

**UNIVERSIDADE ESTADUAL PAULISTA
MESTRADO EM CIÊNCIA E TECNOLOGIA DE MATERIAIS
CAMPUS DE BAURU**

**“DOCKING E ANÁLISE DO MODO DE LIGAÇÃO DE
TRÊS MOLÉCULAS PEQUENAS,
UM BENZIMIDAZOL E DOIS COMPOSTOS DE CRÔMIO,
NOS SULCOS DO DNA 5'-CGCGAATTCGCG-3' ”**

ESTHER CAMILO DOS REIS

**Bauru
2008**

ESTHER CAMILO DOS REIS

**“DOCKING E ANÁLISE DO MODO DE LIGAÇÃO DE
TRÊS MOLÉCULAS PEQUENAS,
UM BENZIMIDAZOL E DOIS COMPOSTOS DE CRÔMIO,
NOS SULCOS DO DNA 5'-CGCGAATTCGCG-3' ”**

Dissertação apresentada como requisito à obtenção do Título de Mestre em Ciência e Tecnologia dos Materiais do Programa de Pós Graduação em Ciência e Tecnologia de Materiais da Universidade Estadual Paulista “Julio de Mesquita Filho”.

Orientadora: Profa. Dra. Ignez Caracelli

Bauru
2008

**DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO
UNESP - BAURU**

Reis, Esther Camilo dos.

Docking e análise do modo de ligação de três moléculas pequenas, um benzimidazol e dois compostos de crômio, nos sulcos do DNA 5'-CGCGAATTCGCG-3' / Esther Camilo dos Reis, 2008.

xii, 84 f. il.

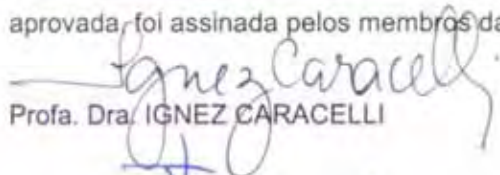
Orientador: Ignez Caracelli.

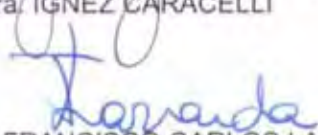
Dissertação (Mestrado) - Universidade Estadual Paulista. Faculdade de Ciências, Bauru, 2008.

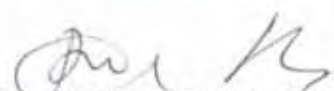
1. Docking. 2. DNA. 3. Programa GOLD. 4. Crômio. I. Universidade Estadual Paulista. Faculdade de Ciências. II. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE ESTHER CAMILO DOS REIS, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA E TECNOLOGIA DE MATERIAIS, DO(A) FACULDADE DE CIÊNCIAS DE BAURU.

Aos 10 dias do mês de abril do ano de 2008, às 14:00 horas, no(a) Anfiteatro da Pós-graduação/FC, reuniu-se a Comissão Examinadora da Defesa Pública, composta pelos seguintes membros: Profa. Dra. IGNEZ CARACELLI do(a) Departamento de Física / Faculdade de Ciências de Bauru, Prof. Dr. FRANCISCO CARLOS LAVARDA do(a) Departamento de Física / Faculdade de Ciências de Bauru, Profa. Dra. REGINA HELENA DE ALMEIDA SANTOS do(a) Instituto de Química de São Carlos / Universidade de São Paulo, sob a presidência do primeiro, a fim de proceder a arguição pública da DISSERTAÇÃO DE MESTRADO de ESTHER CAMILO DOS REIS, intitulada "DOCKING E ANÁLISE DO MODO DE LIGAÇÃO DE TRÊS MOLÉCULAS PEQUENAS, UM BENZIMIDAZOL E DOIS COMPOSTOS DE CROMIO, NOS SULCOS DO DNA 5'-CGCGAATTGCG-3'". Após a exposição, a discente foi arguida oralmente pelos membros da Comissão Examinadora, tendo recebido o conceito final: APROVADA . Nada mais havendo, foi lavrada a presente ata, que, após lida e aprovada, foi assinada pelos membros da Comissão Examinadora.


Profa. Dra. IGNEZ CARACELLI


Prof. Dr. FRANCISCO CARLOS LAVARDA


Profa. Dra. REGINA HELENA DE ALMEIDA SANTOS

AGRADECIMENTOS

À minha orientadora, Profa. Dra. Ignez Caracelli, por ter me colocado em contato com um ramo da ciência tão rico e ainda inexplorado. Por seus conselhos de vida, agradeço.

Ao Prof. Dr. Mauricio Angel Vega Teijido, do Grupo de Pesquisa BioMat, FC-UNESP/Bauru pela paciência e valiosa colaboração na parte de cálculos e sugestões na análise de resultados.

Ao aluno de iniciação científica Sergio Pizano Rodrigues, do Grupo de Pesquisa BioMat, FC-UNESP/Bauru pela paciência e colaboração com a visualização dos resultados.

À aluna de doutorado Cristiane Cabral de Melo, do Laboratório de Cristalografia, Estereodinâmica e Modelagem Molecular – DQ – UFSCar, São Carlos, pela ajuda nos cálculos.

Ao Prof. Dr. Julio Zukerman Schpector do Laboratório de Cristalografia, Estereodinâmica e Modelagem Molecular – DQ – UFSCar, São Carlos, pela busca de estruturas de pequenas moléculas utilizando sua licença CSD.

À Secretaria Estadual de Educação pela Bolsa de Estudos da Secretaria da Educação do Estado de São Paulo

Aos professores do programa de pós-graduação pelos ensinamentos.

Aos membros da banca por suas sugestões

Aos amigos por torcerem pelo meu sucesso.

A toda a minha família pelo apoio, carinho e paciência sempre.

REIS, Esther Camilo dos. **Docking e análise do modo de ligação de três moléculas pequenas, um benzimidazol e dois compostos de crômio, nos sulcos do DNA 5'-CGCGAATTCGCG-3'**. 2008. Dissertação (Mestrado em Ciências e Tecnologia dos Materiais) – UNESP, Faculdade de Ciências, Bauru, 2008.

RESUMO

Neste trabalho foi estudado o modo de ligação de três moléculas pequenas ao DNA 5'-CGCGAATTCGCG-3', através da simulação computacional e análise em tela gráfica. Verificou-se que é possível utilizar o programa GOLD, baseado em algoritmos genéticos, em estudos envolvendo o DNA, uma vez que até o momento há apenas dois trabalhos publicados na base de dados Scopus (Portal de Periódicos da CAPES) onde esta metodologia é utilizada. Também foi verificado que 1vzk, complexo cristalográfico do DNA com 2-(5-{4-[amino (imino)metil]fenil}-2-tienil)-1h-benzimidazol-6-carboximida (DB818), obtido do PDB é uma estrutura representativa para estudos no sulco menor. Foram investigados os modos de ligação dos seguintes ligantes: (a) DB818, obtido a partir da estrutura cristalográfica 1vzk; (b) DAZJOE; [Cr(1,2-bis(naftilidenoamino)etano)(H₂O)₂], estrutura obtida do Cambridge Structural Database e (c) CR3NC, [Cr(2,3-bis{[(2-hidroxi-4-dietilamino) (fenil) (metileno)]amino}2-butenodinitrila) (H₂O)₂], cuja estrutura foi obtida por modelagem molecular, tomando como modelo DAZJOE. Os estudos de docking do DAZJOE em DNA sugerem que seu modo de ligação seria no sulco menor, porém sua interação não se dá com as bases AATTC como no caso do DB818, mas sim entre as bases ATTC. Quando os cálculos de docking foram realizados com a molécula CR3NC, que tem volume maior e mais flexibilidade nas extremidades, observou-se que a mesma posiciona-se no sulco maior. Este resultado está em concordância com dados experimentais disponíveis. Um mecanismo de quebra do DNA é sugerido, não devido à influência do crômio, mas sim devido à presença dos nitrogênios. Observou-se que a influência do metal é menor do que a da molécula complexada ao íon nas interações.

Palavras-chave: docking, DNA, programa GOLD, crômio.

REIS, Esther Camilo dos. **Docking e análise do modo de ligação de três moléculas pequenas, um benzimidazol e dois compostos de crômio, nos sulcos do DNA 5'-CGCGAATTCGCG-3'**. 2008. Dissertação (Mestrado em Ciências e Tecnologia dos Materiais) – UNESP, Faculdade de Ciências, Bauru, 2008.

ABSTRACT

In this work we studied the binding mode of some small molecules in the DNA 5'-CGCGAATTCGCG-3' through computer simulation and analysis in graphical screen. We found that it is possible to use the GOLD program, based on genetic algorithms in studies involving DNA, despite we found only two papers in the database Scopus (CAPES Portal of Journals) where this approach is used. It was also found that 1vzk, crystallographic complex of the DNA with 2-(5-(4-[amino (imino) methyl] phenil) -2- thienyl)-1h-benzimidazol-6-carboximide (DB818), obtained from PDB is a representative structure for studies in the minor groove. We investigated the binding mode of the following ligands: a) DB818, obtained from the 1vzk crystallographic structure, (b) DAZJOE, [Cr(1,2-bis(naphthylideneamino) ethane)(H₂O)₂], structure obtained from Cambridge Structural Database and (c) CR3NC, [Cr((2,3-bis[(2-hydroxy-4-phenyl) (diethylamino) (methylene)] 2-amino butenedinitrile)(H₂O)₂], whose structure was obtained by molecular modeling, taking DAZJOE as model. The DAZJOE docking's studies in DNA suggest that its binding mode would be in the minor groove, but their interaction is not given with the bases AATTC as for the DB818, but between the bases ATTC. When the docking calculations were performed with CR3NC molecule, which has larger volume and more flexibility in the extremities, we observed that it take place in major groove. This result is in agreement with experimental data available. It is suggested a DNA break mechanism, but not due to the influence of chromium, but the nitrogens. It was observed that the influence of the metal is lower than the ionophore in interactions

Keywords: docking, DNA, GOLD program, chromium.

LISTA DE FIGURAS

	página
Figura 1.1 Níveis de organização desde o organismo até os pares de base do DNA. Nos seres humanos há 23 pares de cromossomos. As histonas são proteínas globulares onde o DNA se enovela.	5
Figura 1. 2 Cristais de DNA	8
Figura 1. 3 Esquema ilustrando o modo de obtenção do padrão de difração por mono-cristal de DNA.	8
Figura 1. 4 Padrão de difração para o B-DNA, obtido por Franklin, apresentando um “X”.	9
Figura 1. 5 Padrão de difração usual para moléculas protéicas.	9
Figura 1. 6 Padrão de difração onde se observam os “diamantes”.	10
Figura 1. 7 As camadas, na foto de difração, com destaque para a camada 4 (direita).	10
Figura 1. 8 Representação esquemática da estrutura tridimensional do DNA.	11
Figura 1. 9 Estrutura cristalográfica do B-DNA 1vzk e sua representação esquemática (direita).	12
Figura 1. 10 (a) Formas A, B e Z do DNA considerando o eixo de enrolamento e embaixo uma visão perpendicular ao eixo. (b) Diferença entre os eixos de enrolamento das fitas.	13
Figura 1. 11 As quatro bases nitrogenadas do DNA: Timina e Adenina, Guanina e Citosina. As linhas pontilhadas indicam ligações de hidrogênio entre as bases C e G e as bases A e T.	14
Figura 1. 12 Nomenclatura padrão dos átomos em formato pdb.	15
Figura 1. 13 O gráfico da energia em função da conformação mostra que a mínima energia também resulta no melhor resultado receptor-ligante em um caso de docking rígido.	19
Figura 1. 14 Fluxograma de um GA básico (MIRANDA, 2000).	20
Figura 1. 15 Cruzamento do cromossomo 1 com o cromossomo 2. Um ponto de cruzamento representado por uma barra é escolhido aleatoriamente. Copia-se tudo o que vem antes de um dos pais e tudo o que vem depois do outro.	22
Figura 1. 16 Exemplo de mutação	22
Figura 1. 17 (a) Benzimidazol ligado ao DNA no sulco menor (código PDB:1vzk). (b) Antraciclina ligada ao DNA por intercalação (código PDB: 1nab). Observa-se uma deformidade das bases. (c) A enzima recombinase atuando no sulco maior do DNA (código PDB: 1jkq).	25
Figura 2. 1 Fluxograma de trabalho	27
Figura 2.2 Tela de Entrada do algoritmo de busca do NDB, com os parâmetros adotados.	28
Figura 2.3 Ligante 1 – DB818 - com a numeração dos átomos utilizada neste trabalho.	31
Figura 2.4 Ligante 2 - DAZJOE - com a numeração dos átomos utilizada neste trabalho	32
Figura 2.5 Ligante 3 - CR3NC - com a numeração dos átomos utilizada neste trabalho.	33
Figura 2.6 Docking	34

