresentam perda de contatos com resíduos do domínio F quando da mutação Q919R.

Resíduo		% de tempo de contato	
F	Contato	Nativa	Q919R
V911	V818 (S3/H9)	89	83
K912	Y915 (F)	31	15
F916	919R (F)	10	3
919R	K861 (H10)	26	20

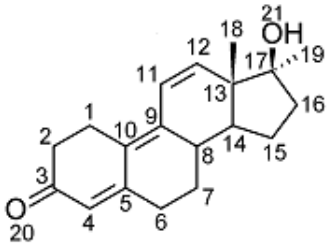
Tabela 22: Resíduos que apresentam ganho de contatos com resíduos do domínio F quando da mutação Q919R.

Resíduo		% de tempo de contato	
F	Contato	Nativa	Q919R
P913	I906 (H12)	50	62
Y915	Q867 (H10)	81	94
Y915	P868 (H10)	76	89
Y915	T918 (F)	20	34
Y915	919R (F)	67	72
F916	R831 (H9)	85	95
F916	F856 (H10)	66	74
H917	D864 (H10)	87	94
919R	D864 (H10)	35	41
919R	S865 (H10)	1	20
919R	P868 (H10)	7	44

Anexo II: Parâmetros do Ligante R1881

O esquema de numeração do ligante R1881 é mostrado na Tabela 23, juntamente com os tipos atômicos atribuídos a cada átomo e suas respectivas cargas parciais calculadas.

Tabela 23: Tipos atômicos atribuídos e cargas parciais calculadas para todos os átomos do ligante R1881.



Átomo	Tipo	Carga (u.a.)	Átomo	Tipo	Carga (u.a.)
C1	CTL2	+0,085	**HC12**	HEL1	+0,193
HC1	HAL2	+0,011	**C13**	CTL1	+0,791
HC1	HAL2	+0,032	**C14**	CTL1	-0,395
C2	CTL2	-0,271	**HC14**	HAL1	+0,098
HC2	HAL2	+0,076	**C15**	CTL2	+0,028
HC2	HAL2	+0,073	**HC15**	HAL2	+0,013
C3	C	+0,764	**HC15**	HAL2	+0,029
C4	CEL1	-0,506	**C16**	CTL2	-0,276
HC4	HEL1	+0,180	**HC16**	HAL2	+0,072
C5	CEL1	+0,121	**HC16**	HAL2	+0,074
C6	CTL2	-0,105	**C17**	CTL1	+0,393
HC6	HAL2	+0,053	**C18**	CTL3	-0,574
HC6	HAL2	+0,079	**HC18**	HAL3	+0,133
C7	CTL2	-0,112	**HC18**	HAL3	+0,150
HC7	HAL2	+0,028	**HC18**	HAL3	+0,123
HC7	HAL2	+0,055	**C19**	CTL3	-0,464
C8	CTL1	+0,270	**HC19**	HAL3	+0,112
HC8	HAL1	+0,017	**HC19**	HAL3	+0,113
C9	CEL1	-0,138	**HC19**	HAL3	+0,106
C10	CEL1	-0,039	**O20**	ON1	-0,590
C11	CEL1	-0,172	**O21**	OHL	-0,787
HC11	HEL1	+0,156	**HO21**	HOL	+0,468
C12	CEL1	-0,460			

Os arquivos de parâmetros e de topologia do ligante R1881 estão reproduzidos abaixo.

