

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

**INVESTIGAÇÃO COMPUTACIONAL DOS MECANISMOS DE
INTERAÇÃO ENTRE BASES DE TRÖGER E O DNA**

CLARISSE GRAVINA RICCI

Porto Alegre, Fevereiro de 2010

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

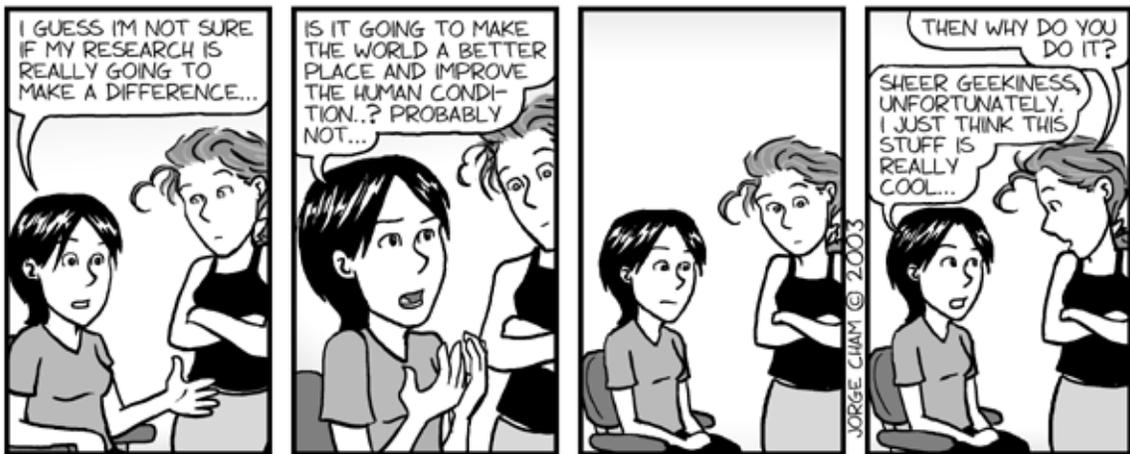
CLARISSE GRAVINA RICCI

**INVESTIGAÇÃO COMPUTACIONAL DOS MECANISMOS DE
INTERAÇÃO ENTRE BASES DE TRÖGER E O DNA**

Dissertação apresentada como requisito parcial para a
obtenção do grau de Mestre em Química

Prof. Dr. Paulo Augusto Netz
Orientador

Porto Alegre, Fevereiro de 2010



www.phdcomics.com

"Piled Higher and Deeper" by Jorge Cham
www.phdcomics.com

À minha mãe.

AGRADECIMENTOS

Ao meu pai, que me explicou pela primeira vez, quando eu tinha 11 anos, a estrutura molecular do DNA.

Ao meu orientador, Paulo Netz, que sempre se mostrou disponível, porém ao mesmo tempo confiante da minha capacidade, deixando que eu andasse pelas minhas próprias pernas. Agradeço por ter compartilhado conosco – alunos da Química Teórica – as suas opiniões e o seu conhecimento sobre os mais variados temas durante os cafés das 3h da tarde no Antônio. Principalmente, gostaria de agradecê-lo por ser um grande (e raro) exemplo de pesquisador íntegro e genuinamente interessado no crescimento dos seus alunos que eu tive a sorte de encontrar no meio acadêmico.

À minha mãe - com quem sempre tive as melhores conversas - por estar sempre por perto, me convidando para chás da tarde e cafés.

Ao Hermes.

À Melina, William, Jones e Helen, pelo companheirismo nos congressos e nas disciplinas.

Aos demais colegas e pesquisadores do laboratório de Química Teórica, por garantirem um bom ambiente de pesquisa.

Ao Marcelo, meu companheiro de estudos desde o processo de ingresso no mestrado, pois juntos jamais esqueceremos de que a resposta nem sempre é “o ovo”. Também agradeço pelas lições de vida, especialmente o grande lema de persistência “Atura o Parabéns”.

À Capes, pela concessão da Bolsa de Estudo.

Trabalhos gerados a partir dos resultados da dissertação

Ricci, C.G.; Netz, P.A. Docking studies on DNA-ligand interactions: building and application of a protocol to identify the binding mode. *Journal of Chemical Information and Modeling*, v. 49, p. 1925-1935, 2009. (Apêndice A).

ÍNDICE

1. INTRODUÇÃO	1
1.1 Ligantes de DNA	1
1.2 Bases de Tröger	5
2. OBJETIVOS	8
2.1 Gerais	8
2.2 Específicos	8
3. ESTUDOS DE DOCAGEM MOLECULAR	9
3.1 Introdução	9
3.1.1 Docagem molecular	9
3.1.2 Algoritmo genético e o <i>AutoDock</i> 4.0	10
3.1.3 Docagem aplicada a ácidos nucleicos	12
3.2 Metodologia	15
3.3 Resultados e discussão	18
3.3.1 Docagens diretas	19
3.3.2 Docagens cruzadas	21
3.3.3 Análise conformacional dos receptores	22
3.3.4 Comparação entre protocolos de docagem	26
3.3.5 Predição do modo de ligação das bases de Tröger	27
3.4 Conclusões	32
4. ESTUDOS DE DINÂMICA MOLECULAR	34
4.1 Metodologia	34
4.2 Resultados e discussão	36
4.2.1 Estabilidade dos oligonucleotídeos	36
4.2.2 Análise global das simulações	40
4.2.3 Análise conformacional da cadeia principal	50
4.2.3.1 Ângulos α e γ	52
4.2.3.2 Ângulos ϵ e ζ	58
4.2.4 Análise de parâmetros de pares de bases	62
4.2.5. Análise dos <i>Bending-dials</i>	74
4.3 Conclusões.....	80
5. CONSIDERAÇÕES FINAIS	82
6. BIBLIOGRAFIA	86

LISTA DE FIGURAS

Figura 1. Principais modos de interação com o DNA.....	02
Figura 2. Ligantes de DNA.....	04
Figura 3. Enantiômeros de uma base de Tröger simétrica com substituintes proflavina.....	05
Figura 4. Bases de Tröger que interagem com o DNA.....	06
Figura 5. Exemplo do conceito dos algoritmos genéticos.....	10
Figura 6. Esquema de funcionamento do algoritmo genético no <i>AutoDock 4.0</i>	11
Figura 7. A-DNA e B-DNA.....	14
Figura 8. Perfis de <i>clusters</i> para as docagens A-H.....	20
Figura 9. Conformações de melhor energia livre de interação para as docagens A-H..	22
Figura 10. Perfis de <i>clusters</i> para as docagens da bases de Tröger com quatro diferentes receptores.....	31
Figura 11. Conformações de melhor energia livre para as docagens das bases de Tröger com o DNA canônico com <i>gap</i> (2A-2D).....	32
Figura 12. Gráficos de RMSD.....	37
Figura 13. Número de pares de bases ao longo das simulações para os oligonucleotídeos simulados.....	39
Figura 14. Grupamentos monitorados para a Tröger simétrica (a) e para a Tröger assimétrica (b) na análise das distâncias de interação com o DNA.....	40
Figura 15. Evolução temporal do complexo de intercalação da Tröger simétrica (COMPL-INT-SIM).....	41
Figura 16. Evolução temporal do complexo de intercalação da Tröger assimétrica (COMPL-INT-ASS).....	42
Figura 17. Evolução temporal do complexo de ligação ao sulco da Tröger simétrica (COMPL-SUL-SIM).....	45
Figura 18. Evolução temporal do complexo de ligação ao sulco da Tröger assimétrica (COMPL-SUL-ASS).....	48
Figura 19. Ângulos torcionais da cadeia principal e da ligação glicosídica.....	51

Figura 20. Perfis dos ângulos α e γ para os complexos de intercalação com a Tröger simétrica (A) e com a Tröger assimétrica (B).....	53
Figura 21. Análise temporal dos ângulos α e γ	55
Figura 22. Perfis dos ângulos ε e ζ para os complexos de intercalação com a Tröger simétrica (A) e com a Tröger assimétrica (B).....	59
Figura 23. Análise temporal dos ângulos ε e ζ e do parâmetro $\varepsilon - \zeta$	61
Figura 24. Parâmetros de pares de bases.....	62
Figura 25. Modelo da transição $B \rightarrow A$ do DNA, por Calladine e Drew.....	63
Figura 26. <i>Rise</i> para o DNA-GAP (A) e para o COMPL-INT-SIM (B).....	66
Figura 27. <i>Rise</i> para o DNA-GAP (A) e para o COMPL-INT-ASS (B).....	66
Figura 28. <i>Rise</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B).....	67
Figura 29. <i>Rise</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B).....	67
Figura 30. <i>Twist</i> para o DNA-GAP (A) e para o COMPL-INT-SIM (B).....	68
Figura 31. <i>Twist</i> para o DNA-GAP (A) e para o COMPL-INT-ASS (B).....	68
Figura 32. <i>Twist</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B).....	69
Figura 33. <i>Twist</i> para o DNA-SEMGAP (A) e para o COMPL-SUL-ASS (B).....	69
Figura 34. <i>Slide</i> para o DNA-GAP (A) e para o COMPL-INT-SIM (B).....	70
Figura 35. <i>Slide</i> para o DNA-GAP (A) e para o COMPL-INT-ASS (B).....	70
Figura 36. <i>Slide</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B).....	71
Figura 37. <i>Slide</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B).....	71
Figura 38. <i>Roll</i> para o DNA-GAP (A) e para o COMPL-INT-SIM (B).....	72
Figura 39. <i>Roll</i> para o DNA-GAP (A) e para o COMPL-INT-ASS (B).....	72
Figura 40. <i>Roll</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B).....	73
Figura 41. <i>Roll</i> para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B).....	73
Figura 42. Combinação dos parâmetros <i>Roll</i> e <i>Tilt</i> para obtenção do parâmetro de <i>bending-dials</i>	75
Figura 43. Análise dos <i>bending-dials</i> dos sistemas contendo <i>gap</i>	78
Figura 43. Análise dos <i>bending-dials</i> dos sistemas sem <i>gap</i>	79

LISTA DE TABELAS

Tabela 1. Resultados para as docagens da Elipticina (EL) e da Netropsina (NE).....	18
Tabela 2. Valores médios de parâmetros estruturais da dupla hélice.....	23
Tabela 3. Resumo dos sistemas simulados.....	36

LISTA DE ABREVIATURAS

DNA	Ácido Desoxirribonucléico
NER	Reparo por Excisão de Nucleotídeos
LGA	Algoritmo Genético Lamarckiano
(d)A	(desoxi)adenilato
(d)T	(desoxi)timidilato
(d)G	(desoxi)guanilato
(d)C	(desoxi)citidilato
NE	Netropsina
EL	Elipticina
Int	Intercalação
Sulco	Ligação ao sulco menor
P-P	Largura do sulco menor
DNA-GAP	DNA com <i>gap</i> de intercalação
DNA-SEM-GAP	DNA sem <i>gap</i> de intercalação
COMPL-INT-SIM	Complexo de intercalação da Tröger simétrica
COMPL-INT-ASS	Complexo de intercalação da Tröger assimétrica
COMPL-SUL-SIM	Complexo de ligação ao sulco menor da Tröger simétrica
COMPL-SUL-ASS	Complexo de ligação ao sulco menor da Tröger assimétrica
PME	<i>Particle-Mesh Ewald</i>
ACPYPI	<i>AnteChamber Python Parses Interface</i>
GAFF	<i>Generalized Amber Force Field</i>
RESP	<i>Restrained Electrostatic Potential</i>
RMSD	Desvio quadrático médio
C	Carbono da ponte diazocina
N1	Grupamento amino 1
N2	Grupamento amino 2
Na	Grupamento amino
Nf	Nitrogênio da fenantrolina

RESUMO

Neste trabalho, a docagem e a dinâmica molecular foram utilizadas para investigar a interação entre oligonucleotídeos e duas bases de Tröger: uma base simétrica - contendo duas proflavinas – e uma base assimétrica – contendo uma proflavina e uma fenantrolina. Nas docagens que utilizaram DNA canônico com *gap* como receptor, os resultados reproduziram corretamente a estereosseletividade da interação, que favorece os isômeros levorrotatórios em relação aos dextrorrotatórios, e apontou para os modos de interação sugeridos pela literatura: ligação ao sulco menor para a Tröger simétrica e intercalação com ligação ao sulco menor (modo misto) para a Tröger assimétrica.

Para as simulações, foram escolhidos quatro complexos contendo os enantiômeros levorrotatórios das bases de Tröger e oligonucleotídeos com e sem *gap*. Embora os complexos representassem pontos de partida diferentes, houve convergência entre aqueles que continham o mesmo oligonucleotídeo como receptor. Os complexos com DNA com *gap* (complexos de intercalação) convergiram para um modo misto de interação, com intercalação de uma proflavina, enquanto os complexos com DNA sem *gap* (complexos de ligação ao sulco menor) convergiram para um modo de interação novo, em que as bases de Tröger se ligam ao sulco menor unicamente através da ponte diazocina, com os substituintes projetados para fora.

Os complexos de intercalação apresentaram tempos de residência altos (20 ns) e diminuíram o ângulo de torção da dupla hélice na região do *gap*, levando os ângulos α e γ da cadeia principal para uma região não-canônica, associada à forma A do DNA. Já nos complexos de ligação ao sulco menor, houve mudança tanto do modo como do sítio de interação e, embora tenham ocorrido eventuais curvaturas no DNA, não houve alterações significativas nos ângulos da cadeia principal, de modo que a estrutura global do DNA se manteve dentro da região considerada B-canônica.

Dessa forma, os resultados sugerem que a intercalação (com adicionais contatos no sulco menor) seja o modo preferencial de interação dessas bases de Tröger com o DNA. O modo de ligação ao sulco menor, por depender quase que exclusivamente da ponte diazocina, seria independente da quiralidade e, portanto, não seria capaz de explicar enantiosselevidade das bases de Tröger.

ABSTRACT

In this work, molecular docking and molecular dynamics simulations were applied to investigate the interaction between DNA oligomers and two Tröger bases: a symmetric Tröger base derived from proflavine and a mixed Tröger base containing proflavine and phenanthroline. The dockings that used canonical DNA with an artificial intercalation gap as receptor correctly reproduced the stronger DNA binding affinity of the levorotatory isomers when compared to the dextrorotatory ones. Moreover, the results pointed out to the binding modes suggested, yet not established, by experimental data: minor groove binding for symmetric Tröger base and intercalation with additional minor groove binding for the asymmetric one.

For the simulations, four complexes were selected, containing the (-)-isomers of the Tröger bases and DNA oligomers with or without intercalation gap. Interestingly, the systems containing the same DNA as receptor converged to similar binding modes: those containing DNA with gap (intercalation complexes) resulted in a mixed binding – with one proflavine intercalated and the other moiety occupying the minor groove – while those containing DNA without gap (minor groove complexes) lead to a new binding mode, in which the Tröger bases interact with the minor groove mainly through the diazocin bridge.

In the intercalation complexes, the ligands presented long residence times (20 ns) and also unwound the double helix in the region of the intercalation gap, leading some backbone angles (α and γ) to assume non-canonical values typical of the A form of DNA. In minor groove complexes, the ligands displayed enhanced mobility – changing both the binding mode and the binding site – and, although they promoted some bending of the double helix, the main chain torsion angles of DNA remained in the B-canonical region.

Altogether, these results suggest that intercalation of proflavine (with additional contacts in the minor groove) might be the stronger DNA binding mode for these Tröger bases. Since the minor groove binding depends almost exclusively of the diazocin bridge, it appears to perform independently of the ligand chirality and, therefore, could not explain the enantioselectivity of the interaction.

1. INTRODUÇÃO

1.1 LIGANTES DE DNA

Há algumas décadas, moléculas capazes de se ligar a ácidos nucleicos vêm sendo usadas como fármacos antibióticos e/ou antitumorais,¹⁻⁴ devido ao seu efeito citotóxico, além de serem importantes para o avanço das pesquisas em biologia molecular. Hoje, com o conhecimento do genoma de diversos organismos e tendo em mente os inúmeros processos celulares que envolvem diretamente ácidos nucleicos, grande atenção está voltada para a descoberta de novas moléculas capazes de interagir com o DNA.⁵ Essa tendência é reforçada pelo fato de que os agentes antitumorais que atuam sobre o DNA estão entre os mais eficazes em uso clínico atualmente, ainda que sejam pouco seletivos e, conseqüentemente, bastante tóxicos.⁶

Na maioria dos casos, embora o modo de interação de muitos fármacos com o DNA seja conhecido, ainda pouco se sabe sobre os mecanismos pelos quais essa interação gera seus efeitos biológicos. As hipóteses mais prováveis sugerem que as distorções provocadas na estrutura do DNA por esses ligantes interfiram em processos de replicação, transcrição ou no reparo do DNA (bloqueando a ação de enzimas como helicases, topoisomerasas, polimerases, etc).⁷ Trabalhos recentes com antitumorais da classe das ecteinascidinas, por exemplo, sugerem que a distorção provocada por essas moléculas à dupla hélice atrai enzimas do sistema de reparo por excisão de nucleotídeos (NER), que são estabilizadas no local da distorção, produzindo danos irreversíveis ao DNA.^{8,9}

Com relação ao modo de ligação, já está bem estabelecido que a grande maioria dos ligantes não peptídicos (fármacos ou agentes tóxicos e carcinogênicos) interage com o DNA através de dois mecanismos distintos: i) intercalação entre pares de bases; ii) ligação ao sulco menor (Figura 1).^{5,10} O sulco maior, embora ofereça mais sítios aceptores e doadores de ligação de hidrogênio, possui uma geometria demasiadamente larga e rasa para o encaixe de ligantes pequenos. Além disso, sendo o sítio de interação

preferencial para a maioria das proteínas,⁵ é provável que o sulco maior seja consideravelmente menos disponível do que o sulco menor no ambiente celular.

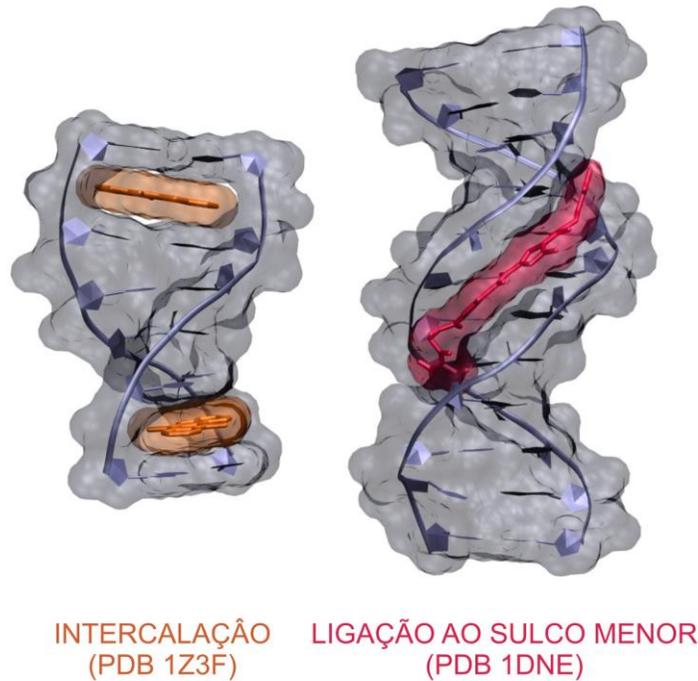


Figura 1. Principais modos de interação com o DNA, exemplificados com a elipticina (laranja) e com a netropsina (rosa).

Dentre os principais modos de interação de ligantes pequenos, a intercalação é o modo mais comum pelo qual moléculas rígidas e aromáticas reconhecem o DNA. Como a intercalação depende principalmente de interações do tipo π -stacking em conjunto com interações eletrostáticas, a maior parte dos intercaladores clássicos (tais como o etídeo e a proflavina (Figura 2A)) não apresenta significativa seletividade de sequência.^{11,12} No entanto, existem também intercaladores mais complexos, como a actinomicina¹³ e a equinomicina¹⁴ que, além de intercalar, fazem contatos adicionais com algum dos sulcos, resultando em modos de interação mistos com maior seletividade de sequência.

É também bem estabelecido que os agentes intercaladores impõem graves alterações estruturais no DNA ao abrir uma “lacuna” (ou *gap*) de intercalação entre dois pares de bases consecutivos.⁵ Além do próprio *gap*, a alteração mais evidente consiste em

uma significativa redução do ângulo de torção (ou *Twist*) da dupla hélice, efeito que é considerado uma assinatura deste modo de interação.^{5,11,15} Esse “desenrolamento” parcial do DNA decorre diretamente do aumento da distância entre pares de bases unidos por uma cadeia principal de comprimento finito e tem como consequência o aumento das interações π -stacking devido à melhor sobreposição entre os anéis aromáticos do intercalador e das bases nitrogenadas.

Portanto, devido a essas fortes alterações estruturais da dupla hélice, o modo de intercalação pode ser considerado, em uma analogia com as interações enzima-substrato, como uma interação do tipo *encaixe-induzido*.

Usando essa mesma analogia, a ligação de agentes como a netropsina e a distamicina (Figura 2B) ao sulco menor pode ser considerada mais semelhante a uma interação do tipo *chave-fechadura*, já que pouca ou nenhuma distorção aparente é observada na dupla hélice após a ligação dessas moléculas.^{16,17} De fato, essas poliamidas aromáticas apresentam uma forma naturalmente curvada que se encaixa facilmente à topologia da dupla hélice.¹² Essa complementaridade de formato permite um encaixe bem ajustado do ligante ao sulco menor, reforçando as interações de van der Waals com as paredes do sulco, bem como as interações eletrostáticas entre grupamentos catiônicos do ligante e os fosfatos da cadeia principal.^{10,18}

Além disso, a netropsina e a distamicina pertencem a uma classe de ligantes de sulco que atraiu considerável atenção da comunidade científica em função da sua significativa seletividade por sequências ricas em adenina e timina, mostrando que é possível obter ligantes de DNA sequência-específicos.^{12,19} De acordo com Bailly e Chaires,¹⁸ essa seletividade se deve a uma combinação de diferentes fatores: (i) ligações de hidrogênio entre os grupamentos NH da poliamida e as bases nitrogenadas (O2 da timina e N3 da adenina); (ii) adequada complementaridade geométrica favorecendo as interações de van der Waals com as paredes do sulco; e (iii) interações eletrostáticas.

No entanto, o fator preponderante parece ser o fator geométrico,¹⁷ pois sequências ricas em adenina e timina produzem um estreitamento do sulco menor de 4-5 Å em relação a sequências ricas em guanina e citosina.¹⁰ Além disso, a presença de um grupamento NH₂ apontando para o sulco menor nas guaninas (N2) torna o sulco menor de sequências poli(CG) mais raso em relação ao sulco menor de sequências poli(AT).¹⁶

Ou seja, o aumento das interações de van der Waals em um sulco estreito e profundo é considerado como a principal origem da seletividade de sequência das poliamidas aromáticas, que por isso são chamadas de agentes *formato*-seletivos.^{5,10,16,20}

Em resumo, intercalação e ligação ao sulco menor consistem nos principais modos de interação de moléculas pequenas com os ácidos nucleicos. Claramente, a intercalação é um mecanismo do tipo de *encaixe-induzido*, enquanto que a ligação não-covalente de agentes *formato*-seletivos parece seguir um mecanismo tipo *chave-fechadura*. Existem ainda ligantes capazes de se ligar covalentemente ao sulco menor - como a antramicina²¹ e a saframicina A²² - e ligantes que combinam mais de um modo de ligação - como, a actinomicina,¹³ a equinomicina¹⁴ e a aflatoxina B₁.²³ No entanto, por se tratarem de modos de interação mais complexos, não é possível classificar de um modo geral a interação desses ligantes em *encaixe-induzido* ou *chave-fechadura*, mesmo no caso de agentes estruturalmente relacionados.

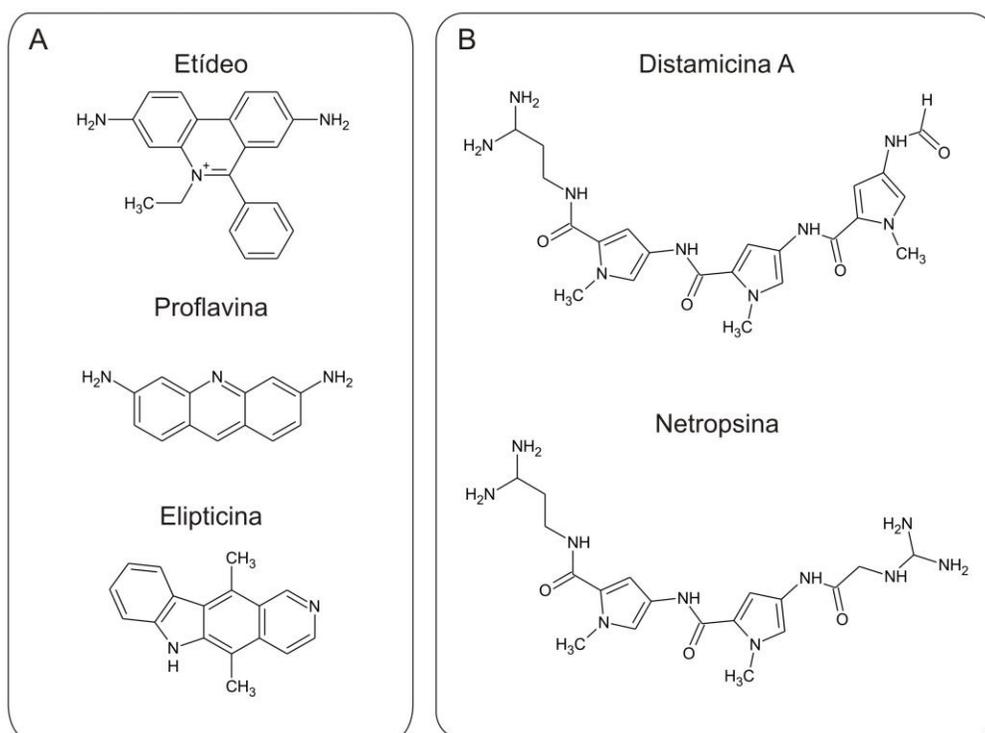


Figura 2. Ligantes de DNA. A) Intercaladores clássicos; B) Ligantes de sulco menor formato-seletivos.

para esses compostos, já que ainda não foram realizados ensaios a fim de compreender os efeitos biológicos dessa interação (citotoxicidade, dano ao DNA, mutagenicidade, etc).

Como pode ser observado na Figura 3, a geometria da base de Tröger garante aos seus derivados um formato de hélice que pode ter sentido igual ou oposto ao sentido das hélices do DNA. Em 1991, Yashima *et al.* já haviam sugerido que o enantiômero levorrotatório, sendo de mão direita, resultaria em um melhor encaixe às hélices de um B-DNA.²⁷ Essa previsão foi de fato confirmada por Tatiboüet *et al.*²⁸ e posteriormente por Bailly *et al.*²⁹ e Baldeyrou *et al.*,³⁰ que mostraram através de estudos de desnaturação térmica que o enantiômero (-)-(R,R) possui maior afinidade pelo B-DNA do que o enantiômero (+)-(S,S).

A Figura 4 mostra as duas bases de Tröger cuja interação com o DNA foi mais bem investigada até o momento: a) uma base de Tröger simétrica (com dois substituintes proflavina) e b) uma base de Tröger assimétrica (com substituintes proflavina e fenantrolina). Embora a interação dessas moléculas com o DNA tenha sido confirmada por ensaios de desnaturação térmica, o modo como essas moléculas interagem com o DNA ainda não foi devidamente elucidado.

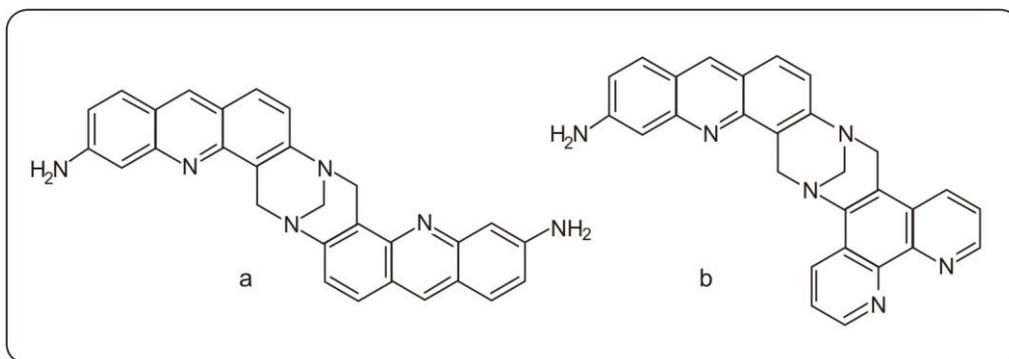


Figura 4. Bases de Tröger que interagem com o DNA. a) Base de Tröger simétrica contendo dois substituintes proflavina. b) Base de Tröger assimétrica contendo substituintes proflavina e fenantrolina.

Estudos de *DNase footprinting* revelaram que o enantiômero (-) da base simétrica apresenta moderada seletividade por determinadas sequências²⁹ e que o enantiômero (-) da base assimétrica apresenta moderada seletividade apenas quando em baixas concentrações.³⁰ Por outro lado, os enantiômeros (+) são completamente desprovidos de seletividade de sequência, o que sugere que os enantiômeros (-) e (+) de cada derivado

apresentam modos de interação diferentes com o DNA.^{29,30} Além disso, estudos bioquímicos e dados espectroscópicos sugerem que o enantiômero (-) da base de Tröger simétrica interage com o DNA através de ligação não covalente ao sulco menor, apesar de conter substituintes conhecidos como intercaladores.²⁹ Estudos semelhantes realizados por Baldeyrou *et al.* sugerem que o enantiômero (-) da Tröger assimétrica interage com o DNA através de um mecanismo misto (com intercalação da proflavina e encaixe da fenantrolina no sulco menor).³⁰

Neste contexto, Abella *et al.* sintetizaram bases de Tröger contendo 2-(4'-amino-2'-hidroxifenil)benzazóis que, devido a sua fluorescência, poderiam ser explorados como novas sondas de DNA.³³ Apesar de não conterem substituintes sabidamente capazes de interagir com ácidos nucleicos, experimentos preliminares mostraram que estes derivados fluorescentes são capazes de interagir com o DNA (comunicação pessoal do Prof. Valter Stefani, UFRGS). Isso não só ressalta o potencial desses compostos como moléculas marcadoras, como levanta a hipótese de que possivelmente seja a geometria intrínseca das bases de Tröger, e não os seus substituintes, que garante a interação com os ácidos nucleicos.

A elucidação dos mecanismos moleculares envolvidos na interação de diferentes derivados de Tröger com o DNA é certamente de grande importância para descobrir se existe um mecanismo único de interação entre derivados de Tröger e o DNA - o que resultaria em uma nova classe de ligantes - ou se a estrutura em “V” da Tröger consiste em um versátil arcabouço capaz de modificar, combinar ou mesmo potencializar as propriedades de interação dos seus substituintes.

2. OBJETIVOS

2.1 GERAIS

Elucidar computacionalmente os mecanismos moleculares envolvidos na interação entre derivados de Tröger e o DNA, visando propor modos de interação mais prováveis e determinar possíveis sítios ou sequências preferenciais de interação.

2.2 ESPECÍFICOS

Construir e validar um protocolo para adaptar a metodologia de docagem molecular aos ácidos nucléicos, utilizando para isso ligantes de DNA com modo de interação conhecido.

Obter, através de docagem molecular, complexos prováveis entre bases de Tröger e segmentos de DNA para serem utilizados como ponto de partida nas simulações de dinâmica molecular.

Investigar a estabilidade dos complexos entre bases de Tröger e oligonucleotídeos através de simulação por dinâmica molecular, bem como as alterações estruturais provocadas pela ligação desses compostos à dupla hélice.

3. ESTUDOS DE DOCAGEM MOLECULAR

3.1 INTRODUÇÃO

3.1.1 Docagem Molecular (baseada em Kitchen *et al.*;³⁴ Moitessier *et al.*³⁵)

A docagem molecular (do termo em inglês *molecular docking*) consiste em um método computacional que “atraca” ligantes pequenos a seus alvos macromoleculares, e pontua (*scores*) a potencial complementaridade entre eles.

Basicamente, um programa de docagem é composto por um algoritmo de busca e por uma função de escore. O algoritmo de busca explora o espaço de fases, buscando possíveis conformações e orientações do ligante, enquanto a função de escore pontua os resultados obtidos pelo algoritmo de busca. Dessa forma, a função de escore não somente direciona o processo de *docagem* para o modo de ligação mais favorável, como também fornece um *ranking* dos resultados obtidos ao final do processo e, muitas vezes, é formulada para prever a atividade biológica dos ligantes.

Os algoritmos de busca podem ser divididos em três classes: métodos de simulação, métodos sistemáticos e métodos estocásticos. Os métodos de simulação incluem a simulação por dinâmica molecular e métodos de minimização de energia, mas, embora permitam um tratamento detalhado do sistema, dificilmente são capazes de transpor barreiras energéticas entre mínimos locais em um tempo computacionalmente viável. Dessa forma, geralmente são utilizados apenas como algoritmos de busca local para complementar algoritmos de busca global.

Os métodos sistemáticos se propõem a explorar sistematicamente todo o espaço de fases, mas enfrentam o problema da explosão combinatorial quando o ligante é uma molécula flexível. Já os métodos estocásticos operam aplicando alterações aleatórias para cada ligante, que são avaliadas por uma função de escore e podem ou não ser aceitas em cada etapa. O programa *AutoDock*,³⁶ por exemplo, oferece tanto o algoritmo de Monte Carlo conhecido como *Simulated Annealing* como também algoritmos genéticos, que adaptam princípios de competição biológica e de dinâmica de populações e que têm se mostrado promissores na exploração aleatória do espaço de fases.

3.1.2 Algoritmos Genéticos e o *AutoDock 4.0* (baseado em Morris *et al.*³⁶)

Fundamentalmente, os algoritmos genéticos se baseiam no conceito evolutivo de que a solução para um determinado problema adaptativo se encontra espalhada no *pool* gênico de uma população de indivíduos. Esse conceito pode ser melhor compreendido através do exemplo mostrado na Figura 5, que trata de uma população de predadores carnívoros. Para esses predadores, seria vantajoso possuir dentes bastante pontiagudos e uma visão aguçada, características fenotípicas que são codificadas pelos genes azuis e vermelhos, respectivamente. A solução para o problema evolutivo neste exemplo seria, portanto, encontrar um indivíduo que possuísse os dois alelos ativos para ambas as características, o que lhe daria dentes muito pontiagudos e visão muito aguçada. Embora inicialmente não exista nenhum indivíduo com este genótipo, podemos ver que a solução está presente na população como um todo, pois uma operação de *crossing-over* entre os indivíduos C e D poderia levar a um indivíduo com o genótipo ideal. O papel do algoritmo genético é justamente aplicar operadores genéticos (como *crossing-over* e mutação gênica) em uma população inicial, produzindo novas gerações de indivíduos até que uma solução satisfatória seja encontrada.

