

**INSTITUTO DE FÍSICA DE SÃO CARLOS  
UNIVERSIDADE DE SÃO CARLOS**

LEONARDO HENRIQUE FRANÇA DE LIMA

**Estudos por modelagem e dinâmica molecular integradas a técnicas físicas para biomoléculas em solução - interação de receptores nucleares a elementos responsivos no DNA e dinâmica inter-domínios da celobiohidrolase I**

São Carlos

2011



LEONARDO HENRIQUE FRANÇA DE LIMA

**Estudos por modelagem e dinâmica molecular integradas a técnicas físicas para biomoléculas em solução - interação de receptores nucleares a elementos responsivos no DNA e dinâmica inter-domínios da cellobiohidrolase I**

Tese apresentada ao Programa de Pós-graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para a obtenção do título de Doutor em Ciência.

Área de Concentração: Física Aplicada  
Opção: Versão Original  
Orientador: Prof. Dr. Leandro Martinez  
Co-orientador:

Versão Original

São Carlos

2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Lima, Leonardo Henrique França de.

Estudos por modelagem e dinâmica molecular integradas a técnicas físicas para biomoléculas em solução - interação de receptores nucleares a elementos responsivos no DNA e dinâmica inter-dominios da cellobiohidrolase I./ Leonardo Henrique França de Lima; orientador Leandro Martinez. -- São Carlos, 2011.  
233p.

Tese (Doutorado em Ciência - Área de concentração: Física Aplicada – Opção Biomolecular) – Instituto de Física de São Carlos da Universidade de São Paulo, 2011.

1. Dinâmica molecular. 2. Anisotropia de fluorescência. 3. Espalhamento de raios-x a baixos ângulos. 4. Receptores nucleares. 5. Receptor do ácido 9-cis retinóico. I. Estudos por modelagem e dinâmica molecular integradas a técnicas físicas para biomoléculas em solução - interação de receptores nucleares a elementos responsivos no DNA e dinâmica inter-dominios da cellobiohidrolase I.

Lado dedicado à folha de aprovação. Apagar isso antes de imprimir a versão oficial



Dedico este trabalho aos meus pais Arcanjo Jacinto de Lima e Maria Jose França Magalhães de Lima.

# AGRADECIMENTOS

- Primeiro agradeço a Deus, pois sem sua benção e misericórdia jamais teria chegado até aqui.
- Aos meus pais Arcanjo e Maria José, pelo constante apoio em todas as fases da minha vida, me fazendo sentir confortável até nos momentos mais difíceis da minha vida, mas sempre me estimulando a seguir em frente .
- Ao Prof. Dr. Leandro Martinez pelo companheirismo, solidariedade e dedicação para meu desenvolvimento acadêmico e humano.
- Aos meus irmãos Giovanni, Giselle e Rafael pelo companheirismo, auxílio e amparo nas decisões. Um agradecimento especial ao meu irmão Rafael, sem cuja o estímulo e auxílio este trabalho literalmente não estaria aqui.
- Aos demais familiares pelo apóio. A minha vó Silvia, tio Eurico e tia Marcia.
- Aos colegas do grupo de biotecnologia estrutural molecular e estrutural do Instituto de Física de São Carlos pelo bom ambiente de trabalho e coleguismo sempre.
- Aos amigos “mais chegados” do IFSC: Fernando Bachega (“Goirdo”), Lucas Bleincher (“Otro goirdo”), Napoleão Valadares, Livia Salum, Tavin, Mário, Vivi, Ana, e outros tantos que não acabariam em uma única folha.
- A todos que acreditaram e me fizeram acreditar muito obrigado.
- A todos que duvidaram... Meu igual muito obrigado, eu com certeza não estaria aqui se não fosse por vocês.
- Agradeço a CAPES pelo auxílio financeiro para desenvolvimento do meu doutorado.

- A todos que direto ou indiretamente colaboraram para que esse trabalho se tornasse possível meus mais sinceros agradecimentos.



*“(...)Onde houver erro, que eu leve a verdade (...) onde houver trevas, que eu leve a luz...”*

— FRANCISCO DE ASSIS

*“A ciência é feita de fatos como uma casa é feita de tijolos, mas um acúmulo de fatos não é mais ciência do que um monte de tijolos é uma casa.”*

— HENRI POINCARÉ



## RESUMO

LIMA, L. H. F. Estudos por modelagem e dinâmica molecular integradas a técnicas físicas para biomoléculas em solução - interação de receptores nucleares a elementos responsivos no DNA e dinâmica inter-domínios da celobiohidrolase I. 2011. 232p. Tese (Doutorado em Ciência)-Instituto de Física de São Carlos, Universidade de São Paulo, 2011.

Movimentos coletivos prestam um papel fundamental na dinâmica e energética de biomoléculas em solução. Estes movimentos permitem o acoplamento de regiões significativamente distantes, apresentando considerável influência, por exemplo, no alosterismo para a formação de complexos macromoleculares e no funcionamento integrado de proteínas multidomínios como “máquinas moleculares”. Neste trabalho de doutoramento, serão apresentados os resultados referentes à aplicação conjunta de técnicas experimentais biofísicas, de modelagem estrutural e de dinâmica molecular no estudo de dois sistemas para os quais estes movimentos coletivos demonstram considerável importância funcional. Em primeiro, nós estudamos a interação do receptor nuclear do ácido 9-cis-retinóico com seu elemento responsável específico no DNA (HRE) utilizando uma combinação de estudos de dinâmica molecular e de ensaios de afinidade através da técnica de anisotropia de fluorescência. Os resultados sugerem que diferenças nos movimentos coletivos conferem a baixa afinidade para a associação dos monômeros e a alta colaboratividade na associação do dímero ao HRE. A baixa afinidade monomérica é mais proeminente para o monômero 5'. Isto ocorre tanto devido a um *empilhamento* naturalmente menos efetivo no último passo de pares de bases no meio-sítio 5', quanto ao fasseamento dos dois meios-sítios de ligação na topologia do DNA, que impõem modos coletivos de movimento que tendem a ocultar o sítio de ligação 5'. Este perfil, por sua vez, é condizente com a bem conhecida polaridade 3' e com a menor especificidade de ligação à seqüência ao meio-sítio 5' para o monômero do hRXR $\alpha$ . Este mesmo padrão impõe um mecanismo “chave e fechadura” dependente da interação do dímero completo. Em segundo, um estudo integrado de espalhamento de raios X a baixos ângulos e modelagem estrutural baseada em dinâmica molecular foi realizado buscando compreender os movimentos interdomínios da celobiohidrolase I de *Trichoderma harzianum*. Este estudo permitiu a elaboração de um modelo estrutural de maior resolução para esta enzima de alto potencial biotecnológico. Também foram constatados aparentes mecanismos moleculares a partir dos quais as glicosilações no peptídeo conector impõem restrições à orientação e modos vibracionais para a enzima completa, condizentes com um mecanismo de deslize sobre a superfície da celulose. Este mesmo mecanismo é fundamental para a processividade da enzima na hidrólise da celulose microcristalina.

Palavras-chave: Dinâmica molecular. Anisotropia de fluorescência. Espalhamento de raios-X a baixos ângulos. Receptores Nucleares. Receptor do ácido 9-cis-retinóico. Celobiohidrolase I. *Trichoderma harzianum*.

## ABSTRACT

LIMA, L. H. F. Integrated experimental biophysics and molecular dynamics simulations of biomolecules in solution - the interaction of nuclear receptors with DNA response elements and the inter-domain dynamics of Cellobiohydrolase I. 232p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, 2011.

Collective motions play a fundamental role in solution biomolecule dynamics and energetics. These movements can couple very distant regions in the protein structures affection, for instance, allosteric mechanisms, the establishment of macromolecular complexes, and on the integrated function of multidomain proteins as “molecular machines”. In this thesis, we present results concerning to the joint use of experimental biophysical techniques, structural modeling and molecular dynamics simulations on the study of two systems for which these collective motions have substantial importance. First, we study the interaction of the nuclear retinoid X receptor with its specific DNA hormone response element (HRE) using a combination of molecular dynamics simulations and affinity assays performed by using fluorescence anisotropy. We find out that collective motions mediate the low binding affinity of monomers and the high cooperative binding of HRE dimers. The lower binding affinity of the monomer is more prominent for 5' monomers. This occur due to an natural ineffective stacking of the last base pair step at the 5'-half-site and to the phasing of the two binding half-sites in the DNA topology, that impose a collective motions that tends to occlude the 5' binding site. This behavior, in turn, is concurrent with the well known 3' polarity and the decreased binding specificity to the 5' half site for the hRXR $\alpha$  monomer. This same pattern impose a lock-and-key mechanisms dependent on the binding of the full dimer. Second, an integrated Small angle X ray scattering and molecular dynamics based structural modeling was used to comprehend the interdomain motions of cellobiohydrolase I of *Trichoderma harziannum*. We manage to build a refined model for this enzime, with important biotechnological potential. We also provide insights into molecular mechanisms of linker and glycosylation imposed restraints on the orientation and vibrational modes of the full-length enzyme, supporting a mechanism of sliding of on the cellulose surface. This mechanism is fundamental for the high processivity on the hydrolysis of microcrystalline cellulose.

Keywords: Molecular dynamics. Fluorescence anisotropy. Small angle X-ray scattering. Nuclear receptors. Retinoid X receptor. Cellobiohydrolase I. *Trichoderma harziannum*.



# ***LISTA DE FIGURAS***

Figura 1.1 - Representações esquemáticas dos elipsóides termicos da leucina L32 da ferricitocromo C, conforme recuperado em simulações de dinâmica molecular nos tempos de <i>a</i> 32 ps e <i>b</i> 0.2 ps. Retirado de (1) . . . . .	34
Figura 1.2 - Direções dos deslocamentos atômicos decorrentes de dois modos vibrationais em baixas freqüências para o inibidor de tripsina pancreático conforme obtido por análise de modos normais (retirado de (2)) . . . . .	36
Figura 1.3 - Dogma central da biologia: A síntese de proteínas tem basicamente três etapas: a transcrição, processamento intermediário e tradução. Na transcrição uma fita de RNA <sub>m</sub> produzida a partir do DNA. No processamento, enquanto a molécula de pré-mRNA é transportada para o citosol, ela sofre os <i>splicings</i> necessários para a tradução final. Na tradução a seqüência de nucleotídeos do RNA <sub>m</sub> é finalmente lida pela maquinária provida pelos complexos de tRNAs, codificando a seqüência de aminoácidos da proteína.(retirado de (3)) . . . . .	49
Figura 1.4 - Representação esquemática da estrutura modular de um receptor nuclear.	51
Figura 1.5 - Superposição das estruturas cristalográficas do cerne dos DBDs de RXR (vermelho), RAR (ciano), e TR (azul), retiradas de suas respectivas estruturas cristalográficas associadas ao DNA (4–6). Cada um dos dois zincos nos dois dedos de zinco é mostrado em laranja e coordenado por quatro cisteínas em stick em amarelo. As hélices A e B e os dois dedos de zinco são rotulados. (retirado de (7)). . . . .	54

Figura 1.6 - Dados de SAXS da literatura para a celobiohidrolase I de <i>Trichoderma harziannum</i> . . . . .	56
Figura 4.1 - Isotermas de interação do hRXR $\alpha$ -ΔAB a diferentes elementos responsivos de DNA . . . . .	87
Figura 4.2 - Características anisotrópicas e parâmetros de ajuste numérico para o complexo de hRXR $\alpha$ -ΔAB a diferentes HREs . . . . .	90
Figura 4.3 - Experimentos de competição para a interação do hRXR $\alpha$ -ΔAB a diferentes HREs . . . . .	96
Figura 4.4 - Principais parâmetros e informações provenientes da análise dos experimentos de competição . . . . .	97
Figura 4.5 - Ajuste para os termos divergentes das interações inespecíficas . . . . .	101
Figura 4.6 - Curvas Estequiométricas para a interação do hRXR $\alpha$ -ΔAB com diferentes DR1, DR4 e F2 . . . . .	106
Figura 4.7 - Correção do termo inespecífico “comum” das isotermas para os HREs anticooperativos . . . . .	111
Figura 4.8 - Correção do termo inespecífico “comum” para as isotermas do tipo I <sub>1</sub> . .	113
Figura 4.9 - Parâmetros de energia livre recuperados através do ajuste numérico global pelo BIOEQS e comparação com o ajuste de Hill . . . . .	120
Figura 4.10 - Estruturas iniciais para as simulações de monômero DBD de hRXR $\alpha$ em solução . . . . .	127
Figura 4.11 - Progressão do Equilíbrio para as simulações de hRXR $\alpha$ -DBD para diferentes configurações iniciais em solução. . . . .	129
Figura 4.12 - Progressões das energias de hidratação para as trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais.	132

Figura 4.13 - Análise de componentes principais para as trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais . . . . .	136
Figura 4.14 - Distinções para a mobilidade do cerne globular do domínio DBD do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais. . . . .	140
Figura 4.15 - Projeções dos autovetores recuperados na análise de componentes principais para a trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir da configuração inicial RXR <sub>RMN</sub> <sup>I</sup> . . . . .	143
Figura 4.16 - Projeções das componentes principais para a trajetória do monomero DBD na conformação RXR <sub>RMN</sub> <sup>XX</sup> . . . . .	145
Figura 4.17 - Projeções das componentes principais para a trajetória do monomero DBD na conformação RXR <sub>crys</sub> . . . . .	146
Figura 4.18 - Diferenças de fases entre os dois meio sitios para diferentes HREs . . . . .	154
Figura 4.19 - Progressões do equilíbrio para os HREs DR1 e DR4 de forma representativa para todos os quatro HREs analisados neste estudo . . . . .	156
Figura 4.20 - Componentes principais para os HREs DR1 e DR4. . . . .	158
Figura 4.21 - Menor empacotamento no passo C-A 3' terminal do primeiro meio-sítio confere perturbações no sulco maior 5' para DR1 e DR4 . . . . .	161
Figura 4.22 - Progressão da trajetórias de equilíbrio para um único monômero associado a DR1. A e B - trajetória cA:DR1-A. C e D - trajetória cA:DR1-B. E e F - trajetória cB:DR1. . . . .	167
Figura 4.23 - Movimento vibracional do DNA e competição com os íons em solução tende a deslocar o T-box 3' na ausência do monômero 5'. . . . .	169

Figura 4.24 - Análise de componentes principais para os complexos de monômero a DR1. Esquerda para a direita: Componentes 1-3. A- trajetória CA:DR1-A; B- trajetória cA:DR1-B; C - trajetória cB:DR1. . . . .	170
Figura 4.25 - Variações na largura do sulco menor ao longo da cadeia de DR1 nas componentes principais para as trajetórias com o monômero individualmente ligado a cada sítio. De cima para baixo: Componentes 1 a 3. A, C e D: Componentes para cA:DR1-B. B, D e E: Componentes para cB:DR1. Direção 5'-3' a partir da origem no eixo das ordenadas. . . . .	172
Figura 4.26 - Progressão do equilíbrio para as trajetórias do homodímero de hRXR $\alpha$ associado a DR1. <i>Topo</i> : trajetória S-RXRd:DR1 <sub>c</sub> ; <i>Fundo</i> : trajetória S-RXR:DR1 <sub>l</sub> . . . . .	177
Figura 4.27 - Estimativa para o balanço das componentes entálpicas entre as espécies de hRXR $\alpha$ -DBD e DR1 combinadas com diferentes graus de complexidade, conforme recuperado pelas simulações de Dinâmica Molecular. . . . .	179
Figura 4.28 - Análise das componentes principais para o homodímero de hRXR $\alpha$ -DBD associado a DR1 na trajetória S-RXRd:DR1 <sub>c</sub> . . . . .	182
Figura 4.29 - Análise das componentes principais e flexão do eixo maior do DNA com maior ajuste superficial para o homodímero de hRXR $\alpha$ -DBD associado a DR1 na trajetória S-RXRd:DR1 <sub>l</sub> . . . . .	183
Figura 4.30 - Superfícies de correlação entre as diferentes energias não ligadas envolvidas na manutenção do complexo homodimérico em DR1 . . . . .	186
Figura 4.31 - Perfil simétrico em relação ao centro de massas das variações nas larguras do sulco menor descritas pelas principais componentes da simulação S-RXR:DR1 <sub>l</sub> . . . . .	188
Figura 4.32 - Resíduos com alta informação mútua apresentam papel crucial na estabilização vibracional do complexo homodimérico em DR1. . . . .	190

Figura 4.33 - Empacotamento entre o T-box 3' e o D-box 5' estabiliza o complexo homodimérico em DR1 próximo a seu centro de massas. . . . . 192

Figura 4.34 - Trajetória para o homodímero hRXR $\alpha$ -DBD em DR1 na configuração cristalográfica original sugere uma significativa menor estabilização do complexo na ausência do empacotamento entre o T-box 3' e o D-box 5' devido à instabilidade do sulco menor no passo C-A na extremidade 3' do primeiro meio-sítio. . . . . 193

Figura 4.35 - Aparente direcionamento para as interações entre os LBDs nos padrões de simetria da estrutura completa cristalográfica pela componente principal recuperada para a trajetória “estável” S-RXR:DR1<sub>l</sub>. As estruturas foram geradas sobrepondo-se as cadeias terminais do domínio D da cadeia A da estrutura 3DZU e os T-box de *instantâneos* com ordem crescente de projeção sobre a componente 1 desta trajetória. Na parte inferior da figura, a estrutura completa é ilustrada para fins de comparação. . 197

Figura 5.1 - Sobreposição dos modelos estruturais completos para a ThCel7A obtidos por modelagem computacional. O modelo estrutural obtido pelo I-Tasser utilizando as restrições topológicas do modelo DAM obtido pelo tratamento dos dados de SAXS é comparado ao mesmo na figura. Nota-se que a estrutura recuperada apresenta um peptídeo conector (o *linker*) semi-extendido. As conformações I a V, representativas dos *decoys* recuperados a partir das simulações de MD são mostradas. Os valores de RMSD médios computados em relação ao modelo inicialmente obtido pelo I-TASSER para cada *decoy* são mostrados. . . . . 200

Figura 5.3 - Análise energética das conformações extraídas das simulações de MD.  
a- Soma das energias não ligadas para o LP e “vizinhanças”; b- energia de hidratação do LP; c-Interações proteína-proteína para o LP e vizinhanças; d- interações não ligadas envolvendo as extensões O-glicosídicas.205

Figura 5.4 - Refinamento do modelo estrutural utilizando anelamento simulado por dinâmica molecular. *A*- Modelo dos átomos *Dummy* (DAM) obtido a partir dos dados experimentais. *B*- Mapa volumétrico promediado sobre o tempo e pesado pelas massas para os átomos de todas as estruturas da simulação com uma *grid* de resolução de 5 Å, sobreposto aos cinco confôrmeros médios recuperados pelo SA/MD. Nota-se a representatividade difusa para as O-glicosilações. *C* Reprodução do padrão de espalhamento de raios X a baixos ângulos calculado para um modelo estrutural a partir de uma média ponderada para as conformações de I a III. *D* Sobreposição do melhor modelo de alta resolução recuperado ao envelope do DAM.

Figura 5.7 - A - Contribuição das entropias conformacionais à energia livre do LP e da proteína inteira nos diferentes confôrmeros, conforme recuperado por uma análise quasiharmônica (8). B - Capacidades caloríficas recuperadas pela mesma análise quasiharmônica. A conformação V apresenta uma relativa baixa entropia conformacional e pequena entropia para o peptídeo conector, refletindo sua maior rigidez inter-domínios. . . . . 219



## *Lista de Tabelas*



# **LISTA DE ABREVIATURAS**

<i>DNA</i>	Ácido desoxirribonucleico
<i>RNA</i>	Ácido ribonucléico
<i>mRNA</i>	RNA mensageiro
<i>rRNA</i>	RNA ribossômico
<i>tRNA</i>	RNA transportador
<i>CNTP</i>	Condições normais de temperatura e pressão
<i>MD</i>	Dinâmica molecular
<i>SA/MD</i>	Anelamento simulado por dinâmica molecular
<i>RNAP</i>	RNA polimerase
<i>NR</i>	Receptor Nuclear
<i>LBD</i>	Domínio de ligação com o ligante
<i>DBD</i>	Domínio de ligação com o DNA
<i>HRE</i>	Elemento responsivo no DNA
<i>SAXS</i>	Espalhamento de raios-X a baixo ângulo
<i>TR</i>	Receptor do Hormônio Tireoideano
<i>H</i>	Hélice
<i>RAR</i>	Receptor do Ácido Retinóico

<i>PPAR</i>	Receptor ativado por proliferador de peroxissomo
<i>CHARMM</i>	<i>Chemistry at Harvard Molecular Mechanics</i>
<i>NAMD</i>	<i>Not Another Molecular Dynamics (package)</i>
<i>RXR</i>	Receptor X-retinóide
<i>9cRA</i>	Ácido 9-cis retinóico
<i>vdW</i>	Van der Waals
<i>RMSD</i>	Desvio quadrático médio
<i>CBH</i> –	<i>I</i> Celobiohidrolase I
<i>Cel7A</i>	Celulase 7A
<i>CBM</i>	Domínio de ligação à celulos
<i>CCD</i>	Domínio do cerne catalítico
<i>PCA</i>	Análise de componentes principais
<i>WLC</i>	<i>Worm like chain model</i>
<i>L</i>	Autovalor
<i>L/Ltot</i>	Relação autovalor/autovalores totais



# **SUMÁRIO**

<b>1</b>	<b>Introdução</b>	<b>33</b>
1.1	Movimentos coletivos de macromoléculas em solução e sua importância para processos biológicos . . . . .	33
1.2	O estudo do comportamento de macromoléculas em solução através de simulações computacionais por dinâmica molecular . . . . .	37
1.2.1	Porque é possível descrever sistemas moleculares em termos clássicos - a aproximação de Born-Oppenheimer e a localização do pacote de ondas para núcleos termalizados . . . . .	38
1.2.2	Aproximações nos Campos de Forças comumente utilizados . . . . .	44
1.2.3	Resolvendo as trajetórias - Integração das equações de movimento . . . . .	47
1.3	Receptores Nucleares e seu papel na transcrição gênica . . . . .	48
1.3.1	A transcrição gênica e os fatores de transcrição . . . . .	48
1.3.2	Receptores Nucleares . . . . .	50
1.3.3	A interação de receptores nucleares a elementos responsivos . . . . .	52
1.3.4	O receptor do ácido 9- <i>cis</i> -retinóico e seu papel como parceiro “promíscuo”	53
1.4	Celulases e sua importância em biotecnologia . . . . .	54
1.5	A técnica de anisotropia de fluorescência e sua aplicabilidade para estudos de formação de complexos macromoleculares em solução . . . . .	56
1.6	Espalhamento de Raios-X a baixos ângulos - SAXS . . . . .	58

<b>2</b>	<b>Objetivos</b>	<b>61</b>
<b>3</b>	<b>Materiais e Métodos</b>	<b>63</b>
3.1	Metodologia Experimental para os estudos de afinidade do hRXR $\alpha$ - $\Delta$ AB a diferentes elementos responsivos no DNA . . . . .	63
3.1.1	Reagentes . . . . .	63
3.1.2	Expressão e purificação do hRXR $\alpha$ - $\Delta$ AB . . . . .	63
3.1.3	Oligonucleotídeos sintéticos . . . . .	65
3.1.4	Interação do hRXR $\alpha$ - $\Delta$ AB com elementos responsivos de DNA . . . . .	66
3.1.5	Anisotropia de fluorescência - detalhes técnicos e considerações pré análises . . . . .	68
3.1.6	Análise termodinâmica quantitativa das isotermas de interação . . . . .	70
3.2	Metodologia dos estudos de dinâmica molecular para o alosterismo na interação do hRXR $\alpha$ -DBD a diferente elementos responsivos no DNA . . . . .	73
3.3	Metodologia para o estudo integrado por SAXS e modelagem estrutural baseada em dinâmica molecular para a celobiohidrolase I de <i>Trichoderma harzianum</i> em solução . . . . .	73
3.3.1	Obtenção da amostra . . . . .	73
3.3.2	ensaios de SAXS . . . . .	74
3.3.3	Análise dos dados de SAXS . . . . .	74
3.4	Metodologia da modelagem - criação de um modelo para a proteína completa, dinâmica molecular e comparação com os dados de SAXS . . . . .	74
3.4.1	Construção de um modelo de alta resolução através do programa I-TASSER . . . . .	74

3.4.2	Adição das glicosilações . . . . .	77
3.4.3	Metodologia da simulação de dinâmica molecular . . . . .	79
<b>4</b>	<b>Resultados e Discussões I - Bases Moleculares para o Alosterismo da Interação do Receptor Nuclear do Ácido 9-cis-Retinóico com Elementos Responsivos no DNA</b>	<b>85</b>
4.1	Estudos de Interação do Receptor de Ácido 9-cis-Retinóico a Elementos Responsivos de DNA por Anisotropia de Fluorescência . . . . .	85
4.1.1	Promiscuidade e presença de interações inespecíficas no reconhecimento de elementos responsivos pelo hRXR $\alpha$ - $\Delta$ AB . . . . .	85
4.1.2	Diferenças no alosterismo e na correlação rotacional dos complexos de hRXR $\alpha$ - $\Delta$ AB com elementos responsivos específicos e inespecíficos de DNA . . . . .	89
4.1.3	Experimentos de competição confirmam a inespecificidade das isotermas do tipo I <sub>1</sub> e indicam a existência de um intermediário e de um maior número de interações inespecíficas para o complexo com F2 . . . . .	94
4.1.4	Curvas estequiométricas comprovam a ligação dimérica do hRXR $\alpha$ - $\Delta$ AB a DR1, DR4 e F2 na ausência do agonista e sugerem, em conjunto com os experimentos de competição, a existência de diferentes caminhos de associação do dímero a F2 . . . . .	103
4.1.5	Correção das isotermas para HREs não cooperativos a partir do perfil de interação com a sequência alvo do p53 corrobora a hipótese de um intermediário monomérico para sequências inespecíficas . . . . .	110
4.1.6	Modelagem numérica global multi-equações corrobora a hipótese de que a interação do hRXR $\alpha$ - $\Delta$ AB com HREs específicos e não específicos se diferencia pela distribuição de monômeros e dímeros associados, sendo estas distribuições influenciadas pelo ligante . . . . .	117

4.2	Estudos de modelagem estrutural e dinâmica molecular das interações do receptor nuclear do ácido 9- <i>cis</i> - retinóico a Elementos Responsivos de DNA . . . . .	123
4.2.1	Considerações prévias a respeito das componentes do sistema - dinâmica conformacional do domínio DBD e HREs e suas possíveis implicações na formação do complexo . . . . .	123
4.2.2	Conhecendo as componentes do sistema (I) - Desempacotamento hidrofóbico decorrente da desestruturação do T-box na conformação de interação do hRXR $\alpha$ -DBD ao DNA induz a mudanças na dinâmica em nanosegundos e na conformação média de todo o domínio de forma possivelmente funcional . . . . .	126
4.2.3	Conhecendo as componentes do sistema (II) - Mobilidade intrínseca, blindagem eletrostática por íons e natureza dos modos normais das sequências de HREs conferem penalidades adicionais para a interação do DBD . .	151
4.2.4	Conhecendo as componentes do sistema (III) - A ligação de um único monômero distorce as componentes vibracionais da seqüência adicionando aparentes penalidades energéticas, com efeito mais notório na interação ao sítio 5' . . . . .	164
4.2.5	O complexo homodimérico do hRXR $\alpha$ a DR1 - A interação simultânea dos dois domínios DBD otimiza o acoplamento interno ao longo de todo o complexo, estabiliza as componentes vibracionais com uma maior simetria em relação ao centro de massas do sistema e apresenta componentes de movimento que parecem facilitar a correta interação entre os LBDs . . . . .	175

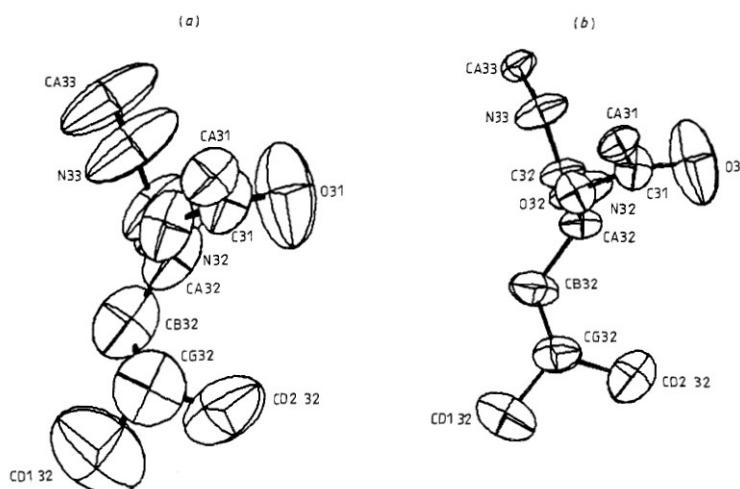
5.0.6	O modelo de alta resolução recuperado para a celobiohidrolase I apresenta o peptídeo conector dobrado em sua extensão N-terminal . . . . .	199
5.0.7	Cinco conformações principais, com diferentes extensões do peptídeo conector, foram recuperadas no procedimento de SA/MD . . . . .	201
5.0.8	Hidratação e interações intramoleculares favorecem a conformação semi-extendida recuperada pelos estudos de SAXS . . . . .	204
5.0.9	Os parâmetros médios de distância retornados pela simulação de SA/MD são condizentes com os parâmetros obtidos nos estudos de SAXS . . .	207
5.0.10	Apenas as regiões de maior ocupância ao longo do tempo para a Th-Cel7A são claramente identificadas no modelo de átomos dummy retornado pelo SAXS . . . . .	209
5.0.11	O modelo tridimensional de átomos dummy corrobora as principais interações computadas para o peptídeo conector na simulação . . . .	211
5.0.12	O peptídeo conector é mantido em um conjunto de conformações vibrationalmente ativas conforme inferido através da análise das componentes principais e quasiharmônica para os diferentes confôrmeros. . . . .	216
<b>6</b>	<b>Conclusões</b>	<b>221</b>
6.1	Conclusão a respeito dos estudos dos mecanismos moleculares para o alostéricismo de hRXR $\alpha$ a seu elemento responsivo DR1 . . . . .	221
6.2	Conclusões a respeito da modelagem estrutural e dinâmica interdomínios da Celobiohidrolase de <i>Trichoderma harzianum</i> em solução . . . . .	222
<b>REFERÊNCIAS</b>		<b>229</b>

# 1    *Introdução*

## 1.1    **Movimentos coletivos de macromoléculas em solução e sua importância para processos biológicos**

Nas últimas décadas, o avanço nas técnicas de biologia molecular estrutural têm permitido a determinação de um sem número de estruturas de macromoléculas e de complexos macromoleculares de importância biológica, permitindo um maior entendimento sobre a estrutura e função de diversos processos celulares, com impactos científicos e tecnológicos para a sociedade. Os avanços nas técnicas de cristalografia de raios X e de espectroscopia por ressonância nuclear magnética (9) , bem como de técnicas experimentais com menor resolução e ferramentas computacionais de modelagem por homologia permitiram um conhecimento cada vez mais profundo a respeito da estrutura e função de proteínas, ácidos nucléicos e outros constituintes celulares com resolução atômica. Uma informação fundamental que este “olhar cada vez mais de perto” provido por este conjunto de técnicas tem proporcionado acerca destes constituintes macromoleculares dos sistemas biológicos, em particular as proteínas, é o fato de que sua extensa diversidade e mesmo maleabilidade funcional é refletida na considerável complexidade em suas estrutura e dinâmica detalhadas (2). Contudo, a despeito da contribuição fundamental provida, por exemplo, pela estrutura tridimensional de uma molécula de proteína ou DNA resolvida em alta resolução, as coordenadas reportadas como representativas para a estrutura destas macromoléculas são, na verdade, *médias* sobre um conjunto conformacional, muitas vezes significativamente mais extenso que a informação relativamente limitada da estrutura única permite supor. Além disto, no caso de estruturas resolvidas por cristalografia de raios-X, esta mesma média conformacional é tomada sobre temperaturas consideravelmente mais baixas e em ambi-

entes físico-químicos significativamente distintos das condições fisiológicas sob as quais estas macromoléculas exercem seu extenso conjunto de funções. Mesmo para um conjunto típico entre 10-20 confôrmeros recuperados em um estudo de RMN de uma macromolécula biológica (9), informações detalhadas acerca da dinâmica e da transição entre estes confôrmeros, e sobretudo das restrições energéticas que regem estas transições são muitas vezes perdidas. De fato uma representação esquemática mais realista para o comportamento de macromoléculas biológicas em solução é dada pelo modelo clássico do *elipsóide térmico*(2), geralmente obtido para as secções cônicas dos movimentos dos átomos em uma simulação de *dinâmica molecular*<sup>1</sup> (Fig. 1.1).



**Figura 1.1** – Representações esquemáticas dos elipsóides termicos da leucina L32 da ferricitocromo C, conforme recuperado em simulações de dinâmica molecular nos tempos de *a* 32 ps e *b* 0.2 ps. Retirado de (1)

Na representação esquemática da Fig. 1.1-A e B, retirada de um estudo de 1983 (1), as respectivas secções cônicas descritas pelos movimentos térmicos de cada átomo do resíduo L32 da proteína ferricitocromo C (obtidas através de uma simulação por dinâmica molecular) são representadas considerando-se apenas a elipse na extremidade desta secção (gerando a representação gráfica a que se denomina *elipsóide térmico*) para os intervalos de tempo de 32 ps e de 0.2 ps. Nota-se que para o intervalo de tempo mais longo (sendo o tempo de 32 ps ainda extremamente curto em face aos intervalos usualmente abordados em estudos de dinâmica molecular de proteínas atualmente (10)) os eixos maiores dos elipsóides representativos dos movimentos

<sup>1</sup> Simulações por dinâmica molecular e sua aplicabilidade no estudo de macromoléculas em solução serão melhor discutidas na próxima seção.

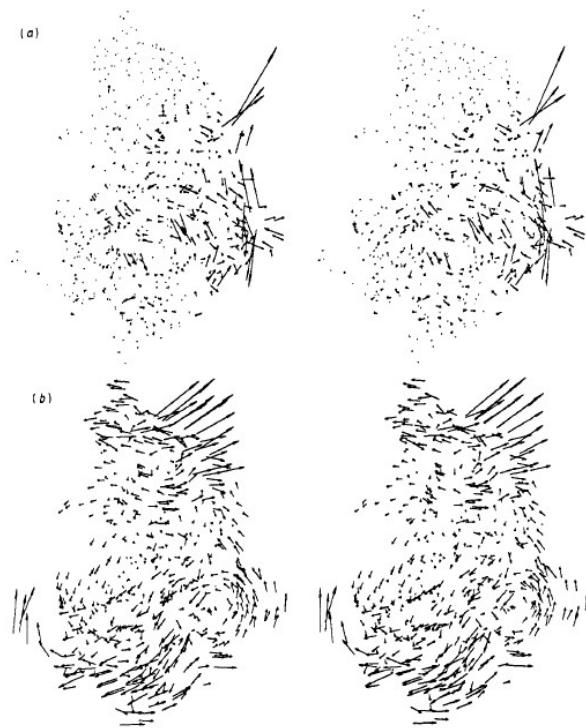
térmicos de cada átomo tendem a apresentar orientações similares. Duas informações fundamentais acerca do comportamento de proteínas (e de outras macromoléculas biológicas) em solução podem ser tiradas a partir da análise desta simulação relativamente simples. Uma, é que proteínas em solução são consideravelmente móveis, dada a abrangência significativa das áreas ocupadas pelos elipsóides nos intervalos de tempo consideravelmente curtos abordados nesta simulação. A outra, é que os movimentos térmicos dos átomos constituintes destas moléculas tendem a ser correlacionados a distâncias consideravelmente longas, sendo os movimentos locais dominados pelos *movimentos coletivos* envolvendo aglomerados substancialmente numerosos de átomos ao longo de toda a molécula.

De fato, os movimentos coletivos em proteínas e outras macromoléculas biológicas em solução envolvem aglomerados de átomos muito maiores que os contidos em um único resíduo de aminoácido como a L32 da Fig. 1.1. O conceito de *phonons* de baixa freqüência ou de movimentos internos para proteínas foi originalmente proposto por Chou e Chen (11) de forma a resolver o problema do “défice” na energia livre recuperada na formação de complexos interproteicos. Neste trabalho, estudando as energias de interação entre proteínas, tais como a insulina e seu receptor, notou-se que listando as contribuições conhecidas para as mudanças na energia do sistema, tais como entropia translacional e rotacional, reconfiguração das interações de hidrogênio, interações de van der Waals e interações hidrofóbicas, as mudanças observadas em decorrência destas interações não eram completamente explicadas. Assim, foi inferido que o défice adicional poderia ser explicado pela geração de modos vibracionais extras com números de onda significativamente baixos, na faixa entre  $10\text{-}100\text{ cm}^{-1}$ , correspondendo à faixa de freqüências de terahertz (entre  $3 \times 10^{11}$  e  $3 \times 10^{12}\text{ Hz}$ ). Subsequentemente, estes modos vibracionais a baixa freqüência foram observados por espectroscopia Raman para um número considerável de moléculas de proteína (12) e DNA (13). Estes mesmos resultados foram posteriormente confirmados através de experimentos de espalhamento de nêutrons (14). Hoje, o conceito de *modos normais* para os movimentos coletivos de macromoléculas em solução é bem fundamentado e sua correlação com a energética e com os caminhos de associação e reação para vários processos de interesse biológico, como por exemplo na formação de comple-

xos proteína-DNA e nos mecanismos de catálise enzimática e de processividade para um sem número de enzimas é bem documentado.

Uma ilustração dos efeitos dos modos normais em uma proteína é apresentada Na Fig. 1.2. Nesta figura, os vetores correspondentes aos deslocamentos de cada respectivo átomo pesado da molécula do inibidor de tripsina pancreático sob a influência de dois modos normais de baixa freqüência são apresentados. Nota-se que os movimentos de baixas freqüências e altas amplitudes apresentam contribuições dominantes para os movimentos internos da proteína, correlacionando, através destes movimentos, regiões relativamente distantes. Este acoplamento dinâmico a longas distâncias é fundamental para o funcionamento de proteínas (sobretudo) e de biomoléculas, de uma forma geral, como “máquinas moleculares”, utilizando uma definição previamente cunhada por Alberts

. Este acoplamento tem papel crucial, por exemplo, nos mecanismos de associação e de alosterismo para a formação de complexos proteicos e com seus ligantes.



**Figura 1.2** – Direções dos deslocamentos atômicos decorrentes de dois modos vibracionais em baixas freqüências para o inibidor de tripsina pancreático conforme obtido por análise de modos normais (retirado de (2))

## 1.2 O estudo do comportamento de macromoléculas em solução através de simulações computacionais por dinâmica molecular

Dinâmica molecular (*Molecular dynamics - MD*) é uma técnica de simulação computacional que descreve a evolução temporal das posições e velocidades dos átomos constituintes de um sistema (15, 16). Tem sido aplicada com sucesso em diversos problemas em física, química e biologia. Uma particularidade que a torna numericamente acessível a sistemas de dimensões consideráveis (como nos casos de biomoléculas em solução) é sua aproximação dos potenciais em termos de equações *clássicas*, significativamente menos complexas dos pontos de vista matemático e numérico que a descrição rigorosa do sistema à luz da *mecânica quântica*. Se por um lado esta aproximação torna a dinâmica de um sistema contendo milhares de átomos (como uma macromolécula ou complexo macromolecular em ambiente aquoso) numericamente acessível, ela obviamente limita a natureza dos processos que podem ser observados durante as simulações. Nenhum efeito quântico é visualizado em uma simulação de dinâmica molecular. Nenhuma ligação química é rompida, não ocorrem interações entre orbitais, efeitos de ressonância, polarização ou transferência de cargas, por exemplo. À primeira vista, a descrição do comportamento de um sistema molecular não considerando estes efeitos pode parecer inadequada. Contudo, inúmeros dados experimentais podem ser modelados sem a descrição *explícita* destes fenômenos quânticos, sendo os mesmos *implicitamente* embutidos em aproximações clássicas para os potenciais a que os átomos são submetidos (conforme será visto nas próximas subseções). Isto é particularmente verdadeiro para os casos em que se está interessado basicamente na dinâmica dos *núcleos* atômicos em sistemas no equilíbrio químico (não sendo de interesse, por exemplo, a descrição das coordenadas de uma reação química), próximo às *condições normais de temperatura e pressão* (CNTP), e para um sistema cuja dinâmica seja primariamente ditada por átomos pesados (para o caso em que os átomos hidrogenóides, se presentes, tenham a sua dinâmica fortemente atrelada a movimentos coletivos envolvendo estes átomos mais pesados). As duas últimas condições garantem que as distâncias médias entre os átomos na simulação sejam significativamente maiores que os comprimentos de onda térmicos

de de Broglie (17) para os núcleos que regem a dinâmica do sistema, podendo desta forma os mesmos serem considerados como partículas *pontuais* e sua dinâmica propagada classicamente. As propriedades físico-químicas que dependem fundamentalmente da dinâmica dos átomos de hidrogênio (que, por sua vez, fogem a esta última regra)<sup>2</sup> devem ser corrigidas quanticamente (usualmente na forma da adição de *constrições* para as configurações acessíveis pelos átomos de hidrogênio durante a simulação, podendo estas serem ou não acompanhadas de parâmetros auxiliares na descrição dos potenciais para estes átomos) de forma a que a simulação produza resultados comparáveis aos experimentais (17).

### **1.2.1 Porque é possível descrever sistemas moleculares em termos clássicos - a aproximação de Born-Oppenheimer e a localização do pacote de ondas para núcleos termalizados**

Dois arcabouços teóricos fundamentais permitem a aproximação da dinâmica dos núcleos do sistema nas condições acima descritas, *i.e.*, de forma independente da dinâmica eletrônica e com suas trajetórias descritas classicamente. São eles, respectivamente, a *aproximação de Born-Oppenheimer* (19) e a localização dos pacotes de ondas relacionados a núcleos pesados à temperatura ambiente devido à combinação de suas autofunções. Uma descrição rigorosa destes fundamentos teóricos relacionados à mecânica quântica foge ao objetivo do presente trabalho. Contudo, uma descrição consideravelmente simplificada é apresentada nesta subseção. O leitor interessado em exposições mais elaboradas sobre o tema encontrará na literatura um vasto material a respeito, como por exemplo em (19). Seguido à apresentação simplificada dos conceitos de mecânica quântica relacionados à dinâmica molecular, nas próximas subseções serão definidas as principais aproximações quanto aos potenciais que são utilizados na descrição de átomos e moléculas neste tipo de simulação. Em seguida, os principais algoritmos utilizados na execução de uma simulação serão descritos. O leitor já familiarizado com as técnicas de simulações de dinâmica molecular poderá omitir a leitura desta seção sem que se prejudique o

---

<sup>2</sup> para um núcleo de carbono a 298 K, por exemplo, o comprimento de onda térmico de de Broglie é de aproximadamente 0.3 Å, enquanto que a distância mínima entre os átomos na simulação é maior que 1 Å. Já para um núcleo de hidrogênio à mesma temperatura o comprimento de onda de de Broglie é também de aproximadamente 1 Å e, portanto, no limite crítico para a descritividade clássica do sistema, através da estatística de Maxwell-Boltzmann, sendo necessária a descrição quântica através da estatística de Fermi-Dirac para este caso (18).

entendimento deste trabalho.

*A priori*, a descrição da dinâmica de uma partícula é obtida pela resolução da equação de Schröedinger dependente do tempo:

$$\frac{-\hbar^2}{2m} \nabla^2 \Psi(\vec{x}, t) + V(\vec{x}, t) \Psi(\vec{x}, t) = i\hbar \frac{\partial \Psi(\vec{x}, t)}{\partial t} \quad (1.1)$$

Uma interpretação *qualitativa* para esta equação é que a variação com o tempo da *função de onda da partícula* ( $\Psi(\vec{x}, t)$ ) - multiplicada por um *fator de fase*  $i\hbar$  (o que é necessário visto se tratar de uma *função de onda*) no lado direito da equação - é igual ao *operador Hamiltoniano* (lado esquerdo) atuando na função de onda  $\Psi$ . A função  $\Psi$  pode ser descrita como uma *função do estado* de um dado sistema, que é reflexiva para as coordenadas espaciais de todas as partículas consideradas ( $\vec{x}$ , para o caso de um sistema com uma partícula única como na equação acima), do tempo ( $t$ ) e de algumas variáveis específicas da mecânica quântica que determinam o *spin* de cada partícula (15, 16, 19). A função de onda fornece a informação completa acerca do sistema, sendo sua resolução a meta da equação 1.1. A resolução desta função é importante porque o quadrado de sua integração sobre o intervalo espacial  $\vec{x} + \delta\vec{x}$  (sendo  $\delta\vec{x}$  um incremento *infinitesimal*) no tempo  $t$  ( $|\Psi(\vec{x}, t)|^2 \delta\vec{x}$ ) fornece a densidade de probabilidade<sup>3</sup> de se encontrar a partícula nesta região espacial em  $t$ . Em mecânica quântica, este é o máximo de informação a que se pode chegar sobre a partícula em um instante  $t$  *bem definido*: é possível apenas saber a *probabilidade* de que a partícula assuma certos valores com respeito às suas variáveis simultaneamente dependentes de posição e momento - ao contrário do que ocorre para os sistemas que podem ser resolvidos à luz da mecânica clássica, em que, a princípio, estas variáveis são precisamente definidas. O operador Hamiltoniano no lado esquerdo, por sua vez, determina o comportamento da função de onda e consiste na soma de dois outros operadores: o operador *energia cinética* ( $\frac{-\hbar^2}{2m} \nabla^2$ ) e o operador *energia potencial* ( $V(\vec{x}, t)$ ), este último também assumido como possivelmente variável em tempo e espaço, conforme pode ser visualizado pelos índices ( $\vec{x}, t$ )), ambos atuando na função  $\Psi(\vec{x}, t)$ . Desta forma, pode-se dizer que o lado esquerdo da equação 1.1 descreve as deturpações impostas à função  $\Psi(\vec{x}, t)$  devido às energias cinéticas

<sup>3</sup> Na verdade, em uma definição rigorosa, a densidade de probabilidade é obtida multiplicando-se a função  $\Psi$  por seu *conjugado complexo* ( $|\Psi^*(\vec{x}, t)\Psi(\vec{x}, t)| \delta\vec{x}$ ).

(decorrentes do “movimento” das partículas) e potenciais (decorrentes das interações entre as partículas), sendo estas as causas para a variação desta função com o tempo no lado direito da equação. Se pudéssemos obter todo o conjunto de soluções possíveis para a equação diferencial em 1.1 (todas as *autofunções* para a função  $\Psi$ ) aplicada a um sistema atômico ou molecular obteríamos a função de estado  $\Psi(\vec{R}, \vec{r}, t)$  - sendo  $\vec{R}$  e  $\vec{r}$  as coordenadas nucleares e eletrônicas, respectivamente - e teríamos o máximo de informação permissível para o sistema mediante a teoria quântica, podendo calcular qualquer propriedade experimentalmente mensurável do mesmo. Contudo, atualmente a resolução da equação de Schrödinger dependente do tempo não é acessível analiticamente a sistemas significativamente mais complexos que átomos hidrogenóides e mesmo soluções numéricas são restritas a casos extremamente simples, não se extendendo a sistemas macromoleculares, devido ao elevado número de variáveis interdependentes envolvidas.

Uma primeira simplificação da equação 1.1 passível de ser realizada quando os potenciais no sistema não variam com o tempo, é a separação dos termos de  $\Psi$  em suas dependências temporal e espacial, respectivamente:

$$\Psi(\vec{x}, t) = \psi(\vec{x})f(t), \quad (1.2)$$

sendo a parte independente do tempo definida como:

$$\frac{-\hbar^2}{2m} \nabla^2 \psi(\vec{x}) + V(\vec{x})\psi(\vec{x}) = E\psi(\vec{x}), \quad (1.3)$$

onde  $E$  é a energia do estado estacionário, que assume um valor *constante* para cada possível solução desta equação (diferentes valores de  $E$  para cada uma das autofunções de  $\Psi(\vec{x})$ ). Mesmo esta versão “simplificada” da equação de Schrödinger (a equação independente do tempo) ainda é impraticável para sistemas de interesse bioquímico, necessitando portanto de maiores aproximações de forma a permitir sua resolução.

Sistemas atômicos e moleculares, são essencialmente descritos pelo comportamento de dois

tipos de partículas - os elétrons e o núcleo atômico, o último apresentando massa e carga variadas conforme o número atômico do elemento em questão. Estes dois tipos de partículas diferem em suas massas em, no mínimo, 3 ordens de grandeza<sup>4</sup>. Esta substancial diferença de massas implica em que os elétrons e os núcleos de um dado sistema atômico apresentarão graus de liberdade significativamente distintos, podendo assim, em uma boa aproximação, terem suas dinâmicas tratadas separadamente. Esta é a base para a *aproximação de Born-Oppenheimer* (19). Ela leva ao seguinte procedimento para a resolução da equação de Schrödinger independente do tempo:

- Primeiramente, resolve-se a *equação de Schrodinger eletrônica* considerando-se núcleos pontuais e fixos no tempo. Desta forma, a equação diferencial remanescente é simplificada visto que são eliminados do operador Hamiltoniano ( $\mathcal{H}$ ) os operadores respectivamente referentes às energias cinéticas dos núcleos (que são considerados como imóveis nesta primeira abordagem) e às energias potenciais de repulsão internucleares (visto que estas são constantes para um sistema em que a disposição espacial dos núcleos é fixa e visto que a adição ou remoção de uma constante não apresenta qualquer efeito para a integração de uma equação diferencial). As autofunções  $\Psi_{elet}$  resolvidas desta forma refletem unicamente a distribuição dos elétrons e juntamente às energias  $E_{elet}$  correspondentes a cada autofunção passam a apresentarem dependência paramétrica das coordenadas nucleares previamente fixadas. A equação do Hamiltoniano eletrônico ( $\mathcal{H}_{elet}$ ) pode ser descrita, de forma simplificada por:

$$\mathcal{H}(elet)\Psi(elet) = E(elet)(\vec{r}_1, \vec{r}_2, \vec{r}_3, \dots, \vec{r}_n) \quad (1.4)$$

para um sistema contendo  $n$  elétrons, cada um com coordenadas  $\vec{r}_i$ . Esta equação, resolvida para um leque de coordenadas internucleares descriptivas de uma condição de círculo de interesse e em conjunto às restrições decorrentes do princípio de exclusão de Pauli (a antissimetria da função de onda), permite calcular uma superfície de energia po-

---

<sup>4</sup> O núcleo atômico mais leve - o de um átomo de hidrogênio, constituído de apenas um próton - apresenta uma massa cerca de 1836 vezes maior que a do elétron.

tencial gerada meramente pela distribuição da densidade eletrônica para cada uma destas disposições espaciais dos núcleos no sistema.

- A segunda etapa consiste em tratar unicamente a resolução da dinâmica dos núcleos, considerando a superfície de energias obtidas para a solução do problema eletrônico no passo anterior como um potencial de campo efetivo que, em conjunto às repulsões internucleares (tratadas explicitamente) passam a constituir a energia potencial do sistema. Desta forma, o novo Hamiltoniano nuclear ( $\mathcal{H}_{nuc}$ ) definido pela aproximação de Born-Oppenheimer (descrito de forma simplificada) é:

$$\mathcal{H}_{nuc} = - \sum_{A=1}^M \frac{1}{2M_A} \nabla_A^2 + \mathcal{E}_{tot}, \quad (1.5)$$

em que o primeiro termo aconta para o somatório das energias cinéticas para todos os núcleos (também tratadas de forma explícita), enquanto que o segundo termo  $\mathcal{E}_{tot}$  descreve o total da energia potencial “sentida” pelos núcleos - o campo efetivo gerado pela superfície potencial obtida a partir da resolução da densidade eletrônica no passo anterior somado às energias de repulsão internucleares.

Estando-se interessado na trajetória dos núcleos sobre espaços curtos, o termo  $\mathcal{E}_{tot}$  poderia ser calculado *passo a passo*, através de um procedimento *ab initio* ou semi-empírico tradicional. A superfície potencial eletrônica e as repulsões internucleares obtidas a cada passo podem ser utilizadas como potenciais para se propagar a trajetória dos núcleos para o passo seguinte e então estes mesmos potenciais podem ser recalculados, novamente armazenados para o próximo passo e assim sucessivamente, desde que este processo recursivo seja efetuado sobre intervalos pequenos de tempo de integração. Entretanto, esta abordagem é ainda, obviamente impraticável do ponto de vista numérico para sistemas com milhares de átomos, sobretudo para trajetórias sob intervalos da ordem de nanosegundos, sendo estes os intervalos para os quais a maior parte dos efeitos de interesse bioquímico ocorrem (10). Na prática, ao invés de se realizar subsequentes passos de obtenção de  $\mathcal{E}_{tot}$  através do cálculo da função de estado eletrônica na etapa 1 da aproximação de Born-Oppenheimer, seguida da propagação das trajetórias nucleares

na etapa 2, o que se faz para este tipo de sistema é aproximar o valor de  $\mathcal{E}_{tot}$  através de funções analíticas ajustadas a dados empíricos ou *ab initio* de forma a reproduzir a superfície de energia sob a qual os núcleos se movem nas imediações do equilíbrio químico (condições estas, sobre as quais a análise de trajetórias de dinâmica molecular se aplicam, conforme previamente mencionado em 1.2). Os potenciais “sentidos” pelos núcleos nas imediações do equilíbrio químico, por sua vez, se ajustam significativamente bem a funções com “formatação” clássica, de forma muito conveniente para a resolução computacional de suas trajetórias. Os termos repulsivos internucleares são potenciais coulômbicos simples, enquanto que os efeitos decorrentes do campo efetivo eletrônico podem ser aproximados por uma série de funções igualmente simples, conforme o tipo de interação envolvida (3, 15, 16).

Desta forma, o primeiro passo para a descritividade clássica do sistema, o desacoplamento das dinâmicas nuclear e eletrônica e a modelagem dos potenciais atuantes sobre os núcleos através de funções analíticas *clássicas*, se tornam respectiva e automaticamente permissíveis através da aproximação de Born-Oppenheimer e pela ajustabilidade a este tipo de função para a superfície potencial obtida nas condições de contorno da simulação. Note que os efeitos eletrônicos não são *negligenciados* nesta aproximação, eles se encontram originalmente implícitos na superfície potencial descrita em  $\mathcal{E}_{tot}$  e, desde que se escolha um conjunto de funções analíticas que modelem de forma suficientemente acurada esta superfície, *i. e.*, desde que o sistema seja bem *parametrizado*, estes efeitos continuarão implícitos nos potenciais remanescentes na aproximação final da dinâmica do sistema.

O próximo passo para que a trajetória dos átomos no sistema (agora representadas pelas trajetórias dos núcleos sobre o campo efetivo eletrônico e de repulsão internuclear) seja descrita classicamente é que também a *movimentação* dos núcleos ao longo da trajetória possa ser descrita do ponto de vista clássico. Uma descrição quântica da trajetória dos núcleos no potencial aproximado obtido conforme acima seria oriunda da resolução da seguinte equação:

$$\mathcal{H}_{nuc} \Psi_{nuc} = i\hbar \frac{\partial \Psi_{nuc}}{\partial t}. \quad (1.6)$$

Contudo, a resolução desta trajetória quântica não é de todo necessária para átomos pesados e termalizados a temperaturas consideráveis em Kelvins (por exemplo, à temperatura ambiente de 298 K). Isto ocorre devido a um efeito de localização do *pacote de ondas* para estes átomos devido à combinação da contribuição de várias autofunções de maior energia nestas condições:

$$\Psi_{nuc}(\vec{x}, t) = \sum_{i=1}^{\infty} C_i \psi_i(\vec{x}) e^{-iE_i t/\hbar}, \quad (1.7)$$

onde  $\Psi_{nuc}(\vec{x}, t)$  é a função de estado do núcleo na trajetória, descrita por um pacote de ondas formado por todas as autofunções  $\psi_i(\vec{x})$ , cada uma com seu autovalor  $C_i \psi_i$  e contribuindo para o pacote com um “peso”  $e^{-iE_i t/\hbar}$ . Nota-se que as funções mais energéticas (e mais deslocalizadas em  $\vec{x}$ ) contribuem com pesos exponencialmente menores. À temperatura ambiente (prova-se) (19) este efeito culmina com que os núcleos da maioria dos átomos pesados possam ser descritos por pacotes de onda cujos comprimentos de onda térmicos de de Broglie (17) são significativamente menores que a distância média entre as partículas em soluções aquosas a 1 atm, conforme foi ilustrado para o núcleo de carbono na *nota 2* no fim da última subseção. Desta forma, os núcleos podem ser considerados como partículas pontuais se propagando com trajetórias clássicas (lembrando-se que para o caso dos átomos leves, correções quânticas devem ser embutidas no cálculo das trajetórias de forma a se obter resultados acurados para as simulações).

A esta aproximação da trajetória pontual dos núcleos sobre um potencial efetivo aproximado na forma de um conjunto de funções analíticas *clássicas* dá-se o nome de *Dinâmica Molecular* e ao conjunto de funções utilizadas para se descrever os potenciais efetivos denomina-se *Campo de Forças*.

## 1.2.2 Aproximações nos Campos de Forças comumente utilizados

A construção de funções analíticas para a construção de parâmetros *auto-consistentes* na descrição das propriedades dos átomos em uma simulação por dinâmica molecular, comumente

utiliza uma mescla da obtenção de parâmetros por cálculos *ab initio* ou semi-empíricos e de dados experimentais. As funções de energia comumente utilizadas consistem na aproximação dos potenciais nas imediações do equilíbrio como funções *harmônicas*. Isto é válido porque as energias das ligações *estáveis* abordáveis em dinâmica molecular (por exemplo, C-C, H-H, O-O, N-N, C-H, C-O, C-N... dentre outros) são suficientemente menores que a energia térmica ( $\sim kT$ ), da ordem de 0.026 eV à temperatura ambiente. Desta forma, os átomos envolvidos nessas interações tendem a se manter nas imediações do mínimo de potencial, região esta onde o potencial pode ser convenientemente aproximado por funções desta natureza (e que por sua vez são consideravelmente simples de serem resolvidas computacionalmente). Além dos termos que descrevem harmonicamente o comportamento de átomos em distâncias de até 3 ligações (estiramento de uma ligação química, variação de ângulo entre duas ligações e de diedro entre três, respectivamente) é também necessário tratar as interações a longas distâncias. Sobretudo para proteínas e macromoléculas biológicas de uma forma geral este tipo de interação apresenta importância fundamental para os movimentos coletivos. Tanto as interações de longas distâncias *internas* nas macromoléculas quanto as interações intermoleculares (estas de fundamental importância na descrição de qualquer aglomerado de átomos em condições distintas que as de um gás ideal (17)) são tratadas como interações coulombicas (para as interações de longa distância) e de van der Walls (funções de Lennard-Jones) para as interações dispersivas de curtas distâncias nos campos de força comumente utilizados no estudo de proteínas e ácidos nucléicos (20). Desta forma, dentre os parâmetros adicionados aos átomos na construção de um campo de forças, cada átomo  $i$  em um dado tipo de *grupamento químico* apresenta uma *carga* (que pode ser adicionada baseando-se, por exemplo, nas diferenças de densidades eletrônicas para os átomos envolvidos em uma ligação química, conforme obtido em um cálculo *ab initio*) e dois *coeficientes de Lennard-Jones intrínsecos* ( $\epsilon_i$  e  $\sigma_i$ ), que são linearmente ou geometricamente (dependendo do campo de forças) combinados aos outros dois coeficientes de um segundo átomo  $j$ , obtendo-se um coeficiente de Lennard-Jones *do par i-j* para o cálculo das interações dispersivas entre ambos durante a simulação. Estes coeficientes  $\sigma$  e  $\epsilon$  para o cálculo das interações dispersivas entre pares de átomos  $i-j$  por sua vez, são mais difíceis de serem obtidos de forma acurada *ab initio*, geralmente tendo de serem associados a dados experimentais (pode-se por

exemplo, na parametrização de uma nova molécula de forma *consistente* com a do campo de forças previamente parametrizado para o ambiente no qual a molécula será simulada, buscar por parâmetros  $\sigma$  e  $\varepsilon$  que, tornem o comportamento desta molécula condizente com a compressividade e/ou com o calor de vaporização, conforme experimentalmente determinado para um gás ou uma solução contendo a mesma) (15, 16).

para os campos de força CHARMM22 (21) e AMBER (22), utilizados nas simulações dos capítulos 3 e 4, respectivamente, os termos referentes às aproximações harmônicas para os potenciais que regem os movimentos de átomos separados por até 3 ligações (chamados *potenciais ligados*) assumem a seguinte forma:

$$V_{(lig)} = \sum_{lig} K_d(d - d_0)^2 + \sum_{ang} K_\theta(\theta - \theta_0)^2 + \sum_{diedros} K_\chi(1 + \cos n\chi - \delta) + \sum_{improp} K_\phi(\varphi - \varphi_0)^2. \quad (1.8)$$

Nesta equação, os termos do potencial da esquerda para a direita são, respectivamente, os potenciais harmônicos para a variação da distância  $d$  entre dois átomos ligados em relação à distância no equilíbrio ( $d_0$ ), para a variação do ângulo entre as duas ligações para três átomos separados por estas mesmas ligações ( $\theta$ ) relativo ao ângulo no equilíbrio ( $\theta_0$ ) e, de forma similar, para a variação do diedro  $\chi$  entre três ligações definidas por quatro átomos em relação à posição no equilíbrio definida por  $\delta$  e pela multiplicidade  $n$  para o mesmo diedro<sup>5</sup>.

Para o cálculo da superfície potencial final, estes termos são somados aos termos referentes às interações coulômbicas a longas distâncias e os termos referentes às interações dispersivas para átomos separados por mais que 3 ligações (*potenciais não ligados*):

$$V_{(nao-lig)} = \sum_{nao-ligado} \varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{r_{ij}}, \quad (1.9)$$

sendo o primeiro termo do somatório correspondente ao potencial de Lennard-Jones a partir

---

<sup>5</sup> para a diferença energética entre as conformações *gauche* e *eclipsada* definidas pelos diedros dos hidrogênios de cada um dos grupos metila de uma molécula de etano, por exemplo, os ângulos múltiplos inteiros de 60 ° são todos energeticamente permissíveis (e equivalentes) enquanto que os ângulos múltiplos inteiros de 120 ° e o ângulo de 0 ° são todos equivalentemente desfavoráveis.

do qual são calculadas as energias para as interações não ligadas dispersivas, enquanto que o segundo termo corresponde ao potencial coulômbico para o cálculo das interações de natureza eletrostática e/ou polar a longas distâncias<sup>6</sup>.

### 1.2.3 Resolvendo as trajetórias - Integração das equações de movimento

Nas duas últimas subseções foi mostrado como as interações entre as partículas de um sistema podem ser descritas em termos de potenciais *clássicos* e como é permissível resolver as mesmas trajetórias utilizando também uma visão clássica. Desta forma, de posse de uma descrição acurada destes mesmos potenciais, tudo o que necessitamos é de que o sistema se encontre em condições iniciais bem resolvidas para que se possa prever a trajetória das partículas ao longo do tempo. Estas condições iniciais, por sua vez, são as posições e energias de todas as partículas do sistema. Para simulações de biomoléculas, estas posições iniciais geralmente são obtidas a partir de estruturas cristalográficas e, com menor freqüência, a partir de estruturas resolvidas por espectroscopia de resonância nuclear magnética. Em última análise, métodos de modelagem também podem ser utilizados na geração das coordenadas iniciais, na ausência de uma estrutura experimental ou nos casos em que a(s) estrutura(s) resolvida(s) apresenta(m) limitações quanto à sua resolução e/ou informação (23–27). Obviamente, quanto maior o grau de modelagem na preparação de uma condição inicial para a simulação (quão mais distante de um dado experimental original), maior o rigor que deve ser tomado nas etapas prévias de preparação do sistema antes de se gerar o que se chama *dinâmica produtiva* (a fração da simulação de dinâmica molecular que de fato será analisada). As velocidades iniciais por sua vez, geralmente são atribuídas aleatoriamente aos átomos individuais do sistema, mas de forma a que a *distribuição* das velocidades para cada átomo obedeça a uma distribuição de Maxwell-Boltzmann condizente com a temperatura desejada.

Dadas as posições de todos os átomos do sistema e a descrição dos potenciais, a mecânica clássica nos permite prontamente calcular as força atuando sobre os mesmos como a derivada do potencial:

---

<sup>6</sup> No sistema internacional de unidades, usados nas simulações em geral, o potencial eletrostático é multiplicado por  $1/(4\pi\epsilon_0)$ , sendo  $\epsilon_0$  a constante de permissividade no vácuo.

$$\vec{F} = -\nabla \vec{V}(t) \quad (1.10)$$

A segunda lei de Newton, permite calcular a aceleração de cada átomo,  $\vec{a}(t)$ , a partir da força usando

$$-\nabla \vec{V}(t) = m \vec{a}(t) \quad (1.11)$$

sendo  $m$  a massa do átomo.

O algoritmo mais simples para se obter a variação das coordenadas das partículas no tempo consiste em calcular as velocidades no instante  $t + \Delta t$  por

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a}(t)\Delta t \quad (1.12)$$

e obter as posições no instante  $t + \Delta t$  através da equação

$$\vec{x}(t + \Delta t) = \vec{x}(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2}, \quad (1.13)$$

Com as novas posições no instante  $t + \Delta t$  os potenciais podem ser recalculados. A derivação dos novos potenciais define novas forças que, por sua vez, permitem o cálculo de uma nova posição no instante posterior e assim recursivamente, permitindo a resolução da evolução temporal do sistema.

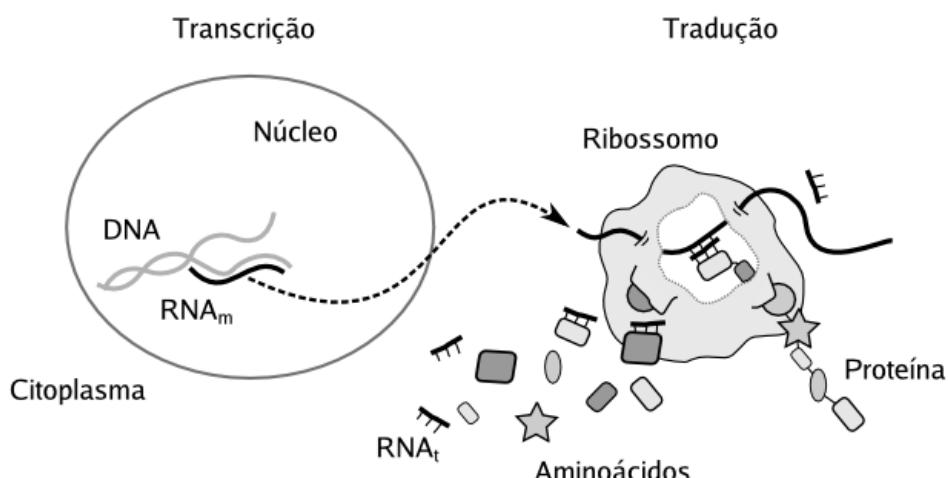
## 1.3 Receptores Nucleares e seu papel na transcrição gênica

### 1.3.1 A transcrição gênica e os fatores de transcrição

A transferência da informação genética de uma seqüência de DNA no núcleo da célula para uma seqüência de aminoácidos em uma proteína, tendo como intermediário uma molécula de RNA, é o que se denomina *expressão gênica*. O processo como um todo apresenta três divisões principais (28):

- A *transcrição gênica*, em que um transcrito primário de pré-RNA mensageiro (pré-mRNA) é formado;
- A etapa de *conversão do transcrito primário* para o mRNA maduro, em que processamentos secundários, como o *splicing* com remoção de seqüências não tradutoras ocorre enquanto a molécula de mRNA é transportada do núcleo para o citosol.
- A *tradução* propriamente dita, em que a síntese de proteínas ocorre no ribosoma.

Esta cadeia de eventos forma o que se conhece como *dogma central da biologia* em que uma determinada seqüência no DNA é associada a uma seqüência protética e só a partir daí correlacionada a uma função biológica<sup>7</sup>



**Figura 1.3 – Dogma central da biologia:** A síntese de proteínas tem basicamente três etapas: a transcrição, processamento intermediário e tradução. Na transcrição uma fita de RNA<sub>m</sub> produzida a partir do DNA. No processamento, enquanto a molécula de pré-mRNA é transportada para o citosol, ela sofre os *splicings* necessários para a tradução final. Na tradução a seqüência de nucleotídeos do RNA<sub>m</sub> é finalmente lida pela maquinária provida pelos complexos de tRNAs, codificando a seqüência de aminoácidos da proteína.(retirado de (3))

Cada uma destas etapas é rigidamente controlada por uma maquinária específica de complexos protéticos, dada a sua óbvia importância na manutenção da homeostase. Porém, como qualquer processo de transferência de informação, a etapa com maior impacto sobre o processo como um todo ocorre na base - na etapa de transcrição. Além da maquinária envolvida

<sup>7</sup> Na verdade, este *dogma* é quebrado quando se consideram as sequências não traduzíveis (por exemplo, os *introns*) e as sequências de transcrição dos RNA ribossômicos e transportadores (rRNA e tRNA, respectivamente), ambos apresentando funcionalidade *per se* (29). Contudo, em linhas gerais, o dogma reflete bem a hierarquia de comandos na associação da informação gênica à função bioquímica.

no processo de transcrição propriamente dita, *i.e.*, as diversas enzimas dos complexos RNA-polimerase, em organismos eucariotos, a *ativação* ou *repressão* da transcrição gênica são ambas finamente controladas por uma classe de proteínas denominada *fatores de transcrição*. Estes por sua vez podem ser de três tipos: fatores gerais, fatores *upstream* e fatores de transcrição indutíveis (28, 30).