II.1. ARQUIVO DE PARÂMETROS

```
CHARMM parameters for components R18

BONDS
!
!V(bond) = Kb(b - b0)**2
!
!Kb: kcal/mole/A**2
!b0: A
!
!atom type   Kb         b0
!

C    CTL2    241.582    1.5030
CEL1 C       344.021    1.4375
ON1  C       737.018    1.2280
CEL1 CTL1    178.149    1.5432
CTL1 CEL1    235.428    1.5190


ANGLES
!
!V(angle) = Ktheta(Theta - Theta0)**2
!
!V(Urey-Bradley) = Kub(S - S0)**2
!
!Ktheta: kcal/mole/rad**2
!Theta0: degrees
!Kub: kcal/mole/A**2 (Urey-Bradley)
!S0: A
!
!atom types        Ktheta    Theta0    Kub      S0
!

C    CTL2 CTL2     17.754    109.94
CEL1 C    CTL2     13.869    113.43
ON1  C    CTL2      8.089    123.81
HAL2 CTL2 C        13.134    112.08
HAL2 CTL2 C         5.140    112.04
CEL1 CEL1 C        17.693    121.62
HEL1 CEL1 C        35.078    121.97
ON1  C    CEL1     35.495    122.67
CEL1 CEL1 CEL1     21.023    122.84
CEL1 CEL1 CEL1     15.562    121.17
CEL1 CTL1 CTL2     11.891    110.21
CEL1 CEL1 CTL1     11.057    120.13
CEL1 CEL1 CTL1     13.922    116.19
CTL1 CTL1 CEL1      5.470    109.75
HAL1 CTL1 CEL1     17.612    109.08
CEL1 CEL1 CEL1     18.293    123.95
CEL1 CEL1 CEL1     20.526    123.64
CTL1 CEL1 CEL1     21.796    121.36
CTL1 CTL1 CEL1     15.707    108.34
CTL1 CTL1 CEL1     22.388    119.93
CTL3 CTL1 CEL1     19.555    107.06
```

```
HEL1 CEL1 CTL1      6.808    125.46
OHL  CTL1 CTL1     11.240    113.87
OHL  CTL1 CTL3     31.109    110.12


DIHEDRALS
!
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))
!
!Kchi: kcal/mole
!n: multiplicity
!delta: degrees
!
!atom types             Kchi     n    delta
!

CEL1 C    CTL2 CTL2     0.1128   6    233.58
ON1  C    CTL2 CTL2     0.1112   6     50.17
CEL1 C    CTL2 HAL2     0.1652   6    108.29
ON1  C    CTL2 HAL2     0.1408   6    284.88
CEL1 C    CTL2 HAL2     0.1370   6    180.00
ON1  C    CTL2 HAL2     0.1264   6    175.43
CEL1 CEL1 C    CTL2     0.8935   2    155.18
HEL1 CEL1 C    CTL2     0.7383   2    335.22
CEL1 CEL1 C    ON1      0.8572   2    338.55
HEL1 CEL1 C    ON1      0.8397   2    158.58
CEL1 CEL1 CTL1 CTL2     0.1010   6      0.00
CEL1 CEL1 CTL1 CTL2     0.1252   6      0.00
CEL1 CEL1 CTL1 CTL1     0.0949   6      0.00
CEL1 CEL1 CTL1 CTL1     0.0761   6      0.00
CEL1 CEL1 CTL1 HAL1     0.1010   6      0.00
CEL1 CEL1 CTL1 HAL1     0.1053   6      0.00
CTL1 CTL1 CEL1 CEL1     0.0683   6      0.00
CTL1 CTL1 CEL1 CEL1     0.1128   6    318.48
CTL3 CTL1 CEL1 CEL1     0.1339   6      0.00
CTL1 CTL1 CEL1 HEL1     0.0627   6      0.00
CTL1 CTL1 CEL1 HEL1     0.0976   6    138.51
CTL3 CTL1 CEL1 HEL1     0.1041   6      0.00

IMPROPER
!
!V(improper) = Kpsi(psi - psi0)**2
!
!Kpsi: kcal/mole/rad**2
!psi0: degrees
!note that the second column of numbers (0) is ignored
!
!atom types             Kpsi        0  psi0
!
ON1   X    X   C         90.0       0  0.0000


NONBONDED nbxmod  5 atom cdiel shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
                !adm jr., 5/08/91, suggested cutoff scheme
!
!V(Lennard-Jones) = Eps(i,j)*[(Rmin(i,j)/r(i,j))**12 - 2(Rmin(i,j)/r(i,j))**6]
```

```
!
!epsilon [kcal/mole]: Eps(i,j) = sqrt(eps(i) * eps(j))
!Rmin/2  [A]:         Rmin(i,j) = Rmin/2(i) + Rmin/2(j)
!
!atom  ignored   epsilon     Rmin/2  ignored   eps,1-4     Rmin/2,1-4
!
! from lipids

HOL      0.0       -0.046     0.2245
HAL1     0.0       -0.022     1.3200 ! alkane, 3/92
HAL2     0.0       -0.028     1.3400 ! alkane, yin and mackerell, 4/98
HAL3     0.0       -0.024     1.3400 ! alkane, yin and mackerell, 4/98
HEL1     0.0       -0.031     1.25   ! alkene, yin,adm jr., 12/95
HEL2     0.0       -0.026     1.26   ! alkene, yin,adm jr., 12/95
!
CTL1     0.0       -0.0200    2.275 0.0 -0.01 1.9 ! alkane, 3/92
CTL2     0.0       -0.0560    2.010 0.0 -0.01 1.9 ! alkane, 4/98, yin, adm jr.
CTL3     0.0       -0.0780    2.040 0.0 -0.01 1.9 ! alkane, 4/98, yin, adm jr.
CEL1     0.0       -0.068     2.09   ! alkene, yin,adm jr., 12/95
C      0.000000  -0.110000     2.000000 ! ALLOW   PEP POL ARO
                ! NMA pure solvent, adm jr., 3/3/93
!
OHL      0.0       -0.1521    1.77
ON1      0.0       -0.1200    1.70

END
```