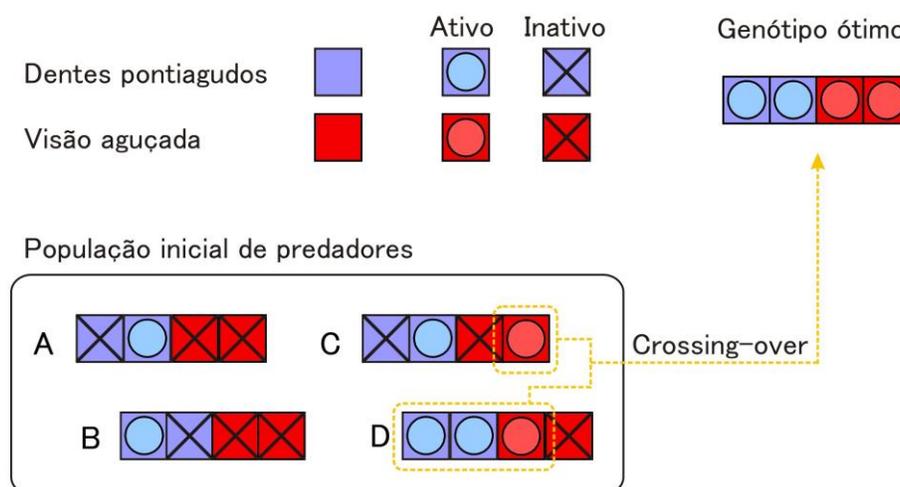


Figura 5. Exemplo do conceito dos algoritmos genéticos. Adaptado de <http://www.rennard.org/alifeenglish/gavintrgb.html>.

No contexto da docagem, cada indivíduo corresponde a uma determinada orientação do ligante no espaço e a solução adaptativa corresponde à melhor orientação do ligante para interagir com o receptor. Cada indivíduo, portanto, é representado por um arquivo (que pode ser considerado como um *cromossomo*), que contém variáveis (ou *genes*) de translação, orientação e conformação. Esse arquivo contendo as variáveis de estado é considerado, em uma analogia com a genética, o *genótipo* do indivíduo, e pode ser “traduzido” para um arquivo de coordenadas atômicas, considerado o *fenótipo*.

Basicamente, os algoritmos genéticos operam de acordo com o esquema mostrado na Figura 6. Inicialmente é gerada uma população de indivíduos (ligantes) posicionados aleatoriamente no espaço. Em seguida, é feita uma estimativa da energia de interação de cada ligante com o receptor, a fim de selecionar os progenitores, ou seja, os indivíduos sobre os quais serão aplicados os operadores genéticos de mutação ou de *crossing-over*. Se os indivíduos gerados produzirem energias de interação mais favoráveis do que os progenitores, eles são selecionados para substituir os piores indivíduos da geração anterior. Dessa forma, são realizados diversos ciclos de gerações até que se atinja o número máximo de gerações ou de avaliações de energia, ambos definidos pelo usuário.

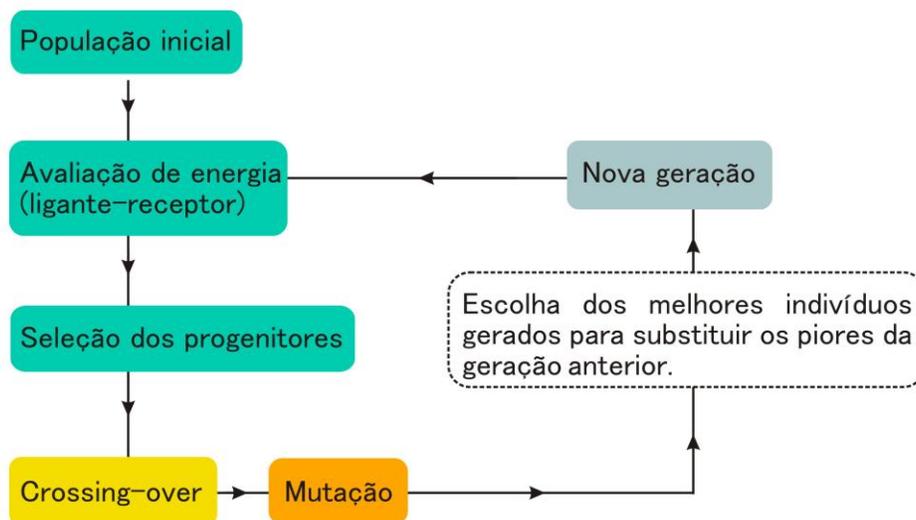


Figura 6. Esquema de funcionamento do algoritmo genético no *AutoDock 4.0*. Adaptado de Morris *et al.*³⁶

Em muitos programas de docagem – não apenas os que usam algoritmos genéticos - a avaliação das conformações obtidas é feita através de funções de escore que foram parametrizadas para estimar energias livres de interação. Essas funções são consideradas semi-empíricas, pois combinam expressões extraídas de campos de força da mecânica molecular, com coeficientes empíricos obtidos a partir de um conjunto de dados experimentais – constantes de afinidade entre ligantes e receptores, por exemplo. Algumas funções de escore estimam também as contribuições entrópicas, como a função de escore do *AutoDock 4.0* que possui, além dos termos entálpicos, que estão baseados no campo de força *AMBER*, dois termos completamente empíricos formulados para estimar a perda de mobilidade conformacional do ligante e a reorganização do solvente após a ligação³⁷. Apesar de ser mais fácil incluir termos entrópicos de forma empírica, a desvantagem desta abordagem é que a eficiência e a confiabilidade das funções de escore empíricas e semi-empíricas são fortemente dependentes do conjunto de dados experimentais utilizados na parametrização.

3.1.3 Docagem aplicada a ácidos nucleicos

A dependência do conjunto de parametrização utilizado é uma das principais limitações da docagem, devido ao fato de que a maior parte das funções de escore disponíveis atualmente foram parametrizadas com proteínas, e os programas validados apenas para essa classe de moléculas e seus ligantes. Muito provavelmente por essa razão, existe ainda um número muito pequeno de trabalhos utilizando a docagem molecular para ácidos nucleicos,³⁸⁻⁴⁸ apesar de o DNA ser sabidamente alvo de uma série de fármacos antibióticos e antitumorais.

Como se sabe, os ácidos nucleicos são estruturalmente bastante diferentes das proteínas e possuem características únicas, como alta densidade de carga e geometria quiral. Além disso, os ácidos nucleicos não apresentam um sítio de ligação único e bem definido para cada ligante, como ocorre com as proteínas. Como consequência, isso nos leva a questionar se os programas de docagem validados para proteínas também produzem resultados razoáveis em docagens com ácidos nucleicos. Essa questão foi recentemente abordada por Holt *et al.*,⁴⁸ que mostraram que o programa *AutoDock*

reproduz corretamente a estrutura cristalográfica de diversos ligantes de sulco e intercaladores ligados ao DNA, com uma resolução de aproximadamente 2Å. No entanto, embora as redocagens (ou seja, docagens usando a estrutura cristalográfica original do receptor) sejam consideradas úteis como uma primeira indicação da acurácia da docagem, esse procedimento tem se mostrado pouco útil na descoberta de novos ligantes.³⁵ Portanto, o emprego de métodos de docagem com DNA para a elucidação de mecanismos de ligação desconhecidos – ou seja, sem dados experimentais prévios e conclusivos – permanece um desafio.

Além do mais, o DNA é uma molécula bastante flexível que pode assumir isoformas estruturalmente distintas,^{49,50} como ilustrado na Figura 7, que mostra as isoformas A e B do DNA. Podemos ver nestas estruturas que a forma A é consideravelmente mais achatada em comparação com a forma B. Além disso, o B-DNA possui sulcos distintos em largura, porém semelhantes em profundidade, enquanto que, no A-DNA, os sulcos diferem bastante em profundidade, sendo o sulco maior muito profundo e o sulco menor bastante raso. Essa diferença de profundidade dos sulcos na forma A decorre de um afastamento dos pares de bases em relação ao eixo central da dupla hélice, que pode ser observado na vista de cima dessa estrutura, onde fica evidente a presença de uma “cavidade” central, que não está presente na forma B.

Apesar de a forma B ser a mais comum em ambiente fisiológico, existem evidências de que a forma A - assim como outras formas ainda mais diferentes, como o Z-DNA⁵¹ – também ocorram no ambiente celular, induzidas pelo estresse topológico ou pela presença de outras moléculas.⁵² Além das transições conformacionais, a forma B do DNA também está sujeita a distorções locais, como curvaturas e desenovelamentos da dupla hélice, que são de grande importância para a interação com proteínas e outros ligantes.^{13,17,53,54} Muitas vezes, a flexibilidade intrínseca de uma sequência para sofrer uma determinada distorção pode ser o fator preponderante no reconhecimento desta sequência por outra molécula.^{17,55} Diferentemente da seletividade sequencial que decorre somente do padrão de ligações de hidrogênio – chamado de leitura direta (*direct readout*) –, esse tipo de seletividade pelo formato da dupla hélice é chamado de leitura indireta (*indirect readout*) e decorre fundamentalmente da flexibilidade do DNA.

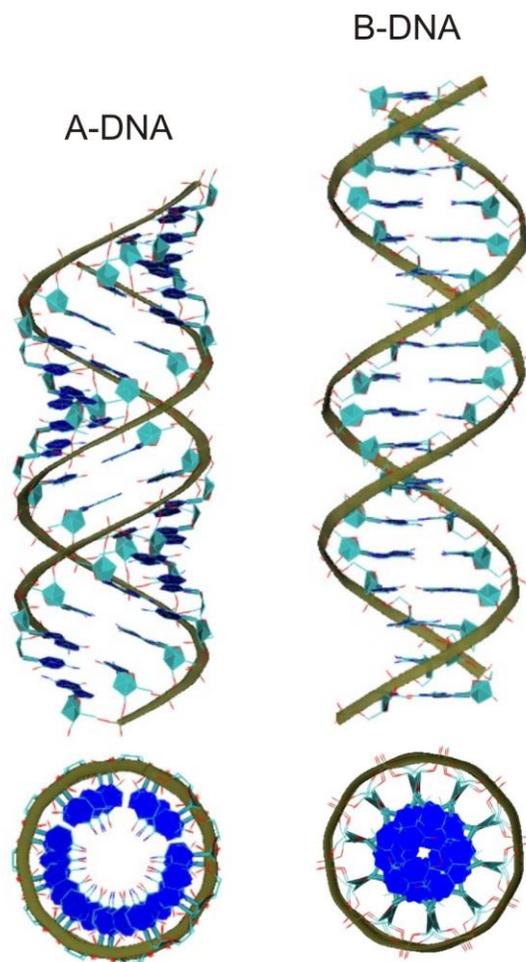


Figura 7. A-DNA e B-DNA. Acima - vistas laterais; abaixo - vistas de cima. As estruturas foram geradas com o X3DNA⁵⁶.

Considerando agora que a maioria dos programas atuais de docagem ainda não leva em consideração a flexibilidade da macromolécula – tratando o receptor como rígido ou semi-rígido – está claro que esse tipo de reconhecimento indireto será completamente desconsiderado, a não ser que se conheça de antemão o tipo de distorção necessária para a interação com o ligante – como no caso de um agente intercalador. Isso nos leva a uma outra questão pouco explorada até o momento: quando se trata de ligantes cujo modo de ligação ao DNA ainda não foi elucidado, que conformação de oligômero deve ser usada como receptor? Claramente, diversos aspectos estruturais da conformação escolhida irão interferir na performance da docagem, especialmente quando não só o reconhecimento

sequencial depende da flexibilidade, mas também a própria interação, como no caso de um mecanismo do tipo encaixe-induzido.

Nesse contexto, nós decidimos utilizar o programa *AutoDock 4.0* para realizar não apenas redocagens como também docagens cruzadas entre dois ligantes (elipticina e netropsina) e quatro oligômeros estruturalmente distintos (dois canônicos e dois cristalográficos) a fim de estabelecer um protocolo padrão capaz de identificar dois dos principais modos de ligação ao DNA: intercalação e ligação ao sulco menor. Posteriormente, esse protocolo será utilizado para a docagem das duas bases de Tröger mais bem estudadas experimentalmente até o momento (ver Figura 3), para as quais existem modos de interação sugeridos pela literatura, embora de forma não conclusiva.

3.2 METODOLOGIA

As docagens foram realizadas com o programa *AutoDock 4.0*, que utiliza uma função de escore empírica baseada na energia livre de ligação.^{36,37} Entre os algoritmos estocásticos oferecidos pelo pacote do *AutoDock*, foi escolhido o algoritmo genético lamarkiano (LGA), que combina um algoritmo de busca global (algoritmo genético) com um algoritmo de busca local (algoritmo de Solis e Wets⁵⁷).

Para a fase de validação, dois complexos cristalográficos foram obtidos do Protein Data Bank: um hexâmero d(CGATCG)₂ complexado com uma molécula de elipticina (1Z3F¹⁵) e um dodecâmero d(CGCGATATCGCG)₂ complexado com uma molécula de netropsina (1DNE⁵⁸). Como anteriormente mencionado, a elipticina é típico agente intercalador com atividade antitumoral, enquanto a netropsina é um antibiótico de origem natural que interage com o DNA através de ligação ao sulco menor (ver seção 1.1).

Etapa 1: redocagens

Primeiramente, a elipticina e a netropsina foram docadas nos oligômeros dos complexos cristalográficos 1Z3F e 1DNE, respectivamente. Para isso, foi feita a separação das coordenadas atômicas dos ligantes e dos oligonucleotídeos, e os hidrogênios polares e aromáticos foram adicionados com o pacote *GROMACS*.⁵⁹ A

seguir, foram atribuídas cargas de Gasteiger-Marsili⁶⁰ para o DNA e para os ligantes com o programa *AutoDock Tools (ADT)*.⁶¹

Utilizando o *ADT*, foi criada uma malha de afinidade com 96x96x96 pontos e uma resolução de 0,375 Å, de modo a englobar o fragmento inteiro de DNA. Após a malha de afinidade ter sido centralizada na macromolécula (DNA), os mapas de afinidade entre os átomos do ligante e da macromolécula foram gerados utilizando o módulo *AutoGrid 4.0*.

Após diversos testes variando parâmetros da docagem a fim de otimizar o processo,⁶² foi estabelecido um protocolo padrão em que cada docagem consiste de 25 corridas, com uma população inicial de 50 indivíduos, 50 x 10⁶ avaliações de energia e um número máximo de 27.000 gerações. Os operadores de mutação e *crossing-over* foram aplicados com taxas de 0,02 e 0,80, respectivamente. Para o algoritmo de busca local, foi escolhido o algoritmo de pseudo Solis e Wets⁵⁷, com um passo translacional de 0,2 Å, um passo de orientação de 5° e um passo torsional de 5°. Para o restante dos parâmetros, foram utilizados valores *default* do programa.

As conformações resultantes que diferiram por menos de 2,0 Å em RMSD foram agrupadas em um mesmo *cluster*, que é representado pela conformação energeticamente mais favorável dentro do *cluster*. Como a malha de afinidade incluiu não apenas o sítio de ligação, mas todo o fragmento de DNA, não foi possível utilizar o RMSD como critério de acurácia. O critério utilizado foi uma classificação do modo de ligação através de inspeção visual em: ligação ao sulco menor, intercalação ou outros (ligação ao sulco maior, interação com os fosfatos da cadeia lateral, etc).

Etapa 2: docagens diretas

Nessa etapa, os receptores cristalográficos foram substituídos por oligonucleotídeos canônicos (B-DNA) gerados pelo programa *X3DNA*,⁵⁶ com sequências semelhantes às dos oligonucleotídeos cristalográficos. Como a macromolécula não pode sofrer nenhuma alteração estrutural durante a docagem, um dos oligômeros canônicos foi modificado para incluir uma lacuna (*gap*) de intercalação. Para isso, o programa *Swiss PDB Viewer*⁶³ foi utilizado para posicionar uma molécula de elipticina entre dois pares de bases do DNA canônico, paralela à orientação dos anéis das bases nitrogenadas. A seguir,

o complexo foi submetido a uma minimização de energia com o método *steepest descent*, utilizando o pacote *GROMACS* e o campo de força GROMOS 53A6⁶⁴. A minimização resultou em um DNA modificado, com os pares de bases que flanqueiam a elipticina separados por 6,50 Å, o que caracteriza uma lacuna de intercalação. A molécula de elipticina foi removida do complexo e o DNA modificado (ver Figura 9 B/F) foi usado como receptor para a docagem direta da elipticina. As sequências dos oligômeros canônicos são d(CGCAATTCGC)·(GCGAATTGCG) (sem *gap*) e d(CG**GC**ATGCCG)₂ (com *gap*, indicado em negrito).

Etapa 3: docagens cruzadas

Nesta última etapa de validação, foram realizadas docagens cruzadas: a elipticina foi docada com os oligonucleotídeos previamente utilizados como receptores para a netropsina (sem *gap*) enquanto a netropsina foi docada com os oligonucleotídeos previamente utilizados como receptores para a elipticina (com *gap*).

Etapa 4: aplicação

Os receptores testados na etapa de validação foram utilizados para a docagem de duas bases de Tröger: uma base simétrica com substituintes proflavina e uma base assimétrica com substituintes proflavina e fenantrolina. Sendo as bases de Tröger compostos quirais, foram construídos os dois enantiômeros (-)-(R,R) e (+)-(S,S) utilizando o programa *GaussView*.⁶⁵ As geometrias foram otimizadas com o programa *Gaussian*,⁶⁶ através de um cálculo Hartree-Fock restrito usando a base 6-31G*. A preparação dos ligantes foi feita como descrito anteriormente e a malha de afinidade foi construída novamente de modo a incluir toda a estrutura do receptor. Os parâmetros da docagem foram os mesmos empregados nas etapas de validação, exceto pelo aumento do número de corridas para 100 em vez de 25.

3.3 RESULTADOS E DISCUSSÃO ⁶²

Dentre as 25 corridas realizadas em cada docagem na etapa de validação, foram analisadas as dez mais favoráveis energeticamente, ou seja, aquelas que apresentaram as energias livres de interação mais negativas. A Tabela 1 apresenta um resumo dos resultados para as docagens da elipticina (A, B, C e D) e da netropsina (E, F, G e H) com os quatro diferentes oligômeros. A/E são docagens com DNA cristalográfico do complexo elipticina-DNA (com *gap*); B/F são docagens com B-DNA canônico modificado para conter um *gap* de intercalação; C/G são docagens com DNA cristalográfico do complexo netropsina-DNA (sem *gap*); e D/H são docagens com B-DNA canônico (sem *gap*). Dessa maneira, A, B, G e H são docagens diretas, enquanto C, D, E e F são docagens cruzadas. Entre as docagens diretas, apenas as docagens A e G são redocagens. Os perfis de *clusters* estão mostrados na Figura 8 e as conformações que resultaram nas melhores energias de interação para cada docagem estão ilustradas na Figura 9.

Tabela 1. Resultados para as docagens da Elipticina (EL) e da Netropsina (NE).

Docagem	Ligante	DNA	N_{sulco} (%)	N_{int} (%)	$\Delta G_{mínima}$	$\Delta G_{média}$
A	EL	Cristalográfico (com <i>gap</i>) ¹	0	100	-8,71 (int)	-8,71
B	EL	Canônico (com <i>gap</i>)	50	50	-8,10 (int)	-7,25
C	EL	Cristalográfico (sem <i>gap</i>) ²	100	0	-7,98 (sulco)	-7,98
D	EL	Canônico (sem <i>gap</i>)	100	0	-6,51 (sulco)	-6,51
E	NE	Cristalográfico (com <i>gap</i>) ¹	20	70	-7,62 (int)	-6,74
F	NE	Canônico (com <i>gap</i>)	90	0	-9,85 (sulco)	-7,17
G	NE	Cristalográfico (sem <i>gap</i>) ²	100	0	-9,97 (sulco)	-9,13
H	NE	Canônico (sem <i>gap</i>)	100	0	-9,47 (sulco)	-8,70

N_{sulco} , porcentagem de corridas resultando em sulco menor; N_{int} , porcentagem de corridas resultando em intercalação, $\Delta G_{mínima}$, energia livre de interação mais favorável da docagem (kcal/mol); $\Delta G_{média}$ energia livre de interação média da docagem (kcal/mol).

¹ original do complexo elipticina-DNA

² original do complexo netropsina-DNA

3.3.1 Docagens diretas

Como esperado, os resultados mostram que o *AutoDock* é bastante eficiente para predizer os modos de ligação da elipticina e da netropsina em redocagens (Tabela 1, docagens A e G). A Figura 8 mostra que todas as corridas com a elipticina resultaram em intercalação, com uma energia livre de interação média de -8,71 kcal/mol (8A), e que todas as corridas com a netropsina resultaram em ligação ao sulco menor, com uma energia livre de interação média de -9,13 kcal/mol (8G).

No entanto, quando os receptores cristalográficos originais foram substituídos por DNA canônico de sequência similar, tendências diferentes foram observadas para cada tipo de ligante (Tabela 1, docagens B e H). A docagem da elipticina com DNA canônico (com *gap*) mostra que a intercalação representou neste caso apenas 50% das corridas (Figura 8B), embora continue sendo o modo de ligação preferencial em termos de energia livre de interação (-8,10 kcal/mol). Já a docagem com a netropsina com B-DNA canônico (sem *gap*) não mostrou nenhuma mudança na porcentagem de corridas que resultaram em ligação ao sulco menor (Figura 8H) em comparação com a docagem G. Com relação às energias livres de interação, o perfil de *clusters* da docagem H se mostrou bastante similar ao da docagem G, com uma energia livre de interação média muito favorável (-8,70 kcal/mol).

Esses resultados estão de acordo com os mecanismos característicos de cada modo de ligação. Considerando que a elipticina interage com o DNA através de um mecanismo do tipo encaixe-induzido que não pode ocorrer durante o processo de docagem, é natural que a docagem desta molécula dependa fortemente da conformação escolhida do receptor. A netropsina, por outro lado, sendo uma molécula bastante flexível e cuja flexibilidade é levada em conta durante a docagem, pode sofrer adaptações estruturais que levam a um encaixe bastante favorável no sulco menor de um DNA canônico, como mostrado pelos resultados da docagem H.

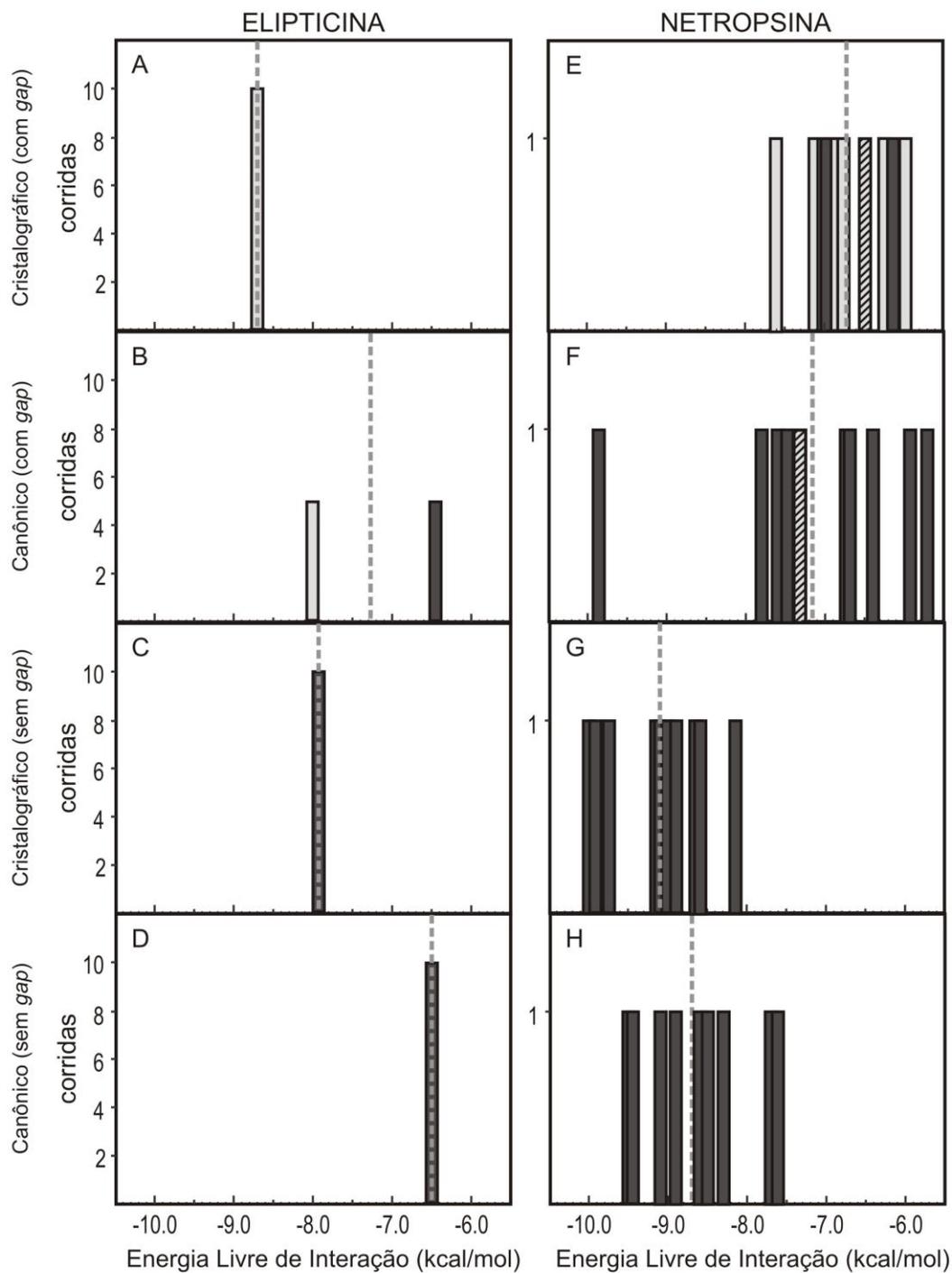


Figura 8. Perfis de *clusters* para as docagens A-H. Cada *cluster* é representado por uma barra em que a altura corresponde ao número de conformações do *cluster* e a cor corresponde ao modo de ligação: ■ sulco menor; □ intercalação; ▨ outro. A linha pontilhada indica a energia livre de interação média.

3.3.2 Docagens cruzadas

Nas docagens cruzadas, os aspectos mecânicos observados para a elipticina e a netropsina se tornaram ainda mais evidentes. Como esperado, os resultados das docagens C e D (Tabela 1) mostraram claramente que a docagem não foi capaz de reconhecer a intercalação quando o DNA não possuía um *gap* de intercalação adequado. Essa limitação de todos os programas de docagem que não levam em conta a flexibilidade do receptor já havia sido comentada na literatura.^{35,41} De fato, quando a elipticina foi docada com DNA cristalográfico (Figura 8C) ou canônico (Figura 8D) sem *gap* de intercalação, todas as corridas resultaram em ligação ao sulco menor, embora com energias livres de interação não tão favoráveis quanto as energias observadas para o redocagem (docagem A).

Por outro lado, as docagens cruzadas com a netropsina mostraram que a ligação ao sulco menor continuou sendo reconhecida como um possível modo de interação mesmo quando o DNA apresentava um *gap* de intercalação (Tabela 1, Figuras 8E e 8F). No entanto, os resultados da docagem E apontaram para intercalação como o modo de interação preferencial para a netropsina tanto quantitativamente (70% das corridas) como qualitativamente (melhor energia livre de interação). Ainda assim, é importante ressaltar que as energias livres de interação desta docagem foram consideravelmente menos favoráveis do que as energias das docagens G e H e, além disso, que a menor energia livre de interação (-7,62 kcal/mol) não foi tão negativa quando as energias resultantes da intercalação da elipticina nas docagens A (-8,71 kcal/mol) e B (-8,10 kcal/mol).

Em comparação com a docagem E, a docagem F apresentou um perfil de *clusters* bastante diferente (Figura 8F), pois a ligação ao sulco menor continuou sendo reconhecida como o modo de interação preferencial e mais acessado (90% das corridas), mesmo na presença do *gap* de intercalação. Além disso, diferentemente da docagem E, na docagem F nenhuma das corridas resultou em intercalação. A energia livre de interação média (-7,17 kcal/mol), embora não tenha sido comparável às energias das docagens G e H, foi ainda assim mais favorável do que as energias livres de interação da docagem E (-6,74 kcal/mol). Também notável foi a menor energia livre de interação resultante da docagem F, de -9,85 kcal/mol, portanto comparável às menores energias das docagens G

e H e também mais negativa do que a menor energia livre de interação resultante do modo de intercalação na docagem E (-7,62 kcal/mol).

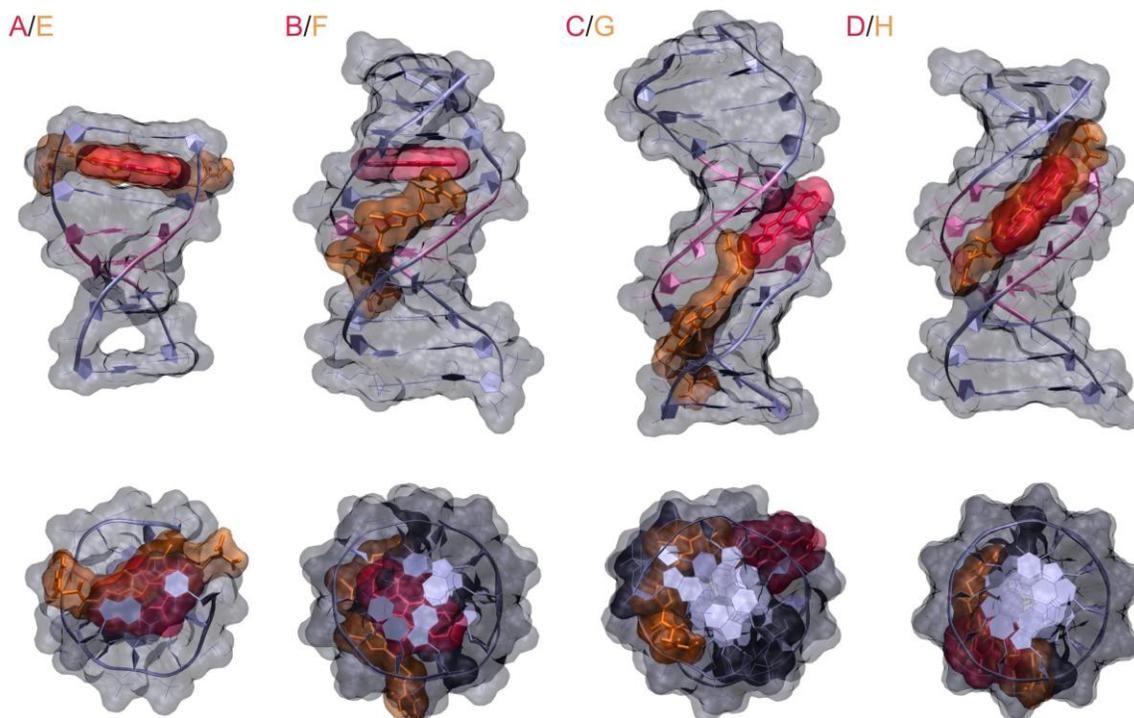


Figura 9. Conformações de melhor energia livre de interação para as docagens A-H. A elipticina é mostrada em rosa e a netropsina em laranja. No DNA, as regiões CG são mostradas em cinza e as regiões AT são mostradas em magenta.

3.3.3 Análise conformacional dos receptores

No intuito de esclarecer os resultados apresentados acima, foi utilizado o programa *X3DNA* para avaliar quantitativamente alguns parâmetros de hélice dos quatro oligômeros utilizados como receptores. Os resultados estão mostrados na Tabela 2, e as definições dos parâmetros *Twist* e *Rise* podem ser encontradas na seção 4.2.4.

Como pode ser concluído a partir dos valores de *Twist*, o DNA cristalográfico original do complexo com a elipticina está numa forma relativamente “desenovelada” ($29,43^\circ$) em comparação com o B-DNA canônico ($35,73^\circ$). De fato, é bem estabelecido que a intercalação impõe um desenovelamento da dupla hélice,^{5,15} que resulta em um

aumento das interações π -stacking devido a uma melhor sobreposição entre as bases nitrogenadas e os anéis do intercalador. Consequentemente, o valor bastante reduzido de *Twist* nesse oligômero é provavelmente um dos principais aspectos que favoreceram a intercalação nas docagens A e E. (É importante ressaltar que, embora a função de escore clássica do *AutoDock* não considere explicitamente a polarizabilidade molecular, os resultados das docagens A e B sugerem que interações do tipo π -stacking são suficientemente bem mimetizadas através de interações de van der Waals).

Tabela 2. Valores médios de parâmetros estruturais da dupla hélice.

DNA		<i>Twist</i> (°)	P-P (Å)	<i>Rise</i> (Å)
Cristalográfico	(com <i>gap</i>) ¹	29,43	16,4	6,88 *
Canônico	(com <i>gap</i>)	34,38	12,4	6,50 *
Cristalográfico	(sem <i>gap</i>) ²	36,22	11,1	3,42
Canônico	(sem <i>gap</i>)	35,73	11,7	3,35

Twist, ângulo de torção entre dois pares de bases consecutivos; P-P, largura do sulco menor; *Rise*, distância entre dois pares de bases consecutivos. A análise desses parâmetros foi feita com o *X3DNA*.

* valor para a os pares de bases que flanqueiam o *gap* de intercalação.

¹ original do complexo elipticina-DNA

² original do complete netropsina-DNA

Por outro lado, considerando que ligantes como a netropsina possuem uma curvatura natural compatível com a do B-DNA, é de se esperar que a afinidade da netropsina pelo DNA seja negativamente afetada por uma redução da torção da dupla hélice. Além disso, o desenovelamento do DNA também afeta a largura do sulco menor, como pode ser observado nas distâncias P-P da Tabela 2. Como anteriormente comentado, as dimensões do sulco menor são aspectos estruturais fundamentais para a ligação de poliamidas *formato-seletivas*, de modo que um sulco menor profundo e estreito permite um encaixe ajustado do ligante, aumentando as interações eletrostáticas e de van der Waals. De forma contrária, o sulco menor raso e largo do DNA do complexo com a elipticina (16,4 Å) diminui a superfície de contato entre a netropsina e as paredes do sulco, diminuindo as interações netropsina-DNA. Em outras palavras, a combinação de características estruturais marcantes observadas para esse oligômero (baixo *Twist* e

sulco menor muito largo) é provavelmente a principal razão por trás do modo de interação incorreto previsto na docagem E, em que dois dos anéis pirróis da netropsina estão inseridos no *gap* de intercalação (ver Figura 9, A/E).

Como também revelado pela Tabela 2, os parâmetros estruturais para os oligômeros sem *gap* de intercalação são bastante similares entre si. Como resultado, não é surpreendente que a docagem da netropsina a esses receptores tenha mostrado perfis de *clusters* semelhantes (docagens G e H).

Ainda assim, as distâncias P-P mostradas na Tabela 2 devem ser consideradas de forma criteriosa, pois elas podem levar à conclusão de que o sulco menor cristalográfico (11,1 Å) é apenas ligeiramente mais estreito do que o sulco menor canônico (11,7 Å). No entanto, as dimensões dos sulcos no DNA cristalográfico não são homogêneas ao longo da sequência, como ocorre no B-DNA canônico. Na realidade, de acordo com que está descrito na literatura, a região central de sequência ATAT no DNA cristalográfico apresenta um sulco menor significativamente mais estreito (10,8 Å) em comparação com as regiões periféricas de sequência CGCG (~13.2 Å) do mesmo oligômero. Essa acessibilidade diminuída da região central é provavelmente a razão pela qual a conformação de melhor energia livre na redocagem (docagem G) mostrou a netropsina docada abaixo da sequência ATAT, numa região em que o sulco não é tão estreito (ver Figura 9, C/G). Ou seja, embora se considere que um sulco menor estreito aumenta a interação netropsina-DNA no complexo formado, pode ser que essa conformação não apresente a geometria intermediária mais favorável para a *aproximação* do ligante.