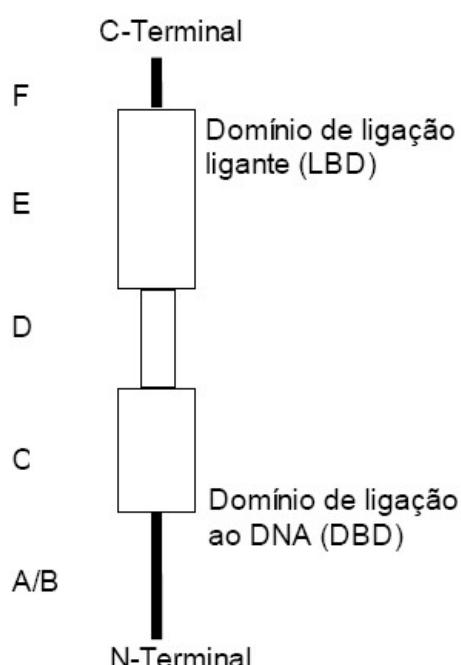
Fatores de transcrição gerais são proteínas que se ligam indiscriminadamente a sítios de DNA em que a transcrição foi ativada por um dos outros dois tipos de fatores, recrutando o complexo RNA-polimerase (RNAP) e permitindo o início da síntese de mRNA. Fatores de transcrição *upstream* reconhecem unicamente a sítios de resposta específicos, localizados na região 5' precursora de genes alvo, não sendo secundariamente ativados (quer por uma acentuação de sua afinidade a estes sítios, quer por sua maior habilitação no recrutamento da maquinária de transcrição) por qualquer agente *externo* à seqüência específica do DNA. Podem agir como ativadores ou repressores da transcrição. A regulação gênica ao nível destes receptores é efetuada basicamente através do controle de sua concentração *in situ*, visto que sua atividade não depende de outros fatores reguladores. Já os fatores de transcrição indutíveis, por sua vez, se ligam também a trechos específicos no DNA (como os fatores *upstream*), mas, além disto, são também regulados pela associação ou não de outras moléculas. Isto confere a este tipo de fator uma plasticidade significativamente maior em seu papel no controle da transcrição, visto que a associação de diferentes moléculas reguladoras implicam em diferentes respostas, podendo permitir, por exemplo, a requisição de diferentes proteínas co-reguladoras da maquinária de transcrição, implicando em diferenças na intensidade ou mesmo na *qualidade* da sinalização aferida (podendo agir na ativação ou na repressão gênica).

### 1.3.2 Receptores Nucleares

Receptores nucleares hormonais (NRs) são fatores de transcrição indutíveis (31). Estes receptores compreendem uma família de proteínas que agem no núcleo da célula, estando envolvidos em praticamente todas as funções fisiológicas do organismo. Se ligam a elementos de resposta específicos no DNA denominados *elementos responsivos hormonais* (HREs), medi-

ando os efeitos pleiotrópicos de seus ligantes, geralmente pequenas moléculas lipofílicas, dentre as quais diversos tipos de hormônios lipofílicos: esteróides, retinóides, vitamina D e hormônios tiroideanos (32). Há também dentre eles um grupo de receptores para os quais não foram identificados ligantes naturais, sendo estes denominados “receptores órfãos” (33). A responsividade destes receptores a diferentes agentes de controle do metabolismo e da homeostase os converte em intermediários terminais entre a influência do meio e a resposta gênica, o que, por sua vez, os torna em alvos farmacológicos óbvios.

Todos os receptores nucleares possuem uma estrutura modular 1.4 constituída de três domínios principais: o domínio N-terminal (região A/B) pouco conservado, responsável pela função de ativação gênica na ausência do ligante (*Activation function 1-AF-1*); um domínio central (região C), responsável pela ligação da proteína ao DNA (*DNA binding domain-DNA*) e o domínio C-terminal (região E), responsável pela ligação ao ligante (*Ligand binding domain-LBD*) e pela função de ativação induzida (*Activation function 2-AF-2*). Há também uma região responsável pela ligação do DBD ao LBD, a região denominada *hinge* (“dobradiça”- região D). O Domínio DBD é o domínio mais significativamente conservado dentre a família, com cerca de 80 % de homologia, seguido pelo domínio LBD com cerca de 30 % de homologia (32).



**Figura 1.4** – Representação esquemática da estrutura modular de um receptor nuclear.

Os domínios LBDs são capazes de se homo ou heterodimerizarem em solução, apresentando esta formação de dímeros um papel funcional (32). Além disto, certos receptores como o RXR e o TR (vide abaixo) apresentam a habilidade de formarem tetrâmeros, aparentemente não funcionais (não se associam ao DNA), em solução e que tendem a se dissociar na presença do ligante ou do elemento responsivo no DNA (34, 35). Foi inferido que estes tetrâmeros funcionariam no “tamponamento” da atividade destes receptores, tornando-os indisponíveis na ausência do ligante ou do sítio correto. Contudo, ainda não se tem comprovações quanto à presença destes tetrâmeros *in situ* e mesmo sua formação a concentrações próximas às celulares apresentam resultados controversos na literatura (34, 36).

Já os DBDs são capazes de se dimerizar unicamente na presença do elemento responsivo no DNA (32), apresentando papel fundamental na seletividade e na interação cooperativa de dímeros de NRs aos elementos responsivos , conforme será visto abaixo.

### **1.3.3 A interação de receptores nucleares a elementos responsivos**

Diferentes receptores nucleares se ligam seletivamente a diferentes elementos responsivos do DNA. Estas seqüências, por sua vez, são constituídas de repetições diretas ou palindrômicas de seqüências de cinco a seis pares de bases (pb) derivadas de uma seqüência *arquetípica*.

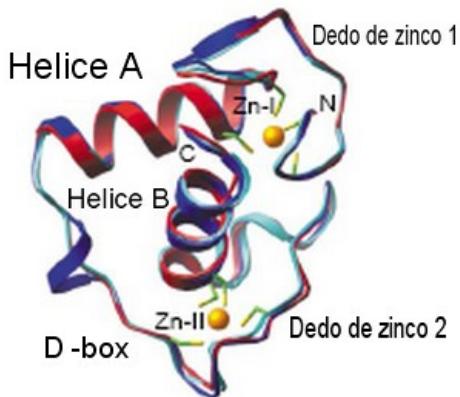
A maioria dos receptores dos não esteróides, e órfãos reconhece HREs compostos de repetições da seqüência consenso 5'-AGGTCA-3', apresentando uma ou duas cópias desta mesma seqüência (7, 37). Este grupo inclui aproximadamente todos os receptores conhecidos como de não esteróides, incluindo o receptor do ácido 9-cis retinóico (*Retinoid X Receptor-RXR*) o receptor do ácido *all-trans*-retinóico (*Retinoic Acid Receptor-RAR*, receptor do hormônio tireoideano (TR), receptor de vitamina D, receptor ativado de proliferação peroxisomal (*Peroxisome Proliferator Activated Receptor-PPAR*) e o fator de crescimento neuronal induzido-B (*Neural Growth Factor Induced-B-NGFI*) (7). As seqüências consenso nos HREs podem estar arranjadas nas formas de repetições diretas (DRs), palindrômicas (Pal) ou na forma de palíndromos invertidos (IPs). A seletividade a estas seqüências geralmente se dá na forma da ligação dos receptores como homo ou heterodímeros. Desta forma, a disposição geométrica de

ambos os monômeros induzida pelo espaçamento e pela orientação das seqüências na topologia de plano inclinado do DNA confere uma interface específica que, por sua vez, permite a seletividade entre um dímero e um HRE específicos. Contudo este mecanismo de seletividade ainda é pouco compreendido. É sabido que a seletividade e mesmo a cooperatividade na ligação ao DNA é regida pela interação entre os DBDs (38). Contudo, este alosterismo e seletividade conferido à dimerização da proteína inteira sobre o HRE por este domínio relativamente pequeno em face aos LBDs é ainda um enigma, sobretudo visto serem as interfaces de dimerização entre os DBDs ligados ao HRE pouco substanciais nas estruturas cristalográficas.

Na Fig. 1.5 as estruturas dos cernes globulares de três domínios DBDs (o dos receptores RXR, RAR, TR) são sobrepostas. Nota-se a alta conservação da topologia entre os diferentes receptores. O domínio DBD é sempre composto de dois motivos dedos de zinco, o primeiro envolvido fundamentalmente na associação ao DNA e o segundo apresentando uma região (o D-box) que é envolvida na interface de dimerização com o parceiro 3', em estruturas de receptores nucleares associados a repetições diretas, e entre ambos os parceiros em estruturas em elementos palindrômicos. A região que estabelece contatos mais significativos com a seqüência responsiva no sulco maior é a hélice C-terminal ao primeiro dedo de zinco (hélice-A). Também uma porção N-terminal do domínio D (*hinge*) - o *T-box* é apontada como crucial para a manutenção da ligação ao DNA e ao parceiro de dimerização em diversos trabalhos (7, 39–41). O papel desta região na ligação colaborativa ao DNA é ainda mais curioso, visto sua localização em uma região consideravelmente desordenada, a interface tênue formada pela mesma com o parceiro nos complexos em DR1, onde ela participa da interface de dimerização (7, 40) e, sobretudo, o fato de que ela não participa no reconhecimento da seqüência específica, sugerindo um possível reconhecimento *topológico* ao ambiente do complexo.

### 1.3.4 O receptor do ácido 9-cis-retinóico e seu papel como parceiro “promiscuo”

Um receptor particularmente interessante no estudo da seletividade de receptores nucleares na interação a diferentes HREs é o RXR. Isto porque, como parceiro de heterodimerização, este receptor se associa a diferentes outros receptores reconhecendo diferentes espaçamentos



**Figura 1.5** – Superposição das estruturas cristalográficas do cerne dos DBDs de RXR (vermelho), RAR (ciano), e TR (azul), retiradas de suas respectivas estruturas cristalográficas associadas ao DNA (4–6). Cada um dos dois dedos de zinco é mostrado em laranja e coordenado por quatro cisteínas em *stick* em amarelo. As hélices A e B e os dois dedos de zinco são rotulados. (retirado de (7)).

em repetição direta para os HREs (7). como homodímero o RXR se associa ao elemento de repetição direta 1 (DR1) (42). Ele também se associa a esta sequência como parceiro no sítio 5' para os receptores RAR, PPAR, COUP e HNF4 (7, 41, 43).

Uma nuance particularmente interessante da associação do RXR como homo ou heterodímero com diferentes parceiros a diferentes NRs é sua polaridade específica na interação com os dois meio-sítios, polaridade esta que é aparentemente (e novamente) ditada pelo DBD e que é mantida com diferente parceiros em um mesmo HRE, mas que é mudada dentre diferentes HREs e que, ao que parece, é relacionada com o tipo de resposta transcrecional induzida: O heterodímero de RXR com RAR em DR2 se dá, obrigatoriamente, com o RXR ligado à posição 3', apresentando este heterodímero uma atividade repressora nesta seqüência. Já o mesmo heterodímero com RAR em DR5, se dá com o RXR associado à posição 5' e apresentando uma atividade ativadora (7).

## 1.4 Celulases e sua importância em biotecnologia

Celulases são enzimas que hidrolizam cadeias de celulose em um substrato microcristalino, convertendo-as em unidades sacarídicas menores (44). Estas enzimas são de suma importância para o metabolismo de microorganismos unicelulares, dentre os mesmos diversas cepas de fun-

gos filamentosos e bactérias, que as utilizam de forma a disponibilizar estas unidades menores sacarídicas como fontes primárias de energia a partir da biomassa vegetal.

Recentemente, um grande interesse tem sido despertado em relação a estas enzimas devido à sua aplicabilidade industrial na conversão de biomassa e materiais “brutos” contendo celulose em glicose, que pode então ser convertida em bioetanol por fermentação (45).

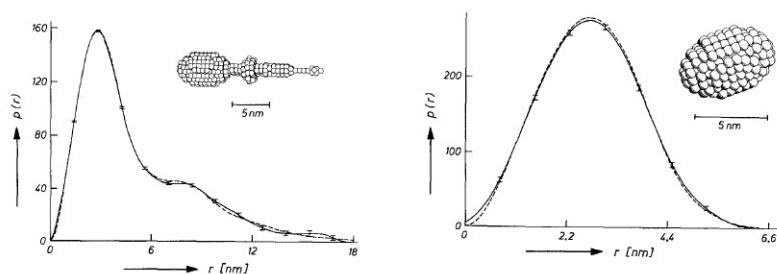
O complexo celulásico hábil na degradação de celulose normalmente consiste de várias enzimas e proteínas auxiliares que agem em sinergismo na solubilização e decomposição dos biopolímeros na biomassa, tais como a celulose. Até agora, um dos mais estudados complexos de celulásicos é o do fungo filamentoso *Trichoderma reesei*, que consiste de mais de vinte diferentes proteínas . O maior componente deste complexo protéico é a enzima celobiohidrolase I (CBH-I, 1.4- $\beta$ -D-glicana celobiohidrolase, E.C. 3.2.1.91), que pertence à família 7 das hidrolases glicosídicas, Cel7A e representa mais de 60 % do total de proteínas secretadas por fungos do gênero *Trichoderma*. Esta enzima degrada cadeias de celulose e celodextrina, liberando unidades de celobiose a partir de substratos não redutores .

Estudos de proteólise limitada para a celobiohidrolase I (CBHI) de *T. reesei* mostram que esta enzima apresenta dois domínios globulares distintos, conectados por uma região móvel, com o domínio C-terminal agindo como módulo de interação à celulose (*Celulose Binding Domain - CBM*), e um segundo domínio N-terminal, substancialmente maior, consistindo do cerne catalítico da enzima (*Catalytic Core Domain*), que é responsável pela clivagem enzimática dos substratos

estes domínios são interligados por um peptídeo conector (LP) flexível, O-glicosilado e rico em seqüências do tipo prolina/serina/treonina. Foi observado que a atividade enzimática é reduzida quando o mesmo peptídeo conector é deletado, sugerindo que o tamanho do LP é importante para assegurar a flexibilidade e correta ação independente dos dois domínios funcionais da enzima

O primeiro envelope molecular a baixa resolução de uma celobiohidrolase foi obtido por dados de espalhamento de raios a baixos ângulos (SAXS). Schmuck e colaboradores mostraram

que a celobiohidrolase I de *T. reesei* apresenta uma conformação de “girino” 1.6, consistindo de uma porção “cabeça” isotrópica e uma “cauda” longa flexível (46). Posteriormente, baseado em análises de SAXS, a mesma arquitetura bimodular foi descrita para uma outra celulase, a de *Cellulomonas fimi*. Estes estudos pioneiros foram limitados em sua precisão e acurácia, devido à carência de estruturas tridimensionais de alta resolução para os domínios isolados, bem como aos equipamentos relativamente modestos para as medidas de SAXS à época.



**Figura 1.6** – Dados de SAXS da literatura para a celobiohidrolase I de *Trichoderma harzianum*

A estrutura do domínio CBM da CBH-I de *T. reesei* foi determinada por RMN em 1989 enquanto a estrutura cristalográfica do domínio catalítico se tornou disponível em 1994 (Fig. EstruturasDomíniosSeparados). Com a disponibilização das estruturas tridimensionais em alta resolução destes dois módulos globulares da enzima, bem como os avanços na tecnologia e nas metodologias de análise para os dados de SAXS e a possibilidade da utilização conjunta de ferramentas mais robustas de modelagem molecular, a obtenção de modelos mais refinados a partir de dados experimentais de menor resolução para estas enzimas de alto potencial de impacto tecnológico se torna hoje uma possibilidade.

## 1.5 A técnica de anisotropia de fluorescência e sua aplicabilidade para estudos de formação de complexos macromoleculares em solução

A fluorescência extrínseca pode ser utilizada de diversas maneiras, entre elas, para a realização de experimentos de anisotropia de fluorescência. Medidas de anisotropia podem ser utilizadas em análises bioquímicas e fornecem informações sobre o tamanho e forma de proteínas ou sobre a mobilidade de diferentes ambientes moleculares. Este tipo de medida é utilizado também para

investigar a associação proteína-proteína ou proteína-DNA. Experimentos de anisotropia estão baseados no princípio de excitação fotoseletiva de fluoróforos pela luz polarizada. Fluoróforos absorvem preferencialmente fôtons cujos vetores elétricos estejam alinhados paralelamente ao momento de transição do fluoróforo. Este, por sua vez, tem sua orientação definida com relação à molécula. Numa solução isotrópica, os fluoróforos estão orientados aleatoriamente. Com a incidência da luz polarizada, aqueles fluoróforos que tiverem seu dipolo de transição alinhados paralelamente ao vetor elétrico da luz incidente serão seletivamente excitados. Esta excitação seletiva resulta em uma população de fluoróforos parcialmente orientada (fotoseleção) e uma emissão fluorescente parcialmente polarizada. Esta emissão também ocorre com a luz polarizada ao longo de um eixo fixo do fluoróforo. O ângulo relativo entre estes momentos determina a máxima anisotropia. Anisotropia e polarização são definidas por (WEBER, 1952):

$$r = (I_p - I_{pp}) / (I_p + 2I_{pp}) \quad (1.14)$$

$$P = (I_p - I_{pp} / (I_p + I_{pp})) \quad (1.15)$$

onde  $r$  é a anisotropia,  $P$  é a polarização,  $I_p$  e  $I_{pp}$  são as intensidades de fluorescência da emissão polarizada verticalmente e horizontalmente, quando a amostra é excitada com luz polarizada verticalmente. Alguns fenômenos podem diminuir a máxima anisotropia medida em uma amostra, sendo o mais comum a difusão rotacional. Essa difusão ocorre durante o tempo de vida do estado excitado e desloca o dipolo de emissão do fluoróforo. A intensidade deste deslocamento dependerá da difusão e portanto, do tamanho e/ou mobilidade da proteína a qual a molécula está ligada. Sendo assim, o observação da anisotropia de fluoróforos ligados ao DNA pode ser utilizada para estudos de associação à proteínas.

## 1.6 Espalhamento de Raios-X a baixos ângulos - SAXS

A técnica de espalhamento de raios-X a baixo ângulo, SAXS, permite a obtenção de informações da estrutura de diversos materiais. Informações como flutuações ou heterogeneidades da densidade eletrônica com dimensões características da ordem de 10 a 1000 Å podem ser obtidas, por exemplo, para soluções de macromoléculas biológicas (proteínas, ribossomos, DNAs, etc).

A interpretação de dados de espalhamento a baixos ângulos de solução de macromoléculas permite a determinação do envelope molecular e, consequentemente, de estados de oligomerização. Dependendo da concentração destas soluções, podemos ter macromoléculas interagindo umas com as outras, efeito inter-partícula. Isso gera uma correlação espacial que altera a intensidade espalhada, gerando um artefato no sinal, dificultando a análise dos dados. Sendo assim, embora o uso de mais altas concentrações seja desejável, para diminuir a dispersão do sinal, é importante que os sinais de baixas e altas concentrações sejam proporcionais, descartando a possibilidade de estar ocorrendo um efeito inter-partícula.

As soluções de macromoléculas biológicas apresentam pouca diferença na densidade eletrônica entre seus constituintes, que são essencialmente compostas por átomos leves (H, C, N, O e P no caso de ácidos nucleicos), e o solvente. Estes tipos de soluções produzem um espalhamento muito fraco necessitando usar feixes de raios-X produzidos por fontes síncrotron, que possuem um fluxo de fótons maior que aparelhos de raios-X convencionais.

A diferença entre a energia dos fótons de raios-X incidentes e a energia de ligação dos elétrons permite tratar estes últimos como elétrons livres. Desta forma, partículas livres carregadas oscilantes produzem ondas espalhadas coerentes com a onda incidente em todas as direções. Em solução, as partículas de macromoléculas estão em movimento browniano e podem assumir todas as orientações possíveis em relação ao feixe incidente. Assim, a intensidade  $I(q)$ , produzida por esse sistema, é proporcional a amplitude de espalhamento de uma macromolécula, promediado para todas as orientações. Sendo  $q = 4\pi \sin(\theta)/\lambda$ , onde  $2\theta$  é o ângulo entre o feixe incidente e o feixe espalhado e  $\lambda$  é o comprimento de onda utilizado. Quando os valores de  $q$  são pequenos, vale a aproximação de Guinier :

$$I(q) = I(0) \exp(-1/3 \cdot Rg^2 \cdot q^2)$$

sendo  $Rg$  o raio de giro. O  $Rg$  é um parâmetro estrutural importante utilizado na análise de mudanças conformacionais de macromoléculas em solução. A equação acima pode ser reescrita na forma:

$$\ln[I(q)] = -(1/3 \cdot Rg^2) q^2 + \ln[I(0)]$$

Sendo assim, ao se fazer a análise gráfica da variação logarítmica de intensidade produzida por um determinado sistema,  $\ln[I(q)]$ , em função de  $q^2$ , obtém-se uma reta, onde o coeficiente angular é  $-(1/3 Rg^2)$  e o coeficiente linear é  $-\ln[I(0)]$ . A partir do valor de  $-\ln[I(0)]$ , possuindo-se um parâmetro de comparação, pode-se calcular a massa molecular do sistema.



## 2 *Objetivos*

Este trabalho de doutoramento tem como objetivo a utilização integrada de abordagens de dinâmica molecular, modelagem estrutural e técnicas experimentais de análise biofísica em solução para o estudo de dois sistemas de interesse biológico, para os quais os movimentos coletivos e o acoplamento energético a longas distâncias prestam papéis fundamentais.

No primeiro sistema, o mecanismo em que a associação do domínio de interação ao DNA (DBD) da isoforma  $\alpha$  do receptor do ácido 9-cis-retinóico humano (hRXR $\alpha$ ) confere a selectividade e alosterismo necessários para a associação dimérica do receptor completo aos elementos responsivos hormonais (HRE) no DNA será avaliado. Para tal, resultados de simulação de dinâmica molecular para o domínio e para as seqüências dos oligonucleotídeos isolados em solução, bem como do complexo com diferentes arranjos sobre o elemento responsivo fisiológico (a repetição direta com espaçamento de um par de bases - DR1) serão analisados em conjunto a resultados experimentais de afinidade do receptor a diferentes HREs através da técnica experimental de anisotropia de fluorescência. Também através das simulações de dinâmica molecular, o mecanismo para a polaridade da interação do receptor a DR1 será averiguado.

Para o segundo sistema, técnicas de modelagem estrutural e de dinâmica molecular por anelamento simulado (*simulated annealing/molecular dynamics - SA/MD*) foram utilizadas na geração, refinamento e validação de modelos estruturais de alta resolução para a celobiohidrolase I de *Trichoderma harzianum*, em sua completa extensão e perfil de glicosilações, a partir de modelos de baixa resolução obtidos através da técnica experimental de espalhamento de raios X a baixos ângulos (*small angle X ray scattering - SAXS*). A análise dos modelos recuperados

por SA/MD será também considerada no que se refere às principais restrições conformacionais e dinâmicas entre os dois domínios da proteína e no impacto das glicosilações a estas mesmas restrições. Esta análise será direcionada almejando melhor compreender os possíveis mecanismos de ação concertada entre os dois domínios desta enzima para a associação e movimento sobre a superfície da celulose, ambos processos fundamentais para sua atividade biológica e de especial interesse para seu emprego em biotecnologia.

## 3 *Materiais e Métodos*

### 3.1 Metodologia Experimental para os estudos de afinidade do hRXR $\alpha$ - $\Delta$ AB a diferentes elementos responsivos no DNA

#### 3.1.1 Reagentes

Todos os reagentes utilizados foram de grade analítica. Todas as soluções foram preparadas logo anteriormente ao uso com água destilada, deionizada a menos que  $1.0 \mu\text{S}$  e filtrada através de uma membrana de  $0.22 \mu\text{m}$  em um sistema de purificação de água. O 9-cRA foi obtido da SIGMA<sup>R</sup> Lote 122367-X8.

#### 3.1.2 Expressão e purificação do hRXR $\alpha$ - $\Delta$ AB

A construção RXR $\alpha$   $\delta$ AB humano ou hRXR $\alpha$ -DL (resíduos 126-462, carecendo do domínio N-terminal) de 42 kDa foi expressa em *Escherichia coli* cepa BL21(DE3) utilizando o plasmídeo pET28a(+) (Novagen<sup>R</sup>). Foi utilizada uma construção contendo uma cauda de seis histidinas (tag(6-His) fusionada *in frame* ao C-terminal. Os protocolos de expressão e purificação tiveram como base os originalmente descritos por Fischer et al. Basicamente, a indução foi feita com 0.5 mM de IPTG em meio de cultura 2XYT, na presença de  $10 \mu\text{M}$  de ZnSO<sub>4</sub> (de forma a estabilizar os dois motivos dedos de zinco no domínio DBD da proteína). O crescimento ocorreu durante 3 horas a  $25^\circ\text{C}$  sob agitação circular a 250 rpm, até que foi atingida a  $\text{DO}_{600\text{nm}} = 0.7 \text{ AU}$ . As células foram coletadas através de centrifugação e ressuspensas em tampão Na<sub>2</sub>HPO<sub>4</sub> 50 mM, pH 8.00 contendo NaCl 100 mM, glicerol 10% V/V, imidazol 10 mM,  $\beta$ -

mercapto etanol 2mM e PMSF 1mM. A células de cada 1 L de meio foram ressuspensas em 50 ml de tampão. Esta suspensão foi então incubada em lisozima (concentração final de 500  $\mu\text{g/ml}$ ) por 30 min a 4 °C para posterior sonicação, também a 4 °C, por cerca de 15 ciclos de 15 segundos, com intervalo de 15 segundos entre cada ciclo. Após este passo, a proteína se encontrava na fase solúvel e pôde ser separada do *pellet* por centrifugação, deixando a amostra sempre à temperatura de 4 °C.

O sobrenadante da quantidade de células de cada 1 L de expressão foi incubado com 1 ml de resina de afinidade Talon Superflow por 1 hora a 4 °C. Após a incubação a resina foi lavada com 4 volumes do mesmo tampão de ressuspensão do lisado. A eluição foi feita com aproximadamente 4 volumes do mesmo tampão, porém com 300 mM de imidazol. Todas as etapas da cromatografia de afinidade foram realizadas com eluição isocrática em bancada, com a resina e amostra incubadas a 4 °C. As frações contendo a proteína recombinante foram identificadas por dosagem pelo método de Bradford e sua pureza verificada por SDS-PAGE 15%.

Seguido à etapa de cromatografia de afinidade, a amostra foi submetida a uma etapa de cromatografia por exclusão por tamanho, de forma a separar os estados oligoméricos da proteína. A amostra foi eluída em frações de 5.5 ml em uma coluna HiLoad Superdex 200 26/60 (Amersham Bioscience) e o perfil cromatográfico, verificado por absorção a 280 nm, foi utilizado para se isolar as frações de interesse. A eluição foi feita em tampão Hepes 20 mM, pH 8,0, com NaCl 200 mM, glicerol 5 % e DTT 1 mM. Esta etapa cromatográfica foi realizada utilizando-se um sistema AKTA Purifier da Amersham Bioscience. As amostras tiveram sua pureza verificada por SDS-PAGE 15 % e as frações provenientes de um mesmo pico (*pool*) cromatográfico foram respectivamente agrupadas.

Para a clivagem da cauda de histidina, a concentração de proteína de cada *pool* foi quantificada pelo método de Bradford e, em seguida, estas amostras foram incubadas por 8 horas à 18 °C com trombina em uma proporção de 0.5 U por 100  $\mu\text{g}$  de proteína. As amostras foram então concentradas (até aproximadamente 50  $\mu\text{M}$ ) e a monodispersidade e estado oligomérico da proteína em cada *pool* checadas por gel nativo de poliacrilamida 15 %. Para a estocagem das amostras destinadas à análise posterior, estas foram aliquotadas, congeladas através de processo

de resfriamento rápido em nitrogênio líquido e armazenadas a 18 °C. Para o descongelamento da amostra, imediatamente anterior à análise, as alíquotas eram deixadas em equilíbrio a 4 °C até sua completa liquefação. Em seguida eram centrifugadas por 10 minutos a 13000 g em centrífuga deixada a 4 °C, tendo então o sobrenadante retirado (embora não tenha sido constatada a presença de precipitados em nenhuma das amostras utilizadas nos experimentos), e a concentração de proteína medida novamente pelo método de Bradford . Só então as amostras eram levadas aos ensaios de anisotropia de fluorescência (vide 2 tópicos abaixo).

### 3.1.3 Oligonucleotídeos sintéticos

Para os estudos de anisotropia de fluorescência foram utilizados oligonucleotídeos dupla-fita sintéticos, com a fita senso contendo uma molécula de isotiocianato de fluoresceína (*fluorescein isothiocyanate* - FITC) ligada à sua extremidade 5' através de uma cadeia polimérica de fosforamidita contendo seis carbonos. Estes HREs marcados com fluoresceína serão referidos, no texto que se segue, pela denotação F-HRE.

As fitas únicas de oligonucleotídeo sintético, purificadas por cromatografia líquida de alta pressão, foram obtidas da Alpha DNA (Canadá). Suas concentrações foram calculadas através do coeficiente de extinção molar a 260 nm. Os F-HREs dupla-fita foram preparados como se segue. Primeiramente, as sequências das fitas senso e anti-senso foram devidamente pareadas em soluções aquosas equimolares. Subsequentemente, este oligonucleotídeo dupla fita foi anelado em um aparato de PCR (Perkin-Elmer 480 *programmable thermocycler*, Perkin-Elmer, USA) por aquecimento a 100 °C por 10 minutos e resfriado lentamente até 25 °C. Os oligonucleotídeos anelados foram armazenados em solução aquosa a –20 °C.

Foram utilizadas nove diferentes sequências de elementos responsivos: *Half site* (HS), o meio sítio constituído de uma única sequência arquetípica AGGTCA sem repetição; DR0-5, repetições diretas destas mesma sequência com espaçamentos de 0 a 5 pares de base, respectivamente; TRE-Pal, elemento palindrômico sem espaçamento, responsável para o receptor de hormônios tireoideanos (TR) ; F2 ou IP6, palíndromo invertido com espaçamento de seis pares de base, é o elemento responsável ao qual o TR $\beta$  se liga fortemente na forma apo, selecionando

co-repressores . Foram também realizados ensaios de afinidade a um oligonucleotídeo marcado para controle negativo e na presença de um agente de competição inespecífica não marcado, conforme referido no próximo tópico. As sequências para cada um dos oligonucleotídeos utilizados é detalhada na Tabela 3.1

**Tabela 3.1** – Oligonucleotídeos utilizados neste estudo.

HS	<i>F-H-aha-5'</i> -AGCTAA <b>AGGTCA</b> GAT-3'
DR0	<i>F-H-aha-5'</i> -AGCTAA <b>AGGTCAAGGTCA</b> GTTAGGA-3'
DR1	<i>F-H-aha-5'</i> -AGCTTC <b>AGGTCA</b> GAGGTCA <b>GAG-3'</b> †
DR2	<i>F-H-aha-5'</i> -AGCTTC <b>AGGTCA</b> GAGGTCA <b>GA-3'</b>
DR3	<i>F-H-aha-5'</i> -AGCTTC <b>AGGTCA</b> CAGAGGTCA <b>G-3'</b>
DR4	<i>F-H-aha-5'</i> -AGCTTC <b>AGGTCA</b> CAGG <b>AGGTCA-3'</b> †
DR5	<i>F-H-aha-5'</i> -AGCTAA <b>AGGTCA</b> GAGTC <b>AGGTCA</b> GTTAGGA-3'
F2 (IP6)	<i>F-H-aha-5'</i> -AGCTTATT <b>TGACCC</b> CAGCTG <b>AGGTCA</b> AGTTACG-3'†
Pal	<i>F-H-aha-5'</i> -ATATT <b>AGGTCA</b> TGAC <b>CTGA</b> ATAT-3'
p53 (controle negativo)	<i>F-H-aha-5'</i> - TAATTAGGCATGTCTAGGAAA - 3'
poli AT (competidor inespecífico)	5'- (AT) <sub>12</sub> - 3'

### 3.1.4 Interação do hRXR $\alpha$ - $\Delta$ AB com elementos responsivos de DNA

A afinidade do hRXR $\alpha$ - $\Delta$ AB a diferente HREs foi analisada através de ensaios de titulação isotérmica utilizando-se como titulante uma solução concentrada da proteína (solução estoque a 50  $\mu$ M ) adicionada em quantidades gradativas (correspondendo a concentrações finais de 1 a 2800 nM) sob uma solução com concentração fixa (5 nM) do F-HRE em uma cubeta de quartzo com quatro faces transparentes e caminho ótico de 1 cm previamente termalizada a 10 °C. Todas as medidas foram realizadas em tampão Hepes 20 mM, pH 8,0 com NaCl 50 mM<sup>1</sup>, ditioreitol (DTT) 3mM e glicerol 5 % V/V. Em todos os casos, a diluição máxima foi menor que 15 % e as variações na concentração do DNA e da proteína foram sempre corrigidas para o fator de diluição. Seguido à adição da proteína, a solução foi homogenizada e após o intervalo de 2 minutos a anisotropia de fluorescência da solução na cubeta foi medida sucessivamente até a sua estabilização.

<sup>1</sup> Os ensaios nestas condições de força iônica relativamente baixa foram previamente padronizados em nosso grupo de pesquisa no trabalho de Doutorado de Rodrigo Villares Portugal . Neste trabalho, em condições de força iônica significativamente maiores, não eram visualizados platôs bem definidos para as interações do hRXR $\alpha$ - $\Delta$ AB com diferentes HREs (DR1, DR4, Pal e F2), não permitindo o correto ajuste numérico das isotermas (vide subseção relacionada aos ajustes numéricos abaixo)

O efeito do agonista ácido 9-cis retinóico (9-cRA) foi verificado através da realização dos mesmos ensaios na presença de excesso do ligante. Para tal, a solução estoque da proteína foi previamente incubada com o agonista por três horas a 4 °C com uma razão molar de 3:1 para 9-cRA:proteína e o mesmo ligante foi adicionado e homogenizado à solução de titulado na cubeta antes do ensaio, mantendo sua concentração fixa em 10  $\mu$ M (5X a concentração máxima de proteína ao longo do ensaio de afinidade).

A habilidade do hRXR $\alpha$ -DL para a discriminação entre sequências específicas e não específicas foi testada através de dois tipos de experimento:

- **Titulação de oligonucleotídeo inespecífico marcado com fluoresceína.** Nestes ensaios, a sequência alvo para o supressor de tumor p53, devidamente marcada com FITC em sua extremidade 5', foi utilizada como um controle negativo para a interação de hRXR $\alpha$ -DL em experimentos semelhantes aos realizados para os F-HREs. Esta sequência de 23 pb não contém nenhum sítio consenso para a interação com o hRXR $\alpha$ -DL.
- **Ensaios de competição com excesso de oligonucleotídeo não marcado.** Foram realizados ensaios de titulação em que as interações do hRXR-DL com os F-HREs DR1, DR4 e F2 foram respectivamente desafiadas com oligonucleotídeo não marcado específico e não específico em concentração 10 X maior (50 nM) que o oligo marcado. Os ensaios foram realizados tanto com excesso do mesmo HRE não marcado (competição específica) quanto com excesso de um oligonucleotídeo poli AT de 24 pb não marcado (competição inespecífica)

A estequiometria do complexo receptor:DNA foi verificada para três HREs cujas isotermas foram tomadas como representativas: DR1, DR4 e F2. Para tal, foram realizadas as respectivas titulações destes elementos em concentrações estequiométricas pelo hRXR $\alpha$ - $\Delta$ AB. Nestes ensaios, além do oligonucleotídeo marcado na concentração de 5 nM, a solução de titulado continha o mesmo oligonucleotídeo não marcado em concentração de 600 nM, concentração esta que superava por um fator maior que 10 X os valores de  $K_D^{app}$  previamente determinados pelo ajuste não linear (vide seção 3.1.6 e ).

Finalmente, foi verificada a influência do íon Mg<sup>2+</sup> através de ensaios na presença de MgCl<sub>2</sub> 3 mM para a afinidade a DR1, F2 e Pal.

### 3.1.5 Anisotropia de fluorescência - detalhes técnicos e considerações pré análises

Todas as medidas de fluorescência foram realizadas em um espectrofluorímetro ISS-PC1 (ISS, Champaign, IL), equipado com polarizadores e com o experimento ajustado para a geometria “L” do aparelho. Para controle da temperatura foi utilizado um banho de circulação RTE-740 (Thermo Neslab). A excitação foi realizada a 480 nm e a emissão monitorada sem o uso de monocromadores com um filtro OG515, com corte de 50 % em 515 nm. A fonte foi utilizada em 15 A.

Todos os dados foram obtidos utilizando-se pelo ao menos três diferentes repetições do experimento realizadas em dias diferentes. Os valores de anisotropia foram calculados por programa intrínseco ao equipamento através da equação . Os valores retornados foram estimados a partir de um procedimento de 5 iterações (em que as posições dos polarímetros de excitação e emissão foram variadas), retornando-se assim 5 diferentes valores de anisotropia de fluorescência para um mesmo ensaio. Os desvios absolutos para estes 5 valores foram sempre menores que 0.002 unidades arbitrárias de anisotropia (A).

Decréscimos sistemáticos na intensidade de fluorescência decorrentes da diluição e/ou de possíveis efeitos de *quenching* cinético e filtro interno , além de variações randômicas relacionadas a oscilações na amperagem da fonte durante o experimento foram monitorados durante todos os ensaios, os primeiros tendo valores máximos correspondentes a menos de 15 % da intensidade inicial (após o término do experimento) e as segundas raramente observadas, apresentando valores máximos próximos a 5 % entre duas medidas consecutivas. Embora o valor de anisotropia não seja diretamente dependente do valor absoluto da intensidade de fluorescência, variações significativas destas mesmas intensidades podem estar relacionadas a alterações do sinal devido a fenômenos de espalhamento da luz incidente (isotrópico, por natureza) ou a efeitos de *quenching* estático do fluoróforo (interações intermoleculares estáveis do titulante com a

sonda fluorescente, de forma a distorcer a geometria molecular do estado fundamental e/ou do estado excitado) , o que por sua vez poderia implicar em mudanças no tempo médio de relaxamento do estado excitado ( $\tau$ ). Ambos são efeitos passíveis de alterar os valores de anisotropia obtidos (vide equação) . O primeiro efeito (espalhamento isotrópico) pode ser considerado como desprezível, devido à presença do filtro "passa alta" OG515.

De forma a certificar quanto a não ocorrência de mudanças significativas no ambiente químico do fluoróforo, os espectros de excitação para diferentes comprimentos de emissão e os de emissão para diferentes comprimentos de excitação foram monitorados para as duas condições extremas de um mesmo experimento: tanto para as soluções iniciais de F-HRE a 5 nM quanto para as soluções finais após a saturação do oligo marcado com a proteína (não mostrado). Estas pré-análises foram realizadas tanto no início, quanto no fim de um mesmo dia de experimentos, de forma a se verificar também quanto a estabilidade da sonda. Também a influência do ligante 9-cRA foi verificada, realizando-se os mesmos ensaios acima mencionados na ausência e na presença do agonista na concentração utilizada nos experimentos (10  $\mu\text{M}$ ). Para todas as condições experimentais e ao longo do intervalo de um dia de experimentos, embora tenham sido observadas variações nos valores absolutos das intensidades (condizentes com os pequenos decréscimos de intensidade ocorridos ao longo dos experimentos de titulação), não foram verificadas mudanças significativas quer para os centros de massa dos espectros de absorção e de emissão do fluoróforo (*blue* ou *redshift*) quer para as intensidades relativas para diferentes comprimentos de onda  $\lambda$  (os perfis dos espectros obtidos não apresentaram distorções). Desta forma, a interação da proteína com o oligonucleotídeo marcado não apresentou evidências de mudanças nos momentos dipolo de absorção ( $\mu$ ) e/ou de emissão ( $v$ ) da fluoresceína, diminuindo consideravelmente a probabilidade de que as pequenas variações observadas para as intensidades de fluorescência estivessem relacionadas a distorções na geometria molecular dos estados fundamental e/ou excitado do fluoróforo ao longo dos experimentos. Assim sendo, foi descartada a hipótese de mudanças significativas no valor de  $\tau$ , passíveis de influenciarem visivelmente nas medidas de anisotropia obtidas na faixa de concentrações utilizadas..

O outro fator passível de alterar o valor de anisotropia da solução sob temperatura constante (sem mudanças significativas no volume molar ou na mobilidade intrínseca da molécula e/ou do fluoróforo e sem mudanças no valor de  $\tau$ ) seria decorrente de variações na viscosidade ( $\eta$ ) da solução (equação) devido à variação na concentração de macromoléculas em solução. Contudo, a concentração máxima alcançada para o hRXR-DL durante os experimentos de titulação foi de  $2.4 \mu\text{M}$ , correspondendo a aproximadamente  $0.1 \text{ mg}\cdot\text{ml}^{-1}$  de proteína. A dissolução de proteínas globulares só começa a afetar significativamente a viscosidade de soluções aquosas a partir de concentrações maiores que  $10 \text{ mg}\cdot\text{ml}^{-1}$ , podendo então o parâmetro  $\eta$  ser tomado como constante ao longo do ensaio.

Desta forma, conforme a equação , os valores de anisotropia obtidos puderam ser devidamente interpretados como decorrentes de mudanças nos valores do tempo médio de correlação rotacional do fluoróforo ( $\theta_h$ ) e para a molécula como um todo (vide equação ), decorrentes da associação da proteína com o F-HRE e permitindo que sua covariância com a concentração de hRXR-DL em solução fosse associada à fração de formação do complexo, conforme as diferentes modelagens expostas na próxima subseção e posteriormente discutidas em 4.1.1.

### **3.1.6 Análise termodinâmica quantitativa das isotermas de interação**

As isotermas de interação foram independentemente analisadas seguindo as duas diferentes abordagens de ajuste não linear comentadas nas próximas duas subseções. Os parâmetros recuperados através destas duas abordagens foram então devidamente comparados, de forma a melhor inferir sobre os mecanismos de associação e alosterismo do hRXR-DL para os diferentes HREs.

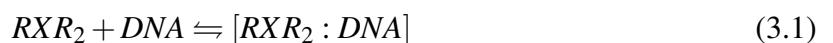
#### **Ajuste não linear com derivação de expressão analítica**

Esta é a forma de análise de dados mais comumente utilizada em estudos de isotermas de interação para processos bioquímicos no equilíbrio . Os dados são ajustados numericamente a uma única expressão analítica "fechada" que descreve a fração de formação do complexo e consequente saturação do titulado segundo um modelo de escolha. Embora inúmeros pro-

cedimentos de linearização existam, as abordagens que recuperam parâmetros com menores margens de erros (e portanto as mais utilizadas) são ajustes não lineares das variáveis independente e resposta contidas nos dados experimentais (aqui, representadas pela concentração de proteína em solução e a intensidade anisotrópica para cada concentração, respectivamente) e dos parâmetros a serem recuperados (neste caso, a constante de dissociação do complexo ( $K_d$ ) e algum parâmetro referente à colaboratividade do processo para modelos que admitam a existência de um ou mais intermediários) além de parâmetros secundários, podendo estes serem mantidos como fixos (por exemplo, a concentração de DNA titulado, fixa a 5 nM durante o experimento e, portanto, tida como uma constante durante o ajuste para o modelo de dois estados (vide abaixo)) ou flutuantes (neste caso, representados pelos valores de anisotropia para o platô inicial [fração de ligação = 0] e final [fração de ligação = 1], que, devido à sua natureza assimptótica, são tomados como ajustáveis durante a modelagem). A equação não linear é adequada ao conjunto de dados através de sucessivas aproximações numéricas, geralmente através de algum algoritmo de minimização de função de desvios quadráticos. Dentre estes, o algoritmo de Maquardt-Levenberg se encontra entre os mais utilizados pelos softwares convencionais e é, em particular, o algoritmo empregado nos dois softwares utilizados neste estudo.

Para a análise não linear foram considerados dois modelos: o modelo simples de equilíbrio entre dois estados reversíveis para o dímero de hRXR-DL ( $RXR_2$ ) e o DNA e o formalismo de Hill para as isotermas que apresentaram comportamento cooperativo evidente. Ambos os ajustes foram realizados, independentemente, utilizando o programa SigmaPlot

O modelo de dois estados admite o equilíbrio simples entre hRXR dimérico e DNA livres em solução de um lado da reação e o complexo entre o dímero e o DNA do outro:



Desta forma, a constante de dissociação  $K_d$  é definida de acordo com

$$K_d = \frac{RXR_2 * D}{[RXR_2 : D]} \quad (3.2)$$

Onde D é a concentração molar de DNA dupla fita livre e  $[RXR_2:D]$  é a concentração molar do complexo formado entre o dímero de RXR e o DNA.

A formação do complexo em solução pode ser relacionada aos dados experimentais por

$$[RXR_2 : D] = Dt * \alpha \quad (3.3)$$

Onde  $Dt$  é a concentração total de DNA (livre e ligado) e  $\alpha$  é a fração de DNA (ou  $RXR_2$ ) ligada, que, por sua vez, é relacionada às medidas de anisotropia por

$$\alpha = (A_{obs} - A_i) / (A_f - A_i) \quad (3.4)$$

sendo  $A_{obs}$  a intensidade anisotrópica observada, e  $A_i$  e  $A_f$ , respectivamente, os limites assintóticos inferior e superior para os valores de anisotropia obtidos através do ajuste das isotermas. Segue-se então que

$$A_{obs} = A_i + (A_f - A_i) * \frac{\sqrt{(K_d + RXR_2 t + Dt) - ((K_d + RXR_2 t + Dt)^2 - 4RXR_2 t * Dt)}}{2Dt} \quad (3.5)$$

onde  $RXR_2 t$  é a concentração total de hRXR dimérico, de forma análoga a  $Dt$  para a concentração total de DNA.

A equação 3.5 é a função não linear para a qual os dados são ajustados no modelo de dois estados, embora ela apresente vantagens quanto à sua simplicidade

### **3.2 Metodologia dos estudos de dinâmica molecular para o alosterismo na interação do hRXR $\alpha$ -DBD a diferentes elementos responsivos no DNA**

Todas as simulações apresentadas neste capítulo foram executadas utilizando o programa NAMD . Foram utilizadas topologias do tipo CHARMM22 para todos os átomos (21). O algoritmo integrado

### **3.3 Metodologia para o estudo integrado por SAXS e modelagem estrutural baseada em dinâmica molecular para a celobiohidrolase I de *Trichoderma harzianum* em solução**

Sendo este um trabalho em colaboração, a produção da enzima por fermentação, sua extração e purificação e os ensaios de SAXS foram realizados pelo grupo do Prof. Dr. Igor Polikarpov, Instituto de Física de São Carlos, Universidade de São Paulo. Também os primeiros tratamentos dos dados de SAXS (análise de Guinier, geração do modelo de átomos dummy e modelagem de corpo rígido com os domínios gerados por homologia) foram desenvolvidos pelo pesquisador Dr. Mário de Oliveira Neto e pelo aluno de mestrado Flávio Rodolfo Rossetto, do mesmo grupo, embora o presente autor tenha acompanhado a execução destas análises e tenha frequentemente voltado a elas durante o decorrer das etapas de modelagem e dinâmica molecular. Assim sendo, estas etapas onde a participação do autor foi menos ativa serão descritas de forma mais sucinta nas próximas três subseções, dando-se maior ênfase às etapas de modelagem estrutural e dinâmica molecular para o refinamento dos modelos obtidos e melhor compreensão da dinâmica dos movimentos coletivos entre os dois domínios da proteína baseado nos resultados de SAXS.

#### **3.3.1 Obtenção da amostra**

Basicamente, a proteína foi expressa diretamente no fungo (expressão homóloga) em um bioreator com volume operante de 5 L, com agitação contínua. O nível de oxigênio, o pH e a temperatura foram devidamente regulados para 60 %, pH 5.0 e 28°C, respectivamente,

e a produção da enzima foi induzida pela adição de celulose microcristalina (Avicel). Após a produção, o extrato extracelular foi filtrado, concentrado e a proteína purificada por troca iônica. O protocolo completo das etapas de fermentação, extração e purificação está descrito no artigo

### **3.3.2 ensaios de SAXS**

Os ensaios de SAXS foram realizados no Laboratório Nacional de Luz Síncrontom (Campinas, Brazil). Os dados foram coletados para a proteína e para seu domínio catalítico em concentrações

### **3.3.3 Análise dos dados de SAXS**

## **3.4 Metodologia da modelagem - criação de um modelo para a proteína completa, dinâmica molecular e comparação com os dados de SAXS**

### **3.4.1 Construção de um modelo de alta resolução através do programa I-TASSER**

Um modelo em alta resolução para a molécula de ThCel7A em sua completa extensão e totalmente glicosilada foi gerado utilizando como base o modelo de corpo rígido recuperado no último passo do tratamento dos dados de SAXS. Para a geração deste modelo, foi utilizada a ferramenta computacional I-TASSER (Iterative Threading/ASSEmly/Refinement) (47). Este pacote de programas recupera modelos acurados para a estrutura tridimensional de proteínas através de uma sequência multi-etapas constituída basicamente dos três seguintes passos:

- **Alinhamento multi-fragmentos (“multiple threading alignment”).** Diferentes janelas da sequência primária da proteína são alinhadas contra uma base de dados não redundante de estruturas e sequências (não apresentando, na base original, sequências com homologia acima de um dado valor limite) (48, 49). Quatro algoritmos diferentes categorizam cada resíduo dos fragmentos alvo quanto à probabilidade de se encontrarem em um dado contexto de estrutura super-secundária hélice/fita/desordenado. Estas probabilidades são

relacionadas aos perfis de "vizinhanças" entre resíduos de aminoácidos nas sequências primárias alvo, comparados a estes mesmos perfis nas sequências molde com estruturas conhecidas na base de dados (este procedimento é denominado alinhamento perfil - perfil (profile-profile alignment - PPA)). As probabilidades para os diferentes consensos de estrutura calculados por cada algoritmo são somadas e, ao fim deste passo, dois algoritmos de programação dinâmica são usados para identificar o melhor alinhamento entre os fragmentos alvo e as sequências molde. Ao término desta etapa, estas diferentes regiões da sequência alvo estarão associadas a contextos distintos de estrutura super-secundária com respectivas probabilidades (que por sua vez serão tanto maiores quanto maior for a significância dos alinhamentos no PPA).

- **Agrupamento (“assembly”).** Nesta etapa são efetuadas tanto a reunião dos fragmentos gerados no passo anterior em contextos maiores de estrutura terciária, quanto a modelagem *ab initio* das regiões não alinhadas (principalmente regiões de estrutura desordenada). A geração de modelos completos tridimensionais é efetuada através de simulações de Monte Carlo do tipo troca entre réplicas. Os campos de força são descritos a partir de uma série de potenciais empíricos descrevendo a propensão de interação entre cadeias laterais (SC), correlações entre carbonos  $\alpha$  ( $C\alpha$ ) a curtas distâncias, propensão a estruturas secundárias consenso, predição resíduo-baseada por rede neural da acessibilidade das cadeias laterais ao solvente, otimização das distâncias e ângulos para as pontes de hidrogênio para motivos específicos de estruturas secundárias e super-secundárias, além das constrições para contatos entre as SC e distâncias entre os  $C\alpha$  recuperadas no alinhamento multi-fragmentos da etapa anterior. Além disto, podem ser adicionadas restrições “externas” para contatos, distâncias e estruturas moldes pelo usuário. Os potenciais empíricos utilizados neste passo são descritos em (50). Em casos extremos onde nenhuma sequência molde tenha sido propriamente identificada na primeira etapa, a molécula inteira é modelada *ab initio* a partir deste passo utilizando os potenciais empíricos (e/ou as restrições externas) acima descritos. As conformações com aparentes menores energias livres são identificadas através da clusterização dos principais confôrmeros recuperados

na simulação de Monte Carlo.

- **Refinamento (“refinement”).** Na terceira e última etapa, o agrupamento por troca de réplicas é realizado novamente a partir dos cinco principais confôrmeros obtidos no passo anterior. Além dos potenciais acima descritos, são utilizadas como constrições as distâncias internas dos confôrmeros nas centróides dos *clusters* recuperados no último passo. Esta etapa de reagrupamento tem por finalidade principal a eliminação de impedimentos estéricos. Ao fim deste novo processo de agrupamento, são recuperados até cinco modelos (correspondentes às centróides de cinco *clusters*). As cadeias laterais são então adicionadas na íntegra (até então, os aminoácidos são representados como átomos “unidos” para os  $C\alpha$ , os  $C\beta$  e para o centro de massa das SC) e os rotâmeros são otimizados em um último passo do refinamento.

A validade dos modelos recuperados pelo programa I-TASSER é avaliada pela função *score* de confidênciia (C-score):

(3.6)

Onde  $M$  é a multiplicidade do *cluster* referente ao modelo;  $M_{tot}$  é a multiplicidade total de estruturas (para todos os *clusters*);  $\langle rmsd \rangle$  é a média sobre o desvio quadrático médio (*root mean square deviation*) para os átomos de todos os confôrmeros à centróide do *cluster*;  $Z(i)$  é o mais alto valor da função Z-score (a energia em relação à média em unidades de desvio padrão) dos fragmentos molde gerados pelo  $i$ -ésimo algorítimo de PPA (entre os quatro algorítimos utilizados) e  $Z_0(i)$  é um valor de corte para a função Z-score específico para cada algorítimo. Os dois primeiros fatores da equação 3.6 referem-se ao grau de convergência estrutural no algorítimo de clusterização, que por sua vez se correlaciona à consistência das restrições para contatos e distâncias oriundas do primeiro passo ou adicionadas pelo usuário e dos potenciais intrínsecos do programa. O terceiro fator é uma medida da qualidade do passo inicial de alinhamento multi-fragmentos. O logarítmico na equação 3.6 serve para ajustar os valores de C-score a uma distribuição aproximadamente simétrica (normal). O valor de C-score varia tipicamente

de -5 (para um modelo extremamente incurado) a 2 (para um modelo extremamente acurado). Esta função de *score* foi extensamente textada pelos autores do programa (47), mostrando ter uma forte correlação com a qualidade dos modelos preditos. Para valores de C-score acima de -1.5, a taxa de insucessos durante a fase de testes do pacote (falsos positivos ou negativos para sequências de estruturas conhecidas) foi menor que 0.1.

Em nossa modelagem da ThCel7A, a estrutura dos dois domínios posicionados sob o DAM no modelo de corpo rígido foi adicionada como uma restrição externa para o cálculo do I-TASSER. Também foram adicionadas restrições de distância relacionadas às pontes de hidrogênio das folhas  $\beta$  do CBM, bem como para as duas pontes dissulfeto deste domínio. Estas restrições para o CBM mostraram-se necessárias para a recuperação de um modelo topologicamente acurado para o sistema CCD-LP-CBM. Sendo o CBM significativamente menor e relativamente menos estruturado que o CCD, além de ser vizinho a uma outra sequência de 22 resíduos pobre em informação estrutural (LP), em tentativas prévias sem restrição os algorítimos do pacote demonstraram dificuldades na recuperação de conformações tridimensionais acuradas e bem individualizadas para os dois domínios em uma mesma estrutura.

As seguintes estruturas foram usadas como molde pelo programa (os códigos do PDB são mostrados entre parênteses): O domínio CCD da celobiohidrolase I de *Trichoderma reesei* (1celA); o domínio CCD de *Hypocrea jecorina* (CEL7A); as celobiohidrolases de *Phanerochaet chrysosporium* (1gpiA), de *Talaromyces emersoni* (1q9hA), e de *Melanocarpus albomyces* (2rfwA).

### **3.4.2 Adição das glicosilações**

Não existem dados cristalográficos para a real extensão das cadeias de carboidratos em cada sítio. Desta forma, as O-glicosilações do peptídeo conector e as N-glicosilações do domínio catalítico foram adicionadas seguindo as sugestões de estudos prévios de espectrometria de massa e cromatografia de troca iônica na literatura (51, 52). Embora estes estudos tenham sido focados na CBHI de *T. reesei* (TrCel7A), eles se apresentam como estimativas razoáveis para as mesmas glicosilações em ThCel7A, dada a alta homologia dentre as sequências dos sítios de

glicosilação destas duas enzimas. Desta forma, foi adicionado um único resíduo de  $\beta$ -D-manoze O-ligado a cada resíduo de serina e adicionadas cadeias lineares contendo dois ou três resíduos da mesma  $\beta$ -D-manoze O-ligadas aos resíduos de treonina do LP (o tamanho da extensão glicosídica para as treoninas foi maior ou menor conforme estas tivessem como vizinho N-terminal um outro resíduo glicosilado ou não, respectivamente). De forma similar, foram adicionados segmentos ramificados de  $(\alpha$ -D-glicose)<sub>1</sub>-( $\alpha$ -D-manoze)<sub>6</sub>- $\beta$ -D-manoze-(beta-D-N-acetil-glicosamina)<sub>2</sub> (de agora em diante, denotados apena por  $(Glc\alpha)_1-(Man\alpha)_6-Man\beta-(GlcNac\beta)_2$ ) N-ligados aos resíduos de asparagina 126, 283 e 397, ambos sítios de N-glicosilação preditos para o CCD de *T. harzianum* e estruturalmente homólogos aos sítio Asn45, Asn270 e Asn384 em *T. reesei*. Os sítios de glicosilação e a estrutura das cadeias são melhor ilustrados na figura

As extensões polissacarídicas foram adicionadas utilizando a ferramenta computacional do servidor GLYCAM , que por sua vez utiliza topologias do tipo AMBER para otimizar a geometria das mesmas.

O LP de ThCel7A apresenta cinco resíduos de serina e seis de treonina, tendo portanto um possível sítio extra de O-glicosilação comparado a TrCel7A (com oito resíduos de treonina e apenas dois de serina). Contudo, um dos aparentes sítios de O-glicosilação, o resíduo Thr470 de ThCel7A, pertence à sequência na interface entre o LP e o CBM. A única estrutura em alta resolução disponível para o TrCBM (que serviu de molde para a predição da estrutura homóloga do ThCBM neste estudo) foi obtida por RMN para um domínio sintetizado in vitro e portanto não glicosilado. A estrutura resolvida para este domínio sintético apresenta este resíduo de treonina interagindo diretamente com o cerne do CBM. Assim, em nosso modelo, não foram adicionadas extensões O-glicosídicas para a Thr470, visto que preferimos correr o risco de não glicosilar um resíduo originalmente contendo esta cadeia a adicionar uma falsa extensão sacarídica a uma região com o potencial de interagir diretamente com o cerne do CBM.

As extensões glicosídicas em nosso modelo foram utilizadas de modo a simular da forma mais acurada o possível (considerando as limitações das informações estruturais sob as quais o modelo foi construído) o ambiente químico e as restrições impostas pela múltiplas cadeias de carboidratos à conformação e dinâmica dos dois domínios. Este modelo não tem, no entanto,

a pretensão de ilustrar uma composição precisa para o teor das glicosilações para a CBH-I de *T. harzianum* visto carecerem, para tal, suficientes dados experimentais. Uma estratégia similar foi previamente utilizada em dois estudos computacionais na construção de um modelo tridimensional para a estrutura completa da celobiohidrolase I de *T. reesei*, contendo as O-glicosilações do LP (mas sem as N-glicosilações do CCD) e ligada à celulose (53, 54).

### **3.4.3 Metodologia da simulação de dinâmica molecular**

A estrutura gerada através das etapas de modelagem acima mencionadas foi submetida a procedimentos de minimização, equilíbrio e anelamento simulado por dinâmica molecular (SA/MD).

#### **Campos de Força**

Foram utilizados campos de força AMBER para descrever o sistema. A proteína e as águas foram modeladas pelo campo ff99SB e os resíduos de carboidratos pelo campo GLYCAM06. (55, 56) Estes campos de força consideram todos os átomos explicitamente. As moléculas de água foram representadas utilizando a implementação do modelo TIP3P para o AMBER7 (57)

#### **Configurações iniciais**

Os átomos da proteína tiveram suas cargas padronizadas para o pH 7.0, resultando em uma carga líquida de -16. As extensões glicosídicas eram essencialmente neutras, não contribuindo portanto para a carga total do sistema. Foram adicionados 32 íons sódio e 16 íons cloreto, interagindo com a molécula de acordo com a sua superfície potencial, de forma a se alcançar a neutralidade. Todos os 24 resíduos de cisteína envolvidos em pontes dissulfeto (20 no CCD e 4 para o CBM) foram devidamente "assinalados" como tal e as geometrias destas pontes devidamente otimizadas durante as etapas de minimização descritas abaixo. O sistema foi colocado em uma caixa retangular de moléculas de água com dimensões iniciais de 128.1 por 100.0 por 91.0 Å e todas as moléculas de água que se sobreponham a átomos pesados foram deletadas.

Desta forma, a simulação resultante continha 154152 átomos no total, com 48563 moléculas de água, 6869 átomos para a cadeia protéica e 1057 átomos para as extensões glicosídicas. A força iônica do sistema (calculada pelo número de íons dividido pelo volume inicial) era próxima a 50 mM, similar à dos experimentos de SAXS.

### **Condições da simulação**

A simulação foi conduzida no ensemble NPT utilizando-se condições periódicas de contorno. O ensemble NPT foi mantido utilizando-se de um termostato de Langevin com coeficiente de fricção de  $2 \text{ ps}^{-1}$ . Embora a temperatura de equilíbrio do sistema tenha sido 300 K para toda a simulação, esta mesma temperatura foi extensamente variada durante o protocolo de SA/MD 3.4.3. A pressão do sistema foi mantida a 1 atm, usando o algorítmo de Nosé-Hoover com a flutuação do piston controlada por dinâmica de langevin. O período de oscilação e o tempo de decaimento do piston foram, respectivamente, de 100 fs e 50 fs.

A integração numérica foi realizada utilizando do algorítmo Velocity-Verlet implementado no pacote NAMD, com passos de integração de 2 fs e tendo as coordenadas da simulação coletadas a cada 4 ps. As ligações covalentes envolvendo átomos de hidrogênio foram mantidas fixas em suas posições de equilíbrio utilizando do algorítimo de constrição SHAKE e as moléculas de água TIP3P tiveram tanto os ângulos H-O-H quanto as distâncias O-H mantidos fixos através do algorítimo SETTLE

Para todos os passos da simulação, as interações não ligadas foram truncadas a partir da distância de 14 Å.

### **Preparação do sistema**

Uma vez que a estrutura inicial foi gerada a partir de subsequentes passos de modelagem, um procedimento gradual e extensivo de minimização e equilíbrio foi realizado antes da etapa produtiva da simulação por SA/MD. Estas etapas de minimização/equilíbrio foram todas executadas em  $T = 300 \text{ K}$  e  $P = 1 \text{ atm}$ .

- 1. Minimização por 60000 passos de gradiente conjugado (CG) com restrição harmônica de expoente 2 para todos os átomos da proteína e carboidratos;
- 2. 72 ps de dinâmica com restrição harmônica para a proteína e carboidratos;
- 3. Nova minimização sem restrição por 20000 passos de gradiente conjugado (CG);
- 4. Nova dinâmica com restrição harmônica por mais 60 ps;
- 5. Dinâmica sem restrição para os carboidratos e as cadeias laterais dos resíduos glicosilados por 300 ps;
- 6. Desativação das restrições também para o LP (resíduos 445 a 472) e dinâmica por mais 200 ps;
- 7. Nova ativação das restrições para os resíduos 445 a 467, deixando a região N-terminal do LP e o CBM (resíduos 468 a 505) relaxarem por 52 ps.

O protocolo acima foi montado de forma a permitir o relaxamento das águas e íons, bem como um relaxamento gradual das regiões da molécula com maior grau de modelagem (sem uma estrutura homóloga de alta resolução resolvida) - o LP e as extensões glicosídicas.

### **Anelamento simulado por dinâmica molecular**

Um protocolo de anelamento simulado por dinâmica molecular (SA/MD) com ciclos de aquecimento e resfriamento relativamente rápidos foi realizado após as etapas de minimização e otimização de geometria descritas na última subseção. O objetivo deste protocolo foi amostrar um número significativo de conformações com diferentes extensões do LP, a princípio a região mais móvel da molécula. Em uma etapa subsequente, as energias envolvidas na manutenção das diferentes conformações do LP foram analisadas e utilizadas no refinamento do modelo de SAXS. Apenas foram analisadas as conformações notoriamente em um platô de temperatura/energia/mobilidade próximo a 300 K.

As configurações iniciais para o procedimento de SA/MD foram tomadas das últimas coordenadas e parâmetros de célula periódica obtidos da etapa de MD restrita, sendo as velocidades randomicamente reiniciadas conforme uma distribuição de Boltzmann para uma temperatura de 300 K. Os ciclos de aquecimento e resfriamento foram executados através da reatribuição periódica das temperaturas do termostato de Langevin e do barostato (de forma a se manter a configuração NPT) a uma taxa de  $5 \text{ K}\text{ps}^{-1}$  até se alcançar a temperatura alvo. Não foram utilizadas restrições estruturais de qualquer tipo, dada a rápida duração dos ciclos de aquecimento e resfriamento, a faixa relativamente estreita de temperaturas utilizada (as temperaturas mínima e máxima foram respectivamente de 250 e 550 K, comparado a temperaturas máximas próximas a 1000 K comumente usadas para simulações do tipo SA/MD com restrição (como por exemplo no estudo de (58))), bem como o alto teor de estrutura secundária e de pontes dissulfeto para os dois domínios globulares. Desta forma, a proteína é suficientemente estável, comparada ao incremento adicionado à energia térmica durante nossa simulação, sendo as maiores modificações esperadas apenas para as regiões naturalmente mais móveis. Durante toda a simulação, o plot de Ramachandran por resíduo e o mapa de contatos para a proteína foram checados, não tendo sido notados indícios de desnaturação.

A simulação de SA/MD foi dividida em três etapas principais, separadas por dois passos em que o sistema foi aquecido de 300 a 550 K e resfriado novamente.

- **Primeira etapa.** Esta etapa foi subdividida em dois estágios a 300 K (de 540 e 366 ps, respectivamente) separados por um ”degrau” de 50 ps a 425 K (não contando os tempos de aquecimento e resfriamento) e 4000 passos de minimização por CG entre o passo de 425 K e o início do segundo estágio.
- **Primeiro passo de aquecimento 300-550 K.** Após esta etapa, o sistema foi aquecido até 550 K, mantido nesta temperatura por apenas 4 ps e, então, gradualmente resfriado até 300 K, com passos a 460, 425 e 350 K (30, 50 e 130 ps, respectivamente);
- **Segunda etapa.** Após o resfriamento do primeiro passo de aquecimento principal (300-550 K), foi permitido ao sistema explorar uma faixa mais estreita de temperaturas, bem

como temperaturas menores que 300 K, de forma a acessar modificações conformacionais menores e favorecer estruturas mais compactas. Assim, a simulação prosseguiu a 300 K por 300 ps e então o sistema foi submetido a uma sequência de passos de aquecimento e resfriamento curtos: 320 K por 60 ps, 280 K por 200 ps, 320 K por 40 ps, 250 K por 120 ps, 280 K por 200 ps e, finalmente, aquecido de volta a 300 K e simulado por mais 200 ps.

- **Segundo passo de aquecimento 300-550 K.** Após o segundo ciclo, o sistema foi novamente aquecido de 300 a 550 K, mantido neta temperatura por 80 ps e resfriado novamente até 310 K.
- **Terceira etapa.** Nesta última etapa, o sistema foi mantido a 310 K por 500 ps e finalmente resfriado até 300 K e simulado por mais 500 ps.

Somando os tempos de aquecimento, resfriamento e dinâmica, foram executados 3985 ps de simulação.

### **Análise dos resultados de SA/MD e comparação com os dados de SAXS**

platôs de energia e mobilidade da simulação foram identificados como Zonas metaestáveis, a partir das análises da energia total da molécula, do rmsd sobre a média conformacional da simulação e do rmsd progressivo a cada 100 ps (relativo à conformação 100 ps anterior). Para os cálculos de rmsd, as estruturas da trajetória foram superpostas espacialmente à conformação inicial, tomando-se como referência as regiões menos móveis da molécula (em nossa simulação, esta região foi representada pelo CCD -resíduos 16 a 443- não levando-se em conta o *loop* contendo os resíduos 414 a 426, na boca do túnel catalítico, que se mostrou significativamente mais móvel que o cerne do domínio). As zonas metaestáveis identificadas foram tomadas como *decoys* do protocolo de anelamento e analisadas quanto à sua conformação e comportamento dinâmico médios.