LISTA DE FIGURAS *(continuação)*

	página
Figura 2.7 Redocking.	35
Figura 2.8 Cross-docking	36
Figura 2.9 A função score dentro do programa GOLD.	38
Figura 3. 1 O dodecâmero d(C-G-C-G-A-A-T-T-C-G-C-G) ₂ das estruturas 1vzk e 2b3e. Os centros dos cálculos são mostrados	48
Figura 3. 2 Estruturas tridimensionais 1vzk e 2b3e, com seus respectivos ligantes DB818 e DB819, dentro do sulco menor.	48
Figura 3. 3 Estruturas tridimensionais 1vzk e 2b3e, com seus respectivos ligantes DB818 e DB819, dentro do sulco menor.	49
Figura 3. 5 O ligante DB819 dentro do sulco menor do receptor 2b3e, mostrando a ligação de hidrogênio.	50
Figura 3. 6 Os ligantes DB818 e DB819.	50
Figura 3. 7 Duas orientações representam os dois grupos obtidos no redocking em torno do centro O2/T19.	54
Figura 3. 8 Superposição da estrutura cristalográfica do complexo DNA-DB818 (stick) com a orientação selecionada no redocking (ball-stick). As interações análogas entre essas estruturas são apresentadas.	55
Figura 3. 9: Sobreposição da solução obtida em cálculos com a versão anterior do GOLD e a solução obtida utilizando-se a última versão. Um RMSD de apenas 0,16 Å foi encontrado, evidenciando a compatibilidade das duas versões.	57
Figura 3. 10 Na orientação 6 o anel aromático é perpendicular ao restante da molécula o que dificulta sua entrada no sulco.	58
Figura 3. 11 Observa-se que o resultado do cálculo de redocking(em ball-stick) está em concordância com a estrutura cristalográfica (em stick). No entanto observa-se uma aproximação do N32 do O4 do açúcar.	61
Figura 3. 12 Superposição da estrutura cristalográfica do ligante DB818 (em preto, ao fundo) com as 6 soluções selecionadas nos 6 cálculos de redocking (em branco)	62
Figura 3. 13 Orientação 4 no sítio da 2b3e resultante do cross-docking no centro T8 em 2b3e. O ligante não fica totalmente plano, mas acomoda-se no sulco menor entre as bases AATTC.	66
Figura 3. 14 Orientação 7 no sítio da 2b3e resultante do cross-docking no centro T20 em 2b3e. Melhor resultado comparando-se com a simulação em T8.	67
Figura 3. 15 Interações entre os nucleotídeos T8, T19 e T20 e do receptor 1vzk e o ligante DAZJOE. Estas interações estão descritas na Tabela 3.14.	71
Figura 3. 16 Mecanismo de quebra do DNA pelo composto Cr3NC sugerido por Vaidyanathan <i>et al.</i> (2000)	72
Figura 3. 17 Orientações representativas dos grupos 1 e 2 das soluções do docking de cr3nc em 1vzk	73
Figura 3. 18 Posição das três orientações do ligante CR3NC em 2b3e. A interação ocorre no sulco maior, porém há contato maior com uma das faces.	74
Figura 3. 19 Medida da abertura do sulco maior das estruturas 1vzk e 2b3e.	75

LISTA DE TABELAS

		página
Tabela 1. 1	As bases nucleotídicas do DNA	6
Tabela 1. 2	Formas do DNA e suas características geométricas (pb = par de base)	13
Tabela 1. 3	Resumo das interações consideradas na função escore do Programa GOLD	24
Tabela 3. 1	Levantamento de complexos de DNA CGCGAATTCGCG e pequenos ligantes	43
Tabela 3. 2	Descrição dos ligantes dos complexos DNA-moléculas pequenas	45
Tabela 3. 3	Resultados do redocking para o cálculo com o centro T19, raio 10 Å.	54
Tabela 3. 4	Resultados do redocking para o cálculo com centro T19, raio 25 Å.	56
Tabela 3. 5	Resultados do redocking para o cálculo com o centro C15, raio 25 Å.	57
Tabela 3. 6	Resultados do redocking para o cálculo com o centro T8, raio 25 Å.	59
Tabela 3. 7	Resultados do redocking para o cálculo com o centro C21, raio 25 Å.	60
Tabela 3. 8	Resultados do redocking para o cálculo com o centro C21 raio 25 Å.	61
Tabela 3. 9	Solução selecionada do cálculos de redocking e valor de Fitness.	62
Tabela 3. 10	Resultados do cross- docking para o cálculo centrado em T8.	65
Tabela 3. 11	Resultados do cross-docking para o cálculo centrado em T20.	67
Tabela 3. 12	Resultados de docking da DAZJOE em 1vzk centrado em T8.	69
Tabela 3. 13	Resultados de docking da DAZJOE em 1vzk centrado em T19.	70
Tabela 3. 14	Interações entre a DAZJOE e 1vzk	70
Tabela 3. 15	Resultados de docking de CR3CN em 1vzk.	73
Tabela 3. 16	Interações entre a orientação 6 e o DNA	74
Tabela 3. 17	Resultados de docking de CR3NC em 2b3e.	75

LISTA DE ABREVIATURAS E SÍMBOLOS

α	polarizabilidade
ϵ_0	permissividade no vácuo
A	adenina
AMP	adenosina monofosfato
C	citossina
CSD	<i>Cambridge Structural Database</i>
DNA	ácido desoxirribonucleico
dTMP	deoxitimidina monofosfato
EOR	espécies de oxigênio reativo
G	guanina
GA	<i>Genetics Algorithms</i> (Algoritmo Genético)
GMP	guanosina monofosfato
GOLD	<i>Genetic Optimisation for Ligand Docking</i>
GTF	<i>Glucose Tolerance Factor</i>
hb	<i>hydrogen bond</i> (ligação de hidrogênio)
k	constante de Boltzmann
LaCrEMM	Laboratório de Cristalografia, Estereodinâmica e Modelagem Molecular
MC	Monte Carlo
NDB	<i>Nucleic Acid Database</i>
NMR	Ressonância Magnética Nuclear
p	dipolo
pb	par de base
PBS	<i>Public Broadcasting Service</i>
PDB	<i>Protein Data Bank</i>
pic	picolinato
q	carga
r	distância entre as cargas
R	raio
RMSD	<i>root mean square deviation</i>
RNA	<i>ribonucleic acid</i>
T	timina
T	temperatura
vdw	van der Waals
int	interno
ext	externo
ΔG	energia livre de Gibbs

SUMÁRIO

	Página
1 INTRODUÇÃO	1
1.1 CIÊNCIA DOS MATERIAIS E MATERIAIS BIOLÓGICOS	1
1.2 DNA – UM POUCO DE HISTÓRIA	2
1.3 COMPOSIÇÃO QUÍMICA	4
1.4 DESVENDANDO A ESTRUTURA TRIDIMENSIONAL DO DNA	5
1.5 A ESTRUTURA TRIDIMENSIONAL DO DNA	7
1.6 BANCOS DE DADOS DE ESTRUTURAS DE MOLÉCULAS BIOLÓGICAS, ORGÂNICAS E ORGANOMETÁLICAS	14
1.7 DOCKING MOLECULAR	17
1.8 ALGORITMO GENÉTICO (GA)	20
1.9 INTERAÇÕES COM DNA	22
1.10 COMPOSTOS DE CRÔMIO	25
2 METODOLOGIA	27
2.1. RECEPTOR	27
2.1.1. Busca nos Bancos de dados de informações estruturais do DNA	27
2.1.2. Levantamento dos complexos de DNA-ligante	28
2.1.3. Seleção e Análise do DNA	29
2.1.4. Seleção de sítios para o cálculo de docking	29
2.1.5. Adição de Hidrogênios	29
2.2. LIGANTES	31
2.2.1. Ligante 1	31
2.2.2. Ligante 2	32
2.2.3. Ligante 3	33
2.2.4. Adição de Hidrogênios	33
2.3. DOCKING MOLECULAR	34
2.3.1. O que é docking?	34
2.3.2. Redocking ou validação	35
2.3.3. Cross-docking	36
2.3.4. O Método Computacional	36
2.4. VISUALIZAÇÃO GRÁFICA	39
2.5. EXPERIMENTO <i>IN SILICO</i> × DADOS EXPERIMENTAIS	41

3	RESULTADOS E DISCUSSÃO	42
3.1.	LEVANTAMENTO NO PDB DAS ESTRUTURAS TRIDIMENSIONAIS DE DNA 5'-CGCGAATTCGCG-3' COMPLEXADOS COM MOLÉCULAS PEQUENAS	42
3.2.	ANÁLISE DAS ESTRUTURAS CRISTALOGRAFICAS DOS COMPLEXOS 1VZK E 2B3E DO PDB	47
3.3.	ESTUDOS DE LIGAÇÃO NO SULCO MENOR DO DNA.	51
3.3.1.	Resultados do redocking de DB818 no sulco menor de 1vzk	51
3.3.1.a	Resultados do redocking de DB818 no receptor R-1vzk, centrado em T19, com raio 10 Å	53
3.3.1.b	Resultados do redocking de DB818 no receptor R-1vzk, centrado em T19, com raio 25 Å	55
3.3.1.c	Resultados do redocking de DB818 no receptor R-1vzk, centrado em C15, com raio 25 Å	57
3.3.1.d	Resultados do redocking de DB818 no receptor R-1vzk, centrado em T8, com raio 25 Å	58
3.3.1.e	Resultados do redocking de DB818 no receptor R-1vzk, centrado em C21, com raio 25 Å	59
3.3.1.f	Resultados do modelo-2 de DB818 no receptor R-1vzk, centrado em T19, com raio 25 Å	60
3.3.1.g	Comparacao dos resultados de redocking	62
3.3.2.	Cross-docking	64
3.3.2.a	Cross-docking do modelo M2 no receptor R-2b3e centrado em T8 com raio 25 Å	64
3.3.2.b	Cross-docking do modelo M2 no receptor R-2b3e centrado em T20 com raio 25 Å	64
3.3.3.	Docking do ligante DAZJOE no receptor R-1vzk	68
3.3.3.a	Docking do ligante DAZJOE no receptor R-1vzk, centrado em T8, com raio 25 Å	69
3.3.3.b	Docking do ligante DAZJOE no receptor R-1vzk, centrado em T19, com raio 25 Å	69
3.3.3.b	Comparacao dos resultados de docking centrado em T8 e T19 e análise das interações	69
3.4.	ESTUDOS DE LIGAÇÃO NO SULCO MAIOR DO DNA	71
3.4.1.	Docking do ligante CR3CN no receptor R-1vzk	72
3.4.2.	Docking do ligante CR3CN no receptor R-2b3e	74
4	CONCLUSÕES	77
5	REFERÊNCIAS	79

1 INTRODUÇÃO

1.1 CIÊNCIA DOS MATERIAIS E MATERIAIS BIOLÓGICOS

Com o desenvolvimento das ciências, o conhecimento foi se subdividindo em áreas cada vez mais específicas. Em tal situação, um pesquisador da área da Física tinha pouco ou nenhum conhecimento sobre os avanços da biologia, química e computação e vice-versa. Com o rápido avanço tecnológico e a necessidade moderna do desenvolvimento de novos materiais essas áreas tiveram que trabalhar cada vez mais próximas, o que culminou em uma área interdisciplinar: a ciência e tecnologia dos materiais. Essa área possibilitou o estudo dos materiais não somente sob o ponto de vista macroscópico, mas também a partir do conhecimento da estrutura tridimensional das moléculas.