II. 2. ARQUIVO DE TOPOLOGIA

```
MASS     1    C 12.01100  C !
MASS     2 CEL1 12.01100  C !
MASS     3 CTL1 12.01100  C !
MASS     4 CTL2 12.01100  C !
MASS     5 CTL3 12.01100  C !
MASS     6 HAL1  1.00800  H !
MASS     7 HAL2  1.00800  H !
MASS     8 HAL3  1.00800  H !
MASS     9 HEL1  1.00800  H !
MASS    10  HOL  1.00800  H !
MASS    11  OHL 15.99900  O !
MASS    12  ON1 15.99900  O !

RESI  R18         +0.000 ! type from CHARMM
GROUP
ATOM   C1 CTL2     +0.084777  ! type from CHARMM
ATOM   C2 CTL2     -0.270800  ! type from CHARMM
ATOM   C3    C     +0.763667  ! type from CHARMM
ATOM   C4 CEL1     -0.506621  ! type from CHARMM
ATOM   C5 CEL1     +0.121131  ! type from CHARMM
ATOM   C6 CTL2     -0.105266  ! type from CHARMM
ATOM   C7 CTL2     -0.112395  ! type from CHARMM
ATOM   C8 CTL1     +0.270251  ! type from CHARMM
ATOM   C9 CEL1     -0.138436  ! type from CHARMM
ATOM  C10 CEL1     -0.039214  ! type from CHARMM
```

```
ATOM  C11 CEL1     -0.171700  ! type from CHARMM
ATOM  C12 CEL1     -0.460090  ! type from CHARMM
ATOM  C13 CTL1     +0.790626  ! type from CHARMM
ATOM  C14 CTL1     -0.395219  ! type from CHARMM
ATOM  C15 CTL2     +0.028267  ! type from CHARMM
ATOM  C16 CTL2     -0.275921  ! type from CHARMM
ATOM  C17 CTL1     +0.393163  ! type from CHARMM
ATOM  C18 CTL3     -0.574688  ! type from CHARMM
ATOM  C27 CTL3     -0.463999  ! type from CHARMM
ATOM  O83  ON1     -0.598953  ! type from CHARMM
ATOM  O97  OHL     -0.787213  ! type from CHARMM
ATOM H12A HAL2     +0.011309  ! type from CHARMM
ATOM H11A HAL2     +0.032480  ! type from CHARMM
ATOM  H22 HAL2     +0.076152  ! type from CHARMM
ATOM  H21 HAL2     +0.072863  ! type from CHARMM
ATOM   H4 HEL1     +0.180257  ! type from CHARMM
ATOM  H62 HAL2     +0.053490  ! type from CHARMM
ATOM  H61 HAL2     +0.079037  ! type from CHARMM
ATOM  H72 HAL2     +0.028670  ! type from CHARMM
ATOM  H71 HAL2     +0.055492  ! type from CHARMM
ATOM   H8 HAL1     +0.017268  ! type from CHARMM
ATOM  H11 HEL1     +0.155577  ! type from CHARMM
ATOM  H12 HEL1     +0.193455  ! type from CHARMM
ATOM  H14 HAL1     +0.097749  ! type from CHARMM
ATOM H152 HAL2     +0.013403  ! type from CHARMM
ATOM H151 HAL2     +0.029947  ! type from CHARMM
ATOM H162 HAL2     +0.072343  ! type from CHARMM
ATOM H161 HAL2     +0.073849  ! type from CHARMM
ATOM H183 HAL3     +0.133451  ! type from CHARMM
ATOM H182 HAL3     +0.150006  ! type from CHARMM
ATOM H181 HAL3     +0.123282  ! type from CHARMM
ATOM H273 HAL3     +0.111703  ! type from CHARMM
ATOM H272 HAL3     +0.113030  ! type from CHARMM
ATOM H271 HAL3     +0.106104  ! type from CHARMM
ATOM  H97  HOL     +0.467738  ! type from CHARMM
BOND   C2   C1   C10   C1  H12A   C1  H11A   C1
BOND   C3   C2   H22   C2   H21   C2    C4   C3
BOND  O83   C3    C5   C4    H4   C4    C6   C5
BOND  C10   C5    C7   C6   H62   C6   H61   C6
BOND   C8   C7   H72   C7   H71   C7    C9   C8
BOND  C14   C8    H8   C8   C10   C9   C11   C9
BOND  C12  C11   H11  C11   C13  C12   H12  C12
BOND  C14  C13   C17  C13   C18  C13   C15  C14
BOND  H14  C14   C16  C15  H152  C15  H151  C15
BOND  C17  C16  H162  C16  H161  C16   C27  C17