Por outro lado, provavelmente porque o B-DNA canônico usado na docagem H apresenta uma geometria de hélice uniforme, com um sulco menor não tão estreito, a docagem conseguiu posicionar a netropsina corretamente na região central AATT (ver Figura 9, D/H), que é conhecida por ser uma sequência preferencial de ligação pelo padrão de doadores e aceptores de hidrogênio. Dessa forma, parece que em docagens com receptor rígido, as ligações de hidrogênio possuem um papel apenas secundário na predição do sítio de interação, que é determinado principalmente pela geometria das hélices.

Esse significativo estreitamento do sulco menor na região ATAT do DNA cristalográfico sem *gap* pode ser também a razão por trás das energias livres de interação

relativamente favoráveis para a elipticina na docagem C, já que a elipticina foi posicionada exatamente no local em que a distância P-P atinge o valor mínimo de 10.8 Å (ver Figura 9, C/G). O mesmo efeito não ocorreu quando a elipticina foi docada no DNA canônico (docagens B e D), pois nessas docagens as energias livres de interação para ligação ao sulco menor foram bem menos favoráveis do que as energias resultantes da intercalação, claramente indicando que a ligação ao sulco menor não é o modo de ligação preferencial para a elipticina, mas sim um artefato decorrente das limitações da docagem.

Com relação ao B-DNA modificado (com *gap*), a Tabela 2 mostra que esse oligômero apresenta um valor de *Twist* quase canônico (34,38 Å), e distâncias P-P apenas ligeiramente mais largas (12,4 Å) do que as distâncias de um sulco menor canônico (11,7 Å). Portanto, o B-DNA modificado é mais similar ao B-DNA canônico do que ao DNA cristalográfico do complexo com a elipticina, o que indica que o mecanismo artificial aplicado para abrir o *gap* de intercalação no B-DNA canônico impõe adaptações estruturais muito mais sutis do que as alterações causadas pela interação real com um intercalador. Sendo assim, é perfeitamente compreensível que a docagem da netropsina com esse oligômero tenha apontado para um modo de interação muito semelhante ao modo resultante das docagens com o B-DNA canônico e cristalográfico sem *gap*, como revelado pelos perfis de clusters das docagens F, G e H.

Finalmente, como pode ser observado a partir da comparação dos valores de *Rise* dos oligômeros, a altura do *gap* criado artificialmente (6,50 Å) não é tão alta quanto a do *gap* cristalográfico (6,88 Å), causado pela interação real com um intercalador. Essa comparação contribui para explicar porque os resultados para a elipticina na docagem B não foram tão satisfatórios quanto os resultados da docagem A, confirmando a proposta de que a docagem de intercaladores ao DNA é extremamente sensível ao formato do *gap*, que por sua vez depende fortemente de aspectos estruturais como o ângulo de torção da hélice (*Twist*) e altura entre pares de bases (*Rise*).

3.3.4 Comparação entre protocolos de docagem

A fim de avaliar a habilidade de cada protocolo de docagem na identificação dos modos de ligação, é interessante analisar cada um dos pares de docagem, ou seja, o mesmo receptor docado com dois diferentes ligantes, como está mostrado na Figura 9.

O oligômero cristalográfico com *gap* de intercalação usado nas docagens A/E claramente favoreceu a intercalação, provavelmente em função da sua estrutura fortemente distorcida em relação ao DNA canônico (Figura 9, A/E). Por outro lado, os oligômeros sem *gap* de intercalação usados nas docagens C/G e D/H não foram capazes de reconhecer a intercalação como um modo de ligação possível, resultando em ligação ao sulco menor para ambos os ligantes (Figura 9, C/G e D/H). No entanto, é importante salientar que o perfil de clusters para a netropsina na docagem E apresentou energias livres de interação que são, em média, menos favoráveis quando comparadas às energias de ligação ao sulco menor das docagens F, G e H ou quando comparadas às energias de intercalação de um verdadeiro intercalador, como na docagem A. Analogamente, o perfil de clusters (Figura 8) para a elipticina na docagem D mostrou energias livres de interação que são, em média, menos favoráveis quando comparadas às energias de intercalação das docagens A e B ou quando comparadas às energias de ligação ao sulco menor de um verdadeiro ligante de sulco na docagem H. Portanto, os perfis de clusters da Figura 8 indiretamente sugerem que o modo de interação previsto para a elipticina e para a netropsina nas docagens D e E não são os modos preferenciais de ligação para essas moléculas, mas artefatos decorrentes das limitações metodológicas da docagem.

Finalmente, a versão modificada do B-DNA (contendo *gap* de intercalação) que foi utilizada nas docagens B e F levou, para cada um dos ligantes, ao modo de interação correto (Figura 9, B/F). Esse oligômero mostrou-se, portanto, um receptor padrão promissor para a discriminação entre agentes intercaladores e ligantes de sulco em docagens rígidas.

3.3.5 Predição do modo de ligação das bases de Tröger

Os receptores testados na etapa de validação foram utilizados para a docagem de duas bases de Tröger. Dessa forma, cada um dos receptores foi docado com os quatro seguintes ligantes: A) enantiômero (-)-(R,R) da Tröger simétrica; B) enantiômero (+)-(S,S) da Tröger simétrica; C) enantiômero (-)-(R,R) da Tröger assimétrica; e D) enantiômero (+)-(S,S) da Tröger assimétrica.

Dentre as 100 corridas realizadas em cada docagem, foram analisadas as 25 que apresentaram as melhores energias livres de interação. As conformações resultantes de cada docagem foram classificadas por inspeção visual em cinco categorias: ligação parcial ao sulco menor (com apenas um dos substituintes da Tröger encaixado no sulco); ligação completa ao sulco menor (com ambos substituintes encaixados no sulco); intercalação; modo misto (com um dos substituintes intercalado e o outro encaixado em um dos sulcos) e intercalação parcial (com a Tröger paralela ao último par de bases do oligômero, na extremidade). Os perfis de *clusters* estão mostrados na Figura 10 e as conformações de melhor energia livre para as docagens com o DNA canônico com *gap* estão ilustradas na Figura 11. As conformações de melhor energia para o restante das docagens estão ilustradas no apêndice B, onde também podem ser observados os cinco modos de interação descritos acima.

Como pode ser observado na Figura 10, as docagens com o DNA canônico sem *gap* (1A-D) claramente apresentaram as energias livres de interação menos favoráveis do conjunto (-6,96 a -7,38 kcal/mol) e resultaram somente em modos de interação parciais (sulco menor parcial e intercalação parcial). Além disso, as energias livres de interação previstas para os dois enantiômeros de cada base foram extremamente próximas, indicando que a docagem com este receptor não é capaz de reproduzir estereosseletividade observada experimentalmente para as bases de Tröger.^{29,30}

Já a docagem com DNA cristalográfico sem *gap* (3A-D), embora tenha apresentado energias livres de interação consideravelmente mais favoráveis (-7,93 a -9,47 kcal/mol), ainda resultou principalmente em ligação parcial ao sulco menor. Além disso, as energias livres de interação previstas para os enantiômeros (+)-(S,S) (3B e 3D) foram mais negativas do que as energias livres de interação dos enantiômeros (-)-(R,R) (3A e

3C), o que indica que, também com este receptor, a docagem não é capaz de reproduzir a maior estereoespecificidade do enantiômero (-)-(R,R) em relação ao (+)-(S,S).

O mesmo não ocorreu com as docagens que utilizaram receptores (canônico ou cristalográfico) contendo *gap* de intercalação (2A-D e 4A-D). Tanto para as bases simétricas como para as bases assimétricas, os perfis de *clusters* com esses receptores mostraram energias livres de interação mais negativas para os enantiômeros levorrotatórios (-9,6 a -10,35 kcal/mol) do que para os dextrorrotatórios (-8.48 a -9.35 kcal/mol), o que está de acordo com a literatura.²⁸⁻³⁰ Além disso, a Figura 10 mostra para esses receptores um perfil de modos de interação (predominantemente intercalação e modo misto) diferentes em relação aos modos de interação previstos para os receptores sem *gap* (predominantemente ligação parcial ao sulco menor).

Portanto, essa análise preliminar do conjunto de docagens sugere que, por resultarem em melhores energias livres de interação, por reproduzirem a estereosseletividade dos enantiômeros e por fornecerem modos de interação mais completos, os receptores contendo *gap* de intercalação são os mais adequados para a investigação do modo de interação das bases de Tröger com o DNA. Particularmente interessante é o caso do receptor canônico com *gap* que, além de ter se mostrado a alternativa mais adequada para investigação de ligantes desconhecidos na etapa de validação, forneceu resultados bastante promissores com as bases de Tröger, que serão discutidos a seguir.

Como já comentado, as docagens usando esse receptor (2A-D) sugeriram uma maior afinidade do enantiômero levorrotatório, o que está plenamente de acordo com resultados de estudos de desnaturação térmica²⁸⁻³⁰ e coerente com a geometria intrínseca dos compostos, já que os enantiômeros (-) apresentam um formato de hélice de mão direita semelhante ao do B-DNA enquanto os enantiômeros (+) apresentam um formato de hélice de mão esquerda, contrário ao do B-DNA. De fato, essa complementaridade de formato é provavelmente a razão pela qual a docagem do enantiômero (-) da Tröger simétrica (2A) resultou na ligação dos dois substituintes proflavina ao sulco menor com uma energia livre de interação muito favorável (-9,90 kcal/mol) enquanto as docagens com o enantiômero (+) desta mesma base (2B) resultaram somente na intercalação de um

dos substituintes proflavina, com a outra proflavina projetada para fora do sulco maior (ver Figura 11, vistas de cima).

Embora a docagem deste mesmo receptor com o enantiômero (-) da Tröger simétrica (2A) também tenha resultado em intercalação (-9,67 kcal/mol, 7 corridas), a ligação ao sulco menor foi o modo de ligação preferencial tanto qualitativamente (-9,90 kcal/mol) como quantitativamente (18 corridas). Isso está de acordo com estudos de dicroísmo linear e com a ausência de efeitos de desenovelamento do DNA, ambos reportados por Bailly *et al.*, que sugerem que o enantiômero (-) da base de Tröger simétrica interage com o DNA através de ligação ao sulco menor.²⁹

Ainda assim, considerando que a ligação ao sulco menor ocorreu exatamente na região do *gap*, não deve ser descartada a hipótese de que esse resultado seja um artefato das alterações estruturais introduzidas pelo *gap* e que, por essa razão, não ocorreu na docagem com o B-DNA canônico sem *gap*.

Ainda em relação à Tröger simétrica, estudos de *DNase footprinting* provaram que apenas o enantiômero (-) apresenta seletividade de sequência, o que sugere um modo de interação diferente para o enantiômero (+).²⁹ Essa hipótese foi corroborada pelos resultados das docagens com o receptor canônico com *gap* (2B e 2D), que indicaram que os enantiômeros (+) das bases de Tröger interagiram com o DNA através de intercalação (um modo de ligação geralmente associado à falta de seletividade de sequência).

Considerando o enantiômero (-) da Tröger assimétrica (2C), todas as corridas com o receptor canônico com *gap* apontaram para um modo de ligação misto, com uma energia livre de interação muito negativa (-9,84 kcal/mol), em que o substituinte fenantrolina encontra-se intercalado enquanto o substituinte proflavina está encaixado no sulco menor (ver Figura 11). Mais uma vez, esse modo de ligação só foi possível porque a quiralidade do ligante é similar à quiralidade da dupla hélice. Por outro lado, as docagens do mesmo receptor com o enantiômero (+) da base assimétrica (2D) resultaram apenas em intercalação, com uma energia livre de interação de -9,35 kcal/mol, portanto não tão favorável quanto as energias do modo de interação misto do enantiômero (-).

Já havia sido sugerido que o enantiômero (-) da base assimétrica não interage através de ligação ao sulco menor, pois os ensaios de *DNase footprinting* mostraram seletividade de sequência apenas moderada para essa molécula.³⁰ Essa seletividade

moderada poderia ser explicada pela intercalação de um dos substituintes, como sugerido pelos resultados das docagens (2C e 4C) e também por ensaios de inibição de topoisomerase I, reportados por Baldeyrou *et al.*³⁰ No entanto, os resultados da docagem usando o receptor canônico com *gap* (2C) concordaram apenas parcialmente com os dados experimentais, já que estudos de dicroísmo circular e linear sugerem um modo misto de interação com o substituinte proflavina intercalado e o substituinte fenantrolina encaixado no sulco menor.

É possível que os resultados da docagem com o receptor canônico com *gap* (2C) tenham apontado para um modo misto alternativo (com a fenantrolina intercalada ao invés proflavina) devido às limitações de flexibilidade da docagem. Como discutido anteriormente, o sulco menor de um B-DNA canônico com *gap* artificial não é tão largo quanto o sulco menor de um DNA cristalográfico com *gap* decorrente da presença real de um intercalador. Em outras palavras, o sulco menor do DNA com *gap* artificial é provavelmente muito estreito para comportar o substituinte fenantrolina, o que significa que a conformação do receptor pode ter guiado a docagem 2C para um modo de interação análogo em que a proflavina – ao invés de fenantrolina – se ligou no sulco menor. De fato, essa hipótese é corroborada pelos resultados da docagem do enantiômero (-) da base assimétrica com o DNA cristalográfico com *gap* (4C), que apontou para o modo misto sugerido pela literatura (ver apêndice B), com uma energia livre de interação extremamente favorável, de -10,35 kcal/mol.

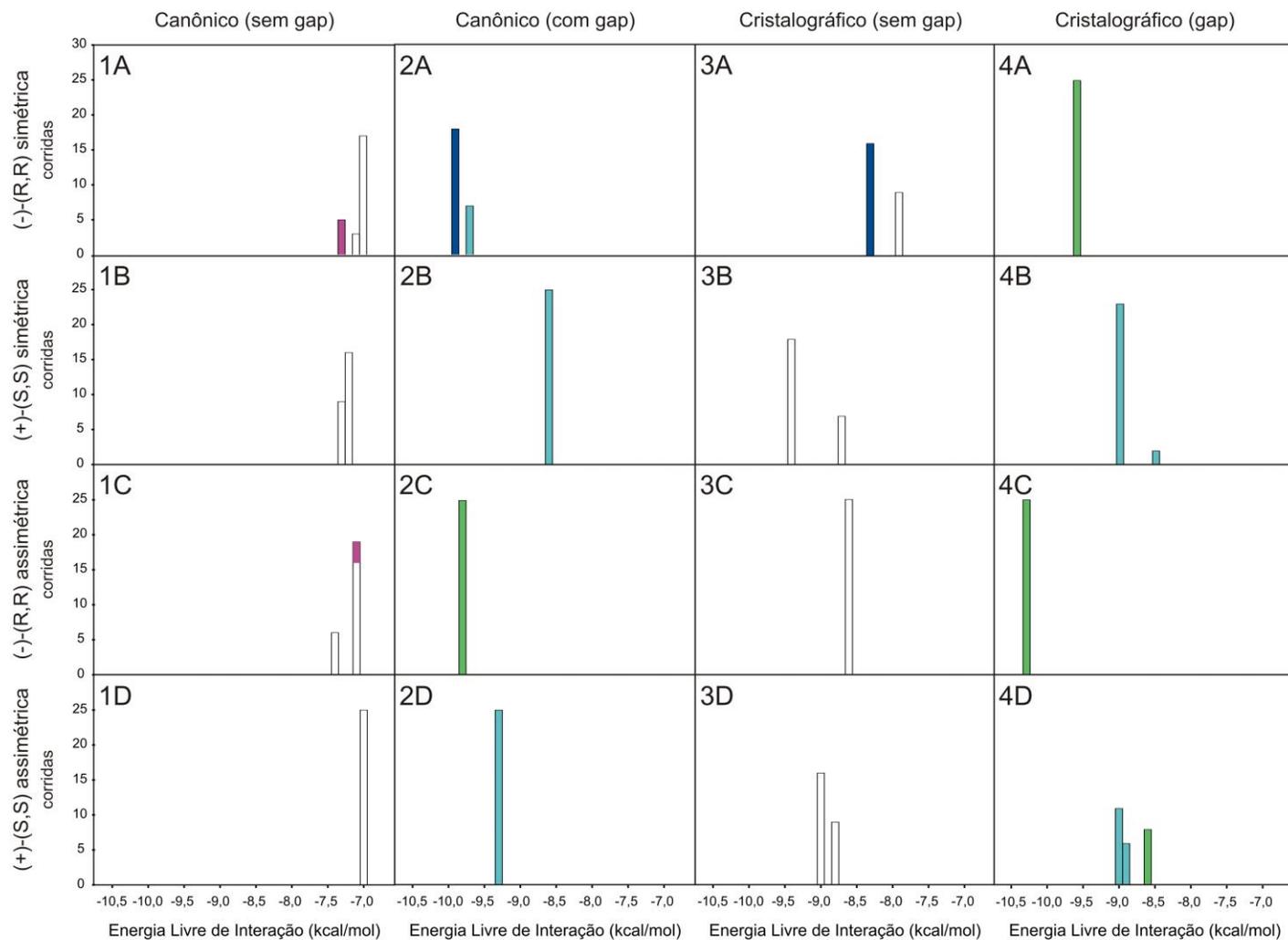


Figura 10. Perfis de *clusters* para as docagens das bases de Tröger com quatro diferentes receptores. Cada *cluster* é representado por uma barra em que a altura corresponde ao número de conformações do *cluster* e a cor corresponde ao modo de ligação: □ sulco menor parcial; ■ sulco menor completo; ■ intercalação; ■ misto (intercalação + sulco menor); ■ intercalação parcial (na extremidade).

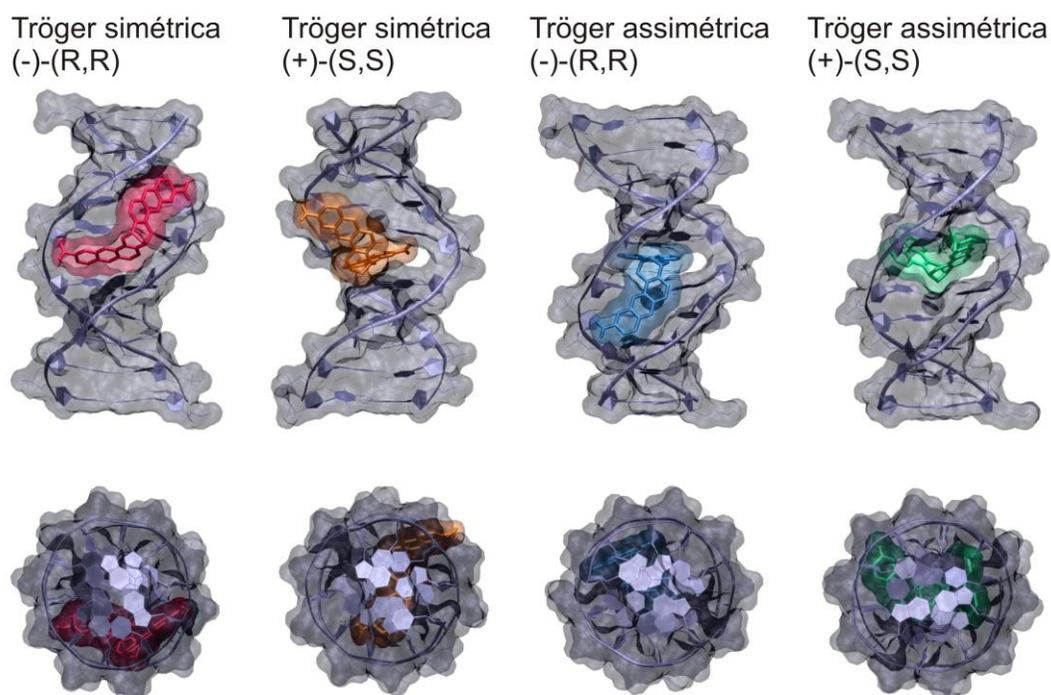


Figura 11. Conformações de melhor energia livre de interação para as docagens das bases de Tröger com o DNA canônico com *gap* (2A-2D).

3.4 CONCLUSÕES

Foram realizados diversos estudos de docagem usando ligantes que interagem com o DNA através de intercalação (elipticina) ou através de ligação ao sulco menor (netropsina). Quatro estruturas diferentes de DNA foram usadas como receptores: DNA cristalográfico de um complexo com a elipticina; DNA cristalográfico de um complexo com a netropsina; B-DNA canônico; e B-DNA modificado contendo um *gap* de intercalação. Foram feitas docagens diretas e cruzadas utilizando um conjunto otimizado de parâmetros com o programa *AutoDock 4.0*.

Os resultados das docagens dessa etapa de validação estão de acordo com cada um dos mecanismos de interação (encaixe-induzido ou chave-fechadura) quando o sistema é submetido às limitações de flexibilidade do método: a docagem com a elipticina mostrou-se mais sensível à conformação do oligômero do que a docagem com a netropsina. Nas docagens cruzadas, essa observação foi mais evidente, pois a netropsina

conseguiu se ligar mesmo ao sulco menor de um DNA estruturalmente distorcido, enquanto que a elipticina não conseguiu intercalar quando o receptor não possuía um *gap* de intercalação apropriado.

Ainda assim, a atual limitação de flexibilidade dos métodos de docagem pode ser contornada por uma escolha apropriada da conformação do receptor. Foi mostrado que, desde que o receptor apresente um *gap* de intercalação gerado artificialmente, a docagem é capaz de apontar, para ambos os ligantes, para o modo de interação correto como o resultado energeticamente mais favorável. Isso sugere, também, que a função de escore do *AutoDock* é eficiente para avaliar interações ligante-DNA pelo menos de forma qualitativa.

Esses quatro tipos de receptores foram posteriormente aplicados para investigar possíveis modos de ligação de duas bases de Tröger ao DNA: uma base simétrica com substituintes proflavina e uma assimétrica com substituintes proflavina e fenantrolina, cada uma com seus dois enantiômeros ópticos.

Os resultados dessas docagens mostraram que os oligômeros contendo *gap* de intercalação são mais adequados para interação com as bases de Tröger, resultando em energias livres de interação muito negativas e reproduzindo qualitativamente a estereosseletividade dos enantiômeros ópticos. Também nesta etapa, o B-DNA modificado com *gap* mostrou-se particularmente promissor, pois os modos de ligação sugeridos pelas docagens com esse receptor estão de acordo com os modos sugeridos pelos dados experimentais encontrados na literatura (ligação ao sulco menor para a Tröger simétrica e modo de interação misto para a Tröger assimétrica).

Portanto, nós sugerimos que este protocolo padrão utilizando um B-DNA com *gap* de intercalação artificial pode ser aplicado para investigar outros ligantes cujo modo de interação com o DNA permanece sem elucidação experimental, fornecendo pontos de partida promissores para estudos teóricos subsequentes, tais como estudos de simulação por dinâmica molecular.

4. ESTUDOS DE DINÂMICA MOLECULAR

A simulação do tipo dinâmica molecular é uma das ferramentas computacionais mais utilizadas no estudo de sistemas biológicos atualmente. Por ser uma metodologia já bem estabelecida, cujos fundamentos e aplicações podem ser facilmente encontrados em um grande número de revisões de alta qualidade, julgamos desnecessário apresentar aqui uma revisão teórica desta metodologia. A quem interessar, sugerimos a leitura de um trabalho recente de van Gunsteren *et al.*;⁶⁷ que fazem uma revisão bastante abrangente da aplicação da dinâmica molecular para a simulação de sistemas biológicos.

4.1 METODOLOGIA

Primeiramente, foram selecionados alguns dos complexos obtidos através das docagens da etapa anterior. Dentre os 16 sistemas de docagens, foram escolhidos os seguintes complexos:

Tröger **simétrica** R e DNA canônico **com gap** (2A, cluster 2);

Tröger **assimétrica** R e DNA canônico **com gap** (2C, cluster 1);

Tröger **simétrica** R e DNA cristalográfico **sem gap** (3A, cluster 1);

Tröger **assimétrica** R e DNA cristalográfico **sem gap** (3C, cluster 1).

O cluster de menor energia livre do sistema 2A (cluster 1) também foi simulado, e o modo de interação resultante foi exatamente o mesmo que ocorreu na simulação do cluster 2. Por essa razão, os resultados para esse complexo não serão apresentados. Também foram simulados os complexos 4A e 4C, contendo o DNA cristalográfico com *gap*. Esses sistemas, no entanto, mostraram-se muito instáveis, provavelmente devido ao tamanho muito curto do oligonucleotídeo (6 pares de bases). Sendo assim, esses resultados foram descartados e também não serão apresentados no trabalho.

Além dos complexos, foram também simulados os receptores e os ligantes de forma isolada, totalizando os 8 sistemas que estão apresentados na Tabela 3. Para facilitar a discussão dos resultados, os complexos que foram simulados não serão mais referidos

como 2A, 2C, 3A e 3C, mas sim de acordo com as novas referências e abreviaturas contidas na Tabela 3.

Tabela 3. Resumo dos sistemas simulados.

Sistema	Nova referência	Abreviatura
Complexo 2A	Complexo de intercalação da Tröger simétrica	COMPL-INT-SIM
Complexo 2C	Complexo de intercalação da Tröger assimétrica	COMPL-INT-ASS
Complexo 3A	Complexo de ligação ao sulco da Tröger simétrica	COMPL-SUL-SIM
Complexo 3C	Complexo de ligação ao sulco da Tröger assimétrica	COMPL-SUL-ASS
DNA canônico com <i>gap</i>	DNA com <i>gap</i>	DNA-GAP
DNA cristalográfico sem <i>gap</i>	DNA sem <i>gap</i>	DNA-SEM-GAP
Tröger Simétrica R	Tröger simétrica	TRO-SIM
Tröger Assimétrica R	Tröger assimétrica	TRO-ASS

Os sistemas foram simulados utilizando os potenciais do campo de força ffamber03⁶⁸ portados para o pacote de simulação *GROMACS*⁵⁹ de acordo com Sorin e Pande.⁶⁹ Foram utilizadas condições periódicas de contorno e o tratamento das interações eletrostáticas foi feito de acordo com o método PME (*Particle-Mesh Ewald*).⁷⁰

Enquanto os oligonucleotídeos foram parametrizados com o campo de força ffamber03, os parâmetros estruturais para os ligantes foram obtidos automaticamente a partir do GAFF (*Generalized Amber Force Field*),⁷¹ utilizando o script ACPYPI (*Antechamber Python Parses Interface*).⁷² O ACPYPI é uma ferramenta baseada em Python que utiliza o programa *Antechamber*⁷³ para gerar topologias a partir dos parâmetros do GAFF (ver apêndice C). Como arquivo de entrada, o ACPYPI requer somente o arquivo de coordenadas atômicas do ligante e, entre os arquivos de saída, estão arquivos de topologia adequados para a simulação com o pacote *GROMACS* (ver apêndice D). Após o cálculo dos potenciais eletrostáticos em nível HF/6-31G* com o *Gaussian*⁶⁶, as cargas pontuais dos ligantes foram obtidas através do método RESP (*Restrained Electrostatic Potential*).⁷⁴

Os sistemas foram colocados em caixas cúbicas de tamanho suficiente para que a distância mínima entre a molécula do soluto e as paredes da caixa fosse 18 Å. A seguir, foram adicionados contra-íons sódio aos sistemas contendo DNA, que então foram submetidos a uma minimização de energia usando o método *steepest descent*, seguida de

uma simulação de 100 ps no vácuo com restrição de posições para os átomos pesados do(s) soluto(s) – DNA e ligantes.

Todos os sistemas foram solvatados com moléculas de água TIP3P e submetidos a uma nova minimização de energia utilizando o método *steepest descent*. Em seguida, foram adicionados íons sódio e cloreto em concentração fisiológica, e os sistemas foram submetidos a uma simulação de 100 ps com restrição de posições para os átomos do soluto. Os sistemas foram então aquecidos linearmente de 270K a 310K em 500 ps de simulação NVT. A etapa de produção consistiu em 20 ns de simulação NVT para os sistemas contendo DNA e em 10 ns de simulação (NVT) para os sistemas contendo somente os ligantes.

Durante a análise, foi utilizado o módulo *g_rms* do *GROMACS* para obtenção dos gráficos de desvio quadrático médio (RMSD) e o módulo *g_dist* para o monitoramento das distâncias entre alguns átomos dos ligantes e dos receptores. A análise da estabilidade dos pares de bases dos oligômeros foi feita a partir do programa *X3DNA*, assim como a análise de alguns dos ângulos torcionais e de alguns dos parâmetros de pares de bases.

4.2 RESULTADOS E DISCUSSÃO

4.2.1 Estabilidade dos oligonucleotídeos

A fim de verificar se as simulações atingiram o equilíbrio e de avaliar a estabilidade dos oligonucleotídeos frente à desnaturação (rompimento das ligações de hidrogênio entre pares de bases, que resulta na separação das duas fitas), foi monitorados o desvio quadrático médio em relação à estrutura de partida e o número de pares de bases conservados ao longo da simulação.

De acordo com a análise do RMSD (Figura 12), todos os sistemas atingiram um patamar de estabilidade nos últimos 5 ns das simulações, exceto o complexo de intercalação da Tröger simétrica (COMPL-INT-SIM), para o qual está claro se o sistema realmente atingiu o equilíbrio nos últimos 2 ns. Além disso, os valores de RMSD foram maiores para os oligonucleotídeos pertencentes aos complexos de intercalação do que para os complexos de ligação ao sulco menor. No complexo de intercalação da Tröger

simétrica, o RMSD do DNA (verde) se manteve, durante toda a simulação, consideravelmente mais alto ($>0,3$ nm) em comparação com o RMSD do DNA isolado (ciano), indicando que a presença do ligante promoveu alterações estruturais mais intensas do que as decorrentes da relaxação da estrutura isolada. No complexo de intercalação da Tröger assimétrica, embora o RMSD do DNA também tenha sido mais alto do que o do DNA isolado durante a maior parte da simulação, nos últimos 5 ns eles passaram a ser semelhantes em magnitude.

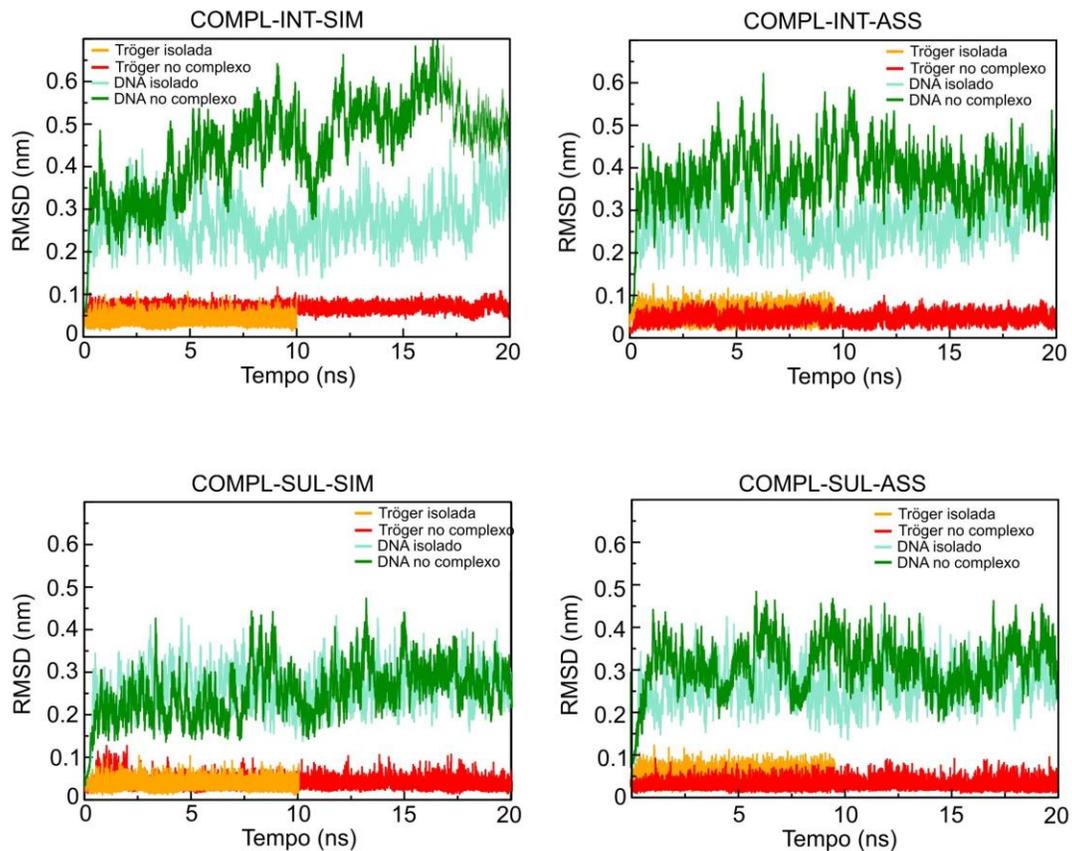


Figura 12. Gráficos de RMSD. DNA e Tröger nos complexos (linhas verde e vermelha) em comparação com o DNA e Tröger isolados (linhas ciano e laranja).

Já no caso dos complexos de ligação ao sulco menor, os valores de RMSD do DNA (verde) foram menores ($\sim 0,3$ nm) e se mantiveram, durante toda a simulação, próximos dos valores do DNA isolado (ciano), indicando que as alterações estruturais provocadas pela ligação ao sulco menor são de intensidade comparável às alterações decorrentes da simples relaxação da estrutura isolada.

A análise do número de pares de bases ao longo das simulações (Figura 13) mostra que não houve desnaturação (separação das duas fitas) em nenhum dos oligonucleotídeos simulados, já que todos mantiveram a maioria dos pares de bases estáveis ao longo das simulações. É possível ver que o decâmero com *gap* (coluna da esquerda) apresentou dez pares de bases durante praticamente toda a simulação, tanto quando simulado isoladamente, como quando nos complexos de intercalação. Dessa forma, conclui-se que os valores mais altos de RMSD observados para o DNA nesses complexos decorreram realmente de alterações estruturais e não simplesmente da desnaturação da dupla hélice. Eventualmente, no entanto, ocorreu o despareamento de um dos pares de bases e/ou a formação de pares não-canônicos (que não seguem estritamente as ligações de hidrogênio descritas por Watson e Crick⁷⁵).

O dodecâmero sem *gap* (coluna da direita), apresentou um perfil semelhante, com eventual perda de até dois pares de bases, no caso do complexo de ligação ao sulco menor da Tröger simétrica, e com uma maior proporção de pares de bases não-canônicos durante as simulações. Ainda assim, esse efeito foi pequeno em relação ao tamanho do oligonucleotídeo (12 pares de bases) e, além disso, ocorreu tanto para o dodecâmero isolado como para o dodecâmero interagindo com os ligantes no sulco menor.

Em resumo, os oligonucleotídeos se mantiveram estáveis em relação ao número de pares de bases, o que não apenas serve como critério de validação das simulações como também indica que a presença dos ligantes não induziu à desnaturação da dupla hélice.