Dentre os materiais que podem ser estudados, encontram-se os materiais biológicos, como as proteínas, o DNA e o RNA. A compreensão de como os sistemas biológicos funcionam, pode auxiliar tanto no tratamento de doenças, como também no desenvolvimento de outros materiais com aplicações tecnológicas. Com relação ao tratamento de doenças, atualmente, o planejamento racional de fármacos tem à disposição importantes ferramentas computacionais necessárias ao aumento da rapidez e à diminuição dos custos envolvidos no processo de desenvolvimento desses produtos. Já no desenvolvimento de novos materiais com aplicações tecnológicas, os organismos vivos são fontes inesgotáveis de informações e em pleno século XXI muitos mistérios, como o enovelamento de proteínas, os mecanismos de desenvolvimento do câncer ou como são armazenadas as informações no cérebro, continuam sendo desafios da ciência que permanecem sem uma completa explicação. Novas informações que contribuam para a compreensão da lógica da vida podem levar ao desenvolvimento de tecnologias, novos materiais e computadores mais inteligentes.

Percebe-se hoje que a solução para os problemas modernos não pode mais ser encontrada somente por pesquisadores de uma única área do conhecimento. É preciso que pessoas com diferentes formações trabalhem com o objetivo de desvendar e/ou modificar o sistema em estudo.

O sistema biológico em estudo nesta dissertação é um fragmento de DNA, na sua forma B, que é a forma de DNA que aparece em condições fisiológicas. Em geral, compostos que podem se ligar ao B-DNA no sulco menor, de forma não-covalente, podem ter aplicações como agentes contra câncer, vírus, e algumas infecções. É o caso, por exemplo, de pentanamida, que demonstrou atividade contra leishmaniose e contra pneumonia causada pelo *Pneumocystis carinii*, doença oportunista que aparece freqüentemente em pacientes portadores do vírus HIV (EVANS, NEIDLE, 2006).

No caso presente, o sistema em estudo é o fragmento de DNA 5'-CGCGAATTCGCG-3', do qual se pretende analisar o modo de ligação de pequenos ligantes. Inicialmente, foi feito um estudo do complexo de código PDB (Protein Data Bank) 1vzk, no qual o fragmento de B-DNA tem ligada, no sulco menor, uma molécula de 2-(5-{4-[amino(imino)metil]fenil}-2-tienil)-1h-benzimidazol-6-carboximida (DB818) (MALLENA *et al.*, 2004).

O modo de ligação de outros dois compostos de crômio com o DNA também foram também estudados: $[\text{Cr}(1,2\text{-bis}(\text{naftilidenoamino})\text{etano})(\text{H}_2\text{O})_2]$, cujo código no Cambridge Structural Database (CSD) é DAZJOE, e $[\text{Cr}(2,3\text{-bis}\{[(2\text{-hidroxi-4-dietilamino})(\text{fenil})(\text{metileno})]\text{amino}\}2\text{-butenodinitrila})(\text{H}_2\text{O})_2]$ (CR3CN), modelado a partir de DAZJOE.

Para estudar os modos de ligação destes compostos ao DNA, foram feitos cálculos de docking, utilizando métodos de algoritmos genéticos, e a análise dos resultados foi feita através de visualização gráfica. Isto permitiu o estudo de cada molécula e comparações entre moléculas e sítios possíveis de ligação.

1.2 DNA – UM POUCO DE HISTÓRIA

Em 1869, Friedrich Miescher, durante a guerra da Criméia, enquanto trabalhava com feridos e colocava bandagens nas feridas, isolou células de pus. Miescher adicionou uma solução alcalina fraca ao pus que rompeu as membranas celulares e permitiu a precipitação dos núcleos. A partir deles, isolou-se uma única substância química, a qual Miescher chamou de nucleína. Posteriormente, a nucleína foi encontrada em todas as células testadas e era rica em fósforo.

Miescher pensou que a função principal das nucleínas era armazenar átomos de fósforo na célula. Entretanto, seu papel genético e sua estrutura permaneceram desconhecidos por muito tempo (ACOT, 2003). Naquela época, não havia como demonstrar qualquer outra função biológica para a substância. Mesmo sem saber, Miescher tinha obtido o primeiro extrato de DNA bruto.

Depois da descoberta de Miescher, outros cientistas desenvolveram técnicas de purificação mais refinadas e a substância “pura” ficou conhecida como ácido desoxirribonucleico, por causa do açúcar presente em sua estrutura, a desoxirribose, e suas propriedades de natureza ácida. No início do ano de 1900, sabia-se que as nucleínas de Miescher eram uma mistura de proteínas e ácidos nucleicos, havendo dois tipos: DNA (sigla do inglês “deoxyribonucleic acid” - ácido desoxirribonucleico), encontrado principalmente no núcleo; e RNA (“ribonucleic acid” – ácido ribonucleico), encontrado principalmente no citoplasma (Acot, 2003).

Cientistas, como Phoebus Aaron Theodor Levene (1869-1940), começaram a estudar o DNA decompondo-o em seus componentes e descobriram que era constituído de nucleotídeos (estrutura composta por grupos fosfato, o açúcar desoxirribose e bases nitrogenadas) de quatro tipos diferentes (Adenina - **A**, Timina - **T**, Citosina - **C** e Guanina - **G**). Esses resultados não chegaram a provocar interesse, pois o DNA foi considerado uma molécula muito simples para ser um bom candidato à molécula responsável pelas características hereditárias. Em 1909, por exemplo, o químico lituano Phoebus Levene assegurou que os nucleotídeos do DNA ocorriam em quantidades idênticas e cunhou a chamada “hipótese do tetranucleotídeos” que afirmava que essas subunidades mantinham-se unidas em uma sequência repetida e invariável ao longo da molécula. Dessa forma, se Levine estivesse certo, o DNA não teria potencial para ser o repositório das informações genéticas. Essa visão modificou-se a partir da década de 40 após alguns importantes experimentos. Inicialmente, o médico de origem canadense Oswald Avery (1877-1955) trabalhando com cepas virulentas e não virulentas de *Pneumococcus* mostrou que o DNA era um fator chave para as diferenças herdadas entre esses organismos. Posteriormente, no final de 1940, Erwin Chargaff (1905-2002) demonstrou a incorreção da “hipótese do tetranucleotídeos”. Chargaff mostrou que as frequências de nucleotídeos diferem entre espécies e,

mesmo quando esses valores são mantidos analiticamente constantes, poderia ocorrer um número enorme de variações nessas seqüências e mesmo entre regiões diferentes de uma mesma molécula de DNA. Deste modo, os ácidos nucléicos poderiam apresentar um grau elevado de complexidade em suas seqüências, apesar de serem compostos por apenas 4 tipos de bases diferentes. Foi o começo do estudo e entendimento sobre a composição química do DNA (SIMONI, HILL, VAUGHAN, 2002; SEMENZA, 2003).

Essa complexidade dos ácidos nucléicos e sua relação com o armazenamento e transmissão da informação genética tornou-se clara após a elucidação da estrutura tridimensional do DNA em 1953, feita por difração de raio X por Franklin, Wilkins, Watson e Crick (Maddox, 2002).

1.3 COMPOSIÇÃO QUÍMICA

Todos os organismos vivos, desde os vírus aos seres humanos, têm as informações genéticas armazenadas nos ácidos nucléicos (DNA ou RNA como no caso de alguns vírus), em particular, as células dos eucariontes têm o seu DNA organizado na forma de cromossomos que se encontram no núcleo celular (Figura 1.1).

Um cromossomo pode ser conceitualmente dividido em genes – blocos funcionais do DNA, cada um dos quais codifica uma proteína em particular. Cada gene está posicionado em um locus (posição) particular no cromossomo que é conservada de geração em geração. Muito superficialmente, pode-se pensar no gene como o codificador de uma característica, como por exemplo, a cor dos olhos. O conjunto dos genes que resultam em determinada característica são denominadas alelos.

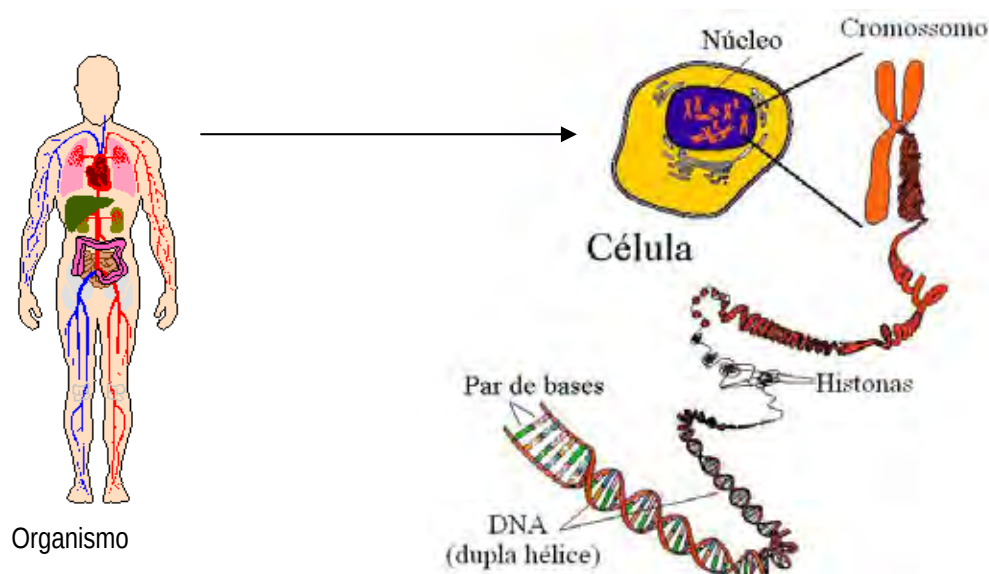
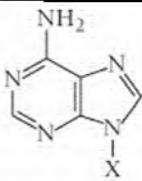
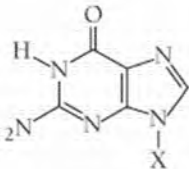
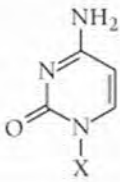
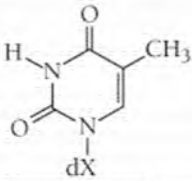


Figura 1.1 Níveis de organização desde o organismo até os pares de base do DNA. Nos seres humanos há 23 pares de cromossomos. As histonas são proteínas globulares onde o DNA se enovela.

Os genes são formados por uma sequência de pares de nucleotídeos. Por sua vez, essas unidades moleculares são constituídas por um açúcar (desoxirribose) ligado por um lado a um grupo fosfato e por outro a uma das quatro bases nitrogenadas (Tabela 1.1). A sequência de pares de bases ao longo da molécula de DNA é a responsável pela informação genética. O código genético é justamente a leitura destas sequências, a qual especifica a sequência linear dos aminoácidos das proteínas. Embora os nucleotídeos que constituem o DNA sejam muito pequenos, o DNA de todos os cromossomos de uma única célula é superior a um metro em humanos, pois cada molécula contém aproximadamente 3×10^9 milhões de nucleotídeos. (VOET, VOET, PRATT, 2000).