BOND  O97  C17  H183  C18  H182  C18  H181  C18
BOND H273  C27  H272  C27  H271  C27   H97  O97
ANGLE   C3   C2   C1    H22   C2   C1    H21   C2   C1
ANGLE   C5  C10   C1     C9  C10   C1    C10   C1   C2
ANGLE H12A   C1   C2   H11A   C1   C2     C4   C3   C2
ANGLE  O83   C3   C2    H22   C2   C3    H21   C2   C3
ANGLE   C5   C4   C3     H4   C4   C3    O83   C3   C4
ANGLE   C6   C5   C4    C10   C5   C4     H4   C4   C5
ANGLE   C7   C6   C5    H62   C6   C5    H61   C6   C5
ANGLE   C9  C10   C5    C10   C5   C6     C8   C7   C6
ANGLE  H72   C7   C6    H71   C7   C6    H62   C6   C7
ANGLE  H61   C6   C7     C9   C8   C7    C14   C8   C7
```

```
ANGLE   H8   C8   C7    H72   C7   C8    H71   C7   C8
ANGLE  C10   C9   C8    C11   C9   C8    C13  C14   C8
ANGLE  C15  C14   C8    H14  C14   C8    C14   C8   C9
ANGLE   H8   C8   C9    C12  C11   C9    H11  C11   C9
ANGLE H12A   C1  C10   H11A   C1  C10    C11   C9  C10
ANGLE  C13  C12  C11    H12  C12  C11    H11  C11  C12
ANGLE  C14  C13  C12    C17  C13  C12    C18  C13  C12
ANGLE  H12  C12  C13    C15  C14  C13    H14  C14  C13
ANGLE  C16  C17  C13    C27  C17  C13    O97  C17  C13
ANGLE H183  C18  C13   H182  C18  C13   H181  C18  C13
ANGLE   H8   C8  C14    C17  C13  C14    C18  C13  C14
ANGLE  C16  C15  C14   H152  C15  C14   H151  C15  C14
ANGLE  H14  C14  C15    C17  C16  C15   H162  C16  C15
ANGLE H161  C16  C15   H152  C15  C16   H151  C15  C16
ANGLE  C27  C17  C16    O97  C17  C16    C18  C13  C17
ANGLE H162  C16  C17   H161  C16  C17   H273  C27  C17
ANGLE H272  C27  C17   H271  C27  C17    H97  O97  C17
ANGLE  O97  C17  C27   H11A   C1 H12A    H21   C2  H22
ANGLE  H61   C6  H62    H71   C7  H72   H151  C15 H152
ANGLE H161  C16 H162   H182  C18 H183   H181  C18 H183
ANGLE H181  C18 H182   H272  C27 H273   H271  C27 H273
ANGLE H271  C27 H272
DIHED   C3   C2   C1  C10    H22   C2   C1  C10
DIHED  H21   C2   C1  C10     C3   C2   C1 H12A
DIHED  H22   C2   C1 H12A    H21   C2   C1 H12A
DIHED   C3   C2   C1 H11A    H22   C2   C1 H11A
DIHED  H21   C2   C1 H11A     C5  C10   C1   C2
DIHED   C9  C10   C1   C2     C5  C10   C1 H12A
DIHED   C9  C10   C1 H12A     C5  C10   C1 H11A
DIHED   C9  C10   C1 H11A     C4   C3   C2   C1
DIHED  O83   C3   C2   C1     C4   C3   C2  H22
DIHED  O83   C3   C2  H22     C4   C3   C2  H21
DIHED  O83   C3   C2  H21     C5   C4   C3   C2
DIHED   H4   C4   C3   C2     C5   C4   C3  O83
DIHED   H4   C4   C3  O83     C6   C5   C4   C3
DIHED  C10   C5   C4   C3     C6   C5   C4   H4
DIHED  C10   C5   C4   H4     C7   C6   C5   C4
DIHED  H62   C6   C5   C4    H61   C6   C5   C4
DIHED   C7   C6   C5  C10    H62   C6   C5  C10
DIHED  H61   C6   C5  C10     C1  C10   C5   C4
DIHED   C9  C10   C5   C4     C1  C10   C5   C6
DIHED   C9  C10   C5   C6     C8   C7   C6   C5
DIHED  H72   C7   C6   C5    H71   C7   C6   C5
DIHED   C8   C7   C6  H62    H72   C7   C6  H62
DIHED  H71   C7   C6  H62     C8   C7   C6  H61
DIHED  H72   C7   C6  H61    H71   C7   C6  H61
DIHED   C9   C8   C7   C6    C14   C8   C7   C6
DIHED   H8   C8   C7   C6     C9   C8   C7  H72
DIHED  C14   C8   C7  H72     H8   C8   C7  H72
DIHED   C9   C8   C7  H71    C14   C8   C7  H71