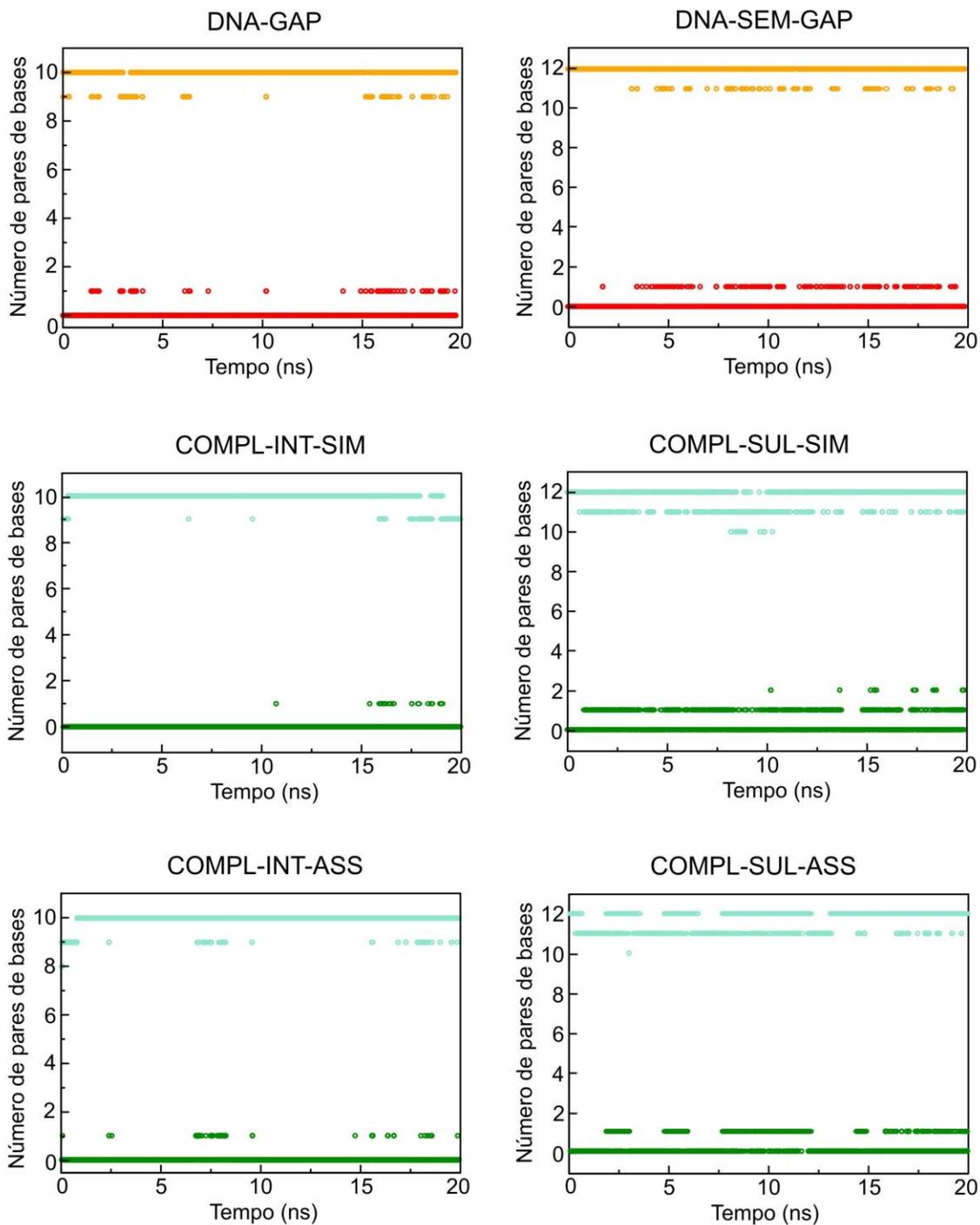


Figura 13. Número de pares de bases ao longo da simulação para os oligonucleotídeos simulados. À esquerda, sistemas contendo DNA com *gap* (decâmero); à direita, sistemas contendo DNA sem *gap* (dodecâmero). Os pares de bases totais são mostrados em laranja (oligonucleotídeos isolados) e ciano (oligonucleotídeos nos complexos), e os pares de bases não-Watson-Crick são mostrados em vermelho (oligonucleotídeos isolados) e verde (oligonucleotídeos nos complexos).

4.2.2 Análise global das simulações

A fim de obter o tempo de residência dos ligantes e de identificar os principais eventos que ocorreram durante as simulações, foi feita uma análise global para cada um dos quatro complexos simulados. Essa análise global consistiu no monitoramento das distâncias entre a base de Tröger e alguns dos pares de bases do DNA, bem como na análise visual de algumas estruturas extraídas de diferentes pontos da trajetória.

Para a análise das distâncias, foram escolhidos três grupamentos pertencentes a cada base de Tröger, cuja distância foi medida em relação ao centro de massa dos pares de bases pertencentes ao(s) sítio(s) de interação. No caso da Tröger simétrica, foi escolhido o carbono da ponte diazocina (C) e os nitrogênios dos grupamentos amino (N1 e N2), como mostra a Figura 14a. Para a Tröger assimétrica, foi escolhido, além do carbono da ponte diazocina (C), o nitrogênio do grupamento amino (Na) e um dos nitrogênios pertencentes aos anéis aromáticos do substituinte fenantrolina (Nf), como mostrado na Figura 14b.

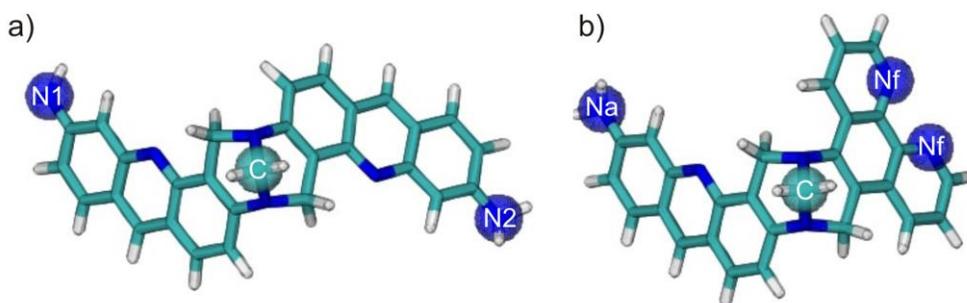


Figura 14. Grupamentos monitorados para a Tröger simétrica (a) e para a Tröger assimétrica (b) na análise das distâncias de interação com o DNA.

Complexo de Intercalação da Tröger Simétrica (COMPL-INT-SIM)

No início da simulação, a Tröger simétrica se encontrava intercalada no DNA, com a ponte diazocina inserida no *gap* (dado não mostrado). No entanto, rapidamente houve uma reorientação do ligante e, a partir de então, um dos substituintes proflavina permaneceu intercalado no sulco, enquanto o outro ficou interagindo no sulco menor, como esquematizado na Figura 15A. Pelo monitoramento das distâncias (Figura 15B), foi

possível observar: i) uma baixa movimentação da ponte diazocina (C), que permaneceu bastante próxima do par de bases 6A-15T; ii) também baixa movimentação do grupo amino 1 (N1), que se mostrou aproximadamente equidistante dos pares 5T-16A e 6A-15T, portanto inserido no *gap*; iii) maior movimentação do grupo amino 2 (N2), que manteve interações mais próximas com os pares de bases 5T-16A e 4T-17A.

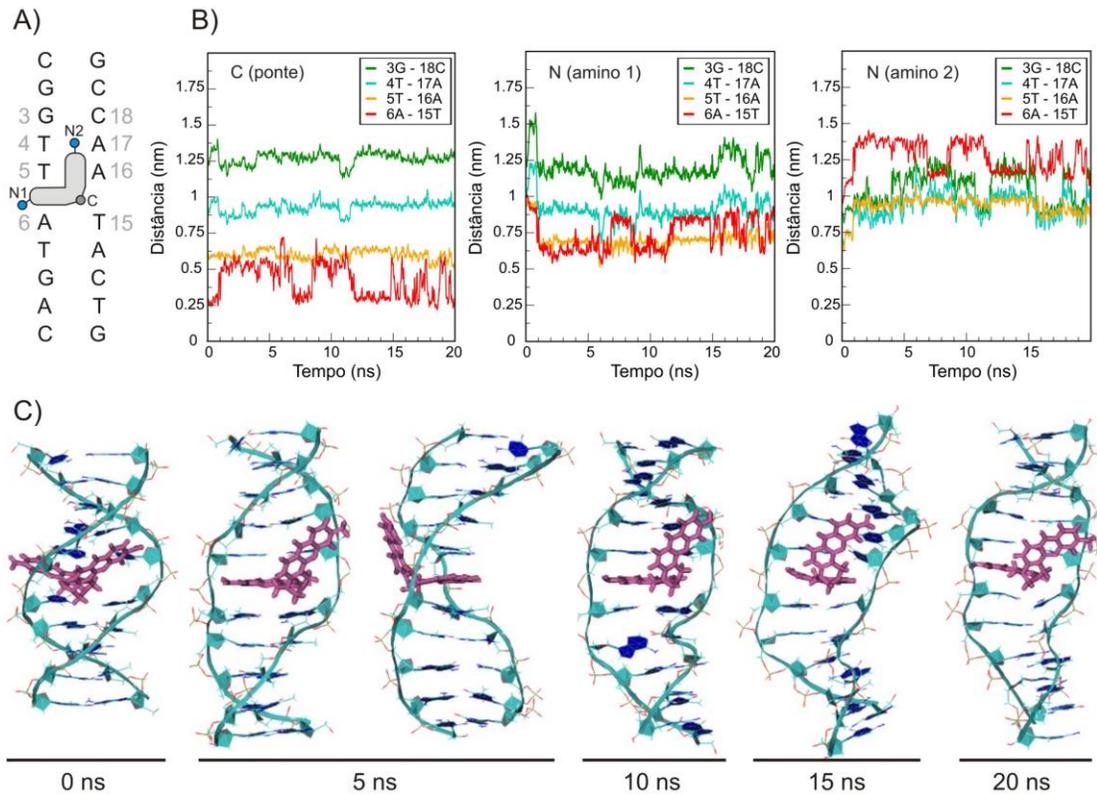


Figura 15. Evolução temporal do complexo de intercalação da Tröger simétrica (COMPL-INT-SIM). A. Esquema do sítio e do modo de interação; B. distâncias da ponte diazocina (C ponte) e dos grupamentos amino (N1 e N2) em relação ao centro de massa dos quatro pares de bases do sítio de interação; C. estruturas obtidas a partir da trajetória da simulação.

Para os três grupamentos, as maiores flutuações ocorreram para o par de bases 6A-15T e foram decorrentes da movimentação do próprio par de bases, já que as distâncias da ponte e do grupo N1 em relação aos outros pares de bases se mantiveram aproximadamente constantes. Além disso, o padrão de flutuação do par 6A-15T foi o mesmo para a ponte diazocina e para o grupo N2 (no sulco), mas inverso para o grupo N1 (no *gap*) (ver Figura 15B, linhas vermelhas). Isso significa que, ao se aproximar dos grupamentos N2 e da ponte diazocina, o centro de massas do par 6A-15T se afastou do

grupamento N1, e vice-versa. Dessa forma, podemos concluir que a movimentação do centro de massa do par 6A-15T ocorreu em um plano paralelo aos planos dos pares de bases, e não ao longo do eixo da dupla hélice – já que, neste último caso, o par de bases deveria se afastar ou se aproximar *simultaneamente* dos três grupamentos monitorados.

Muito provavelmente, essa reorientação do par de bases inferior do *gap* decorre de uma diminuição do ângulo de torção da dupla hélice (*Twist*), o que leva ao aumento da sobreposição entre os anéis aromáticos do ligante e das bases nitrogenadas do DNA. De fato, esse efeito de diminuição da torção da hélice pode ser observado em algumas estruturas retiradas da trajetória da simulação (Figura 15C) e confirmado pelas análises que serão apresentadas na seção 4.2.4.

Além deste, outros eventos podem ser identificados visualmente na Figura 15C. Em 5 ns, por exemplo, observa-se que já houve uma reorientação da base de Tröger no *gap*, que agora apresenta uma proflavina posicionada de forma paralela aos pares de bases do *gap*. Esse evento pode ser correlacionado com o brusco decréscimo observado em 1 ns nas distâncias do grupamento N1 (Figura 15B), mostrando que o grupamento N1 se aproximou simultaneamente do centro de massa dos quatro pares de bases monitorados, ou seja, “entrou” para dentro do *gap*. Nesse estágio, também já se observa um certo “desenrolamento” da dupla hélice, que se mostra mais intenso a partir de 15 ns. Em 10 ns, já é possível ver distorções bastante intensas da cadeia principal na região inferior ao *gap*, que podem decorrer justamente da reorientação do par 6A-15T. Além disso, no final da simulação o último par de bases do DNA se desfez e as duas bases se “empilharam” como se fossem pares de bases consecutivos (ver 15 e 20 ns). A estrutura final, em 20 ns, mostra uma dupla hélice bastante desenrolada e com fortes distorções em forma de zigue-e-zague na cadeia principal.

Complexo de Intercalação da Tröger Assimétrica

Durante a simulação do complexo de intercalação da Tröger assimétrica, o substituinte fenantrolina permaneceu intercalado no *gap*, enquanto a proflavina ficou interagindo no sulco menor do DNA, como esquematizado na Figura 16A.

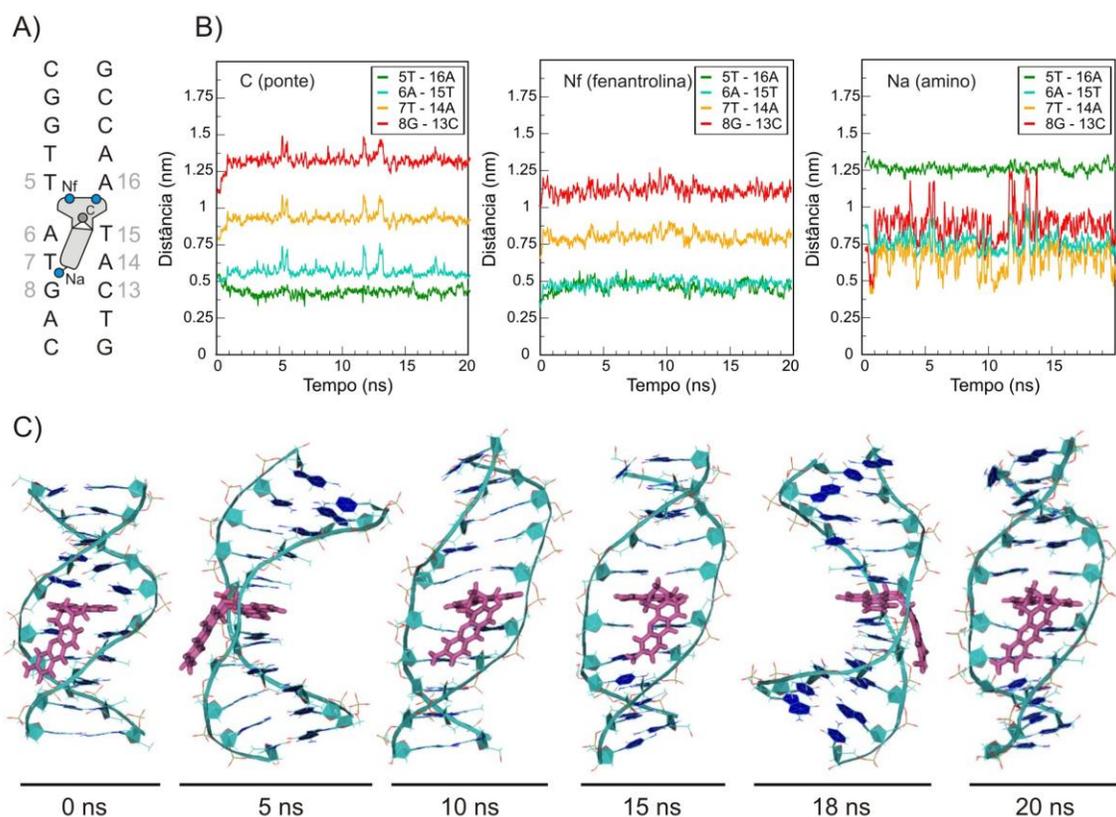


Figura 16. Evolução temporal do complexo de intercalação da Tröger assimétrica (COMPL-INT-ASS). A. Esquema do sítio e do modo de interação; B. distâncias da ponte diazocina (C ponte), do grupamento amino (Na) e de um dos nitrogênios da fenantrolina (Nf) em relação ao centro de massa dos quatro pares de bases do sítio de interação; C. estruturas obtidas a partir da trajetória da simulação.

Através do monitoramento das distâncias (Figura 16B), pode-se observar que praticamente não houve movimentação da ponte diazocina (C) e nem do nitrogênio da fenantrolina (Nf). Ambos se mantiveram interagindo de forma praticamente equidistante com os pares de bases envolvidos no *gap* (linhas verde e ciano), confirmando a intercalação. O grupamento amino da proflavina (Na) se manteve mais distante do DNA de uma forma geral e foi o que apresentou maior movimentação, interagindo de forma mais próxima com o par de bases 7T-14A. Neste caso, porém, as flutuações do grupamento amino não estão correlacionadas com flutuações semelhantes nos outros grupamentos monitorados (ponte e Nf), indicando não serem decorrentes de uma movimentação global da Tröger nem de uma movimentação do próprio DNA. Dessa forma, esse resultado mostra que, apesar de possuir uma estrutura essencialmente rígida, a base de Tröger apresenta algum grau de flexibilidade.

As estruturas ao longo da simulação (Figura 16C) se mostraram relativamente conservadas, sem as distorções em zigue-e-zague de cadeia principal que ocorreram no complexo de intercalação da Tröger simétrica. Em alguns pontos da simulação, no entanto, foi observada uma leve curvatura em direção ao sulco maior do DNA, como pode ser visto nas estruturas em 5 e 18 ns e confirmado pelas análises que serão apresentadas na seção 4.2.5. Além disso, houve uma diminuição do ângulo de torção da dupla hélice, que pode ser observada nas estruturas em 10, 15 e 20 ns. Neste caso, porém, o “desenrolamento” parece ter sido local, ou seja, mais restrito aos pares de bases do *gap* de intercalação, sem propagação deste efeito para as extremidades do oligonucleotídeo.

Em resumo, o monitoramento das distâncias, juntamente com a análise visual das estruturas, apontaram para tempos de residência altos para as bases de Tröger nos complexos de intercalação (>20 ns), em que os ligantes apresentaram baixa mobilidade, especialmente no caso da Tröger assimétrica. A simulação de ambos os complexos resultou em uma diminuição do *Twist* da dupla hélice, o que é uma característica comum a dos intercaladores^{5,10,11}, e que parece ter sido mais intensa no caso da Tröger simétrica. Aparentemente, a estrutura do DNA no complexo com a Tröger assimétrica se manteve mais conservada em relação à do complexo com a Tröger simétrica. No entanto, isso não necessariamente indica uma maior capacidade da Tröger simétrica de deformar o DNA, podendo ser simplesmente um artefato do ponto de partida utilizado para cada simulação.

Complexo de Ligação ao Sulco da Tröger Simétrica

A simulação do complexo de ligação ao sulco menor da Tröger simétrica iniciou com o ligante docado no sulco menor, bastante próximo à extremidade inferior do oligonucleotídeo e com a ponte diazocina apontando para fora do sulco menor, como ilustrado esquematicamente na Figura 17A. Embora a Tröger tenha interagido com o DNA durante toda a simulação, o monitoramento das distâncias (Figura 17B), bem como a análise visual da simulação (Figura 17C) mostra uma significativa reorientação do ligante no sulco menor. De fato, pela análise das distâncias da ponte diazocina em relação aos pares de bases do sítio de interação (Figura 17B, C (ponte)), foi possível identificar

quatro transições bem definidas, em aproximadamente 8, 10, 12 e 15,4 ns, que estão indicadas pelas linhas verticais tracejadas que dividem gráfico das distâncias da ponte em cinco patamares relativamente estáveis, chamados de regiões I, II, III, IV e V. Como era de se esperar, algumas dessas transições estão correlacionadas com transições simultâneas observadas para os grupamentos amino, refletindo a rigidez estrutural das bases de Tröger.

Considerando que essas transições podem indicar mais de um sítio ou de mais de um modo de interação sendo amostrados durante a simulação, convém fazer uma análise mais detalhada das 5 diferentes regiões, como apresentado a seguir.

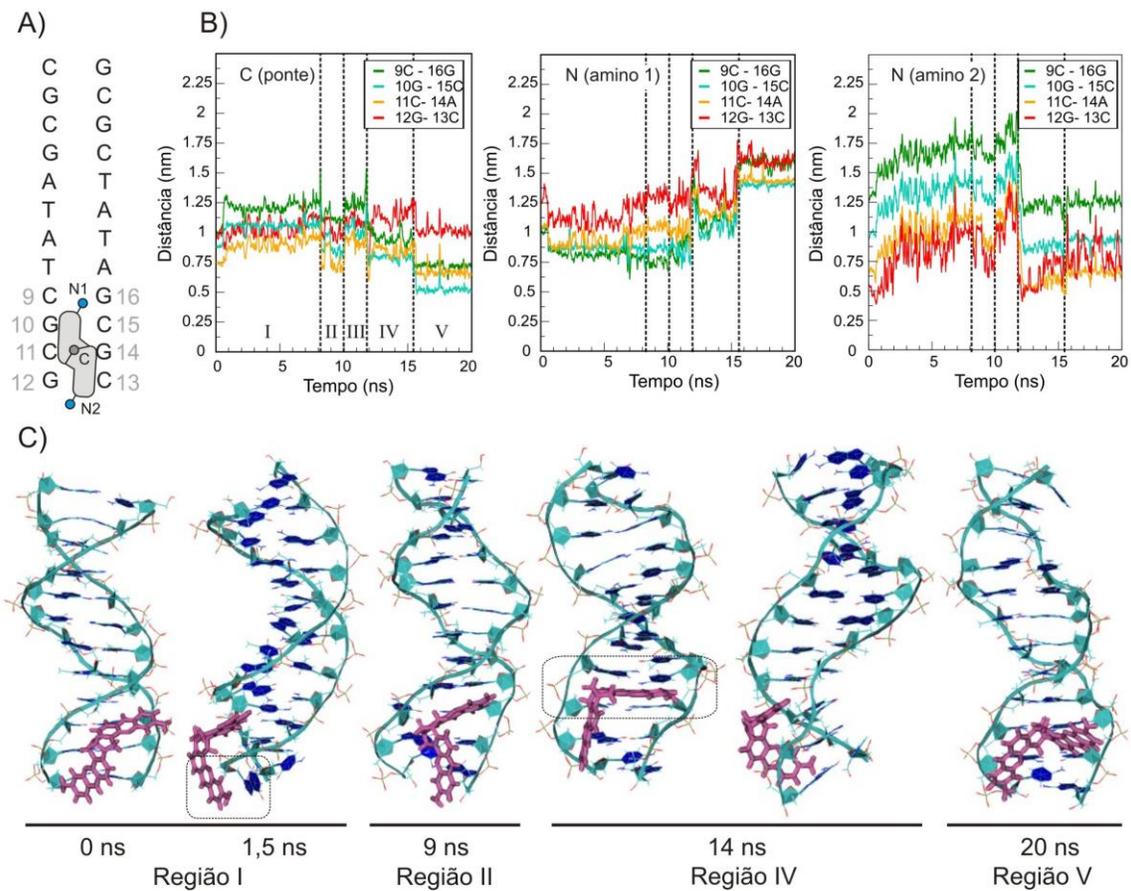


Figura 17. Evolução temporal do complexo de ligação do sulco da Tröger simétrica (COMPL-SUL-SIM). A. Esquema do sítio e do modo de interação inicial; B. distâncias da ponte diazocina (C ponte) e dos grupamentos amino (N1 e N2) em relação ao centro de massa dos quatro pares de bases do sítio de interação; C. estruturas obtidas a partir da trajetória da simulação.

Nas três primeiras regiões (I a III), a Figura 17B mostra que a ponte diazocina permaneceu mais próxima do par de bases 11C-14A (linha laranja), sem que houvesse qualquer inversão nas distâncias referentes aos diferentes pares de bases. Isso indica, portanto, que o sítio de interação permaneceu praticamente o mesmo durante essa etapa da simulação. Além disso, as transições da ponte não foram acompanhadas de alterações significativas nas distâncias relativas aos grupamentos amino, indicando que também não houve mudança do modo de interação. De fato, a única alteração significativa consistiu em uma diminuição da distância da ponte em relação aos pares 9C-16G, 10G-15C e 11C-14A na região II, o que indica uma suave movimentação da Tröger em direção à região central da sequência do oligonucleotídeo, mas que é revertida na região III. Essas conclusões estão de acordo com as estruturas de 1,5 ns (região I) e 9ns (região II), mostradas na Figura 17C. Um outro efeito que somente foi observado através da inspeção visual das estruturas foi que, em alguns momentos da simulação na região I (ver 1,5 ns, área tracejada), ocorreu a desnaturação do par de bases da extremidade inferior e uma das bases passou a interagir de forma paralela com um dos anéis aromáticos da base de Tröger, de forma semelhante ao que ocorre durante a intercalação. Obviamente, este é um artefato decorrente do tamanho do fragmento de oligonucleotídeo e da proximidade da base de Tröger em relação à extremidade inferior do DNA. No entanto, apesar de artificial, pode estar relacionado com a capacidade intrínseca do substituinte proflavina de interagir através de intercalação.

Na região IV, o monitoramento das distâncias aponta para uma reorientação mais significativa do ligante: i) a ponte diazocina passou a se localizar mais próxima do par de bases 10G-15C (linha ciano) ii) o grupamento amino 1 se afastou do centro de massa dos pares 9C-16G e 10G-15C e iii) o grupamento amino 2 se aproximou consideravelmente de todos os pares de bases monitorados, indicando uma movimentação deste grupamento em direção à região central da sequência do oligonucleotídeo. Como pode ser visto pelas estruturas da região IV mostradas na figura 17C, essa “subida” do grupamento amino 2 produziu um encaixe mais efetivo de um dos substituintes proflavina no sulco menor, simultaneamente posicionando o outro substituinte de forma perfeitamente paralela entre os pares de bases 9C-16G e 10G-15C (ver 14 ns, área tracejada). Em uma visão lateral

desta mesma estrutura, é possível ver que essa orientação do ligante é bastante favorável à intercalação.

Embora o ligante tenha permanecido nesta orientação por pouco mais de 2 ns, a intercalação não chegou a ocorrer. Em aproximadamente 15,4 ns, houve uma brusca aproximação da ponte diazocina em relação ao DNA – como pode ser visto pela diminuição simultânea das distâncias referentes aos quatro pares de bases monitorados – que foi acompanhada de forma inversa por um brusco afastamento do grupo amino 1 (região V). O modo de interação resultante está ilustrado na estrutura de 20 ns da Figura 17C, que mostra que a Tröger terminou a simulação interagindo com o DNA principalmente através da ponte diazocina, que está voltada, não mais para fora, mas sim para dentro do sulco menor.

Com relação à estrutura do DNA, pode-se observar visualmente que ocorreram algumas distorções locais e globais, sugerindo que o DNA foi afetado pela presença do ligante. Em 1,5 ns, por exemplo, houve uma aparente curvatura do DNA em direção ao sulco maior; e, em uma das estruturas em 14 ns, o DNA apresentou um alargamento do sulco menor, que remete à forma A do DNA. No entanto, essas alterações estruturais parecem ser reversíveis e não acumulativas ao longo da simulação, já que não se mantiveram presentes na estrutura final (20 ns).

Complexo de ligação ao sulco da Tröger Assimétrica

A simulação do complexo de ligação ao sulco da Tröger assimétrica iniciou com o ligante ligado parcialmente no sulco menor, como mostra o esquema da Figura 18A. Assim como no caso do complexo de ligação ao sulco da Tröger simétrica, houve bastante movimentação da base de Tröger no sulco menor, que é melhor acompanhada pelas distâncias relativas a ponte diazocina, já que este grupamento apresentou flutuações menores em comparação aos outros grupamentos monitorados.

Inicialmente (região I), a ponte se encontrava mais próxima do par de bases 9C-16G (ciano), a uma distância de aproximadamente 0,8 nm. Após 1 ns (região II), houve uma reorientação em que: i) a ponte se aproximou dos pares de bases 9C-16G (ciano), 10G-15C (laranja) e 11C-14G (vermelho); ii) a ponte passou a interagir

preferencialmente com o par 10G-15C (laranja), a uma distância de aproximadamente 0.5 nm. Isso indica não somente uma alteração do sítio de interação como também sugere uma mudança no próprio modo de interação. De fato, as estruturas de 5 e 10 ns mostram que, embora a fenantrolina continue próxima do sulco menor, houve uma reorientação da ponte diazocina, que se encontra mais voltada para o interior do sulco menor (Figura 18C). Esse modo de interação se manteve estável por aproximadamente 15 ns, com relativa mobilidade do grupamento amino (Na) e do nitrogênio da fenantrolina (Nf), como pode ser observado pelas flutuações das distâncias da Figura 18B.

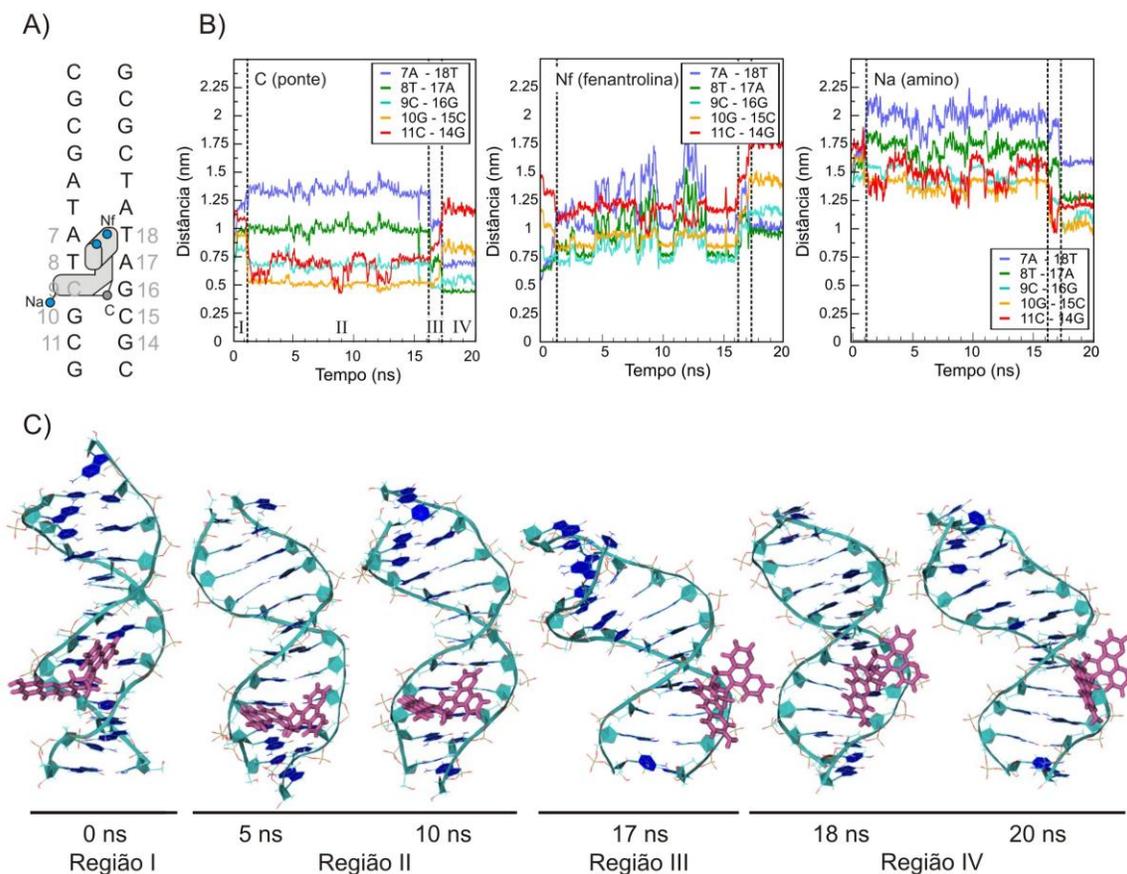


Figura 18. Evolução temporal do complexo de ligação do sulco da Tröger assimétrica (COMPL-SUL-ASS). A. Esquema do sítio e do modo de interação inicial; B. distâncias da ponte diazocina (C ponte), do grupamento amino (Na) e de um dos nitrogênios da fenantrolina (Nf) em relação ao centro de massa dos quatro pares de bases do sítio de interação; C. estruturas obtidas a partir da trajetória da simulação.

Após 15 ns (região III), nota-se que a ponte diazocina voltou a interagir preferencialmente com o par de bases 9C-16G (ciano), porém de forma mais próxima do

que ocorreu na região I. Essa transição é correlacionada com transições simultâneas nos outros grupamentos monitorados (Na e Nf) que mostram que houve um afastamento do grupamento fenantrolina em relação ao sulco menor e uma aproximação do grupamento amino. Esse novo modo de interação está ilustrado na estrutura de 17 ns, onde se observa claramente o substituinte fenantrolina projetado para fora do sulco menor, com a Tröger interagindo com o DNA principalmente através da ponte diazocina (Figura 18C).

Logo após 17 ns (região IV), houve uma terceira e última reorientação, em que a ponte passou a interagir preferencialmente com o par de bases 8T-17A (verde), de forma ainda mais próxima do que nas regiões anteriores. Simultaneamente, houve um novo afastamento do nitrogênio da fenantrolina, de modo que este grupamento terminou a simulação significativamente mais afastado do sulco menor do que estava inicialmente. Por outro lado, houve uma aproximação do grupamento amino, que terminou mais próximo do sulco menor do que no início da simulação. De modo geral, no entanto, o monitoramento das distâncias mostrou uma tendência do sistema para um modo de interação na qual a base de Tröger interage com o DNA basicamente através da ponte diazocina (0,5 nm), com pouca ou nenhuma contribuição dos nitrogênios dos substituintes (~1 nm), como pode ser observado nas estruturas de 18 e 20 ns da Figura 18C. Convém observar também que esse modo de interação é bastante semelhante ao modo de interação final da Tröger simétrica no sulco menor.

Com relação à estrutura do DNA, no entanto, a Tröger assimétrica parece ter induzido alterações estruturais mais consistentes do que as observadas no complexo de ligação ao sulco da Tröger simétrica. Ainda na estrutura de 5 ns, já é possível observar um alargamento do sulco menor e um “achatamento” da dupla hélice, que pode ser tanto devido a uma curvatura local como a uma alteração conformacional, levando à forma A do DNA. Esses efeitos são observados de forma ainda mais evidente na estrutura de 17 ns, que mostra uma estrutura bastante semelhante à forma A do DNA, com um sulco menor mais raso e um sulco maior mais profundo. Diferentemente das distorções observadas no COMPL-SUL-SIM, neste caso as distorções se mantiveram até o final da simulação (20 ns).

Comparando com os complexos de intercalação, portanto, os complexos de ligação ao sulco menor apresentaram tempos de residência menores, com alteração do modo ou mesmo de sítio de interação ao longo das simulações. Para ambas as bases de Tröger, houve uma reorientação dos ligantes no sulco menor, levando a um modo de interação comum para os dois sistemas, em que a base de Tröger interagiu com o DNA através da ponte diazocina. No caso da simulação com a Tröger simétrica, foi também amostrada uma orientação paralela do ligante em relação a dois pares de bases, que se mostrou bastante propícia para a intercalação. Apesar deste efeito poder ser considerado como um indício de intercalação, não foi possível esclarecer se a base de Tröger simétrica é ou não capaz de induzir a abertura do *gap* necessário para que este modo de interação ocorra.

Além disso, as alterações provocadas pela ligação ao sulco menor não foram, aparentemente, tão intensas como nos complexos de intercalação. Essa observação, juntamente com os tempos de residência em cada tipo de interação, é um indício de que o modo de ligação ao sulco menor consiste em um modo de interação mais fraco do que a intercalação.