Esses nucleotídeos se agrupam por ligações dos grupos fosfatos às posições 3' e 5' da ribose e estas ligações assimétricas significam que a fita de DNA tem uma direção. Em uma dupla hélice a direção dos nucleotídeos em uma fita é oposta à direção dos nucleotídeos na outra fita. Este arranjo das fitas do DNA é chamado antiparalelo e as terminações assimétricas das fitas de DNA são chamadas terminais 5' e 3'.

Tabela 1. 1 As bases nucleotídicas do DNA

Base	Base X = H	Nucleosídeo X = ribose	Nucleotídeo X = ribose fosfato
	Adenina A	Adenosina A	Ácido Adenílico AMP
	Guanina G	Guanosina G	Ácido Guanílico GMP
	Citosina C	Citidina C	Ácido Citidílico CMP
	Timina T	Timidina dT	Ácido Deoxitimidílico dTMP

1.4 DESVENDANDO A ESTRUTURA TRIDIMENSIONAL DO DNA

Após muitos anos de investigações, em 1953, James Watson e Francis Crick publicaram o célebre artigo apresentando o primeiro modelo tridimensional da molécula de DNA que foi amplamente aceito pela comunidade científica. Este modelo era compatível com todas as características químicas conhecidas da molécula e levava em consideração vários resultados experimentais obtidos previamente por outros pesquisadores. A publicação desse artigo só foi possível graças à contribuição de cientistas como Linus Pauling, que usou seu conhecimento de química e uma nova técnica poderosa chamada Cristalografia de raio X, para descobrir uma estrutura com formato de saca-rolhas encontrada em muitas proteínas – a alfa-hélice. Watson e Crick seguiram a abordagem química de Pauling e os modelos de difração de raio X para desenvolver o modelo da estrutura do DNA. A difração de raio X pode fornecer muitas informações sobre o formato e estrutura de uma molécula (ACOT, 2003).

A cientista Rosalind Franklin teve um papel fundamental para essa descoberta, integrante da equipe de John Randall, no King's College de Londres, onde também trabalhava Wilkins. Na Universidade de Cambridge, onde estudava física e química, Rosalind Franklin teve contato com Sir William Lawrence Bragg, que usava a difração por raio X para revelar a estrutura de cristais. Em 1951 foi incumbida de analisar a estrutura do DNA, um campo novo para ela, e sua primeira tarefa foi montar um equipamento de difração de raio X, com o qual obteve imagens de DNA com qualidade cada vez melhor, auxiliada pelo estudante de doutorado Raymond Gosling. Porém, as desavenças entre Franklin e Wilkins contribuíram para a ruptura dos dois, que chegaram a trabalhar sobre o mesmo tema, mas isolados. Após dois anos, Franklin decidiu ir embora. Muitas especulações são feitas sobre até onde Franklin chegou na sua interpretação das imagens que obteve do DNA. Porém, o que se sabe é que a cientista obteve imagens de excelente qualidade em 1952, em especial a de número 51, que Gosling deu a Wilkins, e este mostrou a Watson sem que Franklin soubesse. Seguiu-se daí, uma série de manobras que minimizaram a contribuição de Franklin. Para ter crédito sobre a descoberta, Wilkins escreveu um texto sobre DNA para a revista "Nature" sem mencionar o trabalho de Franklin. Em 1958, Rosalind morreu de câncer de ovário. Em 1962, o norte-americano James Watson, o britânico Francis Crick e o neozelandês Maurice Wilkins ganham o prêmio Nobel e nesse contexto Franklin não passa de uma mera colaboradora (MADDOX, 2002).

Progressos continuam sendo feitos no estudo do DNA e o Projeto Genoma Humano foi um grande avanço em sua pesquisa. Iniciado em 1990 e finalizado em 2003, esse projeto seqüenciou as 3164,7 milhões de bases nucleotídicas (A, C, T e G) que constituem o DNA humano (HUMAN GENOME PROJECT). Ainda há muito a fazer no estudo do DNA e o papel das ferramentas computacionais é crucial, visto o grande número de dados experimentais disponíveis para análise.

1.5 A ESTRUTURA TRIDIMENSIONAL DO DNA

A estrutura tridimensional do DNA pode ser desvendada graças aos experimentos realizados utilizando-se a difração por raio X.

Para realizar os estudos utilizando esta metodologia, o primeiro passo é a obtenção de um cristal da macromolécula em questão. A Figura 1.2 mostra algumas fotos de cristais de DNA (CAVALIERE, 2004).

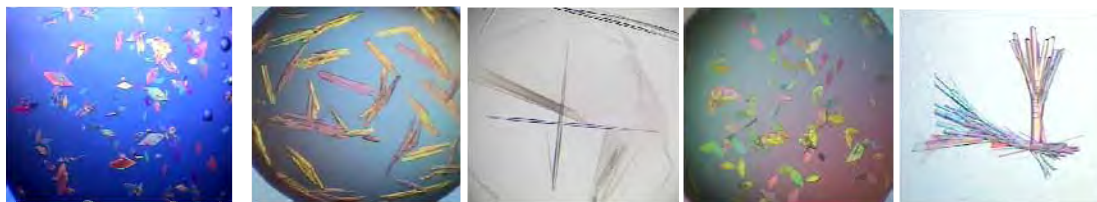


Figura 1. 2 Cristais de DNA

Um cristal, quando submetido a um feixe de raio X gera um padrão de difração representado de forma simplificada na Figura 1.3.

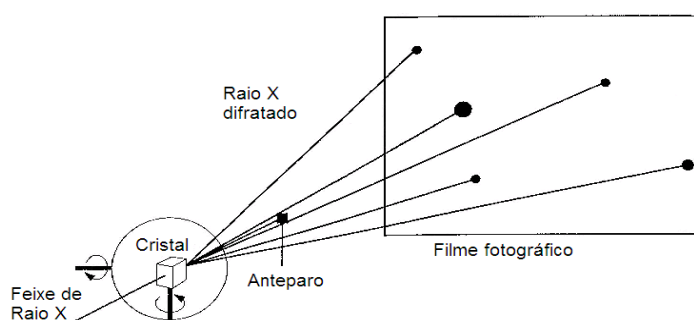


Figura 1. 3 Esquema ilustrando o modo de obtenção do padrão de difração por mono-cristal de DNA.

Rosalind Franklin, na determinação da estrutura do DNA, conseguiu obter uma fibra de DNA na forma B, cujo padrão de difração é apresentado na Figura 1.4 (PBS FOUNDATION).

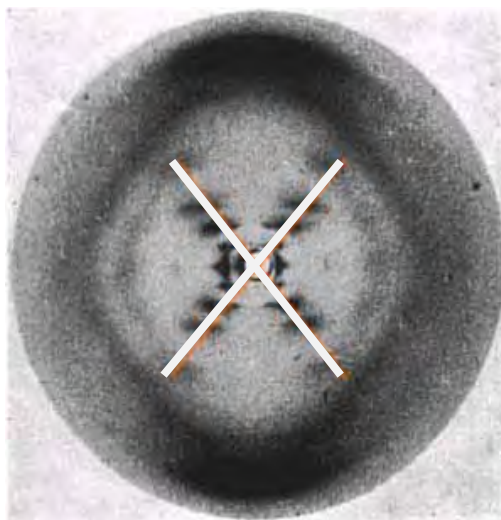


Figura 1. 4 Padrão de difração para o B-DNA, obtido por Rosalind Franklin, apresentando um “X”.

O passo importante do trabalho foi tentar entender porque o padrão de difração apresentava um “X” característico, ao invés do padrão usual para outras moléculas biológicas como o apresentado na Figura 1.5.

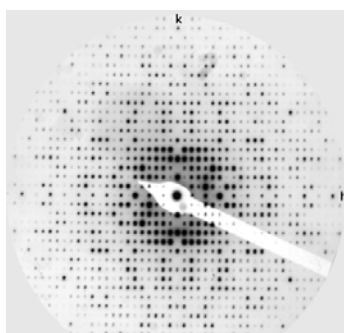


Figura 1. 5 Padrão de difração usual para moléculas protéicas.

A interpretação das fotos de DNA, feita por Watson, Crick e Wilkins, mostra que o “X” estava relacionado com a estrutura química do DNA. O padrão de “X” foi relacionado com a dupla hélice do DNA.

No difratograma do DNA outro aspecto observado, além do “X”, foram os diamantes (Figura 1.6) (PBS Foundation). Os diamantes foram relacionados com a existência e organização dos açúcares e fosfatos do DNA, posicionados no lado externo das hélices.



Figura 1. 6 Padrão de difração onde se observam os “diamantes”.

Outro aspecto observado foi em relação às manchas que apareciam nos “X”, pois o X, não se apresenta como uma linha contínua (Figura 1.7) (PBS Foundation). As descontinuidades foram relacionadas com camadas, e as camadas com as bases que formam os degraus da hélice. Além disso, foi observado também que uma das camadas, a chamada camada 4, ocorre devido ao fato que as hélices correm em direções contrárias, mas se enrolam em relação a um eixo comum ocorrendo desse modo, a formação de dois sulcos (groove, em inglês), chamados de sulco maior e sulco menor.

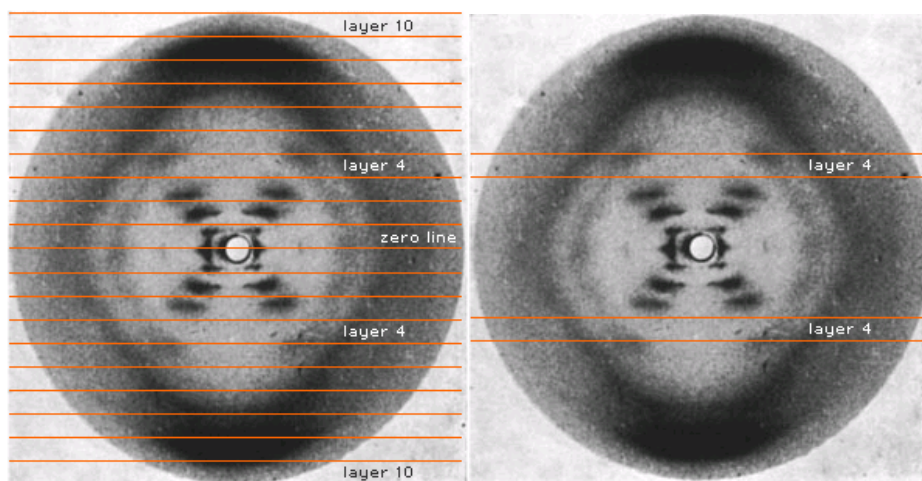


Figura 1. 7 As camadas, na foto de difração, com destaque para a camada 4 (direita).

Na Figura 1.8, observa-se a estrutura do DNA, que lembra uma escada com seus degraus e os bastões que representam os pares de bases (adenina, citosina, guanina e timina). Também aparece o corrimão de cada hélice, mostrando que tem

um eixo comum e que correm em direções opostas, como indicado pelas setas. Observa-se a formação do sulco maior e do sulco menor.

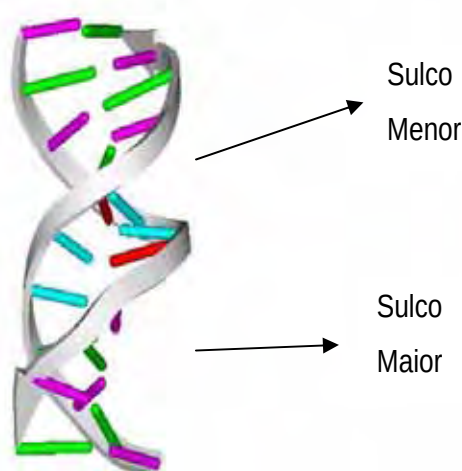


Figura 1. 8 Representação esquemática da estrutura tridimensional do DNA.