A análise dos resultados da simulação foi focada basicamente nos seguintes parâmetros, seu inter-relacionamento e com os dados de SAXS:

- Os valores médios por conformação para o raio de giro ( $R_g$ ) e para a distância entre o centro geométrico da superfície "inferior" do CCD e o centro de massa do CBM ( $D_{CCD-CBM}$ )
- O conjunto de interações relacionadas a diferentes extensões do LP (as interações "não ligadas" entre as cadeias de O-glicosídeos do LP, proteína-proteína e O-glicosídeos-proteína; hidratação dos resíduos de aminoácido e de O-glicosídeos e interação com os íons em solução para o LP e vizinhança);
- O número médio de interações de hidrogênio para o *backbone* do LP;
- A densidade das camadas de hidratação para o LP e vizinhança em diferentes extensões;
- A atividade vibracional do LP e para a molécula como um todo, checadas a partir de análise de componentes principais e por estimativas da entropia e entalpia vibracionais e capacidade calorífica para diferentes conformações através da abordagem quasiharmônica desenvolvida por Andricioaei e Karplus

As energias foram calculadas utilizando o pacote NAMD e o pacote VMD como interface gráfica. (59) As estruturas foram analizadas também utilizando-se o pacote VMD ou o pacote PyMOL. Os valores de rmsd,  $R_g$  e  $D_{CCD-CBM}$  foram calculados utilizando o pacote Wordom (60). Também a geração e diagonalização das matrizes  $3Nx3N$  para as análises de componentes principais e quasiharmônica foram efetuadas utilizando este mesmo programa.

Uma vez analisados os parâmetros acima mencionados, as dimensões moleculares ( $R_g$ ,  $D_{CCD-CBM}$  e  $D_{Max}$ ) devidamente ponderadas quanto as energias previamente calculadas foram comparadas às mesmas dimensões recuperadas no experimento de SAXS. Também as principais conformações para a molécula acessadas durante o SA/MD foram fitadas contra o DAM recuperado no tratamento de SAXS e a relação entre a discrepância do fitting e a energia calculada para a conformação, bem como a mobilidade das regiões melhor adequadas à extensão do envelope molecular foram avaliadas de forma a se inferir quanto à acurácia do modelo final obtido.

## **4    *Resultados e Discussões I - Bases Moleculares para o Alosterismo da Interação do Receptor Nuclear do Ácido 9-cis-Retinóico com Elementos Responsivos no DNA***

### **4.1    Estudos de Interação do Receptor de Ácido 9-cis-Retinóico a Elementos Responsivos de DNA por Anisotropia de Fluorescência**

#### **4.1.1    Promiscuidade e presença de interações inespecíficas no reconhecimento de elementos responsivos pelo hRXR $\alpha$ - $\Delta$ AB**

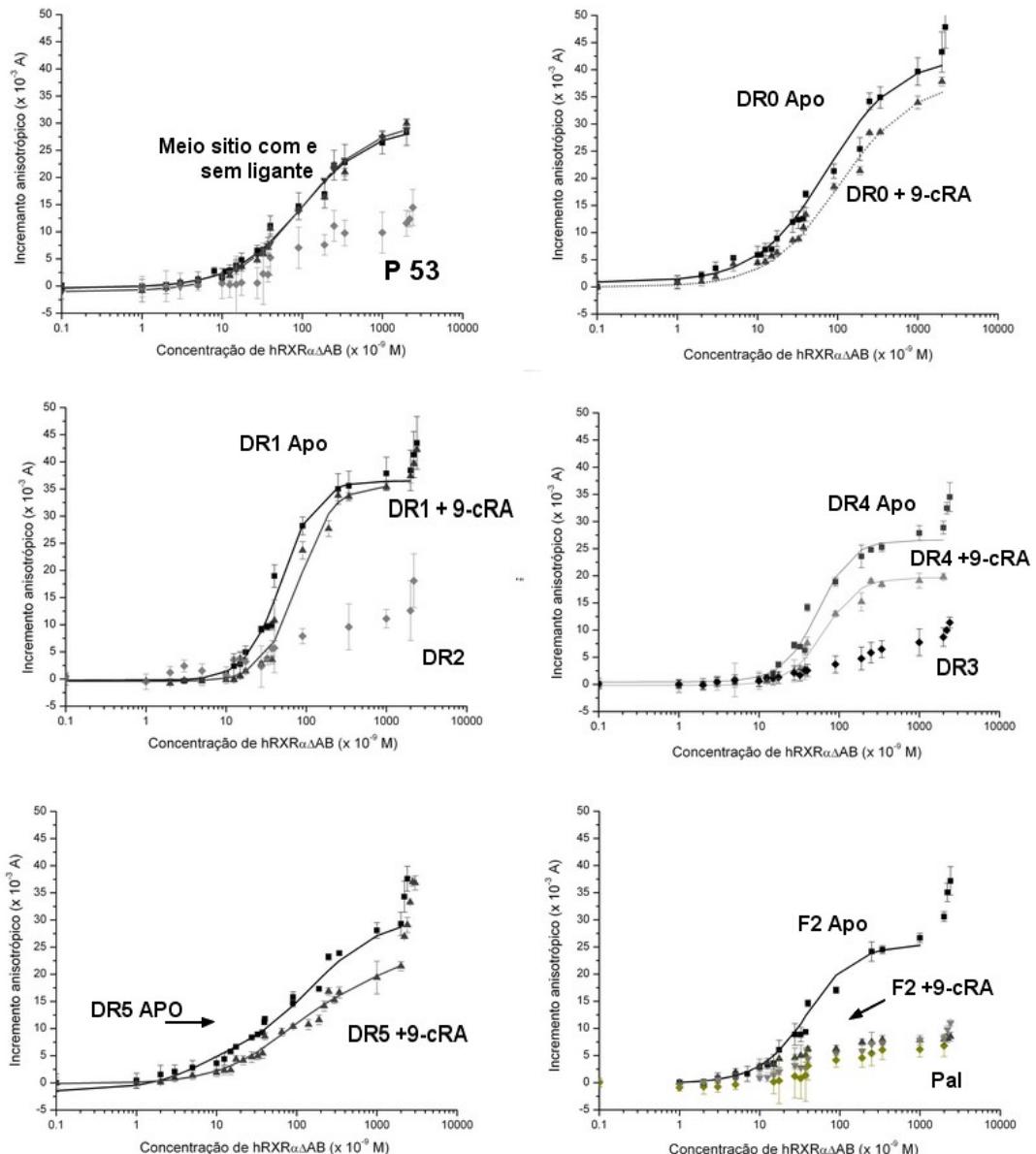
A técnica de anisotropia de fluorescência foi utilizada para caracterizar a interação da proteína hRXR $\alpha$ - $\Delta$ AB com diferentes sequências de DNA responsivas a hormônios (HREs). A adição de concentrações crescentes de proteína à solução contendo o oligo marcado promoveu um incremento significativo na intensidade anisotrópica (Fig. 4.1), indicando um aumento no tempo de correlação rotacional da molécula decorrente da formação do complexo proteína:DNA.

Em uma primeira análise, nota-se a presença de pelo ao menos três diferentes tipos de complexos hRXR:DNA formados com a titulação do DNA marcado pela proteína. Para a sequência alvo do supressor de tumor p53 (inespecífica para o hRXR $\alpha$ ) e para quatro elementos responsivos não específicos (DR2, DR3, Pal e F2 na presença de ligante), o valor de anisotropia cresce pouco em face dos desvios entre as medidas, com aparentes termos assimptóticos próximos

a 10 mA (mA = mili-unidades arbitrárias de anisotropia) após a adição de hRXR $\alpha$ -ΔAB em concentrações entre 100 e 1000 nM (Fig. 4.1-A, C, D e F e Fig. 4.2-A). Estas curvas apresentaram um perfil relativamente inclinado sobre o eixo da variável independente em escala logarítmica e o aparente segundo platô é pouco pronunciado, considerando o incremento anisotrópico total e os desvios relativos. No texto que se segue, as interações do hRXR $\alpha$ -ΔAB com os HREs que apresentaram este padrão de incremento anisotrópico serão referidas como I<sub>1</sub>. Para o meio-sítio (HS), DR0, DR1, DR4 e para DR5 e F2 na ausência de ligante, foram observados incrementos anisotrópicos significativamente maiores e com termos assimptóticos mais bem definidos (de um modo geral) com valores de incremento anisotrópico entre 20 e 40 mA (Fig. 4.1-A-F e Fig. 4.2-A). Estas isotermas serão referidas como I<sub>2</sub>.

Para DR5 na presença de ligante, foi verificado um comportamento aparentemente intermediário entre o das isotermas I<sub>1</sub> e I<sub>2</sub>, considerando-se em conjunto o segundo platô para o incremento anisotrópico (próximo a 22 mA) e a inclinação da curva voltada para o eixo da variável independente, com o referido segundo platô mais fracamente definido em comparação às outras curvas que apresentaram incrementos comparáveis (I<sub>2</sub>).

Para a interação do hRXR $\alpha$ -ΔAB com todos os oligonucleotídeos foi verificado um segundo padrão de ascensão do valor de anisotropia a altas concentrações ( $\geq 2 \mu\text{M}$ ) (Fig. 4.1). Este novo incremento anisotrópico não apresentou sinais de saturação, não atingindo um novo platô mesmo após a adição de hRXR $\alpha$ -ΔAB em concentrações  $\geq 4 \mu\text{M}$  (não mostrado), demonstrando ser decorrente de interações inespecíficas da proteína em altas concentrações com o DNA. O RXR $\alpha$ , como diversos outros Fatores de transcrição, apresenta uma densidade de cargas positivas relativamente alta em seu domínio de ligação ao DNA, o que favorece sua interação inespecífica com os fosfatos de carga negativa no *backbone* de ácidos nucléicos quando em altas concentrações e sob condições de baixa força iônica. Além disto, os fragmentos relativamente pequenos de oligonucleotídeos utilizados neste estudo, entre 15 e 32 pb, permitem uma promiscuidade e adaptabilidade conformacional à superfície da proteína que não seriam permissivas em fragmentos com dimensões da ordem do material genômico, possibilitando a ocorrência de interações inespecíficas que não necessariamente seriam observadas



**Figura 4.1 – Isotermas de interação do hRXR $\alpha$ - $\Delta$ AB a diferentes elementos responsivos de DNA**

em condições fisiológicas<sup>1</sup>. Interações desta natureza foram também observadas em estudos de afinidade para outros receptores nucleares por DNA através de anisotropia de fluorescência . A presença de tais associações inespecíficas não representa um problema, desde que estas possam ser acuradamente distintas das interações específicas e que apenas estas últimas sejam analisadas e consideradas quanto às implicações fisiológicas do processo. Os ajustes numéricos para as isotermas deste trabalho foram realizados considerando-se sempre os dados até onde as interações inespecíficas a altas concentrações ainda não se tornavam evidentes (até 1000 nM ou 2000 nM, conforme o perfil da isoterma). Também, a influência (e abrangência) das interações inespecíficas foi diretamente avaliada através de experimentos de competição (vide 4.1.3).

As interações decorrentes das isotermas tipo I<sub>1</sub> também demonstraram um evidente caráter inespecífico, uma vez que este tipo de isoterma é visualizada para a interação do hRXR $\alpha$ -ΔAB com a sequência alvo para o p53, que não apresenta a seqüência arquetípica 5'-AGGTCA-3'. Estas isotermas também apresentam um comportamento razoavelmente assimptótico entre as concentrações de 100 e 1000 nM, conforme mencionado previamente (embora os erros relativos tornem difícil a caracterização deste aparente platô e a acurada distinção do mesmo das interações inespecíficas a altas concentrações). Desta forma, uma primeira pergunta a ser respondida é se as interações do tipo I<sub>1</sub> são representativas de um intermediário inespecífico bem caracterizado para as curvas de maior incremento anisotrópico (por exemplo, monômeros de hRXR $\alpha$ -ΔAB se ligando inespecificamente ao DNA) ou se apenas se tratam de primeiras manifestações das interações inespecíficas visualizadas a altas concentrações em todas as curvas, tornando-se também evidentes a baixas concentrações na ausência de interações de caráter mais específico.

Visando um tratamento sistemático das interações aqui visualizadas, foram primeiramente realizados ajustes numéricos com derivação de expressão analítica (vide 3.1.6) das curvas de

---

<sup>1</sup> As dimensões dos oligonucleotídeos utilizados são ideais para a realização de experimentos de anisotropia de fluorescência. Fragmentos significativamente maiores que os utilizados tendem a apresentar tempos de correlação rotacional com ordens de grandeza consideravelmente maiores que o  $\tau$  do fluoróforo e a exibir variações pouco significativas, ou mesmo desprezíveis, na anisotropia de fluorescência após a associação com a proteína . Além disto, embora dimensões maiores aumentem a penalidade energética relacionada a conformações locais relativamente *impróprias* para a estrutura terciária do DNA (limitando sua adaptabilidade conformacional a superfícies inespecíficas) , elas também implicam em um maior número de sítios inespecíficos disponíveis, terminando por ambos os efeitos se compensarem

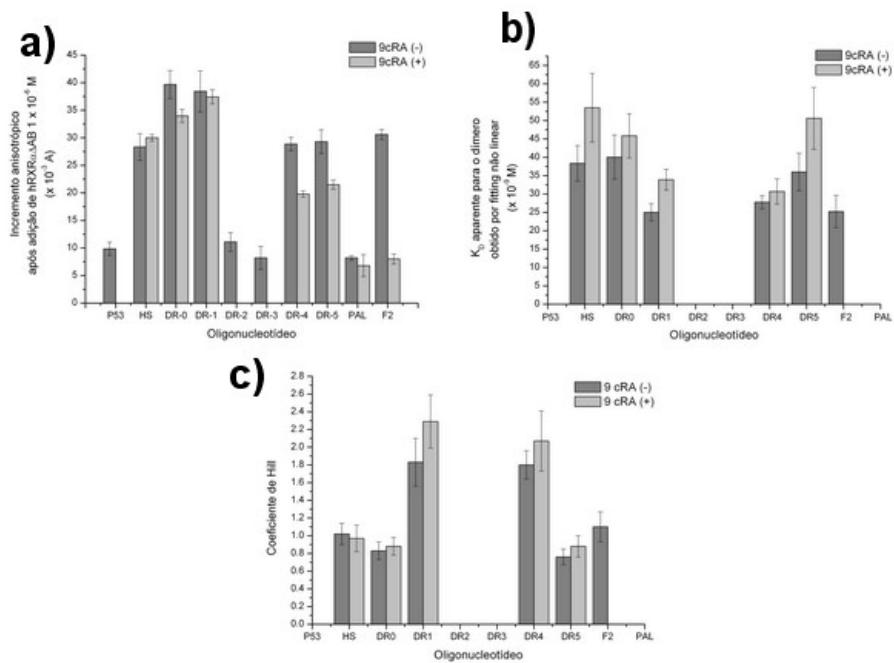
maior incremento anisotrópico - I<sub>2</sub> e DR5 na presença de ligante (que aparenta um comportamento intermediário entre os dois tipos de isotermas) (Fig. 4.2-B e C), conforme será discutido na próxima subseção.

#### **4.1.2 Diferenças no alosterismo e na correlação rotacional dos complexos de hRXR $\alpha$ - $\Delta$ AB com elementos responsivos específicos e inespecíficos de DNA**

O ajuste não linear das isotermas do tipo I<sub>2</sub> conforme o modelo de Hill recuperou valores de K<sub>D</sub><sup>app</sup> com diferenças estatisticamente pouco ou não significativas para a interação do hRXR $\alpha$ - $\Delta$ AB com os diferentes HREs (Fig. 4.2-B). Embora sejam observados valores moderadamente menores de K<sub>D</sub><sup>app</sup> para DR1, DR4 e F2 em relação aos outros HREs, estas diferenças, considerando-se os desvios, são pouco substanciais do ponto de vista termodinâmico.

Apesar da semelhança em valores de K<sub>D</sub><sup>app</sup>, o hRXR $\alpha$ - $\Delta$ AB demonstrou diferenças significativas no que concerne ao alosterismo para a interação com os diferentes HREs com isotermas do tipo I<sub>2</sub>. Dois elementos do tipo repetição direta - DR1 (o elemento responsável específico para o homodímero de hRXR $\alpha$ ) e DR4 - apresentaram isotermas cujos eventos de ligação inicial e de saturação de sítios (delimitados pela distância entre o platô inicial e o final, respectivamente) ocorreram ao longo de aproximadamente uma unidade logarítmica de concentração monomérica de proteína (entre 10 e 100 nM, Fig. 4.1-C e D). A fim de comparação, um evento de ligação que segue um padrão normal segundo o modelo de dois estados ocorre sobre 1.908 unidades logarítmicas de concentração. Esta cooperatividade diretamente evidente na inspeção visual dos dados "brutos" se refletiu nos valores de coeficientes de Hill (*n*) aproximadamente iguais a 2 recuperados para as isotermas destes dois elementos responsivos (Fig. 4.2-C). Assumindo a interação da proteína essencialmente na forma dimérica com o DNA (ao menos para a faixa de concentrações do platô final)- o que foi extensivamente demonstrado em trabalhos prévios da literatura ( ) sobretudo para amostras oriundas da fração correspondente ao monômero da gel filtração para procedimentos de purificação similares aos utilizados neste trabalho () - e assumindo que, ao menos para a faixa de concentrações iniciais e medianas dos

ensaios de afinidade, uma fração significativa da proteína em solução se encontra na forma monomérica, podendo, a princípio, se associar aos HREs como tal (o que também é direta ou indiretamente averiguado em , dentre outros), os valores de  $n$  próximos a 2 recuperados para DR1 e DR4 indicam uma interação do tipo "tudo ou nada" do dímero de hRXR $\alpha$ - $\Delta$ AB a estes dois elementos responsivos, sem a presença significativa de um intermediário monomérico, ou de espécies não cooperativas com maior grau de oligomerização. Contudo, o incremento anisotrópico acentuadamente maior no platô para a interação da proteína a DR1 indica uma significativa diferença conformacional entre os complexos do hRXR-DL com este elemento e com DR4 (Figs. 4.1-C e D e 4.2-A), de tal forma que o tempo de correlação rotacional da sonda fluorescente para hRXR $\alpha$ - $\Delta$ AB <sub>$n$</sub> :DR1 é significativamente maior que para hRXR $\alpha$ - $\Delta$ AB <sub>$n$</sub> :DR4<sup>2</sup>. A discussão a respeito destas diferenças será retomada e sua natureza melhor explorada posteriormente ao longo desta seção.



**Figura 4.2** – Características anisotrópicas e parâmetros de ajuste numérico para o complexo de hRXR $\alpha$ - $\Delta$ AB a diferentes HREs

Para as demais isotermas do tipo I<sub>2</sub>, o ajuste ao modelo de Hill recuperou valores de  $n$  sempre próximos a 1 (não cooperativo) (Fig. 4.2). Contudo, para DR0 e DR5, sobretudo

<sup>2</sup> O índice "n" é utilizado nesta parte do texto ao invés do índice "2", pois, até este ponto da análise, ainda não é assumido categoricamente que ambos os complexos apresentam a mesma estequiometria dimérica.

na ausência do ligante, os valores de  $n$  recuperados (considerando os desvios) foram ambos pouco menores que 1, indicando um comportamento moderadamente anti-cooperativo para a associação do hRXR $\alpha$ - $\Delta$ AB a estes dois HREs. Novamente esta anti-cooperatividade é evidente na inspeção visual dos respectivos conjuntos de dados: ambas as isotermas apresentaram eventos de ligação e saturação abrangendo visivelmente mais de 2 unidades logarítmicas de concentração do monômero (o segundo platô é ainda pouco definido após este intervalo de concentrações) (Fig. 4.1-B e E). Este padrão foi particularmente mais evidente para DR5: na ausência de ligante, a isoterma para este elemento responsivo apresentou uma relativa depressão entre os platôs inicial e final (Fig. 4.1-E, seta cheia), sugerindo uma curva de caráter bimodal e a presença de um intermediário relativamente estável para a titulação do mesmo com o hRXR $\alpha$ - $\Delta$ AB antes da saturação dos sítios. Na presença do ligante, além de um relativo deslocamento da isoterma para a "direita", a curva para DR5 se torna mais inclinada em direção ao eixo das abscissas e o platô final ainda menos definido, caracterizando um aparente comportamento intermediário entre as isotermas do tipo I<sub>1</sub> e I<sub>2</sub> conforme previamente mencionado. É ainda interessante notar que, cruzando as duas isotermas na região delimitada pelas duas setas (cheia e vazia) na Fig. 4.1-E (até a concentração de 190 nM na isoterma com o ligante), obtém-se um perfil de incremento anisotrópico semelhante ao das isotermas para DR2 e F2 na presença de ligante (eventos iniciais de ligação próximos a 1 nM de receptor), enquanto que seguindo apenas a isoterma com o 9-cRA, da região inicial até o mesmo intervalo delimitado pelas duas setas, o padrão é similar ao obtido para a titulação de P53, DR3 e Pal (eventos iniciais de ligação próximos a 10 nM de receptor).

Embora a presença do agonista não altere significativamente os valores de  $K_D^{app}$  e  $n$  recuperados pelo modelo de Hill, foram observadas diferenças significativas nos valores dos incrementos anisotrópicos para a maioria dos HREs após a adição do mesmo (Fig. 4.2-C): Excetuando HS e DR1, as isotermas para a interação de hRXR-DL com os HREs na presença do 9-cRA apresentaram reduções no valor do incremento anisotrópico para a espécie final que variaram entre pequenas (aproximadamente 14,4 % para DR0) até quase três quartos do incremento original (cerca de 73,9 % para F2). Estas diferenças indicam variações nas conformações e/ou

distribuições de espécies oligoméricas para os diferentes HREs após a adição do ligante. A variação mais acentuada neste quesito ocorreu para F2, que na ausência de ligante apresentou um incremento anisotrópico condizente com uma isoterma não cooperativa ( $n = 1.10 \pm 0.17$ ) do tipo I<sub>2</sub> (incremento anisotrópico =  $30.60 \pm 0.90$  mA), enquanto que na presença do 9-cRA apresentou um padrão do tipo I<sub>1</sub> (incremento anisotrópico =  $8.00 \pm 0.90$  mA) (Figs 4.1-F e 4.2-B e C).

É interessante notar que, além do considerável comportamento cooperativo, o complexo de hRXR $\alpha$ - $\Delta$ AB com DR1 demonstrou uma significativa inércia rotacional e/ou conformacional, conforme demonstrado pelo valor do incremento anisotrópico relativamente alto conferido pela proteína ao se complexar a este HRE (apenas DR0 apresentou valores similares). Além disto, o complexo com DR1 é o único que aparenta manter esta mesma estabilidade rotacional/conformacional na presença do ligante, visto que a adição do 9-cRA não afetou o seu acentuado valor de anisotropia (Fig. 4.1-C e 4.2-A). Tomados em conjunto o comportamento cooperativo, a inércia mecânica rotacional e/ou conformacional e a estabilidade destas mesmas propriedades na presença do ligante, o complexo hRXR $\alpha$ - $\Delta$ AB:DR1 apresentou propriedades únicas dentre todos os outros complexos hRXR $\alpha$ - $\Delta$ AB:DNA abordados neste estudo.

O outro HRE do tipo repetição direta para o qual o receptor demonstrou o mesmo comportamento cooperativo - DR4 - não exibiu a mesma inércia mecânica aparente em DR1. Além disto, o complexo ternário proteína:ligante:DR4 apresentou uma redução de aproximadamente 5% em sua anisotropia (indicando um aumento na já alta mobilidade do fluoróforo), embora tenha mantido a mesma intensidade em sua cooperatividade.

Conforme discutido em , as diferenças substanciais entre os valores de anisotropia para o complexo com DR1 e os dos complexos com DR4 e F2 poderiam tanto serem decorrentes de diferenças na dinâmica local do fluoróforo, devido a diferenças conformacionais, quanto serem oriundas de mudanças no tempo de correlação rotacional global entre os respectivos complexos, decorrentes de variações nas dimensões moleculares (por exemplo, devido a diferenças na distribuição de espécies oligoméricas para a proteína entre os três complexos). Além disto, as duas hipóteses não são mutuamente excludentes. Na literatura, contudo, para procedimentos

de expressão e purificação similares aos utilizados neste estudo, a formação de complexos de maior ordem que dímero da proteína associada a HREs contendo apenas dois meios sítios somente foi observada de forma inespecífica e para frações provenientes do pico da gel filtração correspondente ao tetrâmero . Estes complexos tetrâmero:HRE tendiam a se converter a complexos do tipo dímero:HRE com o aumento da concentração de proteína e/ou DNA. Além disto, a adição do agonista 9-cRA também promovia a conversão para o complexo dimérico. Também em nosso próprio grupo de pesquisa, no trabalho de Doutorado de Portugal RV , para um procedimento de purificação idêntico ao do presente estudo e utilizando a mesma sequência de DR1, complexos do tipo tetrâmero:DR1 só foram visualizados em gel nativo na forma de traços, após a incubação do HRE com amostras provenientes do pico de tetrâmero. Mesmo para estas frações, a maior parte da população de complexos proteína:DNA observada era de natureza dimérica. O presente autor não encontrou na literatura relatos a respeito da formação espontânea de complexos com estequiometria do RXR maior que a dimérica a partir da mistura de soluções de HREs contendo apenas dois meios sítios a soluções da proteína previamente purificada na forma dimérica, ou mesmo extratos celulares.

Além da procedência do hRXR $\alpha$ - $\Delta$ AB do pico correspondente ao dímero da gel filtração, a manutenção do valor de incremento anisotrópico para o complexo hRXR $\alpha$ - $\Delta$ AB:DR1 após a adição do 9-cRA, torna improvável uma alteração da estequiometria do complexo com este HRE na presença do ligante, o que por sua vez, comparado aos dados prévios da literatura acima mencionados, é um indício de que o complexo cooperativo formado entre a proteína e DR1 se tratam do complexo específico de dímero ligado a DNA.

As análises descritas nas próximas subseções foram realizadas de forma a se obter uma maior informação a respeito da especificidade, estequiometria e mobilidade das espécies finais nos diferentes complexos de hRXR $\alpha$ - $\Delta$ AB:DNA estudados.

#### **4.1.3 Experimentos de competição confirmam a inespecificidade das isotermas do tipo I<sub>1</sub> e indicam a existência de um intermediário e de um maior número de interações inespecíficas para o complexo com F2**

De forma a melhor averiguar quanto à especificidade das interações observadas, a formação dos respectivos complexos com os F-HREs DR1, DR4 e F2 foi desafiada com um excesso de 10 X em solução (50 nM) do próprio HRE ou de um poli A-T inespecífico de 24 pb, ambos respectivamente não marcados.

Como observado na Fig. 4.3-A-C, o excesso de HRE não marcado induziu a um deslocamento das respectivas isotermas para maiores concentrações. Estes resultados, claramente demonstram o status de equilíbrio das interações do hRXR $\alpha$ - $\Delta$ AB com os HREs nas condições experimentais utilizadas . Os valores de  $K_D^{app}$  recuperados para estas três isotermas deslocadas para altas concentrações foram, como esperado, significativamente maiores que os valores calculados na ausência de competição (Fig. 4.4 A).

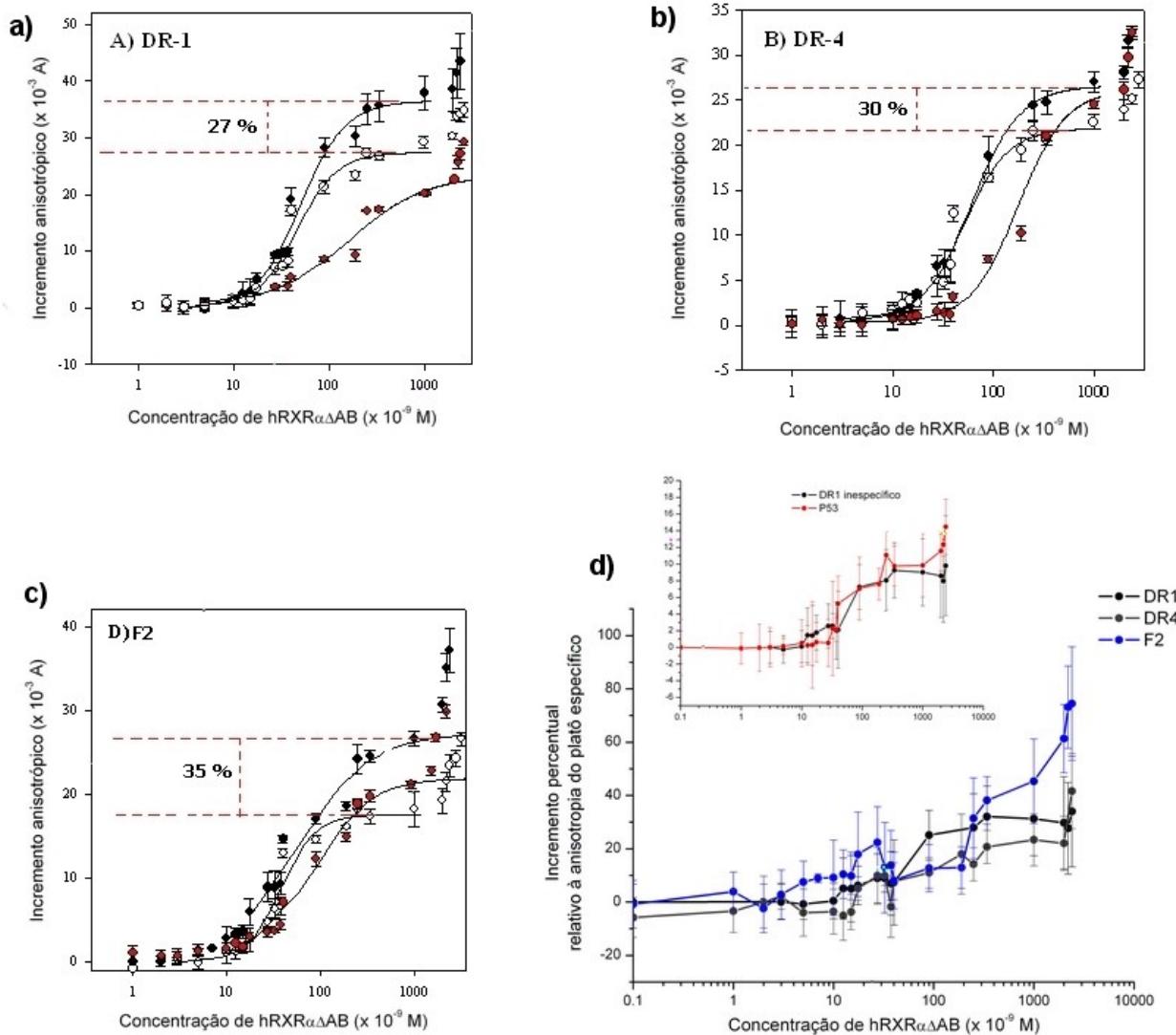
A adição de excesso de poli A-T não marcado, por sua vez, deslocou as isotermas de uma forma assimétrica: para os três HREs testados, a porção correspondente a aproximadamente o terço superior da curva é bruscamente deslocada (Fig. 4.3-A-C), provocando uma redução no valor máximo de incremento anisotrópico observado (Fig. 4.4 B). As porções mais fortemente deslocadas das respectivas isotermas ocorrem a partir de concentrações próximas à incidência do platô final nas curvas originais (Fig. 4.3-D). Este padrão de deslocamento faz com que, apesar das consideráveis reduções nos valores finais de anisotropia, os valores de  $K_D^{app}$  recuperados para as isotermas na presença do excesso de 10 X do competidor inespecífico sejam similares aos recuperados para as isotermas na ausência de competição (Fig. 4.3-A). Também a hierarquia original dos valores de incremento anisotrópico é mantida na presença do poli A-T, com o incremento para o complexo com DR1 significativamente maior que os mesmos incrementos para DR4 e F2.

Apesar da manutenção dos valores de  $K_D^{app}$  para as três isotermas na presença do poli A-T não marcado, a região inespecificamente deslocada para a isoterma de F2 segue um padrão

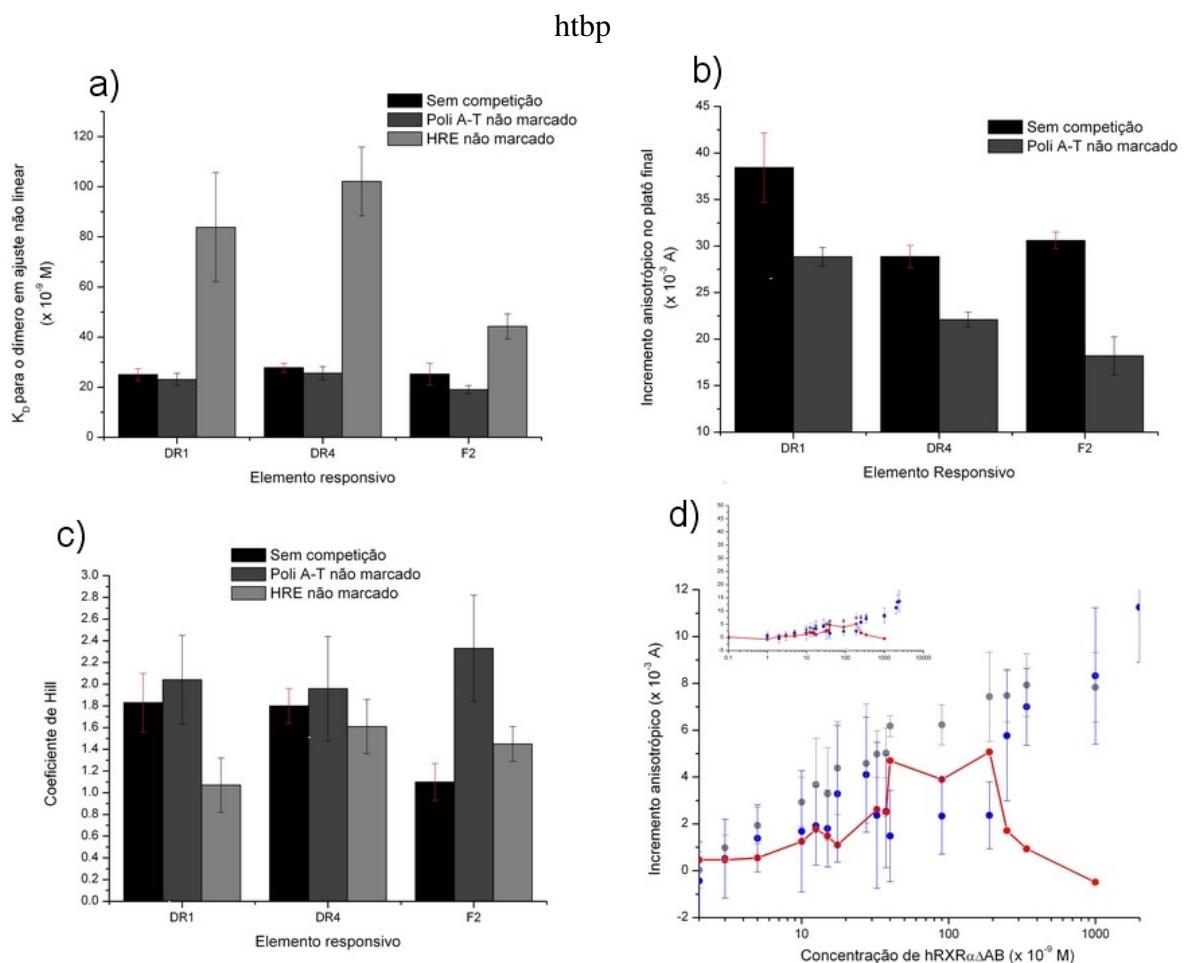
distinto do observado para os complexos cooperativos com os dois DRs. Em primeiro, a redução no incremento anisotrópico do platô final para o complexo hRXR $\alpha$ - $\Delta$ AB:F2 (tanto em valores percentuais quanto absolutos) é significativamente maior que as respectivas reduções para DR1 e DR4, indicando uma maior participação de interações inespecíficas no perfil originalmente obtido para esta isoterma (Figs. 4.3-A-C e 4.4-B). Além disto, enquanto que para DR1 e DR4 a interação só é visivelmente deslocada a partir de concentrações de proteína titulante  $\geq$  100 nM; para F2, existe também um deslocamento da porção inicial, entre 5 e 30 nM, ao passo que a porção mediana da curva praticamente não apresenta deslocamento (mantendo assim o mesmo valor de  $K_D^{app}$ , como previamente mencionado) (Fig. 4.3-A-D). Também o platô entre 100 e 1000 nM se torna mais bem definido para a isoterma com F2 na presença do poli A-T que na ausência de competição. Estes fatores fazem com que a interação com F2 na presença do competidor inespecífico adote o perfil de uma isoterma cooperativa, o que pode ser evidenciado pelo significativo aumento do valor de  $n$  recuperado pelo ajuste de Hill: próximo a 2, semelhante aos valores recuperados para a interação com DR1 e DR4 (Fig. 4.4-C), os quais, por sua vez, não apresentaram mudanças significativas para o valor de  $n$  na presença ou na ausência do competidor inespecífico.

Na figura 4.3-D, são comparados os perfis resultantes das diferenças nas isotermas sem competição e com competição inespecífica de 10 X para DR1, DR4 e F2 (referidos no texto que se segue como *incrementos inespecíficos*), devidamente normalizados pelos valores de incremento anisotrópico nos platôs das respectivas isotermas na presença do competidor inespecífico (estes platôs serão doravante referidos como *platôs específicos*). Os desvios mostrados na figura são os desvios acumulados entre as duas isotermas, sendo os mesmos também devidamente normalizados pelos valores nos respectivos platôs específicos de cada HRE<sup>3</sup> de forma a manterem-se os desvios relativos para cada ponto. Os perfis dos incrementos inespecíficos para DR1 e DR4 se sobrepõem consideravelmente ao perfil da isoterma de interação do hRXR $\alpha$ - $\Delta$ AB com a seqüência alvo para o p53 (o que é melhor evidenciado no *inlet* da Fig. 4.3-D, em que o perfil de incremento inespecífico para DR1 é comparado, em valores absolu-

<sup>3</sup> Os incrementos inespecíficos no *inlet* da Fig. 4.3, bem como os respectivos desvios acumulados, são mostrados em valores absolutos de incremento anisotrópico em mA (não normalizados) (vide texto)



**Figura 4.3 –** Experimentos de competição para a interação do hRXR $\alpha$ - $\Delta$ AB a diferentes HREs



**Figura 4.4** – Principais parâmetros e informações provenientes da análise dos experimentos de competição

tos de anisotropia (mA), ao incremento anisotrópico para a sequência responsiva do p53<sup>4</sup>). A sobreposição da fração da isoterma deslocada destes HREs pelo poli-AT de 24 pb com a isoterma de associação com a sequência alvo do p53 confirma o caráter inespecífico da interação observada para o hRXR $\alpha$ -ΔAB com esta última seqüência e sugere que este mesmo tipo de interação ocorre indiscriminadamente, em maior ou menor grau, pelo ao menos para qualquer sequência de oligonucleotídeo com um número de pares de bases próximo aos das seqüências utilizadas neste estudo. Visto a alta sobreposição dos perfis de incremento anisotrópico para todas as isotermas do tipo I<sub>1</sub> com a isoterma da seqüência responsiva para o p53, pode-se deduzir uma inespecificidade pelo ao menos similar para todas estas isotermas<sup>5</sup>. Outro fator interessante é que o incremento anisotrópico referente ao platô específico para a interação com DR4 apresenta um valor similar ao do platô para a interação do hRXR $\alpha$ -ΔAB com este elemento na presença do ligante (Figs. 4.1-D, 4.2-A e 4.3). Ainda, o fato de que os respectivos incrementos inespecíficos para a interação do hRXR $\alpha$ -ΔAB com DR1 e com DR4 conferem com o perfil de uma isoterma inespecífica previamente caracterizada, são indícios de que estes mesmos incrementos inespecíficos não são decorrentes da formação de oligômeros de maior ordem sobre o dímero especificamente formado<sup>6</sup> visto seguirem o mesmo perfil (inclusive com valores de incremento anisotrópico similares) que uma interação que ocorre sem a necessidade da presença de qualquer espécie de hRXR $\alpha$ -ΔAB previamente associada.

O perfil de incremento inespecífico para a interação com F2, por outro lado, aparenta conter a contribuição de outro(s) tipo(s) de interação(ões) além da interação inespecífica “comum” caracterizada para a sequência alvo do p53. O incremento anisotrópico deslocado da interação do hRXR $\alpha$ -ΔAB com F2 pelo competidor inespecífico (Fig. 4.3-C e D), sugere a existência de uma interação de menor especificidade que começa a se manifestar a baixas concentrações além de um incremento adicional para as interações inespecíficas a altas concentrações, ambos

<sup>4</sup> uma vez que os oligonucleotídeos contendo as respectivas seqüências do DR1 e responsiva para o p53 apresentam aproximadamente o mesmo número de pares de base (24 e 23, respectivamente) os perfis de interação inespecífica podem ser mais acuradamente comparados em termos dos incrementos anisotrópicos em valores absolutos para estes dois HREs

<sup>5</sup> uma discussão mais elaborada a respeito das espécies representativas das isotermas do tipo I<sub>1</sub> será melhor formulada em 4.1.5.

<sup>6</sup> por exemplo, a formação de complexos do tipo tetrâmero-DNA, a partir da interação de um novo dímero com o previamente associado ao HRE

apresentando baixa colaboratividade em relação às interações específicas a concentrações medianas (acarretando a diminuição considerável do coeficiente de Hill quando estas duas porções inespecíficas da isoterma são levadas em consideração (Fig. 4.4-C)).

Na Fig. 4.4-D, o perfil de incremento inespecífico para a interação do hRXR $\alpha$ - $\Delta$ AB com o F2 na ausência de ligante é comparado à isoterma de interação para o receptor com este mesmo elemento na presença do 9-cRA. Nota-se a significativa sobreposição da isoterma na presença do agonista com a interação inespecífica na ausência do mesmo. A única divergência (embora pouco significativa em face aos desvios super-acumulados) é a região correspondente à depressão aparente no perfil de incremento inespecífico, que na ausência do ligante é preenchida pelo incremento correspondente à interação específica e cooperativa a concentrações medianas da Fig. 4.3-C e na presença do mesmo é ocupada por um incremento consideravelmente reduzido (aproximadamente 3 X menor) contínuo às duas regiões que flanqueiam esta depressão (Fig. 4.4-D). Este padrão sugere que a interação do hRXR $\alpha$ - $\Delta$ AB com o F2 se dá com a presença de uma espécie intermediária com significativa maior mobilidade (verificada pelo incremento anisotrópico pouco significativo em relação ao da isoterma completa sem o ligante) e menor especificidade (uma vez que corresponde à porção da isoterma que é deslocada pelo excesso de competidor inespecífico). Na ausência do ligante e na presença do excesso de poli A-T, a interação deste intermediário de menor especificidade foi inibida. Contudo, a interação simultânea deste mesmo intermediário e de uma segunda espécie<sup>7</sup> do hRXR $\alpha$ - $\Delta$ AB com o F2 ainda puderam ocorrer de forma colaborativa com o aumento da concentração da proteína em solução. Aparentemente, a interação mútua destas duas espécies confere um maior grau de complementaridade estereoquímica ao complexo, resistindo assim ao deslocamento pelo excesso de 10 X de oligo inespecífico. A adição do ligante, aparentemente, promove um efeito contrário: inibe a associação cooperativa da(s) espécie(s) com maior grau de oligomerização do hRXR $\alpha$ - $\Delta$ AB a F2, restringindo a formação do complexo à espécie intermediária, e esta por sua vez dá lugar às interações inespecíficas de maior ordem conforme se aumenta a concentração

---

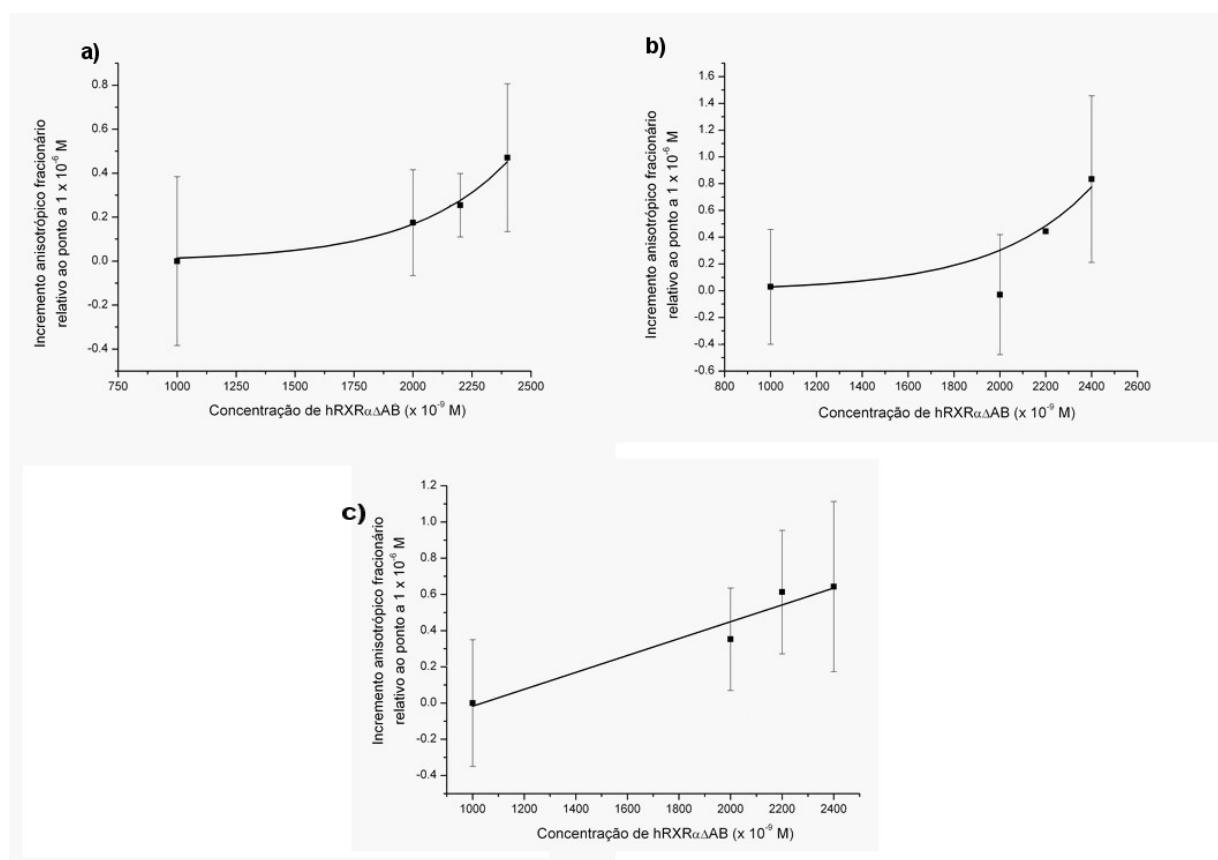
<sup>7</sup> Aqui é utilizada a nomenclatura *espécie* e não *elemento*, *molecula* ou *oligómero*, visto que a primeira não implica na assunção *a priori* de nenhuma estequiometria ou estado oligomérico em particular. Por exemplo, tanto hRXR, hRXR<sub>2</sub>, hRXR<sub>3</sub> e hRXR<sub>4</sub> são espécies oligoméricas de hRXR.

de hRXR $\alpha$ - $\Delta$ AB em solução para concentrações  $\gg$  100 nM (Figs. 4.1-F, 4.3-C e D e 4.4-D). Os complexos altamente cooperativos com DR1 e DR4, por outro lado, não apresentaram evidências da presença significativa desta ou de qualquer outra espécie intermediária, tanto na ausência quanto na presença de excesso de competidor ou do agonista.

O complexo com F2 apresentou também distinções no que concerne ao perfil das interações inespecíficas formadas a partir de maiores concentrações da proteína. Além do já referido maior percentual de incremento anisotrópico decorrente destas interações para o complexo com F2, a inclinação do incremento inespecífico relativo ao platô específico<sup>8</sup> é mais acentuada para este elemento que para DR1 e DR4 (Fig. 4.3-D). Para os dois DRs de isotermas cooperativas, os platôs referentes às interações inespecíficas entre 100 e 1000 nM são aproximadamente paralelos aos respectivos platôs específicos. Já para F2 existe uma inclinação positiva para o incremento inespecífico nesta faixa de concentrações (Fig. 4.3-C e D), indicando a interação de uma espécie inespecífica aparentemente distinta da(s) previamente caracterizada(s) para a sequência responsiva do p53 e para as interações inespecíficas em DR1 e DR4. Este comportamento é melhor visualizado através dos diferentes padrões de divergência das interações inespecíficas a partir de 1000 nM para as isotermas destes três HREs. Para todas as isotermas, as interações inespecíficas a altas concentrações aumentam acentuadamente após o termo assintótico entre 100 e 1000 nM, apresentando um aparente comportamento divergente (Figs. 4.1 e 4.3). Para os incrementos inespecíficos das isotermas de interação com DR1 e DR4, esta divergência segue um padrão similar ao observado para a interação inespecífica com a seqüência responsiva do p53: com a fração de incremento anisotrópico entre 1000 e 2500 nM sendo bem ajustável a uma função exponencial com expoente próximo a 0.002 e fator pré exponencial da ordem de  $10^{-3}$  (na Fig. 4.5-A e B, são comparados os ajustes exponenciais nesta região para as isotermas de interação com a seqüência alvo do p53 e com DR4, respectivamente). Para o incremento inespecífico da isoterma de interação do apo-hRXR $\alpha$ - $\Delta$ AB com o F2, por outro lado, este incremento fracional entre 1000 e 2500 nM é melhor ajustável a uma função linear, e não a uma exponencial (4.5-C). Isto ocorre porque, conforme pode ser verificado na Fig. 4.3,

<sup>8</sup> vide definições de *incremento inespecífico* e *platô específico* no texto acima

para a titulação de DR1 e DR4, tanto os platôs específicos quanto os inespecíficos demonstram sinais de saturações locais na faixa de concentração entre 100 e 1000 nM, seguindo-se depois a região divergente das interações inespecíficas. A sucessão de interações entre a concentração de 1000 nM (que ainda se encontra no platô) até as concentrações maiores (passando da região assintótica para a região divergente) confere, desta forma, o perfil de uma exponencial próxima a seu termo assintótico horizontal a estas regiões das isotermas inespecíficas. Contudo, para F2, enquanto a isoterma específica se encontra completamente saturada nesta região, a isoterma inespecífica apresenta, como mencionado, um padrão assintótico menos definido, com uma saturação não tão evidente<sup>9</sup>, conferindo assim à sucessão de incrementos anisotrópicos entre 1000 e 2500 nM um perfil menos semelhante ao de uma exponencial e mais próximo ao linear (Figs. 4.3-C e D).



**Figura 4.5 – Ajuste para os termos divergentes das interações inespecíficas**

Este perfil corrobora a idéia de que o incremento inespecífico para a interação com o F2 (na

<sup>9</sup> esta saturação menos evidente para a fração inespecífica da isoterma de interação do apo-hRXR $\alpha$ -ΔAB com o F2 colabora, em conjunto com a região de incremento inespecífico entre 5 e 30 nM, para o comportamento não cooperativo da isoterma na ausência de competição

ausência do ligante) a altas concentrações não se restringe à interação inespecífica “comum” caracterizada previamente pela interação com a seqüência responsiva do p53 (que apresenta, também, um perfil de saturação melhor definido entre 100 e 1000 nM, de forma similar aos incrementos inespecíficos observados para as isotermas com DR1 e DR4), demonstrando sinais da associação de uma espécie oligomérica do hRXR $\alpha$ -ΔAB, com menor grau de especificidade e baixa colaboratividade em relação à(s) espécie(s) previamente formada(s) (Fig. 4.3-C).

O perfil singular de interações específicas e inespecíficas do hRXR $\alpha$ -ΔAB com o F2, conforme observado no experimento de competição ilustrado na Fig.4.3-C, levanta duas possibilidades quanto à natureza das espécies da proteína que interagem com este HRE:

- **Hipótese 1- Diferentes espécies oligoméricas com diferentes extensões de especificidade.** Segundo este modelo, cada uma das interações, específicas e inespecíficas, ilustradas na Fig. 4.3-C seriam representativas de um oligômero distinto do hRXR $\alpha$ . Por exemplo, admitindo a possibilidade da interação do receptor com este elemento responsável até a estequiometria de tetrâmero, a espécie intermediária e inespecífica (entre 5 e 30 nM) poderia ser o monômero, a espécie específica e colaborativa a concentrações medianas o dímero e a última espécie, a maiores concentrações, um tetrâmero com menor especificidade e colaboratividade se associando ao dímero previamente ligado (só ocorrendo tal interação na ausência do ligante). O incremento anisotrópico significativamente menor do complexo em F2 relativo ao mesmo incremento para o complexo em DR1 torna esta hipótese particularmente pouco provável no caso de se comprovar que o complexo do hRXR $\alpha$ -ΔAB com DR1 observado em nossos experimentos se trata da ligação da proteína na forma dimérica (em concordância com os dados já extensamente documentados na literatura).
- **Hipótese 2- A proteína se liga com a mesma estequiometria através de duas formas distintas e com diferentes especificidades.** Uma mesma quantidade de sítios do DNA poderiam ser ocupados pela proteína de duas formas diferentes. Por exemplo, admitindo a ligação da proteína exclusivamente na forma dimérica, esta poderia se associar a F2

tanto por um caminho em que os dois monômeros se ligam de forma independente (proporcionando uma isoterma com menor colaboratividade e especificidade) quanto por um caminho em que os dois monômeros se ligam interagindo entre si (proporcionando um maior grau de cooperatividade e especificidade). Na ausência do excesso de DNA competidor, os dois caminhos ocorreriam simultaneamente para a população de HREs em solução, podendo inclusive serem intercambiáveis. Com a adição do excesso de competidor, apenas o caminho cooperativo (com maior ajuste estereoquímico e consequente especificidade) é favorecido.

Obviamente, as duas possibilidades acima mencionadas não são, *a priori*, auto excluientes. Assim sendo, de forma a melhor caracterizar a estequiometria das espécies formadas para DR1, DR4 e F2, foram realizados experimentos de titulação destes três HREs em concentrações estequiométricas pelo hRXR $\alpha$ - $\Delta$ AB.

#### **4.1.4 Curvas estequiométricas comprovam a ligação dimérica do hRXR $\alpha$ - $\Delta$ AB a DR1, DR4 e F2 na ausência do agonista e sugerem, em conjunto com os experimentos de competição, a existência de diferentes caminhos de associação do dímero a F2**

De forma a melhor caracterizar as espécies oligoméricas de hRXR $\alpha$ - $\Delta$ AB presentes nas isotermas do tipo I<sub>2</sub>, foram realizados experimentos de titulação de DR1, DR4 e F2 em concentração de 600 nM ( $\approx$  10X os respectivos valores de K<sub>D</sub><sup>app</sup> expressos em concentração do monômero em nM). Na Fig. 4.6 são apresentadas as curvas de titulação estequiométrica para os três HREs. Embora, para os dados *brutos* de todas as respectivas curvas, seja observada uma notória inflexão próxima à concentração de proteína relativa à saturação de dois sítios no DNA (precisamente, mais acentuada em 1.80 sítios - próximo a 1080 nM de proteína em solução), sugerindo a ocupação destes sítios pela proteína na forma dimérica; em uma primeira análise não são observados os platôs finais bem definidos característicos para curvas estequiométricas. Para os três HREs, a intensidade anisotrópica continua a aumentar além desta aparente inflexão, não alcançando um platô bem definido mesmo após a adição de hRXR $\alpha$ - $\Delta$ AB em concentrações 6X maiores que a concentração fixa de DNA em solução (não mostrado). Isto ocorre devido

à aparente sobreposição dos platôs estequiométricos com os termos divergentes inespecíficos previamente observados para as curvas de afinidade na adição da proteína em concentrações acima de 1000 nM (Figs 4.1, 4.3 e 4.5). Embora a concentração utilizada de 600 nM para o oligonucleotídeo titulado seja estequiométrica para as interações específicas caracterizadas pelo platô final entre 100 e 1000 nM ( $> 10 \times$  os valores de  $K_D^{app}$  recuperados para o monômero nos modelos de Hill para os três HREs), isto obviamente não é verdadeiro no que se refere às interações inespecíficas responsáveis pelos termos divergentes a concentrações  $\geq 1000 \text{ nM}^{10}$ , prosseguindo a interação da proteína com os oligonucleotídeos em solução de forma inespecífica após a saturação dos sítios específicos. Efeitos similares de sobreposição de eventos específicos e inespecíficos em altas concentrações para a interação proteína-DNA em soluções de baixa força iônica são comumente observados na literatura, como por exemplo em , para a interação da isoforma beta do receptor nuclear de estrogênios ( $ER\beta$ ) com o seu elemento responsivo (ERE). Assim sendo, os ajustes numéricos previamente efetuados para os termos divergentes das interações inespecíficas entre 1000 e 2500 nM foram utilizados para se corrigir, de forma semiempírica, a influência destas interações nos platôs estequiométricos. Embora estes ajustes tenham sido realizados para a titulação do DNA marcado em uma concentração  $120 \times$  menor, os mesmos foram realizados considerando-se os incrementos *fracionais* de anisotropia entre 1000 e 2500 nM (tomando como referência o valor de anisotropia a 1000 nM, que para as titulações de DNA a 5 nM em solução se encontra sobre um platô) e não para os incrementos em valores absolutos de anisotropia. Assim, tais ajustes consideram a concentração do DNA titulado de forma implícita, visto ser esperada uma dependência similar da concentração de DNA em solução tanto para o valor de incremento anisotrópico no platô, quanto para a fração dos incrementos relativos de anisotropia para os primeiros pontos das interações inespecíficas

---

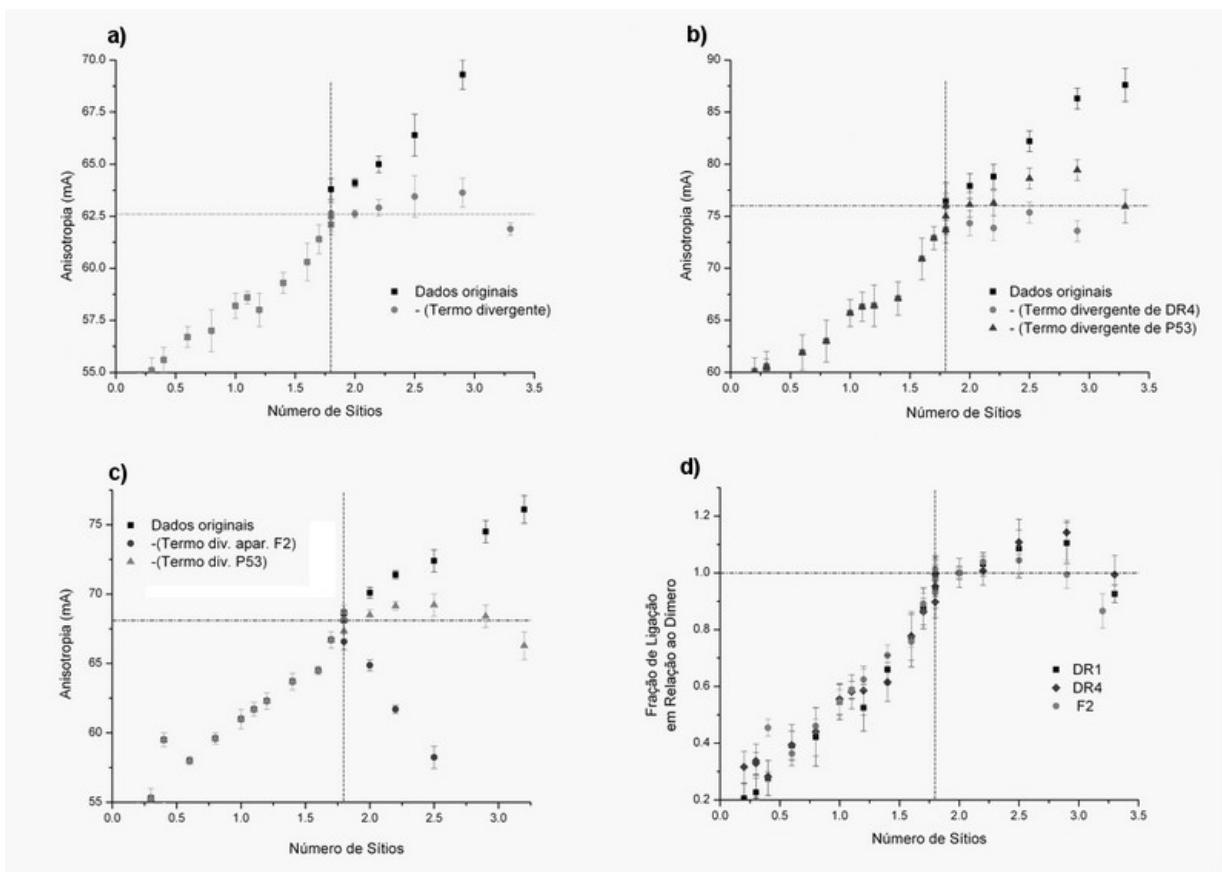
<sup>10</sup> assumindo que estas interações inespecíficas divergentes atinjam um ponto de saturação bem delimitado em alguma concentração qualquer, o que não foi observado previamente nas condições não estequiométricas (vide 4.1.1)

divergentes<sup>11</sup>. A correlação entre as intensidades do incremento anisotrópico a 1000 nM de proteína e dos incrementos entre esta concentração e 2500 nM, bem como a consequente veracidade dos ajustes numéricos da Fig. 4.5 para o incremento fracional nesta região, se tornam claros pela inspeção das curvas da Fig. 4.1. Embora os valores absolutos das intensidades anisotrópicas nos termos divergentes das isotermas do tipo I<sub>1</sub> sejam significativamente menores que as mesmas intensidades para as isotermas do tipo I<sub>2</sub> (em concordância com as diferenças de incremento anisotrópico nos respectivos platôs), os incrementos fracionais relativos ao valor a 1000 nM de proteína para os termos divergentes das isotermas da seqüência responsiva do p53 (I<sub>1</sub>), de DR1 e DR4 (ambos do tipo I<sub>2</sub>) e de pal (I<sub>1</sub>), por exemplo, se ajustam numericamente a funções exponenciais com o mesmo expoente e fatores pré-exponenciais similares (Fig. 4.5-A e B e não mostrado). Desta forma, a correção dos platôs estequiométricos, entre a concentração de proteína titulante de 1000 a 2500 nM, pelos ajustes numéricos ilustrados na Fig. 4.5 é uma aproximação semiempírica razoável. Estas concentrações de proteína correspondem à quantidade necessária para a saturação de um número de sítios entre aproximadamente 1.7 e 4.2 para uma concentração de 600 nM de DNA marcado.

A correção das três isotermas estequiométricas, a partir da concentração de 1000 nM (~ 1.7 sítios), pelo termo divergente da interação inespecífica com a seqüência alvo do p53 estabilizou os valores de anisotropia para concentrações acima do ponto de inflexão correspondente a 1.8 sítios (~ 1080 nM) (Fig. 4.6). As isotermas remanescentes destes ajustes adotam um perfil similar ao da interação de uma proteína ocupando dois sítios no DNA (interação dímero-DNA). Para DR1 e DR4, a correção das curvas originais tanto pelo termo divergente calculado através da isoterma para a seqüência responsiva do p53 quanto pelo calculado a partir dos próprios termos divergentes a altas concentrações nos incrementos inespecíficos destes respectivos HREs produziram resultados similares (Fig. 4.6-D e não mostrado).

---

<sup>11</sup> estas considerações, bem como os próprios ajustes, só são válidos considerando que a concentração do DNA em solução se encontra suficientemente abaixo da concentração *estequiométrica* para tais interações inespecíficas aparentemente divergentes. Contudo, tal assunção é razoável, visto que tanto para as isotermas de afinidade quanto para as estequiométricas, estas interações não apresentaram sinais de saturação mesmo para concentrações próximas a 4  $\mu$ M, conforme previamente mencionado. Desta forma, tanto a concentração de 5 nM quanto a de 600 nM de DNA marcado estão suficientemente abaixo de possíveis valores de  $K_D^{app}$  para estas interações inespecíficas



**Figura 4.6** – Curvas Estequiometricas para a interação do hRXR $\alpha$ -ΔAB com diferentes DR1, DR4 e F2

Para F2, por outro lado, enquanto a correção pelo termo divergente de p53 proporcionou um perfil de saturação estequiométrica de dois sítios do HRE pelo receptor (similar aos obtidos para DR1 e DR4), a correção pelo aparente termo linear calculado para o incremento inespecífico entre 1000 e 2500 nM na isoterma de afinidade a este elemento provocou um perfil decrescente para a anisotropia posterior à inflexão em 1.8 sítios (Fig. 4.6-C). Este perfil sugere que o termo linear ajustado previamente para a interação inespecífica em F2 (Fig. 4.5-C) não modela a interação inespecífica a altas concentrações para este elemento como ocorreu com o ajuste numérico exponencial, aparentemente comum a todos os oligonucleotídeos testados nas condições do experimento. A estabilização do incremento anisotrópico na curva estequiométrica a partir do ponto correspondente a aproximadamente 2 sítios com a correção por este termo divergente “comum” sugere ser o termo linear previamente ajustado para F2 decorrente da sobreposição deste termo exponencial e do padrão assimptótico da interação não cooperativa e de menor especificidade de uma espécie aparentemente dimérica do hRXR $\alpha$ -

$\Delta AB$  a F2, ainda em vias de saturação na concentração protéica de 1000 nM. Com a titulação de uma concentração de DNA  $> 10 \times K_D^{app}$ , a saturação dos sítios de interação para o dímero ocorre de forma mais “direta”, a partir da adição de proteína em concentração o suficiente, sem um padrão assimptótico na eminência desta mesma saturação<sup>12</sup>, de forma a que a divergência dos valores de anisotropia passa a ser bem ajustável pela exponencial, aparentemente comum, caracterizada para este termo na isoterma com a seqüência alvo do p53. A interação desta espécie dimérica não cooperativa segue, obviamente, um caminho de associação distinto que o do dímero cooperativo, já completamente saturado a 340 nM no experimento de competição inespecífica, corroborando a hipótese 2 formulada no fim da subseção 4.1.3.

Assim, enquanto a associação do receptor ocorra no mesmo macroestado dimérico para DR1, DR4 e F2, o complexo formado para as duas repetições diretas apresenta uma natureza obrigatoriamente cooperativa, enquanto que para o palíndromo invertido a associação dos dois monômeros (na ausência do agonista) demonstra uma maior flexibilidade quanto aos mecanismos de associação, adotando o complexo dimérico neste HRE dois micro-estados distintos (pelo ao menos). Aparentemente, o apo-hRXR $\alpha$ - $\Delta AB$  se complexa a F2 tanto na forma de dois monômeros independentes (formando um complexo com menor especificidade e de forma não colaborativa a esta sequência) quanto na forma de um dímero com interação intersítios da mesma ordem que a dos complexos com DR1 e DR4. A interação do dímero cooperativo a F2 apresenta especificidade significativamente maior (comparando sua preferência de ligação entre este HRE e um poli-AT com 24 pares de base) que a ligação dos dois monômeros independentes<sup>13</sup>. Esta maior especificidade indica um maior grau de ajuste estereoquímico a F2 para o complexo cooperativo que para o não cooperativo. Um outro indício de que a formação do dímero cooperativo implica em um relativo maior ajuste estereoquímico dos monômeros aos HREs (e/ou entre si) é o ”arraste” inicial (entre 5 e 30 nM) para a isoterma específica coope-

<sup>12</sup> Isotermas de titulação de espécies em concentrações significativamente maiores que os valores de  $K_D$  para a interação com o titulante (usualmente,  $\geq 10 \times K_D$ ), tendem a apresentar um comportamento mais ”direto” para sua saturação, sendo este o fundamento do experimento de titulação estequiométrica aqui abordado. Isto ocorre porquê sob concentrações significativamente maiores que o  $K_D$  a probabilidade de associação das espécies ao longo da maior parte da titulação deixa de ser uma função dependente da concentração de titulante livre em solução e passa a ser uma função direta do número de sítios de ligação disponíveis ()

<sup>13</sup> pelo ao menos 10 X maior, visto que o complexo na forma de dois monômeros independentes é desafiado na presença de um excesso de 10 X de poli-AT enquanto o dímero cooperativo não o é.

rativa, e que não ocorre para a isoterma inespecífica (Figs. 4.3-C e D). Este arraste sugere que a interação inicial do(s) monômero(s) é menos facilitada na formação do complexo cooperativo que para o complexo não cooperativo e inespecífico, conquanto que no primeiro caso a interação do primeiro monômero facilite significativamente a associação do segundo, o mesmo não ocorrendo para a formação do complexo inespecífico<sup>14</sup> (vide modelagem numérica global em 4.1.6). Desta forma, para a interação do primeiro monômero do hRXR $\alpha$ - $\Delta$ AB com o F2 no caminho que leva à formação do dímero cooperativo faz-se necessário que este sofra alguma mudança conformacional (e/ou induza alguma mudança de natureza dinâmica e/ou conformacional ao DNA) com uma certa penalidade energética demonstrada pela maior dificuldade para a interação deste mesmo primeiro monômero por este caminho<sup>15</sup>. Os valores similares de  $K_D^{app}$  implicam em que, pelo ao menos na faixa de concentrações referentes aos platôs finais (aproximadamente entre 100 e 1000 nM) e na ausência de competição, ambos os microestados e respectivas rotas de associação são igualmente prováveis (embora não se possa dizer, a partir da análise no equilíbrio, se são ou não *intercambiáveis*<sup>16</sup>).