DIHED   H8   C8   C7  H71    C10   C9   C8   C7
DIHED  C11   C9   C8   C7    C10   C9   C8  C14
DIHED  C11   C9   C8  C14    C10   C9   C8   H8
DIHED  C11   C9   C8   H8    C13  C14   C8   C7
DIHED  C15  C14   C8   C7    H14  C14   C8   C7
DIHED  C13  C14   C8   C9    C15  C14   C8   C9
DIHED  H14  C14   C8   C9    C13  C14   C8   H8
```

```
DIHED  C15  C14   C8   H8    H14  C14   C8   H8
DIHED   C1  C10   C9   C8     C5  C10   C9   C8
DIHED   C1  C10   C9  C11     C5  C10   C9  C11
DIHED  C12  C11   C9   C8    H11  C11   C9   C8
DIHED  C12  C11   C9  C10    H11  C11   C9  C10
DIHED  C13  C12  C11   C9    H12  C12  C11   C9
DIHED  C13  C12  C11  H11    H12  C12  C11  H11
DIHED  C14  C13  C12  C11    C17  C13  C12  C11
DIHED  C18  C13  C12  C11    C14  C13  C12  H12
DIHED  C17  C13  C12  H12    C18  C13  C12  H12
DIHED   C8  C14  C13  C12    C15  C14  C13  C12
DIHED  H14  C14  C13  C12     C8  C14  C13  C17
DIHED  C15  C14  C13  C17    H14  C14  C13  C17
DIHED   C8  C14  C13  C18    C15  C14  C13  C18
DIHED  H14  C14  C13  C18    C16  C17  C13  C12
DIHED  C27  C17  C13  C12    O97  C17  C13  C12
DIHED  C16  C17  C13  C14    C27  C17  C13  C14
DIHED  O97  C17  C13  C14    C16  C17  C13  C18
DIHED  C27  C17  C13  C18    O97  C17  C13  C18
DIHED H183  C18  C13  C12   H182  C18  C13  C12
DIHED H181  C18  C13  C12   H183  C18  C13  C14
DIHED H182  C18  C13  C14   H181  C18  C13  C14
DIHED H183  C18  C13  C17   H182  C18  C13  C17
DIHED H181  C18  C13  C17    C16  C15  C14   C8
DIHED H152  C15  C14   C8   H151  C15  C14   C8
DIHED  C16  C15  C14  C13   H152  C15  C14  C13
DIHED H151  C15  C14  C13    C16  C15  C14  H14
DIHED H152  C15  C14  H14   H151  C15  C14  H14
DIHED  C17  C16  C15  C14   H162  C16  C15  C14
DIHED H161  C16  C15  C14    C17  C16  C15 H152
DIHED H162  C16  C15 H152   H161  C16  C15 H152
DIHED  C17  C16  C15 H151   H162  C16  C15 H151
DIHED H161  C16  C15 H151    C13  C17  C16  C15
DIHED  C27  C17  C16  C15    O97  C17  C16  C15
DIHED  C13  C17  C16 H162    C27  C17  C16 H162
DIHED  O97  C17  C16 H162    C13  C17  C16 H161
DIHED  C27  C17  C16 H161    O97  C17  C16 H161
DIHED H273  C27  C17  C13   H272  C27  C17  C13
DIHED H271  C27  C17  C13   H273  C27  C17  C16
DIHED H272  C27  C17  C16   H271  C27  C17  C16
DIHED H273  C27  C17  O97   H272  C27  C17  O97
DIHED H271  C27  C17  O97    H97  O97  C17  C13
DIHED  H97  O97  C17  C16    H97  O97  C17  C27
IMPR   C3   C2   C4   O83
PATCHING FIRS NONE LAST NONE

IC C3   C2   C1   C10   1.503  109.94  -57.00  111.36  1.541
IC H22  C2   C1   C10   1.115  112.03   68.32  111.36  1.541
IC H21  C2   C1   C10   1.116  112.03  177.74  111.36  1.541
IC C3   C2   C1   H12A  1.503  109.94   68.26  111.52  1.116
IC H22  C2   C1   H12A  1.115  112.03 -166.42  111.52  1.116
IC H21  C2   C1   H12A  1.116  112.03  -57.00  111.52  1.116
IC C3   C2   C1   H11A  1.503  109.94  177.71  111.47  1.114
IC H22  C2   C1   H11A  1.115  112.03  -56.97  111.47  1.114
IC H21  C2   C1   H11A  1.116  112.03   52.45  111.47  1.114
IC C5   C10  C1   C2    1.488  114.22   32.30  111.36  1.513
IC C9   C10  C1   C2    1.368  124.60 -148.78  111.36  1.513
```