Muitas características do DNA são importantes para seu funcionamento, entretanto, sua estrutura tridimensional é a característica chave. Essa estrutura foi publicada por Francis Crick e James Watson em 1953 baseando-se em experimentos cristalográficos realizados por Rosalind Franklin. Até então nenhum cristalógrafo havia conseguido separar as formas A e B do DNA, que Franklin percebeu estarem relacionadas com a hidratação, conforme será visto mais adiante. Após obter a imagem de difração de raio X do DNA, Franklin trabalhando com o mapa de densidade eletrônica determinou que os fosfatos ficavam nas extremidades e as bases na parte interna do DNA. (ELKIN, 2003).

Pode-se perguntar por que o DNA se mantém em formato de hélice e porque duas hélices se mantêm unidas. Uma análise da polaridade dos componentes do DNA revela que tanto os açúcares quanto os fosfatos são muito solúveis em água. Porém, as bases são insolúveis. Considerando que todo espaço nas células que não são ocupados por componentes importantes, como DNA, RNA e enzimas, são preenchidos por água, observa-se que o formato helicoidal do DNA está estritamente relacionado à sua interação com moléculas de água. Na estrutura tridimensional do DNA as bases ficam para dentro, evitando interação com a água e os fosfatos e a ribose ficam para fora, fato que pode ser observado através de resultados da difração de raio X do DNA (Figura 1.9).

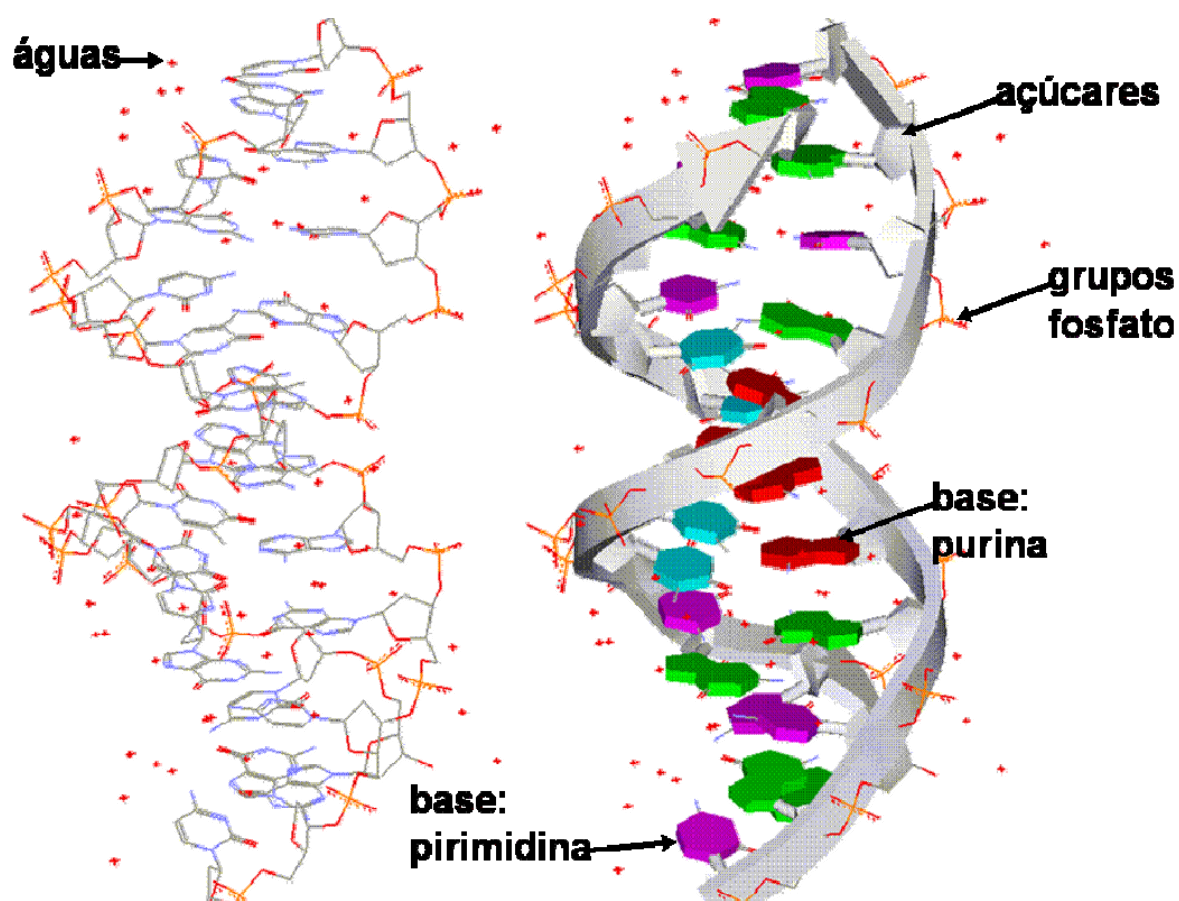
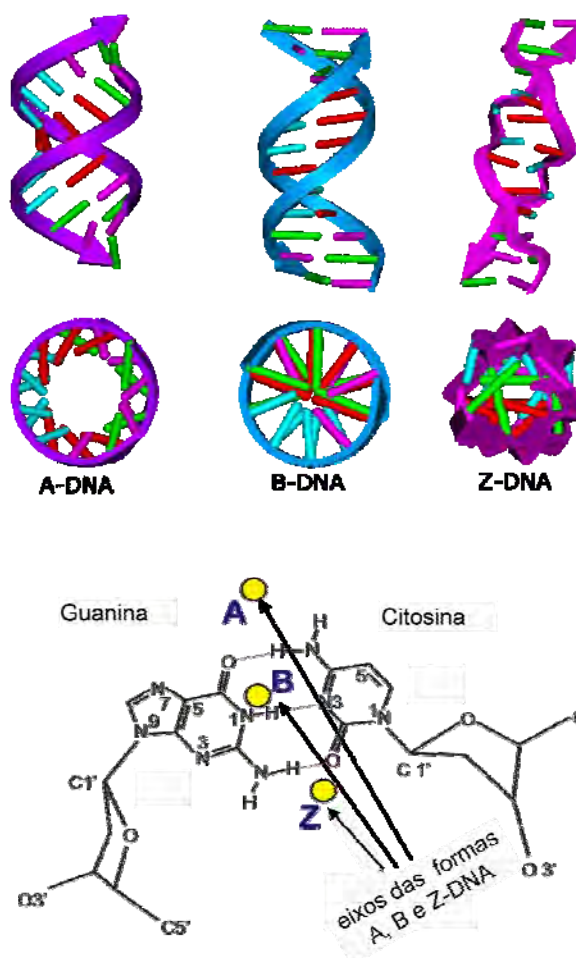


Figura 1. 9 Estrutura cristalográfica do B-DNA 1vzk e sua representação esquemática (direita).

Há três formas biológicas conhecidas do DNA: A-DNA, B-DNA e Z-DNA (Figura 1.10). As formas A e B-DNA são similares do ponto de vista estrutural, uma vez que ambas são hélices de mão direita e as duas fitas de DNA estão em direção opostas: uma das fitas tem a direção exata da sua síntese (5'—3') enquanto que a outra está invertida (3'— 5'). Em condições de baixa hidratação, a forma predominante é A-DNA enquanto que em condições fisiológicas é B-DNA. Na forma Z-DNA as hélices são de mão esquerda. A forma Z parece estar associada à zonas de transcrição e é a que apresenta maior energia entre as três formas (HA *et al.*, 2005). Uma comparação entre as características estruturais das três formas é apresentada na Tabela 1.2.

Tabela 1. 2 Formas do DNA e suas características geométricas (pb = par de base)

Formas	A	B	Z
sentido da hélice	mão direita	mão direita	mão esquerda
unidade de repetição	1 bp	1 bp	2 bp
rotação/pb	33,6°	35,9°	60°/2
pb/volta	10,7	10,0	12,0
inclinação do bp em relação ao eixo	+19,0°	-1,2°	-9,0°
passo/volta da hélice	24,6Å	33,2Å	45,6Å
diâmetro	26 Å	20 Å	18 Å

**Figura 1. 10** (a) Formas A, B e Z do DNA considerando o eixo de enroscamento e embaixo uma visão perpendicular ao eixo. (b) Diferença entre os eixos de enroscamento das fitas.

As quatro bases nitrogenadas do DNA são a Timina (T), Adenina (A), Guanina (G) e Citosina (C). As maiores, A e G, com dois anéis são as purinas e as menores (T e C), as pirimidinas. No DNA existe uma relação $A/T = 1$ e $C/G = 1$. Devido à relação complementar entre as bases A–T e C–G (uma purina associada

com uma pirimidina), o espaçamento entre os corrimãos das hélices contém três anéis. As linhas pontilhadas na Figura 1.11 indicam ligações de hidrogênio, fundamentais para a estabilidade da molécula de DNA. Entre as bases C e G existem três ligações de hidrogênio, enquanto que entre as bases A e T existem somente duas.

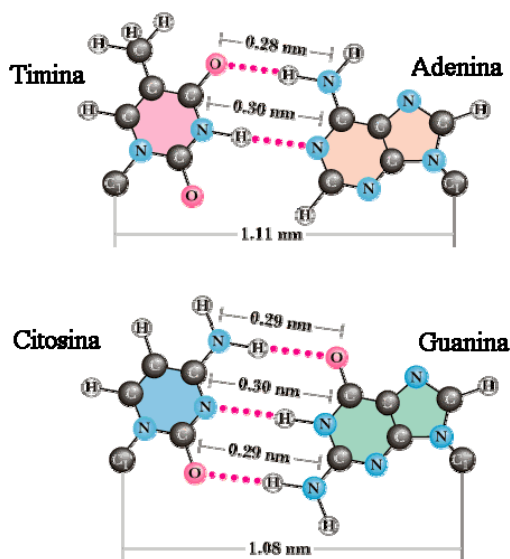


Figura 1. 11 As quatro bases nitrogenadas do DNA: Timina e Adenina, Guanina e Citosina. As linhas pontilhadas indicam ligações de hidrogênio entre as bases C e G e as bases A e T.

1.6 BANCOS DE DADOS DE ESTRUTURAS DE MOLÉCULAS BIOLÓGICAS, ORGÂNICAS E ORGANOMETÁLICAS.

Nos anos 70, um grupo de jovens cristalógrafos, Edgar Meyer, Gerson Cohen e Helen Berman (BERMAN, 2008), reuniu esforços no sentido de estabelecer uma central de coordenadas de estruturas cristalográficas de moléculas biológicas. No início, tratava-se de cerca de uma dúzia de estruturas e hoje já são mais de 50.000 estruturas depositadas (cerca de 75% das estruturas obtidas por difração de raio X, e o restante, por Ressonância Magnética Nuclear – NMR, microscopia eletrônica, difração de elétrons e difração de nêutrons e também por modelos construídos teoricamente). Esse banco de dados, de domínio público, o Protein Data Bank (PDB) era uma base de dados de cristalografia e utilizada por cristalógrafos. Hoje essa base de dados dispõe de dados de estruturas de proteínas, RNA, DNA, e dos complexos formados, como por exemplo, DNA – proteínas, enzimas – inibidores. Outro fato importante, é que seus dados têm sido

acessados não só por cristalógrafos, mas também por pesquisadores de diversas áreas que utilizam dados biológicos estruturais.

Por causa do crescente interesse, o banco de dados PDB também relaciona dados cristalográficos com dados biológicos das moléculas de interesse, com *links* para outros bancos de dados como o de enzimas, dados de sítios ativos, bancos de ligantes, etc. O PDBSum (outro banco que pode ser consultado de forma paralela ao PDB), foi construído com uma forma mais resumida e com uma visualização gráfica mais imediata, permitindo consultar os resultados a uma simples vista (LASKOWSKI, CHISTYAKOV, THORNTON, 2005 e LASKOWSKI, 2007).