Descrições de mecanismos similares envolvendo rotas paralelas de interações inespecíficas e específicas para a formação de complexos proteína-DNA são relativamente frequentes na literatura. Ferreiro e Prat-Gay , através de experimentos do tipo *duplo impulso com interrupção de transferência de calor* demonstraram que a interação do domínio C-terminal da proteína E2 do papilomavírus com sua sequência responsiva ocorre simultaneamente através de uma rota inespecífica, cineticamente favorecida, e de uma outra rota envolvendo um maior grau de rearranjo conformacional e discriminação de seqüência de bases. Para o próprio hRXR $\alpha$ , estudos de RMN para a interação do DBD com um oligonucleotídeo de 14 pb , contendo a sequência de 6 pb do meio sítio flanqueada por duas seqüências inespecíficas de 4 pb, demonstraram uma

<sup>14</sup> Este perfil de arraste e cooperatividade, sugerindo uma dificuldade inicial para a associação do primeiro monômero e facilitação para a entrada do segundo é o que ocorre obrigatoriamente para a interação do hRXR $\alpha$ - $\Delta$ AB com DR1 e DR4

<sup>15</sup> a mudança conformacional obrigatória para a ligação do hRXR $\alpha$ - $\Delta$ AB a F2 pela rota cooperativa pode também ser a pré-formação do dímero em solução para a ligação ao DNA. De fato, em um experimento em que a formação do complexo é medida no equilíbrio, não se pode distinguir entre um mecanismo de cooperatividade envolvendo um intermediário com o dímero obrigatoriamente formado em solução e um envolvendo a formação obrigatória de um complexo monômero-DNA com baixa energia de dissociação .

<sup>16</sup> Contudo, a análise posterior por modelagem estrutural e dinâmica molecular na próxima seção sugere que ambos os tipos de complexo são, de fato, intercambiáveis

alta taxa de troca entre interações específicas e inespecíficas para resíduos na interface proteína-DNA. Estas trocas ocorriam com uma cinética significativamente mais rápida que a estimada para um ciclo de associação-dissociação do DBD ao oligonucleotídeo, demonstrando que tal intercâmbio entre interações específicas e inespecíficas ocorria com o domínio ainda associado ao DNA.

Além da diferença em possibilidade de microestados para F2 na ausência do ligante, a mobilidade local do fluoróforo e/ou global para os complexos diméricos com DR4 e F2 demonstram serem consideravelmente maiores que para o complexo com DR1, visto o significativo maior valor de incremento anisotrópico para o complexo do dímero com este HRE (Figs. 4.2-A e 4.4-B). Desta forma, os dados sugerem que a associação do dímero de hRXR $\alpha$ - $\Delta$ AB ao elemento fisiologicamente responsivo DR1, além de ocorrer de forma altamente colaborativa confere uma significativa inércia mecânica ao complexo final<sup>17</sup>, mantendo este mesmo perfil de acentuada colaboratividade e inércia mecânica na presença do ligante. Tal padrão não ocorreu para o complexo com nenhum outro oligonucleotídeo deste estudo.

A significativa sobreposição do incremento inespecífico inicial referente à formação do complexo não cooperativo [apo-hRXR $\alpha$ - $\Delta$ AB]<sub>2</sub>:F2 e da isoterma de associação do complexo 9-cRA:hRXR $\alpha$ - $\Delta$ AB:F2 (Fig. 4.4-D) sugere que o intermediário inespecífico formado na rota não cooperativa de associação a este HRE na ausência do ligante e a espécie que se associa ao mesmo elemento na presença do 9-cRA se tratam, igualmente, do complexo monômero:F2. Desta forma, enquanto as mudanças conformacionais induzidas pela associação do ligante são permissivas para a formação do complexo dimérico a DR1 e DR4, elas não o são para a formação da mesma interface de dimerização sobre a superfície do palíndromo invertido. Também a associação como dois monômeros independentes é impedida, ocorrendo a associação do receptor a F2 apenas de forma transitória, como uma única espécie monomérica a baixas concentrações de proteína em solução. Com o aumento da concentração do receptor, esta espécie monomérica é consumida, provavelmente pela formação do complexo dimérico em solução, e dá lugar às interações inespecíficas com o DNA observadas a altas concentrações

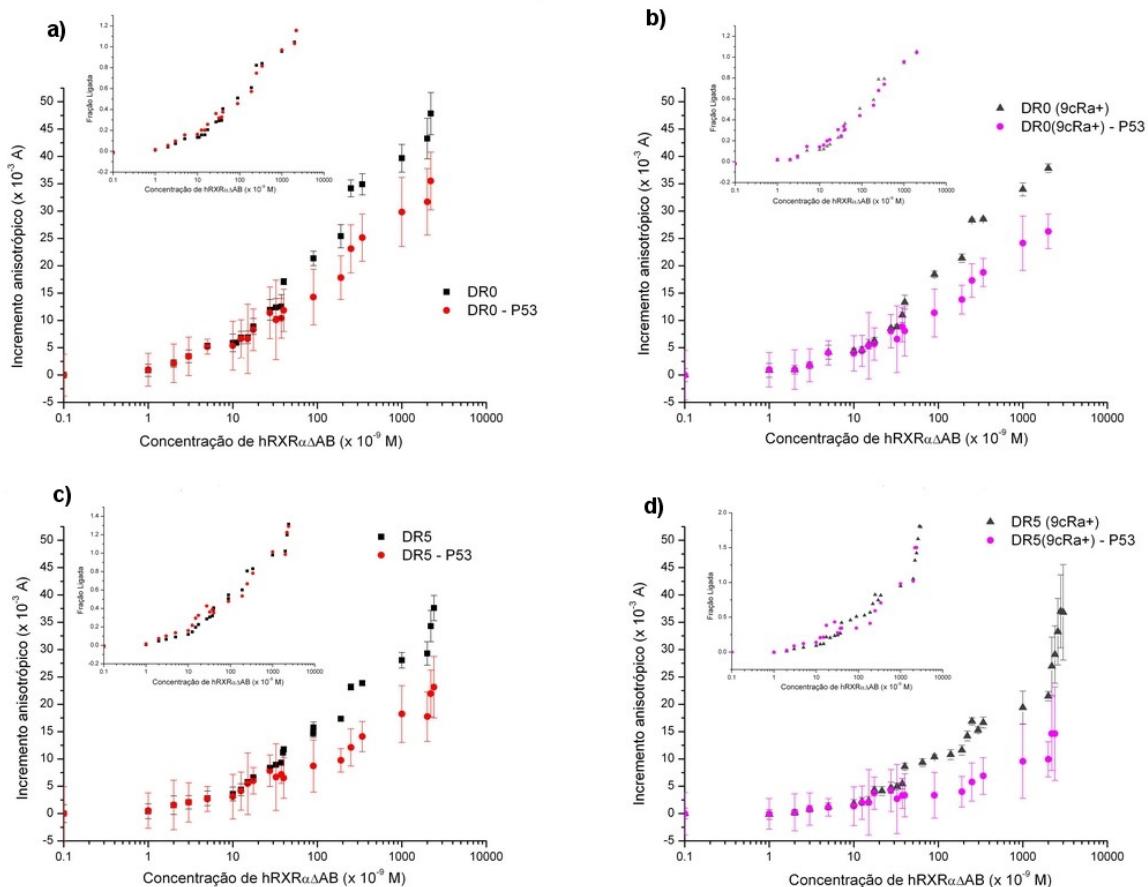
<sup>17</sup> os mecanismos de estabilização dinâmica do complexo hRXR $\alpha$ - $\Delta$ AB-DR1 e aos demais HREs serão melhor discutidos em

para todos os oligonucleotídeos (específicos e inespecíficos) testados nas condições deste estudo (Fig. 4.4-D *inlet*)

#### **4.1.5 Correção das isotermas para HREs não cooperativos a partir do perfil de interação com a sequência alvo do p53 corrobora a hipótese de um intermediário monomérico para sequências inespecíficas**

Uma vez que as análises em 4.1.3 sugerem que a interação inespecífica caracterizada para o complexo com a seqüência alvo do p53 é comum a diferentes oligonucleotídeos, o perfil da isoterma de interação a esta seqüência foi corrigido das interações com o restante dos HREs. Na Fig. 4.7, as duas isotermas anticooperativas de DR0 e DR5, tanto na ausência quanto na presença do 9-cRA, são apresentadas com as respectivas correções em relação à isoterma inespecífica “comum” (p53). Nota-se que, embora esta correção promova uma redução dos valores absolutos de incremento anisotrópico nos platôs finais, os perfis de inclinação e inflexão globais para as respectivas isotermas não são significativamente afetados, o que pode ser visualizado pela normalização destas curvas para os perfis de fração ligada (Fig. 4.7-A-D *inlets*). Contudo, para DR5, a correção do perfil de interação inespecífica ”comum” torna mais claro o caráter bimodal das isotermas (Fig. 4.7-C e D). Para este HRE, esta correção torna evidente a presença de um platô intermediário entre 20 e 100 nM (referido como  $P_{DR51}$ ), na região onde é visualizada a modesta depressão para a isoterma original, e um termo assimptótico final (referido como  $P_{DR52}$ ) a partir de 300 nM, seguindo até as interações inespecíficas divergentes (Figs. 4.1-E e 4.7-C e D). Na ausência do ligante, o platô intermediário apresenta um incremento anisotrópico similar ao do intermediário monomérico caracterizado na isoterma de interação a F2 na presença do 9-cRA (Figs. 4.1-F e 4.4-D). A adição do agonista, aparentemente, desloca toda a isoterma, desde o platô intermediário, para maiores concentrações. Este deslocamento faz com que a espécie originalmente correspondente ao platô médio, com incremento anisotrópico próximo a 5 mA (quando corrigido da interação inespecífica), passe a ocupar grande parte da faixa de concentrações originalmente ocupadas pelo platô final (aproximadamente entre 300 e 1000 nM), enquanto este, por sua vez, apresenta sua maior parte deslocada e sobrepondo-se ao

termo divergente das interações inespecíficas a altas concentrações (Fig. 4.7-D)<sup>18</sup>.



**Figura 4.7** – Correção do termo inespecífico “comum” das isotermas para os HREs anticooperativos

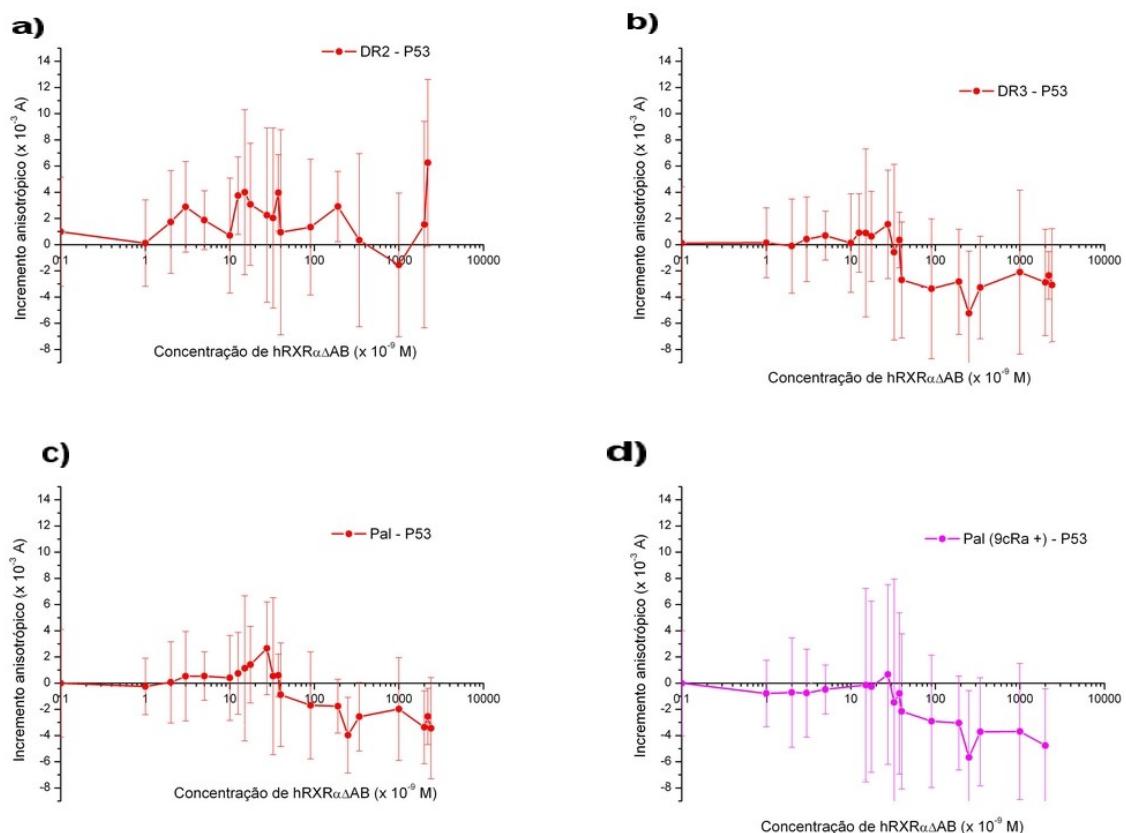
A mesma correção pela interação inespecífica “comum” para as curvas do tipo I<sub>1</sub> confirma a significativa contribuição das interações inespecíficas para estas isotermas (Fig. 4.8). De fato, considerando-se os desvios acumulados, não se observa diferenças significativas entre o perfil da isoterma inespecífica para a seqüência alvo do p53 e as demais isotermas do tipo I<sub>1</sub>. Contudo, considerando-se apenas os comportamentos médios, a diferença entre a isoterma para DR2 e a interação inespecífica “comum” apresenta um pequeno incremento positivo englobando as mesmas faixas de concentrações onde ocorrem as primeiras manifestações do incremento inespecífico em F2 (entre 5 e 30 nM) e a depressão neste mesmo incremento, caracterizando a região aparentemente ocupada pelo intermediário monomérico de menor especificidade em F2 na presença do ligante (Figs 4.4-D, 4.8-A e 4.1-C e F). Também a intensidade

<sup>18</sup> Para fins didáticos, os dois modos aparentes na isoterma de interação com DR5 na presença do ligante serão referidos como P<sub>DR5</sub>1' e P<sub>DR5</sub>2'

média do incremento corrigido em DR2 apresenta, aproximadamente, a mesma intensidade que a diferença entre o incremento da isoterma de interação em F2 na presença do ligante e o termo inespecífico para este HRE na ausência do mesmo (Fig. 4.4-D), bem como o platô intermediário para a isoterma corrigida em DR5 na ausência do ligante (entre 4 e 5 mA) (4.7-C e 4.8-A). Para as demais isotermas do tipo I<sub>1</sub>, a correção pelo incremento inespecífico “comum” (p53) apresenta diferenças ainda menos substanciais, mesmo considerando-se apenas os valores médios (desconsiderando-se os desvios acumulados). Contudo, a única região onde esta diferença média é sempre minimamente *positiva* (*i.e.*, para a qual os valores médios de incremento anisotrópico para as isotermas de HREs do tipo I<sub>1</sub> atingem valores, pelo ao menos, minimamente maiores que os da isoterma para a seqüência alvo do p53) ainda corresponde à faixa de concentrações onde são visualizadas as espécies intermediárias nos complexos com F2, DR5 e (aparentemente) com DR2 (embora os desvios acumulados não permitam afirmar com certeza estatística). É também interessante notar que, enquanto os incrementos anisotrópicos médios para a interação inespecífica com a seqüência alvo do p53 se mostraram, vias de regra, menores que os incrementos para os HREs do tipo I<sub>1</sub> para as concentrações iniciais de proteína; os mesmos incrementos médios para p53 tenderam a se mostrarem iguais ou relativamente maiores a concentrações mais altas<sup>19</sup>.

Tomados em conjunto, estes resultados sugerem que as isotermas do tipo I<sub>1</sub> (inclusive a fração da interação inespecífica “comum” caracterizada para a seqüência alvo do p53 até a concentração de 1000 nM (acima da qual imperam as interações inespecíficas ”divergentes”)) são representadas basicamente por espécies monoméricas de baixa especificidade do hRXR $\alpha$ - $\Delta$ AB se associando ao DNA com variações nas afinidades relativas (embora as diferenças nos parâmetros de afinidade, bem como os valores de incremento anisotrópico para a maioria destas isotermas sejam por demais sutis para serem significativamente diferenciados por esta técnica, ou mesmo apresentarem significância bioquímica). Espécies de baixa especificidade (facil-

<sup>19</sup> para DR2, os incrementos médios são  $\geq$  os de p53 ao longo de essencialmente toda a isoterma, mas a diferença entre ambos tende a decair entre 200 e 2000 nM, após o que segue-se o termo divergente. É interessante notar que para o TRE-Pal (aqui referido apenas como Pal), enquanto os perfis médios relativos das isotermas corrigidas na ausência e presença do ligante sejam similares, na presença do ligante os incrementos médios corrigidos da isoterma p53 se deslocam para valores mais negativos ao longo de toda a isoterma, sugerindo um certo deslocamento por parte da interação com o agonista para a isoterma de ligação a este HRE.



**Figura 4.8** – Correção do termo inespecífico “comum” para as isotermas do tipo I<sub>1</sub>

mente deslocáveis por oligonucleotídeos não contendo a seqüência consenso) com migração similar à do monômero são comumente observadas (com maior ou menor representatividade) em estudos de gel nativo para a associação do hRXR $\alpha$  a HREs não específicos (<sup>20</sup>). Além disto, as taxas relativamente altas de intercâmbio entre interações específicas e inespecíficas previamente observadas por RMN para a interação do hRXR $\alpha$ -DBD com o meio sítio, bem como a abundância significativamente maior de interações dos DBDs com o *backbone* que com as bases nitrogenadas nas estruturas cristalográficas resolvidas para complexos do homodímero e heterodímeros do hRXR $\alpha$  com o DNA, sugerem a possibilidade da formação de complexos monoméricos inespecíficos na ausência de seqüências permissivas para a formação do dímero.

<sup>20</sup>.

O incremento anisotrópico conferido pela aparente interação de um único monômero a estas espécies se mostrou significativamente menor que o mesmo incremento conferido à seqüência contendo o meio sítio<sup>21</sup> (Fig. 4.1-A). Contudo, esta diferença de incremento é esperada, visto que os oligonucleotídeos que apresentaram isotermas do tipo I<sub>1</sub> neste estudo contêm entre 1.5 até mais de 2 X o número de pares de base do oligo com uma única seqüência arquetípica (HS) (Tab. 3.1). Assim sendo, espera-se, de fato, que a associação ao meio-sítio promova uma significativa maior estabilização da mobilidade local do fluoroforo, bem como um acréscimo mais substancial no tempo de correlação rotacional do complexo, comparado à mesma interação para as seqüências que apresentaram isotermas do tipo I<sub>1</sub>. Além disto, a área ocupada pelo DBD do monômero complexado à seqüência contendo um único meio sítio abrange mais as imediações do centro geométrico (e de massa) da molécula de DNA que o mesmo complexo monomérico para os HREs com isotermas I<sub>1</sub> (para o oligo de 15 pb da Tab. 3.1, este centro se localiza no segundo par de bases (G-C) da sequência arquetípica), proporcionando, a princípio, uma maior estabilização mecânica deste centro de massas, com movimentos vibracionais e

<sup>20</sup> É esperado que a formação de tais complexos monoméricos inespecíficos seja particularmente provável em experimentos em soluções com valores de força iônica relativamente baixos, como é o caso do presente estudo

<sup>21</sup> A presença de uma única seqüência consenso (Tab. 3.1), bem como o coeficiente de Hill igual a 1 para a isoterma do oligonucleotídeo contendo o meio-sítio (Fig. 4.2 permitem inferir que o hRXR $\alpha$ -ΔAB se liga na forma de monômero a este fragmento

rotacionais de maior simetria para o complexo como um todo<sup>22</sup>, o que também é esperado colaborar para o maior valor de incremento anisotrópico ()).

É particularmente notório que a interação inespecífica monomérica para o p53 começa a se manifestar a mais altas concentrações que as mesmas respectivas interações para HS, DR2, DR5 e F2 (Figs. 4.1-A, C, E e F, 4.4-D e 4.7-C e D), sugerindo, como esperado, uma afinidade relativamente menor conforme o menor grau de especificidade da interação. É também notório que esta mesma interação monomérica inespecífica se manifestou de forma adicional à interação dos dímeros para DR1, DR4 e F2. De fato, todos os oligonucleotídeos contendo seqüências arquetípicas de HREs neste estudo possuem uma seqüência 5'-precursora, com pelo ao menos 6 pb, que pode servir de arcabouço para interações de natureza inespecífica mesmo estando os meio-sítios especificamente ocupados (vide Tab. 3.1). Para F2, sobretudo, existem três sequências adicionais com pelo ao menos 6 pb além dos dois meio-sítios, respectivamente localizadas na posição 5' ao primeiro meio-sítio, na posição intermediária entre os dois e na posição 3' ao segundo meio sítio.

A interação inespecífica adicional se manifestou sem implicações significativas quanto à energética das interações específicas, conforme pôde ser visualizado pela manutenção dos valores de  $K_D^{app}$  e (pelo ao menos para DR1 e DR4<sup>23</sup>) de  $n$  em sua ausência. Isto é indicativo de que os monômeros interagindo inespecificamente com os oligonucleotídeos, na presença dos dímeros com maior especificidade, ou não interagem com estes dímeros, ou suas interações com os mesmos apresentam, simultaneamente, incrementos positivos e negativos que se compensam na energética global do complexo.

Os perfis anti-cooperativos para DR0 e DR5 (sobretudo para este último, em que os dados indicam a presença de um intermediário monomérico significativamente dominante em uma faixa considerável de concentrações) sugerem que para estes oligos ocorre a interação

<sup>22</sup> As relações entre os centros de massa do DNA, do(s) DBD(s) e do complexo como um todo; bem como as simetrias vibracionais e rotacionais e sua possível influência na formação e estabilização dos complexos serão melhor discutidos nas análises por dinâmica molecular em

<sup>23</sup> mesmo para F2, a análise dos resultados leva a crer que interferência das interações inespecíficas no valor de  $n$  não ocorre devido à interação inespecífica “comum”, caracterizada pela interação com a seqüência alvo do p53, mas sim por interações inespecíficas de outra natureza (vide 4.1.3 e 4.1.4)

inicial do monômero com relativa facilidade, seguida da associação do dímero com maior penalidade energética<sup>24</sup>. Este aparente maior custo energético para a associação do dímero a estes HREs inespecíficos faz com que a porção final da isoterma (a princípio, mais relacionada com a associação do segundo monômero) seja deslocada para maiores concentrações, inclusive se sobrepondo, em maior ou menor grau, ao termo divergente das interações inespecíficas a altas concentrações<sup>25</sup>. O perfil das mesmas isotermas na presença do ligante sugere que a adição do agonista acrescenta uma relativa maior dificuldade para a interação da espécie monomérica inicial, terminando por deslocar a associação da espécie dimérica final ainda mais para as concentrações correspondentes aos termos divergentes inespecíficos. Desta forma, as isotermas monoméricas do tipo I<sub>1</sub> podem ser consideradas como perfis extremos da anticooperatividade visualizada para DR0 e DR5. Aparentemente, a formação da interface de dimerização com simultâneas interações dos dois monômeros com o DNA é desfavorecida energeticamente para estes HREs de uma forma ainda mais acentuada. Assim sendo, não são observados sinais da formação do dímero na faixa de concentrações menores que as das interações inespecíficas divergentes nas condições experimentais utilizadas.

Assim, a análise em conjunto dos dados referentes aos diferentes experimentos de anisotropia de fluorescência aqui abordados sugerem que as interações visualizadas para o hRXR $\alpha$ - $\Delta$ AB com os diferentes HREs em nosso estudo se dá com diferentes distribuições de espécies e microestados oligoméricos (um monômero de baixa especificidade/dois monômeros de baixa especificidade/dímero de alta especificidade) conforme o espaçamento e organização das sequências arquetípicas e a presença ou não do ligante. Tal padrão de interações é por demais complexo para ser ajustado de forma suficientemente acurada a um modelo semi-empírico que utiliza uma expressão analítica "fechada" de fração-ligada, como é o caso do modelo de Hill. Desta forma, visando melhor refinar nossa modelagem da energética de associação e consequentes ocupâncias relativas das diferentes espécies oligoméricas do hRXR $\alpha$ - $\Delta$ AB a diferentes HREs, foi efetuado um tratamento numérico relativamente mais robusto das isotermas, em que as

<sup>24</sup> este perfil é justamente o oposto do observado para a formação dos complexos cooperativos com DR1 e DR4

<sup>25</sup> esta sobreposição, ao menos parcial, da isoterma relativa à espécie dimérica final com as interações inespecíficas divergentes a altas concentrações é também constatada, aparentemente, para a interação do hRXR $\alpha$ - $\Delta$ AB com o F2 pela rota não colaborativa na ausência do ligante (Figs. 4.3-C e D e 4.5-C)

equações relacionando a fração populacional das diferentes espécies às suas respectivas energias de formação são simultaneamente ajustadas a cada conjunto de dados, de forma a melhor corresponder aos próprios dados e às equações de balanço de massa . Os resultados referentes a estas análises são apresentados e discutidos na próxima subseção.

#### **4.1.6 Modelagem numérica global multi-equações corrobora a hipótese de que a interação do hRXR $\alpha$ -ΔAB com HREs específicos e não específicos se diferencia pela distribuição de monômeros e dímeros associados, sendo estas distribuições influenciadas pelo ligante**

Para o ajuste numérico global das respectivas isotermas, foi utilizado o programa BIOEQS . O algorítimo utilizado por este programa na análise numérica global de dados para isotermas de afinidade foi melhor descrito em .

O modelo empregado para o ajuste das isotermas foi o de um sistema em que o receptor pode se ligar ao DNA tanto na forma de monômero (estequiométria 1:1) quanto na forma dimérica (estequiométria 2:1). A primeira energia livre característica para este modelo,  $\Delta G^\circ_1$ , corresponde à energia livre de dissociação de um complexo 1:1 monômero:DNA (RXR:DNA) até o monômero (RXR) e DNA livres, enquanto a segunda energia,  $\Delta G^\circ_2$ , corresponde à energia livre de dissociação de um complexo de estequiométria 2:1 monômero:DNA (RXR<sub>2</sub>:DNA), também até o monômero e DNA livres. De forma a calcular a energia de dissociação do segundo monômero até RXR:DNA e RXR livre,  $\Delta G^\circ_{21}$ , subtrai-se o valor de  $\Delta G^\circ_1$  do valor de  $\Delta G^\circ_2$ . A diferença entre  $\Delta G^\circ_2$  e 2 X o valor de  $\Delta G^\circ_1$  provê a energia livre de colaboratividade entre a interação do primeiro e do segundo monômero ( $\Delta\Delta G^\circ_{diss}$ ).

De forma a evitar tendenciosidades por parte da análise dos dados, os ajustes foram realizados para as respectivas isotermas na forma dos dados "brutos", sem as correções pelas aparentes interações inespecíficas mostradas nas Figs 4.7 e 4.8. As análises foram assim realizadas visto que, primeiramente, os dados indicam que, devido ao incremento anisotrópico relativamente pequeno decorrente destas interações inespecíficas, bem como sua localização após o início do termo assumptótico, a influência das mesmas não deturpa significativamente os perfis das

isotermas específicas. Em segundo, quanto o caráter bimodal das isotermas para DR5 se torne mais claro após estas correções, as mesmas tratam-se de estimativas, visto que não foram realizados experimentos de competição inespecífica também para este HRE. Além disto, os dados sugerem que a influência do ligante nestas interações inespecíficas pode ser variável entre diferentes HREs<sup>26</sup>, de forma a que sua correção indiscriminada das isotermas na ausência e presença do 9-cRA pode acometer em erros. Por último (mas não menos importante), o ajuste numérico destas isotermas sem estas correções torna-se um instrumento de análise adicional e independente para se corroborar (ou não) as conclusões previamente firmadas<sup>27</sup>.

Como mencionado previamente, para que o algoritmo do programa BIOEQS possa incorporar as constrições na forma de multiplicadores de Lagrange relacionadas às equações de balanço de massa, a contribuição de cada espécie para a observável experimental (no caso o incremento anisotrópico) deve ser indicada pelo usuário. Desta forma, as contribuições anisotrópicas correspondentes ao DNA livre e ao complexo final foram igualadas aos valores de anisotropia nos platôs iniciais e finais, respectivamente, e adicionadas como valores flutuantes, a serem otimizados ao longo das subsequentes iterações de minimização por Maquardt-Levenberg. Para uma melhor convergência dos parâmetros recuperados por este tipo de ajuste, necessita-se que o valor de incremento anisotrópico referente ao intermediário seja fixado (preferivelmente, a um valor próximo à metade do incremento entre os dois platôs). Desta forma, primeiramente foram feitos ajustes testes em que o valor de anisotropia referente ao intermediário foi fixado em diferentes valores entre 25 e 75 % da variação entre o platô inicial e o final (não mostrado). Uma vez que não foram observadas mudanças significativas para os va-

---

<sup>26</sup> Repare que, enquanto para DR1 a adição do ligante não muda a intensidade anisotrópica, para DR4 esta adição diminui a intensidade aproximadamente da mesma grandeza que o incremento inespecífico das Figs. 4.3-B e 4.4-B (Figs. 4.1-D e 4.2-A)

<sup>27</sup> Por exemplo, se para as isotermas de DR5 sem as correções ilustradas na Fig. 4.7 a análise numérica recupera um caráter bimodal condizente com o mesmo caráter que se torna evidente após as referidas correções (o que, de fato, se verificou (vide texto abaixo)), esta análise independente reforça a conclusão quanto à existência do intermediário para DR5, bem como a validade das correções apresentadas na Fig. 4.7 (*i.e.*, reforça a hipótese quanto a que a interação inespecífica caracterizada para a sequência alvo do p53, e que se sobrepõe às frações inespecíficas das isotermas para DR1 e DR4, é de fato representativa para as interações inespecíficas dos outros oligonucleotídeos analisados neste estudo).

lores de  $\Delta G^\circ$  recuperados<sup>28</sup>, os valores de incremento anisotrópico referentes aos intermediários foram fixados em 50 % do incremento total para as respectivas isotermas nas quais a presença do mesmo não era evidente e, no caso de F2 e DR5, respectivamente fixados aos valores aparentes nas Figs. 4.1-E e F e 4.4-D. Também o valor da contribuição anisotrópica para o monômero livre (RXR) foi devidamente fixado em zero, visto que, uma vez que a sonda fluorescente se encontra covalentemente associada ao DNA nos experimentos, é esperado que apenas as espécies contendo o DNA contribuam para o sinal.

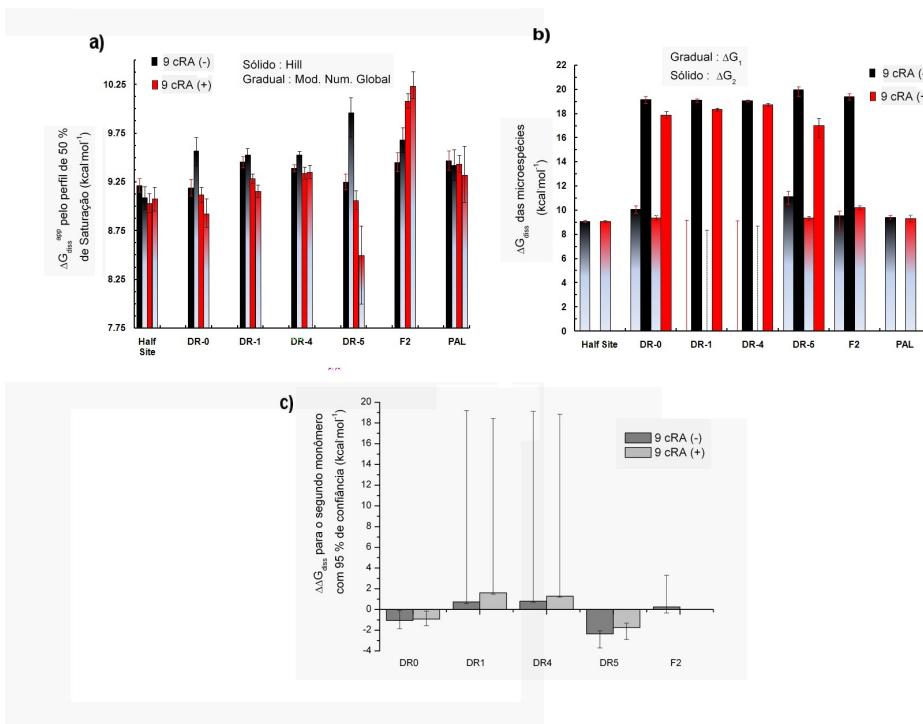
Foram realizados ajustes globais para as isotermas do tipo I<sub>2</sub> e para três isotermas representativas do tipo I<sub>1</sub>: F2 na presença do ligante e Pal na presença e ausência do mesmo<sup>29</sup>. O modelo adotado para a isoterma de ligação do hRXR $\alpha$ -ΔAB ao meio sítio (HS), bem como para as isotermas do tipo I<sub>1</sub> foi o da formação do complexo simples monômero:DNA (RXR:DNA), caracterizado portanto por um único valor de energia livre,  $\Delta G^\circ_{diss} = \Delta G^\circ_1$ . Além disto, para as isotermas visivelmente cooperativas, foram realizados ajustes numéricos globais tanto seguindo o modelo que admite a possibilidade de associação de um primeiro monômero ao DNA quanto um modelo obrigatoriamente cooperativo, em que a formação do complexo RXR<sub>2</sub>:DNA a partir do RXR e DNA livres se dá sem a presença do intermediário RXR:DNA. Os parâmetros, seus respectivos desvios e os valores de discrepância ( $\chi^2$ ) recuperados para os ajustes seguindo estes dois modelos foram devidamente comparados .

Na Fig. 4.9-A os valores de  $\Delta G_{diss}^{app}$  estimados através do ajuste de Hill e do ajuste numérico global com o programa BIOEQS são comparados. Nota-se que os ajustes das isotermas para a maioria do HREs, na ausência e presença do ligante, recuperam valores estatisti-

<sup>28</sup> Isto ocorre porquê, uma vez que os valores correspondentes aos termos assumptóicos inicial e final são adicionados como valores flutuantes ao longo do procedimento de minimização, o algoritmo de ajuste do programa é mais sensível ao perfil global de inclinação e inflexões da isoterma (como, de fato, se espera) que ao valor fixo arbitrariamente atribuído ao intermediário (desde que este valor se encontre, aproximadamente, na faixa entre 25 e 75 % do incremento total - em que existe uma abundância considerável de todas as espécies com significativa representatividade para os dados - o que é uma assunção razoável e condizente com os intermediários aparentemente evidentes em nossos experimentos), permitindo-se que pequenas correções simultâneas para os valores flutuantes de ambos os platôs mantenham os mesmos valores de  $\Delta G^\circ$  para diferentes valores fixos de incremento do intermediário dentro desta faixa.

<sup>29</sup> o perfil anisotrópico de F2 na presença do ligante é mais similar ao de DR2 - com o incremento inicial ocorrendo para concentrações pouco maiores que 1 nM e atingindo um termo assumptólico final próximo a 100 nM-, enquanto a isoterma de interação para Pal se assemelha mais à de DR3 - com um arraste inicial até pouco mais de 10 nM e atingindo um platô final a concentrações um pouco maiores que 100 nM (Figs. 4.1-C,D e F e 4.8).

camente similares de  $\Delta G_{diss}^{app}$  entre os dois métodos numéricos<sup>30</sup>. As duas únicas diferenças significativas são notadas para as isotermas dos dois HREs anti-cooperativos - DR0 e DR5. Embora já se note, para o ajuste não linear de Hill, uma certa tendência para menores valores médios de  $\Delta G_{diss}^{app}$  para as isotermas destes dois HREs na presença que na ausência do ligante, as diferenças recuperadas por meio deste ajuste não são estatisticamente significativas. Contudo, o ajuste numérico global recuperou valores significativamente maiores de  $\Delta G_{diss}^{app}$  na ausência do ligante e menores na presença do mesmo para as isotermas de interação a DR0 e DR5 (sendo esta diferença consideravelmente mais substancial para DR5) (Fig. 4.9-A), indicando uma influência negativa por parte do ligante para a associação global (considerando tanto as microespécies monoméricas quanto diméricas) do hRXR $\alpha$ - $\Delta$ AB a estes dois HREs.



**Figura 4.9 – Parâmetros de energia livre recuperados através do ajuste numérico global pelo BIOEQS e comparação com o ajuste de Hill**

<sup>30</sup> O ajuste global recupera parâmetros com incertezas relativamente maiores. Isto porquê, enquanto as incertezas para o ajuste não linear pelo programa Sigma Plot® (através do qual o ajuste de Hill foi realizado) são recuperadas da forma comumente utilizada - através da diagonalização da matriz de correlação entre os parâmetros; as mesmas incertezas para o ajuste global no programa BIOEQS são recuperadas através de um conjunto mais rigoroso de procedimentos - através do reajuste da curva para cada valor teste de cada respectivo parâmetro, com a simultânea covariância dos demais, checando-se os limites dos intervalos com 67 e/ou 95 % de confiabilidade (respectivamente, um desvio padrão e 95 % da densidade de probabilidade em uma distribuição normal) para cada parâmetro individualmente testado (baseado nos valores de  $\chi^2$  global de cada ajuste). Neste estudo, os valores de incerteza utilizados foram os correspondentes ao intervalo com 67 % de confiabilidade. O método utilizado pelo programa BIOEQS provê intervalos de confiabilidade mais realistas (e maiores) que o utilizado pelo Sigma Plot, além de levar em consideração a correlação entre os parâmetros, o que não ocorre para o método de diagonalização

A natureza da diferença entre os parâmetros de variação de energia livre global aparentes recuperados para os dois HREs anti-cooperativos através dos dois ajustes se explica através da análise dos valores de  $\Delta G^\circ_1$  e  $\Delta G^\circ_2$  recuperados para as isotermas destes elementos através do ajuste global (Fig. 4.9-B). Nota-se que os valores de  $\Delta G^\circ_{21}$  (calculados pela subtração de  $\Delta G^\circ_1$  de  $\Delta G^\circ_2$ ) são significativamente menores que os valores recuperados para  $\Delta G^\circ_1$ , conforme pode ser vislumbrado para os valores negativos de  $\Delta\Delta G^\circ_{diss}$  ilustrados para estes dois HREs na Fig. 4.9-C (sendo este padrão mais substancial e estatisticamente significativo para DR5). Este perfil é condizente com a anti-cooperatividade aparente nos dados para os dois HREs, indicando a presença de um intermediário monomérico significativamente estável durante a titulação dos mesmos com o hRXR $\alpha$ - $\Delta$ AB. Sobretudo, o aparente platô intermediário para a isoterma de DR5 na ausência do ligante é melhor ajustado pela modelagem numérica global que pelo modelo de Hill<sup>31</sup>. Isto ocorre porquê o modelo de Hill não suporta a presença de intermediários estáveis ao longo do processo, sendo os pontos correspondentes ao platô intermediário promediados pela curva e o valor de  $K_D^{app}$  recuperado referente apenas à espécie representativa do termo assimptótico final (no caso, o complexo RXR<sub>2</sub>:DNA). Desta forma, o valor de  $\Delta G_{diss}^{app}$  calculado a partir do  $K_D^{app}$  recuperado no modelo de Hill para um processo de significativa anti-cooperatividade tende a ser menor que o valor de  $\Delta G^\circ_2/2$  para a modelagem global, visto que para o cálculo do primeiro a contribuição da espécie de maior valor de  $\Delta G^\circ_{diss}$  (o complexo RXR:DNA) não é levada em consideração. Já na presença do ligante, nota-se uma diminuição de ambos os parâmetros de energia livre, sendo esta diminuição relativamente mais substancial para o complexo monômero:DNA, sugerindo que a interação com o agonista adiciona uma penalidade adicional para a interação monomérica do hRXR $\alpha$ - $\Delta$ AB a estes HREs (provavelmente devido a uma maior estabilização do dímero em solução) e, por conseguinte, desloca partes consideráveis da interação dimérica (dependente da formação prévia do complexo RXR:DNA e ainda desfavorecida energeticamente em relação a este) para concentrações ainda mais próximas do termo divergente das interações inespecíficas (sendo este comportamento condizente com o sugerido pela análise da isoterma de DR5 corrigida pela interação inespecífica “comum” na

---

<sup>31</sup> O ajuste numérico para DR5 na ausência do ligante, bem como os demais ajustes ilustrados na Fig. 4.1 são os realizados utilizando o programa BIOEQS

Fig. 4.7). Este perfil provoca uma nova divergência entre os valores de  $\Delta G_{diss}^{app}$ , seguindo um padrão contrário à diferença anteriormente listada no ajuste para as isotermas na ausência do ligante: uma vez que o complexo monomérico é mais representativo para o termo assimptótico final nestes elementos na presença do agonista, o valor de  $\Delta G_{diss}^{app}$  estimado através do modelo de Hill é mais reflexivo desta espécie, sendo portanto mais próximo ao valor de  $\Delta G^\circ_1$ , que por conseguinte é significativamente menor que o valor de  $\Delta G^\circ_2/2$  pelo qual  $\Delta G_{diss}^{app}$  é estimado no modelo global<sup>32</sup>. Esta divergência é particularmente mais notória nos parâmetros recuperados para DR5 na presença do ligante, novamente confirmando que o perfil intermediário entre as curvas do tipo I<sub>1</sub> e I<sub>2</sub> para esta isoterna se deve a uma maior fração de espécies monoméricas nas regiões representativas da titulação (entre 25 e 75 % do incremento entre os dois platôs).

Para os dois HREs que apresentaram isotermas altamente cooperativas - DR1 e DR4 - nota-se uma acentuada incerteza nos valores de energia recuperados para a espécie monomérica (Fig. 4.9-B) não sendo possível afirmar com precisão, mediante as estatísticas dos ajustes numéricos, o quão menor é a energia de dissociação da espécie monomérica em relação à dimérica, ou mesmo se ela virtualmente não se forma, ocorrendo a associação do hRXR $\alpha$ - $\Delta$ AB diretamente como dímero, seguindo o padrão que no modelo de Hill é denominado como *cooperatividade infinita* (limite inferior dos respectivos valores de  $\Delta G^\circ_1 \rightarrow 0$ ). Este alto grau de incerteza para a associação da primeira espécie é comum para processos de associação com acentuado caráter cooperativo, uma vez que sob estas condições a espécie intermediária não é suficientemente populada em nenhuma região da isoterna, não se obtendo, portanto, suficientes informações a respeito de sua energética no conjunto de dados . De fato, como anteriormente discutido em 4.1.2 os valores de  $n$  próximos a 2 previamente recuperados pelo ajuste não linear ao modelo de Hill são condizentes com perfis de cooperatividade “infinita” para a associação do hRXR $\alpha$ -

<sup>32</sup> Em uma aproximação simplista, pode-se dizer que para as isotermas na ausência do ligante, em que a influência do intermediário é mais notória no perfil da curva, o valor de  $\Delta G_{diss}^{app}$  estimado a partir do ajuste numérico global é mais próximo a um valor “real” (mais reflexivo da energia livre estimada a partir da concentração para saturação da metade dos sítios de DNA disponíveis em solução), enquanto que na presença do ligante, em que a segunda espécie é mais deslocada para as concentrações finais da isoterna (sobretudo para a interação com DR5) este valor “real” seria mais próximo do valor recuperado pelo ajuste de Hill. Em uma análise mais acurada, contudo, nenhum dos dois valores é, de fato, representativo da ocupância das diferentes espécies ao longo da titulação, estando esta informação melhor contida nos valores individuais de  $\Delta G^\circ_1$  e  $\Delta G^\circ_2$  recuperados no ajuste global, bem como na relação entre ambos.

$\Delta AB$  a estes dois HREs. Apesar das acentuadas incertezas, as diferenças entre os valores de  $\Delta G^\circ_2$  e  $\Delta G^\circ_1$  recuperados nos ajustes numéricos globais para as isotermas de interação com DR1 e DR4 na ausência e presença do ligante, mesmo em seus limites inferiores (0.588, 0.710; 1.48 e 1.19, respectivamente), são condizentes com processos notoriamente cooperativos (Fig. 4.9-C). Além disto, as isotermas para DR1 e DR4 se ajustaram com menores valores de  $\chi^2$  e com valores similares de  $\Delta G^\circ_{diss}$  para o dímero a um modelo do tipo "tudo ou nada", em que não é admitida a presença do intermediário monomérico e o equilíbrio é considerado como do tipo :



,demonstrando novamente a dominância da interação da espécie dimérica, com alta vantagem energética, para as isotermas de associação do hRXR $\alpha$ - $\Delta AB$  a estes dois HREs.

## 4.2 Estudos de modelagem estrutural e dinâmica molecular das interações do receptor nuclear do ácido 9-cis- retinóico a Elementos Responsivos de DNA

### 4.2.1 Considerações prévias a respeito das componentes do sistema - dinâmica conformacional do domínio DBD e HREs e suas possíveis implicações na formação do complexo

Como mencionado em , estudos prévios de RMN indicam que a interação de um único monômero de hRXR $\alpha$ -DBD à sequência do meio-sítio induzem a significativas mudanças dinâmicas e conformacionais na porção N-terminal do *T-box* (resíduos 200 a 209) e no *loop* C-terminal do segundo motivo dedo de zinco (resíduos 182 a 187). Para o T-box, a interação com o meio-sítio induz à desnaturação local da hélice C, observada em solução, de forma a permitir a interação dos resíduos 204 a 209 com o sulco menor na posição 5' anterior ao meio-sítio. Esta região estendida e alocada sobre o sulco menor promove a interface de dimerização com o monômero 5' nas estruturas cristalográficas de homo e heterodímeros de hRXR $\alpha$  em DR1 . Já o *loop* contendo os resíduos 182 a 187 passa de um motivo desordenado e de alta

flexibilidade conformacional em solução para uma conformação de hélice distorcida quando associado ao DNA . Esta região do DBD interage simultaneamente com o *backbone* do DNA e com o monômero 3' nas estruturas cristalográficas. . Os estudos de RMN também revelaram que a ligação de um único monômero ao meio-sítio promoveu significativas mudanças na dinâmica da região do DBD denominada como *D-box* (resíduos Arg172 a Asp176, no "topo" do segundo dedo de zinco e que fazem parte da interface de dimerização com o monômero 3'), bem como do motivo *β-hairpin* na extremidade do primeiro dedo de zinco (que faz parte da interface proteína-DNA) . Embora estes experimentos indiquem que a associação de um único monômero ao DNA induza a mudanças conformacionais e dinâmicas que podem ser, a princípio, permissivas para a associação entre os monômeros, os mecanismos através dos quais estas mudanças, muitas das quais em regiões distantes da interface proteína-DNA, são induzidas e estabilizadas não são claramente compreendidos. Além disto, o foco destes trabalhos se restringe às mudanças induzidas unicamente na dinâmica conformacional da proteína, não sendo abordados os efeitos na topologia e mobilidade dos oligonucleotídeos. Visto o fato da interface proteína-proteína ser relativamente pobre nos complexos dos DBDs ao DNA e serem os DBDs aparentemente determinantes para a cooperatividade e seletividade na associação de NRs a diferentes HREs é esperado que mudanças topológicas e dinâmicas co-induzidas entre a proteína e o DNA sejam diretrizes para estas mesmas cooperatividades e seletividades. Também as conhecidas polaridades 5':3' dos heterodímeros de RXR $\alpha$  com diferentes parceiros em diferente HREs é mantida, vias de regras, para os complexos contendo apenas os respectivos DBDs<sup>33</sup> .

Um outro indício de implicações na estrutura e dinâmica conformacional do DBD para a atividade biológica do complexo como um todo são os estudos de Meijssing et al para a interação de receptores de glucocorticoides (GRs) a diferentes seqüências responsivas a GRs (GBSs). Estes estudos se basearam na comparação de ensaios de afinidade por *gel shift* e de atividade biológica *in situ* para os complexos da construção ΔAB às estruturas cristalográficas dos complexos DBD do receptor a diferentes GBSs. Ficou demonstrado que variações tão sutis quanto um único par de bases na seqüência espaçadora entre os dois meio-sítios induziram a modificações significati-

<sup>33</sup> uma exceção é o complexo de RXR $\alpha$  e VDR em DR3, para o qual foi demonstrado ser necessária a presença de ambos os LBDs para a devida manutenção da interação alostérica e da correta polaridade

vas na conformação e, aparentemente, na dinâmica (conforme inferido pelos respectivos fatores de Debye-Waller e pelas análises dos mapas de densidades omitidas (*omit maps*) das estruturas cristalográficas) no *loop* entre a hélice A e o segundo dedo de zinco (denotado como *lever arm*) que, por sua vez, implicaram em mudanças nas atividades relativas, na utilização de diferentes interfaces de dimerização entre os receptores (tanto para os DBDs quanto para os LBDs) e no recrutamento de diferentes co-ativadores. Embora a correlação entre estas mudanças funcionais e o grau de distorção do *lever arm* tenha ficado óbvia neste trabalho, a mesma correlação nem sempre ocorreu com as afinidades para os diferentes GBSs, sendo observadas mudanças significativas nos ensaios de atividade *in situ* e na estrutura cristalográfica para complexos com valores de  $K_{DS}$  similares. Desta forma, estes estudos sugerem que mudanças conformacionais na *lever arm* de GR amplificam a sinalização inerente a diferenças (por vezes sutis) nas seqüências responsivas e transmitem esta sinalização para os outros domínios, promovendo um ajuste fino da resposta trancicional a diferentes GBSs, ao contrário de uma regulação do tipo *on/off*, como é comumente assumido para NRs e outros fatores de transcrição.

Assim sendo, embora tanto a análise cristalográfica para o complexo das estruturas completas , quanto diversos estudos bioquímicos e biofísicos em solução indiquem que as interações envolvendo os LBDs, ou mesmo o domínio N-terminal corroborem para a estabilização do complexo e/ou para a acentuação da seletividade, fica claro que as bases determinantes para estas mesmas afinidades e seletividades (e, em alguns casos, mesmo para a atividade diferenciada) de NRs a diferentes HREs residem nas diferenças conformacionais e/ou dinâmicas para a associação dos DBDs.

Buscando, então, melhor compreender as bases moleculares para o alosterismo e a seletividade de receptores nucleares na associação ao DNA, foram realizados estudos de modelagem estrutural e dinâmica molecular para o complexo de hRXR $\alpha$ -DBD associado a seu HRE responsável DR1 com diferentes modos de arranjo (monômero ligado ao sítio 5', monômero ligado ao sítio 3' e dímero). Nas análises que se seguem, os resultados referentes às restrições energéticas e dinâmicas e suas possíveis implicações para o alosterismo, seletividade e polaridade na interação destes receptor serão estendidas para outros receptores e seus sítios específicos.

cos. Estes mesmos resultados, em conjunto com a comparação com a estrutura cristalográfica do homodímero de hRXR $\alpha$ :hPPAR $\gamma$  full em DR1 e de dados estruturais de baixa resolução na literatura, serão utilizados para se inferir quanto às implicações dinâmicas e energéticas da associação dos DBDs para os complexos com a proteína completa.

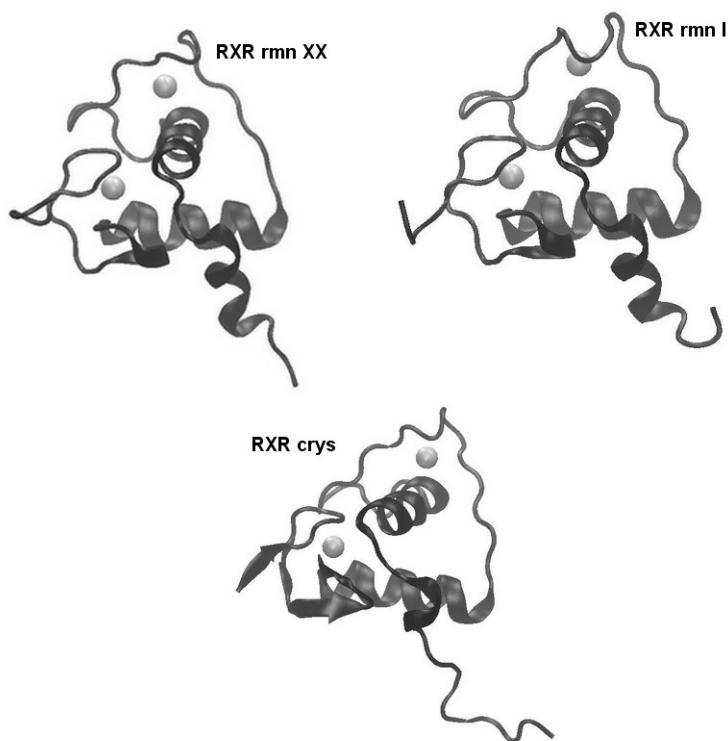
Em uma primeira abordagem, serão exploradas as características dinâmicas referentes às duas componentes do complexo - os respectivos DBDs e HREs - em separado. Assim, nas próximas duas subseções, caracterizações prévias das respectivas mobilidades, correlações dinâmicas e interações relevantes individualmente observadas em estudos de MD, tanto para os DBDs de hRXR $\alpha$ -DBD e de hTR $\beta$ -DBD a partir de diferentes conformações iniciais, quanto para diferentes sequências de HREs livres em solução, serão apresentadas e discutidas. Na terceira subseção, estas análises serão extendidas para os complexos monômero:HREs. Finalmente, nas subseções referentes à terceira parte desta seção, serão analisadas as trajetórias de MD a partir de modelos do homodímeros de hRXR associados a DR1. Ao longo desta seção, sempre que possível e necessário, as respectivas correlações com os dados experimentais previamente apresentados de anisotropia de fluorescência para a cooperatividade da interação do hRXR $\alpha$  serão discorridas.

#### **4.2.2 Conhecendo as componentes do sistema (I) - Desempacotamento hidrofóbico decorrente da desestruturação do T-box na conformação de interação do hRXR $\alpha$ -DBD ao DNA induz mudanças na dinâmica em nanosegundos e na conformação média de todo o domínio de forma possivelmente funcional**

De forma a se obter um melhor conhecimento acerca das implicações dinâmicas decorrentes das mudanças conformacionais do hRXR $\alpha$ -DBD ocorridas durante a associação ao DNA, foram primeiramente realizados estudos de MD do hRXR $\alpha$ -DBD nas conformações comumente adotadas em solução e no complexo.

Foram realizadas simulações a partir de 3 configurações iniciais para o domínio hRXR $\alpha$ -DBD livre em solução(Fig. 4.10A, B e D): as conformações respectivamente denotadas como 1 e 20 das 20 configurações da estrutura resolvida por RMN (código do PDB *lrxr*), referidas

no texto que se segue como  $\text{RXR}_{RMN}^I$  e  $\text{RXR}_{RMN}^{XX}$  respectivamente, e a estrutura contendo o domínio DBD e porção N-terminal do domínio D (*hinge*) do hRXR $\alpha$  do complexo cristalográfico do heterodímero completo de hRXR $\alpha$ /hPPAR $\gamma$  em DR1(código do PDB 3DZY), referida como  $\text{RXR}_{crys}$ , todas as três estruturas consistindo dos resíduos K132 a G212 do receptor. Como previamente mencionado, a região N-terminal do domínio D, o T-box, adota uma conformação em hélice 3<sub>10</sub> (a *hélice C*) nas estruturas resolvidas por RMN para o monômero DBD em solução , enquanto que na estrutura cristalográfica ela se encontra em uma conformação desordenada, interagindo diretamente com o sulco menor e (nos complexos em DR1) com o DBD do monômero 5' .



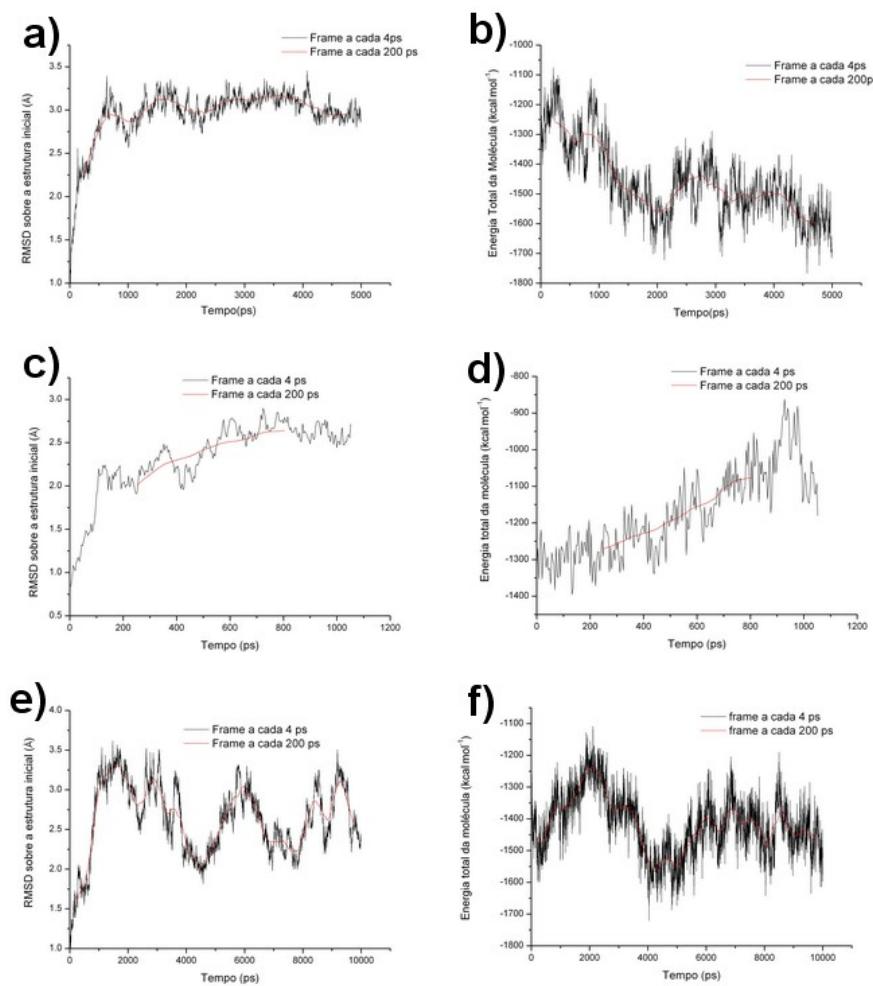
**Figura 4.10** – Estruturas iniciais para as simulações de monômero DBD de hRXR $\alpha$  em solução

Os progressos de rmsd dos átomos pesados da proteína relativos à estrutura inicial e de energia total da molécula durante as duas simulações com configurações iniciais referentes às respectivas conformações resolvidas por RMN (Fig. 4.11-A-D)<sup>34</sup> são condizentes com o relativo maior empacotamento (conforme aparente nas estruturas originais) (Fig. 4.10-A e B)

<sup>34</sup> Como previamente mencionado em , os perfis exibidos e analisados nestas e nas demais figuras relativas às simulações por MD realizadas neste estudo, salvo se explicitamente mencionado o contrário, são referentes às simulações após as etapas iniciais de minimização e relaxamento do solvente por dinâmica com restrição

para a conformação  $\text{RXR}_{RMN}^{\text{XX}}$  em relação à  $\text{RXR}_{RMN}^{\text{I}}$ . Embora a MD para a  $\text{RXR}_{RMN}^{\text{I}}$  tenha sido realizada durante um maior tempo de simulação computacional que o da  $\text{RXR}_{RMN}^{\text{XX}}$  (5000 e 1050 ps, respectivamente), acima de 600 ps, o rmsd em relação à configuração inicial para a segunda demonstra uma tendência de saturação a um valor intermediário entre 2.5 e 2.75 Å, enquanto que para a primeira, no mesmo intervalo de tempo, o rmsd se aproxima de 3.00 Å, se mantendo entre este valor e aproximadamente 3.10 Å durante o restante da simulação (Fig. 4.11-A e C). Além disto, ao longo dos aproximadamente 1000 ps de simulação para a conformação  $\text{RXR}_{RMN}^{\text{XX}}$ , a energia potencial da proteína tende a aumentar (de aproximadamente 300 kcal·mol<sup>-1</sup> no decorrer da MD), não demonstrando uma saturação clara até o final da janela de tempo da modelagem (Fig. 4.11-D). Este padrão indica que, decorridos 1050 ps após terem sido retiradas as restrições harmônicas dos átomos pesados da proteína, o calor ainda "flui" preferencialmente do sistema para a molécula de forma a alcançar a termalização, demonstrando uma considerável resistência térmica para o confôrmero  $\text{RXR}_{RMN}^{\text{XX}}$  o que, por sua vez, é condizente com uma configuração inicial com relativa rigidez. Para a conformação  $\text{RXR}_{RMN}^{\text{I}}$ , por outro lado, no intervalo dos primeiros 1000 ps de simulação, a energia da molécula oscila desta mesma quantidade (aumenta do mesmo valor aproximado de 300 kcal·mol<sup>-1</sup> em apenas 250 ps, diminui deste valor nos próximos 250 ps (tempo de simulação de 500 ps), se mantém por mais 250 ps e aumenta novamente), indicando uma maior mobilidade e consequente maior taxa de trocas térmicas logo a partir da conformação inicial (pós minimização e relaxamento do solvente e íons) (Fig. 4.11-B *inlet*). Decorrido este intervalo de tempo inicial, o valor médio da energia potencial "relaxa" de aproximadamente 250 kcal·mol<sup>-1</sup>, atingindo um aparente platô em 2000 ps e oscilando sobre um valor médio de -1517.3 com amplitude média de variação  $\pm 67.5$  kcal·mol<sup>-1</sup> nos últimos 3000 ps da simulação, demonstrando que o equilíbrio térmico foi atingido (Fig. 4.11-B).

Os mesmos gráficos de rmsd sobre a estrutura inicial e de energia total da molécula Vs tempo de simulação para o monômero na conformação oriunda da estrutura cristalográfica (estrutura induzida pela associação com o DNA) demonstram uma significativa maior mobilidade global da molécula em solução (Figs 4.11-E e F), com movimentos de maior ampli-



**Figura 4.11** – Progressão do Equilíbrio para as simulações de hRXR $\alpha$ -DBD para diferentes configurações iniciais em solução.

tude e menores frequências<sup>35</sup> comparado às duas configurações resolvidas por RMN (representativas da conformação global comumente adotada em solução), em concordância com a desnaturação da porção C-terminal da hélice C (o T-box) e seu consequente desempacotamento do cerne do domínio, conforme evidente na estrutura inicial (Fig. 4.10-C). A energia potencial média da molécula no equilíbrio oscila com maiores amplitudes, porém sobre um valor médio também maior (em aproximadamente 100 kcal·mol<sup>-1</sup>) que o da configuração de RMN RXR<sub>RMN</sub><sup>I</sup> (valor médio da energia da molécula e respectiva amplitude média de variação para a trajetória do monômero RXR<sub>crys</sub> no equilíbrio de -1416.9 ± 79.0 kcal·mol<sup>-1</sup>), sugerindo que a desnaturação do T-box, com consequente menor empacotamento deste sobre o cerne do DBD, confere ao domínio uma maior entropia conformacional, porém com uma considerável penalidade entálpica.

A acentuada contribuição das interações hidrofóbicas para o empacotamento e estabilização da conformação do domínio em solução pode ser constatada comparando-se os gráficos de energia total da molécula nas Figs. 4.11-B,D e E com as energias de interação com as moléculas de água do solvente<sup>36</sup> (Fig. 4.12). Nota-se que para as duas simulações mais longas (RXR<sub>RMN</sub><sup>I</sup> e RXR<sub>crys</sub>), quanto o valor médio da energia total da molécula no equilíbrio diminua (RXR<sub>RMN</sub><sup>I</sup>), ou se mantenha aproximadamente constante (RXR<sub>crys</sub>) em relação ao valor inicial, as energias de hidratação da proteína aumentam consideravelmente em ambas as trajetórias (Fig. 4.12-A e C), denotando uma tendência da liberação de moléculas de água para o meio durante o processo de termalização do monômero que é, no caso do confôrmero RXR<sub>RMN</sub><sup>I</sup>, acompanhada por um evidente aumento no empacotamento interno da molécula (Vide a anticorrelação entre

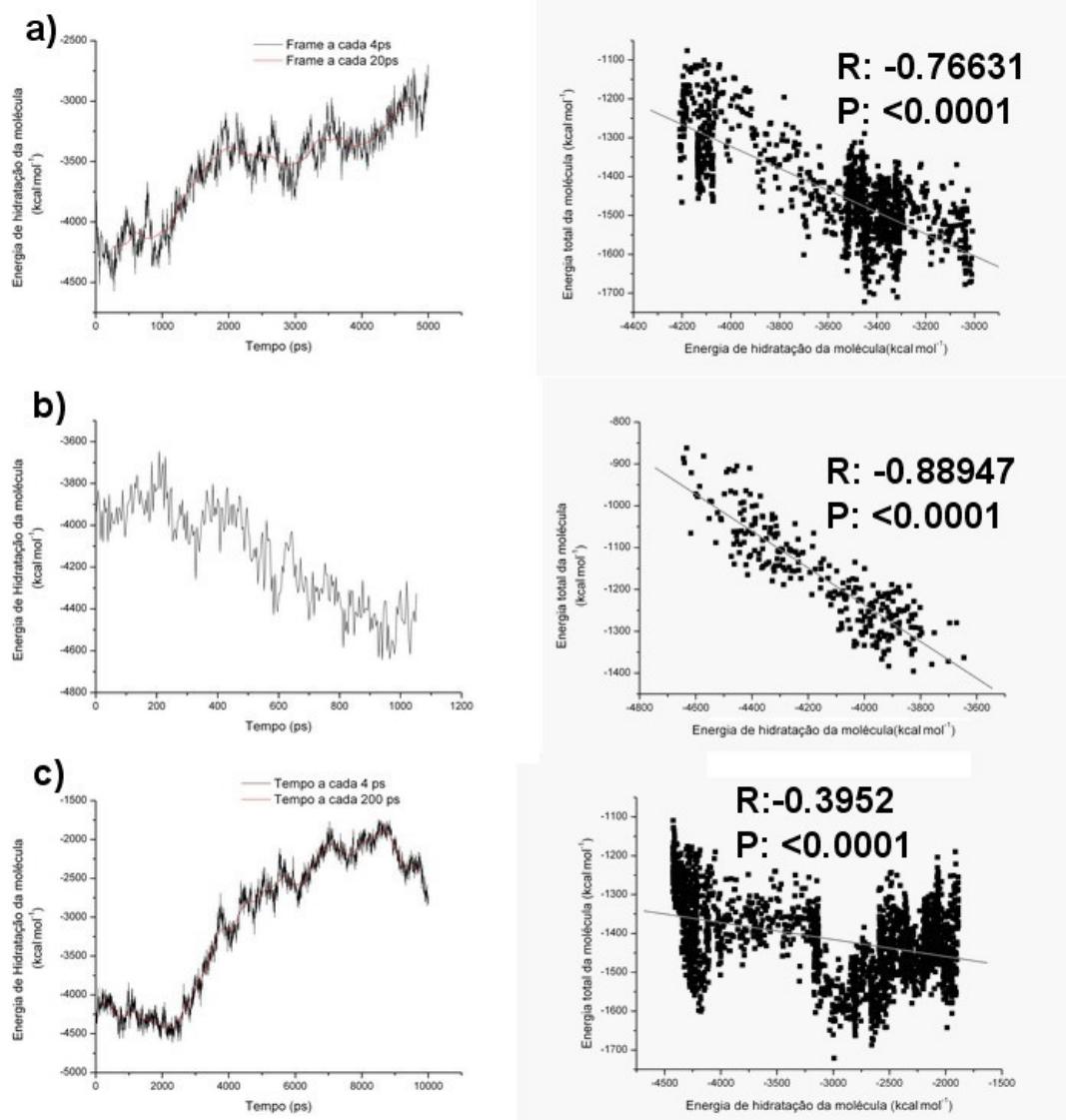
<sup>35</sup> Devido aos períodos de correlação significativamente maiores para os movimentos do monômero DBD na conformação RXR<sub>crys</sub>, a simulação para o mesmo foi realizada até o intervalo de 10000 ps

<sup>36</sup> Embora não seja o termo mais correto, a soma das energias de interação de caráter coulombico da molécula com as águas em um raio de 12 Å e de caráter dispersivo (vdw) com as águas em um raio de 5 Å, ambas calculadas pelo programa NAMD ao longo da trajetória da simulação, serão doravante referidas como *energias de hidratação* da molécula, meramente para fins de simplificação. Em verdade, a energia de hidratação, no sentido em que o termo é comumente utilizado em textos de físico-química (como sinônimo de *energia de solvatação* do sistema em solução aquosa), seria a soma das diferenças por mol dos respectivos termos entálpicos e entrópicos para a molécula em uma solução aquosa com diluição *infinita* em relação aos dois sistemas em separado (*i.e.*, a molécula de proteína no vácuo e uma solução de igual número de moléculas de água na ausência da proteína). Além do termo entálpico referente às interações acima mencionadas para a proteína com as moléculas de água em solução, esta diferença leva em consideração as penalidades entálpicas e entrópicas para a formação da *cavidade de solvatação* de forma a permitir a acomodação da molécula de soluto no meio, e também, para o caso de uma macromolécula como a proteína, as variações destes termos decorrentes de mudanças nas interações intramoleculares devido ao estabelecimento de novas interações com o solvente.

a energia de hidratação da molécula e a energia total da molécula evidenciada no *inlet* de Fig. 4.12-A). Também para o confôrmero  $\text{RXR}_{RMN}^{\text{XX}}$ , apresentando um maior grau de rigidez inicial o aumento relativamente lento da energia potencial da molécula decorrente da termalização (conforme observado na Fig. 4.11-B) apresenta uma forte correlação inversa com a energia de hidratação da molécula (da mesma ordem que a anticorrelação evidenciada para o confôrmero  $\text{RXR}_{RMN}^{\text{I}}$  - Figs. 4.12-A e B), que diminui ao longo de toda a simulação até valores próximos aos iniciais para as configurações  $\text{RXR}_{RMN}^{\text{I}}$  e  $\text{RXR}_{\text{crys}}$ . Desta forma, o aparente maior empacotamento inicial desta configuração proporciona com que a energia térmica lentamente absorvida durante os 1050 ps de sua trajetória seja "canalizada" em movimentos de menor amplitude (conforme pode ser constatado pela menor progressão do valor de rmsd em relação à estrutura inicial), que são associados ao rompimento de interações relativamente estáveis (conforme indicado pelo lento aumento das energias potenciais ao longo de toda a simulação) e, em sua maioria, de natureza hidrofóbica (condizente com a notória correlação entre o rompimento destas interações e o aumento da hidratação total da molécula)<sup>37</sup>

Para o confôrmero  $\text{RXR}_{\text{crys}}$ , que apresenta o T-box na configuração desnaturada e (inicialmente) extendida presente nas estruturas dos complexos cristalográficos, a energia de hidratação nos primeiros 2500 ps apresenta um relativo decréscimo em seu valor médio, de -4120.3 a -4323.6 kcal·mol<sup>-1</sup>, em concordância com sua configuração inicial relativamente mais exposta ao solvente (Figs. 4.10-C e 4.12-C). No entanto, decorrido este intervalo inicial, a mesma energia de hidratação para o confôrmero aumenta substancialmente, em uma amplitude condizente com a acentuada mobilidade demonstrada para a molécula nesta conformação através dos gráficos de variação do rmsd em relação à estrutura inicial e de energia potencial da molécula

<sup>37</sup> As duas outras componentes da energia potencial da molécula - as energias conformacionais e de interação com os íons em solução - variam muito pouco e permanecem oscilando com baixa amplitude em torno de valores médios ao longo de todas as três simulações (não mostrado), não sendo portanto consideradas nestas análises. Isto é esperado, visto que as trajetórias tratam-se de simulações de equilíbrio NPT para configurações oriundas de estruturas experimentais previamente solvatadas e neutralizadas, seguido da minimização do sistema (não sendo portanto esperadas variações adicionais significativas nas energias conformacionais) e relaxamento das águas e íons à mesma temperatura por dinâmica de restrição prévia à dinâmica produtiva. Além disto, foram adicionados apenas os sete íons cloreto necessários para se neutralizar a carga do sistema (devido à carga líquida + 7 do DBD), proporcionando uma força iônica relativamente baixa ( $\sim 17$  mM). Isto e o fato do DBD não apresentar regiões com densidades de carga suficientes para proporcionarem significativa imobilização e/ou *clusterização* locais dos íons  $\text{Cl}^-$ , permanecendo estes dispersos em solução, proporciona a variação igualmente pouco significativa da energia global de interação da proteína com os íons em solução durante a dinâmica produtiva.