4.2.3 Análise conformacional da cadeia principal

Para melhor avaliar e quantificar as distorções estruturais observadas visualmente em cada modo de interação, foram monitorados alguns dos ângulos torcionais da cadeia principal, bem como uma série de parâmetros estruturais da dupla hélice que serão apresentados e discutidos nas seções a seguir (4.2.4).

O DNA possui seis ângulos torcionais na cadeia principal (Figura 19), o que torna a análise conformacional de ácidos nucleicos bastante mais complexa do que de proteínas.^{49,50} No entanto, alguns destes ângulos são mais importantes do que outros da descrição das conformações. A combinação dos ângulos α e γ , por exemplo, é particularmente relevante na descrição de conformações não-canônicas de baixo *Twist*, que são importantes durante o reconhecimento do DNA por proteínas e outras moléculas.⁷⁶ Além disso, o acúmulo de conformações α/γ não-canônicas pode levar a

fortes distorções estruturais da dupla hélice e, por isso, o monitoramento destes ângulos se torna também um critério de estabilidade em simulações por dinâmica molecular.⁷⁷

Além dos ângulos α e γ , foram também monitorados os ângulos ε e ζ , cuja combinação define as duas sub-formas do B-DNA canônico, conhecidas como B-I (mais frequente) e B-II (menos frequente).⁷⁸ Além disso, os ângulos ε e ζ são os que apresentam maior variabilidade no B-DNA⁷⁹ – sendo, portanto, mais flexíveis. Dessa forma, o monitoramento dos ângulos ε e ζ não é exatamente um critério de estabilidade, mas sim uma maneira de analisar a flexibilidade do DNA. Considerando que a flexibilidade do DNA é importante tanto na discriminação entre ligantes como nas adaptações estruturais exigidas pela ligação, é importante que ocorra uma amostragem equilibrada entre as conformações BI e BII nas simulações.

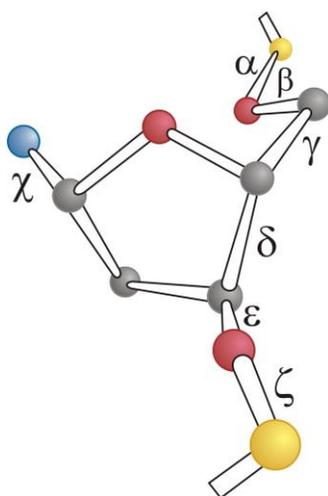


Figura 19. Ângulos torcionais da cadeia principal (de α a ζ) e da ligação glicosídica (χ).

Como α e γ são correlacionados entre si, foram plotados gráficos mostrando a distribuição de α versus γ para cada sistema (Figura 20), com as regiões de mínimo global e de mínimos locais indicadas pelas linhas sólidas e tracejadas, de acordo com Varnai *et al.*⁷⁶ A fim de ilustrar a frequência bem como a reversibilidade das transições conformacionais, a Figura 21 mostra a análise temporal dos ângulos α/γ para uma das bases de cada oligonucleotídeo simulado.

Também foram plotados gráficos de ε versus ζ para cada sistema (Figura 22), com as regiões correspondentes às conformações BI e BII indicadas de acordo com Schneider *et al.*⁷⁸ Uma distinção mais grosseira entre as duas conformações pode ser feita através do parâmetro $\varepsilon-\zeta$, sendo BI quando $\varepsilon-\zeta < 0$ e BII quando $\varepsilon-\zeta > 0$, que é o que está indicado pela linha diagonal tracejada nos gráficos da Figura 22. A análise temporal para os ângulos ε e ζ está mostrada na Figura 23.

Nas figuras 20 e 22, foram representados os ângulos dos pares de bases que compõem a tríade de interação com as bases de Tröger. Já os ângulos dos pares de bases não envolvidos diretamente na interação com os ligantes foram concatenados em um mesmo gráfico que descreve a região chamada de *região não-interagente*.

4.2.3.1 Ângulos α e γ

Sistemas isolados (DNA-GAP e DNA-SEM-GAP)

Nas simulações do DNA com *gap* isolado (Figura 20, A e B), os ângulos α/γ popularam principalmente a região g^-/g^+ , que consiste no mínimo global da superfície de energia e, portanto, corresponde à conformação canônica. No entanto, também foram observadas algumas transições para as outras regiões conhecidas como mínimos locais, principalmente as regiões g^-/t e g^+/t . De fato, o campo de força parm99, no qual está baseado o ffamber03, é conhecido por levar o sistema demasiadamente ao mínimo local g^+/t , mas o acúmulo dessas transições só gera distorções severas em simulações muito longas (>20ns).⁷⁷ Além disso, embora o DNA tenha visitado regiões de mínimo local, o sistema não ficou “preso” em nenhuma dessas conformações não-canônicas, já que as transições foram reversíveis, como mostra a análise temporal da Figura 21A.

Na simulação do DNA sem *gap* isolado (Figura 20, C e D), o sistema também se manteve principalmente na região de mínimo global, embora tenha ocorrido maior exploração das regiões de mínimo local em comparação com o oligonucleotídeo com *gap*. Além disso, pôde-se observar uma pequena tendência a popular a região t/g^+ , que não é descrita como um mínimo local pela literatura, podendo ser um artefato do campo de força.

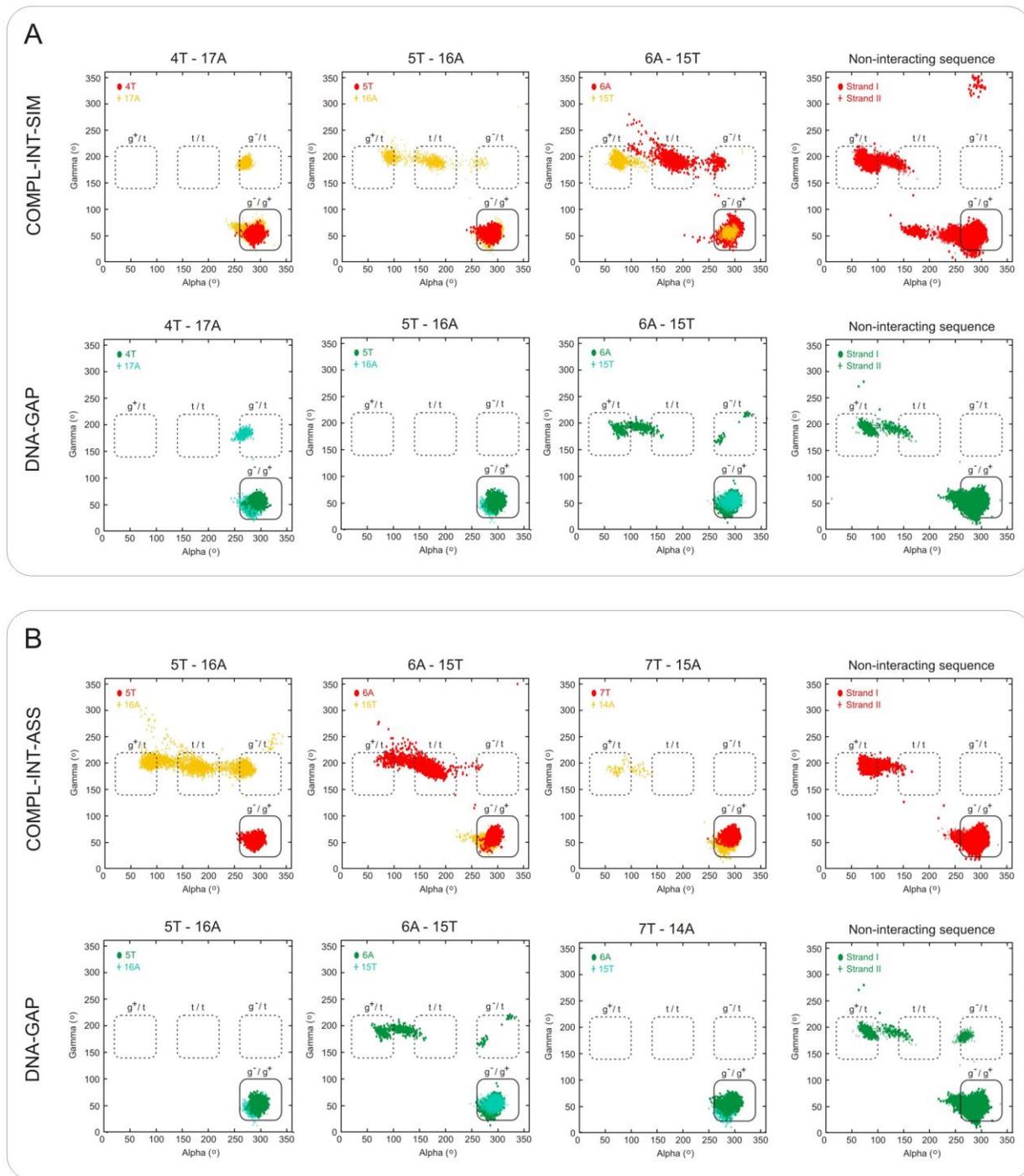


Figura 20. Perfis dos ângulos α e γ para os complexos de intercalação com a Tröger simétrica (A) e com a Tröger assimétrica (B). Em vermelho e amarelo, estão mostrados os perfis para o DNA no complexo; em verde e ciano os perfis para o DNA quando simulado isoladamente. As regiões do DNA que não interagiram diretamente com a base de Tröger durante a maior parte das simulações, foram chamadas de regiões não-interagentes.

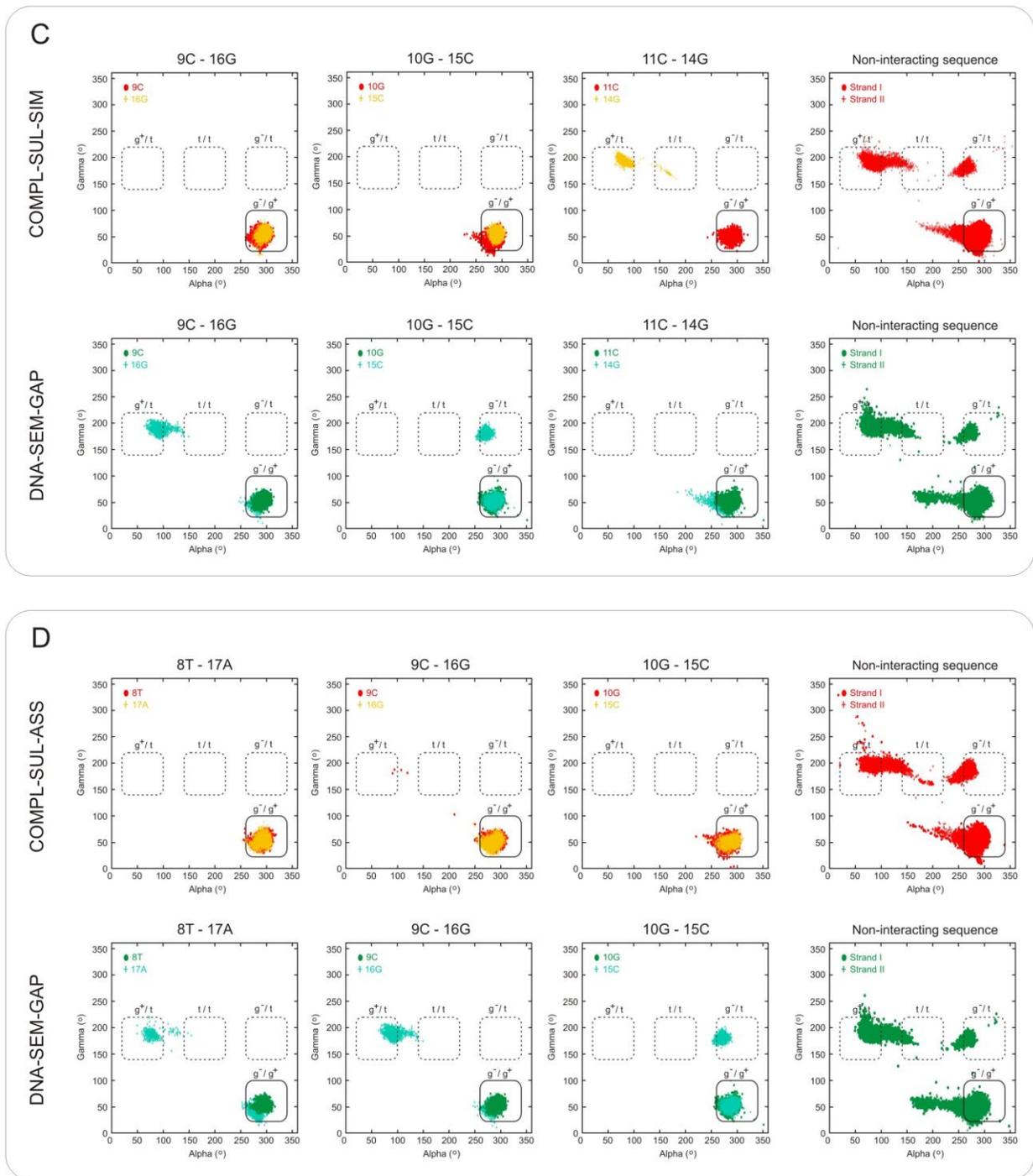


Figura 20 (continuação). Perfis dos ângulos α e γ para os complexos de ligação ao sulco com a Tröger simétrica (C) e com a Tröger assimétrica (D). Em vermelho e amarelo, estão mostrados os perfis para o DNA no complexo; em verde e ciano os perfis para o DNA quando simulado isoladamente.

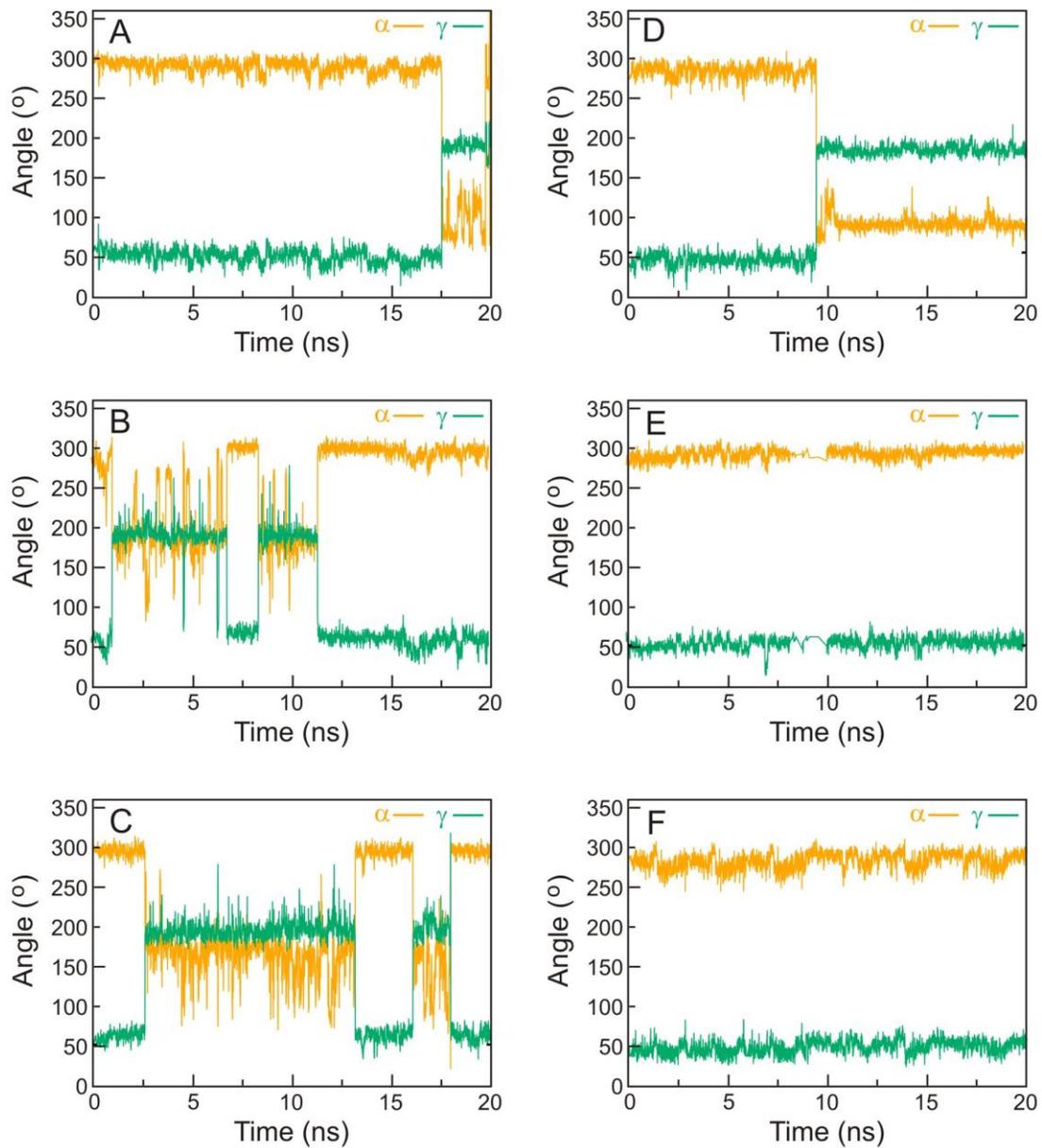


Figura 21. Análise temporal dos ângulos α (laranja) e γ (verde). Na coluna da esquerda, estão mostrados os resultados para a base 6A nos sistemas: DNA-GAP (A); COMPL-INT-SIM (B); COMPL-INT-ASS (C). Na coluna da direita, estão mostrados os resultados para a base 16G nos sistemas: DNA-SEM-GAP (D); COMPL-SUL-SIM (E); COMPL-SUL-ASS (F).

Complexos de intercalação (COMPL-INT-SIM e COMPL-INT-ASS)

Claramente, as alterações mais significativas dos ângulos α/γ nos oligonucleotídeos dos complexos de intercalação ocorreram justamente nas bases que flanqueiam o *gap* de intercalação (Figura 20, A e B). Por outro lado, poucas alterações foram observadas nos perfis das bases que interagem somente com a porção da base de Tröger que não se encontra intercalada (4T-17A, no caso da Tröger simétrica; 7T-15A, no caso da Tröger assimétrica).

Para as bases que flanqueiam o *gap* de intercalação (5T-16A e 6A-15T), foi bastante significativo o aumento da população de pontos nas regiões de mínimo local, principalmente na região trans/trans - que quase não foi visitada quando o oligonucleotídeo foi simulado isoladamente. Apesar de corresponder a um mínimo local, a conformação trans/trans é associada a conformações de baixo *Twist* que diferem do B-DNA canônico, sendo bastante característica da forma A do DNA.⁸¹

No complexo com a Tröger simétrica (Figura 20, A), o perfil correspondente à região das bases não-interagentes também mostrou um aumento da amostragem da região trans/trans e de outras regiões (não-mínimos), em comparação com o oligonucleotídeo isolado. Isso confirma a suposição feita a partir da análise visual das estruturas (seção 4.2.2), de que as alterações estruturais originadas no *gap* se propagaram ao longo do restante da cadeia principal neste complexo.

De acordo com a análise temporal (Figura 21, B e C), as transições também se mostraram reversíveis nestes sistemas, com frequências maiores em relação aos sistemas contendo DNA isolado.

Complexos de ligação ao sulco menor (COMPL-SUL-SIM e COMPL-SUL-ASS)

Nos complexos em que as bases de Tröger interagiram através de ligação ao sulco menor, não houve diferenças muito significativas nos perfis α/γ das tríades de interação. O par de bases 11C-14G, no complexo com a Tröger simétrica (Figura 20C), é o único caso em que houve amostragem de uma região de mínimo local (g^+/t). Para o restante dos pares de bases das tríades, os sistemas se mantiveram estabilizados na região de mínimo

global (g^-/g^+). De fato, houve mais transições para os mínimos locais g^-/t e g^+/t na tríade do oligonucleotídeo isolado do que na dos oligonucleotídeos que interagiram com as bases de Tröger. Isso não necessariamente é um efeito de estabilização provocado pelas bases de Tröger, mas indica que a presença do ligante no sulco menor não induz o DNA a adotar conformações α/γ não-canônicas, possivelmente inibindo essas transições.

Também não houve nenhum efeito de distorção nas regiões adjacentes à ligação, já que os perfis das regiões não-interagentes foram muito semelhantes no DNA isolado e nos complexos.

Em resumo, através da análise conformacional dos ângulos α e γ , pôde-se confirmar algumas tendências que já haviam sido observadas visualmente nas estruturas da simulação. Quando intercaladas, as bases de Tröger de fato afetam mais intensamente a cadeia principal do que quando ligadas ao sulco, levando o DNA a adotar conformações α/γ não canônicas com frequência maior do que o usual. Esse aumento ocorre principalmente nas transições para a região trans/trans, provavelmente como uma adaptação estrutural ao *gap* de intercalação. No caso do complexo com a Tröger simétrica, esse efeito não se restringe somente ao sítio de interação, mas se propaga ao longo da cadeia principal. O mesmo não ocorre quando as bases de Tröger interagem através de ligação ao sulco menor, pois não se observa nenhuma tendência de o sistema visitar regiões não-canônicas com frequência maior do que o DNA isolado, muito antes pelo contrário.

4.2.3.2 Ângulos ϵ e ζ

Diferentemente da análise dos ângulos α e γ , não é possível chegar a conclusões claras em relação ao efeito do ligante sobre os ângulos ϵ e ζ . De modo geral, foram observadas transições BI \rightarrow BII para todos os sistemas, o que indica que a flexibilidade da cadeia principal está sendo amostrada durante a simulação (Figura 22, A-D). Como mostrado na Figura 23, essas transições também foram reversíveis, porém mais frequentes do que as transições α/γ , o que está de acordo com a maior variabilidade relatada para os ângulos ϵ e ζ na literatura.⁷⁹

No entanto, não foi possível observar diferenças significativas entre os perfis dos oligômeros que interagiram com os ligantes e os perfis dos oligômeros simulados isoladamente. Dessa forma, concluiu-se que a presença das bases de Tröger, tanto quando intercaladas como quando ligadas ao sulco menor, não alterou significativamente o perfil dos ângulos ϵ e ζ . É mais provável, portanto, que as alterações que foram observadas visualmente na cadeia principal do DNA nos complexos de intercalação (seção 4.2.2) estejam correlacionadas com os ângulos α e γ , e não com os ângulos ϵ e ζ .

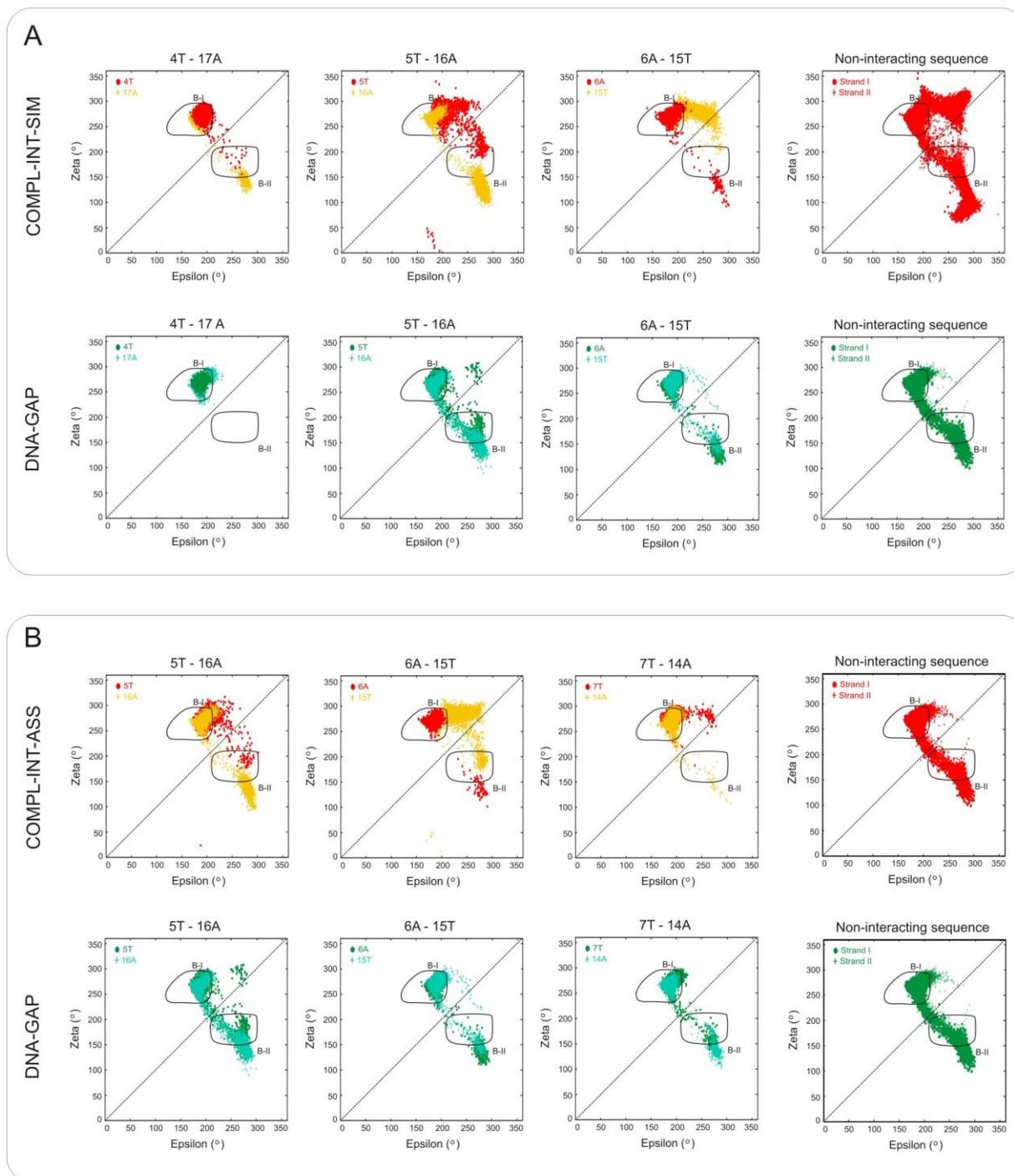


Figura 22. Perfis dos ângulos ϵ e ζ para os complexos de intercalação com a Tröger simétrica (A) e com a Tröger assimétrica (B). Em vermelho e amarelo, estão mostrados os perfis para o DNA no complexo; em verde e ciano os perfis para o DNA quando simulado isoladamente. As regiões do DNA que não interagiram diretamente com a base de Tröger durante a maior parte das simulações, foram chamadas de regiões não-interagentes.

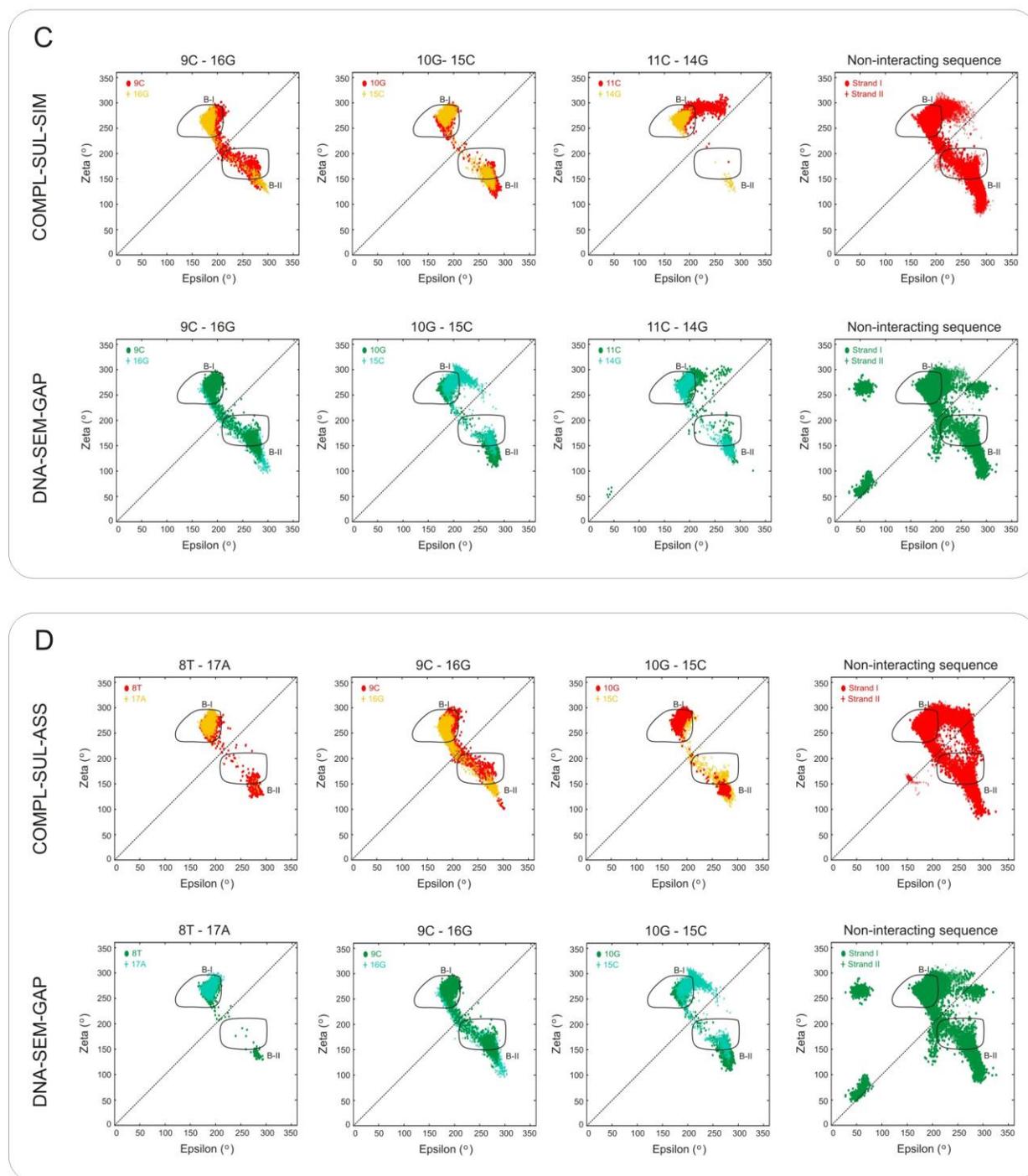


Figura 22 (continuação). Perfis dos ângulos ϵ e ζ para os complexos de intercalação com a Tröger simétrica (C) e com a Tröger assimétrica (D). Em vermelho e amarelo, estão mostrados os perfis para o DNA no complexo; em verde e ciano os perfis para o DNA quando simulado isoladamente.

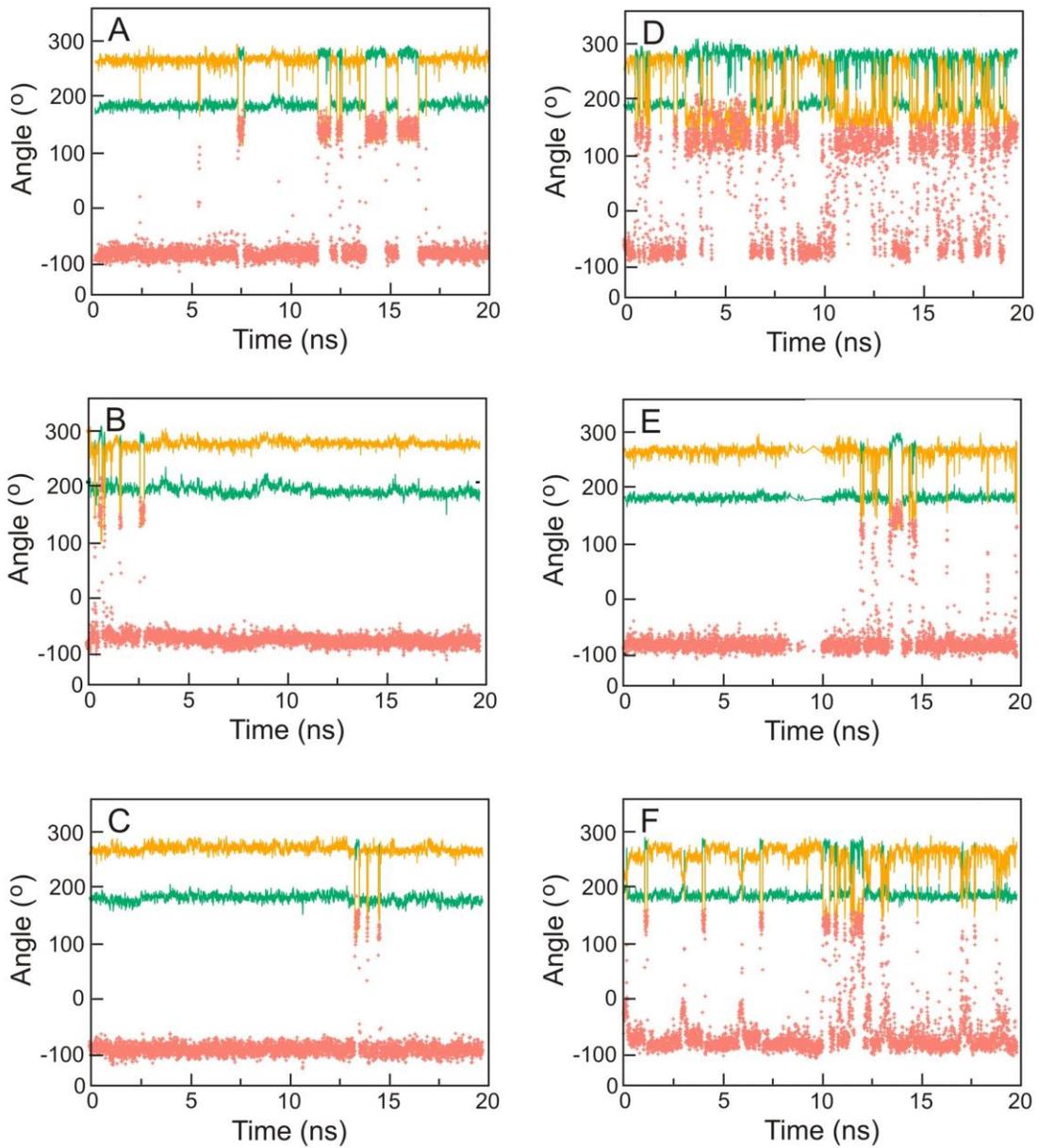


Figura 23. Análise temporal dos ângulos ϵ (verde) e ζ (laranja) e do parâmetro $\epsilon - \zeta$, mostrado em rosa. Na coluna da esquerda, estão mostrados os resultados para a base 6A nos sistemas: DNA-GAP (A); COMPL-INT-SIM (B); COMPL-INT-ASS (C). Na coluna da direita, estão mostrados os resultados para a base 16G nos sistemas: DNA-SEM-GAP (D); COMPL-SUL-SIM (E); COMPL-SUL-ASS (F).

4.2.4 Análise de parâmetros de pares de bases

Para quantificar algumas das distorções estruturais identificadas visualmente nas estruturas dos oligonucleotídeos – como desenrolamentos, curvaturas e a presença do *gap* –, foram selecionados alguns parâmetros de pares de bases, que estão ilustrados na Figura 24, de acordo com Dickerson *et al.*⁸² Esses parâmetros descrevem movimentos de translação ou de rotação de um par de bases em relação ao par de bases vizinho. Os parâmetros *Twist* e *Roll*, por exemplo, quantificam movimentos de rotação nos eixos *z* e *y*, respectivamente, enquanto que os parâmetros *Rise* e *Slide* quantificam movimentos de translação sobre os eixos *z* e *y*.