Outro banco, o Nucleic Acid Database - NDB (BERMAN *et al.*, 1992), é um repositório de informações sobre a estrutura tridimensional de ácidos nucleicos. Hoje em dia, contém cerca de 3750 estruturas, envolvendo o DNA, RNA, DNA-RNA e complexos de DNA com proteínas, vírus e outras moléculas.

O PDB, PDBSum e NDB estão interligados, de forma que toda a informação pode ser obtida em qualquer um deles, com diferentes visualizações. Todas as informações sobre as moléculas de DNA utilizadas neste trabalho foram obtidas nestes bancos de dados: informações sobre seqüências, coordenadas cristalográficas, presença ou não de ligantes, dados de resolução das estruturas. Quando as informações obtidas são as coordenadas cristalográficas, elas aparecem com o chamado formato PDB.

Por exemplo, os dados referentes ao nucleotídeo citosina (base + desoxirribose + grupo fosfato) aparecem da seguinte forma:

ATOM	39	P	C A	3	25.585	25.276	18.534	1.00	30.62
ATOM	40	O1P	C A	3	27.060	25.288	18.677	1.00	37.04
ATOM	41	O2P	C A	3	24.889	25.966	17.423	1.00	33.38
ATOM	42	O5*	C A	3	25.081	23.755	18.538	1.00	28.67
ATOM	43	C5*	C A	3	25.162	22.978	19.736	1.00	23.94
ATOM	44	C4*	C A	3	24.072	21.936	19.787	1.00	23.50
ATOM	45	O4*	C A	3	22.800	22.540	20.100	1.00	23.10
ATOM	46	C3*	C A	3	23.820	21.202	18.477	1.00	25.60
ATOM	47	O3*	C A	3	24.686	20.071	18.379	1.00	32.06
ATOM	48	C2*	C A	3	22.367	20.817	18.534	1.00	26.01
ATOM	49	C1*	C A	3	21.762	21.715	19.594	1.00	24.49
ATOM	50	N1	C A	3	20.727	22.632	19.093	1.00	22.43
ATOM	51	C2	C A	3	19.390	22.374	19.389	1.00	22.08
ATOM	52	O2	C A	3	19.100	21.374	20.073	1.00	23.13
ATOM	53	N3	C A	3	18.457	23.230	18.918	1.00	24.96
ATOM	54	C4	C A	3	18.829	24.290	18.187	1.00	30.21
ATOM	55	N4	C A	3	17.888	25.125	17.730	1.00	34.79
ATOM	56	C5	C A	3	20.191	24.574	17.870	1.00	27.52
ATOM	57	C6	C A	3	21.097	23.716	18.346	1.00	26.15

onde os átomos da citosina são representados por N1, C2, O2, N3, C4, N4, C5 e C6 os da desorribose são O5*, C5*, C4*, O4*, C3*, O3*, C2* e C1*, e os do grupo fosfato são P, O1P e O2P. Em todos os nucleotídeos, os grupos fosfato e açúcar se repetem quanto à nomenclatura. Na Figura 1.12 estão representados os grupos de cada nucleotídeo e os respectivos nomes dos átomos.

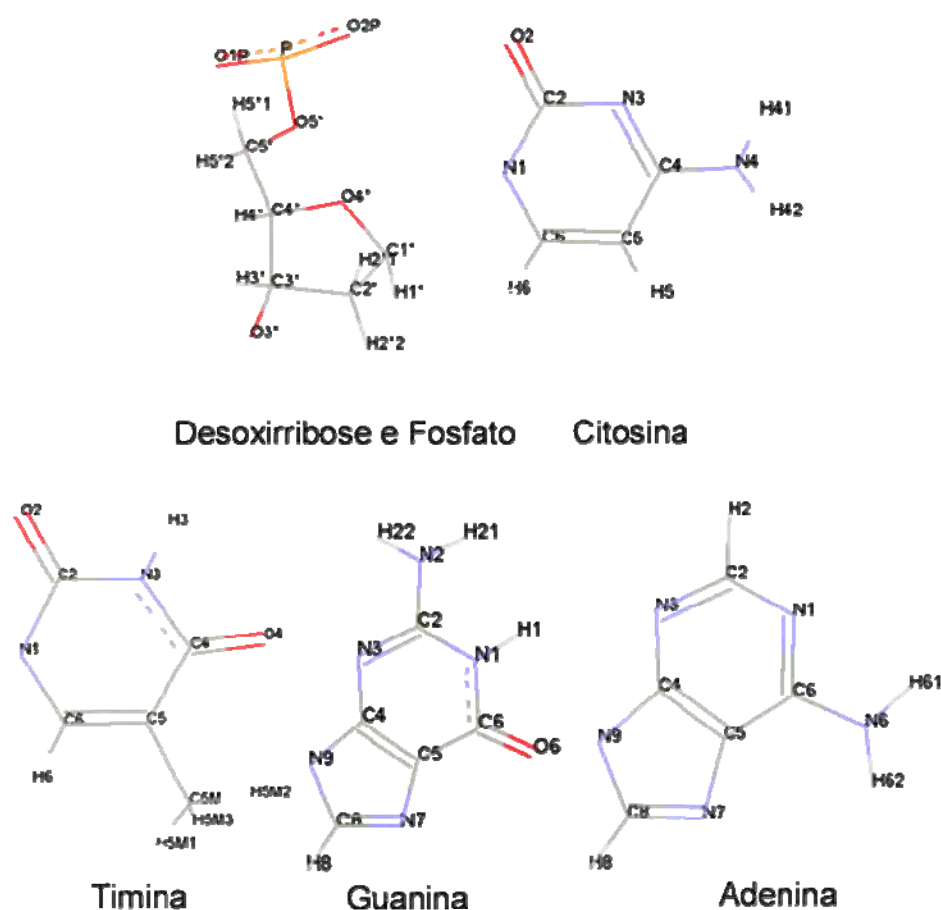


Figura 1. 12 Nomenclatura padrão dos átomos em formato pdb.

Estas informações são fundamentais para os softwares gráficos. A primeira coluna contém o tipo de informação da linha. Esta, pode ser cabeçalho (HEADER), comentário (REMARK), informações sobre os átomos da macromolécula (ATOM) ou seu ligante (HETATM) entre outras. A segunda coluna indica o ID (número de identificação) do átomo. A terceira, contém o nome do átomo o qual segue o seguinte padrão para ácidos nucleicos: todos os átomos são nomeados por seu símbolo acompanhados de um número, conforme mostra a Figura 1.12. A quarta coluna contém o nome do nucleotídeo ao qual pertence o átomo. A quinta coluna contém o nome da cadeia, na sexta seu número. As sétima, oitava e nona colunas

contêm as coordenadas x, y e z do átomo. A décima contém a ocupação, a décima primeira, o fator de temperatura ($10^4 \times \text{\AA}^2$) (PDB, 1998). Toda esta representação recebe o nome de formato pdb.

Um ponto importante a ressaltar, é que o PDB (banco de dados) ou o PDBSum não trazem em seus arquivos os átomos de hidrogênio. Então, na realização de cálculos, medidas de interações ou outros, é necessário utilizar outros programas que adicionam hidrogênios aos arquivos de coordenadas.

Outro banco de dados muito importante, mas não-público, é o Cambridge Structural Database (CSD) que dispõe de uma coleção de estruturas de pequenas moléculas orgânicas e organometálicas (ALLEN, 2002), obtidas por difração de raio X ou difração de nêutrons. Esta base de dados contém cerca de 322.000 estruturas depositadas, sendo atualizado a cada dois meses. Como não é um banco de dados de acesso gratuito, todas as informações e apoio a este trabalho foram resultado de consultas realizadas pelo Prof. Dr. Julio Zukerman Schpector do LaCrEMM – DQ – UFSCar, que possui uma licença para acesso ao CSD.

1.7 DOCKING MOLECULAR

Docking molecular, ou simplesmente docking, é o processo de se encontrar o melhor ajuste para o encaixe entre duas moléculas tridimensionais. Inicialmente esta busca era realizada visualizando as moléculas em programas gráficos, onde se tentava manualmente obter o melhor ajuste para as duas moléculas formarem um complexo. Normalmente uma molécula é chamada de receptor (proteínas, enzimas, DNA, etc.) e a outra molécula é chamada de ligante (em geral, a molécula menor).

Hoje em dia, há vários programas que permitem realizar essa busca de forma mais automatizada. Podemos dividir o processo de “docking” em três partes:

- planejamento do experimento *in silico* (ou seja, realizado utilizando programas e simulações computacionais);
- realização do “experimento” com objetivo de formar complexos;
- avaliação dos resultados obtidos no experimento.

Para a parte de planejamento, necessita-se primeiro buscar quais as moléculas serão estudadas, informações disponíveis sobre as mesmas (por exemplo, dados de cinética enzimática, se for o caso de enzimas, dados de bioquímica, etc.). A escolha também depende da disponibilidade de dados cristalográficos das moléculas envolvidas. No caso dos receptores biológicos, a consulta será realizada no PDB. Para moléculas orgânicas, o desejável é que se disponha das estruturas cristalográficas das moléculas, ou, se não for possível, obtidas por modelagem molecular. Depois da escolha das moléculas, procura-se o sítio de ligação a ser estudado, que será o centro de uma esfera de raio R , em torno do qual os programas computacionais realizam sua busca. Terceiro, adicionar hidrogênios às moléculas a serem estudadas.

Depois se utiliza um programa computacional que será uma ferramenta na realização de cálculos. Há basicamente três possibilidades metodológicas utilizadas pelos programas computacionais: o docking rígido, o docking semiflexível e o docking flexível. No primeiro caso, as duas moléculas são consideradas como corpos rígidos e o programa busca a melhor orientação de um corpo em relação a outro. No semiflexível, permite-se que o ligante possa variar seus ângulos torsionais. No último caso, permite-se que o receptor também sofra flexibilização parcial ou total. Os programas computacionais variam também no que diz respeito à sua essência metodológica de tal forma que a busca de conformações pode ser feita utilizando métodos de mecânica molecular, métodos estocásticos e algoritmos genéticos ou outros. Outro fator que também varia de programa para programa é a chamada “função escore” que determina a melhor orientação e/ou conformação das moléculas. A função escore pode estar baseada em métodos empíricos, campos de força ou potenciais variados.

No docking rígido, o algoritmo de busca explora diferentes posições para o ligante no sítio ativo do receptor usando os graus de liberdade translacional e rotacional (Figura 1.13). O docking de ligantes flexíveis adicionalmente, explora os graus de liberdade torcionais do ligante nesse processo. Uma função escore deve ter uma qualidade de predição que forneça uma pontuação favorável à formação de complexos realísticos (essa qualidade será analisada mais adiante quando for tratado o redocking). Usualmente, a função escore avalia ambos, a

complementaridade estérica entre o ligante e o receptor e a sua complementaridade química (BROOIJMANS, KUNTZ, 2003).

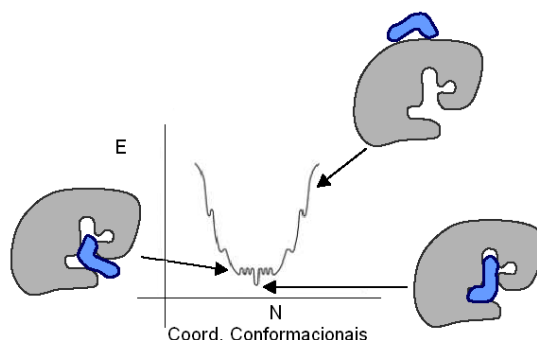


Figura 1.13 O gráfico da energia em função da conformação mostra que a mínima energia também resulta no melhor resultado receptor-ligante em um caso de docking rígido.