**Figura 4.12** – Progressões das energias de hidratação para as trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais.

nas Figs. 4.11-E e F. Desta forma, mesmo na conformação substancialmente mais exposta e flexível correspondente ao  $\text{RXR}_{\text{crys}}$ , o monômero DBD demonstra ainda uma considerável fugacidade em relação às configurações de maior hidratação, confirmando o caráter hidrofóbico predominante em suas principais interações. Contudo, para este confôrmero não é observada a intensa anticorrelação entre a energia de hidratação da molécula e a energia total previamente constatada para as estruturas em solução provenientes dos dados de RMN (compare a correlação linear no *inlet* da Fig. 4.12-C com as correlações análogas nos *inlets* das Figs. 4.12-A e B). Desta forma, o rompimento dos contatos originais devido à desnaturação do T-box e seu consequente desempacotamento do cerne hidrofóbico do DBD é ainda permissivo para a liberação de moléculas de água para o meio<sup>38</sup> mas não é igualmente permissivo para a formação de interações estáveis que confirmam uma compensação entálpica para a liberação destas moléculas de água da molécula (de forma análoga à que é vislumbrada com o empacotamento da hélice C nas conformações comumente adotadas em solução), confirmando a participação crucial das interações hidrofóbicas para a não manutenção desta configuração na ausência dos contatos mediados pela formação dos complexos ao HRE (a despeito de sua aparente maior entropia conformacional).

O comportamento de mobilidade, troca de energia térmica com o meio e contribuição das interações hidrofóbicas para o empacotamento e principais movimentos do monômero hRXR $\alpha$ -DBD em solução a partir das três respectivas configurações iniciais abordadas neste estudo pôde ser melhor compreendido a partir da análise das principais componentes de movimento (*Principal Component Analysis - PCA*) observadas em cada simulação (Fig. 4.13). Embora as diferenças nos tempos de simulação para os quais o método de PCA foi aplicado para cada sistema (tempo de PCA para a conformação  $\text{RXR}_{\text{RMN}}^{\text{XX}} < \text{RXR}_{\text{RMN}}^{\text{I}} < \text{RXR}_{\text{crys}}$ )<sup>39</sup> impli-

<sup>38</sup> a elevada mobilidade do T-box nesta conformação permite uma certa *contração* deste segmento, bem como sua interação com a face “inferior” da hélice A, sendo este movimento aparentemente favorecido pela liberação de moléculas de água destas interfaces para o meio (vide Figs. 4.13-C e 4.17).

<sup>39</sup> O método de PCA, a partir da diagonalização das matrizes de covariância 3N X 3N para os átomos das respectivas cadeias principais (C $\alpha$ , C, O e N) foi empregado para as simulações das duas estruturas de RMN a partir do intervalo de simulação de 600 ps e para a configuração oriunda da estrutura cristalográfica a partir de 760 ps. Estes tempos referem-se aos intervalos a partir dos quais o perfil de rmsd para cada respectiva simulação sugere um comportamento de mobilidade assintótica e/ou um padrão de oscilação sobre uma média conformacional, conforme verificado pelos gráficos representados nas Figs. 4.11-A, C e E, bem como pela análise dos respectivos *plots* para a variação do rmsd em relação à estrutura média (não mostrado).

que em diferenças nas respectivas amplitudes e freqüências dos movimentos descritos pelos autovetores passíveis de serem recuperados, as diferenças de tempo de *PCA* para cada sistema foram, *a priori*, condizentes com as respectivas mobilidades sugeridas para cada conformação pela análise dos perfis de rmsd e energia total Vs tempo (amplitudes dos movimentos para a conformação  $\text{RXR}_{RMN}^{\text{XX}} < \text{RXR}_{RMN}^{\text{I}} < \text{RXR}_{\text{crys}}$  e freqüências na ordem inversa). Em conjunto, conforme se verá na discussão subsequente, as componentes tomadas para as três conformações permitiram inferir quanto à liberdade e correlação de movimentos intramoleculares para o monômero nas conformações comumente adotadas em solução e na conformação permissiva para a interação com o DNA<sup>40</sup>.

Para a conformação  $\text{RXR}_{RMN}^{\text{I}}$ , o autovetor correspondente à principal componente demonstra que, apesar de seu relativo empacotamento inicial sobre o cerne hidrofóbico formado pela hélice A e pelo motivo  $\beta$ -*hairpin* do primeiro dedo de zinco, a hélice C ainda apresenta uma considerável liberdade de movimentos locais de extensão, rotação e translação<sup>41</sup> (Fig. 4.13-A-*Componente I*). De fato, a principal componente para a trajetória referente a esta configuração no equilíbrio (aparente na simulação) é representada pelos movimentos da hélice C e do motivo  $\beta$ -*hairpin* do primeiro dedo de zinco, variando o empacotamento destes dois motivos entre si e com a hélice A, de forma a influenciar a conformação e dinâmica interna de todo o domínio DBD (Fig. 4.13-A-*Componente I*). Na conformação correspondente ao "máximo" desta componente o motivo  $\beta$ -*hairpin* se encontra colapsado sobre as hélices A e C, encontrando-se assim os tês motivos em orientações e/ou conformações não permissivas para sua correta associação

<sup>40</sup> É, também, importante frisar que a PCA para a trajetória referente ao confôrmero  $\text{RXR}_{RMN}^{\text{I}}$  trata-se de uma análise ainda fora do equilíbrio (conforme indicado pelo perfil do gráfico de energia potencial na Fig. 4.11-B). Contudo, as interpretações meramente referentes aos acoplamentos mecânicos dos movimentos intramoleculares, conforme explorado no texto abaixo, e o fato da estrutura original oriunda dos dados experimentais de RMN sugerir ser este um microestado acessível e com relativa densidade populacional no equilíbrio, tornam esta uma abordagem válida.

<sup>41</sup> A disposição em espiral da estrutura da  $\alpha$ -hélice e de suas interações de hidrogênio da cadeia principal, bem como seu formato longilíneo, promovem com que uma porção significativa dos modos normais em solução para segmentos com este tipo de estrutura secundária se assemelhem aos movimentos de uma mola. Desta forma, os acoplamentos entre movimentos vibracionais, rotacionais, translacionais e (dependendo da extensão da hélice) de movimentos de torção do eixo principal, são modos normais comumente observados. Estes acoplamentos ocorrem sobretudo quando a hélice se posiciona em regiões terminais, sendo conectada a um cerne globular (de um domínio ou da proteína como um todo) através de um segmento com relativa mobilidade (como é o caso da hélice C do DBD em solução). Em particular, o empacotamento parcial destes segmentos com considerável atividade vibracional a regiões hidrofóbicas do cerne do domínio permite com que seus modos normais se acoplem aos movimentos coletivos da estrutura como um todo, tendendo, no caso de domínios globulares relativamente pequenos, a influenciar substancialmente nos movimentos globais.

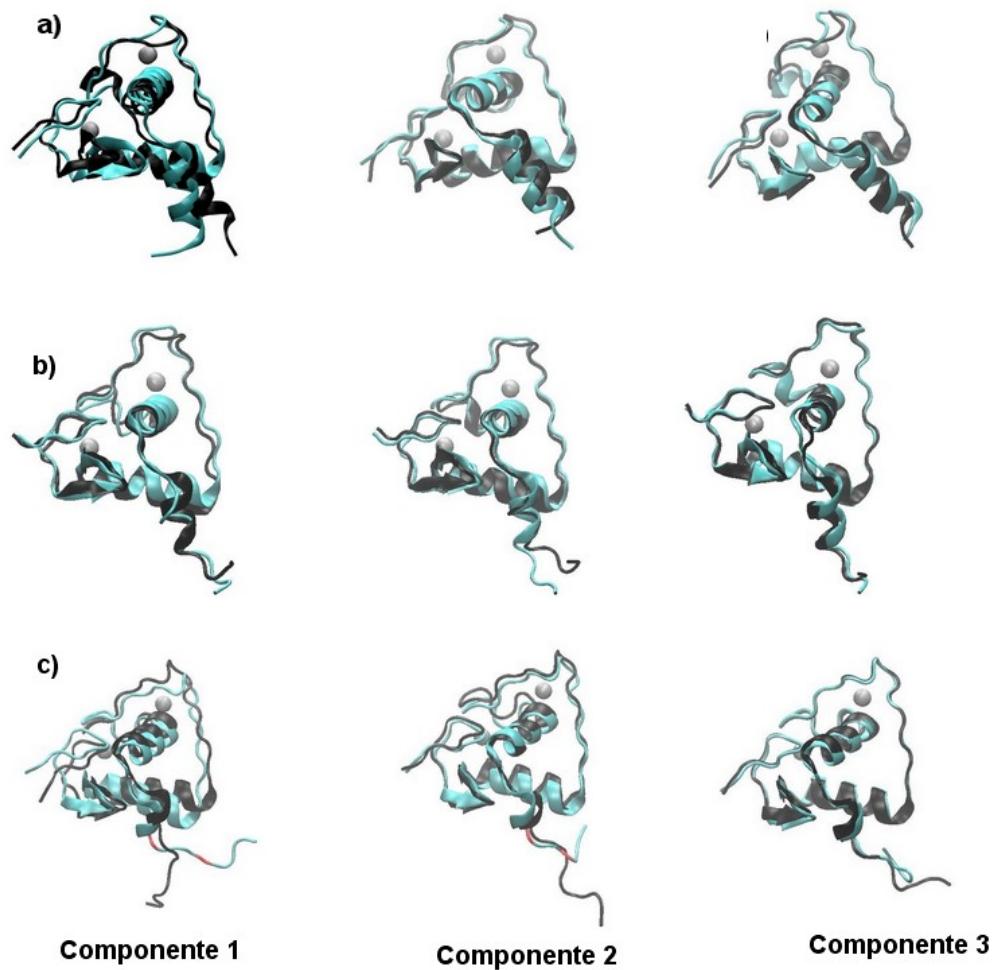
ao HRE. Nota-se que o empacotamento destes três motivos (hélice A,  $\beta$ -hairpin e hélice C) é acoplado a um movimento de rotação destróbira e uma pequena contração da hélice C<sup>42</sup>, que por sua vez, mediante o reposicionamento e estiramento do *coil* formado pelos resíduos G199-M201, entre esta e a hélice B, se correlaciona a um pequeno movimento de "elevação" da última (de forma a se afastarem as cadeias principais das hélices B e A). Além disto, os movimentos de aproximação dos motivos  $\beta$ -hairpin e hélice C e destes motivos à hélice A são correlacionados a pequenos movimentos de estiramento e rotação da última, que por sua vez, corroboram para o movimento previamente mencionado de elevação da hélice B (empacotada sobre a hélice A). Por sua vez, os movimentos das hélices A e B em conjunto corroboram para um pequeno movimento de distorção do D-box (que sofre uma rotação *destróbira* de aproximadamente 15° do *loop*), chegando o efeito desta rotação até o segmento formado pelos resíduos D180-C187<sup>43</sup>, que por sua vez conecta o segundo dedo de zinco à hélice B (Fig. 4.13-A). A rotação do segundo dedo de zinco, por sua vez, se correlaciona a uma mudança na estrutura secundária do segmento D180-C187, que passa de uma conformação próxima à de uma hélice distorcida para uma estrutura mais desordenada<sup>44</sup>. Esta distorção de estrutura secundária, por sua vez, é permissiva para a inclinação previamente mencionada para a hélice B (C-terminal a este segmento) nesta componente principal, fechando-se assim o ciclo de acoplamentos na dinâmica interna do domínio. As duas próximas componentes principais (com alto-valores relativos significativamente menores, indicando a prevalência considerável da componente 1 na descrição dos movimentos globais da molécula durante a trajetória) descrevem pequenos movimentos de rotação e extensão da hélice do T-box (sem a translação significativa da mesma região) sobre o espaço entre o motivo  $\beta$ -hairpin e a hélice A (Fig. 4.13-A *componentes 2 e 3*). Nota-se que mesmo estes movimentos em pequena escala parecem ser acoplados a movimentos

<sup>42</sup> estes movimentos de rotação destróbira e contração são condizentes com o padrão de interações de hidrogênio ( $3 + i \longrightarrow i$ ) para a cadeia principal, próprio de uma hélice  $3_{10}$  concordando com as estruturas resolvidas por RMN), em detrimento do padrão do tipo ( $4 + i \longrightarrow i$ ) (mais expandido) de uma  $\alpha$ -hélice comum.

<sup>43</sup> Os resíduos K181-N185 neste mesmo segmento interagem com o *backbone* do DNA e com o parceiro na posição 3' em diversos complexos de homo e heterodímeros, além de terem sua conformação e dinâmica fortemente influenciadas pela associação de um único monômero ao meio-sítio, conforme verificado por estudos de RMN na literatura e previamente mencionado

<sup>44</sup> Esta mudança é exatamente condizente com a mudança sugerida para este segmento pelos estudos de RMN durante a passagem da conformação do DBD associada ao DNA para as conformações comumente adotadas em solução

vibracionais no D-box.



**Figura 4.13** – Análise de componentes principais para as trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais

Para a conformação RXR $_{RMN}^{XX}$ , o significativo maior empacotamento da hélice C sobre o cerne hidrofóbico formado pela hélice A e pelo motivo  $\beta$ -hairpin promove com que os autovalores relacionados às principais componentes para esta configuração sejam associados a movimentos vibracionais de maior frequência para a mesma hélice C (com menor grau de liberdade translacional e rotacional), semelhantes à vibração de uma mola fixa em uma de suas extremidades (Fig. 4.13-B)<sup>45</sup>. A considerável maior rigidez desta configuração relacionada aos confôrmeros RXR $_{RMN}^I$  e RXR $_{crys}$  é novamente evidente nas diferenças de representatividade para os autovalores das diferentes componentes (conforme se nota pela fração destes em

<sup>45</sup> No trabalho de Holmbeck et al (a) , decréscimos nos deslocamentos químicos do  $^{13}\text{C}_{\alpha}$  a partir da G206 atestam quanto à flexibilidade da porção C-terminal da hélice C em solução, de uma forma condizente com as componentes observadas para os dois confôrmeros RXR $_{RMN}$  (sobretudo para o RXR $_{RMN}^{XX}$ ) no presente estudo.

relação à soma total dos autovalores para os movimentos da molécula, assinalados para cada componente na Fig. 4.13). Enquanto que para os confôrmeros de menor empacotamento os movimentos relacionados às respectivas primeiras componentes apresentam uma contribuição cerca de quatro vezes maior que os da segunda para o movimento total da molécula, denotando uma considerável maior contribuição dos movimentos mais abrangentes (envolvendo mudanças conformacionais globais na estrutura); para a conformação  $\text{RXR}_{RMN}^{XX}$  a maior representatividade é atribuída quase que igualmente para a primeira e a segunda componentes (com um acréscimo de apenas 7 % na representatividade total para a componente 1), indicando uma maior distribuição dos movimentos da molécula em modos normais de maior freqüência e baixa amplitude. Em particular, o movimento vibracional referente à componente 1 se relaciona a uma maior extensão da hélice C, Proporcionando uma considerável distorção de sua estrutura secundária (Fig. 4.13-B-*Componente 1*). Este movimento vibracional, com a consequente perturbação de um número substancial de interações de hidrogênio da cadeia principal e de contatos das cadeias laterais envolvidas na manutenção da hélice e de seu empacotamento, responde por uma fração considerável da energia potencial ininterruptamente absorvida durante os 1050 ps desta simulação (Fig. 4.11-B), bem como por sua evidente anti-correlação com a energia de hidratação (a expansão da hélice e seu consequente desempacotamento sobre o cerne hidrofóbico aumenta a exposição destas regiões ao solvente). Nota-se também que estes movimentos vibracionais da hélice C são novamente acoplados à vibração da molécula como um todo (devido ao mesmo acentuado empacotamento hidrofóbico desta hélice ao cerne do domínio), sendo particularmente interessante a correlação dos movimentos de expansão da hélice C e do D-box na face oposta do DBD que, como já extensamente mencionado ao longo desta seção, é envolvido em contatos com o monômero 3' nos complexos cristalográficos.

A comparação das componentes principais para o confôrmero  $\text{RXR}_{RMN}^{XX}$  com as da conformação  $\text{RXR}_{crys}$  (Fig. 4.13-B e C), sugere que a configuração desnaturada do T-box nos complexos cristalográficos com o DNA pode se originar a partir de um movimento vibracional de expansão decorrente dos modos normais naturalmente exibidos pelo domínio em solução (semelhante ao movimento característico para a componente 1 do confôrmero  $\text{RXR}_{RMN}^{XX}$ ), sendo esta ex-

pansão superestimulada pelo ambiente químico do complexo até a completa desnaturação da hélice (Figs. 4.13-B e C -*Componentes 1*). Nos complexos cristalográficos com o DR1, esta configuração desnaturada é estabilizada pelos contatos da mesma com o sulco menor, sendo de particular relevância as interações dos resíduos Q206 e R209 (cujas cadeias laterais se encontram voltadas para a face oposta na estrutura em hélice (vide Fig. 4.16) e cujas cadeias principais são assinaladas nas ilustrações para as diferentes componentes na Fig. 4.13-C) e por contatos com o parceiro 5' ()<sup>46</sup>. A desnaturação do T-box, conforme indicado pelas componentes principais na Fig. 4.13-C, proporciona um considerável desacoplamento entre os movimentos vibracionais e translacionais desta região e a dinâmica interna do cerne do domínio. De fato, na projeção do autovetor correspondente à componente principal para a simulação do monômero RXR<sub>crys</sub> (Fig. 4.13-C-*Componete 1*), todo o cerne do DBD rotaciona em conjunto, variando sua posição global (aproximando o seu centro de massa) em relação ao T-box (devido a um movimento de translação deste *coil* C-terminal que tende a se aproximar da face “inferior” da hélice A) não sendo observadas mudanças significativas nas distâncias internas para as diferentes regiões do cerne globular do domínio em circunstância deste movimento. A componente 2, com significativa menor amplitude, é representada por movimentos translacionais do T-box e vibrações pouco pronunciadas nas duas outras extremidades estruturais do domínio - os *loops* nos respectivos “topos” dos dois dedos de zinco - enquanto que a componente 3 (com autovalor uma ordem de grandeza menor) é referente à oscilação do T-box sem movimentos significativos sob qualquer grau de liberdade interno (translacional, rotacional ou vibracional) para o cerne globular do DBD. O perfil das componentes principais na Fig. 4.13 indica que, com o desacoplamento da dinâmica interna do DBD dos movimentos oscilatórios do T-box (uma aparente consequência da desnaturação deste último) o cerne globular do domínio adota uma configuração relativamente mais rígida (sem a presença significativa de movimentos internos em suas componentes). Este padrão é confirmado a partir da análise dos

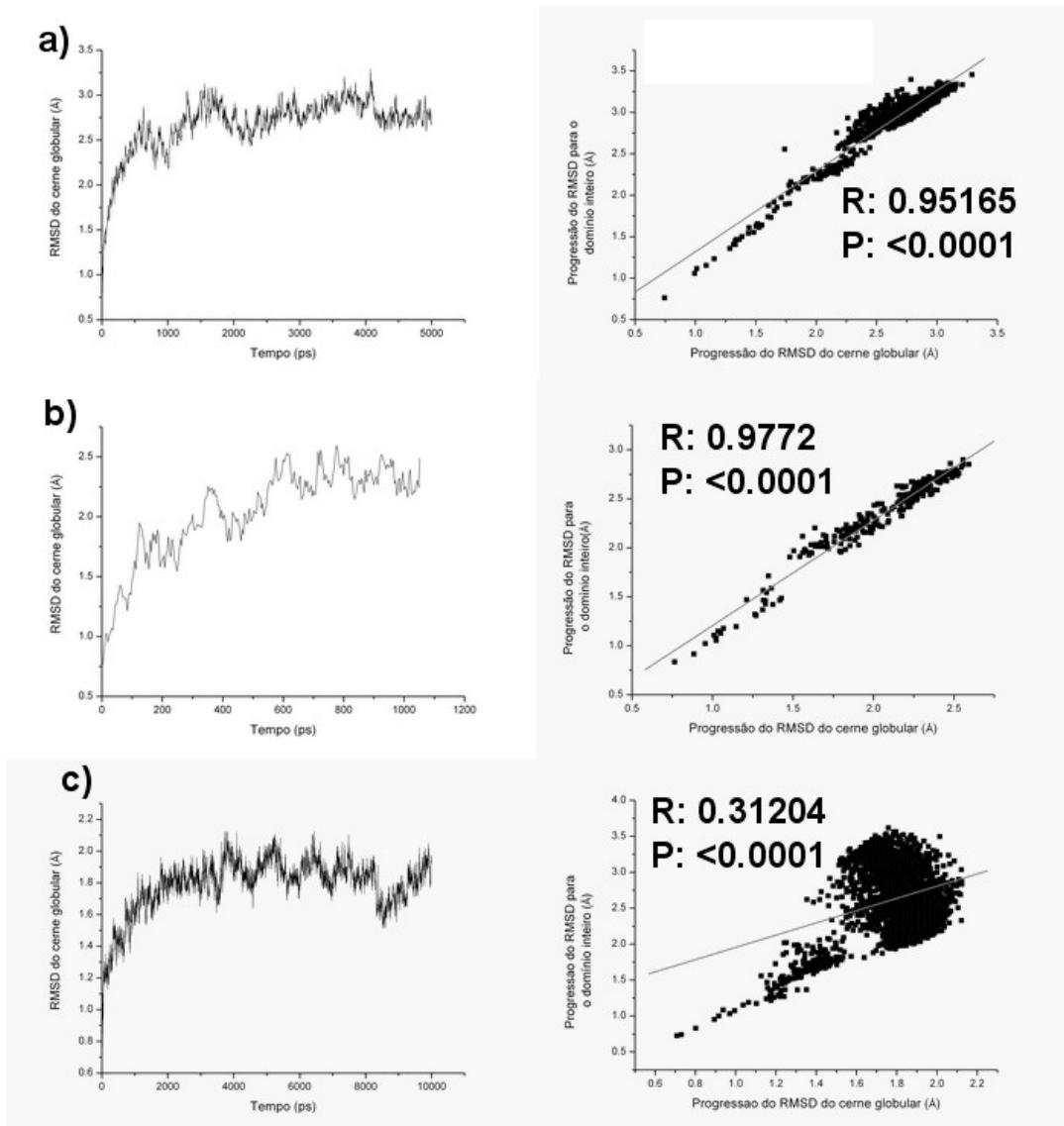
<sup>46</sup> Os trabalhos de RMN previamente citados e as duas estruturas cristalográficas contendo os DBDs de hRXR $\alpha$  na posição 5' (sem contato do T-box com o parceiro de dimerização) e contendo esta região também visivelmente desnaturada e interagindo com o sulco menor (estrutura do homodímero de hRXR $\alpha$ -DBD em DR1 ), ou não sendo resolvível devido à baixa densidade eletrônica (indicando alta mobilidade) (estrutura do heterodímero hTR $\beta$ /hRXR $\alpha$ -DBDs em DR4 ), sugerem que a interação do DBD com o meio-sítio é suficiente para induzir a uma configuração desnaturada do T-box, sendo, portanto, os contatos com o sulco menor aparentemente determinantes para a estabilização desta conformação (Vide texto abaixo e seções ).

respectivos perfis de progressão do rmsd em relação às configurações iniciais, considerando-se o DBD apenas até a região prévia ao segmento C-terminal que sofre desnaturação na formação do complexo (*i.e.*, considerando-se apenas o cerne globular do DBD - resíduos K132-A204) (Fig. 4.14)<sup>47</sup>. Nota-se que, enquanto os perfis de rmsd referentes à estrutura da molécula como um todo (Figs. 4.11A, C e E) indicam uma acentuada maior mobilidade para o domínio na configuração RXR<sub>crys</sub> (devido à oscilação com alta amplitude do T-box, mais livre conformacionalmente nesta configuração), o cerne globular do DBD nesta mesma configuração apresenta um perfil significativamente mais estável para o rmsd ao longo da simulação que os respectivos perfis apresentados para os confôrmeros oriundos das estruturas de RMN, representativas do *ensemble* conformacional comumente adotado em solução (Fig. 4.14)(*i.e.*, o rmsd para o cerne globular do confôrmero RXR<sub>crys</sub> se estabiliza sob um menor valor médio em relação à conformação inicial e oscila com menor amplitude sobre este valor médio). Além disto, os perfis de progressão do rmsd para o cerne globular das estruturas de RMN são significativamente similares (tanto nos valores alcançados quanto nos padrões de oscilação) aos perfis para a estrutura como um todo - inclusive com respectivas correlações lineares se aproximando da função identidade (vide Fig. 4.14-*inlets*) - enquanto que os mesmos perfis para o confôrmero RXR<sub>crys</sub> claramente não o são. Este padrão vem novamente confirmar que, enquanto nas conformações adotadas em solução o T-box (na configuração em hélice 3<sub>10</sub>) e o cerne do DBD apresentam padrões internos de oscilação concertados, com um significativo acoplamento dos modos de rotação, extensão e (em menor escala) translação da hélice C-terminal com a dinâmica interna do domínio como um todo; para a conformação de interação ao DNA, a desnaturação local do T-box desacopla consideravelmente as dinâmicas destas duas regiões, que passam a se comportar como dois corpos conectados um ao outro, porém com movimentos internos independentes. Desta forma, os movimentos de alta amplitude para o RXR<sub>crys</sub> descritos nos graficos de rmsd e energia potencial na Fig. 4.11-E e F são decorrentes, principalmente, de movimentos dos

---

<sup>47</sup> Nos cálculos de progressão dos rmsds, o alinhamento prévio em relação à estrutura de referência (configuração inicial) foi sempre realizado considerando-se apenas as regiões para cujos átomos pesados as respectivas progressões foram calculadas; *i.e.*, no cálculo da progressão do rmsd de todos os átomos pesados da molécula, o alinhamento foi realizado considerando-se igualmente todos estes átomos pesados, enquanto que para a progressão referente apenas aos átomos dos resíduos K132-A204, somente estes mesmos átomos foram previamente alinhados.

respectivos centros de massa do T-box e do cerne globular do domínio DBD (este, por sua vez, com uma menor mobilidade interna nesta conformação), um em relação ao outro.



**Figura 4.14** – Distinções para a mobilidade do cerne globular do domínio DBD do monômero de hRXR $\alpha$ -DBD em solução a partir de diferentes configurações iniciais.

A natureza das diferenças no acoplamento dinâmico-conformacional entre o empacotamento da hélice C e o restante do domínio nas conformações comumente adotadas em solução e na conformação de associação ao DNA, conforme evidente pelas PCAs para as três trajetórias para diferentes confôrmeros do hRXR $\alpha$ -DBD aqui analisadas, pode ser melhor compreendida com uma análise mais refinada das estruturas correspondentes a *instantâneos* destas trajetórias com diferentes valores de projeção sobre as componentes principais (Figs. 4.15, 4.16 e 4.17). Na trajetória para o confôrmero RXR $_{RMN}^I$ , as conformações do domínio com maiores valores

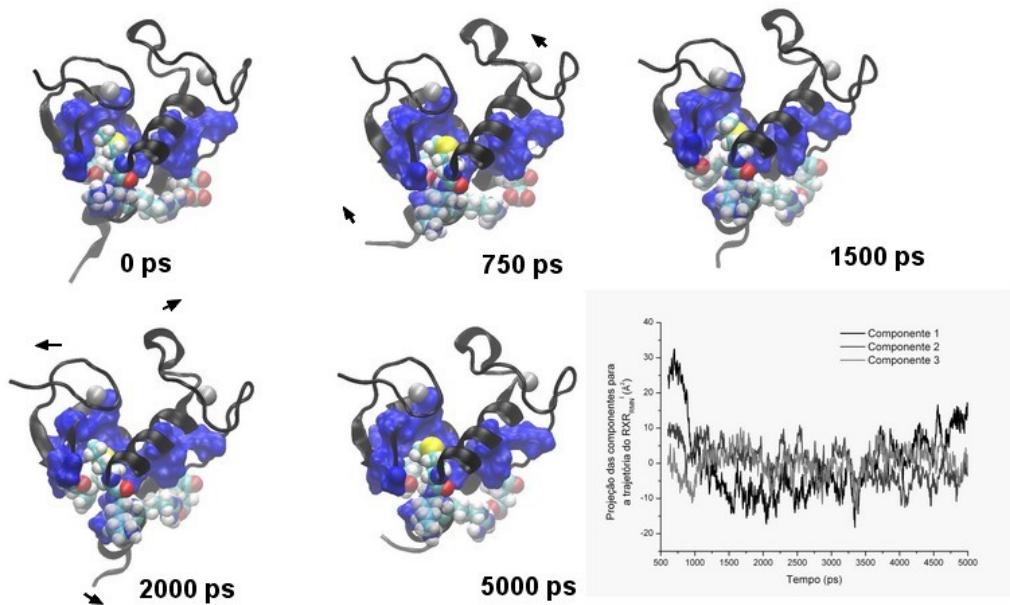
de projeção positiva para a componente 1 demonstram que com a rotação da hélice C no sentido destrógiro, acoplada também a um movimento de translação e de contração da mesma, o resíduo M200 aumenta o seu grau de empacotamento sobre a cavidade hidrofóbica formada por resíduos do primeiro dedo de zinco (sobretudo resíduos do motivo  $\beta$ -*hairpin*, como a V149) e da hélice A. Também ocorre, com este movimento, uma aproximação dos átomos N da cadeia principal da K201, na hélice C, e O da cadeia principal da Y147, na extremidade da  $\beta$ -*hairpin*, bem como das cadeias laterais desta mesma Y147 e da A204 (também na hélice C). Esta interação "sequestra" a cadeia lateral da Y147, promovendo também seu maior empacotamento hidrofóbico com resíduos da hélice A (em particular, com a F159), "fechando" desta forma a cavidade entre os três motivos estruturais ( $\beta$ -*hairpin*, hélice A e hélice C) e acentuando o empacotamento da M200 entre estes três motivos. Esta "captura" da Y147 parece ser de fundamental importância na penalidade energética para a interação com o DNA (na ausência de um parceiro de dimerização específico), visto que enquanto nas estruturas resolvidas por RMN para o monômero em solução este resíduo se apresenta associado ao cerne hidrofóbico do domínio (de forma análoga à observada nos *instantâneos* com maior projeção positiva para a componente 1 nesta simulação), nas diversas estruturas cristalográficas tanto do hRXR quanto de DBDs de outros NRs, o mesmo resíduo (ou seu homólogo) se encontra "liberado" deste cerne e interagindo diretamente, ou mediado por contatos com moléculas intermediárias de água, com o *backbone* do DNA (adotando uma orientação similar à vislumbrada nos *instantâneos* com maior projeção negativa para a mesma componente 1). De fato, os contatos hidrofóbicos envolvidos nesta rede de interações respondem por uma fração considerável da correlação entre a estabilização do sistema no equilíbrio e a liberação de moléculas de água da interface de interação com a proteína para o seio da solução (Figs. 4.11-B e 4.12-C). O empacotamento da hélice C entre a  $\beta$ -*hairpin* e a porção N-terminal da hélice A proporciona uma superfície relativamente mais estável para a interação entre esta face N-terminal do domínio com a porção C-terminal do segundo dedo de zinco e o segmento posterior, D180-C187, entre este segundo dedo de zinco e a hélice B (Fig. 4.15-A). Em contrapartida, na face oposta da hélice C, a extremidade da cadeia lateral da R202 (cuja porção alifática mediana também faz contatos de Van der Walls com os grupos volumosos da região C-terminal da hélice A) forma uma ponte salina com o resíduo D166, o

primeiro resíduo do *lever arm*, C-terminal à hélice A. Este contato (que também é constatado nos experimentos de RMN) aumenta o acoplamento entre os movimentos do *lever arm* e os da hélice C na componente 1, bem como o acoplamento deste mesmo *loop* com toda a face N-terminal do domínio nas conformações em que a hélice C se encontra mais empacotada. É interessante notar que o duplo mutante K201T,R200A (com modificações nos dois resíduos básicos, um em cada face da hélice C, envolvidos em contatos que colaboram para o empacotamento simultâneo da hélice C ao motivo  $\beta$ -*hairpin* do primeiro dedo de zinco e à hélice A) se liga ao DNA como dímero, mas com reduzida afinidade, indicando que as interações envolvendo estes dois resíduos têm participação fundamental na interação ao HRE . Visto que em nenhuma das estruturas cristalográficas são observados contatos entre estes resíduos e o DNA ou ao DBD do monômero 5', ambos se encontrando em orientações médias similares às resolvidas em solução nos experimentos de RMN e condizentes com os movimentos principais observados nas simulações deste estudo, pode-se inferir que o papel destes dois resíduos no empacotamento e consequente acoplamento dinâmico entre a porção N-terminal da hélice C (que não é desnaturada na interação ao DNA) e o cerne globular do DBD representa um fator funcional na formação dos complexos RXR-DBD:DNA <sup>48</sup>.

Conforme previamente analisado através da PCA para a trajetória do confôrmero RXR<sub>RMN</sub><sup>XX</sup> (Fig. 4.13-B), o acentuado empacotamento da hélice C nesta configuração reduz drasticamente sua liberdade translacional, o que faz com que a energia cinética proveniente da termalização seja utilizada por este segmento quase que exclusivamente para movimentos de extensão acoplados à sua rotação<sup>49</sup>. Comparando-se, na simulação referente a esta conformação, a estrutura a 0 ps com o *instantâneo* a 600 ps, que por sua vez apresenta o máximo de projeção positiva sobre a componente 1 (Fig. 4.16), observa-se que a extensão do T-box provoca a rotação dos resíduos Q206 e R209 na direção de suas respectivas posições na conformação do RXR<sub>crys</sub>. Contudo, a conformação em hélice não permite o posicionamento destes resíduos na configuração final do

<sup>48</sup> a contribuição destes (e de outros) resíduos na estabilização do complexo ao DNA será melhor discutida nas subseções referentes às simulações destes complexos .

<sup>49</sup> Em concordância com a proximidade dos autovalores, as projeções da trajetória sobre os autovetores relativos às três componentes demonstram a representatividade aproximadamente igual para os diferentes modos vibracionais de alta freqüência recuperados pela PCA na descrição dos movimentos do confôrmero RXR<sub>RMN</sub><sup>XX</sup> durante os 1050 ps de simulação

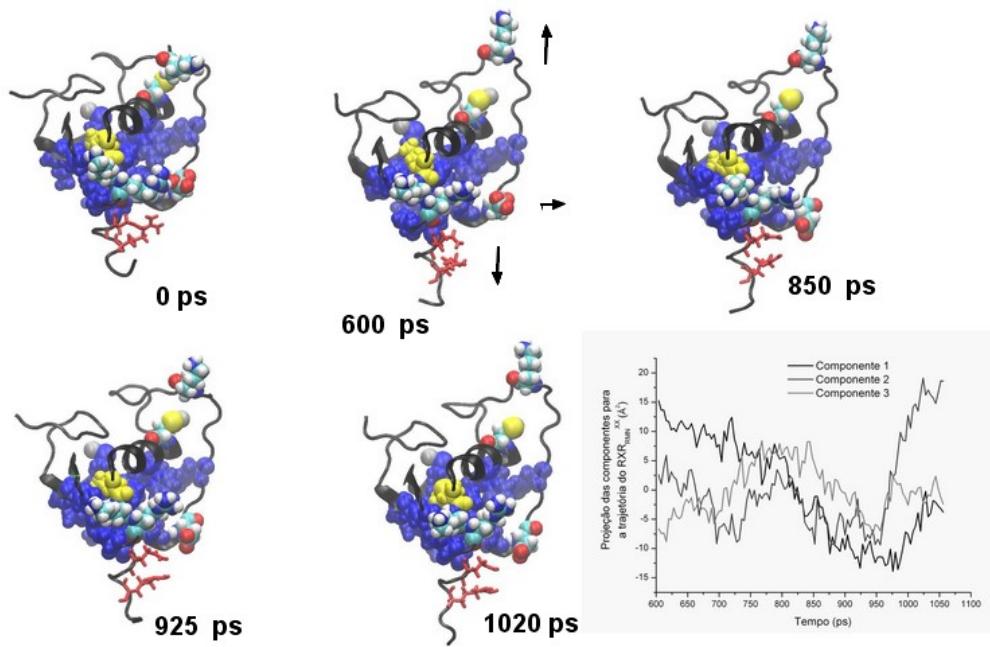


**Figura 4.15** – Projeções dos autovetores recuperados na análise de componentes principais para a trajetórias do monômero de hRXR $\alpha$ -DBD em solução a partir da configuração inicial RXR<sub>RMN</sub><sup>I</sup>.

complexo. Além disto, nas conformações mais contraídas da hélice C as cadeias laterais destes dois resíduos fazem contatos de vdw, tanto entre si, quanto com resíduos da porção N-terminal desta mesma hélice e com resíduos da hélice A, de forma a que o total afastamento destes grupos volumosos confere um acréscimo na penalidade energética para a rotação deste segmento. Desta forma, não são simultaneamente permissíveis o correto posicionamento para a interação do T-box com o sulco menor do DNA, a configuração em hélice para o T-box como um todo e a manutenção do empacotamento hidrofóbico da porção N-terminal da hélice C com o cerne do DBD (altamente favorável, conforme a comparação da progressão da energia total com a energia de hidratação da molécula nas diferentes simulações). No que concerne ao acoplamento vibracional desta região com o restante do domínio, nota-se que o movimento de extensão da hélice C no confôrmero RXR<sub>RMN</sub><sup>XX</sup> provoca rotações nos ângulos  $\phi$  e  $\psi$  da cadeia principal, que terminam por afastar as cadeias laterais dos resíduos K201 e R202 de seus respectivos contatos com o motivo  $\beta$ -hairpin e com a hélice A. Em particular, a ponte salina entre as cadeias laterais da R202 e do D166 é rompida neste movimento, diminuindo a conectividade entre a hélice C e o N-terminal do lever arm. É também notório, nos instantâneos representados na Fig. 4.16, que com a extensão da hélice C a cadeia lateral da M200 (na figura, representada

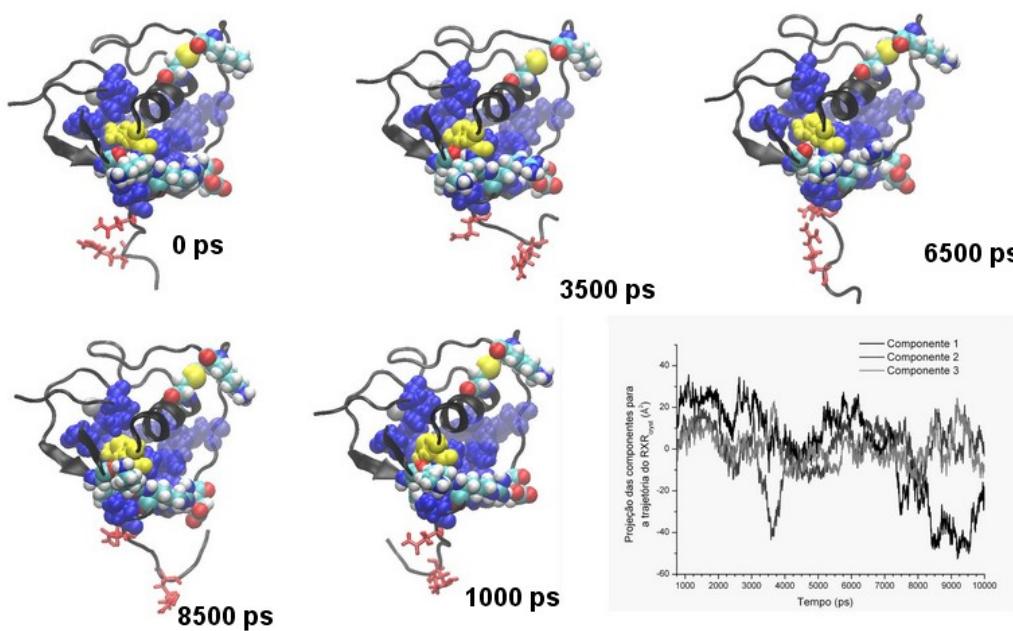
em amarelo) - altamente empacotada sobre o cerne hidrofóbico formado principalmente por resíduos volumosos do motivo  $\beta$ -*hairpin*, da hélice A e pela cadeia lateral da A195 na hélice B - se inclina sobre este mesmo cerne hidrofóbico, alterando uma rede de empacotamentos em direção ao C-terminal da hélice A, envolvendo principalmente resíduos aromáticos: os resíduos F158 e 159 (na hélice A), que por sua vez contribuem para o empacotamento da Y192 (localizada na hélice B e voltada para a hélice A) e chegando este efeito até o *lever arm*, através da Y169 (seta na Fig. 4.16-A-600 ps) que faz interações de hidrogênio com o grupo hidroxila da cadeia lateral da T162 (cuja porção alifática também faz contatos de vdw com a Y192) bem como contatos de vdw ocasionais com a F158 e Y192. Este "efeito dominó", juntamente com movimentos relativamente pequenos de extensão e rotação da hélice A (em contato com a hélice C (sob expansão) e sem a restrição da ponte salina entre as cadeias laterais dos resíduos R202 e D166, que é perdida com a extensão e rotação da mesma hélice C), acopla a extensão do T-box à do D-box, culminando com a perda dos contatos de vdw entre as cadeias laterais da K175 no "topo" do *loop* do segundo dedo de zinco e da C190 (a quarta cisteína na direção N→C-terminal do segundo dedo de zinco), localizada na hélice B. O acoplamento entre os ciclos de extensão e contração da hélice C com o afastamento e aproximação da K175 à C190 é notório nos *instantâneos* com diferentes projeções sobre a componente 1 para o confôrmero  $\text{RXR}_{RMN}^{XX}$  na Fig. 4.16. Conforme será visto na subseção , os estudos de MD para o complexo do homodímero de hRXR $\alpha$ -DBD em DR1 apresentados neste trabalho de doutoramento sugerem que o estabelecimento da dobra na cadeia principal no "topo" do D-box decorrente da interação entre as cadeias laterais destes dois resíduos é importante para a acomodação estereoquímica do T-box do monômero 3', proporcionando um maior empacotamento e comunicação da dinâmica conformacional dos dois sítios, além de poder ter implicações quanto às restrições impostas pela dimerização dos DBDs para a dimerização dos respectivos LBDs.

Já para a trajetória do confôrmero oriundo do complexo cristalográfico ( $\text{RXR}_{crys}$ ), os *instantâneos* com diferentes valores de projeção sobre as componentes principais (Fig. 4.17) mostram que, com a desnaturação da porção C-terminal da hélice C, os movimentos oscilatórios do T-box deixam de ser acoplados à rotação da porção N-terminal desta mesma hélice (onde ocor-



**Figura 4.16** – Projeções das componentes principais para a trajetória do monômero DBD na conformação RXR<sub>RMN</sub><sup>XX</sup>.

rem os maiores empacotamentos sobre o cerne hidrofóbico do domínio) e, por conseguinte, não apresentam mais a significativa influência sobre a dinâmica interna do DBD observada para os confôrmeros nativos (RXR<sub>RMN</sub><sup>I</sup> e <sup>XX</sup>). Desta forma, o novo *coil* formado na extensão C-terminal do DBD (posterior à A204) e contínuo ao domínio D (a *hinge*, igualmente desordenada), pode dissipar sua energia térmica oscilando livremente, sem implicações quanto ao empacotamento e à mobilidade do cerne globular do domínio. Em particular, nota-se que o D-box, livre da influência vibracional da hélice C, se mantém em uma conformação mais “alargada”, permissiva para sua interação simultânea com o sulco menor e (no caso do monômero 5') com o parceiro de dimerização. Nota-se também que a K175, na ausência dos movimentos de extensão e contração do D-box (devido à ausência dos movimentos de rotação da tríade M200, K201 e R202, que permanecem no passo de hélice remanescente, mantendo seus respectivos contatos originais, agora mais estáveis), permanece próxima à C190 ao longo de toda a trajetória e, durante sua maior parte, empacotada sobre este resíduo. De forma análoga, também não é mais observado o acoplamento entre os movimentos do T-box e o “abrir e fechar” dos contatos entre o primeiro e o segundo dedos de zinco aparente na primeira componente para o confôrmero RXR<sub>RMN</sub><sup>XX</sup>.



**Figura 4.17** – Projeções das componentes principais para a trajetória do monomero DBD na conformação RXR<sub>crys</sub>.

Esta ordenação do D-box na conformação de associação ao DNA é condizente com os dados experimentais prévios da literatura. Nos estudos de RMN de Holmbeck et al(a) a não observação dos NOEs relativos aos prótons da cadeia lateral, bem como o padrão das ressonâncias para o H<sup>1</sup> e para o N<sup>15</sup> do grupo amida na cadeia principal da K175 significativamente mais fracas que as esperadas segundo as conformações observadas nos complexos cristalográficos, sugere que este resíduo de lisina sofre flutuações conformativas no DBD em solução que parecem não se apresentarem nas conformações adotadas na interação ao DNA (o que também é constatado na simulações aqui apresentadas). Já os estudos posteriores de Tilborg et al mostram que a interação de um único monômero ao meio sítio promove um acentuado decréscimo nas taxas de relaxação transversa para o N<sup>15</sup> da cadeia principal e um significativo aumento nos parâmetros de ordem de Lipari-Szabo (S<sup>2</sup>) para os resíduos D173-K175 no D-box, sugerindo que a interação ao DNA (sem a necessidade do parceiro de dimerização) provoca um aumento na ordenação deste segmento, consideravelmente distante da interface proteína:DNA, com um decréscimo substancial de sua mobilidade (de forma condizente ao que é observado nos estudos de MD aqui apresentados como simples consequência da desnaturação do T-box). Como previamente mencionado, neste mesmo estudo, bem como no anterior de Holmbeck et

al(b) foi também constatado que, com a interação de um único monômero DBD ao meio-sítio, o T-box sofre uma mudança conformacional condizente com a desnaturação local deste segmento observada nas diferentes estruturas de homo e heterodímeros de hRXR $\alpha$  associados ao DNA (adotando a conformação em *coil* exibida pelo confôrmero RXR $_{crys}$ ). Também foi observada uma maior ordenação para o segmento 181-187 (entre o segundo dedo de zinco e a hélice B) e uma mudança da inclinação do motivo  $\beta$ -*hairpin* (ambas regiões que se contactam ao *backbone* do sulco maior em faces opostas), bem como um decréscimo na dinâmica do DBD em linhas gerais. O presente estudo sugere que as mudanças conformacionais e decréscimos na dinâmica do cerne do DBD previamente documentadas para a interação de um único monômero à seqüência do meio sítio (envolvendo mudanças em regiões distantes da interface proteína:DNA) ocorrem como consequência direta da desnaturação do T-box seguida da perda do acoplamento dinâmico interno entre estas duas regiões. Os resultados indicam que o empacotamento hidrofóbico do T-box ao cerne do DBD não contribui apenas para a estabilização da estrutura secundária da hélice 3<sub>10</sub> nesta região na conformação comumente adotada em solução, mas também proporciona um alto grau de correlação na dinâmica dos diferentes motivos internos do domínio (ao menos na faixa entre ps e ns), acoplando inclusive, em seus movimentos coletivos, regiões de interação ao DNA e regiões fisicamente distantes e envolvidas na interface de dimerização. Além da não disponibilização da hélice A e do motivo  $\beta$ -*hairpin* (ambos envolvidos no empacotamento do T-box na conformação em hélice) para seu correto posicionamento na associação ao meio-sítio; a mobilidade aferida pelo acoplamento dinâmico ao cerne globular do DBD induz a movimentos não permissivos para a correta associação deste domínio ao DNA (como a maior mobilidade do segmento 181-187 e do próprio motivo  $\beta$ -*hairpin*) e para a formação da interface de dimerização entre os dois DBDs na interação com o elemento responsivo (como a maior mobilidade para o D-box, envolvido na interface de dimerização com o monômero 3', na conformação nativa). Desta forma, mais do que apenas permitir os contatos deste segmento com o sulco menor e (nas repetições diretas com espaçamento entre 1 e 2 pb) com o parceiro 5', a desnaturação do T-box na formação do complexo com o DNA parece ter um papel crucial ao permitir que o cerne globular do DBD alcance a estabilização conformacional e dinâmica necessárias para a correta associação ao meio-sítio e ao parceiro de dimerização.

Em particular, o acoplamento dinâmico entre o T-box e o segmento 181-187 observado nas simulações deste estudo é condizente com o aparente alto acoplamento evolutivo entre estas duas regiões segundo o estudo recentemente publicado por Willis e Griffin . Neste trabalho, foi utilizado o algorítimo de informação mútua *mutual information - MI* de forma a identificar regiões com alto grau de co-evolução ao longo de toda a seqüência do receptor em um alinhamento para uma base de dados de 2094 seqüências, apresentando representantes de todas as espécies para as quais a “assinatura molecular” dos receptores nucleares (constituída de 20 posições chaves) é identificada. Surpreendentemente, dos três resíduos identificados por esta técnica com o mais alto padrão de acoplamento evolutivo para toda a seqüência e dentre todas as subfamílias, dois se encontram na região globular do DBD e justamente sobre este segmento -resíduos 139 e 144 na janela de alinhamento multisequências (*Multiple Sequence Alignment - MSA*), correspondendo aos resíduos R182 e Q183 no DBD de hRXR $\alpha$  - e o terceiro se encontra no T-box - resíduo 195 no MSA, correspondendo ao resíduo E207 em hRXR $\alpha$ -DBD. Na estrutura cristalográfica do heterodímero completo de hPPAR $\gamma$ :hRXR $\alpha$  em DR1, os três resíduos homólogos desta classificação para cada um dos dois receptores estão tanto envolvidos na interface de dimerização (os MSA 139 e 144 de hPPAR $\gamma$  e o MSA 195 de hRXR $\alpha$ ), quanto interagindo com a seqüência 5' flankeadora obrigatória em elementos consenso do hPPAR $\gamma$  (PPRE) (o MSA 195 do hPPAR $\gamma$  (F182)). Conforme se verá nas subseções , os resultados de MD para o homodímero de hRXR $\alpha$  associado a diferentes HREs sugerem que estes três resíduos apresentam papel chave no estabelecimento de uma interface de dimerização no complexo em DR1 condizente com a estabilização da principal componente vibracional do DNA, facilitando desta forma a associação individual de cada monômero com seu respectivo meio-sítio. Desta forma, o acentuado acoplamento destas duas regiões com aparentes participações cruciais para o mecanismo de associação do hRXR $\alpha$  a seu elemento responsivo (e cuja importância e colaboratividade de função é, ao que indicam análises recentes, altamente conservada dentre diferentes NRs) e sua “liberação” mútua na conformação de associação ao DNA sugerem fazer parte de um mecanismo “chave e fechadura” para a interação com o HRE e parceiro corretos (vide abaixo).

A simultânea maior liberdade conformacional inter-motivos e desacoplamento dos modos normais internos do cerne do DBD em relação aos movimentos do T-box na conformação de interação ao DNA apresenta vantagens óbvias no que concerne a possíveis caminhos de formação dos complexos: além dos diferentes motivos estruturais do DBD nesta conformação apresentarem a independência conformacional e dinâmica desejável para se ajustarem a seus respectivos sítios de interação na formação dos complexos com diferentes parceiros e/ou diferentes HREs (O cerne globular do domínio apresentando o necessário menor padrão vibracional para melhor estabilizar a cavidade do sulco maior, onde se encontra a seqüência responsiva (com a possibilidade para o D-box no segundo dedo de zinco de se ajustar também à interface de dimerização com o parceiro 3') e o T-box com a plasticidade e independência em relação ao restante do domínio necessárias para se ajustar ao sulco menor e/ou à interface de dimerização com o parceiro na posição 5'), o desacoplamento vibracional entre os dois motivos torna-os, também, individualmente mais sujeitos às influências conformacionais e/ou dinâmicas decorrentes da interação com os diferentes sítios do DNA, diferentes parceiros de dimerização, ou mesmo diferentes proteínas reguladoras (e reciprocamente, ambos os motivos podem influenciar de forma independente estes diferentes sítios, parceiros e/ou reguladores). Desta forma, pode ser conferida uma maior maleabilidade para as necessárias mudanças configuracionais para a formação de diferentes complexos<sup>50</sup>. Assim, o desacoplamento dinâmico e conformacional constatado nos estudos de MD aqui apresentados entre o T-box e o cerne globular do DBD na conformação induzida pelo DNA pode ser um fator determinante para a conhecida promiscuidade do RXR para a interação com diferentes HREs na presença de diferentes parceiros de dimerização. Uma outra vantagem aparente é que, quanto a desnaturação do T-box promova o seu desacoplamento da dinâmica interna do DBD, a componente principal ilustrada na Fig. (4.13-C-Componente I) indica uma maior correlação entre as orientações do T-box e do centro de massas de todo o cerne globular do domínio. Assim sendo, uma vez que a porção mediana do DBD seja mecanicamente estabilizada (o que acontece com a alocação do segmento Q206-R209

<sup>50</sup> Nas seções será mostrado que, na formação do complexo específico, um novo acoplamento dinâmico/conformacional em escala significativamente maior ocorre, envolvendo as diferentes regiões dos dois domínios e os dois meios-sítios do DNA, da extremidade 5' à 3' do complexo (um acoplamento *intermolecular*). Também será mostrado como os resultados sugerem poder este acoplamento chegar, inclusive, à interface de dimerização dos LBDs.

- cujos resíduos em sua extremidade são assinalados em vermelho nas Figs. 4.13-C, 4.16 e 4.17 - no sulco menor) sua extremidade C-terminal pode ser utilizada como uma “alavanca molecular”, cuja orientação - ditada pela interação com o parceiro 5’ e pelos movimentos vibracionais do DNA) é correlacionada ao correto posicionamento do cerne do domínio sobre o meio-sítio 3’ (vide subseções ). A necessidade da desestruturação do T-box para que a interação estável do cerne globular ao meio-sítio ocorra se apresenta como uma primeira etapa no mecanismo “chave/fechadura” para a interação específica do RXR ao parceiro e HRE corretos, visto que, dado o custo energético substancial para a manutenção desta conformação desestruturada em solução (devido aos efeitos hidrofóbicos ilustrados na Fig. 4.12) é esperado que ela apenas se mantenha estável no ambiente químico provido pelos parceiros adequados. Assim, os resultados apresentados nesta subseção são os primeiros deste trabalho a indicar que este mesmo mecanismo “chave fechadura” para a interação do RXR-DBD dimérico ao elemento responsivo no DNA envolve uma energética substancialmente maior que a inferida pela esparsa interface de dimerização entre os domínios nas estruturas cristalográficas.

Na próxima subseção, as propriedades dinâmicas e conformacionais da outra unidade do complexo - a seqüências de oligonucleotídeo contendo o elemento responsivo - serão analisadas de forma a se inferir quanto aos tipos de penalidade energética impostas por este elemento para a associação do DBD. Em sequida, a natureza da necessidade do parceiro de dimerização será inferida a partir da análise de simulações de MD para modelos de monômeros interagindo com cada um dos dois meio sítios (construídos a partir de estruturas cristalográficas existentes para complexos diméricos (Vide Seção )). Finalmente, após esta primeira etapa de análise para as componentes do sistema, serão apresentados os resultados de modelagem estrutural e dinâmica molecular referentes aos complexos do dímero associado a diferentes elementos responsivos.

#### **4.2.3 Conhecendo as componentes do sistema (II) - Mobilidade intrínseca, blindagem eletrostática por íons e natureza dos modos normais das sequências de HREs conferem penalidades adicionais para a interação do DBD**

Na seção foi realizada uma pequena revisão a respeito das propriedades físico-químicas do DNA em solução e de seu impacto quanto à formação de complexos proteína-DNA. Como visto, duas características fundamentais do DNA conferem penalidades energéticas à associação de fatores de transcrição a seqüências responsivas: sua considerável mobilidade em solução - que adiciona um peso entrópico conformacional na formação de complexos proteína:DNA, *a priori*, com menor liberdade conformacional - e sua natureza polieletrólica -tendendo a acumular uma camada bem localizada de íons e moléculas de água em suas imediações (formando uma barreira de ativação a ser vencida para a associação da proteína) . Além disto, o contexto da seqüência de pares de bases implica em eventuais direções preferenciais de movimento (flexões e torções anisotrópicas e compressões preferencialmente sobre determinadas regiões) e, eventualmente, em conformações médias em solução com diferentes comprimentos de persistência da cadeia (embora este último efeito seja mais notório para seqüências contendo acima de 140 pb - o comprimento de persistência médio (em unidades de pares de base) para uma molécula de DNA em solução). Isto ocorre devido às divergências quanto a estabilidade do empacotamento hidrofóbico para diferentes passos de base (vide Tab.) . Os três tipos principais de simetria de movimentos ao longo de uma cadeia de DNA (flexão, torção e compressão), bem como as eventuais direções preferencias destes movimentos e/ou conformações mais favorecidas, impõem também uma relativa limitação quanto à natureza das mudanças conformacionais passíveis de serem induzidas a uma dada seqüência pela associação do complexo protéico. Finalmente, estas características tendem a ser interdependentes, visto existir um considerável números de estudos (computacionais e experimentais) demonstrando a correlação entre a mobilidade e mesmo entre conformações preferencias para uma dada seqüência de DNA com sua interação aos eletrólitos em solução com especial atenção para o papel do íon Mg<sup>2+</sup>

De particular relevância para a restrição da interação de proteínas que se ligam na forma

dimérica ao DNA (como é o caso dos complexos formados pela maioria dos receptores nucleares membros do subgrupo dos não esteróides e, em particular, do hRXR $\alpha$  e seus parceiros) é a simetria do complexo formado comparado à simetria dos modos normais dos movimentos coletivos da seqüência responsiva. Como previamente mencionado na seção , a quebra de simetria de vibração para os movimentos coletivos de uma molécula longa (como uma cadeia de desoxi-oligonucleotídeo), com a consequente substituição de modos globais por modos locais de maior freqüência, confere uma penalidade energética adicional para a formação de complexos macromoleculares . Em contrapartida, a disponibilização de novos modos globais de baixa freqüência, envolvendo movimentos coletivos do sistema como um todo, acrescentam um défice adicional na energia livre de formação do complexo (visto que estes movimentos tendem a conferir um aumento significativo na entropia configuracional do sistema com pequenos acréscimos na entalpia vibracional, enquanto que movimentos locais de alta frequência tendem a apresentar o efeito contrário), favorecendo, portanto, o processo de associação.

Para uma seqüência de DNA dupla fita próxima à forma B, os movimentos envolvendo contrações do sulco maior e abertura do sulco menor têm mais implicações quanto aos movimentos coletivos de alta amplitude para a molécula como um todo que os movimentos em sentido contrário. Isto ocorre devido à considerável maior liberdade para a molécula de DNA se curvar na direção do sulco maior, sem aumentar significativamente a repulsão entre os fosfatos, que na direção do sulco menor . Este efeito é substancialmente acentuado sob a influência de pequenos cátions multivalentes (em particular o Mg<sup>2+</sup>) com capacidade de se inserir na entrada do sulco maior do DNA na forma B, blindando a repulsão entre os fosfatos das duas fitas complementares e facilitando a flexão da molécula de DNA na direção do sulco maior (vide seção e texto abaixo). Além disto, a periodicidade estrutural de uma determinada seqüência dupla fita de DNA pode ser medida pelo número de unidades de sulcos maiores que ela atravessa<sup>51</sup> . Desta forma, se uma dada seqüência responsiva constituída de dois meios-sítios e um número variável de espaçadores pode ser distribuída sobre um número inteiro de “unidades de sulco

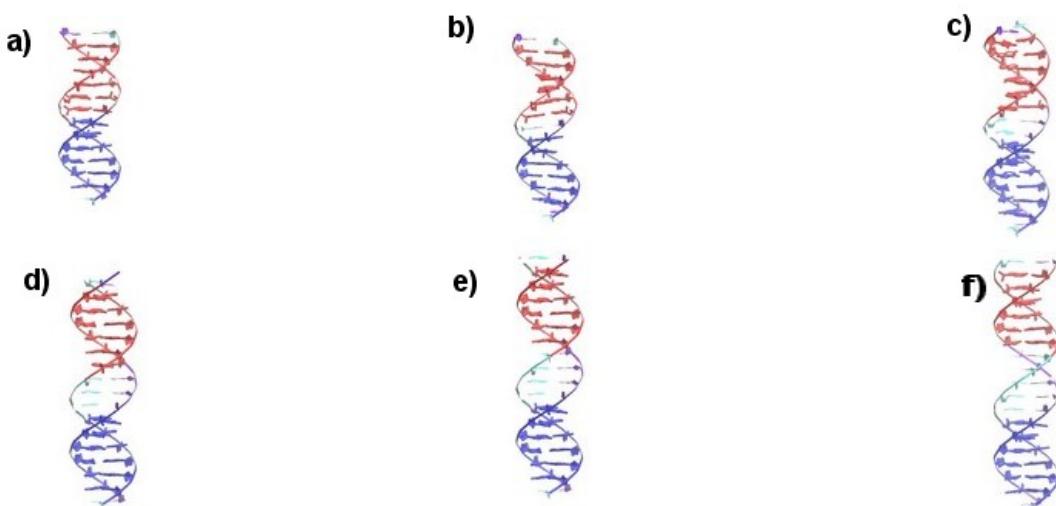
---

<sup>51</sup> Uma volta completa do DNA dupla-fita sobre o eixo principal da hélice equivale a cerca de 10.4-10.5 pb, correspondendo a duas unidades de sulco maior (este, por sua vez, contendo um número médio de cerca de 5.2 pb)

maior”, existirão um maior número de harmônicos vibracionais passíveis de percorrerem a extensão da referida seqüência com movimentos simétricos de ambos os sítios (aumentando, por exemplo, a probabilidade de que um dado monômero, ao se aproximar da seqüência, encontre ambos os sítios “abertos” e com a mesma acessibilidade de interação). No caso dos elementos responsivos compostos por repetições do hexâmero AGGTCA separadas por um número variável de pares de bases, os modos envolvendo distorções na geometria do sulco maior têm implicações particularmente relevantes. Nota-se que esta seqüência arquetípica composta de 6 pb ocupa uma janela topológica correspondente a aproximadamente 1.15 unidades de sulco maior em um fragmento de DNA na forma B. Desta forma, se imaginarmos um fragmento de DNA na forma B com os dois meios-sítios posicionados em seqüência e sem espaçamento (como no caso de DR0, considerando os dois meios-sítios no arranjo de repetição direta, ou Pal, para uma repetição palindrômica simples) e tomando um posicionamento arbitrário para o fragmento de DNA de forma a que o primeiro par de bases do primeiro meio-sítio se encontre na extremidade 5’ do sulco maior, o sexto e último par de bases, por sua vez, extrapola o número de bases acomodáveis no mesmo sulco por um fator de aproximadamente 0.15 (Fig.4.18). Desta forma, a seqüência arquetípica do meio-sítio 1 “invade” o segundo sulco maior (onde reside o meio-sítio 2), deslocando também a fase da segunda seqüência arquetípica em uma fração de 0.15 na direção 3’ deste segundo sulco. Embora para este primeiro caso a diferença de fases seja pouco significativa (podendo-se admitir que ambos os meio-sítios se encontram em janelas topológicas praticamente idênticas na estrutura do DNA e que o segundo meio sítio se encontra na face oposta do primeiro<sup>52</sup>), conforme se aumenta a seqüência espaçadora, o deslocamento de fase do meio sítio 2 no segundo sulco maior também aumenta. Assim, embora ambos os sítios contenham a mesma seqüência de seis pares de bases, eles apresentarão topologias diferentes em uma mesma “janela de leitura” da estrutura do DNA e terão suas configurações médias alteradas de formas também distintas sob a influência de diferentes modos vibracionais coletivos para o fragmento como um todo. Para DR4, os quatro pares de base na seqüência espaçadora ( $\sim 0.77$  unidades de sulco maior), somados aos dois fatores de defasagem de 0.15

<sup>52</sup> Na verdade devido ao fator de defasagem de 0.15, mesmo com o espaçamento 0, os dois meio sítios não se encontram exatamente em faces opostas ( $180^\circ$ ), apresentando um acréscimo de  $27^\circ$  de rotação no sentido da dupla hélice

para cada seqüência arquetípica, colocam os dois meio-sítios praticamente na mesma face do DNA e ocupando topologias similares. Conforme se verá nos resultados apresentados e discutidos abaixo, os diferentes padrões vibracionais oriundos das diferentes fases entre as seqüências arquetípicas, bem como a diferenças locais quanto ao empacotamento entre os diferentes passos de pares de bases nas duas seqüências arquetípicas têm implicações quanto à disponibilidade média de cada um dos dois meios-sítios para o primeiro evento de ligação, o que por sua vez sugere influências quanto à afinidade de ligação do primeiro monômero e mesmo para a polaridade dos complexos de receptores nucleares com diferentes HREs.



**Figura 4.18** – Diferenças de fases entre os dois meios sítios para diferentes HREs

As mesmas diferenças topológicas entre os dois meios sítios, com eventuais diferenças quanto aos padrões de simetria vibracional, implicam também quanto ao tipo de interface permitida entre os monômeros (no caso da interação dimérica) e, sobretudo, quanto à natureza dos movimentos que devem ser induzidos à molécula de DNA de forma a permitir a correta e simultânea associação dos dois domínios com a formação da interface de dimerização. O balanço energético entre as modificações topológicas e vibracionais necessárias e as novas interações e modos de movimentos coletivos introduzidos, aconta para a cooperatividade e afinidade nas interações dos complexos diméricos.