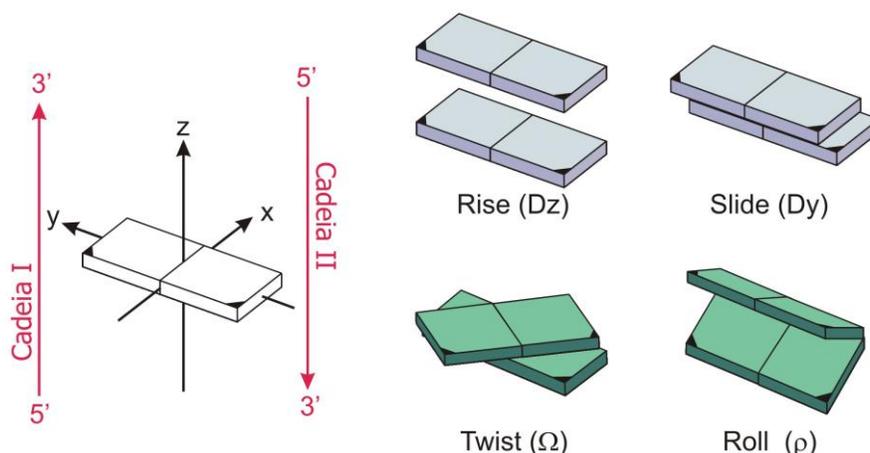


Figura 24. Parâmetros de pares de bases (de acordo com a convenção proposta por Dickerson *et al.*).⁸²

Esses parâmetros são importantes, pois é a combinação deles em cada passo de hélice que confere ao DNA a sua estrutura global. Um bom exemplo disso é o modelo de Calladine & Drew⁸³ para a transição B→A do DNA, que mostra que a combinação de valores negativos de *Slide* com valores positivos de *Roll* leva o DNA a adotar uma estrutura muito semelhante à forma A (Figura 25). Esse modelo de transição ilustra claramente o efeito da aplicação do parâmetro *Slide* na largura da hélice – levando a uma estrutura mais larga e achatada – e na profundidade dos sulcos – levando a uma assimetria de profundidade, em que o sulco menor é mais raso e o sulco maior mais

profundo. A aplicação de valores positivos de *Roll* também afeta as dimensões do sulco, não na profundidade, porém na largura – estreitando o sulco maior e alargando o menor.

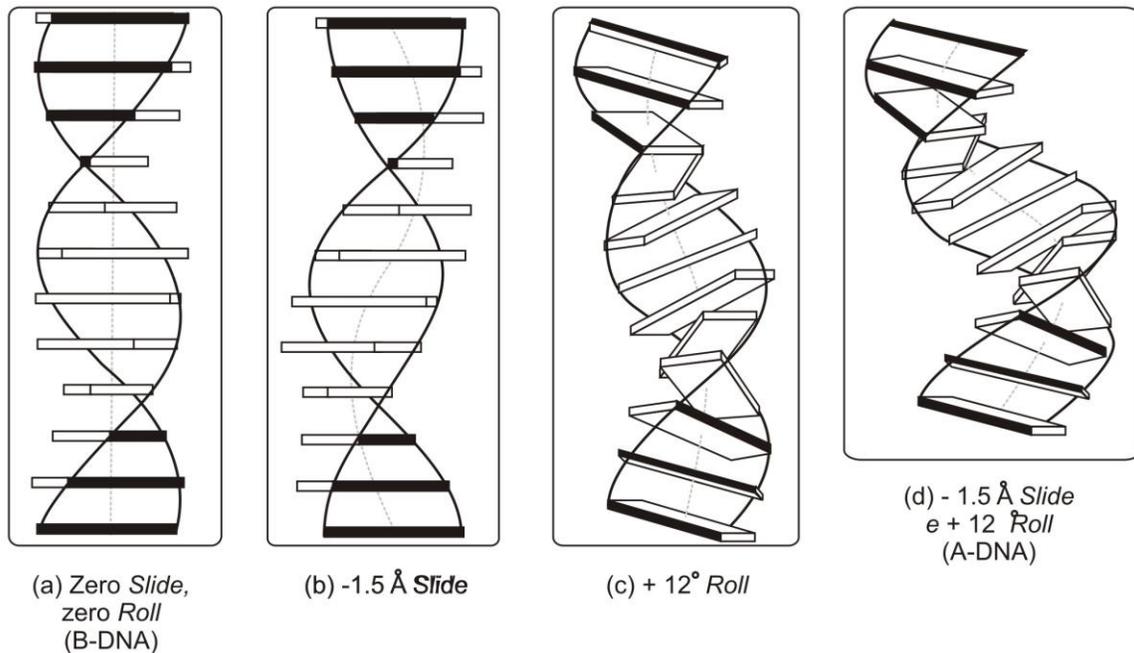


Figura 25. Modelo da transição B→A do DNA, por Calladine e Drew.⁸³ (a) B-DNA ideal. (b) Estrutura intermediária após aplicação uniforme de -1,5 Å de *Slide*. (c) Estrutura intermediária após aplicação uniforme de + 12° de *Roll*. (d) Estrutura semelhante ao A-DNA, após a aplicação de -1,5 Å de *Slide* e de + 12° de *Roll*. Adaptado de Dickerson and Ng.⁸⁴

Dessa forma, esses parâmetros são úteis para descrever mudanças na geometria dos sulcos, ocasionadas pela presença de um ligante. Além disso, a compressão ocasionada sobre um dos sulcos pela aplicação de *Roll* leva a uma curvatura local do DNA sobre o sulco menor (*Roll* negativo) ou sobre o sulco maior (*Roll* positivo)⁵³. Portanto, o parâmetro *Roll* também serve como um indicador para curvaturas que, dependendo da combinação, podem levar a uma curvatura global.

Já as contribuições dos parâmetros de *Twist* e *Rise* para a estrutura da dupla hélice são mais facilmente compreendidas. O parâmetro *Twist*, por quantificar o grau de torção da dupla hélice, é um bom indicador de possíveis desenrolamentos provocados, por exemplo, por agentes intercaladores. Já o parâmetro *Rise*, que mede a distância entre

pares de bases consecutivos, é um bom indicador da presença de um *gap* de intercalação, como já mencionado anteriormente.

Para esses quatro parâmetros selecionados – *Rise*, *Twist*, *Slide* e *Roll* –, foram monitorados os valores referentes aos passos de hélice contendo a tríade de interação, bem como o valor médio obtido a partir de todos os passos de hélice de cada oligonucleotídeo. (Cada passo de hélice significa dois pares de bases consecutivos, como ilustrado na Figura 24). Essas análises estão ilustradas nas Figuras 26 – 29 (*Rise*), 30-33 (*Twist*), 34-37 (*Slide*) e 38-41 (*Roll*). Em cada figura, são monitorados os parâmetros para o oligonucleotídeo simulado isoladamente (A) e no complexo (B). Além disso, estão indicados os valores esperados para as formas canônicas A e B, de acordo Reddy *et al.*⁸⁵

A análise do parâmetro *Rise* mostra que, na ausência de um ligante, o *gap* de intercalação do sistema DNA-GAP rapidamente se fechou em menos de 1 ns de simulação (Figuras 26 e 27, A). Na presença das bases de Tröger, porém, o alto valor de *Rise* para o par TA/TA mostra que o *gap* se manteve aberto durante toda a simulação, o que está de acordo com o alto tempo de residência bases de Tröger no *gap* de intercalação (Figura 26 e 27, B). Para os sistemas sem *gap* (Figuras 28 e 29), os valores de *Rise* se mantiveram, em média, próximos ao valor do B-DNA canônico.

Em todos os sistemas, os valores de *Twist* flutuaram bastante e se mantiveram, em média, abaixo do valor esperado para o B-DNA canônico - mais próximos do valor encontrado no A-DNA (Figuras 30-33). Isoladamente, esse efeito não indica uma transição B→A, mas decorre do próprio campo de força *AMBER*, que é conhecido por subestimar os valores de *Twist* em 3-4° em comparação com os valores cristalográficos.⁸⁶

No caso dos complexos de intercalação (Figuras 30 e 31, B), o passo de hélice contendo o *gap* de intercalação (linha amarela) apresentou valores extremamente baixos de *Twist*, indicando um forte desenrolamento da dupla hélice nessa região, o que decorre do maior afastamento dos pares de bases envolvidos no *gap*, como anteriormente comentado. No entanto, é interessante observar que, no caso da intercalação da Tröger simétrica (Figura 30, B), esse desenrolamento parece ter se propagado ao longo do restante da dupla hélice, pois se observou um progressivo decréscimo do valor médio, que terminou significativamente mais baixo do que o da forma A-canônica. O mesmo não ocorreu no complexo de intercalação da base assimétrica (Figura 31, B) que, mesmo

tendo apresentado um desenrolamento local na região do *gap*, manteve um valor médio de *Twist* próximo ao da forma A-canônica - de forma semelhante ao que ocorreu com o oligonucleotídeo isolado (Figura 31, A). Portanto, a análise do parâmetro *Twist* confirmou que, enquanto no complexo de intercalação da Tröger assimétrica o desenrolamento causado pelo ligante se restringiu à região do *gap*, no complexo da Tröger simétrica houve uma propagação desse efeito para outras regiões da hélice – o que já havia sido sugerido pela análise visual das estruturas na seção 4.2.2.

Com relação ao parâmetro *Slide*, a maioria dos sistemas manteve, em média, valores intermediários entre as formas A e B canônicas (Figuras 34-37), exceto pelo DNA no complexo de intercalação da Tröger simétrica, que apresentou valores mais positivos (Figura 34, B). Isso significa que, com exceção do COMPL-INT-SIM, houve em média um pequeno afastamento das bases em relação ao eixo da dupla hélice, tornando o sulco menor mais raso e o sulco maior mais profundo. Esse efeito, porém, dificilmente poderia ser identificado visualmente devido às altas flutuações ocorridas em cada passo de hélice - principalmente nos complexos de ligação ao sulco menor (Figuras 36 e 37, B).

Finalmente, em relação ao parâmetro *Roll*, pôde-se constatar que os oligonucleotídeos apresentaram valores médios próximos ao valor do B-DNA canônico (Figuras 38-41). No entanto, no caso dos complexos de ligação ao sulco menor, alguns dos passos de hélice envolvidos da região de interação apresentaram valores mais positivos no final das simulações, especialmente no COMPL-SUL-SIM (Figura 40, B). Isso indica que houve uma compressão sobre o sulco maior do DNA na região de interação com a Tröger simétrica, tornando o sulco menor mais raso e mais exposto à interação.

De modo geral, a análise de alguns parâmetros referentes aos pares de bases permitiu confirmar e quantificar algumas das distorções observadas visualmente nos oligonucleotídeos durante as simulações. Como esperado, as maiores alterações estruturais foram observadas para os oligômeros pertencentes aos complexos de intercalação, nos parâmetros de *Twist* e *Rise* - mostrando que a intercalação afeta principalmente a distância entre pares de bases que flanqueiam o ligante e o grau de torção da dupla hélice. Por outro lado, os complexos de ligação ao sulco mostraram

alterações estruturais mais sutis, que foram mais perceptíveis nos parâmetros de *Slide* e *Roll* – mostrando que esse tipo interação tende a afetar principalmente a largura e a profundidade dos sulcos.

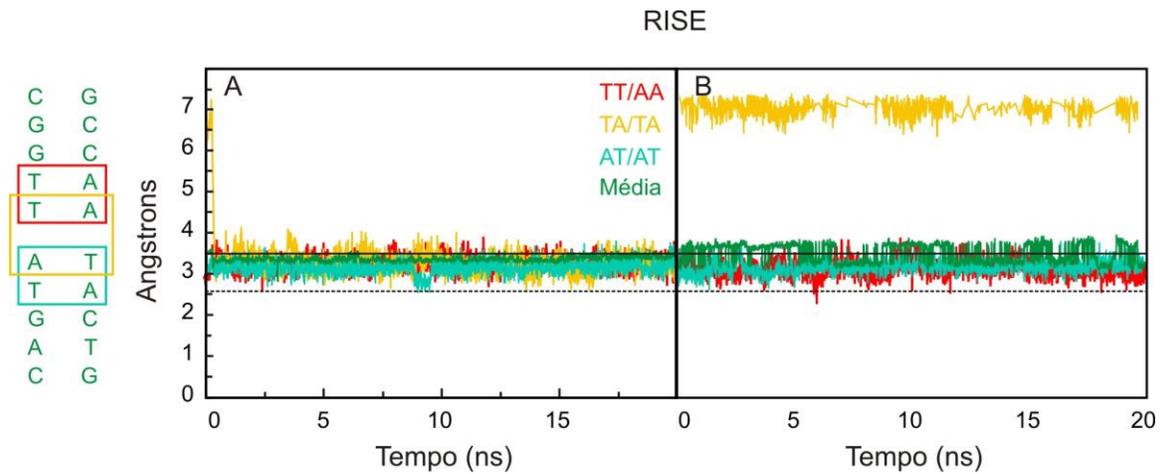


Figura 26. *Rise* para o DNA-GAP (A) e para o COMPL-INT-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

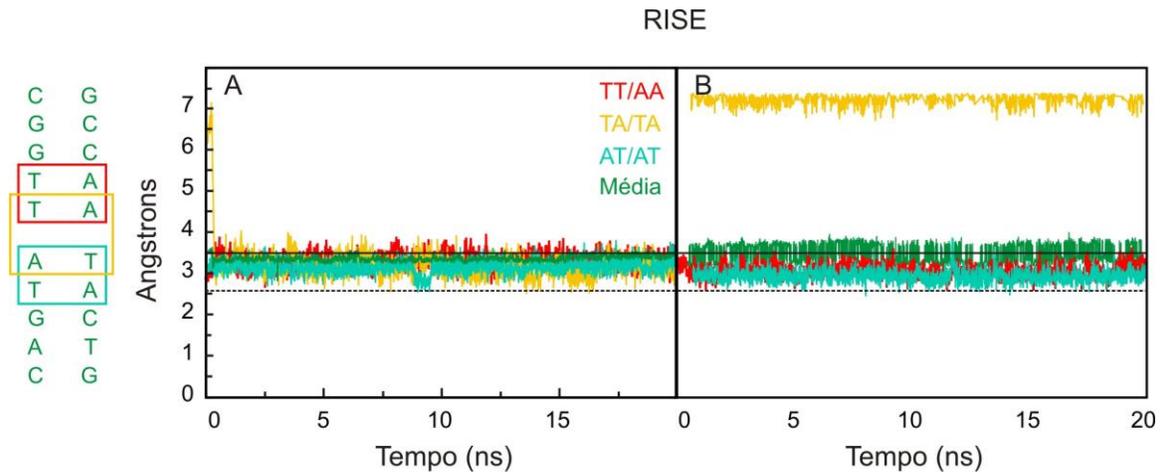


Figura 27. *Rise* para o DNA-GAP (A) e para o COMPL-INT-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

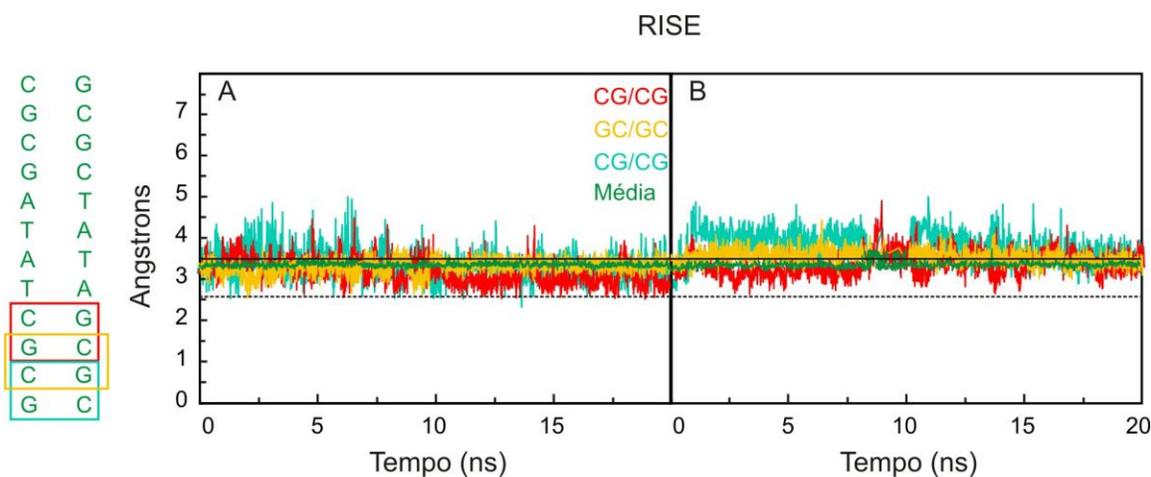


Figura 28. Rise para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

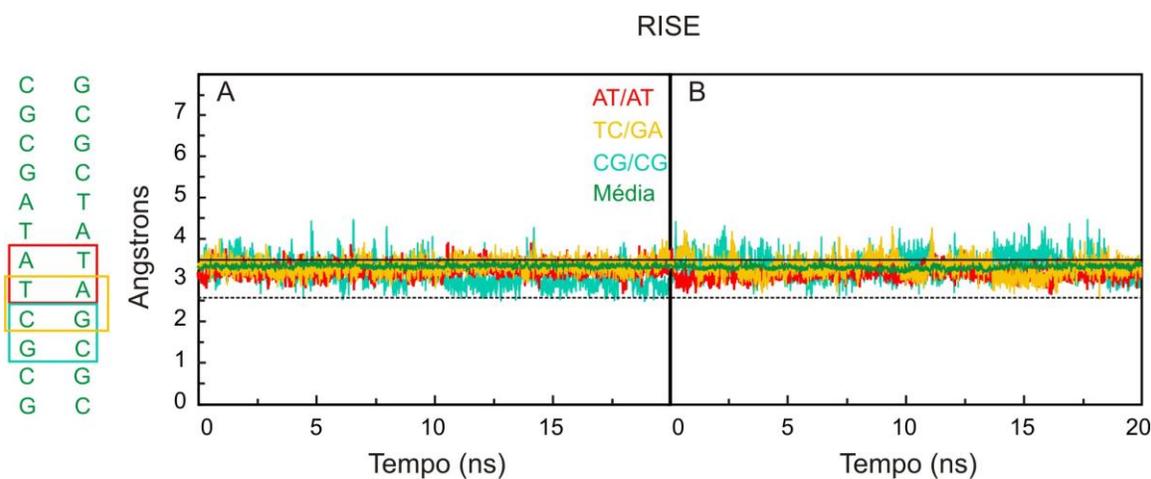


Figura 29. Rise para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

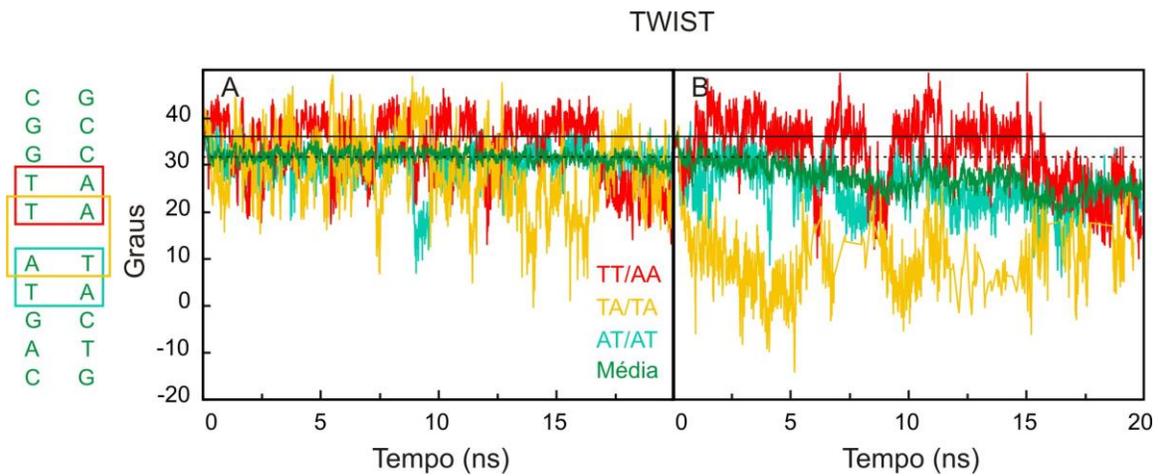


Figura 30. *Twist* para o DNA-GAP (A) e para o COMPL-INT-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

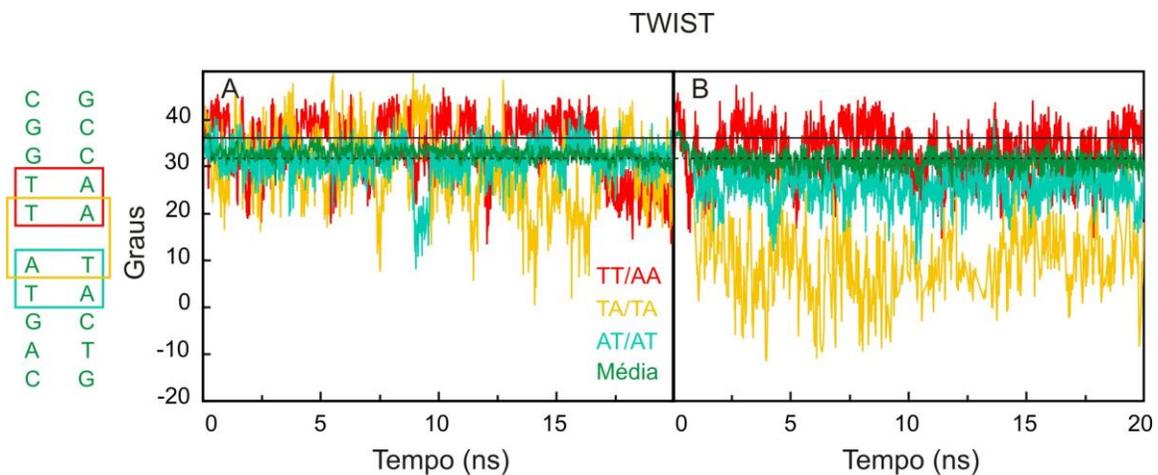


Figura 31. *Twist* para o DNA-GAP (A) e para o COMPL-INT-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

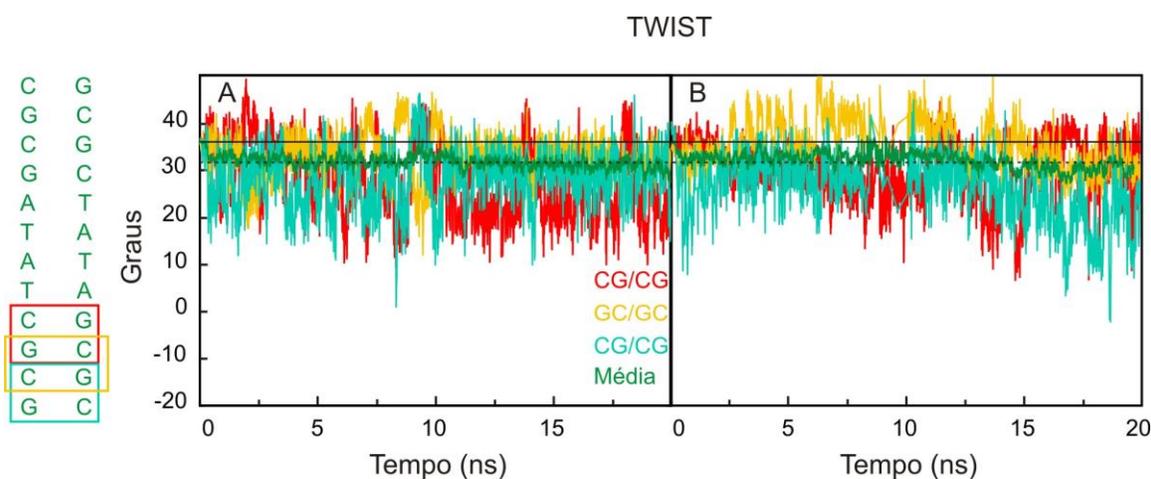


Figura 32. *Twist* para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

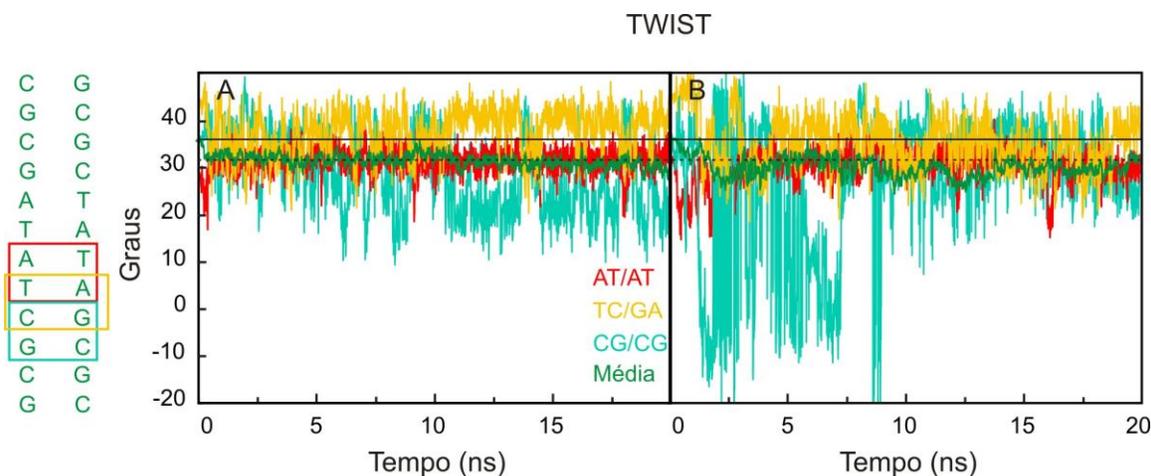


Figura 33. *Twist* para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

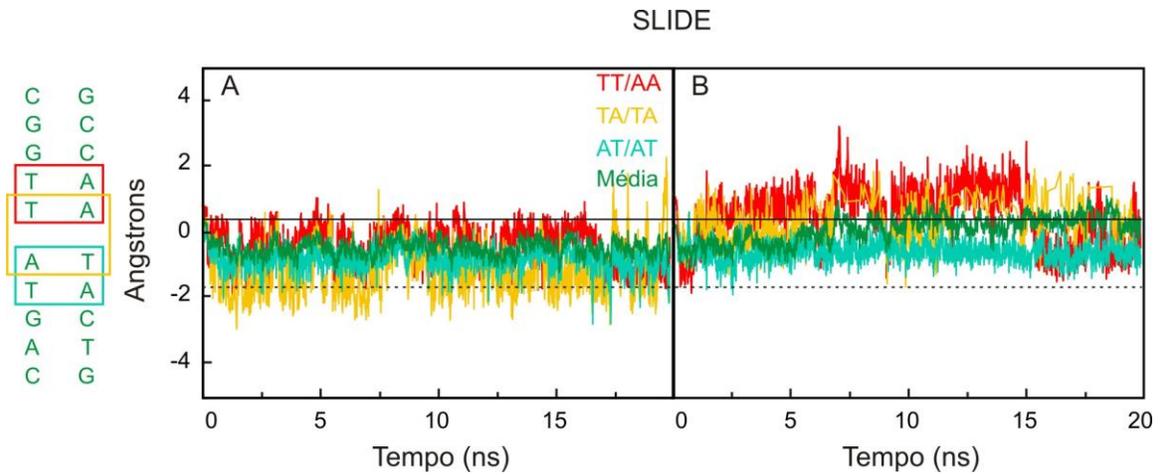


Figura 34. *Slide* para o DNA-GAP (A) e para o COMPL-INT-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

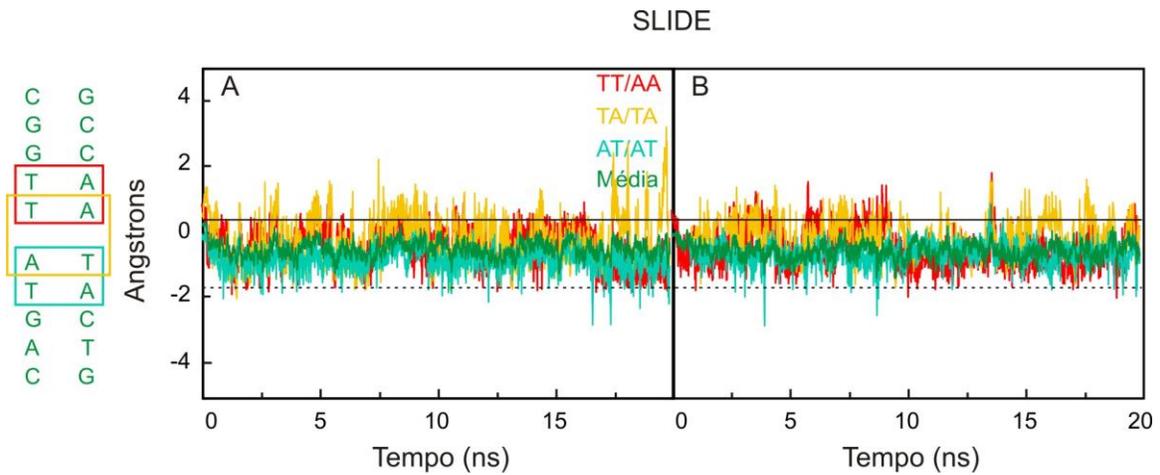


Figura 35. *Slide* para o DNA-GAP (A) e para o COMPL-INT-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

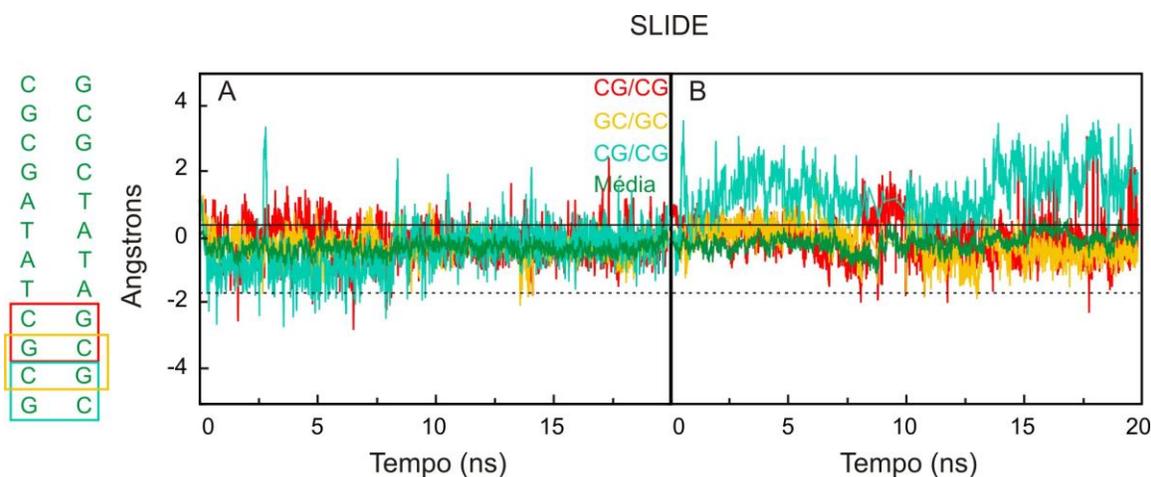


Figura 36. Slide para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

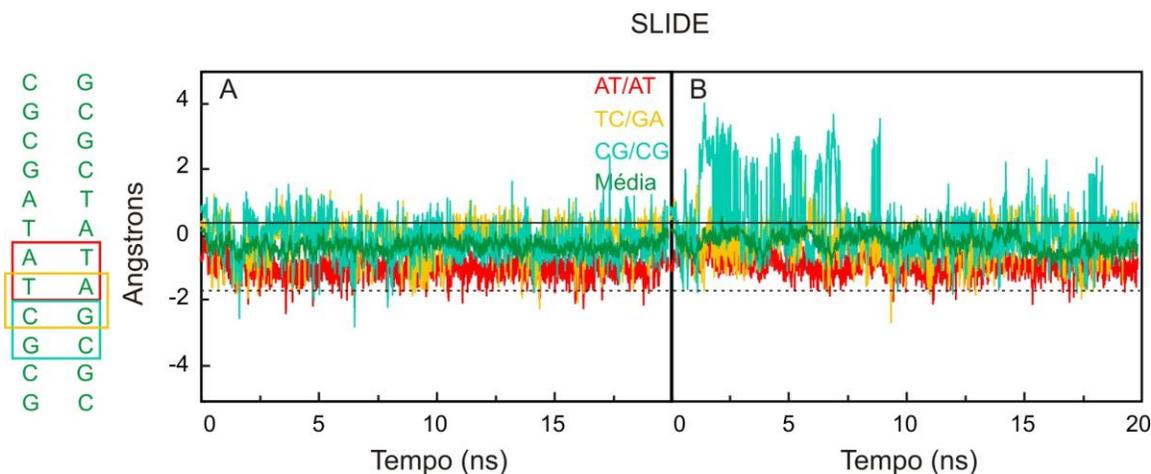


Figura 37. Slide para o DNA-SEM-GAP (A) e para o COMPL-SUL-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

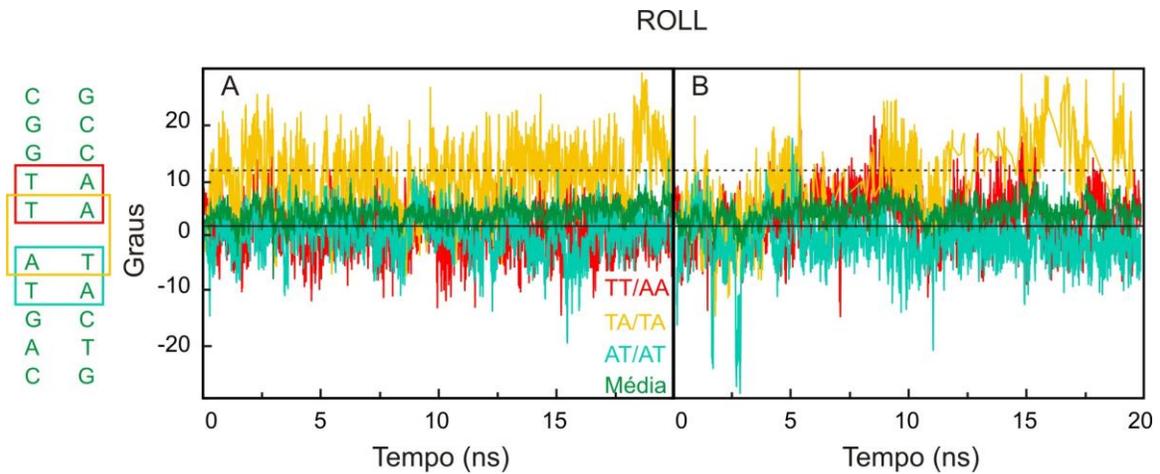


Figura 38. *Roll* para o DNA-GAP (A) e para o COMPL-INT-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

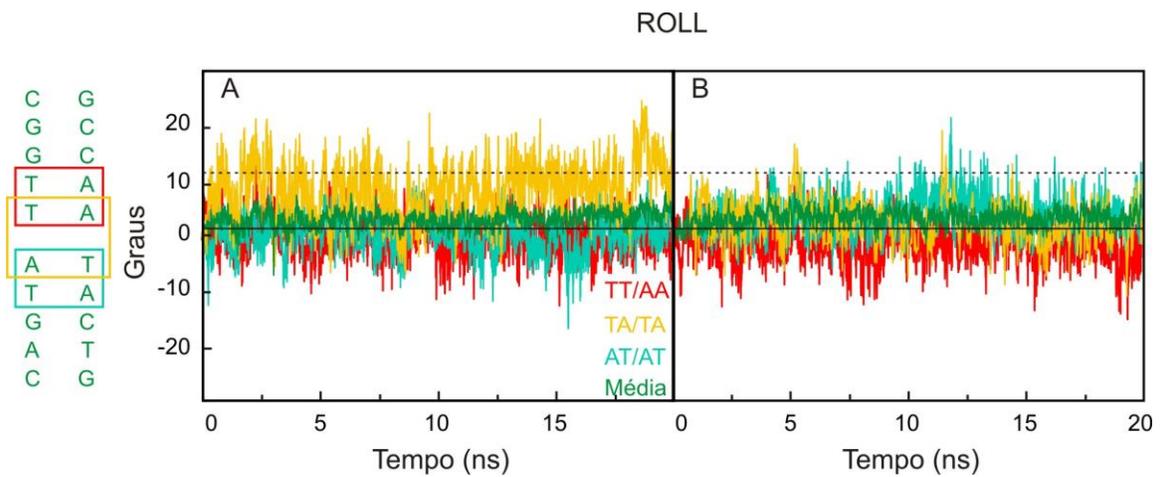


Figura 39. *Roll* para o DNA-GAP (A) e para o COMPL-INT-ASS (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

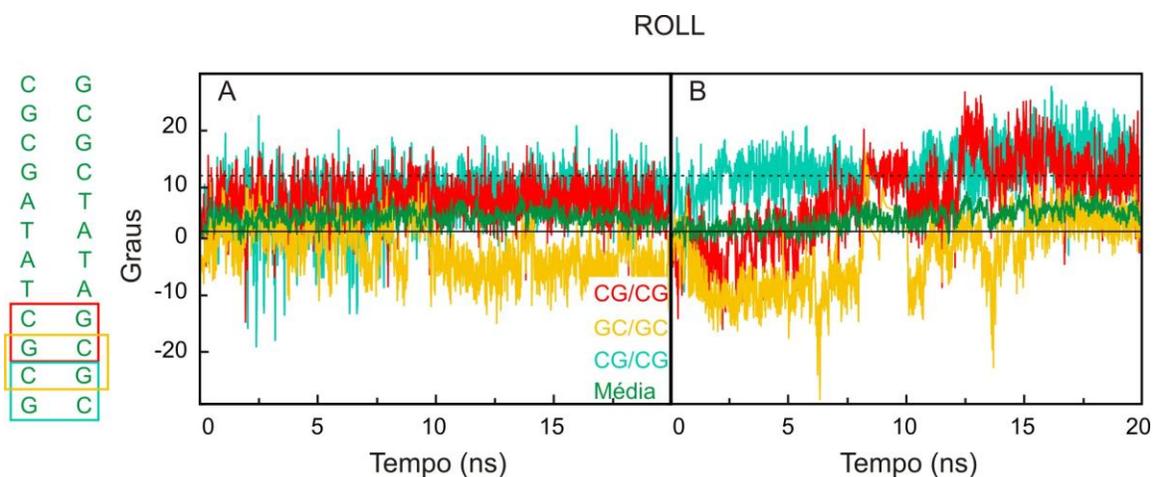


Figura 40. *Roll* para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

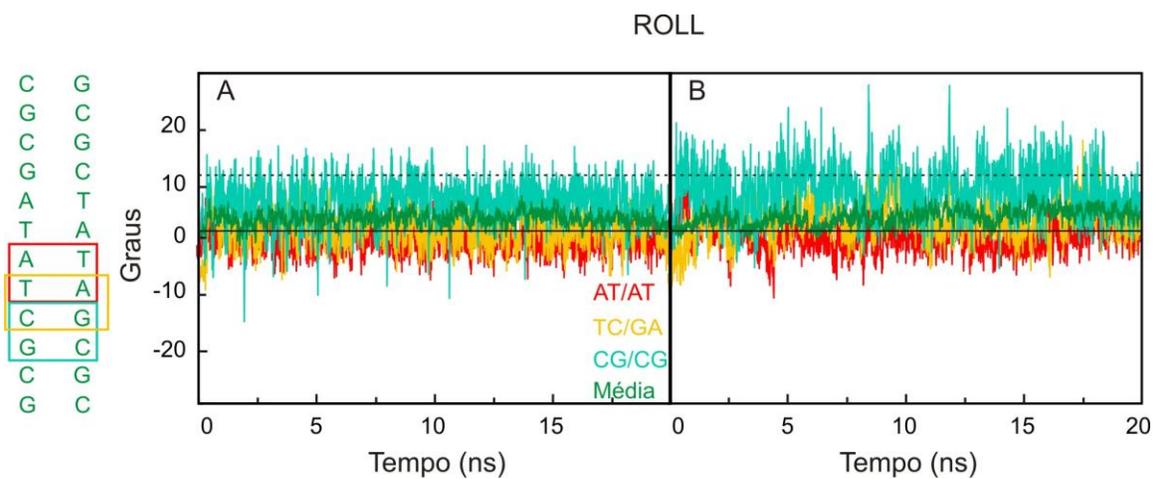


Figura 41. *Roll* para o DNA-SEM-GAP (A) e para o COMPL-SUL-SIM (B). Os valores médios para as formas B e A do DNA são indicados pela linha preta sólida e tracejada, respectivamente.