Os primeiros procedimentos de docking começaram a acontecer logo depois do desenvolvimento dos programas gráficos moleculares, dado que eles permitiram aos pesquisadores manipular duas moléculas e encontrar diferentes conformações que pareciam complementares geométrica e quimicamente. Os programas de docking molecular permitiram a automação da busca e um modo mais objetivo de avaliar o ajuste entre duas moléculas. O número de modos que duas moléculas podem interagir é grande e para tornar a busca eficiente, o problema do docking molecular é simplificado e somente certos graus de liberdade são explorados.

Os principais métodos estocásticos são Monte Carlo (MC) e Algoritmo Genético (GA). No primeiro é utilizado especificamente o método Metrópolis, onde o movimento aleatório do ligante no sistema é aceito ou rejeitado com base na probabilidade de Boltzmann. O segundo é baseado na teoria da evolução de Darwin e é uma poderosa ferramenta para a solução de problemas complexos. Essa técnica imita o processo de evolução manipulando uma coletânea de dados de estrutura chamados cromossomos. Cada uma dessas estruturas codifica uma possível solução isto é, uma possível orientação do ligante dentro do sítio da macromolécula ao problema do docking, e tem relacionado a si uma pontuação (escore). Boas soluções, ou “indivíduos” melhor encaixados com energia mais baixa, sofrerão mutação e recombinação para a geração de um novo grupo de

soluções. Este ciclo se repete até que a variação de energia entre diversas conformações esteja dentro de valor de corte (JONES *et al.*, 1997).

Neste trabalho será utilizado o programa GOLD – Genetic Optimisation for Ligand Docking (JONES, WILLETT E GLEN, 1995, JONES *et al.*, 1997) algoritmo genético desenvolvido para ligantes flexíveis. Na próxima seção, será feita uma breve descrição da idéia existente neste tipo de algoritmo.

1.8 ALGORITMO GENÉTICO (GA)

É um algoritmo estocástico utilizado para solucionar problemas de otimização. A estrutura básica de um GA está apresentada na forma de um fluxograma na Figura 1.14. Neste trabalho, a situação a ser otimizada é a modo de ligação entre algumas pequenas moléculas e o DNA 5'-CGCGAATTCGCG-3'.

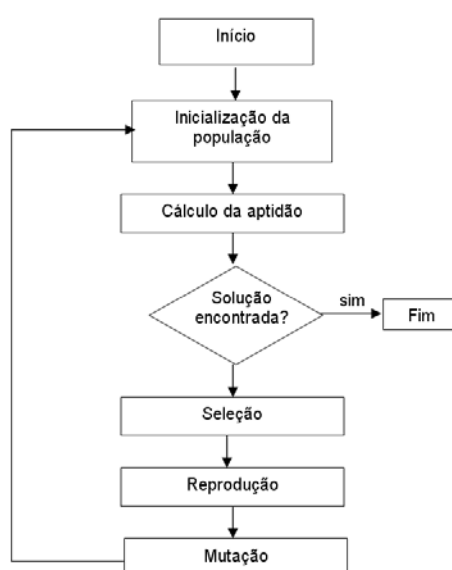


Figura 1.14 Fluxograma de um GA básico (MIRANDA, 2000).

Os principais conceitos utilizados em GA são: a) **gene**, que representa cada parâmetro a ser explorado na otimização; b) **cromossomo**, também chamado de indivíduo, que é a cadeia de genes que representa uma solução possível para o problema e c) **população**, que é o conjunto de soluções no espaço de busca.

O processo de solução adotado nos algoritmos genéticos consiste em gerar um grande número de indivíduos de forma a promover uma varredura tão extensa quanto necessária do espaço de soluções.

O procedimento básico de um GA consiste em inicializar uma população de n indivíduos aleatoriamente. Cada um dos indivíduos da população representa uma possível solução para o problema. Abaixo um exemplo de uma população com 3 indivíduos, cada um representado por uma cadeia de 8 bits.

1	0	0	1	1	0	1	0	1	1	0	1	1	0	0	1	1	0	1	1	0	1	1	0
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

Posteriormente é calculada a aptidão de cada indivíduo conforme a função objetivo, que depende das especificações de problema. Neste trabalho, cada indivíduo representa uma conformação do ligante no sítio do DNA, cuja saída fornece uma pontuação (ou score) do complexo formado o qual permite ao algoritmo genético o cálculo da aptidão do indivíduo. Ainda nesta fase os indivíduos são ordenados conforme a sua aptidão, isto é, maior pontuação. Nesta fase os indivíduos mais aptos da geração atual são selecionados. Esses indivíduos são utilizados para gerar uma nova população por cruzamento, onde cada indivíduo tem uma probabilidade de ser selecionado proporcional à sua aptidão. Como consequência, a seleção de indivíduos pode conter várias cópias de um mesmo indivíduo enquanto outros podem desaparecer, devido a sua probabilidade ser menor. Desta maneira, os mais aptos tem maior probabilidade de serem selecionados.

O cruzamento (*crossover*) é o processo de embaralhar aleatoriamente todos os indivíduos selecionados na etapa anterior, criando-se, desta forma, uma segunda lista, chamada lista de parceiros. Cada indivíduo selecionado é então cruzado com o indivíduo que ocupa a mesma posição na lista de parceiros. A forma como se realiza este cruzamento, é ilustrada na Figura 1.15. Os cromossomos de cada par de indivíduos a serem cruzados são particionados em um ponto, chamado ponto de corte, sorteado aleatoriamente.

Um novo cromossomo é gerado permutando-se a metade inicial de um cromossomo com a metade final do outro.

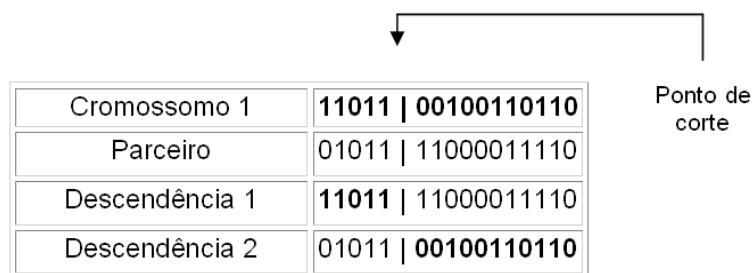


Figura 1. 15 Cruzamento do cromossomo 1 com o cromossomo 2. Um ponto de cruzamento representado por uma barra é escolhido aleatoriamente. Copia-se tudo o que vem antes de um dos pais e tudo o que vem depois do outro.

A operação de mutação (Figura 1.16) consiste em alterar aleatoriamente o valor de um determinado gene para assegurar uma maior varredura do espaço de soluções e evitar que o algoritmo genético convirja muito cedo para mínimos locais.

Descendência Original 1	110 1 111000011110
Descendência Original 2	1101100 1 00110110
Descendência Mutada 1	110 0 111000011110
Descendência Mutada 2	110110 1 100110110

Figura 1. 16 Exemplo de mutação

Há muita bibliografia sobre o tema já que este tipo de algoritmo tem sido empregado em vários tipos de aplicações. Uma descrição mais detalhada pode ser encontrada em Mitchell (1996).

1.9 INTERAÇÕES COM DNA

Recentes desenvolvimentos têm sido feitos no estudo de modelos da dupla hélice investigando-se principalmente o papel da sequência de bases em sua flexibilidade (LANKAS *et al.*, 2000). No entanto, a importância do papel do DNA na vida celular só pode ser considerado quanto à sua interação com outras moléculas. Dentre as interações que poderiam ser citadas, destacam-se:

- complexos com proteínas que se ligam especificamente a diferentes partes da sequência do DNA para realizar importantes tarefas biológicas, tais como, a repressão e ativação de genes, desenovelamento da hélice para cópia e replicação, entre outras;
- as interações consigo próprio no processo de espiralamento;
- complexos com RNA no processo de replicação (CALLADINE *et al.*, 2004).
- complexos com pequenas moléculas orgânicas ou inorgânicas.

Neste trabalho foi dada ênfase ao último tipo de interação, sendo que a compreensão de como elas ocorrem pode fornecer subsídios que permitam trabalhar com estas moléculas para atuarem como fármacos, seja com fins terapêuticos ou ainda para evitar a formação de complexos que danifiquem o DNA. Como exemplo da primeira aplicação pode ser citada a cisplatina empregada no tratamento de câncer (VAN GARDEREN, VAN HOUTE, 1994), e como exemplo da segunda a bleomicina (ZHAO *et al.*, 2002), um antibiótico que quando utilizado indiscriminadamente pode causar câncer. As informações obtidas a partir da visualização gráfica das várias estruturas de DNA-ligante depositadas no PDB permitem verificar que as estruturas tridimensionais do DNA e do ligante têm grande influência na formação do complexo.

O estudo e *design* de moléculas que possam reconhecer seqüências específicas é fundamental para o reconhecimento molecular, bem como para o desenvolvimento de novas terapias e reagentes para a biotecnologia. Parece que o desenvolvimento de novos fármacos pode ser promissor utilizando ácidos nucleicos como possíveis alvos (MUNDE *et al.*, 2007).

As pequenas moléculas que se ligam ao DNA ocupam papel central na farmacologia e bioquímica devido ao papel-chave que ocupam nos processos de replicação, transcrição, recombinação e mutagenese. A habilidade de pequenas moléculas interferirem nas funções do DNA as tornam um alvo para interação com fármacos. No caso de ligantes aromáticos que se ligam ao DNA, pode ocorrer intercalação por sua interação tipo empilhamento (*stack*) que interfere na interação do DNA com proteínas, ou o ligante pode entrar no sulco menor, podendo causar pequenas distorções no DNA pela acomodação que as duas moléculas sofrem ao formar o complexo. (BERA *et al.* 2008).

Os ligantes que se ligam de modo não-covalente ao DNA no sulco menor tem demonstrado larga atividade contra câncer, antivírus, antiprotozoária e contra infecções (EVANS, NEIDLE, 2006, CAMPBELL *et al.*, 2006).

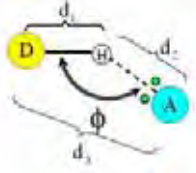
Estruturas cristalinas de um grande número de complexos de DNA com possíveis fármacos, foram determinadas (ver o resultado do levantamento nas Tabelas 3.1 e 3.2). Isto constitui uma fonte significativa de dados para o

desenvolvimento, análise e compreensão de mecanismos de reação de fármacos, mecanismos de interação dependentes de formas e seqüências de DNA, tendo como alvo moléculas de DNA (EVANS, NEIDLE, 2006).

As forças intermoleculares são as principais responsáveis pela forma tridimensional das macromoléculas e também por sua interação com outras moléculas, sendo que as principais são: interações eletrostáticas, interações dipolares, ligações de hidrogênio, interações hidrofóbicas e interações de van der Waals. Essas interações ocorrem quando as moléculas se organizam como moléculas individuais e também quando se associam a outras formando complexos. O estudo destas interações permite que se analise os complexos virtuais formados através de simulações computacionais de docking. (ERMONDI, CARON, 2006).

O Programa GOLD, utilizado neste trabalho para a realização dos cálculos de docking, tem uma função escore que envolve as energias decorrentes das ligações de hidrogênio e as interações de van der Waals, que normalmente tem um peso maior na energia de ligação da formação dos complexos.

Na Tabela 1.3 são apresentadas as forças intermoleculares utilizadas na função escore do programa GOLD. Esta tabela foi adaptada do trabalho de Ermondi e Caron, 2006.