Nesta subseção, estas características dinâmicas e conformacionais no que concerne aos HREs DR1-4, bem como suas prováveis implicações quanto à energética do processo de associação

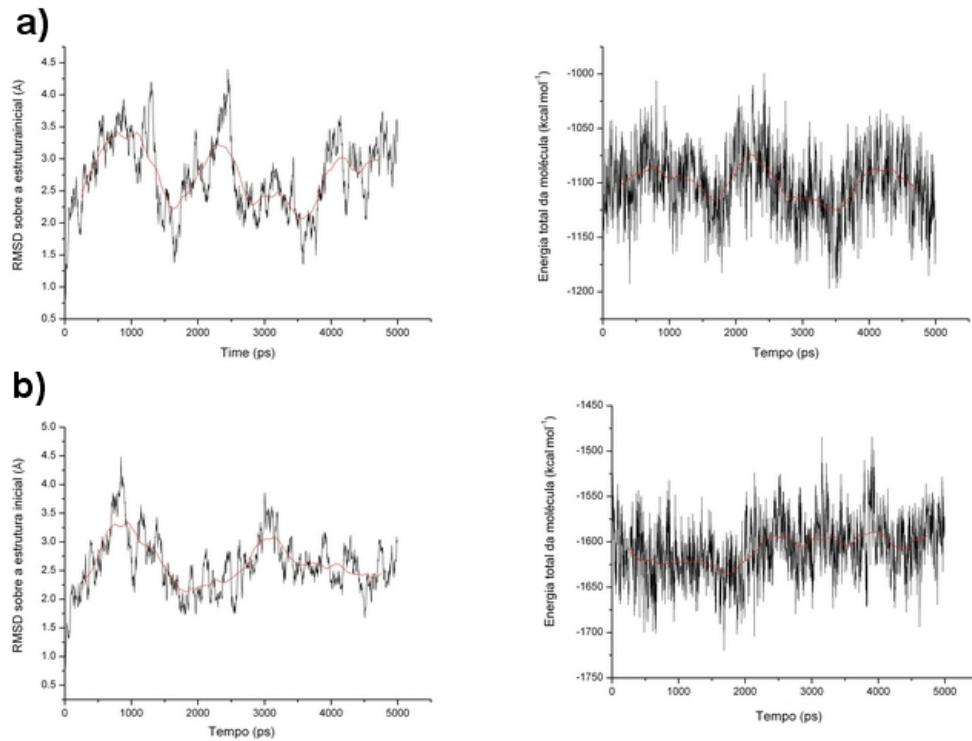
do(s) DBD(s) serão discutidas através da análise de trajetórias de MD, extendendo-se as conclusões referentes à análise dos resultados para os outros quatro HREs abordados nos estudos experimentais.

**Assimetrias nos padrões vibracionais entre os dois meios-sítios sugerem implicações quanto às afinidades diferenciadas para a interação do primeiro monômero e para a polaridade de interação dos complexos do hRXR a diferentes HREs**

O papel significativo das entropias conformacionais para a energética das seqüências de DR1-4 livres em solução<sup>53</sup> pode ser inferido a partir da comparação dos perfis de evolução dos rmsds dos átomos pesados em relação às respectivas estruturas iniciais com a evolução da energia total da molécula durante as trajetórias (Fig. 4.19). Nota-se que no curto intervalo de tempo das simulações (aproximadamente 5 ns para todos os HREs), os oligonucleotídeos demonstraram elevada mobilidade, com movimentos de baixa e média amplitudes (conforme verificado pelo padrão oscilante dos valores de rmsd, tanto locais (4 ps) quanto para os valores médios sobre janelas de 200 ps), acessando uma considerável abrangência de configurações sem alterar significativamente os valores da energia potencial média. Tanto as trajetórias dos modelos estruturais na forma B (DR1, DR2 e DR4) quanto a da estrutura de DR3 (retirada de um modelo cristalográfico experimental) demonstraram um rápido acesso ao equilíbrio térmico aparente, com padrões similares de rmsd oscilantes e estabilização da energia potencial da molécula<sup>54</sup>, demonstrando, desta forma, a validade da comparação destas simulações com os dos modelos gerados a partir de complexos cristalográficos nas próximas subseções.

<sup>53</sup> Conforme mencionado em as configurações iniciais dos HREs DR1,2 e 4 foram construídos a partir do algorítimo do programa Coot (pacote CCP4) para a construção de modelos estruturais de DNA dupla fita na forma B. As seqüências senso utilizadas corresponderam, respectivamente, à seqüência de DR1 e à “pseudo-seqüência de DR2” da estrutura cristalográfica relatada em , bem como à seqüência de DR4 da estrutura comunicada em , esta última tendo sua seqüência espaçadora mutada de 5'-TTTC-3' para 5'-CAGG-3', de forma a coincidir com a seqüência espaçadora utilizada nos estudos experimentais de anisotropia de fluorescência (vide subseção 3.1.3 ). Isto foi realizado tanto devido ao tamanho relativamente extenso desta seqüência espaçadora, quanto pela particularidade do hRXR $\alpha$ - $\Delta$ AB ter demonstrado perfis de afinidade e cooperatividade similares aos observados para o elemento responsável DR1 para o DR4 espaçado pela mesma seqüência nos estudos relatados na seção 4.1. Já para DR3, a estrutura inicial foi retirada diretamente da estrutura cristalográfica do heterodímero VDR:RXR em DR3 (- código do PDB 1YNW). Além das duas seqüências dos meio-sítios e das respectivas seqüências espaçadoras, os quatro oligonucleotídeos continham também dois pares de base flanqueadores, respectivamente nas posições 5' e 3'. As seqüências de DR1 a DR4 continham de 15 a 18 pb, respectivamente.

<sup>54</sup> Para a seqüência de DR4, de maior tamanho, a energia potencial ainda demonstra uma leve inclinação ascendente decorridos os 5 ns de simulação.



**Figura 4.19** – Progressões do equilíbrio para os HREs DR1 e DR4 de forma representativa para todos os quatro HREs analisados neste estudo

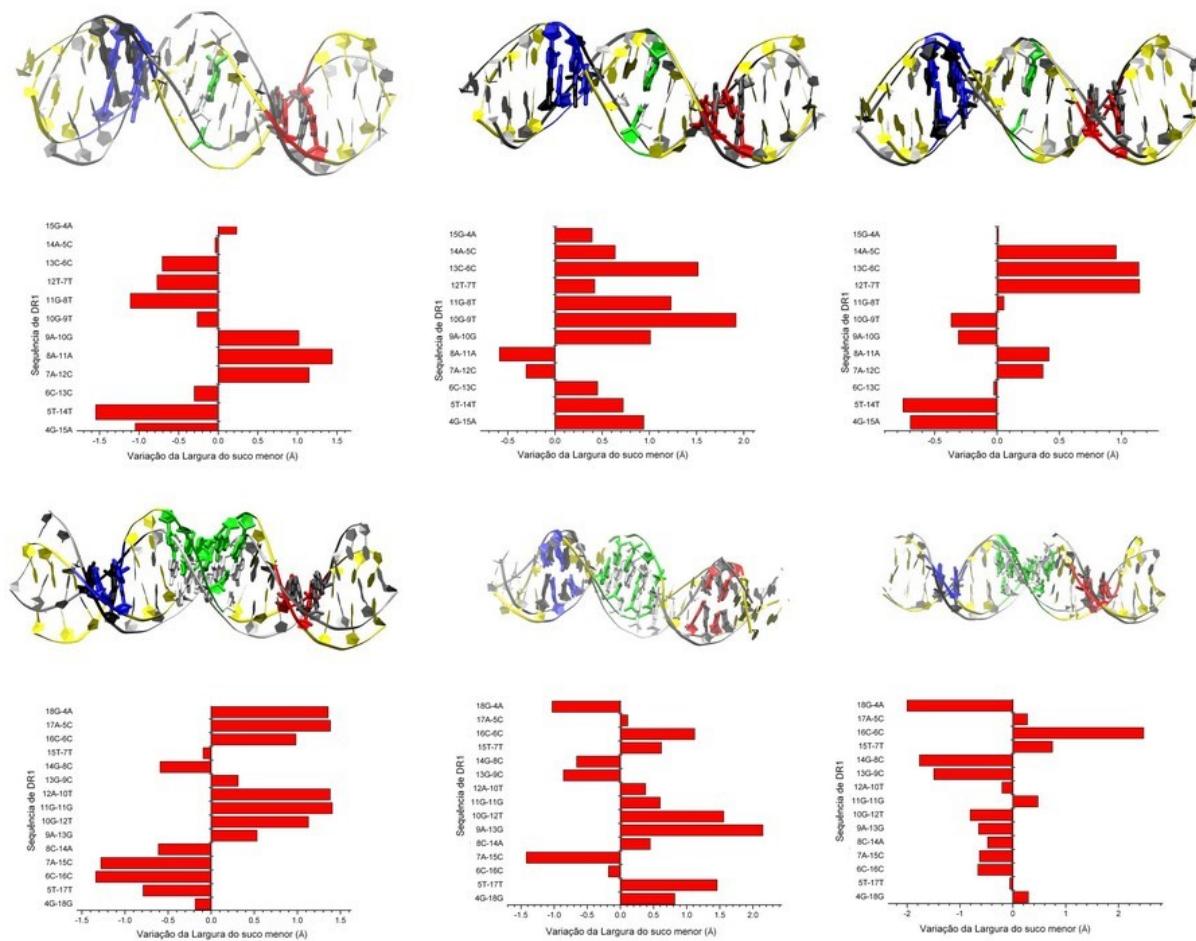
Para todos os oligonucleotídeos, os perfis de rmsd também sugerem uma certa periodicidade para os movimentos térmicos da molécula, com um padrão de alteração das distribuições de distâncias semelhante ao de um oscilador harmônico amortecido. Este padrão, por sua vez, é condizente com o modelo *worm-like-chain* (WLC) de Porod-Kratsky, comumente ajustável para a elasticidade entrópica de cadeias de DNA dupla fita em solução . O comportamento WLC é também sugerido pelo padrão das componentes principais ilustradas para os HREs com espaçamento mínimo e máximo (respectivamente, DR1 e DR4), ambos ilustrados na Fig. 4.20. Nota-se que os movimentos térmicos de ambos os HREs são compostos de movimentos de flexão e torção da molécula sobre o eixo principal, além de pequenos movimentos locais de compressão deste mesmo eixo, que tendem a diminuir a distância média entre as extremidades comparado à estrutura *clássica* do DNA na forma B (que, por sua vez, condiz com as configurações iniciais para DR1, DR2 e DR4 e é significativamente próxima à configuração inicial para DR3, respectivamente utilizadas nas simulações). Nota-se também que os autovetores relacionados aos diferentes padrões de movimento produzem variações nas larguras do sulco menor e maior ao longo da molécula. Os padrões de variações na largura do sulco me-

nor ao longo da molécula decorrentes das componentes para DR1 e da primeira componente para DR4 são similares aos modos de vibração de uma corda, com diferenças de freqüências e fases para cada modo ao longo dos segmentos de DNA dupla-fita (Figs. 4.20-B e D). Já as duas componentes de maior ordem para DR4 (sobretudo a terceira) apresentam significativos graus de distorções ou “ruídos” (afastando-se da simetria típica de modos normais). Isto ocorre devido a diferenças locais de mobilidade (decorrentes, sobretudo, das distinções nas energias de empacotamento para diferentes passos de base, conforme exposto na Tab. ) que se fazem mais notórias nos movimentos de maior freqüência (para os quais pequenas variações locais apresentam maior relevância) e sobretudo para sistemas com um maior número de graus de liberdade (como é caso da seqüência de DR4 em relação à de DR1, visto os três pares de base adicionais da primeira em relação à segunda), para os quais existe um maior número de possíveis combinações de movimentos nos modos de menor amplitude em uma dada configuração da molécula. O movimento total é razoavelmente bem distribuído entre as três primeiras componentes principais (sendo que as duas primeiras descrevem movimentos coletivos de amplitudes consideráveis, envolvendo a molécula em toda a sua extensão), conforme pode ser constatado pelos autovalores relativos e pelas respectivas projeções das componentes nas trajetórias (Fig. 4.20-A, C, E e F), novamente indicando um sistema com significativa possibilidade de arranjos no espaço de fases e, portanto, apresentando considerável entropia conformacional.

As diferenças de simetria e fase entre os dois meio-sítios<sup>55</sup> para os dois respectivos HREs, bem como as distinções no empacotamento entre as bases na interface entre os dois meios-sítios, implicam igualmente em diferenças de simetria para os movimentos coletivos descritos pelas respectivas primeiras componentes principais. Para DR1, os respectivos centros geométricos dos dois meios sítios se encontram em faces distintas do DNA, com simetria próxima à trigonal (ângulo próximo a 240 ° no sentido da hélice e próximo a 120 ° no sentido horário). Também a fase da segunda seqüência arquetípica é deslocada de aproximadamente 1/3 na direção 3' do segundo sulco maior. Desta forma movimentos simétricos de expansão e/ou contração si-

---

<sup>55</sup> A palavra *simetria*, no presente contexto, é utilizada referindo-se apenas à orientação do meio sítio na topologia *macroscópica* da molécula de DNA, *i.e.*, a face da molécula sobre a qual o meio sítio se orienta. Não confundir com o sentido *microscópico* em que a palavra é utilizada quando se referindo à orientação de seqüências na forma de repetições diretas ou palíndromos.



**Figura 4.20 – Componentes principais para os HREs DR1 e DR4.**

multâneas para ambos os meios-sítios ou de flexão do eixo principal da molécula com ambos os sítios se curvando de forma simétrica em direção ao sulco maior ou menor se tornam pouco prováveis. Assim, o autovetor referente à primeira componente principal para a trajetória do DR1 (que, aparentemente, reflete um modo normal de vibração para este fragmento de oligonucleotídeo) descreve um movimento de torção, com superelicoidização a partir do passo C-A no fim do meio-sítio 5' (Fig. 4.20-A-*Componente 1*). Nota-se que o movimento (tomando-se como referência a estrutura mais próxima ao DNA na forma B) provoca uma acentuada inversão de largura do sulco menor nas imediações dos meios sítios com a seqüência espaçadora (Fig. 4.20-B-*Componente 1*), com um pronunciado estreitamento do sulco na tríade AAA, na porção mediana da molécula, e alargamento deste sulco nas respectivas regiões medianas das seqüências dos dois meios-sítios. A expansão do sulco menor para a região do meio-sítio 5' é consideravelmente mais pronunciada que para o meio sítio 3'. Este movimento provoca um estreitamento e aprofundamento do sulco maior 5' (que assume uma geometria local próxima à de uma molécula de DNA na forma A), de forma a “esconder” o meio-sítio 5', enquanto o sulco maior 3' é, em contrapartida, localmente expandido, ocasionando assim uma maior exposição da seqüência arquetípica 3'. A torção assimétrica para os dois meios-sítios tem por base, além da assimetria de fases previamente mencionada, a assimetria do empacotamento entre os passos de base para as seqüências de ambos os meios-sítios na interface com a seqüência espaçadora. Comparando-se a seqüência arquetípica AGGTCA com as energias de empacotamento entre passos de base listadas na Tab. , nota-se que o passo C-A na extremidade 3' apresenta um empacotamento hidrofóbico (com energia  $-0.55 \text{ kcal}\cdot\text{mol}^{-1}$ ) substancialmente menos pronunciado que os empacotamentos em suas imediações<sup>56</sup>. Desta forma, o movimento de torção descrito pela componente 1 se acumula de forma mais notória neste “elo mais fraco” do empacotamento da seqüência de DR1, provocando simultâneos movimentos de *tilt*, *rolling* e de *abertura* na seqüência senso e antisenso deste passo de bases (com rotações dos átomos C2' e O4' dos respectivos anéis ribose para o passo antiseno na direção do sulco menor (o que, por sua vez, causa o afastamento dos fosfatos no *backbone* e inclinação das bases também na direção do sulco me-

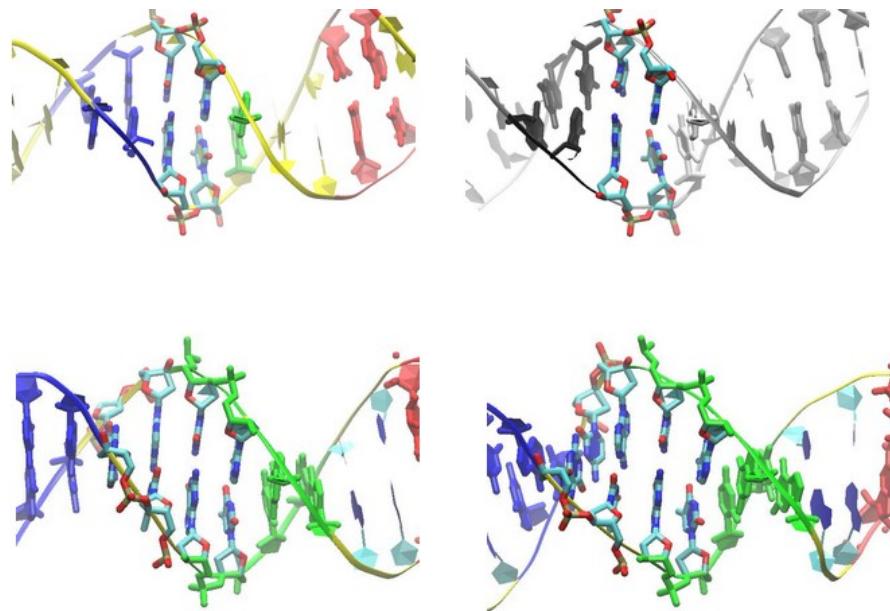
<sup>56</sup> Este efeito se torna mais pronunciado quando a seqüência espaçadora é constituída por uma base púrica (A ou G), sendo estas, justamente, as seqüências que conferem maior afinidade para o homodímero de hRXR $\alpha$ , bem como para os heterodímeros com isoformas de PPARs e RARs em DR1 .

nor) e um movimento aproximadamente inverso para os grupos análogos na seqüência senso (Fig. 4.21-A), ocasionando então a maior abertura do sulco na seqüência 5' a esta posição e seu estreitamento nas imediações 3' à mesma (com um consequente movimento inverso considerando os sulcos maiores). Conforme acima mencionado, este padrão de movimento provoca com que o meio-sítio 5' permaneça, a maior parte do tempo de simulação, consideravelmente menos exposto que o meio sítio 3'. Este padrão vibracional, com o principal modo coletivo tornando o meio-sítio 5' menos exposto que o 3', é condizente com a polaridade 3' de ligação do hRXR $\alpha$  nos complexos com DR1 e é uma explicação razoável para a observação de complexos monoméricos de hRXR $\alpha$ , em experimentos de *gel-shift*, com seqüências curtas de DR1 em que o sítio 5' é mutado para uma seqüência não responsiva (deixando-se a seqüência 3' inalterada), em contrapartida à não observação de qualquer complexo em experimentos onde a mutação é introduzida na seqüência 5'. É também condizente com o “arraste” inicial para a interação do primeiro monômero (seguido da ligação cooperativa do dímero) com a seqüência de DR1 nos experimentos de anisotropia de fluorescência apresentados na subseção 4.1.1. Segundo este padrão, para o primeiro evento de ligação, apenas o meio-sítio 3' se encontra consideravelmente disponível, o que limita o número de microestados acessíveis para a ligação do primeiro monômero comparado ao que ocorre, por exemplo, para F2 e DR5 (vide discussões a respeito dos elementos com repetições  $\geq 5$  no texto abaixo) cujas isotermas apresentaram perfis sugestivos para um primeiro evento de ligação relativamente mais facilitado. Seguindo a mesma lógica, o mecanismo para a cooperatividade observada nos experimentos com DR1 da subseção 4.1.1 deve passar, obrigatoriamente, por uma contraposição ao movimento descrito pela primeira componente principal da Fig. 4.20-A com a interação de ambos os monômeros a DR1<sup>57</sup>. Um movimento relativamente mais bem distribuído para ambos os sítios, só ocorre na segunda componente, com um estreitamento do sulco menor na região próxima ao par de bases espaçador, o que provoca uma moderada flexão do eixo principal da hélice e rotação antissimétrica de cada respectivo meio-sítio na direção do sulco maior (Fig. 4.20-B-Componente 2). A componente 3, como esperado, reflete movimentos de maior freqüência, com aberturas e

---

<sup>57</sup> Os mecanismos moleculares da cooperatividade decorrente do complexo dos dois monômeros serão analisados na subseção .

estreitamentos locais do sulco menor ao longo da cadeia (Fig. 4.20-B-*Componente 3*).



**Figura 4.21** – Menor empacotamento no passo C-A 3' terminal do primeiro meio-sítio confere perturbações no sulco maior 5' para DR1 e DR4

Para a seqüência de DR4, conforme previamente mencionado, as fases dos dois meio-sítios e da sequência espaçadora se distribuem de tal forma na topologia do fragmento correspondente à seqüência do elemento responsivo que ambos os sítios se encontram aproximadamente na mesma face da molécula e com um espaçamento quase condizente com o de um sulco maior entre as duas seqüências (Fig. 4.18-D). Desta forma, o movimento principal da molécula é descrito primeiramente pela flexão do eixo da hélice, de forma aproximadamente simétrica sobre a região central (correspondente à seqüência espaçadora), com os dois meio-sítios oscilando sobre esta região central com um movimento aproximadamente em *U* (como dois corpos fixos nas extremidades de um oscilador harmônico com grau de liberdade correspondente à torção de seu eixo principal) (Fig. 4.20-C e D). Devido à geometria de plano inclinado da molécula de DNA, estes movimentos de flexão são também acompanhados por rotações antissimétricas das extremidades dos dois respectivos meios-sítios, conservando-se assim o momento angular no centro de massas da molécula (note as direções opostas das variações de largura do sulco menor em cada uma das extremidades no diagrama para a primeira componente na Fig. 4.20-D-*Componente 1*). O movimento torsional com maior contribuição para a contração do sulco

maior correspondente ao meio sítio 5' também é observado, mas passa a ser descrito pela segunda componente, com menor representatividade nos movimentos coletivos da molécula ao longo da simulação. Para esta segunda componente, também é observado um efeito aparentemente devido à seqüência espaçadora utilizada (5'-CAGG-3'). Nota-se que o efeito de rotação e contração mais acentuada para o primeiro sulco maior (e consequente abertura do sulco menor na face oposta) a partir da região prévia ao passo C-A na extremidade 5' do primeiro meio-sítio também ocorre, mas é consideravelmente atenuado devido ao fechamento do sulco menor na região imediatamente correspondente ao passo C-A e à pronunciada abertura do mesmo sulco na região da seqüência espaçadora (Fig. 4.20-C e D -*Componente 1*). No detalhe mostrado na Fig. 4.21-B, observa-se que este efeito é devido a uma contraposição dos movimentos de *tilt roolling* e *abertura* do passo C-A na extremidade 5' do primeiro sítio com os mesmos movimentos para o passo C-A na extremidade 3' da seqüência espaçadora. Desta forma, a curvatura do primeiro sulco maior se torna mais bem distribuída, diminuindo a intensidade da oclusão do primeiro meio-sítio devido a este movimento. Este padrão ilustra o papel da seqüência espaçadora, sobretudo para HREs que apresentam esta mesma seqüência em dimensões consideráveis (como é o caso de DR4). Além disto, a fase do segundo meio sítio significativamente mais deslocada na direção 3' comparado a DR1 provê com que a rotação decorrente do movimento descrito pela segunda componente provoque um certo “recobrimento” da porção mediana deste segundo meio sítio por sua extremidade 3'<sup>58</sup>. Como se verá abaixo, este padrão também tem implicações quanto à conformação média do segundo meio sítio de DR4 em solução, o que por sua vez aparenta ser condizente com o ajuste deste mesmo sítio à superfície do parceiro 3' na estrutura cristalográfica do complexo do heterodímero hRXR $\alpha$ :hTR $\beta$  em DR4. Desta forma, as duas principais componentes de movimento da cadeia de DNA a serem estabilizadas para a correta manutenção do complexo dimérico em DR4 são o afastamento

De particular relevância para as restrições impostas à associação de receptores nucleares é o fato de que os movimentos de maior amplitude, para os quais este alto grau de contribuição entrópica é sugestivo, diminuem a acessibilidade dos sítios de associação do monômero: os

<sup>58</sup> De forma a tornar mais clara a origem desta diferença, uma seta na Fig. 4.20-C-*Componente 2* indica a região na topologia da molécula onde o segundo meio-sítio é localizado em DR1.

sulcos maiores dos respectivos oligonucleotídeos (onde estão localizadas as seqüências responsivas), sob o efeito dos movimentos descritos pelas componentes principais (em particular as duas primeiras), se encontram em alta mobilidade, além de serem submetidos a significativas distorções tanto em seu eixo interno (induzindo a valores de profundidade e largura do sulco maior não ajustáveis à associação da proteína) quanto no eixo global da molécula (que por sua vez correlaciona os dois meio-sítios), diminuindo, desta forma, a probabilidade para configurações inter-sítios permissivas à formação do complexo com correta interação entre os monômeros (as distorções intra e inter sítios decorrentes de cada componente podem ser constatadas comparando-se, respectivamente, as diferenças nas posições dos dois pares de base medianos das seqüências arquetípicas de cada meio-sítio (G-T na seqüência senso) e dos pares de base das seqüências espaçadoras entre as configurações nos dois extremos de cada respetivo autovetor, conforme assinalado nas Figs. 4.20-A e C). Desta forma, a formação de um complexo estável receptor:DNA só se torna possível sob pelo ao menos uma de duas condições chave: ou o complexo proteína:HRE ocorre de forma permissiva para a manutenção de uma significativa parcela dos movimentos coletivos originais da molécula de DNA (de forma a não introduzir um custo entrópico acentuado decorrente da perda destes graus de liberdade internos<sup>59</sup>) e ainda assim condizentes com a correta geometria e energias de associação receptor:meio-sítio (uma condição particularmente pouco provável, visto a elevada ocupação da área superficial da molécula de DNA na formação dos complexos NR:HRE nas estruturas cristalográficas resolvidas, sobretudo nas regiões do sulco maior); ou as interações introduzidas na formação do complexo ocorrem de tal forma a estabilizar os movimentos do DNA, induzindo a uma conformação estável e apropriada para a adaptação à superfície protéica e para a correta orientação entre os monômeros, sendo necessário, para tal, a adição de uma significativa compensação energética referente aos graus de liberdade removidos. Além disto, o comportamento altamente cooperativo inerente à formação de complexos específicos de receptores nucleares ao DNA (vide seção 4.1) sugerem que o mecanismo de estabilização do complexo é dependente da associação conjunta dos dois monômeros.

---

<sup>59</sup> Não considerando as contribuições para os graus de liberdade translacionais e rotacionais do sistema como um todo. Estes, forçosamente, são atenuados devido à substancial adição de massa e volume molar no sistema decorrentes da associação da proteína.

A diferença das fases para os dois meio-sítios no plano inclinado da molécula de DNA entre os diferentes HREs, bem como a disposição relativa de passos de bases com empacotamento mais fraco, induzem também a diferenças quanto à representatividade de movimentos simétricos e assimétricos ao longo do eixo da molécula nas principais componentes coletivas.

#### **4.2.4 Conhecendo as componentes do sistema (III) - A ligação de um único monômero distorce as componentes vibracionais da seqüência adicionando aparentes penalidades energéticas, com efeito mais notório na interação ao sítio 5'**

Nas duas últimas subseções, os dois elementos primários para a formação do complexo hRXR $\alpha$ -DBD:HRE - o domínio DBD e o elemento responsivo - tiveram suas trajetórias de MD livres em solução separadamente analisadas. Estas análises isoladas, *per se*, proporcionaram várias introspecções acerca dos mecanismos de reconhecimento e alosterismo na interação dos DBDs com os elementos responsivos adequados. As trajetórias para o monômero DBD em configurações representativas da conformação média em solução e na conformação de reconhecimento ao elemento responsável, demonstraram um acentuado acoplamento vibracional entre o T-box na conformação de hélice (comumente adotada em solução) e o cerne globular do DBD, chegando este acoplamento a regiões estruturalmente distantes e envolvidas na interface de dimerização 3'. Com a desnaturação do T-box e consequente adoção da conformação de interação ao HRE, este alto acoplamento dinâmico é interrompido, alcançando o cerne globular do domínio a necessária estabilidade de movimentos internos e conformacionais para a interação com o meio-sítio e se encontrando o T-box, simultaneamente, na configuração apropriada para se ajustar ao sulco menor e se expandir até a interface com o monômero 3'. Esta desnaturação, por sua vez, apresenta um significativo custo em solução, referente às interações hidrofóbicas da hélice C com o cerne do domínio, o que sugere a necessidade do HRE e parceiros corretos para que a mesma ocorra de forma estável. Já as trajetórias para os HREs DR1-4 demonstraram que as diferenças de fase para os dois meio-sítios na topologia estrutural do DNA dentre diferentes elementos responsivos implica em distinções também nas componentes vibracionais destes elementos, que por sua vez parecem estar envolvidas na seletividade e polaridade

para a interação com diferentes NRs. Foi demonstrado também o papel das interações com os íons em solução na manutenção destas mesmas componentes. O íon  $Mg^{2+}$  mostrou uma maior especificidade que o íon  $Na^+$  para a interação com as regiões centrais nos sulcos maiores (correspondendo também aos centros de massa e regiões medianas para as seqüências dos dois meio-sítios) e para a região próxima ao centro de massa dos HREs (nas imediações das respectivas seqüências espaçadoras) - ambas regiões de crucial importância para o reconhecimento do DBD - bem como uma interação mais estável e com menor distorção da cadeia de DNA; em concordância com o conhecido papel do primeiro na inibição de interações inespecíficas a seqüências responsivas e também conforme este mesmo papel foi demonstrado nos resultados de anisotropia de fluorescência da subseção 4.1.6.

Na presente subseção, serão analisadas as trajetórias dos complexos monoméricos de hRXR $\alpha$ -DBD aos sítios 3' e 5' de DR1, buscando melhor compreender as restrições impostas para a interação de cada um dos dois sítios, a polaridade 3' da interação do RXR-DBD e a necessidade do parceiro de dimerização. Foram realizadas três simulações: Duas para o complexo do monômero com o meio sítio 5', oriundas de duas conformações iniciais independentes e com respectivas trajetórias de 5000 (doravante referida como cA:DR1-A) e 10000 ps (referida como cA:DR1-B), e uma única simulação de 15000 ps para o complexo com o meio sítio 3' (referida como cB:DR1)<sup>60</sup>.

Os perfis de evolução dos rmsds em relação às respectivas estruturas iniciais para os complexos monoméricos demonstram uma considerável fugacidade ao longo de toda a trajetória (Fig. 4.22-A, C e E), não sendo evidenciado o estabelecimento do equilíbrio mesmo decorrido os intervalos de 10000 ou 15000 ps (para a cA:DR1-B e cB:DR1, respectivamente). Conforme será mostrado na próxima subseção, este comportamento é oposto ao observado para as tra-

<sup>60</sup> Conforme previamente mencionado em , para as trajetórias do complexo do DBD com o meio sítio 5', foi utilizada uma conformação inicial “construída” diretamente a partir de estruturas cristalográficas (estrutura do DNA retirada do complexo cristalográfico do homodímero de hRXR $\alpha$ :DR1 - código do PDB:IBY4 - e estrutura do DBD retirada dos resíduos K132-G212 do complexo do heterodímero hPPAR $\gamma$ :hRXR $\alpha$  em DR1 - código do PDB:3DZU - estruturalmente alinhada sobre o monômero 5' (cadeia A) do PDB IBY4 utilizando o algorítmico do programa Pymol ), para a qual foi efetuada a trajetória de 5000 ps (cA:DR1-A) e uma segunda conformação inicial construída a partir do *instantâneo* de 10 ps da trajetória *longa* do homodímero em DR1 (Vide subseções ) (cA:DR1-B). Para o complexo do monômero com o meio-sítio 3', foi realizada apenas uma simulação com duração de 15000ps a partir de uma configuração inicial construída de forma análoga à cA:DR1-A, porém com alinhamento sobre o monômero 3' (cadeia B) do PDB:IBY4 (cB:DR1).

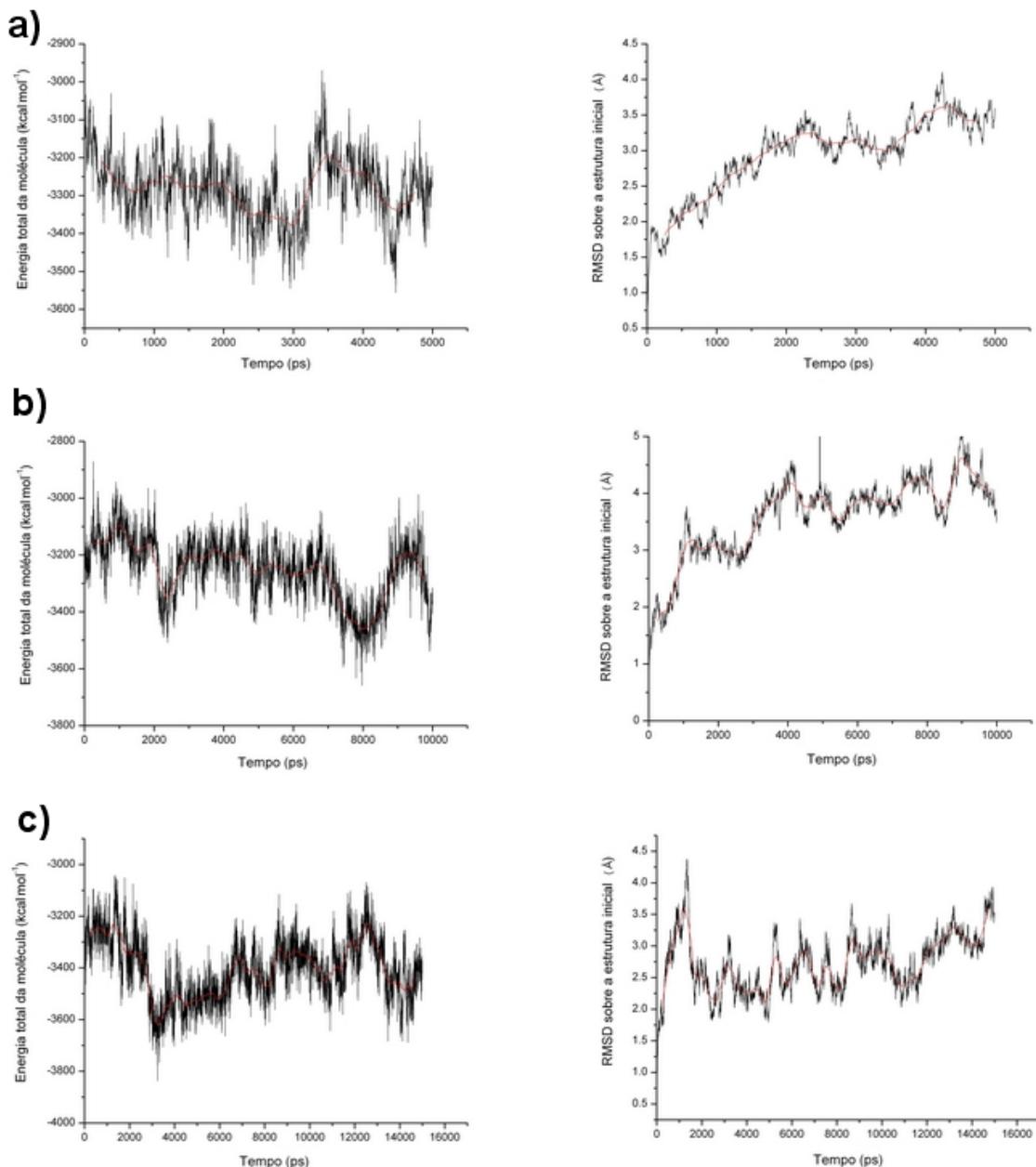
jetórias do complexo dimérico em DR1, para o qual tanto os perfis de progressão do rmsd quanto da energia potencial do sistema demonstram, respectivamente, a manutenção do equilíbrio conformacional e um alto favorecimento da energia potencial (correspondente à maior fração da componente entálpica para a energia livre) do sistema no equilíbrio. Esta fugacidade demonstrada para as trajetórias dos complexos monoméricos não aparenta ser devida a efeitos repulsivos ou atrativos, visto que a energia potencial do sistema permanece oscilando sobre um valor médio durante toda a simulação para os três respectivos modelos (Fig. 4.22-B,D e F), a despeito dos progressos contínuos dos valores de rmsd. Desta forma, pode-se inferir que a fugacidade demonstrada pelo progresso de rmsd é decorrente da termalização de um sistema cujas configurações iniciais são não permissivas para os modos de movimento interno induzidos ao mesmo (ou às suas partes) pela “injeção” da energia termo-cinética na componente estocástica da simulação, de forma a que estes mesmos movimentos térmicos tendem a afastar os respectivos sistemas de suas posições iniciais no espaço de fases (sugerindo uma causa entrópica para a referida fugacidade)<sup>61</sup>.

Nota-se que a fugacidade ilustrada pelos perfis de evolução do rmsd é mais acentuada, desde os primórdios da simulação, para as trajetórias dos complexos monoméricos com o sítio 5' que com o sítio 3', com perfis similares para ambas as trajetórias de complexos cA:DR1, com diferentes configurações iniciais. Para a simulação do complexo cB:DR1, os progressos de rmsd sugerem a adoção de uma configuração com relativa metaestabilidade entre o intervalo de 2000 ps (após a termalização inicial) e 12000 ps (após o que nota-se a manifestação de um evento que tende a afastar o complexo desta configuração metaestável).

Na Fig. 4.23, verifica-se que o evento que caracteriza a maior fugacidade a partir de 12000 ps na trajetória para o complexo com o meio-sítio 3' (cB:DR1) é o “escape” do T-box do sulco menor na interface com o sítio 5' devido aos movimentos vibracionais da cadeia de DNA e à competição com a interação com os íons Mg<sup>2+</sup> na interface DNA:solução. Na conformação inicial (condizente com a configuração destes elementos no complexo dimérico), os resíduos

---

<sup>61</sup> Segundo o modelo proposto, pode-se fazer uma analogia (qualitativa, visto a extensão da presente análise) entre tal fugacidade das dimensões globais da molécula relativas à configuração inicial, após serem retiradas as restrições harmônicas da pré-dinâmica (vide ), com o modelo extensamente utilizado em livros-textos de Físico-Química (como por exemplo em ) para a expansão livre de um gás previamente confinado.



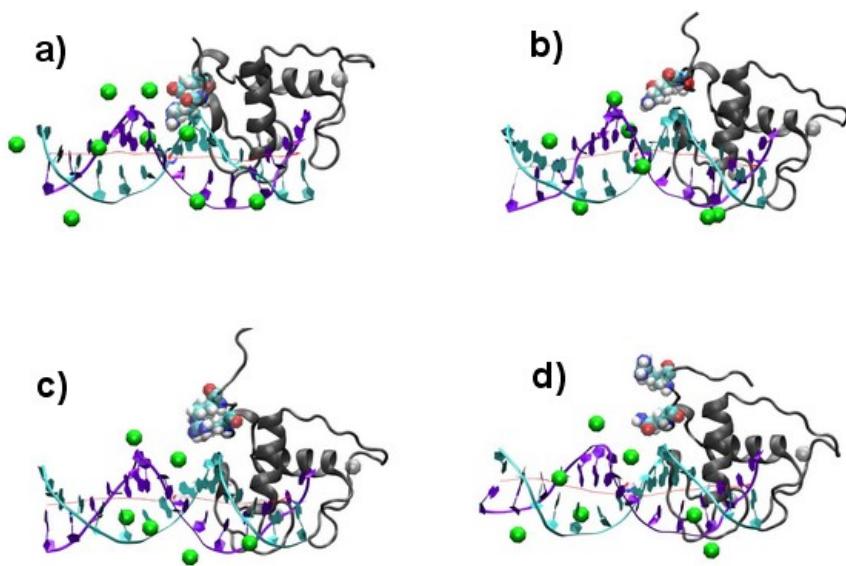
**Figura 4.22** – Progressão da trajetória de equilíbrio para um único monômero associado a DR1. *A* e *B* - trajetória cA:DR1-A. *C* e *D* - trajetória cA:DR1-B. *E* e *F* - trajetória cB:DR1.

Q206 e R209 se associam a uma posição no sulco menor próxima ao centro de massas do DNA (indicado na Fig. 4.23)<sup>62</sup>. Esta posição (conforme previamente mostrado nas análises da subseção 4.2.3) é foco tanto de interações bem localizadas com os íons em solução (em particular o Mg<sup>2+</sup>) quanto (e não por acaso, visto o acoplamento entre os dois fenômenos) de maiores modificações conformacionais decorrentes dos movimentos coletivos do DNA. Com o ajuste do T-box sobre esta região, tanto a mobilidade quanto a interação com os íons são atenuadas, proporcionando o comportamento metaestável evidente na Fig. 4.22-E. Para a interação com o sítio 5', o T-box só se ajusta ao sulco menor na posição flanqueadora (vide ilustrações das componentes na Fig. 4.24-A e B), apresentando, portanto, menores implicações quanto à deturpação das componentes vibracionais do segmento de DR1, previamente ilustradas na Fig. 4.20-A.

Os padrões de simetria vibracional para os respectivos complexos podem ser melhor visualizados através da análise das componentes principais, cujas estruturas nos extremos dos respectivos autovetores são ilustradas na Fig. 4.24 e cujos padrões associados de variações nas larguras do sulco menor ao longo da cadeia de DNA-dupla fita<sup>63</sup> são mostrados na Fig. 4.25. Para ambas as trajetórias, a ocupação de um único dos dois meio-sítios pelo monômero distorce as componentes originais do DNA, instituindo um padrão vibracional mais assimétrico. Contudo, esta menor simetria vibracional é mais saliente, desde o primeiro auto vetor, para as configurações representativas do complexo com o sítio 5' (Figs. 4.24-A e B e 4.25-A). Uma vez que a primeira componente original para o fragmento de DR1 livre em solução se mani-

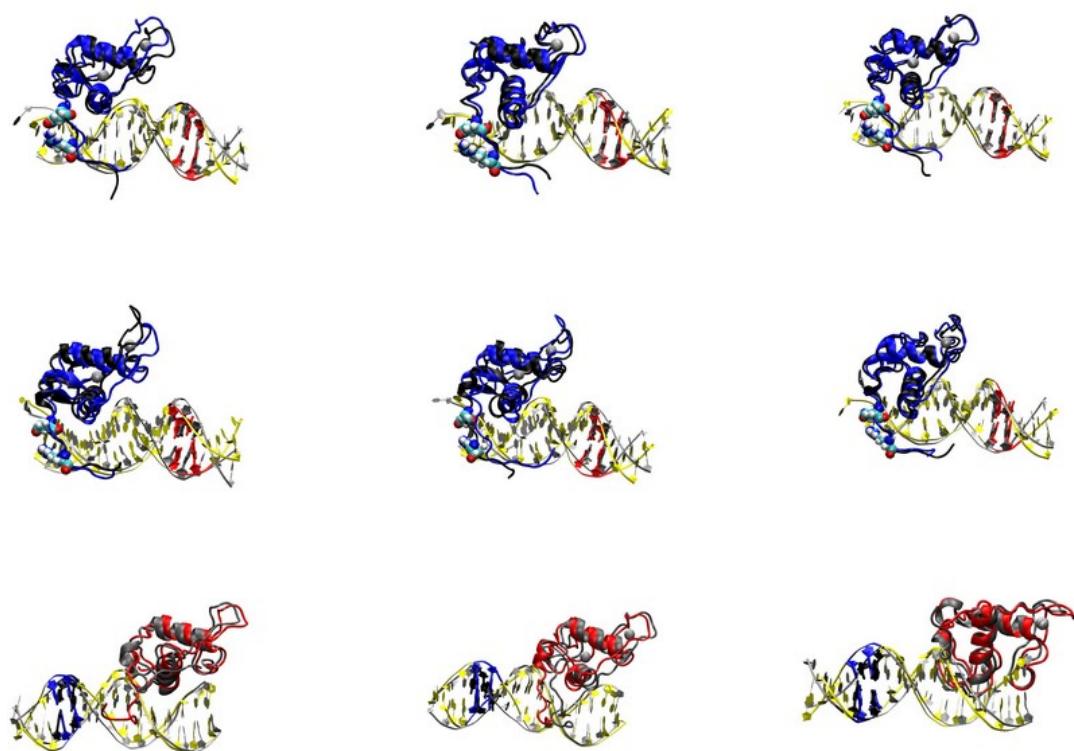
<sup>62</sup> Também é indicado na Fig. 4.23 o centro de massas do complexo como um todo, deslocado para as imediações da interface entre o monômero e o sítio 3'. Esta assimetria do complexo, com o centro de massas deslocado da porção mediana do DNA para uma posição mais próxima à sua extremidade (representativa do deslocamento de centro de massas para todos os complexos monoméricos com HREs contendo dois meio-sítios), com consequentes difusões rotacionais assimétricas em relação ao eixo principal da dupla hélice, bem como a elevada mobilidade da porção 5' (que aumenta, por sua vez, a secção cônea de rotação local do fluoróforo ancorado e esta extremidade nos experimentos de anisotropia de fluorescência da seção 4.2 (Vide ) explica os menores incrementos anisotrópicos observados na mesma seção para os complexos monoméricos aos HREs com dois meios sítios em comparação ao complexo com o oligonucleotídeo contendo apenas um (para o qual o centro de massas global é mais próximo ao centro geométrico e de massas da molécula de DNA, proporcionando assim maior simetria rotacional sobre o eixo da hélice e diminuindo a mobilidade local da porção 5').)

<sup>63</sup> No que concerne aos complexos cA:DR1, esta disposição em gráfico bidimensional dos padrões de simetria vibracional da molécula de DNA é mostrada apenas para o cA:DR1-B (com trajetória mais longa). Contudo, as ilustrações dos autovetores (e respectivos autovalores associados) correspondentes às duas trajetórias nas Figs. 4.24-A e B, bem como dos perfis de progressão dos valores de rmsd sobre a estrutura inicial e de energia potencial dos complexos ao longo das duas simulações nas Fig. 4.22-A-D, demonstra a similaridade dinâmica para ambas as configurações do complexo monomérico ao sítio 5'.



**Figura 4.23** – Movimento vibracional do DNA e competição com os íons em solução tende a deslocar o T-box 3' na ausência do monômero 5'.

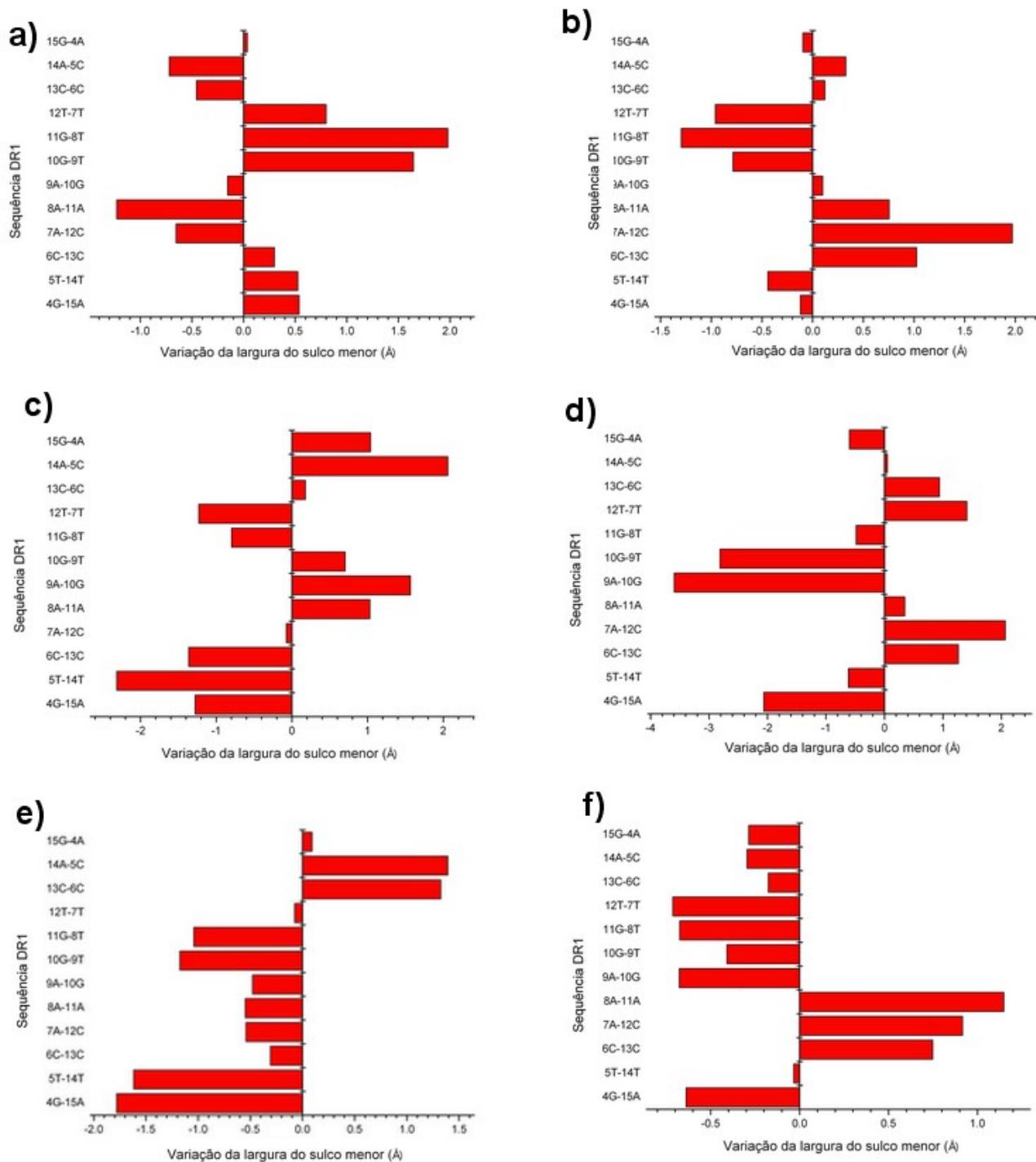
festa mais acentuadamente sobre a curvatura do sulco menor referente ao meio-sítio 5' (vide subseção 4.2.3), a abertura “forçosa” deste sulco pela interação com o monômero 5' provoca um acentuado deslocamento da energia vibracional da molécula para a extremidade 3' (setas nas Figs. 4.24-A e B e 4.25-A), se acumulando as maiores modificações no sulco menor sob o segundo meio-sítio. Este padrão é sugestivo de um acentuado custo no que se refere à entropia e entalpia vibracionais do complexo, visto que a energia vibracional é forçosamente acumulada em uma única extremidade (ocasionando vibrações com maiores freqüências e com um maior teor de interferências destrutivas, conforme pode ser constatado pela comparação dos padrões de abertura do sulco menor nas componentes para as trajetórias dos complexos cA:DR1 e para o fragmento original de DNA (Figs. 4.25-A e 4.20-B, respectivamente)). Além disto, as maiores modificações na conformação do DNA ocorrem contra a tendência ditada pelo empacotamento *nativo* entre os passos de base da molécula, visto a maior perturbação para o sulco menor na região dos passos -A-G-G-T do segundo meio-sítio (todos com alto valor de empacotamento hidrofóbico) que para o passo C-A, na extremidade 3' do primeiro (com empacotamento substancialmente menor).



**Figura 4.24** – Análise de componentes principais para os complexos de monômero a DR1. Esquerda para a direita: Componentes 1-3. A- trajetória CA:DR1-A; B- trajetória cA:DR1-B; C - trajetória cB:DR1.

Para o complexo monomérico com o meio sítio 3', nota-se uma primeira componente com maior simetria vibracional ao longo do HRE (Fig. 4.25). Isto ocorre devido à interação do T-box com o sulco menor na região espaçadora (próxima ao centro de massas da molécula de DNA), que introduz um “nó” nesta região central, de forma a melhor distribuir a vibração entre as duas extremidades da molécula (Fig. 4.24-B). Desta forma, a importância da associação do T-box com o sulco 5' na manutenção do complexo é reafirmada. Além disto a primeira componente original do fragmento de DR1 já é naturalmente deslocada para a porção 5' da molécula, de forma a que o padrão vibracional vislumbrado para a primeira componente do complexo cB:DR1 é mais semelhante à primeira componente do DNA livre, porém com um certo deslocamento de fase para a porção 5' devido ao posicionamento do monômero no sítio 3'. Já para as componentes de maior ordem deste complexo são vislumbrados padrões de assimetria vibracional semelhantes aos constatados para as trajetórias dos complexos cA:DR1. Em particular, a segunda componente do complexo cB:DR1 é mais assimétrica que a do complexo cA:DR1, com as maiores deturpações assimétricas no sulco menor localizadas justamente na interface do sítio 3' com a seqüência espaçadora, onde o T-box do monômero 3' se insere.

Desta forma, embora a interação de um único monômero com qualquer um dos dois meio-sítios de DR1 introduza deturpações na simetria vibracional do HRE, com o acúmulo de freqüências e amplitudes vibracionais em regiões localizadas da molécula em detrimento dos movimentos coletivos (o que, conforme previamente discutido em 4.2.3, introduz penalidades energéticas para a manutenção do complexo macro molecular), justificando assim a necessidade do parceiro de dimerização (vide 4.2.5); estas deturpações para a interação com o sítio 3' são introduzidas em modos mais internos que para o sítio 5' (acontando menos para as penalidades entrópicas e entálpicas vibracionais), corroborando mais uma vez a polaridade 3' de ligação do hRXR $\alpha$  a DR1 nos complexos com os diversos parceiros de interação e a capacidade do monômero se associar unicamente ao sítio 3' em experimentos de *gel shift* com um dos dois meio-sítios mutado. Estes diferentes padrões de deturpação de simetria vibracional explicam também a maior fugacidade demonstrada para a trajetória dos dois diferentes confôrmeros do complexo monomérico ao sítio 5' em comparação ao complexo com o sítio 3', conforme ilustrado na Fig.



**Figura 4.25** – Variações na largura do sulco menor ao longo da cadeia de DR1 nas componentes principais para as trajetórias com o monômero individualmente ligado a cada sítio. De cima para baixo: Componentes 1 a 3. A, C e D: Componentes para cA:DR1-B. B, D e E: Componentes para cB:DR1. Direção 5'-3' a partir da origem no eixo das ordenadas.

#### 4.24.

Por último (mas não menos importante) as ilustrações dos auto-vetores para as três trajetórias na Fig. 4.24 demonstram o acentuado acoplamento entre os movimentos do monômero no meio-sítio de ligação e a oscilação do meio-sítio disponível na extremidade oposta do HRE. Em particular, os movimentos de oscilação do T-box e movimentos com menor amplitude para o segmento 181-187, para o motivo  $\beta$ -hairpin e para a hélice 1 (alocada no sulco maior e interagindo especificamente com os pares de base da seqüência responsiva) são condizentes com a elevada flexibilidade para o primeiro e consideráveis taxas de troca de ambiente químico (sem a desassociação do DNA) para as demais regiões citadas, conforme vislumbrado para o complexo do monômero-DBD ao HRE nos experimentos de RMN de Tilborg et al . Como previamente comentado, esta facilidade de troca de interações na interface com o DNA (particularmente associada à desnaturação do T-box) é condizente com as frações de inespecificidade relativamente altas para as interações do monômero do hRXR $\alpha$  com diferentes HREs nos experimentos de anisotropia de fluorescência da seção 4.1. . Nota-se também que, dentre os movimentos de baixa amplitude do monômero, não são constatadas variações significativas na largura, extensão ou ângulo de rotação do D-box em relação ao cerne do DBD (devido à imobilização das bases do domínio sobre o sulco maior correspondente ao meio-sítio de associação) corroborando as evidências da maior rigidificação desta região no monômero DBD associado ao DNA, conforme sugerido pelos experimentos de RMN prévios da literatura .

Assim, encerrando as análises prévias dos elementos individuais do complexo do homodímero do hRXR $\alpha$ -DBD a HREs, concluimos que o mecanismo “chave e fechadura” na interação cooperativa, específica e com a polaridade correta ao elemento responsável adequadamente tem como barreiras energéticas iniciais (podendo estas serem consideradas as bases moleculares para a “fechadura”) três eventos principais:

- A transição do domínio DBD de uma conformação com maior grau de empacotamento e acoplamento internos para uma configuração em que suas partes individuais de interação com as duas regiões cruciais para a seletividade ao HRE específico (o meio-sítio e a

região 5' flanqueadora (onde, por sua vez, pode estar localizado o parceiro 5') ) estão devidamente disponibilizadas (sendo esta conformação, por sua vez, menos favorecida em solução);

- O deslocamento da camada de íons interagindo fortemente com o DNA na interface com a solução. Esta camada de íons por sua vez tende a interagir mais fortemente e de forma mais localizada (embora não *específica*) quando formada a partir de pequenos íons divalentes, com a razão massa/carga e raio atômico adequados para a alocação com significativa ocupância sobre as entradas dos sulcos maiores, onde estão localizados os meiosítios, e no sulco menor entre ambos. Dentre os íons com as propriedades físico-químicas adequadas, se destaca o Mg<sup>2+</sup> (cujo papel fisiológico na estabilização do DNA e, particularmente, na regulação da síntese de proteínas *in situ* é extensamente documentado ), condizendo com a alta eficiência do mesmo no desafio às interações inespecíficas e deslocamento da afinidade de interação do complexo a diferentes HREs em estudos experimentais da literatura e em nossos próprios (vide 4.1);
- As mudanças vibracionais do fragmento de HRE para um novo perfil de movimentos coletivos cujas simetrias sejam, pelo ao menos, tão favorecidas no complexo final quanto as originais o são para o HRE livre em solução. Os resultados de MD para o elemento DR1 livre em solução e em associação com o monômero para cada um dos dois meiosítios demonstraram em conjunto que esta mesma mudança de simetria presta um papel na indução da polaridade correta para o primeiro evento de ligação, visto que as mudanças dinâmicas provocadas pela interação do monômero na polaridade 3' (específica para o hRXR $\alpha$  em DR1) se mostraram menos dispendiosas que para a interação com o sítio do parceiro 5'<sup>64</sup>. As conformações médias e movimentos principais para as simulações com outros HREs livres (sobretudo DR4) sugerem que esta seletividade quanto à polaridade de interação ocorre também para os complexos envolvendo outros NRs e outros HREs que não só os complexos de hRXR $\alpha$  em DR1.

---

<sup>64</sup> Considerando, obviamente, a possibilidade de que este “parceiro” 5’ seja outro monômero do hRXR $\alpha$  (sendo este o caso que será analisado a partir de resultados de MD na próxima subseção )

Nas próximas subseções, passaremos para a discussão de como o parceiro de dimerização (“a chave”) estabiliza as mudanças conformacionais e dinâmicas para a correta manutenção da interação dos dois monômeros, cada um sobre seu respectivo sítio, com o HRE. Serão analisados tanto os mecanismos de como a “fechadura” correta (o HRE específico para o complexo dimérico em questão) se ajusta à esta mesma “chave”, quanto também as nuances que impedem a associação estável do complexo dimérico sobre as “fechaduras” não específicas; ambos eventos fundamentais para a seletividade nos processos de transcrição gênica controlados por NRs e para a consequente manutenção da homeostase.

#### **4.2.5 O complexo homodimérico do hRXR $\alpha$ a DR1 - A interação simultânea dos dois domínios DBD otimiza o acoplamento interno ao longo de todo o complexo, estabiliza as componentes vibracionais com uma maior simetria em relação ao centro de massas do sistema e apresenta componentes de movimento que parecem facilitar a correta interação entre os LBDs**

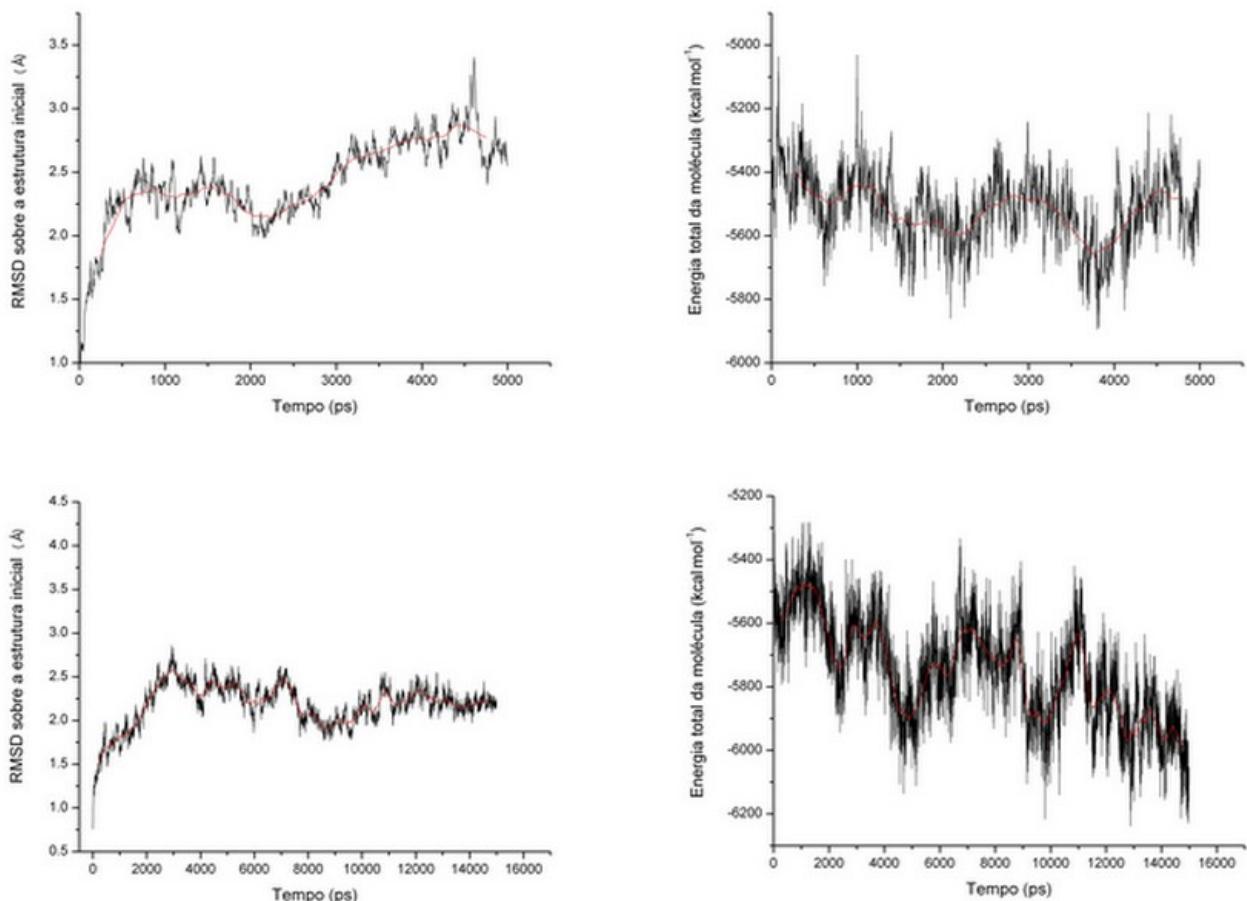
Foram realizadas duas simulações de MD para o complexo homodimérico de hRXR $\alpha$ -DBD a DR1, uma de 5000 ps e outra de 15000 ps (vide ), doravante referidas, respectivamente, como S-RXRd:DR1<sub>c</sub> (uma abreviação de *simulação curta para o dímero de RXR em DRI*) e S-RXRd:DR1<sub>l</sub> (*simulação longa para o dímero de RXR em DRI*). O desenvolvimento das simulações de equilíbrio para as duas respectivas trajetórias é mostrado através dos gráficos de progressão do rmsd em relação à estrutura inicial e de energia total para o sistema proteína-DNA na Fig. 4.26. A configuração inicial para a S-RXRd:DR1<sub>c</sub> é o modelo construído a partir das estruturas cristalográficas PDB:1BY4 e PDB:3DZU, de forma similar à utilizada na geração da maioria dos modelos utilizados para as diferentes simulações deste estudo (vide ). No PDB:1BY4, o fragmento de DR1 utilizado na geração da configuração inicial para a

simulação S-RXRd:DR1<sub>c</sub> apresenta o par de bases TA<sup>65</sup> em sua extremidade 5' envolvido em contatos promovidos pelo empacotamento cristalino. Estes contatos estabilizam o par de bases no complexo cristalográfico em uma orientação superempacotada que é pouco provável em solução, em se tratando de um passo T-A terminal<sup>66</sup> (visto que o o passo T-A apresenta o menor dos valores de empacotamento hidrofóbico listados na Tab. , com apenas -0.19 kcal·mol<sup>-1</sup>). Esta orientação superempacotada é contrária ao movimento induzido a esta extremidade 5' da seqüência pela principal componente verificada para o complexo em ambas as simulações (vide abaixo), além do que, o mesmo movimento é relacionado à interação da extremidade C-terminal do T-box do monômero 5' com o sulco menor, estando esta extremidade C-terminal ausente na construção original do domínio DBD utilizada para a cristalização do complexo . Assim, na simulação S-RXRd:DR1<sub>c</sub>, o par de bases na extremidade 5' se “abre” no decorrer da trajetória, introduzindo uma pequena, porém notória, mudança na configuração e dinâmica global do sistema que é bem visível no perfil de progressão do rmsd acima de 2000 ps (Fig. 4.26) e que demonstrou um certo impedimento para as orientações com maior estabilização do complexo ao longo da simulação. Desta forma, uma nova configuração inicial foi gerada a partir do *instantâneo* a 2000 ps (para o qual o par de bases na extremidade 5' se encontra melhor empacotado ao DNA, mas demonstrando também uma certa *adaptação* ao ambiente em solução (diferente da configuração cristalográfica inicial)), e teve suas velocidades gradativamente reiniciadas para a distribuição correspondente ao *ensemble* NPT com temperatura de 300 K e pressão de 1 atm, sendo esta a trajetória correspondente à simulação S-RXRd:DR1<sub>l</sub>, cujos perfis de progressão do rmsd relativo à configuração inicial e de energia total para as macromoléculas ao longo da simulação são ilustrados nas Figs. 4.26-C e D. Nota-se que a trajetória a partir desta nova configuração inicial, com o empacotamento do passo de bases 5' terminal relativamente mais

<sup>65</sup> Não confundir com a notação T-A utilizada neste trabalho para um *passo de bases T-A*. De forma a não deixar dúvidas: ao longo desta tese, a notação Nu<sup>SS</sup>Nu<sup>AS</sup> (onde Nu simboliza um nucleotídeo) é utilizada para denotar o *par de bases* formado entre os nucleotídeos Nu<sup>SS</sup> na fita senso e Nu<sup>AS</sup> na fita antisenso. Já a notação Nu<sup>5'</sup>-Nu<sup>3'</sup> é utilizada denotando o *passo* entre o par de bases que apresenta o nucleotídeo Nu<sup>5'</sup> na fita senso e o segundo par de bases na direção 3' ao primeiro, apresentando o nucleotídeo Nu<sup>3'</sup> também na fita senso. Em alguns casos, para uma *seqüência* de bases (fita única) ou de pares de bases (fita dupla) também é utilizada a notação Nu<sup>1</sup>Nu<sup>2</sup>Nu<sup>3</sup>... e assim sucessivamente, na ordenação comumente utilizada da extremidade 5' para a 3'. Na maioria das vezes, para este caso, é utilizada a notação 5'- Nu<sup>1</sup>Nu<sup>2</sup>Nu<sup>3</sup>...-3' e, quando não o for, estará bem evidente no contexto se tratar de uma seqüência.

<sup>66</sup> Neste caso sim, se referindo ao *passo* entre este par de bases inicial e o segundo.

estabilizado, demonstrou em contrapartida uma maior estabilização dinâmica (evidente no perfil de progressão do rmsd) e, sobretudo, um gradativo e acentuado decréscimo na energia potencial do sistema, condizente com uma maior otimização de suas interações internas.



**Figura 4.26** – Progressão do equilíbrio para as trajetórias do homodímero de hRXR $\alpha$  associado a DR1.  
Topo: trajetória S-RXRd:DR1 $c$ ; Fundo: trajetória S-RXR:DR1 $l$ .

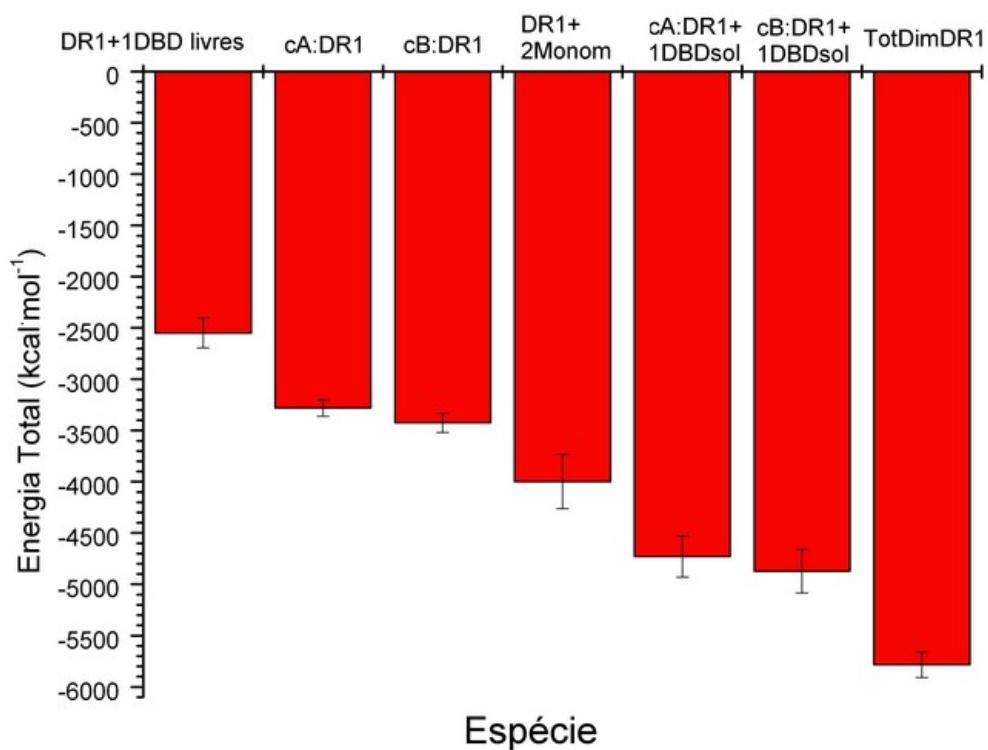
Uma estimativa quanto ao balanço entálpico na interação do dímero pode ser visualizada a partir da comparação dos valores médios e dos desvios para as energias totais<sup>67</sup> dos sistemas com diferentes graus de complexidade (Fig. 4.27)<sup>68</sup>. Nota-se que para duas espécies mo-

<sup>67</sup> Nesta análise, são considerados apenas os termos da energia total do sistema para as macromoléculas em solução (energias ligadas, conformacionais e não-ligadas internas e com o meio para a(s) molécula(s) de proteína e DNA).