IC	C5	C10	C1	H12A	1.488	114.22	-92.97	111.50	1.116
IC	C9	C10	C1	H12A	1.368	124.60	85.96	111.50	1.116
IC	C5	C10	C1	H11A	1.488	114.22	157.55	111.56	1.114
IC	C9	C10	C1	H11A	1.368	124.60	-23.53	111.56	1.114
IC	C4	C3	C2	C1	1.437	113.43	53.58	109.94	1.513
IC	O83	C3	C2	C1	1.228	123.81	-129.83	109.94	1.513
IC	C4	C3	C2	H22	1.437	113.43	-71.71	112.08	1.115
IC	O83	C3	C2	H22	1.228	123.81	104.88	112.08	1.115
IC	C4	C3	C2	H21	1.437	113.43	178.83	112.04	1.116
IC	O83	C3	C2	H21	1.228	123.81	-4.57	112.04	1.116
IC	C5	C4	C3	C2	1.361	121.62	-24.82	113.43	1.503
IC	H4	C4	C3	C2	1.102	121.97	155.22	113.43	1.503
IC	C5	C4	C3	O83	1.361	121.62	158.55	122.67	1.228
IC	H4	C4	C3	O83	1.102	121.97	-21.42	122.67	1.228
IC	C6	C5	C4	C3	1.520	118.56	177.82	121.62	1.437
IC	C10	C5	C4	C3	1.488	122.84	-0.95	121.62	1.437
IC	C6	C5	C4	H4	1.520	118.56	-2.21	116.41	1.102
IC	C10	C5	C4	H4	1.488	122.84	179.02	116.41	1.102
IC	C7	C6	C5	C4	1.511	111.32	153.15	118.56	1.361
IC	H62	C6	C5	C4	1.114	111.51	-81.62	118.56	1.361
IC	H61	C6	C5	C4	1.115	111.47	27.91	118.56	1.361
IC	C7	C6	C5	C10	1.511	111.32	-28.03	118.59	1.488
IC	H62	C6	C5	C10	1.114	111.51	97.20	118.59	1.488
IC	H61	C6	C5	C10	1.115	111.47	-153.27	118.59	1.488
IC	C1	C10	C5	C4	1.541	114.22	-3.19	122.84	1.361
IC	C9	C10	C5	C4	1.368	121.17	177.84	122.84	1.361
IC	C1	C10	C5	C6	1.541	114.22	178.04	118.59	1.520
IC	C9	C10	C5	C6	1.368	121.17	-0.93	118.59	1.520
IC	C8	C7	C6	C5	1.540	109.11	57.22	111.32	1.520
IC	H72	C7	C6	C5	1.114	112.34	-177.52	111.32	1.520
IC	H71	C7	C6	C5	1.116	112.34	-68.03	111.32	1.520
IC	C8	C7	C6	H62	1.540	109.11	-68.00	111.51	1.114
IC	H72	C7	C6	H62	1.114	112.34	57.26	111.51	1.114
IC	H71	C7	C6	H62	1.116	112.34	166.75	111.51	1.114
IC	C8	C7	C6	H61	1.540	109.11	-177.58	111.54	1.115
IC	H72	C7	C6	H61	1.114	112.34	-52.32	111.54	1.115
IC	H71	C7	C6	H61	1.116	112.34	57.17	111.54	1.115
IC	C9	C8	C7	C6	1.543	110.21	-58.57	109.11	1.511
IC	C14	C8	C7	C6	1.499	110.46	179.98	109.11	1.511
IC	H8	C8	C7	C6	1.114	108.41	60.75	109.11	1.511
IC	C9	C8	C7	H72	1.543	110.21	176.17	112.34	1.114
IC	C14	C8	C7	H72	1.499	110.46	54.72	112.34	1.114
IC	H8	C8	C7	H72	1.114	108.41	-64.51	112.34	1.114
IC	C9	C8	C7	H71	1.543	110.21	66.68	112.33	1.116
IC	C14	C8	C7	H71	1.499	110.46	-54.77	112.33	1.116
IC	H8	C8	C7	H71	1.114	108.41	-174.00	112.33	1.116
IC	C10	C9	C8	C7	1.368	120.13	30.84	110.21	1.540
IC	C11	C9	C8	C7	1.464	116.19	-151.21	110.21	1.540
IC	C10	C9	C8	C14	1.368	120.13	152.71	109.75	1.499
IC	C11	C9	C8	C14	1.464	116.19	-29.34	109.75	1.499
IC	C10	C9	C8	H8	1.368	120.13	-88.07	109.08	1.114
IC	C11	C9	C8	H8	1.464	116.19	89.88	109.08	1.114
IC	C13	C14	C8	C7	1.533	112.52	179.75	110.46	1.540
IC	C15	C14	C8	C7	1.541	116.91	-60.07	110.46	1.540
IC	H14	C14	C8	C7	1.115	100.31	58.13	110.46	1.540
IC	C13	C14	C8	C9	1.533	112.52	58.03	109.75	1.543
IC	C15	C14	C8	C9	1.541	116.91	178.21	109.75	1.543