4.2.5 Análise dos *Bending-Dials*

Assim como o parâmetro *Roll* indica se o DNA está se comprimindo em direção a um dos sulcos, o parâmetro *Tilt* (ver Figura 42) indica se o DNA está se comprimindo em direção a uma das cadeias fosfatadas. Embora esse tipo de distorção seja energeticamente menos favorável e, portanto, menos frequente do que a compressão dos sulcos,⁵³ ele ainda assim deve ser considerado na análise da curvatura do DNA.

Uma das maneiras de realizar essa análise é através do parâmetro de *bending-dials*, que combina matematicamente os parâmetros de *Roll* e *Tilt* para cada passo de hélice, fornecendo um gráfico de coordenadas polares em que o ângulo ϕ indica a direção da curvatura e o parâmetro θ , a magnitude (Figura 42). (A fundamentação matemática dessa metodologia pode ser encontrada no trabalho de Young *et al.*⁸⁷). Dessa maneira, uma maior população do hemisfério norte indica uma curvatura local sobre o sulco maior, enquanto que uma maior população do hemisfério sul indica uma curvatura local sobre o sulco menor, como ilustrado na Figura 42. Os hemisférios leste e oeste, por sua vez, estão associados a curvaturas locais em direção a uma das cadeias fosfatadas. Como o parâmetro de *bending-dials* se restringe a um passo de hélice (ou seja, um par de pares de bases consecutivos), ele nos fornece apenas curvaturas locais, e a curvatura global deve ser interpretada a partir do conjunto formado pelos *bending-dials* para cada passo de hélice.

Os gráficos de *bending-dials* para os sistemas simulados estão ilustrados nas Figuras 43 (DNA-GAP e complexos de intercalação) e 44 (DNA-SEM-GAP e complexos de ligação ao sulco menor). Para descartar as flutuações decorrentes da mobilidade dos segmentos terminais da dupla hélice, os passos de hélice pertencentes às extremidades 5' e 3' foram excluídos da análise.

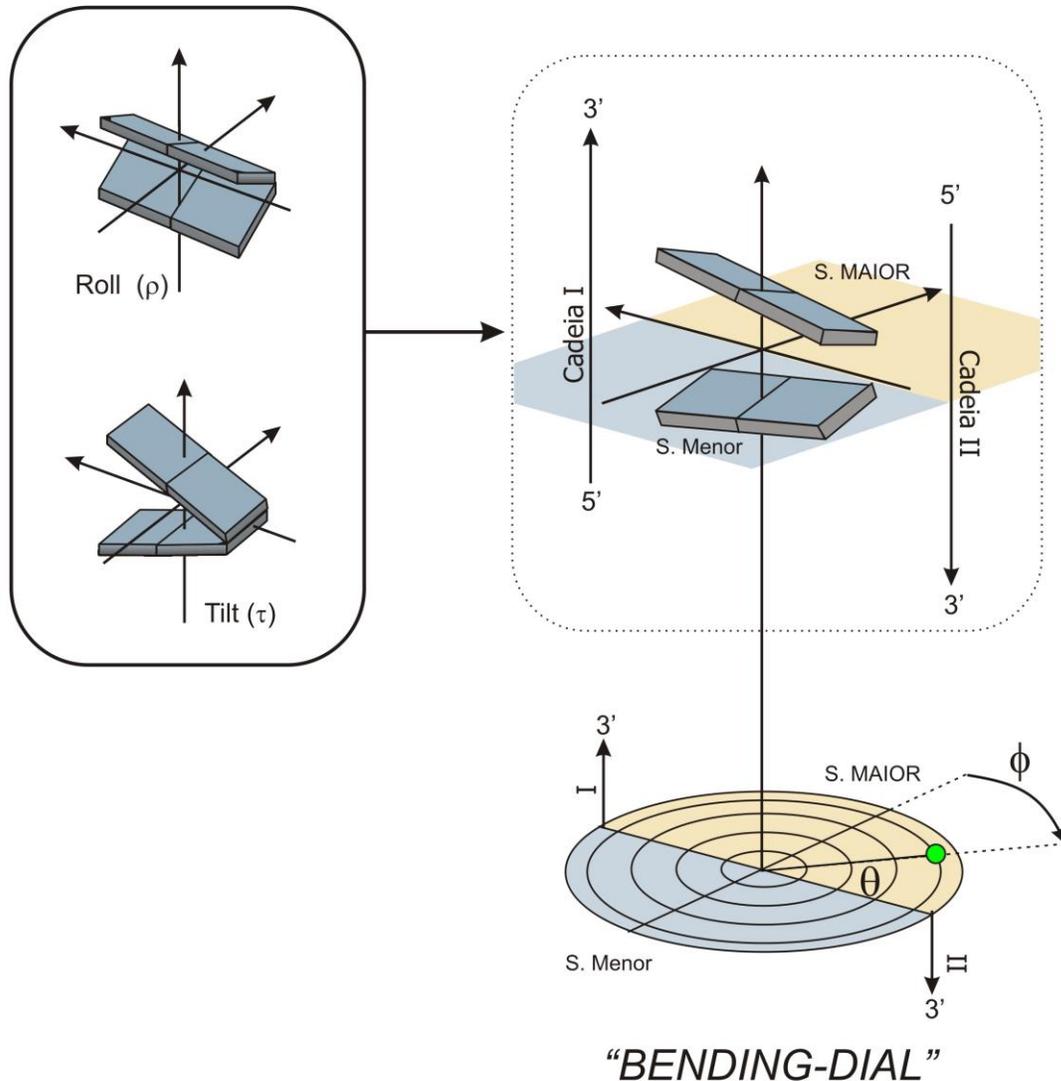


Figura 42. Combinação dos parâmetros *Roll* e *Tilt* para obtenção do parâmetro de *bending-dials*. A variação em *Roll* resulta em curvatura sobre um dos sulcos e a variação em *Tilt* resulta em curvatura sobre uma das cadeias fosfatadas. No gráfico polar, a direção e a magnitude da curvatura são indicadas pelo o ângulo ϕ e pelo parâmetro θ , respectivamente. Hemisfério norte – curvatura sobre o sulco maior; hemisfério sul – curvatura sobre o sulco menor; leste e oeste – curvatura sobre uma das cadeias fosfatadas. Adaptado de Young *et al.*⁸⁷

Sistemas contendo gap (DNA-GAP e complexos de intercalação)

No oligonucleotídeo isolado (DNA-GAP), a maioria dos passos de hélice não apresentou curvaturas locais, como pode ser observado pelos gráficos da Figura 43, com pontos pouco dispersos e simetricamente distribuídos entre os quatro hemisférios. As únicas exceções ocorreram nos passos TG/CA e TA/TA, que apresentaram curvaturas locais em direção ao sulco maior do DNA, como pode ser visto pela maior população do hemisfério norte do gráfico. Também pode ser observado pelos gráficos que a curvatura do passo TA/TA apresentou uma magnitude maior do que o passo TG/CA. Esse resultado é consistente com estudos experimentais, que mostram que passos do tipo TA/TA apresentam valores bastante positivos de *Roll*,^{88,89} induzindo curvaturas em direção ao sulco menor que são importantes, por exemplo, na interação da enzima TATA-Box Binding Protein (TBP) com sequências contendo o motivo TATA.^{55,90}

Essa curvatura do passo TA/TA em direção ao sulco maior também ocorreu no complexo de intercalação da Troger simétrica – embora com uma adicional curvatura em direção a uma das cadeias fosfatadas – mas foi suprimida no complexo de intercalação da Tröger assimétrica. Ainda no complexo de intercalação da Tröger simétrica, a maior perturbação parece ter ocorrido no passo AT/AT (adjacente ao *gap*), que mostrou uma curvatura local em direção ao sulco maior que não foi observada no oligonucleotídeo isolado e, portanto, não é uma curvatura intrínseca da sequência. De modo geral, no entanto, a presença dos ligantes intercalados não promoveu curvaturas significativas na estrutura do DNA.

Sistemas sem gap (DNA-SEM-GAP e complexos de ligação ao sulco menor)

No oligonucleotídeo isolado (DNA-SEM-GAP), além da curvatura já esperada para o passo TA/TA, também foram observadas curvaturas em direção ao sulco maior para os passos CG/CG (Figura 44). No entanto, diferentemente do que ocorreu no passo TA/TA – em que os passos adjacentes (AT/AT) não apresentaram nenhuma curvatura em relação ao eixo central – no caso dos passos CG/CG, a curvatura em direção ao sulco maior parece ter sido compensada, pelo menos parcialmente, pelos passos adjacentes

GC/GC – que apresentaram curvaturas locais na direção contrária, ou seja, em direção ao sulco menor.

Em comparação com os perfis de *bending-dials* dos complexos de ligação ao sulco (COMPL-SUL-SIM e COMPL-SUL-ASS), podemos concluir que as alterações proporcionadas na curvatura da dupla hélice por esses ligantes se restringiram à região de interação (em amarelo). No caso do complexo com a Tröger simétrica (COMPL-SUL-SIM), os pontos pertencentes à região de interação mostraram uma maior população do hemisfério norte para o passo CG/CG e do hemisfério sul para o passo GC/GC. No entanto, é interessante observar que, em ambos os casos, os pontos apresentaram uma dispersão vertical que penetra tanto no hemisfério norte quanto no hemisfério sul. Isso significa que, durante a simulação, esses passos de hélice não se mantiveram sempre curvados em uma única direção, podendo ter variado a direção da curvatura diversas vezes. Esse resultado sugere, portanto, não exatamente um aumento da *curvatura* local nessa região, mas sim um aumento da *flexibilidade* – reflexo, provavelmente, de uma perturbação introduzida pela presença da Tröger simétrica.

No caso do complexo com a Tröger assimétrica (COMPL-SUL-SIM), as maiores perturbações também ocorreram nos passos CG/CG e GC/GC, que apresentaram curvaturas locais em direção ao sulco maior. Neste caso, porém, a direção das curvaturas se mostrou mais consistente ao longo da simulação (o que pode ser deduzido pelos poucos pontos no hemisfério sul dos gráficos). Como essas curvaturas locais não foram compensadas nos passos adjacentes, é bastante provável que seja apenas esse efeito de curvatura – e não uma transição conformacional – o responsável pela semelhança visual do DNA nesse sistema com a forma A (ver seção 4.2.2).

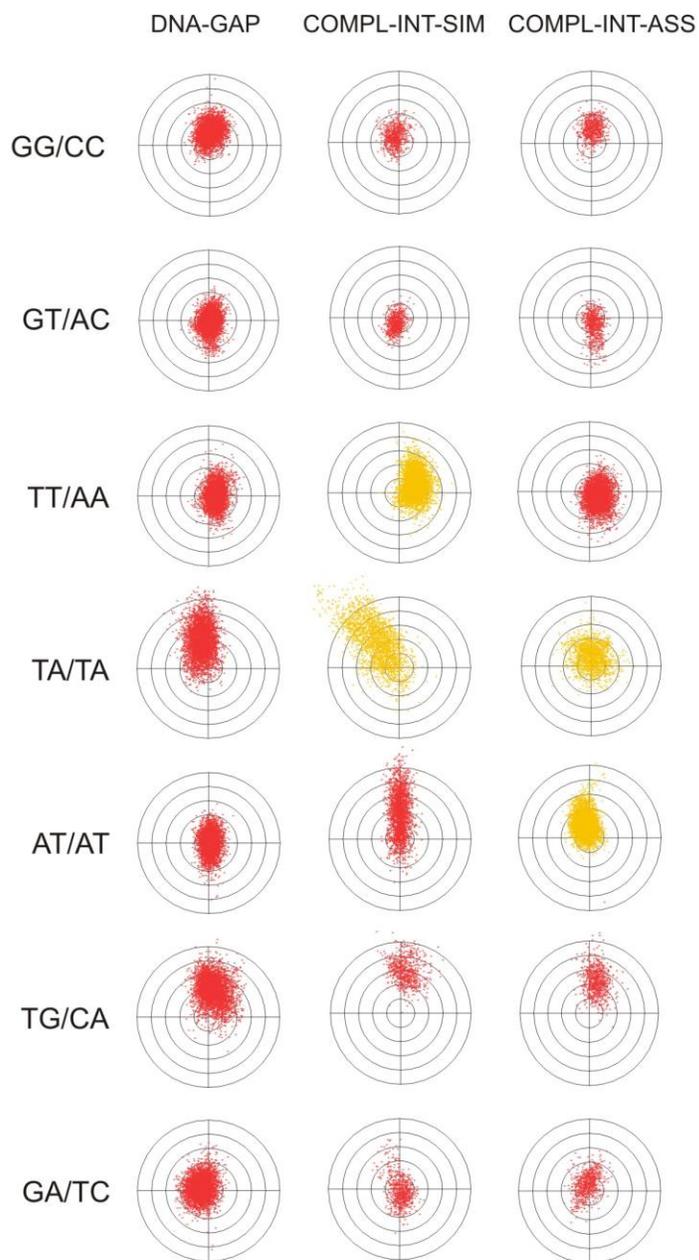


Figura 43. Análise dos *bending-dials* dos sistemas contendo *gap*. Cada um dos círculos concêntricos representa uma magnitude de curvatura de 5° em relação ao eixo central do dupla hélice. Em vermelho, passos de hélice de regiões não-interagentes; em amarelo, passos de hélice de regiões que interagiram com os ligantes.

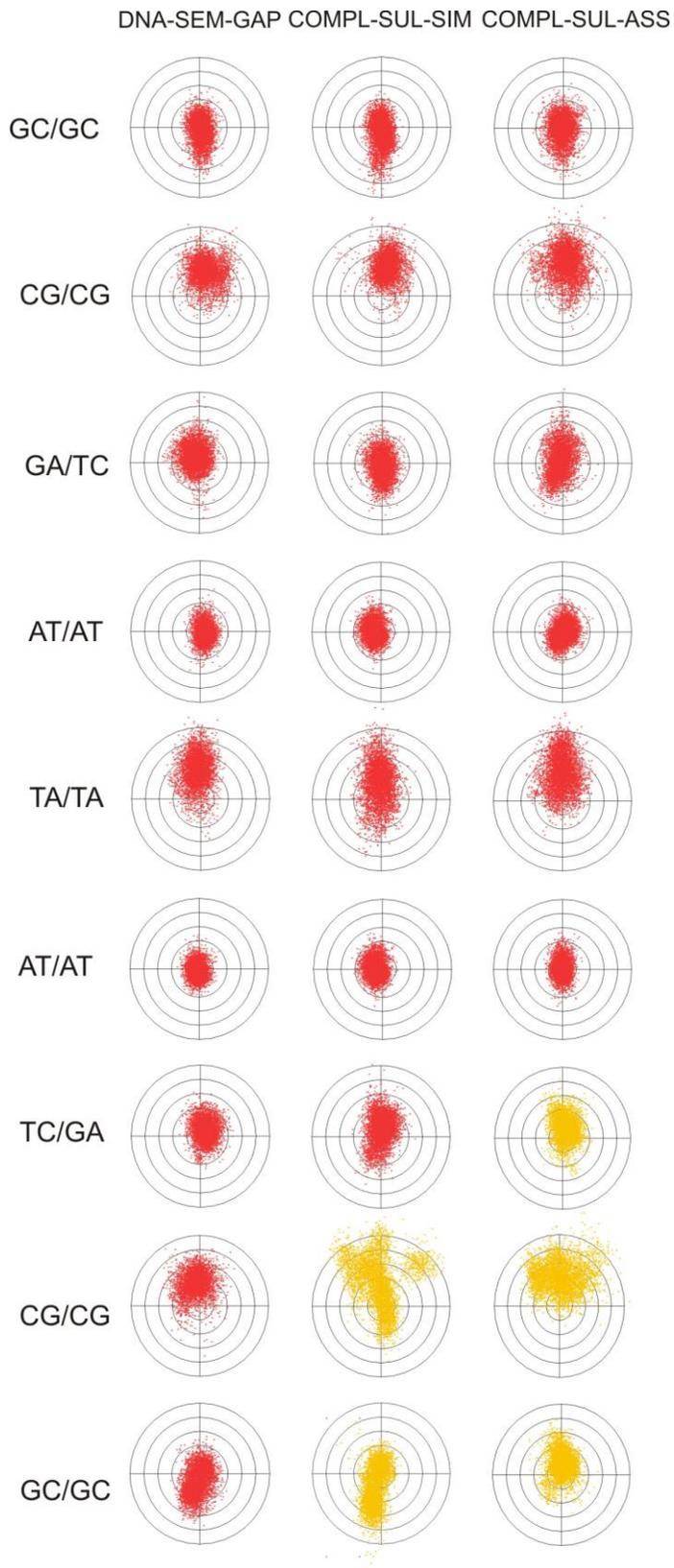


Figura 44. Análise dos *bending-dials* dos sistemas sem *gap*. Cada um dos círculos concêntricos representa uma magnitude de curvatura de 5° em relação ao eixo central da dupla hélice. Em vermelho, passos de hélice de regiões não-interagentes; em amarelo, passos de hélice de regiões que interagiram com os ligantes.

4.3 CONCLUSÕES

Foram realizadas simulações do tipo dinâmica molecular com quatro sistemas envolvendo as bases de Tröger simétrica e assimétrica: dois complexos de intercalação (com um substituinte interagindo no *gap* e o outro no sulco menor do DNA) e dois complexos de ligação ao sulco menor (com pelo menos um dos substituintes encaixado no sulco menor do DNA). Para poder discriminar entre os efeitos provocados pela presença dos ligantes e as propriedades intrínsecas do campo de força ou da composição (sequência) dos oligonucleotídeos, também foram simulados os oligonucleotídeos isoladamente.

Em todas as simulações, a dupla hélice do DNA se mostrou estável – ou seja, não houve desnaturação significativa dos pares de bases –, confirmando que o campo de força *ffamber03* é adequado para a simulação de ácidos nucléicos. Esse fato também foi confirmado por uma amostragem equilibrada das conformações da cadeia principal para os oligonucleotídeos isolados.

Nos complexos de intercalação, as bases de Tröger apresentaram tempos de residência bastante altos (>20 ns) e provocaram o desenrolamento da dupla hélice na região do *gap*, um efeito que é uma assinatura deste modo de interação. Através da análise dos ângulos torcionais da cadeia principal, foi possível mostrar que essas conformações locais de baixo *Twist* exigiram transições nos ângulos alfa e gama para uma região característica da forma A do DNA. No caso da Tröger simétrica, esse efeito de desenrolamento foi mais intenso, com propagação para as bases adjacentes ao sítio de interação.

Nos complexos em que as bases interagiram somente através da ligação ao sulco menor, as alterações na dupla hélice foram mais sutis e produziram alterações nas conformações da cadeia principal. Na realidade, a presença dos ligantes no sulco menor pareceu inibir as transições α/γ da cadeia principal, limitando o DNA à região B-canônica. Dentro da região B-canônica, porém, a presença dos ligantes parece ter aumentado levemente a flexibilidade intrínseca do B-DNA, como sugerido por um aumento das curvaturas locais.

Finalmente, embora as bases de Tröger tenham interagido com o sulco menor por 20 ns, os tempos de residência não foram tão grandes se considerarmos que houve mudança tanto do modo como do sítio de interação. Surpreendentemente, as simulações destes complexos convergiram para um modo de interação semelhante, em que as bases de Tröger estão ligadas ao DNA quase que exclusivamente através ponte diazocina, com os substituintes projetados para fora do sulco.

5. CONSIDERAÇÕES FINAIS

Neste trabalho, foi estudada a interação de duas bases de Tröger com o DNA, através de docagem e dinâmica molecular. Apesar das evidências experimentais que comprovam a capacidade dessas bases de interagir de forma enantiosseletiva com ácidos nucléicos, o modo de interação dessas moléculas ainda não foi elucidado de forma conclusiva. Além disso, como ambas contém conhecidos ligantes de DNA como substituintes – proflavina e fenantrolina – não está claro qual é o papel da estrutura em V da ponte diazocina durante a interação, que pode decorrer exclusivamente da capacidade de ligação intrínseca dos substituintes.

Devido à falta de resultados experimentais que fornecessem uma estrutura de partida adequada para as simulações do tipo dinâmica molecular, a metodologia de docagem molecular foi utilizada de forma preliminar para obtenção de complexos de interação mais prováveis entre as bases de Tröger e os oligonucleotídeos. Para isso, foi validado um protocolo de docagem para ácidos nucléico com o programa *AutoDock 4.0*, que mostrou que um oligômero canônico com um *gap* de intercalação artificial é uma alternativa adequada para contornar as limitações de flexibilidade da docagem com agentes intercaladores.⁶²

Os resultados da docagem das bases de Tröger com esse receptor se mostraram de acordo com os modos de ligação sugeridos na literatura a partir de estudos indiretos (ensaios bioquímicos e espectros de dicroísmo circular e linear): o enantiômero (-) da Tröger simétrica foi docado preferencialmente no sulco menor do DNA, enquanto que o enantiômero (-) da Tröger simétrica resultou em um modo misto de interação (com intercalação da fenantrolina e ligação ao sulco menor da proflavina). Além disso, os resultados da docagem reproduziram corretamente a estereosseletividade dos enantiômeros levorrotatórios em relação aos dextrorrotatórios.

Para as simulações, foram selecionados quatro complexos, cujos modos iniciais de interação estão indicados entre parênteses: 1) Tröger simétrica e DNA com *gap* (intercalação); 2) Tröger assimétrica e DNA com *gap* (modo misto); 3) Tröger simétrica e DNA sem *gap* (ligação completa ao sulco menor); e 4) Tröger assimétrica e DNA sem *gap* (ligação parcial ao sulco menor). De forma bastante interessante, embora os complexos

representassem quatro pontos de partida diferentes, houve convergência dos modos de interação entre aqueles complexos que continham o mesmo oligonucleotídeo como receptor.

O complexo da Tröger simétrica com DNA contendo *gap* rapidamente convergiu para um modo misto, com um dos substituintes intercalado no *gap* e o outro fazendo contatos adicionais no sulco menor, de forma análoga ao modo de interação da Tröger assimétrica com esse receptor. Embora o tempo de residência das bases no *gap* tenha sido alto, e as distorções observadas na estrutura do DNA tenham sido aquelas esperadas de um agente intercalador, esses resultados isoladamente não garantem que esse seja o modo de interação preferencial dos enantiômeros levorrotatórios das bases de Tröger. Para isso, seria necessário, não apenas mostrar que elas *permanecem* no *gap*, mas também que são moléculas capazes de *induzir* a abertura do *gap* de intercalação – caso contrário os resultados dessas simulações seriam meramente artefatos do ponto de partida utilizado.

Nesse sentido, é interessante salientar que, durante a simulação da base de Tröger simétrica com o oligonucleotídeo sem *gap*, os anéis aromáticos de uma das proflavinas permaneceram por aproximadamente 2 ns paralelos a dois pares de bases do DNA, numa orientação bastante favorável para intercalar. Considerando todas as alterações estruturais necessárias para a abertura do *gap*, é muito provável que esse evento envolva uma barreira energética demasiadamente alta, tornando muito pequena a probabilidade dessa transição ser amostrada em apenas 20 ns de simulação.

Somado a isso, temos o fato de que a proflavina isolada é sabidamente um agente intercalador e que, portanto, é capaz de induzir a abertura do *gap*. Se, mesmo ligada à base de Tröger, ela foi capaz de assumir uma orientação favorável para intercalar, é mais provável que a intercalação não tenha ocorrido de fato por uma limitação da amostragem – portanto, da metodologia.

Uma outra forma de argumentação seria pensar que, se proflavina possui a habilidade intrínseca de intercalar do DNA, só existem duas razões para que a intercalação não seja, pelo menos em parte, o modo de interação das bases de Tröger: i) de alguma maneira a estrutura da base de Tröger impede a proflavina de intercalar; ii) existe um modo de interação ainda mais favorável do que a intercalação (por exemplo, ligação ao sulco menor).

A primeira hipótese não parece ser o caso, pois a própria simulação dinâmica molecular mostrou que a interação de uma das proflavinas no sulco menor posicionou corretamente a outra proflavina para a intercalação e que, além disso, os complexos de intercalação se mostraram estáveis, sem que a base de Tröger fosse expulsa para fora do *gap*.

Resta então a hipótese de que exista um outro modo de interação ainda mais favorável do que a intercalação, como, por exemplo, a ligação ao sulco menor. Geometricamente, o formato em “V” da base de Tröger resulta em uma curvatura muito intensa na estrutura do ligante, que não parece ser compatível com a curvatura mais suave do sulco menor do DNA. Ainda assim, considerando a flexibilidade do DNA, poderia se pensar que a ligação ao sulco menor promoveria alterações na geometria do sulco, adaptando-o à curvatura de quase 90 graus entre os substituintes da Tröger - mas isso não foi o que se observou nas simulações. O DNA se manteve pouco alterado, e ambas as bases de Tröger convergiram para um novo modo de interação completamente diferente, em que a interação ocorre quase que exclusivamente através da ponte diazocina.

De forma contra-intuitiva, esse modo de interação não aproveita a potencial complementaridade geométrica existente entre o sulco menor e a base de Tröger e, portanto, se caracteriza em um modo de interação independente da quiralidade da base de Tröger e pouco dependente dos substituintes envolvidos. Poderia constituir, portanto, em um modo de ligação comum a todas bases de Tröger e responsável pela interação fraca documentada para os enantiômeros dextrorrotatórios, mas não poderia explicar a enantiosseletividade dos enantiômeros levorrotatórios.

De acordo com os nossos resultados, portanto, a ligação ao sulco menor não parece ser o modo de interação preferencial das bases de Tröger estudadas, o que diminui a possibilidade de que exista um modo de interação mais favorável do que a intercalação de um dos substituintes. Dessa maneira, os resultados obtidos neste trabalho sugerem – ainda que longe de serem conclusivos – que a interação preferencial dessas bases de Tröger com o DNA ocorra através da intercalação de um dos substituintes, como também sugerido pelos estudos realizados por Baldeyrou *et al.*³⁰

Futuramente, para obter informações mais conclusivas acerca desta hipótese, seriam necessários não somente cálculos de energia livre para os diferentes modos de interação

amostrados durante as simulações, como também o desenvolvimento de uma metodologia capaz de mapear as barreiras energéticas necessárias para a abertura do *gap* de intercalação.

6. BIBLIOGRAFIA

- (1) Waring, M.J.; Wakelin, L.P.G. Echinomycin – bifunctional intercalating antibiotic. *Nature*, v. 252, p. 653-657, 1974.
- (2) Stubbe, J. Mechanisms of bleomycin-induced DNA degradation. *Chemical Reviews*, v. 87, p. 1107-1136, 1987.
- (3) Arcamone, F.M.; Animati, F.; Barbieri, B.; Configliacchi, E.; D'Alessio, R.; Geroni, C.; Giuliane, F.C.; Lazzari, E.; Menozzi, M.; Nicola, M.; Penco, S.; Verini, M.A. Synthesis, DNA-binding properties, and antitumor activity of novel distamycin derivatives. *Journal of Medicinal Chemistry*, v. 32, p. 774-778, 1989.
- (4) Minotti, G.; Menna, P.; Salvatorelli, E.; Cairo, G.; Gianni, L. Anthracyclines: molecular advances and pharmacologic developments in antitumour activity and cardiotoxicity. *Pharmacological Reviews*, v. 56, p. 185-229, 2002.
- (5) Hannon, M.J. Supramolecular DNA recognition. *Chemical Society Review*, v. 36, p. 280-295, 2007.
- (6) Hurley, L.H. DNA and its associated processes as targets for cancer therapy. *Nature Reviews Cancer*, v. 2, p. 188-199, 2002.
- (7) Nelson, S.M.; Ferguson, L.R.; Denny, W.A. Non-covalent ligand/DNA interactions: minor groove binding agents. *Mutation Research*, v. 623, p. 24-40, 2007.
- (8) Damia, G.; Silvestri, S.; Carrassa, L.; Filiberti, L.; Faircloth, G.T.; Liberi, G.; Foiani, M.; D'Incalci, M. Unique pattern of ET-743 activity in different cellular systems with defined deficiencies in DNA-repair pathways. *International Journal of Cancer*, v. 92, p. 583-588, 2001.
- (9) Takebayashi, Y.; Pourquier, P.; Zimonjic, D.B.; Nakayama, K.; Emmert, S.; Ueda, T.; Urasaki, Y.; Kanzaki, A. Antiproliferative activity of Ecteinascidin 743 is dependent upon transcription-coupled nucleotide-excision repair. *Nature Medicine*, v. 7, p. 961-966, 2001.
- (10) Tse, W.C.; Boger, D.L. Sequence-selective DNA recognition: natural products and Nature's lessons. *Chemistry and Biology*, v. 11, p. 1607-1617, 2004.

- (11) Waring, M.J.; Bailly, C. DNA recognition by intercalators and hybrid molecules. *Journal of Molecular Recognition*, v. 109, p. 109-122, 1994.
- (12) Dervan, P.B. Molecular recognition of DNA by small molecules. *Bioorganic and Medicinal Chemistry*, v. 9, p. 2215-2235, 2001.
- (13) Chen, H.; Liu, X.; Patel, D.J. DNA bending and unwinding associated with actinomycin D antibiotics bound to partially overlapping sites on DNA. *Journal of Molecular Biology*, v. 258, p. 457-479, 1996.
- (14) Cuesta-Seijo, J.A.; Sheldrick, G.M. Structures of complexes between echinomycin and duplex DNA. *Acta Crystallographica Section D-Biological Crystallography*, v. 61, p. 442-448, 2005.
- (15) Canals, A.; Purciolas, M.; Aymami, J.; Coll, M. The anticancer agent ellipticine unwinds DNA by intercalative binding in an orientation parallel to base pairs. *Acta Crystallographica Section D-Biological Crystallography*, v. 61, p. 1009-1012, 2005.
- (16) Kopka, M.L.; Yoon, C.; Goodsell, D.; Pjura, P.; Dickerson, R.E. The molecular origin of DNA-drug specificity in netropsin and distamycin. *Proceedings of the Academy of Sciences USA*, v. 82, p.1376-1380, 1985.
- (17) Neidle, S. Crystallographic insights into DNA minor groove recognition by drugs. *Biopolymers*, v. 44, p. 105-121, 1997.
- (18) Bailly, C.; Chaires, B. Sequence-specific DNA minor groove binders. Design and synthesis of netropsin and distamycin analogues. *Bioconjugate Chemistry*, v. 9, p. 513-538, 1998.
- (19) Pindur, U.; Jansen, M.; Lemster, T. Advances in DNA-ligands with groove binding, intercalating and/or alkylating activity: chemistry, DNA-binding and biology. *Current Medicinal Chemistry*, v. 12, p. 2805-2847, 2005.
- (20) Coll, M.; Frederick, C.A.; Wang, A.D.-J.; Rich, A. A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG)₂ and its complex with distamycin. *Proceedings of the National Academy of Sciences USA*, v. 84, p. 8385-8389, 1987.