<sup>68</sup> As médias e desvios padrões ilustrados na Fig. 4.27 foram tomados considerando-se cada respectiva trajetória a partir dos tempos de simulação em que a análise conjunta dos perfis de rmsd e de energia total sugerem terem sido ultrapassadas as primeiras etapas de termalização do sistema. Os intervalos de tempo correspondentes para cada simulação são listados na legenda da figura. Como previamente mencionado, esta abordagem é uma *estimativa*, visto as amostragens relativamente baixas, o fato de algumas das simulações analisadas ainda se encontrarem relativamente longe do equilíbrio (se destacando, dentre estas, a simulação para o confôrmero RXR $_{RMN}^{XX}$ ) e a não padronização da força iônica dentre os diferentes sistemas. Além disto, os termos entrópicos e as entalpias vibracionais (que, conforme as análises prévias e também as elaboradas nesta seção, apresentam colaborações cruciais) não são explicitamente considerados.

noméricas e uma única espécie do DR1, os valores de défice energético entre as duas combinações envolvendo um único monômero associado ao DNA e o outro livre (com o único monômero associado ao sítio 5' ou ao 3', respectivamente) são pouco significativos em face aos desvios. Considerando-se os custos energéticos esperados referentes às distorções nas simetrias vibrationais para a formação destes complexos monoméricos (sobretudo para a associação ao sítio 5', conforme inferido na última subseção), e que não são considerados na estimativa da Fig. 4.27, a manutenção de uma fração populacional prevalecente dos mesmos em solução se torna pouco provável. Um défice notoriamente mais significativo nas energias do sistema é verificado entre o complexo dimérico e as demais combinações dos dois DBDs e do DR1, ilustrando a colaboratividade para a associação simultânea dos dois monômeros ao elemento responsivo. Nota-se também que o valor absoluto da energia não ligada (termos coulômico e de vdw) entre os dois monômeros apresenta uma contribuição praticamente desprezível nesta colaboratividade (compare a pouca significância em face aos desvios da diferença de energia entre a combinação envolvendo os dois monômeros livres em solução + o DNA e a mesma combinação somada à energia de interação entre os monômeros), indicando que a intensidade das interações proteína-proteína não representam um fator *diretamente* correlacionado ao favorecimento para a interação simultânea dos dois DBDs ao elemento responsável (não sendo as energias entálpicas diretamente associadas a estas interações a causa aparente para a colaboratividade na interação do dímero).

A análise das componentes principais para as duas simulações permite um primeiro vislumbre quanto à natureza do ajuste superficial co-induzido e da consequente colaboratividade para a interação simultânea de ambos os monômeros ao DNA (Figs. 4.28 e 4.29). Os movimentos sutis descritos pela primeira componente para a simulação S-RXRd:DR1<sub>c</sub> sugerem uma tendência para a co-adequação do T-box do monômero 3' e da superfície do D-box do monômero 5' (Fig. 4.28-*Componente 1*). Percebe-se que a curvatura do D-box 5' toma a forma aproximada de uma ‘luva’ que acomoda o T-box 3' em sua posição intermediária, no sulco menor entre as duas seqüências. Para a formação desta aparente superfície de acomodação ao T-box 3', se fazem fundamentais a conformação “alargada” do D-box e a dobra no “topo” do



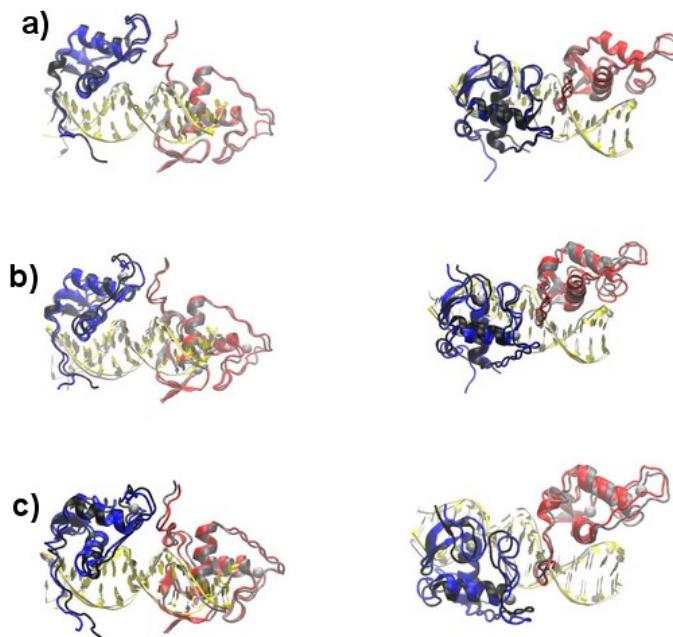
**Figura 4.27** – Estimativa para o balanço das componentes entálpicas entre as espécies de hRXR $\alpha$ -DBD e DR1 combinadas com diferentes graus de complexidade, conforme recuperado pelas simulações de Dinâmica Molecular.

segundo dedo de zinco do monômero 5', favorecida pelo empacotamento da K175 com a C190 na hélice B (sendo ambas as configurações, *i.e.*, tanto a maior extensão do D-box e o empacotamento entre os resíduos K175 e C190, simultaneamente facilitadas pela quebra dos acoplamentos dinâmicos com o T-box na porção C-terminal do mesmo domínio após a desnaturação da hélice C, conforme visto em 4.2.2). A face lateral 5' e a porção mediana desta mesma “luva” no D-box 5' para a acomodação do T-box 3' é proporcionada pelo *loop* formado pelo segmento 181-187, na conformação também estabilizada pela quebra dos acoplamentos dinâmicos e pela interação com o *backbone* na aresta do primeiro sulco maior que faz interface com o par de bases espaçador. A orientação vertical do T-box 3', por sua vez, é crucial para a co-interação deste segmento com o sulco menor e com a “luva” na superfície do D-box 5'. Esta orientação vertical é favorecida pelos contatos dos resíduos Q206 e R209 com o sulco menor (resíduos estes que na conformação nativa em hélice estão orientados para a face oposta, voltados para o cerne globular do domínio, conforme visto em 4.2.2), sobretudo pela alocação da cadeia lateral do resíduo R209 no interior deste sulco, na interface entre o par de bases espaçador (entre as duas seqüências arquetípicas) e o segundo meio-sítio. Além de sua projeção sobre o sulco menor, a cadeia lateral da R209 também se empacota com o resíduo Q185 na porção mediana “inferior” do D-box 5' (que por sua vez, também contacta o *backbone* do DNA na “aresta” 3' do primeiro sítio). Conforme será visto mais adiante, o empacotamento entre estes dois resíduos (análogo aos mesmos empacotamentos verificados nas estruturas cristalográficas dos heterodímeros com PPAR $\gamma$  e RAR $\alpha$  em DR1 (códigos do PDB 3DZU e 1DSZ, respectivamente)<sup>69</sup> proporcionam uma base para uma ampla rede de empacotamentos de vdw e de interações salinas favoráveis nas imediações do centro de massas do DNA e do complexo como um todo que contribuem

<sup>69</sup> Conforme previamente mencionado em 3.2 e no texto acima, a construção do domínio DBD utilizada previamente para a cristalização do complexo homodimérico em DR1 (42) (PDB:1BY4), é “truncada” a partir do resíduo R209 (se encontrando o mesmo na extremidade C-terminal). Desta forma, o T-box 3' não se estende, nesta estrutura, até os resíduos necessários para o estabelecimento da maior parte das interações observadas nas duas trajetórias de MD apresentadas neste estudo com a “luva” formada pelo D-box 5', se encontrando a R209 com acentuada mobilidade (com densidades eletrônicas baixas e relativamente difusas resolvidas para sua cadeia lateral e com toda a região do T-box em suas imediações apresentando valores de B-fator  $\geq 80 \text{ \AA}^2$ ) e, portanto, também não interagindo significativamente com o sulco menor. Ao longo desta seção, serão apresentadas evidências (tanto baseadas na análise conjunta de nossas diferentes simulações quanto a partir das comparações dos resultados de MD com as estruturas cristalográficas para o homo e para os diferentes heterodímeros de hRXR $\alpha$  em DR1 e com dados recentes da literatura sobre o acoplamento evolutivo dos resíduos na interface entre os DBDs (61)) de que as interações proteína-proteína e proteína-DNA no sulco menor observadas em nossas trajetórias são mais reflexivas para as de fato ocorrentes no complexo homodimérico intacto em solução.

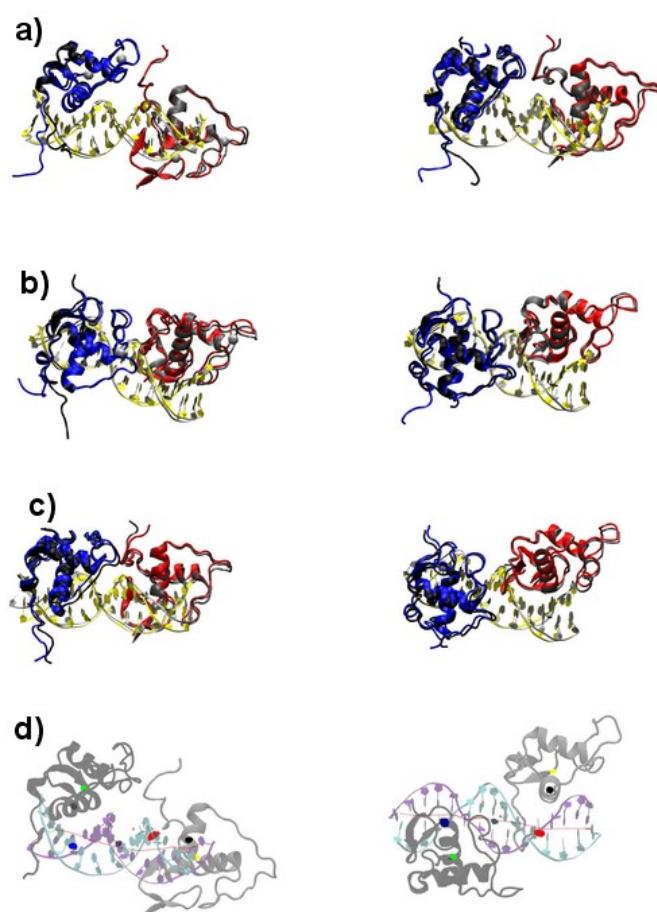
na estabilização das componentes vibracionais do sistema, induzindo a uma maior simetria nos movimentos coletivos do complexo. Além disto, a partir da análise das componentes principais para o complexo do monômero DBD com o sítio 3' na Fig. 4.24 da subseção 4.2.4 pode-se notar que, uma vez que a porção mediana do T-box 3' se encontra devidamente alocada sobre o sulco menor, a inclinação e posicionamento da porção C-terminal desta região de forma condizente com sua interação com o D-box 5', bem como o correto posicionamento da hélice A e do cerne globular do monômero 3' sobre o meio-sítio, são ambos facilitados por conformações em que o sulco menor na região central (onde o T-box 3' se introduz) se encontra com menor largura, o que por sua vez é facilitado pela aproximação do monômero 5' com o simultâneo ajuste dos dois monômeros a seus respectivos sítios. Uma vez que o T-box do monômero 3' e o D-box do monômero 5' estão devidamente auto-ajustados, com ambos os domínios imobilizados em suas posições corretas, o T-box do monômero 5', em sua conformação desnaturada, tende a se ajustar também ao sulco menor 5'. Na ilustração do autovetor referente à principal componente para a simulação S-RXRd:DR1<sub>c</sub>, nota-se que o posicionamento do T-box 5' sobre o sulco menor é correlacionado à correta orientação e empacotamento do par de bases 5' terminal que, conforme previamente mencionado, tendeu a se “abrir” ao longo da trajetória devido à configuração inicial previamente adaptada ao empacotamento no ambiente cristalográfico e, particularmente, a um evento de maior mobilidade regido pela componente estocástica da simulação.

Já para a trajetória S-RXRd:DR1<sub>l</sub>, oriunda de uma configuração em que o empacotamento para o par de bases 5' terminal se encontra previamente mais estabilizado, o autovetor relativo à componente principal demonstra ter sido alcançada uma maior estabilização das interações no empacotamento do T-box 3' sobre a “luva” formada pelo D-box 5' 4.29-*Componente 1*. Esta mudança aparentemente sutil de configurações apresenta implicações consideráveis quanto à estabilização do sistema. Com as interações otimizadas para a tríade monômero 5'-monômero 3'-sulco menor, nota-se que os movimentos descritos pela primeira componente são alocados, principalmente, sobre pequenas reconfigurações ao longo da topologia do DNA que terminam por otimizar o empacotamento global dos passos de base e, sobretudo, proporcionando um melhor ajuste às superfícies protéicas com consequentes interações mais favoráveis proteína-DNA.



**Figura 4.28** – Análise das componentes principais para o homodímero de hRXR $\alpha$ -DBD associado a DR1 na trajetória S-RXRd:DR1<sub>c</sub>

O sulco menor na região da interface monômero 5'-monômero 3' (nas imediações do centro de massas do HRE e do complexo como um todo) sofre um maior estreitamento, o que em contrapartida provoca um sutil maior alargamento do sulco menor nas direções de ambas as extremidades, de forma razoavelmente simétrica, proporcionando com que os dois-meios sítios nos sulcos maiores se adequem de forma otimizada à superfície dos respectivos monômeros. Na extremidade 5', este movimento facilita com que o T-box do monômero associado ao primeiro meio-sítio se aloque sobre o sulco menor. Esta melhor alocação do T-box 5', facilitada também pela maior imobilização do cerne globular deste monômero-DBD (estabilizado pelas interações otimizadas com o DNA e com o monômero 3') e, em particular, pela maior adaptabilidade a este movimento para o passo de bases T-A 5' nesta configuração, permite uma pequena semi-rotação destróbira e inclinação desta extremidade do oligonucleotídeo em direção ao sulco maior, proporcionando com que o *backbone* na face oposta do DNA se ajuste de forma ainda mais complementar à superfície do motivo  $\beta$ -*hairpin* e à hélice A do primerio monômero. Desta forma, nota-se que os movimentos induzidos de uma extremidade a outra do HRE pela primeira componente na S-RXRd:DR1<sub>l</sub> são associados a uma certa flexão do eixo principal na direção do sulco menor central, apresentando esta flexão uma simetria diédrica que é condizente com



**Figura 4.29** – Análise das componentes principais e flexão do eixo maior do DNA com maior ajuste superficial para o homodímero de hRXR $\alpha$ -DBD associado a DR1 na trajetória S-RXRd:DR1<sub>1</sub>

a simetria do complexo (4.29-D), o que culmina por sinergisticamente otimizar as interações proteína-DNA e o empacotamento entre os dois monômeros (vide abaixo), provendo o mecanismo de cooperatividade para a associação dos dois DBDs de hRXR $\alpha$  a DR1. O alinhamento dos centros de massas do DBD e de cada respectivo meio-sítio, bem como a orientação em ângulos dos eixos providos por estes alinhamentos e pelo alinhamento do centro de massas do DNA e do complexo como um todo, sugerem colaborar também na manutenção de uma considerável estabilidade na dinâmica interna e rotacional para o sistema (por sua vez, corroborando o acentuado incremento anisotrópico verificado para a associação da proteína na forma DL a DR1, conforme verificado nas subseções 4.1.2 e 4.1.3<sup>70</sup>).

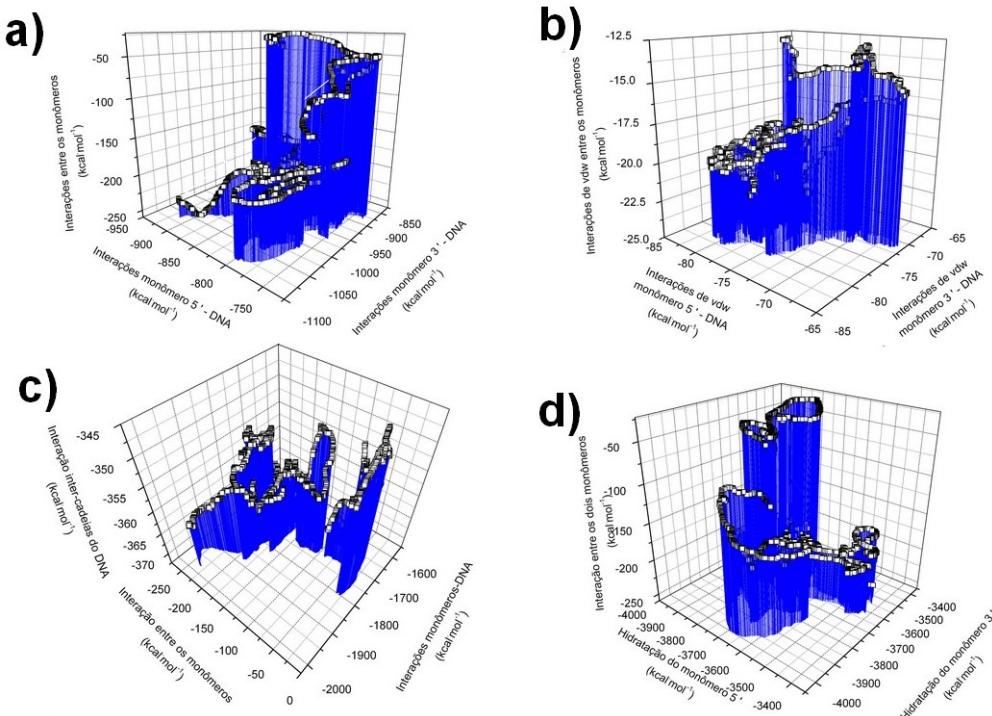
O significativo ajuste estereoquímico do complexo homodimérico a DR1, conforme evidenciado pelas componentes principais nas Figs. 4.28 e 4.29, é também refletido nas consideráveis correlações lineares para as energias de interação no sistema macromolecular (proteína-proteína, proteína-DNA, DNA intercadeias, e a anticorrelação com a hidratação dos monômeros, ilustrando a participação de interações hidrofóbicas na estabilização do complexo (vide abaixo)), conforme ilustrado nas superfícies de correlação da Fig. 4.30. Nota-se que as maiores correlações são obtidas entre as interações em seus termos dispersivos (vdw), demonstrando que o sistema se torna consideravelmente mais acoplado através do ajuste superficial (a distâncias mais curtas) com correlações relativamente menos substanciais para os termos a longas distâncias (interações coulômbicas). Este padrão, por sua vez, é condizente com o alto grau de inespecificidade para as interações não cooperativas e de maior especificidade para as cooperativas, conforme ilustrado nos experimentos prévios de anisotropia de fluorescência (4.1.2 e 4.1.3). A presença considerável de sítios carregados nos elementos constituintes do sistema (devido às altas densidades de cargas negativas decorrentes dos fosfatos no *backbone* do DNA e ao número considerável de resíduos com cargas positivas na superfície de interação da proteína ao DNA) permite com que, a distâncias relativamente mais longas, as superfícies dos respectivos monômeros interajam com os fosfatos do backbone (que não conferem especificidade à interação) de forma me-

---

<sup>70</sup> No decorrer desta subseção serão apresentados resultados referentes à análise das trajetórias de MD que sugerem ser este padrão de simetria diédrica e de colaboratividade também transferido à formação da interface entre os LBDs

nos correlata, sendo nescessários menores co-ajustes para a otimização destas interações. Para interações com maiores ajustes superficiais da proteína ao DNA, contudo, (caso em que são observadas as interações com as bases nitrogenadas encerradas nos sulcos maiores, apresentando portanto maior especidade) as interações para ambos os sítios se tornam mais significativamente acopladas, permitindo-se assim a manutenção, ou não, do complexo específico e colaborativo, conforme a interação ocorra ou não na presença do HRE e do parceiro específicos, respectivamente. Também percebe-se que o grau de correlação entre as interações inter-monômeros e as respectivas interações de cada monômero com o meio-sítio apresenta uma hierarquia invertida quando são considerados os termos coulômbicos e os de vdw, sendo a maior correlação para as interações coulômbicas observadas para o monômero 3', enquanto a maior correlação para os termos de vdw é observada para a interação do monômero 5' (Figs. 4.30-A e B). Isto corrobora a idéia (previamente sugerida na subseção 4.2.2) do T-box 3' como uma “alavanca molecular”, cuja interação simultânea com o sulco menor inter meio-sítios e com o monômero 5' permite o correto posicionamento do cerne globular do DBD-3' sobre a superfície do segundo sulco maior, de forma a acentuar suas interações com o mesmo sítio, sem se fazerem nescessários para tanto ajustes estereoquímicos significativos da superfície do segundo sulco maior ao monômero 3'. Já para a interação otimizada do DBD 5' ao primeiro meio-sítio, as análises prévias e as da presente subseção indicam serem nescessárias maiores modificações configuracionais do sistema, que só se tornam permissivas com um maior ajuste superficial entre ambos os monômeros e destes com o DNA. Para ambas as componentes das energias não ligadas, são recuperadas correlações consideravelmente mais significativas (sobretudo considerando-se as interações a longas distâncias) entre as interações inter-monômeros e as interações simultâneas de ambos os monômeros com o DNA que entre as respectivas interações de cada monômero ao DNA, indicando que os contatos entre os dois monômeros DBD são cruciais para a manutenção da correta colaboratividade no processo de associação dimérica a DR1.

Uma correlação linear de moderada para forte é também vislumbrada entre as interações inter-monoméricas e inter-cadeias do DNA, embora a mesma correlação não seja observada entre as interações dos respectivos monômeros ao DNA e as referidas interações inter-cadeias.



**Figura 4.30** – Superfícies de correlação entre as diferentes energias não ligadas envolvidas na manutenção do complexo homodimérico em DR1

Este perfil demonstra que, uma vez estando ambos os monômeros associados ao seus respectivos meio-sítios, as configurações adotadas pelo oligonucleotídeo em que ambos os monômeros interagem significativamente entre si (a flexão do eixo principal), são mais favoráveis para a dupla-fita. De fato, a inserção do T-box 3' sobre o sulco menor espaçador, de forma permissiva para a interação entre os dois monômeros, permite a configuração do DNA com o seu eixo principal flexionado na direção deste sulco menor central, o que por sua vez diminui a distância média entre as bases, sem no entanto se aproximarem significativamente os fosfatos do *backbone* (A aproximação dos quais confere uma penalidade eletrostática para a adoção de conformações desta natureza com representatividade significativa em solução (vide 4.2.3)).

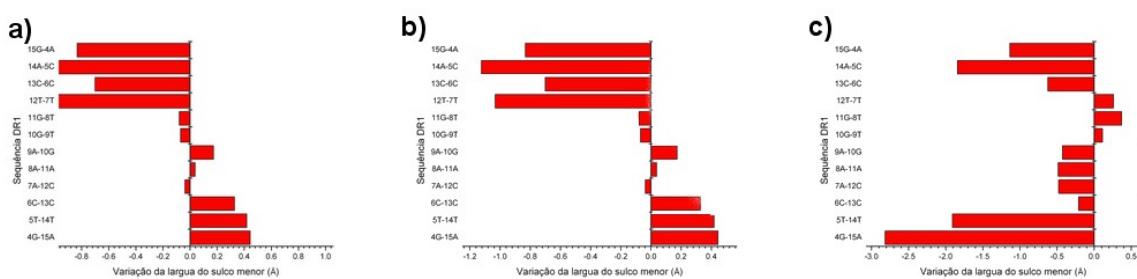
Por fim, a considerável correlação entre os termos de vdw para as interações no complexo dimérico do hRXR $\alpha$  em DR1 é refletida em uma correlação negativa entre as interações monômero-DNA e inter-monômeros no complexo e suas respectivas energias de hidratação<sup>71</sup> (Fig. 4.30-E-I), ilustrando a participação de interações hidrofóbicas na estabilização do com-

<sup>71</sup> O termo *energia de hidratação* é novamente empregado aqui (de forma não tão acurada, mas visando à simplificação) referindo-se à soma das interações com as moléculas de água dentro das distâncias referentes aos respectivos raios de corte para as interações coulombicas e de vdw, em analogia à utilização do mesmo termo em 4.2.2

plexo. Em simetria ao observado para as interações de vdw, a interação entre os dois monômeros é significativamente anticorrelacionada à hidratação do monômero 5', não o sendo para a hidratação do monômero 3', ilustrando novamente a maior dependência da interação do monômero 5' de um maior ajuste superficial, sendo necessário para tal o maior empacotamento entre os monômeros providenciado pela flexão do eixo principal.

A maior estabilidade alcançada para o complexo com o estabelecimento da flexão do eixo e da otimização das interações proteína-DNA e proteína-proteína decorrentes da primeira componente para a S-RXRd:DR1<sub>l</sub> é visível através da comparação das componentes de maior ordem para as simulações S-RXRd:DR1<sub>l</sub> e S-RXRd:DR1<sub>c</sub> nas Figs. 4.28 e 4.29 . Nota-se que, quanto as componentes 2 e 3 para a S-RXRd:DR1<sub>c</sub> ainda demonstrem significativas contribuições para movimentos de maior freqüência da proteína e DNA ao longo de todo o complexo, a segunda e terceira componentes para a S-RXRd:DR1<sub>l</sub> demonstram uma maior estabilização do complexo a partir de seu centro de massas, com os movimentos de maior significância mais deslocados para as extremidades do DNA e regiões mais móveis nas extremidades dos respectivos monômeros (Figs. 4.28 e 4.29 - *Componentes 2 e 3 e 4.31*). Em particular, a terceira componente da S-RXRd:DR1<sub>c</sub> descreve um movimento oscilatório do DNA sobre um autovetor similar ao descrito pela primeira componente do DR1 livre em solução (Fig. 4.20-A-*Componente 1*), demonstrando uma tendência de abertura do sulco menor (e fechamento do sulco maior) na direção 5' ao passo C-A na interface entre o primeiro meio-sítio e o par de bases espaçador, que, por motivos óbvios, confere uma considerável instabilidade (demonstrada pelos movimentos de alta freqüência) ao monômero no sítio 5', sendo transferida esta instabilidade também ao outro monômero em menor escala. Para a trajetória S-RXRd:DR1<sub>l</sub>, as interações otimizadas entre resíduos evolutivamente acoplados (61) do *loop* C-terminal do segundo dedo de zinco do monômero 5' e do T-box do monômero 3' (vide abaixo) estabilizam a região do passo C-A 5', não tendo sido evidenciadas tendências significativas para este mesmo movimento. A estabilização do centro de massas do sistema é refletida sobre uma maior simetria vibracional do fragmento de DNA a partir de seu centro, conforme pode ser inferido pelos padrões de variação na largura do sulco menor ao longo do oligonucleotídeo descritos pelas componentes principais

para a trajetória S-RXRd:DR1<sub>l</sub><sup>72</sup> (Fig. 4.31). É esperado (embora uma prova acurada exija uma análise mais elaborada, fugindo dos objetivos do presente estudo) que este padrão de maior simetria vibracional ao longo do centro de massas do sistema, envolvendo movimentos coletivos de menores freqüências ao longo de todo o complexo, contribua também para o défice de energia livre na associação do hRXR $\alpha$ -DBD dimérico a DR1 (Devido a um decréscimo nos custos entrópicos e, sobretudo, entálpicos vibracionais para a formação deste complexo, conforme discutido em 1.1). Além disto, este acentuado acoplamento energético e dinâmico ao longo do sistema sugere vantagens funcionais óbvias, no que concerne à troca de informações entre os dois sítios. Por exemplo, a informação dinâmica e conformacional decorrente da interação de uma molécula de co-ativador ou co-repressor a um único monômero e/ou a uma única extremidade do DNA pode ser transferida ao longo de todo o complexo, devido aos acoplamentos energéticos e dinâmicos ilustrados nas trajetórias deste estudo. Uma possível transferência de informação decorrente do acoplamento vibracional demonstrado pela primeira componente principal da trajetória S-RXR:DR1<sub>l</sub>, conforme será discutido ao final desta subseção, é que ela sugere um mecanismo de “guia” para a interação entre os LBDs na configuração descrita na estrutura cristalográfica do heterodímero de hPPAR $\gamma$ /hRXR $\alpha$  em DR1 (41).



**Figura 4.31** – Perfil simétrico em relação ao centro de massas das variações nas larguras do sulco menor descritas pelas principais componentes da simulação S-RXR:DR1<sub>l</sub>.

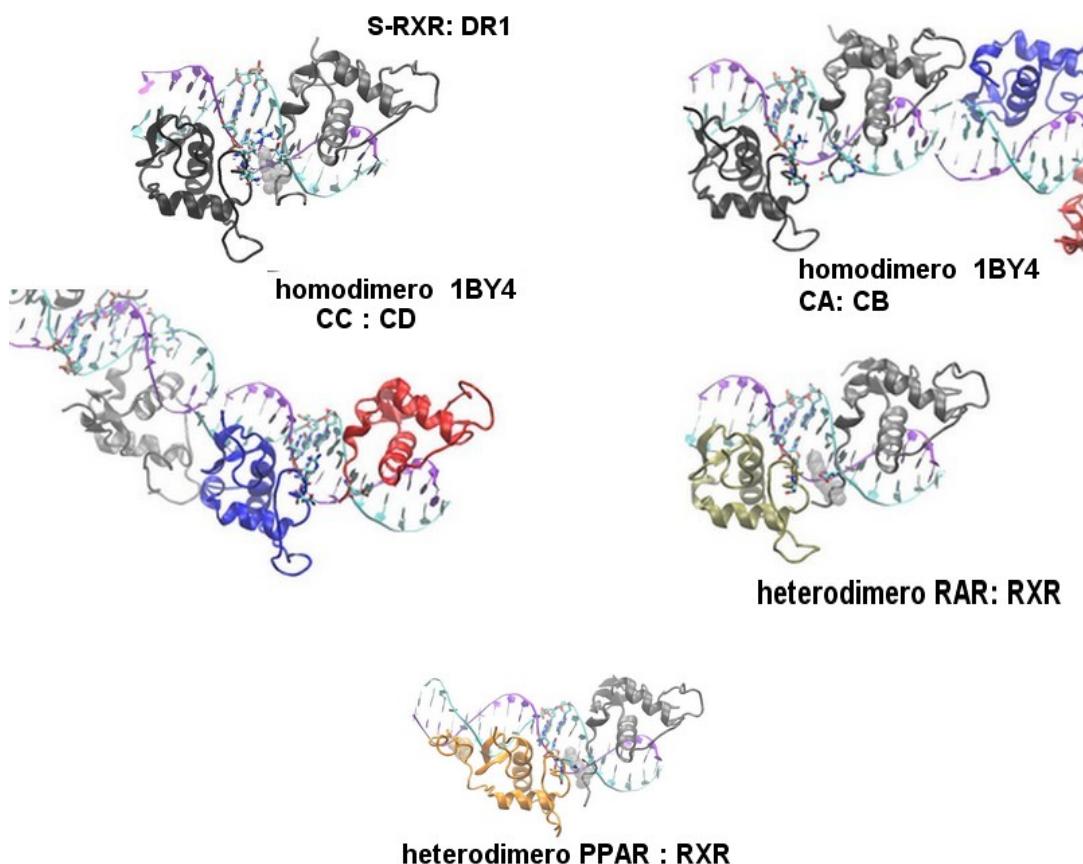
<sup>72</sup> a trajetória para a S-RXR:DR1<sub>c</sub> não foi considerada nesta análise vibracional, devido ao perfil evidentemente menos estável do sistema ao longo da mesma.

**Evidências de que a estabilização da primeira componente torcional é a chave para a seletividade do parceiro 5' em DR1, sendo este mecanismo também evidente nas estruturas cristalográficas dos heterodímeros com hPPAR $\gamma$  e hRAR $\alpha$**

Uma contribuição particularmente relevante para a estabilização dinâmica do complexo é a contraposição do movimento de abertura do sulco menor, ilustrado pela componente 1 do DR1 em solução (Fig. 4.20-A-*ComponenteI*) a partir da interação entre os dois monômeros. Quatro resíduos na interface de dimerização apresentam participação crucial na inibição deste movimento, três deles nas três respectivas posições de MSA identificadas em um estudo recentemente publicado (61) como sendo as que apresentam o maior grau de informação mútua (*Mutual Information - MI*) ao longo das cadeias dos DBDs e LBDs de todos os NRs com respectivas seqüências depositadas na base da dados Pfam (não tendo sido considerados neste alinhamento os domínios N-terminal e o domínio D, ambos com menores índices de homologia) - os resíduos R182 e Q183 no segmento C-terminal do segundo dedo de zinco e o resíduo 207 no T-box. O quarto resíduo, a R209 no T-box, se aloca sobre o sulco menor inter meio-sítios e, quando submetido à mutação R209A, reduz a afinidade na interação dimérica ao DNA (39). Na Fig. 4.32-A verifica-se que, em nossas simulações, a cadeia lateral da R209 se mantém preenchendo o espaço do sulco menor entre os dois meio-sítios, empacotando-se também, na face 5', contra a "base" do segundo dedo de zinco (particularmente, contra o resíduo Q185 (Fig. 4.33)). Este empacotamento, além de fornecer uma maior comunicação física entre os dois sítios do DNA, forma uma "alça" sobre o T-box, conduzindo a extremidade C-terminal a esta alça em uma orientação "vertical". Nesta orientação, a extremidade C-terminal se empacota contra resíduos na região em "luva" do D-box 5' dentre estes, a Q183, que ao longo da trajetória estabelece contatos alternados tanto com resíduos da extremidade N quanto da C-terminal à dobra ocasionada pela alocação da R209 no sulco menor<sup>73</sup>. Ja na extremidade N-terminal do T-box, com a imobilização da R209 no sulco menor, o resíduo E207 é posicionado de forma a otimizar sua interação com a R182 no D-box 5', formando uma ponte salina que "fecha" desta forma a in-

<sup>73</sup> Ao longo da trajetória, a Q183 faz contatos alternados com a cadeia lateral do resíduo D214 (extremidade C-terminal do T-box) e com a cadeia lateral do E207 (extremidade N-terminal de T-box). Os contatos com o E207 são efetuados sem perturbar a forte ponte salina entre as cadeias laterais deste resíduo e da R182, mantidos estáveis ao longo de toda a simulação (vide texto subsequente).

terface “superior” entre os dois monômeros, próximo ao centro de massas do sistema, enquanto o empacotamento da R209 3’ contra a Q185 5’ fecha a interface “inferior” (mais próxima ao DNA) nas imediações deste mesmo centro (Fig. 4.33).



**Figura 4.32** – Resíduos com alta informação mútua apresentam papel crucial na estabilização vibracional do complexo homodimérico em DR1.

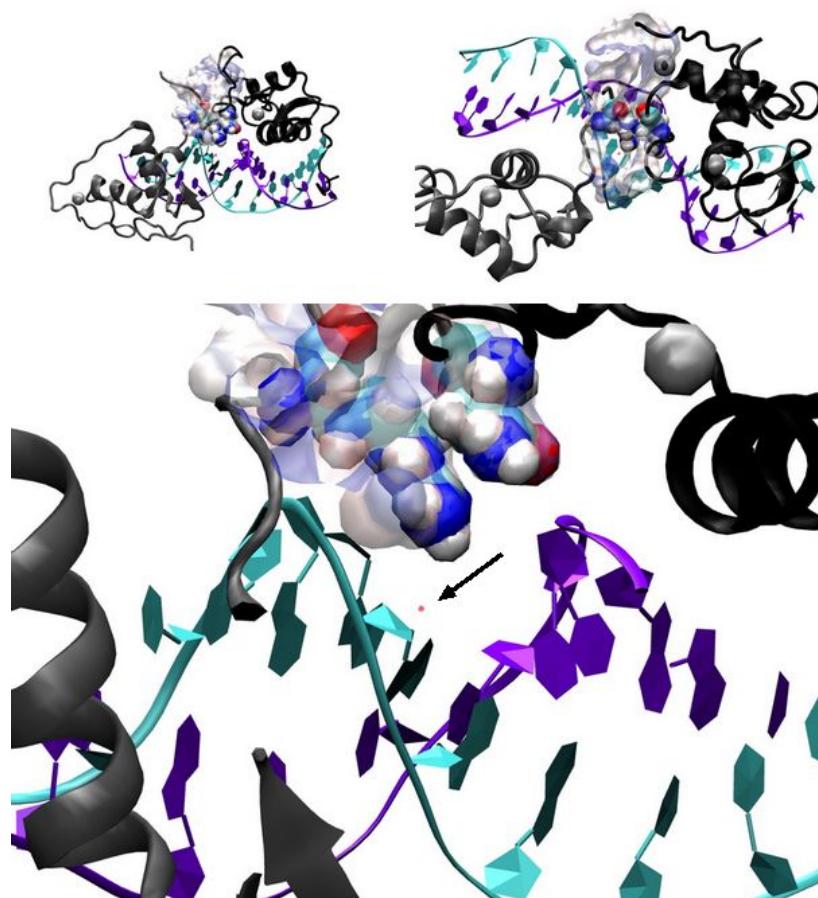
Na estrutura cristalográfica original do homodímero de hRXR $\alpha$  em DR1 (PDB:1BY4), a terminação da construção utilizada para o hRXR $\alpha$ -DBD justamente sobre a R209, conforme previamente mencionado, aumenta a mobilidade desta região, permanecendo a cadeia lateral deste resíduo (no dímero cristalográfico para o qual ela é claramente resolvida<sup>74</sup>) voltada para a face oposta e formando uma ponte salina com o resíduo D207 na mesma cadeia, estando assim ambos os resíduos impedidos de estabelecerem os respectivos contatos com o D-box 5’ obser-

<sup>74</sup> Como mencionado em 3.2, a estrutura cristalográfica com o código do PDB 1BY4 apresenta sua unidade assimétrica constituída por um dímero cristalográfico, com dois homodímeros de hRXR $\alpha$ -DBD associados *in frame*. As estruturas do oligonucleotídeo de DR1, bem como dos monômeros DBDs originais para a simulação S-RXRd:DR1<sub>rig</sub> (vide discussões subsequentes), utilizadas nestes estudos foram retiradas das cadeias E, F (DR1), A e B (monômeros), respectivamente, visto terem sido estas as melhor resolvidas no referido dímero cristalográfico.

vados nas simulações apresentadas neste estudo (Fig. 4.32-B e C). Os contatos estabelecidos pelos dois DBDs, desta forma, são ainda mais esparsos nesta construção, se limitando a um contato entre as cadeias laterais do E208 e da Q183 (contato este não vislumbrado ao longo de nossas simulações) e de uma ponte salina entre o grupo ácido carboxílico do C-terminal (que nesta construção é localizado no resíduo R209) com a cadeia lateral da R186. Este último contato por sua vez não é passível de ocorrer fisiologicamente, visto que na estrutura fisiológica do hRXR $\alpha$ , o C-terminal simplesmente não se encontra neste resíduo (sendo o mesmo seguido do domínio D e do LBD), não estando presentes, portanto, nem a carga e nem a liberdade conformacional que tornam esta ponte salina possível na construção cristalográfica. Além disto, nota-se nas estruturas dos homodímeros do PDB:*1BY4*, ilustradas na Fig. 4.32-B e C, que sem a acomodação do T-box 3' sobre a “luva” formada pelo D-box 5' (devido à ausência da extensão do T-box necessária para tal), o passo de bases C-A na extremidade 3' do primeiro meio-sítio se encontra menos empacotado e apresentando rotações e inclinações nas direções previamente vislumbradas para a primeira componente do DR1 em solução, o que por sua vez contribui para um sulco menor relativamente mais largo na mesma região, sugerindo uma menor estabilidade do complexo próximo a seu centro de massas.

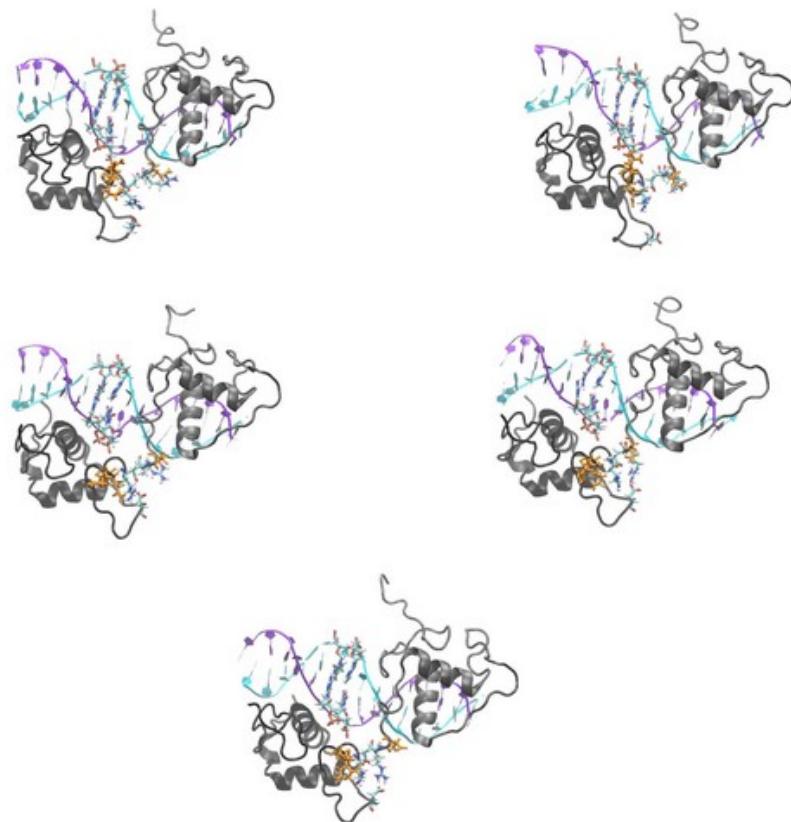
De forma a tirar a prova quanto à estabilidade do complexo na configuração dos DBDs para a estrutura *1BY4*, foi realizada uma simulação utilizando esta estrutura original, porém retirando a influência não fisiológica do grupamento carboxi-terminal na R209, não assinalando este resíduo como terminal no arquivo de estruturas da simulação (seguindo a mesma abordagem que para todas as demais estruturas geradas para as diferentes simulações deste estudo, de não assinalar os resíduos N e C terminais das construções incompletas utilizadas)<sup>75</sup>. A tra-

<sup>75</sup> Não assinalando o resíduo C ou N-terminal como tal na construção do arquivo de estruturas utilizando topologias do tipo CHARMM22 para todos os átomos de proteínas e ácidos nucléicos o resíduo é simplesmente *finalizado* com o grupo amida incompleto, sem a adição do último grupamento NH<sub>3</sub><sup>+</sup> ou O<sup>-</sup>, respectivamente. Desta forma o grupamento contendo a carga terminal não é adicionado e a cadeia principal remanescente do último resíduo permanece neutra, embora apresentando uma topologia “truncada” não condizente com a realidade. A abordagem é válida em uma aproximação de MD para um dado sistema macromolecular, desde que o foco da análise não se concentre nas interações e movimentos vibracionais nas imediações do referido resíduo e/ou de que o estudo em questão não vise a uma modelagem consideravelmente mais rigorosa de mecânica estatística (visto que os potenciais para diedros e ângulos na extremidade da cadeia principal do resíduo se encontrarão relativamente deturpados). Embora, no presente caso, a primeira condição tenha sido violada, uma parte fundamental da pergunta alvo era justamente o quanto a inibição das interações com o grupamento carboxi-terminal no resíduo R209 (não presente na molécula completa) seria deturpativo para a interface observada entre os dois monômeros na estrutura cristalográfica.



**Figura 4.33** – Empacotamento entre o T-box 3' e o D-box 5' estabiliza o complexo homodimérico em DR1 próximo a seu centro de massas.

jetória para esta configuração (doravante referida como S-RXRd:DR1<sub>orig</sub>) demonstrou uma fugacidade consideravelmente maior que a das trajetórias a partir das construções com o T-box completo (sobretudo comparada à trajetória S-RXRd:DR1<sub>l</sub>), conforme pode ser visualizado pelo perfil de progressão do rmsd em relação à estrutura incial na Fig. 4.34-A.



**Figura 4.34** – Trajetória para o homodímero hRXR $\alpha$ -DBD em DR1 na configuração cristalográfica original sugere uma significativa menor estabilização do complexo na ausência do empacotamento entre o T-box 3' e o D-box 5' devido à instabilidade do sulco menor no passo C-A na extremidade 3' do primeiro meio-sítio.

Nota-se que sem o contato da carga negativa provida pelo C-terminal com a R186 no D-box 5', o T-box 3' se torna ainda mais móvel e não mantém a configuração originalmente presente na estrutura cristalográfica, substituindo também a interação do E208 com a Q183 por uma interação com a cadeia lateral da R186. Também a R209 abandona a sua configuração original, voltada para a face oposta ao monômero 5' e passa a interagir com a cadeia lateral do resíduo D176 deste monômero 4.34-B. Além desta evidente mobilidade na interface, todo o sistema é consideravelmente menos estabilizado devido aos movimentos de “abre e fecha” do sulco menor na região do passo C-A terminal do meio sítio 5', que sem o maior empacotamento do T-box

3' ao D-box 5' presente nas trajetórias S-RXRd:DR1<sub>c</sub> e (sobretudo) RXRd:DR1<sub>I</sub> tende ao movimento de torção descrito pela componente 1 do oligonucleotídeo livre (previamente ilustrada na Fig. 4.20), instabilizando assim as interações de ambos os monômeros com o DNA e entre si (sobretudo o monômero 5', cuja hélice A, menos ajustada sobre a superfície móvel do primeiro sulco maior, chega a sofrer pequenos movimentos de semi-rotação, contração e extensão sobre a interface com o meio-sítio, diminuindo a interação com a superfície do DNA e perdendo contatos específicos). Finalmente, o “choque” entre estes movimentos que tendem a dobrar o sulco maior 5' contra a interface proteína-DNA alocada sobre o mesmo terminam por “abrir” o passo T-A 5' terminal (previamente envolvido em contatos cristalinos) de forma análoga à observada para a trajetória RXRd:DR1<sub>c</sub>. Nenhum dos três resíduos com altos índices de MI estabelecem contatos significativos entre os dois monômeros ao longo da simulação. Assim, a interface entre os monômeros na estrutura cristalográfica *1BY4* parece ser decorrente de um artefato devido à construção do DBD utilizada (que introduz uma interação coulombica a mais na interface, contribuindo para amenizar a alta instabilidade vibracional). Além disto, o maior empacotamento das extremidades da seqüência responsiva no dímero da unidade assimétrica devido aos contatos cristalinos contribui para estabilizar o DNA, o que provavelmente torna menos dispendiosa a associação dos monômeros no ambiente cristalográfico. Mesmo conservando a interação coulombica decorrente da extremidade carboxi-terminal, os valores de K<sub>D</sub> significativamente altos (entre 3.5 a 3.9 × 10<sup>-7</sup> M) recuperados por anisotropia de fluorescência para a interação dimérica desta construção ao oligonucleotídeo DR1 em solução no mesmo estudo (42) são sugestivos de que, na ausência da extensão C-terminal do T-box que permite o seu empacotamento contra o D-box 5' sobre o centro de massas do complexo (de forma análoga à vislumbrada em nossas simulações), o complexo dimérico dos DBDs ao elemento responsivo é pouco favorecido.

Por outro lado, as respectivas interfaces entre os dois DBDs nas estruturas cristalográficas tanto para o heterodímero completo com hPPAR $\gamma$  (PDB:3DZU), quanto para o heterodímero-DBD com hRAR $\alpha$  (PDB:1DSZ) (para o qual a construção utilizada para o hRXR $\alpha$ -DBD apresenta a seqüência do T-box prolongada até o resíduo R211) apresentam o empacotamento do

T-box 3' com o D-box 5' envolvendo resíduos análogos aos da trajetória “longa” para o homodímero de hRXR $\alpha$ -DBD em DR1 (Fig. 4.32-D e E). A alocação da cadeia lateral da arginina R209 do T-box 3' no sulco menor e seu empacotamento contra uma glutamina ou asparagina no D-box 5' co-interagindo com o *backbone* fosfato (Q185, N161 e N185, respectivamente, para o homodímero de hRXR $\alpha$  em nossas simulações e para os heterodímeros com hPPAR $\gamma$  e com hRAR $\alpha$ ) é visualizada para os três complexos. É interessante frisar que as respectivas densidades eletrônicas para a cadeia lateral da R209 nas estruturas cristalográficas para estes dois complexos heterodiméricos são substancialmente melhor resolvidas que para o complexo homodimérico na estrutura *IBY4*, sem a alocação da mesma sobre o sulco menor (4, 41, 42). Também para as estruturas cristalográficas dos dois heterodímeros em DR1, os três resíduos com maiores índices de MI no estudo de Griffin e colaboradores (61) se encontram em orientações relativas que sugerem o estabelecimento de contatos nas imediações do centro de massas do complexo, de forma análoga à observada para a simulação S-RXRd:DR1<sub>l</sub>. Em particular, os contatos entre os resíduos equivalentes às posições 139 (no D-box 5') e 195 (no T-box 3') para o MSA utilizado no referido estudo apresentam suas características físico-químicas consideravelmente mais bem conservadas entre o heterodímero com hPPAR $\gamma$  e o homodímero de hRXR $\alpha$  (ambos se tratando de complexos específicos para o HRE DR1) (Fig. 4.32-A e D); sendo que para o hPPAR $\gamma$  a posição 139 do MSA se trata de uma lisina (K157), enquanto que para o hRXR $\alpha$  esta mesma posição, conforme previamente mencionado, se trata da arginina R182 e, para ambos os casos, o resíduo na interface 3', equivalente à posição 195 no MSA, se trata do glutâmico E207 do hRXR $\alpha$ . Esta similaridade de complementaridade eletrostática vem a corroborar que a interface adotada no decorrer da simulação S-RXRd:DR1<sub>l</sub> (para a qual o complexo demonstrou maior estabilidade) é de fato a mais descriptiva para a interface acurada na manutenção do complexo dimérico dos DBDs em DR1. Para o hRAR $\alpha$  o resíduo correspondente à posição 139, a valina V182, é consideravelmente mais curto que seus homólogos em hRXR $\alpha$  e hPPAR $\gamma$  (Fig. 4.32-E), além de não apresentar a mesma complementaridade eletrostática ao E207. Desta forma, é esperado, *a priori*, um empacotamento relativamente menor sobre o centro de massas para o complexo deste heterodímero com DR1, em concordância com o contato menos pronunciado observado na estrutura cristalográfica. Também é visível uma

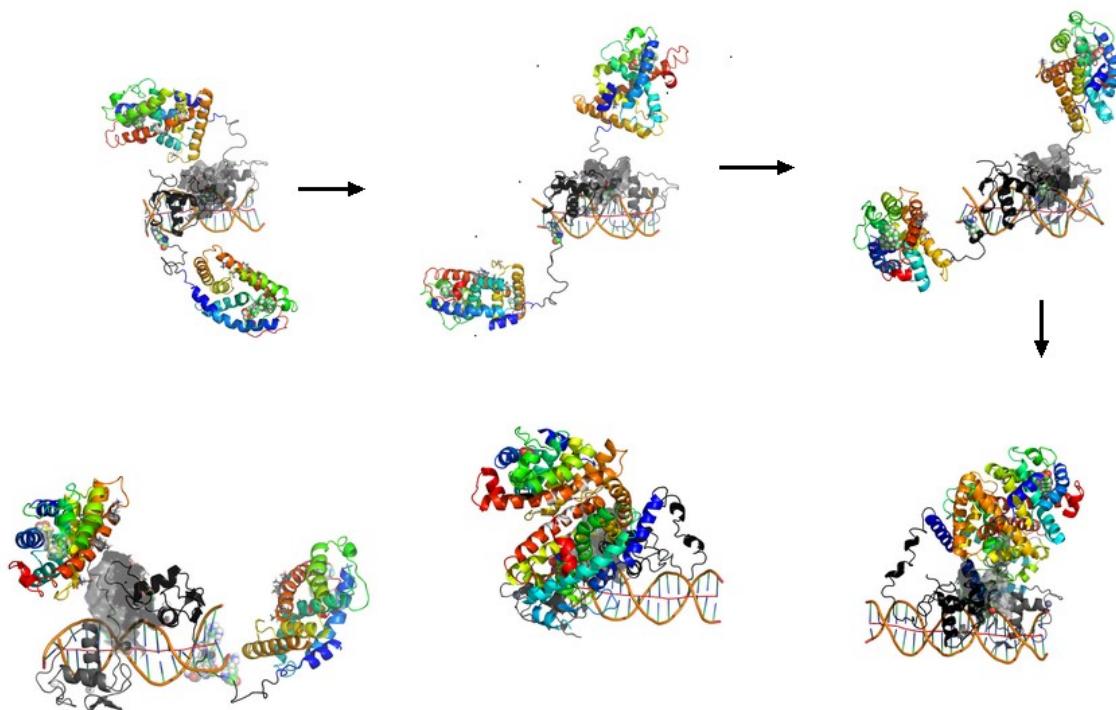
configuração relativamente mais larga para o sulco menor no passo C-A na extremidade 3' do primeiro meio-sítio para o mesmo complexo, novamente corroborando a correlação entre a intensidade do empacotamento envolvendo os resíduos de posição 139, 142 e 195 no MSA e a estabilização deste “defeito” aparente no empacotamento do meio sítio 5' próximo ao centro de massas do HRE. Este perfil não deixa de ser condizente com a menor responsividade do heterodímero hRAR $\alpha$ /hRXR $\alpha$  na transcrição gênica sobre DR1 (7)<sup>76</sup>. Para o hPPAR $\gamma$ , além do empacotamento dos resíduos com maior MI próximo ao centro de massas do sistema (empacotamento este que é relativamente menor que o do homodímero de hRXR $\alpha$ , visto a substituição da R182 por uma lisina, a K157, com menor cadeia lateral e não conferindo a típica configuração *bidentada* para a ponte salina de uma arginina com um resíduo ácido), o contato do T-box do PPAR com a seqüência flanqueadora 5', cuja presença é determinante na afinidade do heterodímero de hPPAR $\gamma$ /hRXR $\alpha$  as seqüências responsivas do PPAR $\gamma$  (PPRE) se posicionam de forma a se contrapor à primeira componente torcional de DR1, adicionando assim um segundo fator estabilizante para a correta topologia do primeiro sulco maior e para o contato com o monômero 5' (e, em consequência, conferindo maior estabilidade a todo o complexo). Este contato do T-box de PPAR $\gamma$  com o sulco menor 5'-flanqueador, por sua vez, é também mediado pelo resíduo na posição 195 do MSA para o hPPAR $\gamma$ , a fenilalanina F182, cuja cadeia lateral volumosa se aloca com significativa complementaridade estereoquímica no mesmo sulco menor, direcionando a rotação e inclinação do eixo principal na extremidade 5' de forma sugestiva tanto para a otimização da adequação superficial do primeiro sulco maior à superfície do DBD de hPPAR $\gamma$  quanto para a estabilização da primeira componente principal recuperada para o oligonucleotídeo DR1 em nossos estudos (Fig. 4.32 - D).

Uma nuance particularmente interessante para os movimentos de alta colaboratividade e simetria vibracional induzidos sinergisticamente ao sulco menor e à proteína na componente principal para a simulação S-RXR:DR1<sub>l</sub> (movimentos estes que contribuem para o maior ajuste simultâneo dos DBDs à superfície de DR1) é o movimento das extremidades dos T-box em uma

---

<sup>76</sup> As principais responsividades para o heterodímero RAR/RXR se dão para os HREs DR2 e DR5, respectivamente nas formas repressora e ativadora e com o RXR nas polaridades 3' e 5', sendo a responsividade em DR1 secundária.

configuração que, aparentemente, é diretiva para a interação dos LBDs nos padrões de simetria e alto empacotamento vislumbrado na estrutura cristalográfica do heterodímero completo com PPAR $\gamma$  (Fig. )



**Figura 4.35** – Aparente direcionamento para as interações entre os LBDs nos padrões de simetria da estrutura completa cristalográfica pela componente principal recuperada para a trajetória “estável” S-RXR:DR1<sub>L</sub>. As estruturas foram geradas sobrepondo-se as cadeias terminais do domínio D da cadeia A da estrutura 3DZU e os T-box de *instantâneos* com ordem crescente de projeção sobre a componente 1 desta trajetória. Na parte inferior da figura, a estrutura completa é ilustrada para fins de comparação.

No capítulo 6 serão discutidas as principais conclusões a respeito das análises aqui discutidas.



## 5 *Resultados e Discussões II - Estudos da dinâmica conformacional em solução da celobiohidrolase I de Trichoderma harzianum*

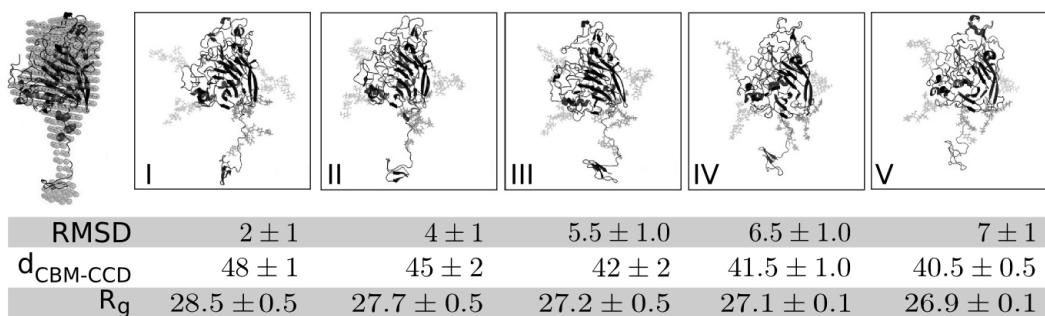
### **5.0.6 O modelo de alta resolução recuperado para a celobiohidrolase I apresenta o peptídeo conector dobrado em sua extensão N-terminal**

O modelo mais acurado recuperado pelo I-TASSER (e satisfazendo, simultaneamente, as restrições de distâncias entre os domínios ditadas pelo RBM) apresentou um valor de C-score de 0.47, sendo considerado, portanto, um modelo fidedigno segundo os critérios de estatística do programa 3.4.1.

O alinhamento pelo programa PyMOL das estruturas cristalográficas originais dos domínios CBM e CCD de *T. reesei* contra os homólogos destes domínios em nosso modelo mostra valores de rmsd relativamente baixos (0.893 Å para o alinhamento de 408 x 408 átomos no caso dos domínios CCD e 0.866 Å para o alinhamento de 32 x 32 átomos para os domínios CBM), confirmando a acurácia do protocolo na recuperação das estruturas tridimensionais destes dois domínios. A razão rmsd/(átomo alinhado) significativamente maior para o alinhamento dos CBMs comparado aos CCDs (0.0271 e 0.0022, respectivamente) se deve em parte à relativa maior flexibilidade e homologia sequencial ligeiramente mais baixa (relacionado à TrCel7A) para o primeiro. Contudo, a maior causa para esta discrepância é o posicionamento da porção N-terminal flexível do CBM (sequência interface entre o LP e o CBM) dobrada sobre o cerne deste domínio na estrutura por RMN para o domínio individualizado, enquanto que em nosso modelo para a enzima completa esta região é contínua à orientação do restante do LP em direção

à porção C-terminal do CCD.

O modelo recuperado confirma a forma de “girino” para a enzima, mas apresenta um peptídeo conector com a extensão N-terminal (resíduos 444 a 458; contendo quatro resíduos O-glicosilados - Thr453, Ser454, Thr455 e Thr456) dobrada sobre a superfície do CCD. A conformação inicial da molécula após a adição das cadeias glicosídicas e após o procedimento de minimização e equilíbrio mostra esta ”dobra” N-terminal estabilizada por interações com uma cavidade hidrofílica rasa na superfície do CCD (a partir de agora, referida como  $\text{CCD}_{\text{surf-pocket}}$ ), constituída pela face externa ”lateral” do motivo  $\beta$ -sandwich do domínio 5.1. Esta configuração dobrada para a porção N-terminal do LP é *Per se* consistente com as dimensões para um LP semi-flexionado sugeridas pela análise dos dados de SAXS.



**Figura 5.1** – Sobreposição dos modelos estruturais completos para a ThCel7A obtidos por modelagem computacional. O modelo estrutural obtido pelo I-Tasser utilizando as restrições topológicas do modelo DAM obtido pelo tratamento dos dados de SAXS é comparado ao mesmo na figura. Nota-se que a estrutura recuperada apresenta um peptídeo conector (o *linker*) semi-extendido. As conformações I a V, representativas dos *decoys* recuperados a partir das simulações de MD são mostradas. Os valores de RMSD médios computados em relação ao modelo inicialmente obtido pelo I-TASSER para cada *decoy* são mostrados.

Dentre as interações polares estabelecidas entre esta porção do LP e o  $\text{CCD}_{\text{surf-pocket}}$ , se destacam as do anel de manose extendido a partir da cadeia lateral do resíduo mono-O-manosilado Ser454. Os grupos polares deste anel interagem com as cadeias laterais dos resíduos Asp143, Thr302 e Asn 433 (todos estruturalmente conservados entre *T. reesei* e *T. harzianum*) no  $\text{CCD}_{\text{surf-pocket}}$  com significativa complementaridade estereoquímica à cavidade. Essas interações persistiram durante toda a simulação, a despeito dos ciclos de aquecimento e resfriamento do protocolo de SA/MD (vide 5.0.7). As extensões polissacarídicas das três treoninas (Thr453 - com dois resíduos de manose; Thr455 e Thr456 - ambas com três resíduos de

manose) permaneceram essencialmente voltadas para o solvente. A dobra do LP nesta região facilitou a simultânea interação destas cadeias tanto com o solvente quanto umas com as outras, contribuindo assim para a estabilização desta conformação local.

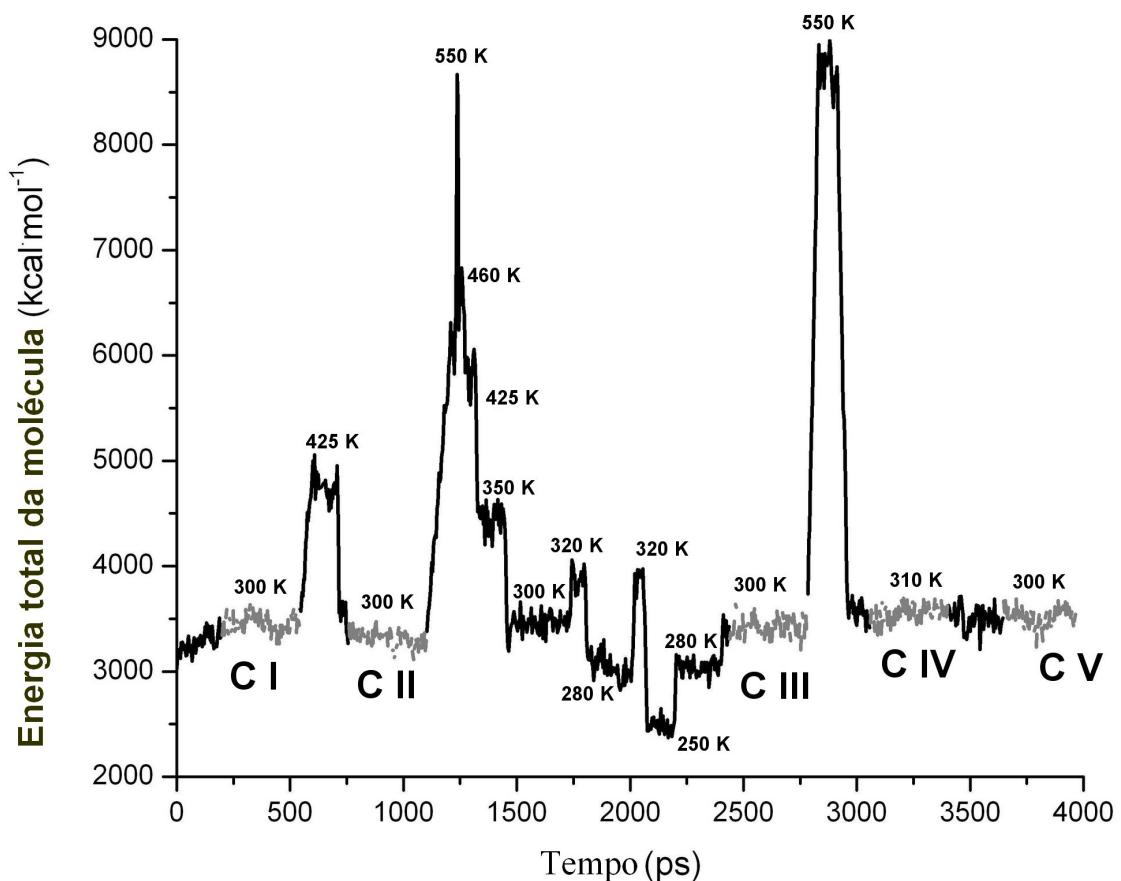
### **5.0.7 Cinco conformações principais, com diferentes extensões do peptídeo conector, foram recuperadas no procedimento de SA/MD**

O protocolo de SA/MD foi utilizado para amostrar diferentes conformações interdomínios para o modelo completo e glicosilado da celobiohidrolase I de *T. harzianum*. É esperado que a energia térmica adicionada à molécula (ou retirada da mesma), seja principalmente alocada para o movimento dos domínios em relação um ao outro, visto que tanto o CCD quanto o CBM são domínios globulares bem estruturados, possuem alto teor de pontes dissulfeto e são conectados pela longa cadeia polipeptídica em uma conformação extendida.

A Fig. 5.2 mostra as energias potenciais para a molécula ao longo da simulação, com conformações representativas da ThCel7A para cada etapa ilustradas sobre o gráfico. Já as Fig. mostra o rmsd sobre a média de todas as estruturas da simulação e o rmsd progressivo a cada 100 ps, respectivamente. Como pode ser visto, os principais passos de aquecimento até 550 K separam a simulação em três regiões principais de menor energia e mobilidade. A primeira região é subdividida em dois estágios, devido ao passo intermediário a 425 K.

Como esperado, a variação de energia termocinética se manifesta na molécula principalmente através de movimentos de extensão e contração do LP. Durante o decorrer dos passos de aquecimento e resfriamento, a molécula sofre consideráveis variações em suas dimensões globais, sempre relacionadas a variações nas configurações do LP (embora a porção C-terminal deste segmento permaneça sempre dobrada e interagindo com a face externa do motivo  $\beta$ -sandwich do CCD, como previamente mencionado).

As extensões máximas do LP ocorrem para os picos de energia e mobilidade relacionados aos passos maiores de aquecimento de 300-550 K. De forma similar, as menores extensões são acessadas pela primeira vez durante as etapas de resfriamento a temperaturas abaixo de 300 K (embora os baixos valores de Rg e D<sub>CBM-CCD</sub> tenham se mantido após o reaquecimento



**Figura 5.2** – Perfil de energia do anelamento simulado por dinâmica molecular utilizada na geração dos modelos estruturais para a ThCel7A completa. As temperaturas alvo para o termostato de Langevin em cada região são indicadas. Os modelos estruturais CI-CV foram extraídos a partir de regiões entre 300 a 310 K abrangendo intervalos de pelo ao menos 500 ps que são indicados em cinza.

a 300 K, indicando que restrições estruturais estabilizam estas conformações mais compactas como zonas metaestáveis ou mínimos locais na superfície de energia para a proteína (vide 5.0.8)). Este comportamento é condizente com um modelo simplificado para a organização da molécula de CBH-I como duas esferas (domínios CCD e CBM) conectadas por uma mola com amortecimento (LP).

Para uma análise mais acurada das energies envolvidas em diferentes configurações do LP, foi gerada uma trajetória “filtrada”, em que os intervalos da simulação contendo estruturas intermediárias aquecidas ou resfriadas foram removidos, permanecendo apenas as regiões da simulação que mais demonstraram estabilidade (ou metaestabilidade) e próximas a 300 K. Foram consideradas metaestáveis as configurações da simulação localizadas sobre aparentes platôs de energia e mobilidade.

A trajetória de MD ”filtrada” gerada desta forma continha cinco subconformeros, (Fig. 5.1), cada um dos quais uma média estrutural de uma diferente subtrajetória metaestável acessada pela ThCel7A durante o SA/MD, sendo todas as 5 estruturas médias para estas subconformações tomadas sobre mesmo intervalo de tempo (340 ps), referente a uma mesma quantidade de estruturas (85), de forma a contribuírem com a mesma amostragem para análises comparativas. Esta cinco principais subtrajetórias serão referidas, no texto que se segue, indiscriminadamente, como subtrajetória I-V, conformação I-V ou *decoy* I-V, denotando tanto cada uma das cinco divisões da trajetória “filtrada” de SA/MD, como as conformações médias representativas de cada subtrajetória. A numeração segue a ordem em que estas subtrajetórias aparecem na simulação original e também coloca as cinco conformações médias em ordem decrescente de extensão do LP e de magnitude de Rg em Å. As regiões do SA/MD original correspondentes a cada subtrajetória deste SA/MD “filtrado” são assinaladas na Fig. 5.2 e suas dimensões médias dispostas na Fig. 5.1.