IC	H14	C14	C8	C9	1.115	100.31	-63.59	109.75	1.543
IC	C13	C14	C8	H8	1.533	112.52	-61.30	108.89	1.114
IC	C15	C14	C8	H8	1.541	116.91	58.87	108.89	1.114
IC	H14	C14	C8	H8	1.115	100.31	177.07	108.89	1.114
IC	C1	C10	C9	C8	1.541	124.60	-179.81	120.13	1.543
IC	C5	C10	C9	C8	1.488	121.17	-0.95	120.13	1.543
IC	C1	C10	C9	C11	1.541	124.60	2.40	123.64	1.464
IC	C5	C10	C9	C11	1.488	121.17	-178.74	123.64	1.464
IC	C12	C11	C9	C8	1.347	123.95	0.90	116.19	1.543
IC	H11	C11	C9	C8	1.103	122.48	-179.16	116.19	1.543
IC	C12	C11	C9	C10	1.347	123.95	178.77	123.64	1.368
IC	H11	C11	C9	C10	1.103	122.48	-1.29	123.64	1.368
IC	C13	C12	C11	C9	1.519	121.36	-0.01	123.95	1.464
IC	H12	C12	C11	C9	1.103	113.18	179.97	123.95	1.464
IC	C13	C12	C11	H11	1.519	121.36	-179.96	113.57	1.103
IC	H12	C12	C11	H11	1.103	113.18	0.02	113.57	1.103
IC	C14	C13	C12	C11	1.533	108.34	26.71	121.36	1.347
IC	C17	C13	C12	C11	1.532	119.93	138.48	121.36	1.347
IC	C18	C13	C12	C11	1.548	107.06	-94.29	121.36	1.347
IC	C14	C13	C12	H12	1.533	108.34	-153.26	125.46	1.103
IC	C17	C13	C12	H12	1.532	119.93	-41.49	125.46	1.103
IC	C18	C13	C12	H12	1.548	107.06	85.74	125.46	1.103
IC	C8	C14	C13	C12	1.499	112.52	-56.31	108.34	1.519
IC	C15	C14	C13	C12	1.541	103.93	176.27	108.34	1.519
IC	H14	C14	C13	C12	1.115	114.08	57.11	108.34	1.519
IC	C8	C14	C13	C17	1.499	112.52	178.16	98.52	1.532
IC	C15	C14	C13	C17	1.541	103.93	50.74	98.52	1.532
IC	H14	C14	C13	C17	1.115	114.08	-68.42	98.52	1.532
IC	C8	C14	C13	C18	1.499	112.52	61.56	112.03	1.548
IC	C15	C14	C13	C18	1.541	103.93	-65.85	112.03	1.548
IC	H14	C14	C13	C18	1.115	114.08	174.98	112.03	1.548
IC	C16	C17	C13	C12	1.555	100.56	-164.77	119.93	1.519
IC	C27	C17	C13	C12	1.506	111.99	-45.48	119.93	1.519
IC	O97	C17	C13	C12	1.429	113.87	80.28	119.93	1.519
IC	C16	C17	C13	C14	1.555	100.56	-47.81	98.52	1.533
IC	C27	C17	C13	C14	1.506	111.99	71.48	98.52	1.533
IC	O97	C17	C13	C14	1.429	113.87	-162.76	98.52	1.533
IC	C16	C17	C13	C18	1.555	100.56	69.74	110.79	1.548
IC	C27	C17	C13	C18	1.506	111.99	-170.97	110.79	1.548
IC	O97	C17	C13	C18	1.429	113.87	-45.21	110.79	1.548
IC	H183	C18	C13	C12	1.115	107.01	179.96	107.06	1.519
IC	H182	C18	C13	C12	1.114	113.11	-54.70	107.06	1.519
IC	H181	C18	C13	C12	1.115	113.09	54.75	107.06	1.519
IC	H183	C18	C13	C14	1.115	107.01	61.33	112.03	1.533
IC	H182	C18	C13	C14	1.114	113.11	-173.34	112.03	1.533
IC	H181	C18	C13	C14	1.115	113.09	-63.89	112.03	1.533
IC	H183	C18	C13	C17	1.115	107.01	-47.61	110.79	1.532
IC	H182	C18	C13	C17	1.114	113.11	77.73	110.79	1.532
IC	H181	C18	C13	C17	1.115	113.09	-172.82	110.79	1.532
IC	C16	C15	C14	C8	1.532	101.96	-156.56	116.91	1.499
IC	H152	C15	C14	C8	1.114	115.07	-31.31	116.91	1.499
IC	H151	C15	C14	C8	1.114	115.06	78.21	116.91	1.499
IC	C16	C15	C14	C13	1.532	101.96	-31.92	103.93	1.533
IC	H152	C15	C14	C13	1.114	115.07	93.33	103.93	1.533
IC	H151	C15	C14	C13	1.114	115.06	-157.14	103.93	1.533
IC	C16	C15	C14	H14	1.532	101.96	90.33	109.49	1.115
IC	H152	C15	C14	H14	1.114	115.07	-144.42	109.49	1.115

```
IC H151 C15  C14  H14   1.114  115.06  -34.89  109.49  1.115
IC C17  C16  C15  C14   1.555  107.62    1.15  101.96  1.541
IC H162 C16  C15  C14   1.114  112.87  126.39  101.96  1.541
IC H161 C16  C15  C14   1.115  112.86 -124.07  101.96  1.541
IC C17  C16  C15  H152  1.555  107.62 -124.11  115.06  1.114
IC H162 C16  C15  H152  1.114  112.87    1.14  115.06  1.114
IC H161 C16  C15  H152  1.115  112.86  110.68  115.06  1.114
IC C17  C16  C15  H151  1.555  107.62  126.38  115.05  1.114
IC H162 C16  C15  H151  1.114  112.87 -108.37  115.05  1.114
IC H161 C16  C15  H151  1.115  112.86    1.17  115.05  1.114
IC C13  C17  C16  C15   1.532  100.56   29.61  107.62  1.532
IC C27  C17  C16  C15   1.506  112.17  -89.55  107.62  1.532
IC O97  C17  C16  C15   1.429  107.76  149.08  107.62  1.532
IC C13  C17  C16  H162  1.532  100.56  -95.62  112.90  1.114
IC C27  C17  C16  H162  1.506  112.17  145.22  112.90  1.114
IC O97  C17  C16  H162  1.429  107.76   23.85  112.90  1.114
IC C13  C17  C16  H161  1.532  100.56  154.82  112.88  1.115
IC C27  C17  C16  H161  1.506  112.17   35.66  112.88  1.115
IC O97  C17  C16  H161  1.429  107.76  -85.71  112.88  1.115
IC H273 C27  C17  C13   1.115  112.05 -180.00  111.99  1.532
IC H272 C27  C17  C13   1.116  111.29  -54.73  111.99  1.532
IC H271 C27  C17  C13   1.115  111.32   54.72  111.99  1.532
IC H273 C27  C17  C16   1.115  112.05  -67.79  112.17  1.555
IC H272 C27  C17  C16   1.116  111.29   57.48  112.17  1.555
IC H271 C27  C17  C16   1.115  111.32  166.94  112.17  1.555
IC H273 C27  C17  O97   1.115  112.05   52.22  110.12  1.429
IC H272 C27  C17  O97   1.116  111.29  177.48  110.12  1.429
IC H271 C27  C17  O97   1.115  111.32  -73.06  110.12  1.429
IC H97  O97  C17  C13   0.950  113.87 -180.00  113.87  1.532
IC H97  O97  C17  C16   0.950  113.87   69.38  107.76  1.555
IC H97  O97  C17  C27   0.950  113.87  -53.26  110.12  1.506
END
```