- (21) Kopka, M.L.; Goodsell, D.S.; Baikalov, I.; Grzeskowiak, K.; Cascio, D.; Dickerson, R.E. Crystal structure of a covalent DNA-drug adduct: anthramycin bound to C-C-A-A-C-G-T-T-C and a molecular explanation of specificity. *Biochemistry*, v. 33, p. 13593-13610, 1994.
- (22) Lown, J.W.; Joshua, A.V.; Lee, J.S. Molecular mechanisms of binding and single-strand scission of deoxyribonucleic acid by the antitumor antibiotics saframycins A and C. *Biochemistry*, v. 21, p. 419-428, 1982.
- (23) Giri, I.; Jenkins, M.D.; Schnetz-Boutaud, N.C.; Stone, M.P. Structural refinement of the 8,9-dihydro-8-(N7-guanyl)-9-hydroxy-aflatoxin B₁ adduct in a 5'-Cp^{AFB}G-3' sequence. *Chemical Research in Toxicology*, v. 15, p. 638-647, 2002.
- (24) Tröger, J. Ueber einige mittelst nascirenden formaldehydes entstehende basen. *Journal für Praktische Chemie*, v. 36, p. 225-245, 1887.
- (25) Spielman, M.A. The structure of Troeger's base. *Journal of American Chemical Society*, v. 57, p. 583-585, 1935.
- (26) Valik, M.; Strongin, R.M.; Král, V. Tröger's base derivatives – new life for old compounds. *Supramolecular Chemistry*, v. 17, p. 347-367, 2005.
- (27) Yashima, E.; Akashi, M.; Miyauchi, N. Chiral bis(1,10-phenanthroline) with Tröger's base skeleton. Synthesis and interaction with DNA. *Chemistry Letters*, p. 1017-1020, 1991.
- (28) Tatibouët, A.; Demeunynck, M.; Andraud, C.; Collet, A.; Lhomme, J. Synthesis and study of an acridine substituted Tröger's base : preferential binding of the (-)-isomer to B-DNA. *Chemical Communications*, p. 161-162, 1999.
- (29) Bailly, C.; Laine, W.; Demeunynck, M.; Lhomme, J. Enantiospecific recognition of DNA sequences by a proflavine Tröger base. *Biochemical and Biophysical Research Communications*, v. 273, p. 681-685, 2000.
- (30) Baldeyrou, B.; Tardy, C.; Bailly, C.; Colson, P.; Houssier, C.; Charmantray, F.; Demeunynck, M. Synthesis and DNA interaction of a mixed proflavine-phenanthroline Tröger base. *European Journal of Medicinal Chemistry*, v. 37, p. 325-322, 2002.

- (31) Valik, M.; Malina, J.; Palivec, L.; Foltýnová, J.; Tkadlecová, M.; Urbanová, M.; Brabec, V.; Král, V. Tröger's base scaffold in racemic and chiral fashion as a spacer for bisdystamycin formation. Synthesis and DNA binding study. *Tetrahedron*, v. 62, p. 8591-8600, 2006.
- (32) Claessens, N.; Pierard, F.; Bresson, C.; Moucheron, C.; Mesmaeker, A.K.-D. Optically active Ru(II) complexes with a chiral Tröger's base ligand and their interactions with DNA. *Journal of Inorganic Biochemistry*, v. 101, p. 987-996, 2007.
- (33) Abella, C.A.M.; Rodembusch, F.S.; Stefani, V. Synthesis of novel Tröger's bases analogues. The first ones fluorescent by excited state intramolecular proton transfer (ESIPT). *Tetrahedron Letters*, v. 45, p. 5601-5604, 2004.
- (34) Kitchen, D.B.; Decornez, H.; Furr, J.R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews Drug Discovery*, v. 3, p. 935-949, 2004.
- (35) Moitessier, N.; Englebienne, P.; Lee, D., Lawandi, J.; Corbeil, C.R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British Journal of Pharmacology*, v. 153, p. S7-S26, 2008.
- (36) Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using Lamarckian Genetic Algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, v. 19, p. 1639-1662, 1998.
- (37) Huey, R.; Morris, G.M.; Olson, A.J.; Goodsell, D.S. A semiempirical free energy force field with charge-based desolvation. *Journal of Computational Chemistry*, v. 28, p. 1145-1152, 2007.
- (38) Grootenhuis, P.D.J.; Kollman P.A.; Seibel, G.L.; DesJarlais, R.L.; Kuntz, I.D. Computerized selection of potential DNA binding compounds. *Anti-Cancer Drug Design*, v. 5, p. 237-242, 1990.
- (39) Grootenhuis, P.D.J.; Roe, D.C., Kollman, P.A.; Kuntz, I.D. Finding potential DNA-binding compounds by using molecular shape. *Journal of Computer-Aided Molecular Design*, v. 8, p. 731-750, 1994.

- (40) Chen, K.; Adelstein, S.J.; Kassis, A.I. Molecular modeling of the interaction of the iodinated Hoechst analogs with DNA: implications for new radio-pharmaceutical design. *Journal of Molecular Structure: THEOCHEM*, v. 711, p. 49-56, 2004.
- (41) Rohs, R.; Bloch, I., Sklenar, H.; Shakked, Z. Molecular flexibility in *ab initio* drug docking to DNA: binding-site and binding-mode transitions in all-atom Monte Carlo simulations. *Nucleic Acids Research*, v. 33, p. 7048-7057, 2005.
- (42) Tuttle, T.; Kraka, E.; Cremer, D. J. Docking, triggering and biological activity of dynemicin A in DNA: a computational study. *Journal of American Chemical Society*, v. 127, p. 9469-9484, 2005.
- (43) Evans, D.A.; Neidle, S. Virtual screening of DNA minor groove binders. *Journal of Medicinal Chemistry*, v. 39, p. 4232-4238, 2006.
- (44) Sobhani, A.M.; Amini, S.R.; Tyndall, J.D.A.; Azizi, E.; Daneshtalab, M.; Khalaj, A. A theory of mode of action of azolylalkylquinolines as DNA binding agents using automated flexible ligand docking. *Journal of Molecular Graphics and Modelling*, v. 25, p. 459-469, 2006.
- (45) Lauria, A.; Patella, C.; Ippolito, M.; Almerico, A.M. Docking and synthesis of pyrrolopyrimidodiazepinone derivatives (PPDs) and their precursors: new scaffolds for DNA-interacting agents. *J. Mol. Struct-THEOCHEM* 2007, 819, 26-31.
- (46) Navarrete, J.T.; Casado, J.; Ramírez, F.J. Electronic spectroscopy study and molecular docking simulation of the interaction of terthiophene with DNA. *Journal of Molecular Structure: THEOCHEM*, v. 834, p. 176-181, 2007.
- (47) Yan, Z.; Sikri, S.; Beveridge, D.L.; Baranger, A.M. Identification of an aminoacridine derivative that binds to RNA tetraloops. *Journal of Medicinal Chemistry*, v. 50, p. 4096-4104, 2007.
- (48) Holt, P.A.; Chaires, J.B.; Trent, J.O. Molecular docking of intercalators and groove-binders to nucleic acids using AutoDock and Surflex. *Journal of Chemical Information and Modeling*, v. 48, p. 1602-1615, 2008.
- (49) Dickerson, R.E.; Drew, H.R.; Conner, B.N.; Wing, R.M.; Fratini, A.V.; Kopka, M.L. The anatomy of A-, B-, and Z-DNA. *Science*, v. 216, p. 475-485, 1982.

- (50) Dickerson, R.E. DNA structure from A to Z. *Methods in Enzymology*, v. 211, p. 67-111, 1992.
- (51) Rich, A.; Zhang, S. Z-DNA: the long road to biological function. *Nature Reviews*, v. 4, p. 566-572, 2003.
- (52) Ricci, C.G. A, B e Z-DNA: caracterização conformacional e importância biológica. Porto Alegre, 2007. Trabalho de conclusão de curso (curso de Farmácia) – Universidade Federal do Rio Grande do Sul.
- (53) Dickerson, R.E.; Chiu, T.K. Helix bending as a factor in protein/DNA recognition. *Biopolymers*, v. 44, p. 361-403, 1997.
- (54) Travers, A.A. The structural basis of DNA flexibility. *Philosophical Transactions of the Royal Society A*, v. 362, p. 1423-1438, 2004.
- (55) de Souza, O.N.; Ornstein, R.L. Inherent DNA curvature and flexibility correlate with TATA box functionality. *Biopolymers*, v. 46, p. 403-415, 1998.
- (56) Lu, X.-J.; Olson, W.-K. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Research*, v. 31, p. 5108-5121, 2003.
- (57) Solis, F.J.; Wets, J.-B. Minimization by random search techniques. *Mathematics of Operations Research*, v. 6, p. 19-30, 1981.
- (58) Coll, M.; Aymami, J.; van der Marel, G.A.; van Boom, J.H.; Rich, A.; Wang, A.H. Molecular structure of the netropsin-d(CGCGATATCGCG)₂ complex: DNA conformation in an alternating AT segment. *Biochemistry*, v. 28, p. 310-320, 1989.
- (59) Lindahl E.; Hess, B.; van der Spoel, D. GROMACS 3.0 : a package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, v. 7, p. 306-317, 2001.
- (60) Gasteiger, J.; Marsili, M. Iterative partial equalization of orbital electronegativity – a rapid access to atomic charges. *Tetrahedron*, v. 36, p. 3219-3228, 1980.
- (61) Sanner, M.F. Python: a programming language for software integration and development. *Journal of Molecular Graphics and Modeling*, v. 17, p. 57-61, 1999.

(62) Ricci, C.G.; Netz, P.A. Docking studies on DNA-ligand interactions: building and application of a protocol to identify the binding mode. *Journal of Chemical Information and Modeling*, v. 49, p. 1925-1935, 2009.

(63) Guex, N.; Peitsch, M.C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis*, v. 18, p. 2714-2723, 1997.

(64) Oostenbrink, C.; Soares, T.A.; van der Vegt, N.F.A.; van Gunsteren, W.F. Validation of the 53A6 GROMOS force field. *European Biophysical Journal*, v. 34, p. 273-284, 2005.

(65) GaussView, Version 4.1, Dennington II, Roy; Keith, Todd; and Millam John; Semichem, Inc., Shawnee Mission, KS, 2007.

(66) Gaussian 98 (Revision A.1x), Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, Jr., J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. Gaussian, Inc., Pittsburgh PA, 2001.

(67) van Gunsteren, W.F; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D.P.; Glättli, A.; Hünenberger, P.H.; Kastenholtz, M.A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N.F.A.; Yu, H.B. Biomolecular modeling: goals, problems, perspectives. *Angewandte Chemie International Edition*, v. 45, p. 4064-4092, 2006.

(68) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins base don condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry*, v. 24, p. 1999-2012, 2003.

- (69) Sorin, E.J.; Pande, V.S. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophysical Journal*, v. 88, p. 2472-2493, 2005.
- (70) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an N-log(N) method for Ewald sums in large systems. *Journal of Chemical Physics*, v. 98, p. 10089-10092, 1993.
- (71) Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a generalized amber force field. *J. Comput. Chem.* 2004, 25, 1157-1174.
- (72) Batista, P.R.; Wilter, A.; Durham, E.H.A.B.; Pascutti, P.G. Molecular dynamics simulations applied to the study of subtypes of HIV-1 protease. *Cell Biochemistry and Biophysics*, v. 44, p. 395-404, 2006.
- (73) Wang, J.; Wang, W.; Kollman, P.A.; Case, D.A. Automatic atom type and bond type perception in molecular mechanics calculations. *Journal of Molecular Graphics and Modelling*, v. 25, p. 247-260, 2006.
- (74) Baily, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *Journal of Physical Chemistry*, v. 97, p. 10269-10280, 1993.
- (75) Watson, J.D.; Crick, F.H.C. Molecular structure of nucleic acids. *Nature*, v. 171, p. 737-738, 1953.
- (76) Várnai, P.; Djuranovic, D.; Lavery, R.; Hartmann, B. α/γ transitions in the B-DNA backbone. *Nucleic Acids Research*, v. 30, p. 5398-5406, 2002.
- (77) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham III, T.E.; Loughton, C.A.; Orozco, M. Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. *Biophysical Journal*, v. 92, p. 3817-3829, 2007.
- (78) Schneider, B.; Neidle, S.; Berman, H.M. Conformations of the sugar-phosphate backbone in helical DNA structures. *Biopolymers*, v. 42, p. 113-124, 1997.
- (79) Bertrand, H.-O.; Ha-Duong, T.; Fermandjian, S.; Hartmann, B. Flexibility of the B-DNA backbone: effects of local and neighbouring sequences on pyrimidine-purine steps. *Nucleic Acids Research*, v. 26, p. 1261-1267, 1998.

- (80) Djuranovic, D.; Hartmann, B. DNA fine structure and dynamics in crystal and in solution: the impact of BI/BII backbone conformations. *Biopolymers*, v. 73, p. 356-368, 2004.
- (81) Gao, Y.-G.; Robinson, H.; Wang, A. H.-J. High resolution A-DNA crystal structures of d(AGGGGCCCT). An A-DNA model of poly(dG)*poly(dC). *European Journal of Biochemistry*, v. 261, p. 413-420, 1999.
- (82) Dickerson *et al.* Definitions and nomenclature of nucleic structure components. *Nucleic Acids Research*, v. 17, p. 1797-1803, 1989.
- (83) Calladine, C.R.; Drew, H.R. A base centred explanation of the B-to-A transition in DNA. *Journal of Molecular Biology*, v. 178, p. 773-782, 1984.
- (84) Dickerson, R.E.; Ng, H.-L. DNA structure from A to B. *Proceedings of the National Academy of Sciences USA*, v. 98, p. 6986-6988, 2001.
- (85) Reddy, S.Y.; Leclerc, F.; Karplus, M. DNA polymorphism: a comparison of force fields for nucleic acids. *Biophysical Journal*, v. 84, p. 1421-1449, 2003.
- (86) Olson, W.K.; Zhurkin, V.B. Modeling DNA deformations. *Current Opinion in Structural Biology*, v. 10, p. 286-297, 2000.
- (87) Young, M.A.; Ravishanker, G.; Beveridge, D.L.; Berman, H.M. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophysical Journal*, v. 68, p. 2454-2468, 1995.
- (88) Goodsell, D.S.; Kaczor-Grzeskowiak, M.; Dickerson, R.E. The crystal structure of C-C-A-T-T-A-A-T-G-C: implications for bending of B-DNA at T-A steps. *Journal of Molecular Biology*, v. 239, p. 79-96, 1994.
- (89) El Hassan, M.A.; Calladine, C.R. Conformational characteristics of DNA: empirical classifications and a hypothesis for the conformational behaviour of dinucleotides steps. *Philosophical Transactions of the Royal Society A*, v. 355, p. 43-100, 1997.
- (90) Nikolov, D.B.; Chen, H.; Halay, E.D.; Hoffmann, A.; Roeder, R.G.; Burley, S.K. Crystal structure of a human TATA box-binding protein/TATA element complex. *Proceedings of the National Academy of Sciences*, v. 93, p. 4862-4867, 1996.

Docking Studies on DNA-Ligand Interactions: Building and Application of a Protocol To Identify the Binding Mode

Clarisse G. Ricci and Paulo A. Netz*

Instituto de Química, Universidade Federal do Rio Grande do Sul, av. Bento Gonçalves 9500, 91501-970, Porto Alegre, Brazil

Received April 30, 2009

Despite DNA being an important target for several drugs, most of the docking programs are validated only for proteins and their ligands. In this paper, we used AutoDock 4.0 to perform self-dockings and cross dockings between two DNA ligands (a minor groove binder and an intercalator) and four distinct receptors: 1) crystallographic DNA without intercalation gap; 2) crystallographic DNA with intercalation gap; 3) canonical B-DNA; and 4) modified B-DNA with intercalation gap. Besides being efficient in self-dockings, AutoDock is capable of correctly identifying two of the main DNA binding modes with the condition that the target possesses an artificial intercalation gap. Therefore, we suggest a default protocol to identify DNA binding modes which uses a modified canonical DNA (with gap) as receptor. This protocol was applied to dock two different Träger bases to DNA and the predicted binding modes agree with those suggested, yet not established, by experimental data. We also applied the protocol to dock aflatoxin B₁ *exo*-8,9-epoxide, and the results are in complete agreement with experimental data from the literature. We propose that this approach can be used to investigate other ligands whose binding mode to DNA remains unknown, yielding a suitable starting point for further theoretical studies such as molecular dynamics simulations.

INTRODUCTION

As the number of biological structures in data banks rapidly increases, molecular docking is becoming an important approach to evaluate or even to elucidate the interaction between potential ligands and their macromolecular targets.^{1,2} It has been shown that several docking methods described so far can correctly reproduce the binding modes of cocrystallized ligands to their protein targets (self-dockings), but none of them can be considered a universally applicable method.²

While there is a large number of studies reporting protein–ligand docking, much less research has been reported on docking of ligands to nucleic acids^{3–13} despite DNA being an important molecular target for a wide number of antibiotics and antitumor drugs.¹⁴ Disappointingly, most of the scoring functions have been parametrized exclusively with protein–ligands sets, and the programs have been validated only for proteins and their ligands.^{2,15} As well-known, nucleic acids differ from proteins due to unique structural features such as high density charge and helix chiral geometry. Also, nucleic acids do not present a single and well-defined binding site (as occur with most of the proteins) and offer more solvent exposed binding pockets.² As a consequence, this leads to the question of whether docking programs validated for proteins can also produce reasonable results in ligand-DNA docking. This issue has been recently approached by Holt et al.,¹³ who has shown that AutoDock and Surfex can accurately reproduce the crystal structure of several ligands (minor groove binders and intercalators) bound to DNA, within a resolution of approximately 2 Å.

Certainly, this had shed new light on the potential of automated docking programs for virtual screening of DNA binding agents. However, although self-dockings (i.e., using the original crystallographic target) are considered useful as a first indication of docking accuracy, they have proved to provide little information about accuracy in real drug discovery.² Indeed, the employment of docking techniques to elucidate unknown DNA binding mechanisms - *without* any conclusive previous experimental data - remains a challenge.

The known fact that DNA is not rigid but rather a very flexible molecule that can assume several structurally distinct isoforms^{16–18} combined with the fact that most of current automated docking programs do not take into account target flexibility leads to a so far insufficiently explored issue: when it comes to the docking of ligands whose binding mode to DNA remains unknown, which oligomer conformation should be used as target? It is likely that several structural features of the chosen oligomer conformation will strongly impact docking performance, especially when ligand and target interact through an induced-fit mechanism.

In general, interaction of small molecules with DNA occurs in two distinct ways: intercalation between base pairs or groove recognition.¹⁹ Although the major groove offers more H-bonding donor and acceptor sites, it has been suggested that it provides a much larger and shallow binding pocket for small molecule binding than the minor groove. Indeed, most of the small ligands bind to the minor groove, while the major groove is the preferential binding site for proteins and peptides.^{19,20}

Among the main DNA binding modes, intercalation is the most common way through which small and rigid aromatic molecules recognize DNA. Most of the classical and simple

* Corresponding author fax: (+55)(51)33087304; e-mail: netz@iq.ufrgs.br.

Apêndice B – Conformações de melhor energia resultantes das docagens com as bases de Tröger.

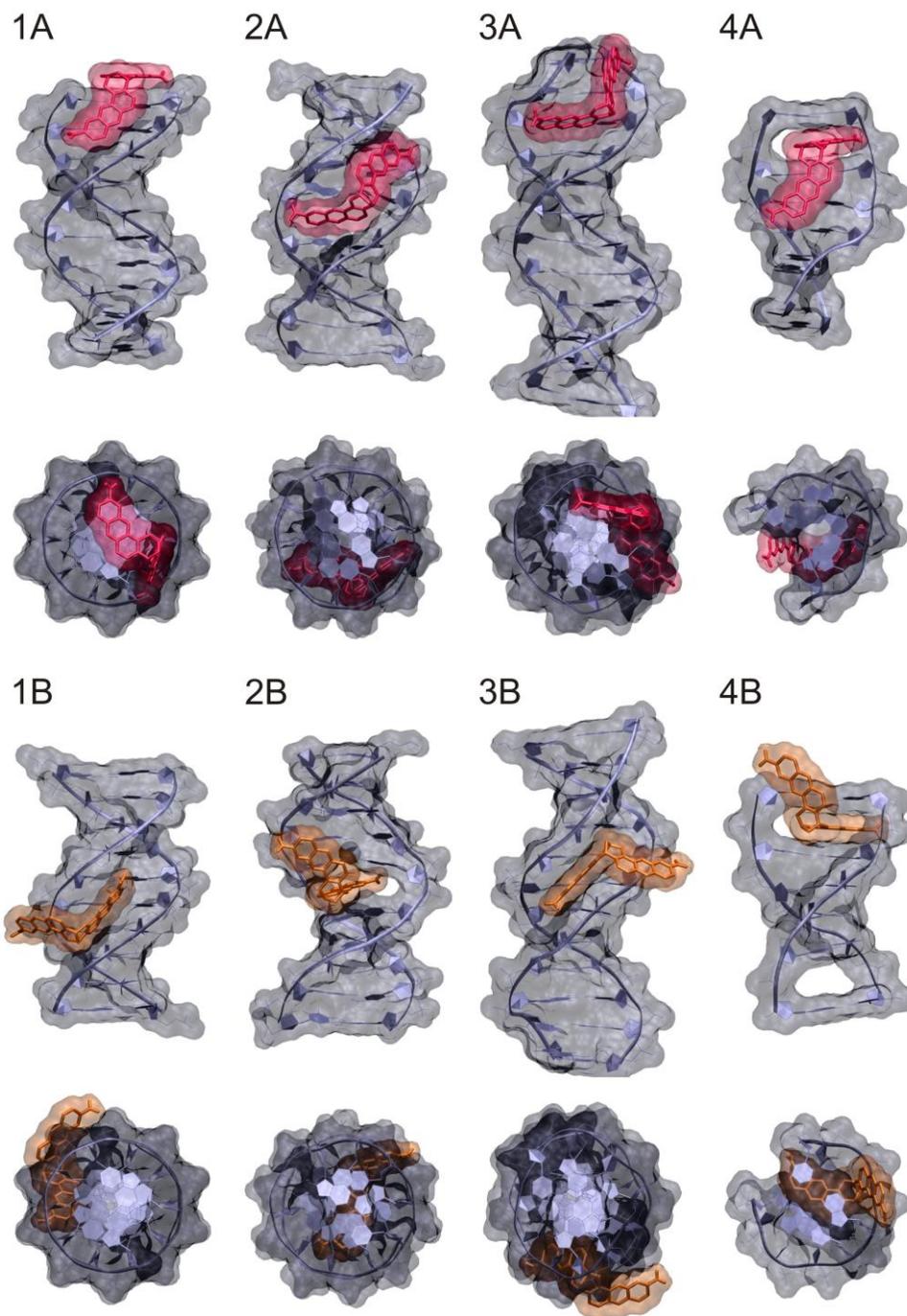


Figura B1. Complexos com melhor energia livre de interação para os dockings com a Tröger simétrica. O enantiômero (-)-(R,R) está representado em rosa e o enantiômero (+)-(S,S) está representado em laranja.

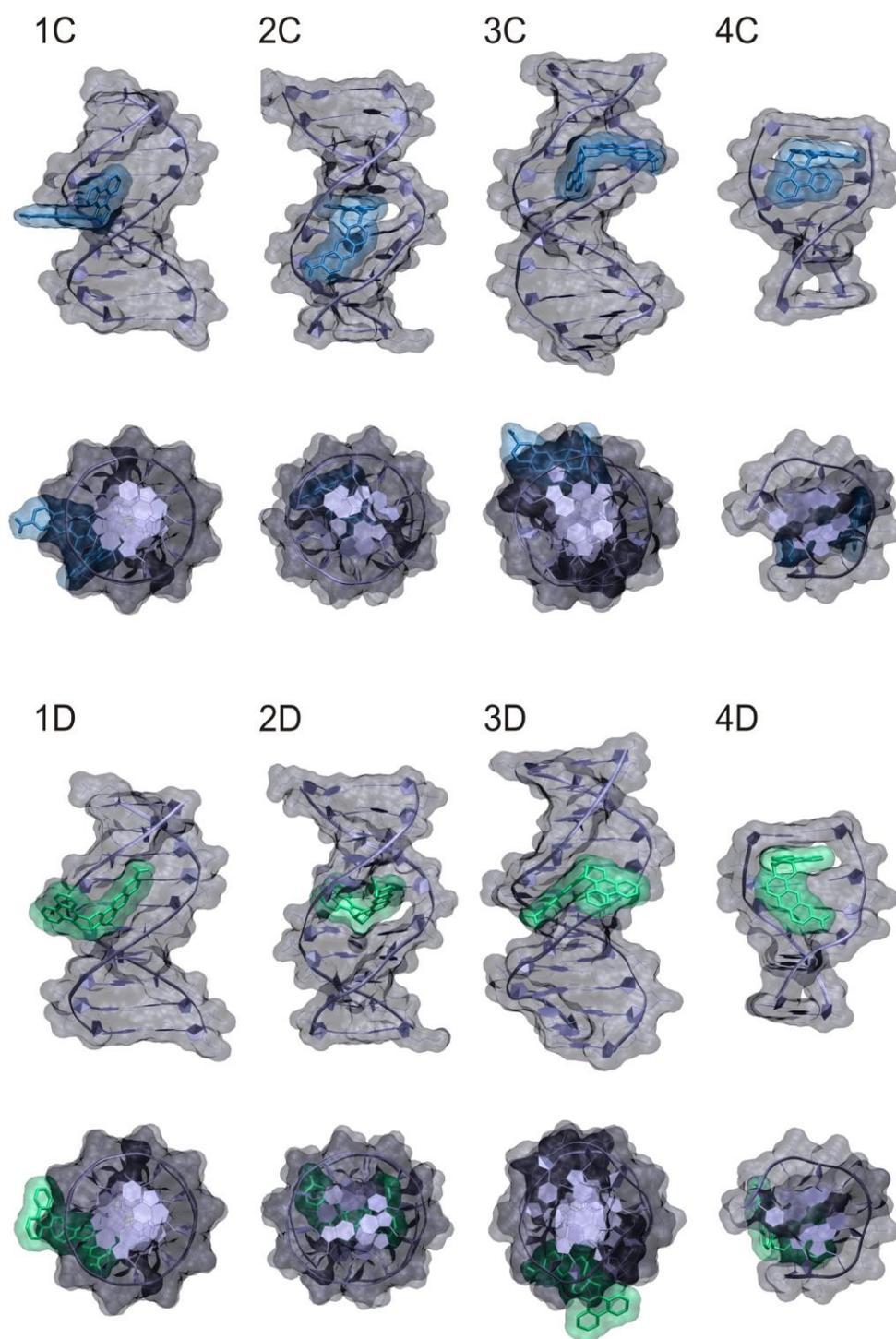


Figura B2. Complexos com melhor energia livre de interação para os dockings com a Tröger assimétrica. O enantiômero (-)-(R,R) está representado em azul e o enantiômero (+)-(S,S) está representado em verde.

Apêndice C – README.txt do ACPYPI

= ACPYPI =
\$Id\$

A tool based in Python to use Antechamber to generate topologies for chemical compounds and to interface with others python applications like CCPN or ARIA.

acypyi is pronounced as "ace + pipe"

Topologies files to be generated so far: CNS/XPLOR, GROMACS, CHARMM and AMBER.

NB: Topologies generated by acypyi/Antechamber are based on General Amber Force Field (GAFF) and should be used only with compatible forcefields like AMBER and its variant.

Several flavours of AMBER FF are ported already for GROMACS (see ffamber: <http://chemistry.csulb.edu/ffamber/>) as well as to XPLOR/CNS (see xplor-nih: <http://amber.scripps.edu/xplor-nih.html>) and CHARMM.

This code is released under GNU General Public License V3.

<<< NO WARRANTY AT ALL!!! >>>

It was inspired by:

- amb2gmx.pl (Eric Sorin, David Mobley and John Chodera) and depends on Antechamber and Openbabel
- YASARA Autosmiles:
<http://www.yasara.org/autosmiles.htm> (Elmar Krieger)
- topolbuild (Bruce Ray)
- xplo2d (G.J. Kleywegt)

For Antechamber, please cite:

1. Wang, J., Wang, W., Kollman P. A.; Case, D. A. "Automatic atom type and bond type perception in molecular mechanical calculations". Journal of Molecular Graphics and Modelling , 25, 2006, 247260.
2. Wang, J., Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. "Development and testing of a general AMBER force field". Journal of Computational Chemistry, 25, 2004, 1157-1174.

If you use this code, I am glad if you cite:

BATISTA, P. R.; WILTER, A.; DURHAM, E. H. A. B.; PASCUTTI, P. G. Molecular Dynamics Simulations Applied to the Study of Subtypes of HIV-1 Protease. Cell Biochemistry and Biophysics, 44, 395-404, 2006.

Alan Wilter S. da Silva, D.Sc. - CCPN Research Associate
Department of Biochemistry, University of Cambridge.
80 Tennis Court Road, Cambridge CB2 1GA, UK.

Apêndice D – Início do arquivo de topologia gerado pelo ACPYPI para a base de Tröger simétrica.

; 10_cluster_1_GMX.itp created by acpypi on Tue Jun 2 22:55:15 2009

```
[ atomtypes ]
;name      bond_type      mass      charge      ptype      sigma      epsilon      Amb
c3         c3              0.00000   0.00000     A          3.39967e-01  4.57730e-01 ; 1.91  0.1094
ca         ca              0.00000   0.00000     A          3.39967e-01  3.59824e-01 ; 1.91  0.0860
nb         nb              0.00000   0.00000     A          3.25000e-01  7.11280e-01 ; 1.82  0.1700
nh         nh              0.00000   0.00000     A          3.25000e-01  7.11280e-01 ; 1.82  0.1700
hn         hn              0.00000   0.00000     A          1.06908e-01  6.56888e-02 ; 0.60  0.0157
h1         h1              0.00000   0.00000     A          2.47135e-01  6.56888e-02 ; 1.39  0.0157
ha         ha              0.00000   0.00000     A          2.59964e-01  6.27600e-02 ; 1.46  0.0150
h4         h4              0.00000   0.00000     A          2.51055e-01  6.27600e-02 ; 1.41  0.0150
h2         h2              0.00000   0.00000     A          2.29317e-01  6.56888e-02 ; 1.29  0.0157
```

```
[ moleculetype ]
;name      nrexcl
10_cluster_1  3
```

```
[ atoms ]
;  nr  type  resi  res  atom  cgnr      charge      mass      ; qtot  bond_type
  1   c3    1   NON   C1    1       0.04110     12.01000 ; qtot  0.041
  2   ca    1   NON   C4    2       0.00080     12.01000 ; qtot  0.042
  3   ca    1   NON   C5    3       0.00890     12.01000 ; qtot  0.051
  4   ca    1   NON   C6    4      -0.05070     12.01000 ; qtot  0.000
  5   ca    1   NON   C8    5      -0.04300     12.01000 ; qtot -0.043
  6   ca    1   NON  C10    6       0.02840     12.01000 ; qtot -0.014
  7   nb    1   NON  N12    7      -0.25280     14.01000 ; qtot -0.267
  8   ca    1   NON  C13    8       0.09720     12.01000 ; qtot -0.170
  9   ca    1   NON  C14    9       0.09860     12.01000 ; qtot -0.071
 10  nb    1   NON  N15   10      -0.25280     14.01000 ; qtot -0.324
 11  ca    1   NON  C16   11       0.02840     12.01000 ; qtot -0.296
 12  ca    1   NON  C18   12      -0.04290     12.01000 ; qtot -0.339
 13  ca    1   NON  C20   13      -0.04930     12.01000 ; qtot -0.388
 14  ca    1   NON  C22   14       0.02390     12.01000 ; qtot -0.364
 15  ca    1   NON  C23   15       0.04590     12.01000 ; qtot -0.318
 16  nh    1   NON  N24   16      -0.30870     14.01000 ; qtot -0.627
 17  c3    1   NON  C25   17       0.04250     12.01000 ; qtot -0.585
 18  ca    1   NON  C28   18       0.01680     12.01000 ; qtot -0.568
 19  ca    1   NON  C29   19       0.07810     12.01000 ; qtot -0.490
 20  nb    1   NON  N30   20      -0.24620     14.01000 ; qtot -0.736
 21  ca    1   NON  C31   21       0.07340     12.01000 ; qtot -0.662
 22  ca    1   NON  C32   22      -0.01500     12.01000 ; qtot -0.677
 23  ca    1   NON  C34   23       0.02660     12.01000 ; qtot -0.651
 24  ca    1   NON  C38   24      -0.04040     12.01000 ; qtot -0.691
 25  ca    1   NON  C40   25      -0.05080     12.01000 ; qtot -0.742
 26  ca    1   NON  C42   26       0.00110     12.01000 ; qtot -0.741
 27  ca    1   NON  C43   27      -0.04340     12.01000 ; qtot -0.784
 28  ca    1   NON  C45   28       0.00140     12.01000 ; qtot -0.783
 29  ca    1   NON  C46   29      -0.05080     12.01000 ; qtot -0.834
 30  ca    1   NON  C48   30      -0.03960     12.01000 ; qtot -0.873
 31  ca    1   NON  C50   31       0.03800     12.01000 ; qtot -0.835
 32  nh    1   NON  N51   32      -0.30930     14.01000 ; qtot -1.145
 33  c3    1   NON  C5H   33       0.07930     12.01000 ; qtot -1.065
 34  nh    1   NON  N35   34      -0.35830     14.01000 ; qtot -1.424
 35  hn    1   NON  H36   35       0.14230       1.00800 ; qtot -1.281
 36  hn    1   NON  H37   36       0.14230       1.00800 ; qtot -1.139
 37  h1    1   NON   H    37       0.05140       1.00800 ; qtot -1.088
 38  h1    1   NON  H1    38       0.05140       1.00800 ; qtot -1.036
```

39	ha	1	NON	H2	39	0.06250	1.00800 ; qtot -0.974
40	ha	1	NON	H4	40	0.06330	1.00800 ; qtot -0.910
41	h4	1	NON	H5	41	0.08300	1.00800 ; qtot -0.827
42	h4	1	NON	H6	42	0.08300	1.00800 ; qtot -0.744
43	ha	1	NON	H7	43	0.06330	1.00800 ; qtot -0.681
44	ha	1	NON	H8	44	0.06250	1.00800 ; qtot -0.619
45	h1	1	NON	H9	45	0.05150	1.00800 ; qtot -0.567
46	h1	1	NON	H10	46	0.05150	1.00800 ; qtot -0.516
47	ha	1	NON	H11	47	0.06570	1.00800 ; qtot -0.450
48	ha	1	NON	H12	48	0.06360	1.00800 ; qtot -0.386
49	ha	1	NON	H13	49	0.06250	1.00800 ; qtot -0.324
50	ha	1	NON	H14	50	0.06310	1.00800 ; qtot -0.261
51	ha	1	NON	H15	51	0.06250	1.00800 ; qtot -0.198
52	ha	1	NON	H16	52	0.06360	1.00800 ; qtot -0.135
53	h2	1	NON	H17	53	0.06730	1.00800 ; qtot -0.067
54	h2	1	NON	H18	54	0.06730	1.00800 ; qtot -0.000