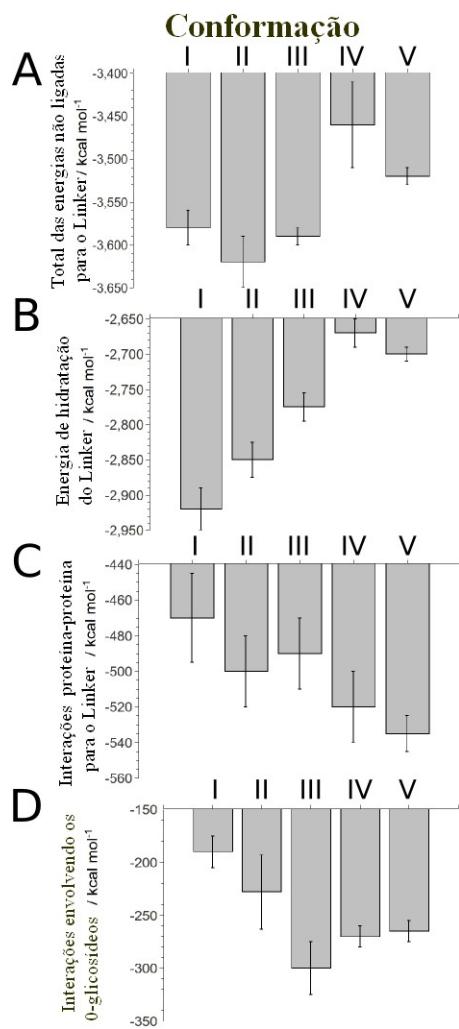
Além de provir uma maior acurácia na análise das energias (uma vez que apenas são consideradas conformações relativamente extáveis), é fácil notar, inspecionando as Figs. 5.1 e 5.2 , que as cinco conformações da trajetória “filtrada” de SA/MD fornecem um objeto de estudo mais diretamente analisável que o perfil “acidentado” da simulação original.

Embora sejam amostradas relativamente poucas configurações para cada subconformação, e embora as conformações III e IV provenham de trajetórias *a priori* fora do equilíbrio a 300 K (os primeiros 140 ps da subtrajetória III foram tomados dos passos finais de aquecimento de 280-300K e a subtrajetória IV foi amostrada a 310 K), todas estas conformações pertencem notoriamente a platôs de energia e mobilidade, com valores de temperatura próximos a 300 K e contêm valores similares de densidade e mobilidade do solvente. Desta forma, a comparação das energias entre as cinco no passo subsequente é uma abordagem razoável.

### **5.0.8 Hidratação e interações intramoleculares favorecem a conformação semi-extendida recuperada pelos estudos de SAXS**

Em uma análise subsequente, as componentes energéticas relacionadas a diferentes extensões do LP foram caracterizadas. Na Fig. o comportamento médio e os desvios destas componentes energéticas, bem como de sua soma, são potados contra a média temporal das dimensões de Rg e DCCD-CBM para a molécula. Como pode ser visto, os gráficos descrevem a energia relacionada às configurações do LP para as diferentes conformações acessadas durante o SA/MD e nos permite inferir sobre a natureza das restrições impostas para diferentes conformações do LP e distâncias entre os domínios. A soma das interações intramoleculares “não ligadas” e a interação com as moléculas de água fazem com que as configurações moderadamente extendidas do LP (entre as conformações I e III) sejam as mais favorecidas. Há um evidente mínimo de potencial com profundidade de  $\tilde{30}$  kcal·mol $^{-1}$  centrado a valores de D<sub>CDD-CBM</sub> e Rg próximos a 46.7 Å e 27.8 Å, respectivamente, correspondendo às subconformações moderadamente extendidas do confôrmero II. Este mínimo é mantido, aparentemente, devido a um balanço favorável entre as energias de hidratação para o LP e vizinhanças e as interações intramoleculares, se destacando entre estas as interações proteína-proteína, que apresentam um mínimo local para estas conformações.

Precedendo a conformação II na superfície de energia (valores de Rg entre 27.1 e 27.45 e valores de D<sub>CDD-CBM</sub> entre 41.5 e 44.25), há uma aparente zona metaestável, correspondendo principalmente às estruturas mais extendidas da conformação III (com um pequeno “degrau” de



**Figura 5.3** – Análise energética das conformações extraídas das simulações de MD. *a*- Soma das energias não ligadas para o LP e “vizinhanças”; *b*- energia de hidratação do LP; *c*-Interações proteína-proteína para o LP e vizinhanças; *d*- interações não ligadas envolvendo as extensões O-glicosídicas.

energia entre estas e as subconformações mais flexionadas da conformação II). A conformação III é amplamente favorecida por interações “não ligadas” intramoleculares dos O-glicosídeos do LP entre si e com a vizinhança

Conformações contraídas para valores de Rg e D<sub>CCD-CBM</sub> abaixo de 27.1 Å e 41.5 Å, respectivamente, foram sujeitas a um forte aumento no valor médio da energia do LP, devido principalmente a uma perda considerável na hidratação dos aminoácidos e glicosídeos deste segmento quando menos distendido e à perda de interações intramoleculares favoráveis envolvendo as O-glicosilações (Fig. 5.3-B e D). Esta última ocorre tanto devido à ausência de distâncias e ângulos que permitam a manutenção de interações de hidrogênio para estas cadeias quanto ao aparecimento de impedimentos estéricos para algumas cadeias umas com as outras e com a superfície da proteína, o que ocasiona uma contribuição positiva no termo dispersivo das interações não ligadas (Van der Walls). A subtrajetória IV é composta principalmente de estruturas pertencentes a um máximo local de energia devido a estes efeitos. A localização da conformação IV em um máximo de potencial é condizente com a subtrajetória aquecida da qual suas sub-estruturas foram amostradas (a subtrajetória IV foi “coletada” de um platô a 310 K (veja 5.0.7).

Embora a contração do LP contribua progressivamente para uma perda das moléculas de água nas camadas de solvatação deste segmento (principalmente no que se refere às camadas dos O-glicosídeos), ela também provê com que as camadas de hidratação dos carboidratos e da superfície protéicas gradualmente se misturem umas com as outras, “aprisionando” uma quantidade considerável de moléculas de água e criando regiões de alta densidade local de solvente próximo ao centro geométrico da molécula na superfície “inferior” do β-sandwich do CCD. Para o LP contraído a partir de uma certa extensão, as moléculas de água desta massa de menor mobilidade interagem mais favoravelmente umas com as outras e com a vizinhança de proteína e glicosídeos, o que adicionado às interações proteína-proteína mais favoráveis, provê um mínimo local centrado em D<sub>CCD-CBM</sub> de 40.5 Å e Rg de aproximadamente 27.0 Å, o que provê a energia média de hidratação relativamente menor para o confôrmero V comparado ao IV (Fig. 5.3-B. Para conformações ainda mais contraídas que a V, é esperado que prevaleçam

os efeitos de impedimento estérico, tanto para as interações “não ligadas” intramoleculares quanto para as camadas de hidratação imobilizadas, “amortecendo” as interações entre os dois domínios globulares, como previamente inferido em (53) nos estudos de simulação para a Tr-Cel7A. Além do mais, este “aprisionamento” de moléculas de água com baixa mobilidade para conformações super contraídas deve adicionar um custo entrópico para o sistema.

Para conformações mais extendidas que as do mínimo da conformação II, há uma região “ascendente” com baixa inclinação para as energias do LP, correspondendo às estruturas da subtrajetória I, com valores de  $R_g$  e  $D_{CCD-CBM}$  variando de 28.1 a 28.3 Å e 46.5 a 49 Å, respectivamente. A conformação I é ligeiramente mais favorecida pela hidratação média dos aminoácidos e O-glicosilações do LP que a conformação II (Fig. 5.3)-B, sendo no entanto menos favorecida pelas interações “não ligadas” intramoleculares (Fig. 5.3- C e D).

Dentre as energias para as interações “não ligadas” consideradas, as interações com os íons em solução parecem ser as menos determinantes para as diferentes extensões do LP acessadas na simulação, não apresentando efeitos significativos nas restrições conformacionais para a proteína. Isto é esperado, visto não haverem regiões contendo grandes densidades de carga nas vizinhanças do LP em nosso modelo. Embora não sejam mostradas as contribuições das energias conformacionais para a proteína e carboidratos, bem como as interações “não ligadas” intra-cadeia para as extensões lineares contendo duas ou três manoses para os resíduos de treonina do LP, estas energias se mostraram pouco ou não relevantes para diferentes configurações do LP, não sendo portanto apresentadas ou discutidas neste trabalho.

### **5.0.9 Os parâmetros médios de distância retornados pela simulação de SA/MD são condizentes com os parâmetros obtidos nos estudos de SAXS**

Uma vez obtido o mapa das energias relacionadas a conformações com diferentes valores de  $D_{CCD-CBM}$  e  $R_g$ , pode-se calcular os valores pesados pelo fator de Boltzmann para estas duas variáveis a 300 K (considerando apenas estas energias “locais” como determinantes para as diferentes conformações acessadas e não considerando o efeito das energias cinéticas. Como será

visto por uma abordagem semiquantitativa na seção 5.0.12 os fatores cinéticos aparentemente favorecem o mesmo conjunto de conformações que as energias potenciais aqui computadas devido a efeitos entrópicos). Os valores médios ponderados pelo fator de Boltzmann podem ser obtidos a partir da equação:

$$\Sigma(((Q_i)\exp(-E_i/kT))/(\Sigma(\exp(-E_i/kT))))$$

Onde  $Q_i$  é o valor da variável coletada Q (que aqui pode ser substituída por  $D_{CCD-CBM}$  ou  $Rg$ ) para a  $i$ -ésima das N estruturas presentes na trajetória “filtrada” do SA/MD,  $E_i$  é a energia relacionada à extensão do LP nesta  $i$ -ésima estrutura, k é a constante de Boltzmann e T é a temperatura do sistema em K (aqui, assumida como 300 K).

Os valores obtidos para  $\langle Rg \rangle$  e  $\langle D_{CCD-CBM} \rangle$  foram respectivamente 27.8 Å e 46.4 Å, ambos correspondentes ao mínimo de potencial para a conformação II. Este valor de  $\langle Rg \rangle$  é significativamente próximo ao valor de  $Rg$  retornado pela análise de Guinier e o DAM baseados nos dados experimentais ( $27.56 \pm 0.50$  e 30.46 Å, respectivamente) e o valor de  $\langle D_{CCD-CBM} \rangle$  é condizente com a distância de 49.5 Å medida para o RBM com os dois domínios globulares.

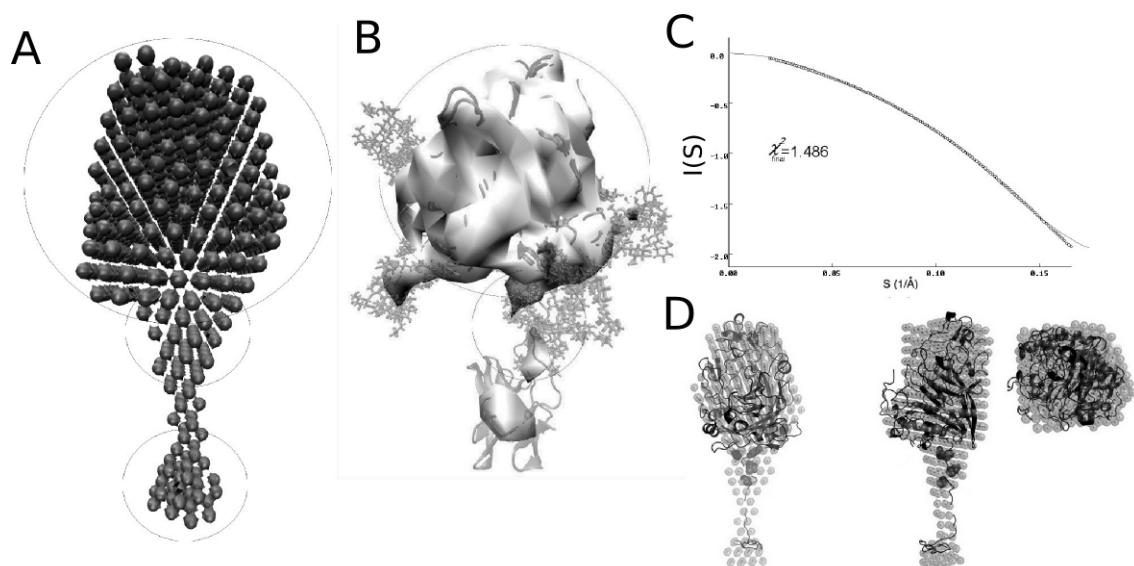
As dimensões máximas para a molécula a 300 K foram acessadas no SA/MD para a subtrajetória da conformação I, com um valor médio de  $D_{Max}$  de  $110.50 \pm 1.58$ . Este valor é condizente com os valores de  $D_{Max}$  recuperados a partir da transformada de Fourier dos dados experimentais na análise de  $P(r)$  e para o DAM ( $110.00 \pm 0.50$  e 110.70 Å, respectivamente). Valores de  $D_{Max}$  significativamente maiores apenas foram alcançados na simulação para os dois picos de aquecimento a 550 K (116.79 e 118.27 Å), retornando prontamente a valores menores durante o resfriamento. Nossa análise energética sugere que conformações significativamente mais extendidas que as da subtrajetória I não são atingidas com estatística significativa no equilíbrio a 300 K (elas não são acessadas nos decoys recuperados a partir do SA/MD).

Desta forma, as dimensões moleculares recuperadas na análise da trajetória “filtrada” do SA/MD corroboraram os dados experimentais.

### 5.0.10 Apenas as regiões de maior ocupância ao longo do tempo para a ThCel7A são claramente identificadas no modelo de átomos dummy retornado pelo SAXS

Os perfis “achatados” para os gráficos das somas das energias entre as conformações I e III representados na Fig. sugerem que uma única conformação estática dificilmente pode representar o comportamento da ThCel7A em solução. Desta forma, nós utilizamos os padrões de mobilidade acessados para as diferentes regiões da ThCel7A em nossa simulação para melhor interpretar os resultados de SAXS.

Pode ser visto que as extensões dos carboidratos são, aparentemente, pouco representadas pelo DAM recuperado no passo de modelagem *ab initio* dos dados de SAXS para a proteína completa (Fig. 5.4).



**Figura 5.4** – Refinamento do modelo estrutural utilizando anelamento simulado por dinâmica molecular. *A*- Modelo dos átomos *Dummy* (DAM) obtido a partir dos dados experimentais. *B*- Mapa volumétrico promediado sobre o tempo e pesado pelas massas para os átomos de todas as estruturas da simulação com uma *grid* de resolução de 5 Å, sobreposto aos cinco confôrmeros médios recuperados pelo SA/MD. Nota-se a representatividade difusa para as O-glicosilações. *C* Reprodução do padrão de espalhamento de raios X a baixos ângulos calculado para um modelo estrutural a partir de uma média ponderada para as conformações de I a III. *D* Sobreposição do melhor modelo de alta resolução recuperado ao envelope do DAM.

As N-glicosilações do CCD são, aparentemente, as menos resolvidas no modelo. Uma possível causa é a alta mobilidade para estes grupos em solução. Devido ao posicionamento

e orientações randômicas para partículas em solução, a intensidade de espalhamento ( $I$ ) monitorada em função do ângulo é proporcional à média ponderada do espalhamento de uma única partícula para todas as orientações acessadas (62). Para um sistema como a ThCel7A, com alta mobilidade interna, é razoável assumir que esta média ponderada produza um perfil de espalhamento que reflita principalmente as regiões da molécula com maior densidade de ocupância atômica, perdendo as informações relacionadas às fases mais externalizadas e de alta mobilidade (como as glicosilações).

Na Fig. 5.4, o DAM recuperado para os dados de SAXS é comparado a um mapa volumétrico (mostrado em estilo *volume* e consistência opaca) para a densidade de massa para a ThCel7A ponderada sobre todas as conformações acessadas para a simulação “filtrada”, com resolução de grade de 5 Å. Esta resolução é usualmente tomada como limite para a informação estrutural útil passível de ser extraída de uma análise de SAXS (62). Também é mostrada na Fig. 5.4 uma sobreposição das estruturas médias para as cinco subtrajetórias da simulação “filtrada” de SA/MD (em estilos *cartoon* e *stick*, para a cadeia principal da proteína e para as glicosilações, respectivamente, e consistência transparente). Pode-se notar uma alta dispersão para as estruturas superpostas das N-glicosilações do CCD, para a maior parte das extensões O-glicosídicas do LP, para a maior parte do próprio LP e para o CBM, devido à alta mobilidade destas regiões da molécula. Isto faz com que estas regiões não sejam claramente representadas no mapa de densidade média. Também o *loop* contendo os resíduos 414-426 na entrada do túnel catalítico e altamente móvel não é envolvido pelo mapa de densidade na figura.

Comparando as duas imagens na Fig., 5.4 pode-se ver que o DAM obtido através do SAXS, aparentemente, reproduz as regiões da molécula com maior ocupância de massa atômica:

- A maior parte do CCD, globular e rígido;
  
  
  
- A porção média do LP, logo “abaixo” do CCD. Durante a simulação, esta região é amplamente ocupada por diferentes configurações do LP e por extensões de O-glicosídeos (vide 5.0.11;

- O centro de massa do CBM.

Desta forma, o DAM pode ser considerado uma aproximação sobre a distribuição geral do “empacotamento” para as diferentes regiões da ThCel7A.

Um outro provável motivo para a não resolução das N-glicosilações é a distribuição heterogênea da extensão destas cadeias ramificadas sobre a população de moléculas em solução. Em estudos anteriores (52), proteínas com o CCD completamente glicosilado (três sítios e na completa extensão das cadeias ramificadas, como representado em nosso modelo) apenas foram isoladas quando no cultivo em meio mínimo (pH final de 2.5), enquanto que para o crescimento em meios mais ricos, o CCD apresentou uma variação no número e na extensão destas cadeias, devido à ação das glicosidases do fungo. É importante frisar, no entanto, que durante toda a nossa simulação, as N-glicosilações permaneceram voltadas para o solvente, virtualmente não interagindo ou interferindo com as interações do LP e vizinhança, para os quais foram efetuados nossos cálculos de energia representados na Fig. 5.3. Desta forma, embora a provável distribuição heterogênea das N-glicosilações possa mascarar o perfil recuperado pelo SAXS, ela não invalida a análise energética obtida para nossa simulação de SA/MD.

Em uma tentativa de melhor comparar os resultados do protocolo de SA/MD como o DAM recuperado para o SAXS, nós utilizamos as regiões de alta ocupância na Fig. 5.4 e as informações para o perfil das energias Vs conformação da Fig. 5.3 em um novo protocolo de modelagem de corpo rígido.

### **5.0.11 O modelo tridimensional de átomos dummy corrobora as principais interações computadas para o peptídeo conector na simulação**

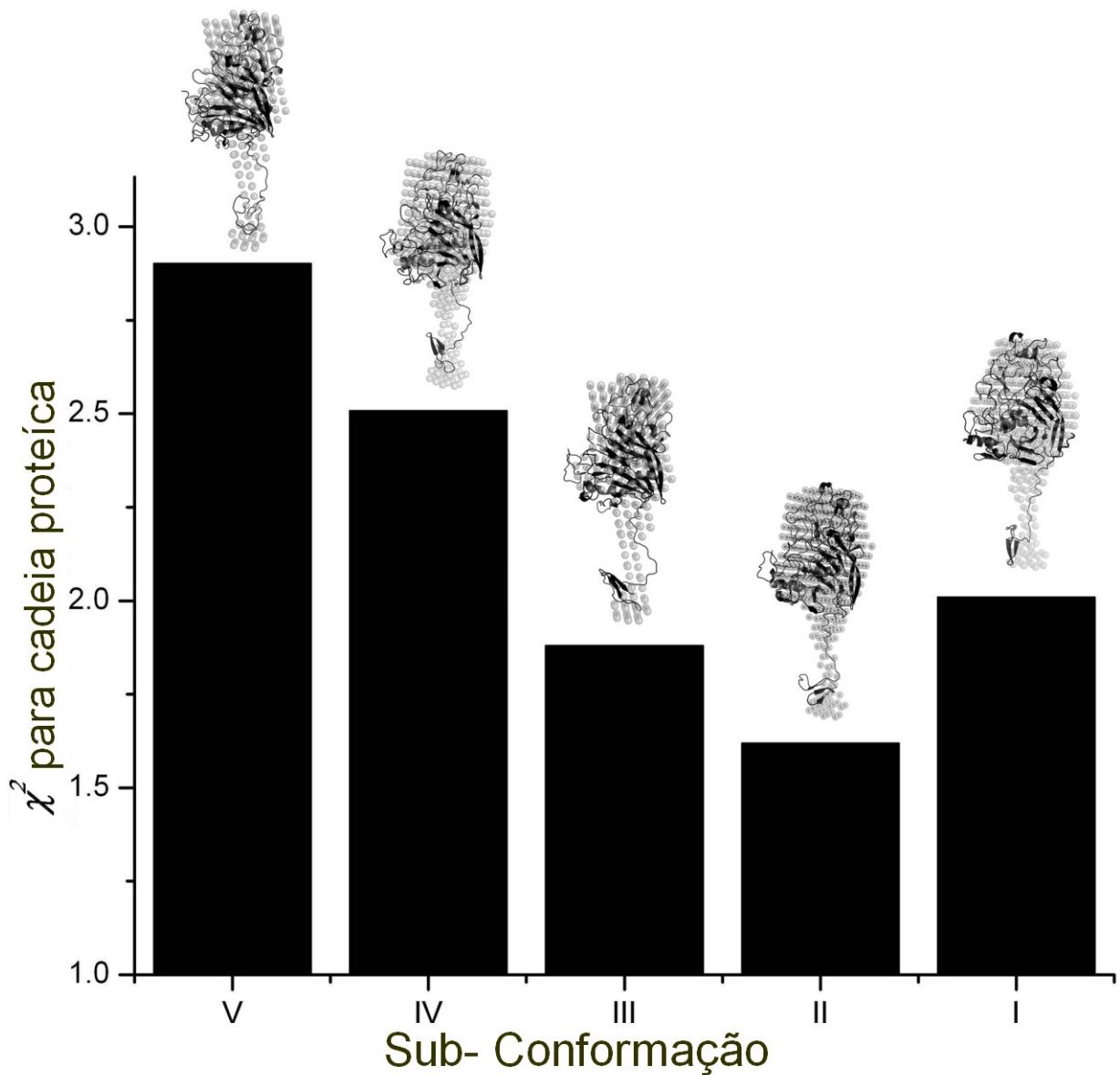
Primeiramente, apenas a cadeia protética dos cinco principais *decoys* foram utilizadas na modelagem. Os resíduos 414 a 426, pertencentes ao *loop* altamente móvel na entrada do túnel catalítico foram descartados.

Na Fig. 5.5 a discrepância ( $\chi^2$ ) para o *fitting* com o DAM para as estruturas médias de cada subtrajetória é ilustrada. Pode-se notar que o perfil do gráfico de  $\chi^2$  por conformação

reproduz o perfil das energias do LP Vs conformação ilustrado nas . O valor mínimo de  $\chi^2$  (1.62) foi obtido para a cadeia protéica da conformação II, que por sua vez apresentou o menor valor médio para as energias relacionadas à configuração do LP. Também as conformações III e II apresentaram, respectivamente, o segundo e terceiro menores valores de  $\chi^2$  (1.88 e 2.00, respectivamente), ambos condizentes com seus respectivos segundo e terceiro menores valores de energia média. As conformações supercontraídas IV e V, com valores médios de energia significativamente mais altos, apresentaram também os maiores valores de  $\chi^2$  (ambos acima de 2.00). Como visto acima, as energias calculadas para estas conformações são muito dependentes das interações “não ligadas” intramoleculares e da hidratação das O-glicosilações do LP. Desta forma, embora os carboidratos do LP não sejam visualizados em sua completa extensão no DAM recuperado através da modelagem para os dados de SAXS; o ajuste deste modelo para as conformações da cadeia protéica se mostrou conivente com a ordem em que estas cadeias foram favorecidas pelas interações destes mesmos glicosídeos em nossa simulação de SA/MD.

Em uma análise subsequente, nós buscamos por uma representação média para a proteína e carboidratos que melhors se adequasse ao DAM. Os gráficos de energia na Fig. 5.3 e para os valores de  $\chi^2$  Vs conformação na Fig. 5.4 sugerem ambos que o espaço conformacional da ThCel7A em solução é melhor representado por conformações em torno do mínimo centrado em valores de  $R_g$  de 27.8 Å e  $D_{CCD-CBM}$  de 46.7 Å. Dada a significativa liberdade vibracional da ThCel7A (ver 5.0.12), é razoável assumir que uma parte considerável da população de moléculas em solução dissipe sua energia térmica oscilando em torno deste mínimo e visitando conformações vizinhas nas regiões “achatadas” da superfície de energia.

Desta forma, nós geramos uma estrutura média para o LP, o CBM e os carboidratos correspondentes às regiões de maior ocupância para estas estruturas na Fig. 5.4. Todas as 85 estruturas da subtrajetória II, com conformação média de menor energia e que se mostrou mais representativa para os dados de SAXS, foram utilizadas na íntegra. Buscando melhor representar as conformações “marginais” ao poço de potencial, foram utilizadas as seguintes estruturas:



**Figura 5.5** – Ajuste dos 5 respectivos confôrmeros recuperados pelo SA/MD (I a V) aos dados experimentais de SAXS. Modelos de corpo rígido destas cinco estruturas médias sobre o DAM são mostrados.

- 10 estruturas da subtrajetória I com os menores valores de rmsd relativos à conformação II;
- 10 conformações para o LP intermediárias entre as conformações I e II, cujas estruturas foram amostradas do passo intermediário a 425 K entre estas duas subtrajetórias na simulação original;
- 10 conformações do LP intermediárias entre as conformações II e III, com estruturas amostradas da região de resfriamento de 425 a 300 K após o primeiro passo de aquecimento na simulação original.

Foram sempre consideradas intermediárias entre duas conformações as estruturas com menores valores de rmsd para uma média sobre estas duas mesmas conformações. Nós acreditamos que estas estruturas são razoavelmente bem representativas de uma conformação para a ThCel7A com o LP oscilando em torno do mínimo centrado sobre as estruturas medianas da subtrajetória II e flanqueado pelas conformações mais contraídas da subtrajetória I e por conformações intermediárias entre a II e III, conforme o gráfico de energia da Fig. 5.3. As O-glicosilações consideradas aqui foram as que mostraram maior ocupância ao longo da simulação:

- O resíduo único de manose O-ligado à Ser454, que permaneceu enterrado no CCD<sub>surf-pocket</sub> ao longo de toda a simulação (vide 5.0.6)
- Os resíduos únicos de manose respectivamente O-ligados às Ser461 e Ser462, ambos interagindo com a superfície “inferior” do  $\beta$ -sandwich do CCD durante a maior parte da simulação;
- O primeiro resíduo de manose diretamente O-ligado à Thr463 trimanosilada e o resíduo único de manose O-ligado à Ser465, ambos mantidos com ocupância relativamente alta na porção mediana da molécula devido a uma rede dinâmica de interações com o solvente e com os glicosídeos vizinhos (principalmente os dois glicosídeos citados no ítem prévio).

Para o CCD, a estrutura média foi calculada apenas sobre as subtrajetórias I e II, ambas apresentando os menores valores de rmsd relativos à estrutura cristalográfica original de *T. reesei*. Novamente, o *loop* móvel na entrada do túnel catalítico não foi considerado.

O modelo de corpo rígido resultante apresentou uma diminuição para o valor de  $\chi^2$  de 1.62 (para o *fitting* apenas da cadeia protética da conformação II) para 1.49 (Fig. 5.4-D). As extensões O-glicosídicas de alta ocupância utilizadas na modelagem se adequam consideravelmente ao DAM. Particularmente notório, é o *fitting* das extensões de carboidratos dos resíduos Ser461, Ser462, Thr463 e Ser465 à porção mais espessa abaixo da região aparentemente relativa ao CCD no DAM. Além disto, o domínio CBM, com alta mobilidade, quando promediado sobre o conjunto de conformações mencionadas apresenta-se com dimensões reduzidas, mais confinado em torno de seu centro de massa, se adequando melhor à pequena região globular na extremidade do DAM. Desta forma, o modelo de átomos dummy gerado a partir dos dados originais de SAXS comprehende as extensões O-glicosídicas com interações mais fortes ao longo do protocolo de SA/MD, bem como as conformações protéticas em torno do aparente mínimo global recuperado no mesmo protocolo. As interações dos carboidratos fitados, por sua vez, apresentaram grande participação nas energias dependentes da conformação do LP representadas na Fig. 5.3.

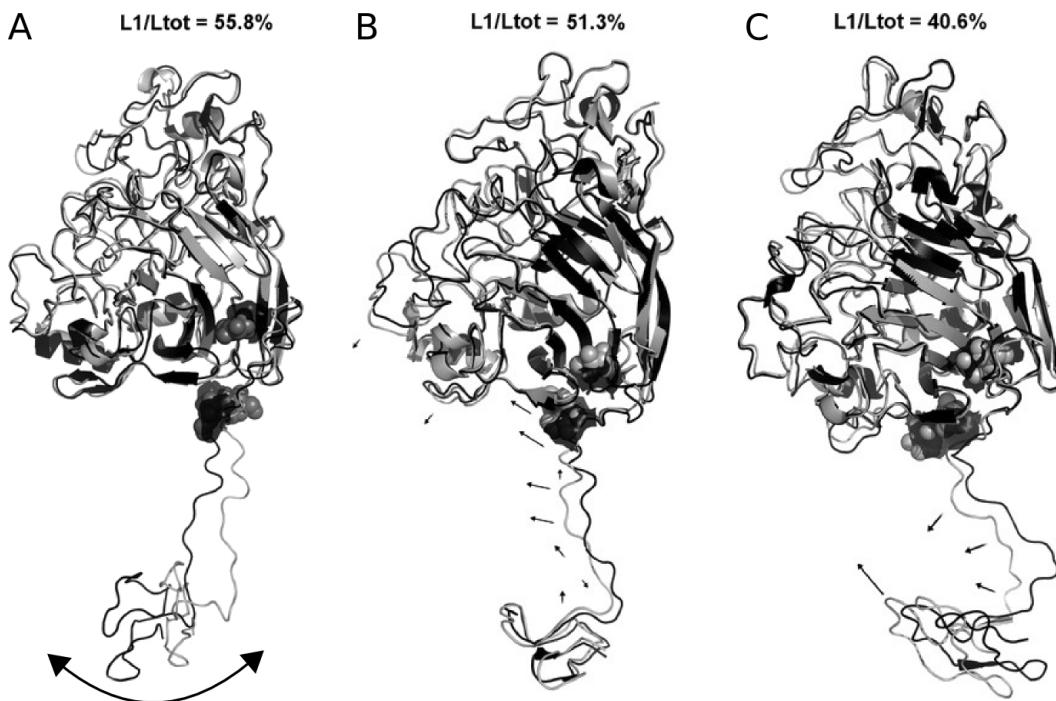
Os três resíduos mono-manosilados de serina cujas extensões glicosídicas interagem diretamente com a superfície do CCD em nosso modelo, são três das quatro únicas substituições de resíduos glicosilados entre os LPs das Cel7A de *T. reesei* e *T. harziannum*. Como será visto nas seções abaixo, a interação destes resíduos com a superfície do domínio catalítico se mostrou crucial para a manutenção do perfil vibracional do LP e sua substituição entre estas duas enzimas homólogas pode ser a causa para a documentada diferença de extensão entre as mesmas.

**5.0.12 O peptídeo conector é mantido em um conjunto de conformações vibracionalmente ativas conforme inferido através da análise das componentes principais e quasiharmônica para os diferentes confôrmeros.**

Nossa análise das energias relacionadas à extensão do LP aponta para um comportamento harmônico para este segmento na conformação mais energeticamente favorável (II): Semiflexionada e localizada em um poço de potencial bem delimitado relativo à sua extensão. Para o LP na conformação II, as cadeias oligosacarídicas do LP permanecem extendidas, de forma a maximizar a interação com o solvente, mas se mantêm preferencialmente orientadas para diferentes direções umas das outras, como os aros de um eixo (orientações similares das cadeias glicosídicas foram visualizadas por Zhong et al (53, 54) para as simulações da TrCel7A interagindo com a celulose microcristalina). Com este posicionamento, as camadas de hidratação das diferentes cadeias O-glicosídicas interferem minimamente umas com as outras. Este arranjo das cadeias de carboidratos promove o surgimento de pequenas “pregas” na cadeia principal do LP, facilitando o aparecimento de interações de hidrogênio temporárias em configurações de “pseudo-hélice”. Estas interações de hidrogênio são essencialmente dinâmicas, durante cada uma apenas poucos picosegundos ao longo da simulação de MD e sempre envolvendo a cadeia principal de pelo ao menos um resíduo glicosilado, como doador e/ou como acceptor (não mostrado).

Esta conformação semiflexionada e a dinâmica interna do peptídeo conector devido à alta mobilidade dos grupamentos glicosídico promove uma maior abundância de movimentos coletivos vibratórios para o LP com modos normais similares ao de um sistema massa mola. Os padrões de movimentos da molécula nas três conformações de menor energia (I, II e III) foram checados em uma análise de componentes principais (*Principal Component Analysis - PCA*) (60). Dada a baixa amostragem conformacional das curtas subtrajetórias analisadas, as componentes recuperadas na análise englobam principalmente modos normais do LP com frequências entre moderadas a altas (Fig. 5.6 ).

Na conformação I, o LP se encontra relativamente “frouxo” e as principais componentes são relacionadas a movimentos de pêndulo do conjunto LP e CBM sobre o CCD (Fig. 5.6-A).



**Figura 5.6** – Movimentos coletivos mais importantes do sistema LP-CBM relativo ao CCD para a média conformacional relativa a *A*- a subtrajetória I; *B*-subtrajetória II e *C*-subtrajetória III.

Na primeira componente principal, o LP e o CBM oscilam translacionamente em relação ao CCD, como em um pêndulo simples. Este movimento corresponde a 55.8 % do movimento total acessado para a molécula na subtrajetória I (ver os autovalores relativos mostrados sobre as estruturas na figura). Já a segunda e terceira componentes (contribuindo, respectivamente, com 9.7 e 7.5 % do movimento da molécula (não mostrado)) tratam-se de movimentos de torção e vibração de mais alta freqüência ao longo do LP. A única interação “não ligada” forte entre carboidrato-superfície do CCD nesta conformação foi observada para o anel único de manose O-ligado à Ser454 interagindo com a superfície do motivo  $\beta$ -sandwich do CCD (no padrão previamente mencionado em 5.0.6) que serve como base para os movimentos de pêndulo acima mencionados. Desta forma, há pouca transferência de energia vibracional entre o LP e a superfície do domínio catalítico nesta conformação.

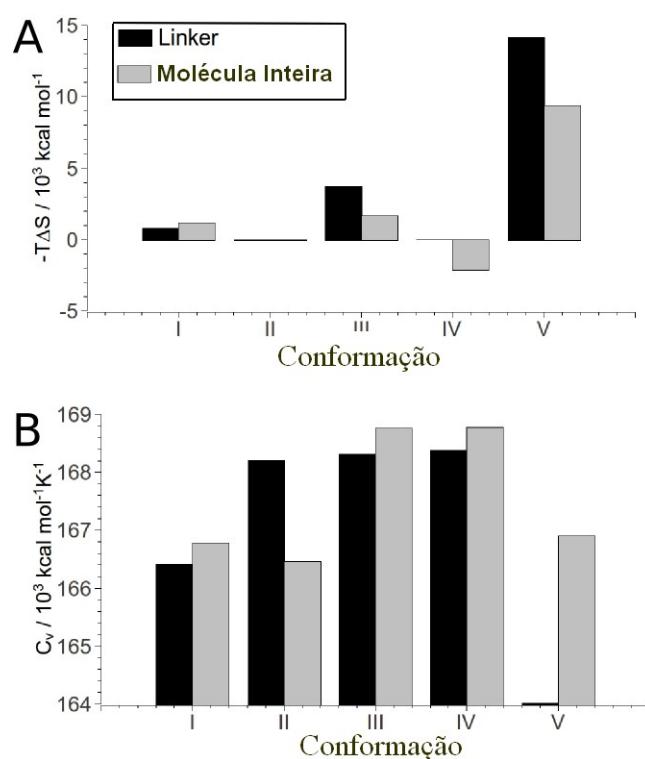
Na conformação II, por outro lado, as extensões únicas de manose O-ligadas, respectivamente, às Ser461 e Ser462 estabelecem (como previamente comentado em 5.0.11 e no tópico anterior) fortes contatos com a superfície do CCD. Isto aliado ao restante dos fatores energéticos previamente comentados em 5.0.8, induzem a uma semi-flexão do LP, com a cadeia principal

rica em “dobras” e pontes de hidrogênio, sobre o catalítico. Esta conformação semi-flexionada faz com que as três componentes principais para esta conformação sejam relacionadas a modos vibracionais semelhante aos de uma mola para o LP (Fig. 5.6-B). Particularmente, a primeira componente é descrita pela resonância entre movimentos de vibração das extensões O-manosídicas das Ser461 e Ser462, em contato com a superfície “inferior” do motivo  $\beta$ -*sandwich*, e um modo vibracional do LP que projeta este segmento e o CBM em direção ao CCD e “para a frente” (considerando o túnel catalítico como a região “frontal” da molécula). Este movimento promove uma grande transferência de energia vibracional do LP para a porção “inferior” rica em folhas  $\beta$  do CCD e desta para os *loops* do túnel catalítico e para o domínio como um todo.

Na conformação III, o LP se encontra relativamente super contraído, surgindo uma torção maior na porção C-terminal deste segmento (Fig. 5.6-C). As três componentes principais recuperadas para esta conformação englobam sobretudo movimentos da região C-terminal dobrada do LP (Fig. 5.6-C e não mostrado). Embora os contatos das extensões O-manosil dos resíduos Ser461 e Ser462 com a superfície do CCD se mantenham, esta mesma região do LP é consideravelmente pressionada contra a face “inferior” deste domínio, perdendo desta forma uma parte significativa de sua liberdade conformacional e dificultando a transferência de energia vibracional do LP para o CCD na conformação III.

A análise vibracional quasiharmônica (8) sobre as quatro subtrajetórias coletadas a partir de regiões estacionárias a 300 K (conformações I, II, III e V) sugere que a maior liberdade vibracional e a maior facilidade para transferência de energia vibracional interdomínios para a conformação II provê um pequeno aumento na entropia configuracional para esta conformação relacionada às demais (Fig. 5.7-A) (embora esta análise tenha sido realizada considerando uma amostragem conformacional pequena, sendo válida apenas no quesito comparativo entre os confôrmeros). A comparação do mínimo de energia potencial obtido para o LP na conformação II na Fig. 5.3 e a maior atividade dinâmica vibracional obtida para este segmento na mesma conformação através das análises de PCA e quasiharmônica (Figs. 5.6 e 5.7), corrobora a idéia de que o conjunto CCD-LP-CBM apresenta uma considerável componente harmônica de

seus movimentos coletivos acerca desta conformação, sugerindo novamente o comportamento deste sistema relativamente semelhante ao modelo de duas esferas conectadas por uma mola. Pode-se notar também na Fig. 5.7-B um valor de capacidade calorífica quasiharmônica significativamente maior para quando apenas o LP é considerado na conformação II que para o cálculo com a molécula inteira (significativamente diferente que o mesmo comportamento para as conformações vizinhas I e III). Isto corrobora a hipótese de que os modos vibracionais internos para a molécula são mais localizados sobre este segmento na conformação II que nas conformações vizinhas (em outras palavras, há um número significativamente maior de “caminhos” para transferência de energia vibracional interna do LP para o CCD nesta configuração, como previamente indicado pelas análises de componentes principais).



**Figura 5.7 – A** - Contribuição das entropias conformacionais à energia livre do LP e da proteína inteira nos diferentes confôrmeros, conforme recuperado por uma análise quasiharmônica (8). **B** - Capacidades caloríficas recuperadas pela mesma análise quasiharmônica. A conformação V apresenta uma relativa baixa entropia conformacional e pequena entropia para o peptídeo conector, refletindo sua maior rigidez inter-domínios.

Como esperado, a conformação V, altamente contraída, possui entropia conformacional significativamente menor que as mais estendidas, adicionando assim um custo entrópico configuracional para o mínimo local com conformações de menor extensão na Fig. 5.7. A maior ca-

pacidade calorífica para as conformações II e III sugere um maior número de graus de liberdade conformacionais termicamente acessíveis (com valores de energia próximos a  $KT$ ) para estas conformações semi-flexionadas comparadas às estruturas mais extendidas ou super-contraídas (conformações I e V, respectivamente). De fato, diversos estudos na literatura já demonstraram que cadeias polipeptídicas em solução em estado semi-flexionado (os chamados *random coil*) tendem a apresentar maior grau de liberdade conformacional que as mesmas cadeias em conformações super-extendidas ou enoveladas (63).

No capítulo seguinte serão apresentadas as principais conclusões decorrente desse trabalho de doutoramento.

## 6 *Conclusões*

### 6.1 Conclusão a respeito dos estudos dos mecanismos moleculares para o alosterismo de hRXR $\alpha$ a seu elemento responsável DR1

Tomando-se em conjunto os resultados para as simulações aqui apresentadas e sua comparação com outras estruturas cristalográficas de dímeros de hRXR $\alpha$  em DR1 com diferentes parceiros, bem como com resultados da literatura referentes ao acoplamento evolutivo entre resíduos aparentemente determinantes na interação com HREs do tipo DR1, conclui-se que a colaboratividade na interação dimérica a este HRE é regida por um mecanismo “chave e fechadura” ditado primariamente pela estabilização de um modo vibracional desta seqüência que tende a desestabilizar o complexo, sobretudo através da diminuição da disponibilidade do sítio para o monômero 5’. Este modo vibracional, por sua vez, é oriundo tanto da disposição relativa das fases dos dois meio-sítios na topologia macroscópica do DNA na seqüência de DR1, quanto de um “defeito” natural no empacotamento dos pares de bases na extremidade 3’ do primeiro meio-sítio. A correção das influências deste “defeito” para os modos vibracionais em DR1, só pode ser devidamente efetuada com a interação concertada do monômero de hRXR $\alpha$  no sítio 3’, disponibilizando o seu T-box desestruturado para a interação simultânea com o sulco menor entre os sítios e com o D-box do parceiro 5’ *correto*. O empacotamento provido por esta interação *correta*, estabiliza o sistema vibracionalmente a partir de seu centro de massas, corrigindo as influências do “defeito” no sítio 5’ e permitindo a associação estável de ambos os monômeros ao HRE, apresentando vantagens energéticas referentes ao empacotamento global do sistema (com acentuadas correlações entre as energias não ligadas de suas componentes) e à maior si-

metria vibracional para o complexo como um todo, com o centro de massas do complexo (onde ocorrem a maior parte de suas interações significativas) consideravelmente estabilizado, e um aumento gradual das amplitudes de movimento em direção às extremidades. Estas vantagens energéticas, por sua vez, compensam os custos entálpicos e entrópicos referentes à desnaturação e desempacotamento do T-box do cerne globular do domínio DBD, ao deslocamento dos íons das imediações dos sulcos menor e maior do DNA e à distorção da simetria original para os movimentos coletivos do segmento de DR1, corroborando o perfil de “arraste” (referente ao custo para a interação do primeiro monômero) e cooperatividade (a vantagem acentuada para a interação do dímero) vislumbrado para as isotermas de associação a este elemento através dos experimentos de anisotropia de fluorescência na seção 4.1.1.

## **6.2 Conclusões a respeito da modelagem estrutural e dinâmica interdomínios da Cellobiohidrolase de *Trichoderma harzianum* em solução**

Primeiramente, os resultados apresentados neste capítulo representam um avanço significativo na qualidade dos dados disponíveis para cellobiohidrolases de espécies de *Trichoderma* em solução. Os modelos prévios de SAXS obtidos para a Cel7A de *Trichoderma reesei* (64) foram construídos quando as estruturas isoladas dos dois domínios eram ainda desconhecidas. Este aumento em informação estrutural de alta resolução somado aos progressos de instrumentação de medida e estatística para o tratamento dos dados de SAXS nas quase três décadas que separam estes estudos, bem como os progressos nas áreas de modelagem molecular, permitiram aliar a técnica experimental de SAXS (de baixa resolução) às ferramentas de modelagem estrutural e dinâmica molecular. Esta integração, por sua vez, proporcionou um considerável aumento na resolução e na qualidade da informação obtida para a estrutura e comportamento dinâmico da molécula.

Nossa análise integrada de SAXS e MD para a ThCel7A em solução permite inferir algumas considerações a respeito de sua biologia.

Os resultados confirmaram a forma de “girino” para a enzima, mas indicaram uma dimensão máxima significativamente menor que os 180 Å obtidos previamente por SAXS para a TrCel7A (46). A ThCel7A, com mais de 80 % de homologia e sequência primária com extensão muito próxima à da TrCel7A (as duas enzimas possuem respectivamente 505 e 513 resíduos), apresenta um valor de  $D_{Max}$  de 110 Å. O valor de  $R_g$  obtido para a enzima de *T. harziannum* (27.56 Å) também é significativamente menor que os 42.7 Å obtidos no mesmo estudo anterior para a TrCel7A completa. Embora um estudo posterior de microscopia eletrônica de alta resolução tenha obtido o valor significativamente menor de  $D_{Max} = 151 \pm 13$  Å para a TrCel7A (sugerindo que os estudos prévios de SAXS foram sujeitos a erros devido à limitada acurácia desta técnica na época), este valor ainda é condizente com um LP completamente extendido para a enzima de *T. reesei*, enquanto que nossos dados de SAXS sugerem um peptídeo semi-flexionado para *T. harziannum*.

Nossos resultados de modelagem estrutural e simulação por SA/MD, comparados ao modelo de átomos dummy recuperado no tratamento dos dados de SAXS, e a análise sequencial das duas enzimas apontam para uma possível explicação para esta aparente diferença de extensão entre estas duas enzimas de alta homologia. Nós assumimos, em nossa explicação que o padrão de O-glicosilações para o LP da ThCel7A é similar ao previamente determinado para a TrCel7A (51): contendo todos os resíduos de serina parcialmente glicosilados com um único resíduo de manose e as treoninas com extensões de cadeias lineares variáveis até três resíduos. Esta mesma premissa foi assumida para a construção de nosso modelo tridimensional de alta resolução que, quando submetido à simulação de SA/MD, produziu resultados condizentes entre a simulação e os dados experimentais de SAXS em nossa análise até aqui. Como visto anteriormente, há três resíduos de serina cujas respectivas extensões de O-manose interagem diretamente com a superfície externa do motivo  $\beta$ -sandwich do CCD durante uma parte significativa de nossa simulação, possuindo todos os três um considerável ajuste com o modelo recuperado para os dados de SAXS. Em nosso SA/MD, as interações destes resíduos com a superfície do CCD contribuíram notoriamente para a estabilização da conformação dobrada da porção N-terminal do LP (extensão da Ser454) e da flexão da região média deste segmento

(extensões da Ser461 e de Ser462), contribuindo assim para a configuração semiflexionada recuperada a partir dos dados de SAXS. No LP da TrCel7A, estes resíduos são todos substituídos por treoninasComo pode ser visto nas Fig. as interações destas extensões de manose com a superfície do CCD apresentaram, em nosso estudo de SA/MD, uma complementaridade estérica que não poderia ser alcançada para as longas cadeias bi ou tri-manosiladas das extensões dos resíduos de treonina. Estas extensões maiores, em nossa simulação, permaneceram preferencialmente voltadas para o solvente, de forma a evitar o impedimento estérico com o restante da molécula e otimizando a hidratação. Não há qualquer outra substituição Thr/Ser entre os LPs de ThCel7A e TrCel7A que pudessem compensar o efeito das três acima mencionadas. Desta forma, é esperado um peptídeo conector relativamente mais extendido e menos rígido para a TrCel7A, mais rica em resíduos de treonina com cadeias lineares longas de manose, comparado ao de ThCel7A, concordando assim com os dados experimentais obtidos independentemente para estas duas proteínas.

Um relacionamento direto entre o comprimento médio do LP e o teor de glicosilação para celulases já havia sido reportado em estudos prévios (65). Como discutido em 5.0.12, as interações citadas das extensões O-manosídicas dos três resíduos de serina com a superfície do CCD apresentaram grande influência sobre a natureza dos modos normais de vibração da enzima (recuperados na análise de componentes principais). Desta forma, se o padrão proposto aqui está correto, são esperadas diferenças significativas para as entalpias e entropias vibracionais e capacidades caloríficas dos LP de ThCel7A e TrCel7A. As possíveis consequências destas diferenças para a processividade relativa das duas enzimas são temas para estudos futuros.

Outro tema que tem atraído considerável interesse na literatura recente se refere aos detalhes da ação conjunta dos dois domínios das celulases no movimento sobre o substrato de celulose microcristalina. Como previamente mencionado em 5.0.6, diversos resultados da literatura baseados em estudos de SAXS sugeriram um perfil flexível e variante entre semi-extendido a extendido para o LP de celulases em solução Este padrão levou os autores a proporem um modelo em que estas enzimas poderiam se mover no substrato cristalino com movimentos semelhantes aos de uma “lagarta”, com o LP funcionando como uma mola conectora entre os dois domínios.

Também foi proposto que a energia livre para a realização deste movimento seria provida pela translação do CBM sobre a superfície da celulose, visto ter sido demonstrado este potencial de difusão do domínio pela superfície da celulose em trabalhos prévios . Contudo, Igarashi et al demonstraram recentemente (66) via microscopia de força atômica de alta velocidade (*High Speed-Atomic Force Microscopy - HS-AFM*), que o CCD seria o único responsável pelo deslizamento da proteína sobre o substrato insolúvel, tendo o CBM a única função de aumentar a concentração da enzima na superfície deste mesmo substrato, como havia sido previamente proposto em . Desta forma, as causas para o significativo incremento conferido pelo LP na processividade da enzima devem ser atribuídas principalmente a mecanismos que quebrem a barreira energética para a passagem da molécula da fase solúvel para a superfície de celulose e que dificultem a desorção da mesma. Os resultados apresentados aqui, não só confirmam a configuração flexível, semi-extendida e com potencial harmônico relativo ao comprimento do LP (conforme previamente sugerido em ) como sugerem possíveis mecanismos de como este mesmo perfil pode facilitar a ruptura da barreira acima mencionada e dificultar a desorção da molécula ao substrato (em concordância com os dados de (66)).

Em primeiro, as conformações mais energeticamente favoráveis recuperadas no SA/MD e, ao mesmo tempo, as que melhor se adequam aos dados de SAXS apresentam os dois domínios em relativa distância um do outro, mas ainda com uma relativa tensão entre ambos, devido ao potencial harmônico aqui obtido. Desta forma, o peptídeo conector é, ao mesmo tempo, suficientemente longo e flexível para aumentar a probabilidade do CBM se ligar à superfície da celulose e suficientemente tenso para “puxar” o CCD em direção a esta superfície uma vez que o CBM se ligue.

Outro dado interessante em nossa análise, são os modos vibracionais para a conformação II, a conformação menos energética e que melhor se ajusta ao modelo de SAXS. Nesta conformação, o LP apresenta uma alta atividade vibracional que é transferida à molécula como um todo. Matthews et al , em seus estudos de dinâmica molecular para a celulose microcristalina I $\beta$ , verificaram uma forte localização de moléculas de água adjacentes na primeira camada de hidratação da superfície da celulose, devido tanto a interações de hidrogênio com o substrato

quanto a efeitos de natureza hidrofóbica. Os autores propuseram que estas moléculas estruturadas de água poderiam se apresentar como uma barreira para a aproximação de celulases e que estas enzimas deveriam apresentar mecanismos moleculares para romper a barreira. Desta forma, os movimentos vibracionais altamente acoplados para os três módulos (CCD-LP-CBM) em conformações próximas à conformação II de nosso estudo poderiam dispersar mais facilmente estas camadas de água fortemente ligadas, de forma a facilitar a aproximação e interação da ThCel7A com a superfície celulósica. É interessante notar que, apesar do maior volume e das interações mais favoráveis, as camadas de hidratação para o LP e vizinhanças na conformação II apresentaram-se com significativa maior mobilidade, relacionado às conformações mais contraídas (notar a menor quantidade de regiões de alta densidade média de moléculas de água para esta conformação na ). Por outro lado, O CCD, de estrutura globular e significativamente rígida, apresenta uma reduzida atividade vibracional intrínseca (compare o perfil vibracional obtido apenas para o LP com o obtido quando a molécula inteira é considerada na Fig. ), e uma afinidade pela celulose consideravelmente menor que o CBM , necessitando portanto da extensão LP-CBM para otimizar sua interação com a mesma.

É também interessante notar que o movimento de “cavalo marinho” ilustrado na primeira componente principal para esta conformação dirige, simultaneamente, a superfície de ligação à celulose do CBM, o LP e o túnel catalítico do CCD para a mesma direção, o que também poderia facilitar a ligação simultânea dos dois domínios ao substrato. Uma vez que a molécula se ligue à superfície microcristalina, a mesma atividade vibracional poderia facilitar a captação de substrato na entrada do túnel catalítico, promovendo uma maior fixação da ThCel7A nesta superfície e o início da reação e do consequente deslizamento sobre a celulose.

Os movimentos da molécula ao longo da região “achatada” da superfície de energia da Fig. se assemelha, de fato, ao movimento de “lagarta” (*caterpillar like motion* proposto por Receveur et al . Além do mais, a variação de extensão do LP ao longo desta mesma região “achatada” (da conformação III à I) acessada em nossa simulação foi de aproximadamente 9.5 Å, próxima portanto da dimensão de uma única unidade de repetição de celobiose na celulose cristalina (10.4 Å). Isto é condizente com a idéia de que os movimentos do LP poderiam direcionar os

dois domínios para “caminhar” sobre a superfície da celulose durante o processo de hidrólise. Contudo, estas duas extensões extremas para o LP na região “achatada” da Fig. são separadas por um poço de potencial relativamente profundo (aproximadamente 30 kcal·mol<sup>-1</sup>).

Considerando que a hidratação do LP e as interações “não ligadas” para os carboidratos desempenham um papel crucial na manutenção da diferença de energia entre as conformações acessadas , é possível (embora isto tenha de ser investigado em trabalhos futuros) que com a ligação à celulose (devido à reduzida atividade das camadas de solvatação nesta fase sólida e às possíveis interações dos O-glicosídeos do LP com as unidades de celobiose) a superfície de energia para as conformações I, II e III se torne ainda mais “achatada”. Desta forma, a molécula poderia transitar livremente entre estas (e outras) diferentes extensões devido ao efeito entrópico. Este possível aumento da flexibilidade conformacional, por sua vez, permitiria tanto uma maior adaptabilidade da molécula às diferentes restrições do relevo da superfície como também que os movimentos de extensão e contração do LP entrassem em fase com o processamento das unidades de celobiose, promovendo assim o proposto movimento de “lagarta” durante a hidrólise do substrato insolúvel.

As considerações a respeito da organização estrutural e dinâmica obtidos no presente estudo refletem, presumivelmente, aspectos de toda a família de enzimas Cel7A. Este conhecimento seria difícil, se não impossível, de ser obtido através da técnica de cristalografia de raios-X, devido à já conhecida flexibilidade do LP e à alta variação em extensão e igual mobilidade das cadeias de glicosídeos (51, 67). Além disto, experimentos de SAXS têm a habilidade de estudar estas estruturas em condições mais próximas às fisiológicas (62), enquanto simulações de dinâmica molecular provêm detalhes acerca da movimentação de suas partes individuais, bem como as energias relacionadas, em uma escala dificilmente acessível às técnicas experimentais (68). Estes dados experimentais de baixa resolução obtidos por SAXS integrados aos dados de simulação de alta resolução obtidos por modelagem estrutural e dinâmica molecular aumentam consideravelmente a quantidade de informação disponível para a organização estrutural e flexibilidade conformacional para a ThCel7A e, de uma forma geral, para celobiohidrolases em solução.

Os resultados deste trabalho foram reunidos no artigo "Hydration and intramolecular interactions impose a spring like effect to the O-glycosilated linker peptide of cellobihidrolase I from *Trichoderma harziannum*" (Lima LHF et al) e estão sendo submetidos para publicação em periódico indexado internacional.

## REFERÊNCIAS

- 1 MORGAN, J. D.; MCCAMMON, A.; NORTHRUP, S. H. Molecular dynamics of ferrocytocrome C: Time dependence of the atomic displacement. *Biopolymers*, v. 22, n. 6, p. 1579–1593, Junho 1983.
- 2 MCCAMMON, J. A. Protein dynamics. *Reports Progress in Physical*, v. 47, p. 1–46, Junho 1984.
- 3 MARTINEZ, L. *Simulações de dinâmica molecular dos receptores de hormônio tireoideano*. Tese (Doutorado) — Instituto de Química. Universidade Estadual de Campinas, 2007.
- 4 RASTINEJAD, F. et al. Structure of the RXRRAR DNA-binding complex on the retinoic acid response element DR1. *The Embo Journal*, v. 19, p. 1045–1054, Outubro 2000.
- 5 ZHAO, Q. et al. Structural Basis of RXR-DNA Interactions. *Journal Molecular Biology*, v. 296, p. 509–520, Outubro 2000.
- 6 RASTINEJAD, F. et al. Structural determinants of nuclear receptor assembly on DNA direct repeats. *Nature*, v. 375, p. 203–211, maio 1995.
- 7 RASTINEJAD, F. Retinoid X receptor and its partners in the nuclear receptor family. *Elsevier Science.*, v. 11, p. 33–38, Junho 2001.
- 8 WERESZCZYNSKI, J.; ANDRICIOAEI, I. Conformational and Solvent Entropy Contributions to the Thermal Response of Nucleic Acid-Based Nanothermometers. *Journal Physical Chemical B*, v. 114, p. 2076–2082, Janeiro 2010.
- 9 WISHARD, M. NMR: prediction of protein flexibility. *Nature protocols*, v. 1, p. 683–688, 2006.
- 10 KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. *nature structural biology*, v. 9, p. 101–123, Setembro 2002.
- 11 CHOU, K. C.; CHENH, N. Y. The biological functions of low-frequency phonons. *Scientia Sinica*, v. 20, p. 447–457, 1977.
- 12 PAINTER, P. C.; MOSHER, L. E.; RHOADS, C. Low-frequency modes in the Raman spectra of proteins. *Biopolymers*, v. 21, p. 1469–1472, 1982.
- 13 URABE, H.; TOMINAGA, Y. Low-lying collective modes of DNA double helix by Raman spectroscopy. *Biopolymers*, v. 21, p. 2477–2481, 1982.
- 14 MARTEL, P. Biophysical aspects of neutron scattering from vibrational modes of proteins. *Progress Biophysical Molecular Biological.*, v. 57, p. 129–179, 1992.

- 15 LEACH, A. R. *Molecular Modelling: principles and applications*. Second Edition. [S.l.]: Pearson education, 2001.
- 16 FIELD, M. J. *Simulation of Molecular Systems*. Second Edition. [S.l.]: Cambridge university press, 2001.
- 17 HANSEN, J. P.; MCDONALD, I. R. *Theory of Simple Liquids*. Second Edition. [S.l.]: Academic press, 1986.
- 18 YAN, Z. General thermal wavelenght and its applications. *European Journal of physics*, v. 21, p. 625–631, 2000.
- 19 SZABO, A.; OSTLUND, N. S. *Modern Quantum Chemistry - Introduction to Advanced Eletronic Structure Theory* . fifth Edition. [S.l.]: Dover Publications, 1996.
- 20 ADCOCK, S. A.; MCCAMMON, J. A. Molecular Dynamics: Survey of Methods for Simulating the Activity of Proteins. *Chemical Review* , v. 106, p. 1589–1615, Maio 2006.
- 21 BROOKS, B. R. et al. Charmm. A Program for Macromolecular Energy, Minimization and Dynamics Calculations. *Journal Computational Chemical*, v. 4, p. 187–217, 1983.
- 22 PEARLMAN, D. A. Amber, a Package of Computer-Programs for Applying Molecular Mechanics, Normal-Mode Analysis, Molecular-Dynamics and Free-Energy Calculations to Simulate the Structural and Energetic Properties of Molecules. *Computational Physics Comments*, v. 91, p. 1–41, 1995.
- 23 ZHANG, Y.; SKOLNICK, J. Automated structure prediction of weakly homologous proteins on a genomic scale. *PNAS.*, v. 101, p. 7594–7599, Abril 2004.
- 24 ZHANG, Y. Template-based modeling and free modeling by I-TASSER in CASP7. *Wiley InterScience*, v. 69, p. 108–117, Junho 2007.
- 25 WU1, S.; SKOLNICK, J.; ZHANG, Y. Ab initio modeling of small proteins by iterative TASSER simulations. *BMC Biology*, v. 5, p. 1741–7007, maio 2007.
- 26 ZHANG, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*, v. 9, p. 1741–2105, Janeiro 2008.
- 27 MARTÍNEZ, J. M.; MARTÍNEZ, L. Packing optimization for automated generation of complex system's initial configurations for molecular dynamics and docking. *Journal Computational Chemical*, v. 24, p. 819–825, 2003.
- 28 KRAUSS, G. *Biochemistry of Signal Transduction and Regulation* . second edition. [S.l.]: Wiley-VCH, 2001.
- 29 NELSON, D. L.; COX, M. M. *Lehninger Principles of Biochemistry*. Fourth Edition. [S.l.]: s.n., 2004.
- 30 VOET, D.; VOET, J. G.; PRATT, C. W. *Fundamentals of Biochemistry*. [S.l.]: John Wiley & Sons, Inc., 1999.
- 31 NETTLES, K. W.; GREENE, G. L. Ligand control to coregulator recruitment to nuclear receptors. *Annual Review Physiology*, v. 67, p. 309–333, 2005.

- 32 GRONEMEYER, H.; GUSTAFSSON, J. A.; LAUDET, V. Principles for modulation of the nuclear receptor superfamily. *Nature Review Drug Discovery*, v. 2, p. 950–964, 2004.
- 33 KLIEWER, S. A.; LEHMANN, M. J.; WILLSON, T. M. Orphan nuclear receptors: shifting endocrinology into reverse. *Science*, v. 284, p. 757–760, 1999.
- 34 CHEN, Z. P. et al. Ligand- and DNA-induced Dissociation of RXR Tetramers. *Journal Molecular Biology*, v. 275, p. 55–65, Janeiro 1993.
- 35 FIGUEIRA, A. C. M. *Oligomerização, estruturas a baixa resolução, ligação ao DNA e ao ligante dos Receptores de Hormônios Tireoidianos*. Tese (Doutorado) — Instituto de Física de São Carlos, da Universidade de São Paulo, 2008.
- 36 KERSTEN, S.; GRONEMEYER, H.; NOY, N. The DNA Binding Pattern of the Retinoid X Receptor Is Regulated by Ligand-dependent Modulation of Its Oligomeric State. *Biophysical Journal*, v. 272, p. 12771–12777, Maio 1997.
- 37 RASTINEJAD, F. et al. Structural determinants of nuclear receptor assembly on DNA direct repeats . *Nature America Inc*, v. 375, p. 5, maio 1995.
- 38 MADER, S. et al. The patterns of binding of RAR, RXR and TR homo-and heterodimers to direct repeats are dictated by the binding specificities of the DNA binding domains . *The EMBO Journal* , v. 12, p. 5029–5041, Agosto 1993.
- 39 LEE, M. S.; SEEM, D. S.; KLIEWER, D. S. Structure of the retinoid X receptor  $\alpha$  DNA-binding domain: a helix required for homodimeric DNA-binding. *Science*, v. 260, p. 1117–1121, julho 1993.
- 40 SIERK, M. L.; ZHAO, Q.; RASTINEJAD, F. DNA Deformability as a Recognition Feature in the RevErb Response Element. *Biochemistry*, v. 40, p. 12833–12843, Outubro 2001.
- 41 CHANDRA, V. et al. Structure of the intact PPAR - $\gamma$ - RXR-  $\alpha$  nuclear receptor complex on DNA . *Nature America Inc*, v. 456, Novembro 2008.
- 42 ZHAO, Q. et al. Structural Basis of RXR-DNA Interactions. *Journal Molecular Biology*, v. 296, p. 509–520, maio 2000.
- 43 BOURGUET, W. et al. Heterodimeric Complex of RAR and RXR Nuclear Receptor Ligand-Binding Domains: Purification, Crystallization, and Preliminary X-Ray Diffraction Analysis . *Protein Expression and Purification* , v. 19, p. 284–288, Fevereiro 2000.
- 44 LEE, I. et al. Substrate-Enzyme Interactions in Cellulase Systems. *Bioresource Technology* , v. 58, p. 163–169, Agosto 1996.
- 45 TSUI, V.; CASE, D. A. Implications of cellobiohydrolase glycosylation for use in biomass conversion. *Journal American Chemical Society*, v. 122, p. 2489–2498, Janeiro 2000.
- 46 ABUJA, P. M. et al. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*- A small angle X-ray scattering study of the intact enzyme and its core. *European Biophysics Journal* , v. 15, p. 339–342, Outubro 1988.
- 47 ROY, A.; KUCURAL, A.; ZHANG, Y. Solvent effect on the ground and excited state dipole moments of fluorescein. *Nature protocols*, v. 5, n. 4, p. 4436–4451, Novembro 2010.

- 48 ZHANG, Y. *NCBI non redundant databank I*. Produced by Y. Zhang. Disponível em: <<ftp://ftp.ncbi.nih.gov/blast/db/nr.00.tar.gz>>. Acesso em: em Fevereiro de 2010.
- 49 ZHANG, Y. *NCBI non redundant databank II*. Produced by Y. Zhang. Disponível em: <<ftp://ftp.ncbi.nih.gov/blast/db/nr.01.tar.gz>>. Acesso em: Fevereiro de 2010.
- 50 ZHANG, Y.; KOLINSKI, A.; SKOLNICK, J. Touchstone II: a new approach to ab initio protein structure prediction. *Biophysical Journal*, v. 85, n. 1, p. 1145–1164, Agosto 2003.
- 51 HARRISON, M. et al. Modified glycosylation of cellobiohydrolase I from a high cellulose producing mutant strain of *Trichoderma reesei*. *European Journal of Biochemistry*, v. 256, p. 119–127, Janeiro 1998.
- 52 STALS, I. et al. Factors influencing glycosylation of *Trichoderma reesei* cellulases. I: Postsecretorial changes of the O- and N-glycosylation pattern of Cel7A. *Glycobiology*, v. 14, n. 8, p. 713–724, Abril 2004.
- 53 ZHONG, L. et al. Interactions of the complete cellobiohydrolase I from *Trichoderma reesei* with microcrystalline cellulose I $\beta$ . *Cellulose*, v. 15, p. 261–273, Janeiro 2008.
- 54 ZHONG, L. et al. Computational simulations of the *Trichoderma reesei* cellobiohydrolase I acting on microcrystalline cellulose I $\beta$ : the enzyme-substrate complex. *Carbohydrate Research*, v. 344, p. 261–273, Janeiro 2009.
- 55 WOODS, R.; DWEK, R.; EDGE, C. Molecular mechanical and molecular dynamical simulations of glycoproteins and oligosaccharides. *Journal of Physical Chemistry*, v. 99, n. 11, p. 3832–3846, Agosto 1995.
- 56 KIRSCHNER, K. et al. GLYCAM06: A new generalizable biomolecular force field. Carbohydrates. *Journal of Computational Chemistry*, v. 29, n. 4, p. 622–655, Julho 2008.
- 57 JORGENSEN, W. et al. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, v. 79, p. 926–935, 1983.
- 58 KERR, I. et al. Parallel helix bundles and ion channels: molecular modeling via simulated annealing and restrained molecular dynamics. *Biophysical Journal*, v. 67, p. 1501–1515, Outubro 1994.
- 59 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual Molecular Dynamics. *Journal Molecular Graphics*, v. 14, p. 33–&, maio.
- 60 SEEGER, M. et al. Wordom: a program for efficient analysis of molecular dynamics simulations. *Bioinformatics*, v. 23, n. 19, p. 2625–2627, Agosto 2007.
- 61 WILLIS, S.; GRIFFIN, P. Mutual information identifies sequence positions conserved within the nuclear receptor superfamily: approach reveals functionally important regions for DNA binding specificity. . *The Open Access Journal of the Nuclear Receptor Signaling Atlas*, v. 9, p. 1, julho 2011.
- 62 SVERGUN, D.; PETOUKHOV, M.; KOCH, M. Determination of domain structure of proteins from X-ray solution scattering. *Biophysical Journal*, v. 80, p. 2946–2953, Jun 2001.

- 63 MA, B.; TSAI, C.; NUSSINOV, R. A systematic study of the vibrational free energies of polypeptides in folded and random states. *Biophysical Journal*, v. 79, p. 2739–2753, Novembro 2000.
- 64 ABUJA, P. et al. Structural and functional domains of cellobiohydrolase I from *Trichoderma reesei*. *European Biophysical Journal*, v. 15, p. 339–342, Outubro 1988.
- 65 OSSOWSKI, I. von et al. Protein disorder: conformational distribution of the flexible linker in a chimeric double cellulase. *Biophysical Journal*, v. 88, p. 2823–2832, Abril 2005.
- 66 IGARASHI, K. et al. High speed atomic force microscopy visualizes processive movement of *Trichoderma reesei* cellobiohydrolase I on crystalline cellulose. *Journal of Biological Chemistry*, v. 284, n. 52, p. 36186–36190, Dezembro 2009.
- 67 NAIDOO, K.; BRADY, J. Molecular dynamics simulations of a glycoprotein: the lectin from *Erythrina corallodendron*. *Journal of Molecular Structure*, v. 395, n. 396, p. 469–475, Aug 1996.
- 68 KARPLUS, M.; MCCAMMON, J. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, v. 9, n. 9, p. 646–652, Sep 2002.