Anexo III: Lista dos Aminoácidos Padrões

Os sítios ácidos e básicos são mostrados na forma majoritária encontrada no pH fisiológico de 7,4.

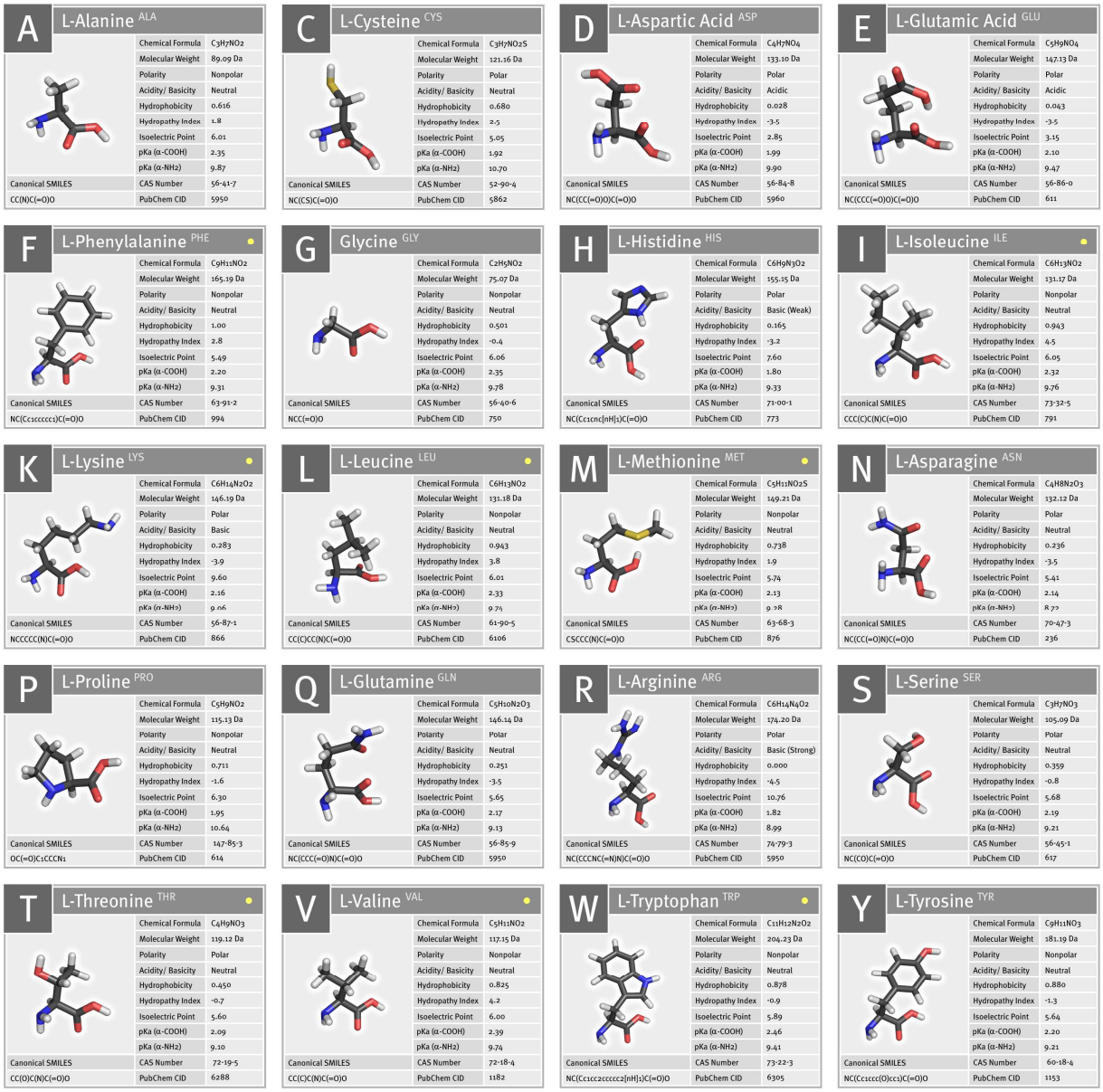


Figura 31: Aminoácidos essenciais estão indicados com círculos amarelos. A polaridade, basicidade e acidez refere-se à cadeia lateral. (retirado de http://www.unc.edu/~bzafer/aminoacids). Acessado em 17 de outubro de 2012.