Tabela 1.3 Resumo das interações consideradas na função escore do Programa GOLD (adaptada de Ermondi e Caron, 2006)	
Categoria	Ordem de energia dos valores totais
ligações de hidrogênio 	Ocorre entre um átomo eletronegativo (A – aceceptor) e um átomo de hidrogênio ligado a um átomo relativamente eletronegativo (D – doador). Energias: 4 a 60 kJ/mol.
interações de van der Waals	Energias: 50 a 500 kJ/mol.

As pequenas moléculas podem ligar-se ao DNA por dois modos principais: ligação nos sulcos (maior ou menor) ou por intercalação, com uma deformação e alargamento do DNA. (Figura 1.17).

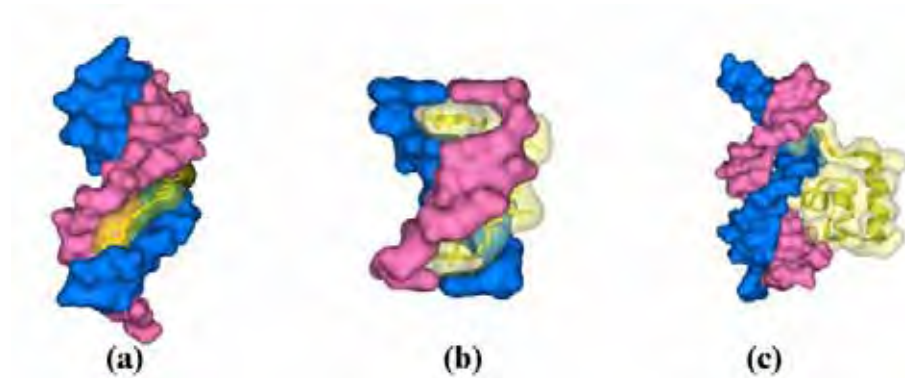


Figura 1. 17 (a) Benzimidazol ligado ao DNA no sulco menor (código PDB:1vzk). (b) Antraciclina ligada ao DNA por intercalação (código PDB: 1nab). Observa-se uma deformidade das bases. (c) A enzima recombinase atuando no sulco maior do DNA (código PDB: 1jkq).

Pode-se observar experimentalmente que a ligação de moléculas nos sulcos perturba pouco a estrutura do DNA, porém uma ligação por intercalação deve distanciar as bases adjacentes tornando o DNA mais alargado. (CHAIRES, 2006).

1.10 COMPOSTOS DE CRÔMIO

Muitos metais desempenham papéis biológicos fundamentais. Os canais de potássio, por exemplo, fazem parte da atividade dos nervos e contração muscular. Já, outros metais auxiliam alguma função do organismo, mas não são considerados essenciais. Exemplos são o Cr(III) que auxilia no metabolismo da insulina e o ouro, que é utilizado no tratamento de artrite reumatóide e nas pesquisas de ponta como as da área da nanobiotecnologia, desenvolvida pelo grupo do Prof. Edward Tiekink, University of Texas, San Antonio, Texas (SARMA *et al.* 2007, BROWN *et al.*, 2007, M., LI *et al.*, 2007).

Uma característica dos metais de transição é a sua capacidade de mudar de estado de oxidação. Se por um lado essa propriedade é benéfica, como no caso do transporte de oxigênio ($\text{Fe (III)} \leftrightarrow \text{Fe(II)}$), a transferência de elétrons feita por enzimas ligadas ao íon, por outro, pode causar sérios danos à saúde. É o caso do íon de crômio (VI), que ao mudar seu estado de oxidação no interior da célula pode causar a clivagem do DNA (FERREIRA, 2002). No estado de oxidação (III), no entanto, o crômio pode auxiliar no metabolismo de ação da insulina. Em 1959, ratos com necrose do fígado e intolerância à insulina foram tratados com sucesso com GTF (Glucose Tolerance Factor), uma molécula com estrutura tridimensional

ainda indefinida, mas que continha o íon crômio (III). Atualmente Crômio(III) tris(picolinato) ou Cr(pic)_3 , é utilizado como suplemento nutricional, porém, em elevadas concentrações foi demonstrado recentemente que pode quebrar o DNA (SPEETJENS *et al.* 1999a).

No entanto, compostos de Cr(III) também podem causar câncer, mas o mecanismo com que isso ocorre é totalmente diferente dos compostos de Cr(VI). Sua interação com o DNA, a qual é muito mais dependente da molécula orgânica que acompanha o íon, do que do próprio íon, é responsável por impedir a atuação de enzimas iniciando um processo carcinogênico (VAIDYANATHAN *et al.*, 2005).

2 METODOLOGIA

Neste trabalho foram feitas simulações computacionais nas quais o receptor estudado foi uma molécula de DNA 5'-CGCGAATTCGCG-3', e os ligantes foram pequenas moléculas, uma vez que a habilidade de pequenas moléculas interferirem na transcrição e replicação do DNA faz com que sejam um alvo importante para o estudo da interação de fármacos. O método de simulação foi o docking molecular, utilizando o programa GOLD, cuja estratégia está baseada em algoritmos genéticos. Um diagrama do fluxo de trabalho está apresentado na Figura 2.1.

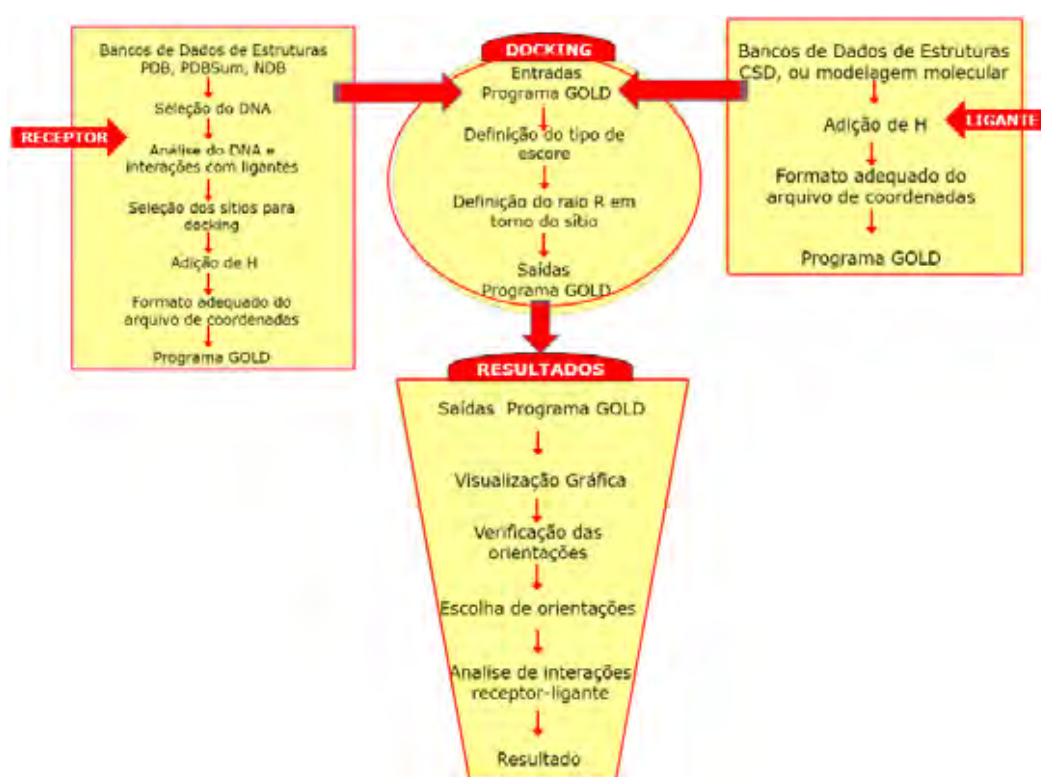


Figura 2. 1 Fluxograma de trabalho

2.1. RECEPTOR

2.1.1. Busca nos Bancos de dados de informações estruturais do DNA

Para escolher a molécula de trabalho, foi feita uma busca no banco de dados NDB, e como o banco está conectado ao PDB e PDBSum, a informação da

estrutura cristalográfica é imediata. A tela de acesso ao NDB, com os parâmetros adotados para a busca está representada na Figura 2.2. Observar que as buscas foram feitas no sentido de encontrar complexos DNA-pequenas moléculas.

The screenshot shows the NDB Search interface with the following sections and parameters:

- General Information:**
 - NDB ID:
 - PDB ID:
 - Author: (ex: Last F or Last; not Last F or F. Last)
 - Citation Year (ex: 2002):
 - Released Since:
- Experimental Type:**
 - Crystal Structure: ☒ Y ☐ N
 - Structure Factors Available: ☐ Y ☒ N
 - Space Group:
 - Resolution better than:
 - R-factor better than <:
 - NMR: ☐ Y ☒ N
 - NMR Restraints Available: ☐ Y ☒ N
- Sequence:**
 - Nucleic Acid Sequence Pattern:
 - Mismatch: ☐ Y ☒ N
- Biomolecule contains:**
 - DNA: ☒ Y ☐ N
 - RNA: ☐ Y ☒ N
 - Protein: ☐ Y ☒ N
 - Ligand: ☒ Y ☐ N
- X-ray Nucleic Acid Modification (You should only search for X-ray structures for these fields!)**
 - Base: ☐ Y ☒ N
 - Sugar: ☐ Y ☒ N
 - Phosphate: ☐ Y ☒ N
- X-ray Structural Features (You should only search for X-ray structures for these fields!)**
 - Double Helix: ☒ Y ☐ N
 - Triple Helix: ☐ Y ☒ N
 - Quadruple Helix: ☐ Y ☒ N
 - Single Strand: ☐ Y ☒ N
 - Internal loop: ☐ Y ☒ N
 - Tetraloop: ☐ Y ☒ N
 - Hairpin loop: ☐ Y ☒ N
 - Bulged bases: ☐ Y ☒ N
 - 3 way junction: ☐ Y ☒ N
 - 4 way junction: ☐ Y ☒ N

Search buttons:

Figura 2.2 Tela de Entrada do algoritmo de busca do NDB, com os parâmetros adotados.

2.1.2. Levantamento dos complexos de DNA-ligante

O PDB contém 2601 estruturas contendo DNA, obtidas por difração de raio X ou Ressonância Magnética Nuclear (NMR). As moléculas de DNA aparecem isoladas, complexadas com proteínas, RNA, moléculas orgânicas ou íons metálicos. Podem ainda ser encontradas nas mais diversas seqüências de pares de bases, bem como em fita simples ou dupla hélice e ainda em outras formas.

Para os propósitos desse trabalho, foram inicialmente selecionados utilizando algoritmo de busca do NDB, todas as estruturas de B-DNA com ligantes sem proteínas e sem RNA e obtidas por difração de raio X. Obteve-se inicialmente 179 estruturas. Dessas, foram selecionadas aquelas cuja seqüência era CGCGAATTCGCG. Posteriormente as estruturas que continham somente íons interagindo com a seqüência foram descartadas, pois não se adequavam aos propósitos do trabalho. Assim foram analisadas as 25 estruturas que estão descritas na Tabela 3.1 e seus respectivos ligantes na Tabela 3.2, do Capítulo 3 de Resultados e Discussão.

2.1.3. Seleção e Análise do DNA

Para fazer a escolha das moléculas a serem utilizadas nos cálculos, foram utilizados programas de visualização gráfica, que permitiram fazer uma inspeção na estrutura cristalográfica do complexo. Os programas de visualização gráfica utilizados estão na Seção 2.4.

2.1.4. Seleção de sítios para o cálculo de docking

A partir da visualização gráfica e medidas dos parâmetros envolvidos nas interações, distâncias e ângulos, foram estabelecidos os possíveis sítios para o docking. Em cada caso, foi selecionado um centro (em torno do qual se determina um raio R) para o docking. Todos os átomos dentro da esfera de raio R foram considerados.

2.1.5. Adição de Hidrogênios

Para realizar os cálculos é necessário obter as coordenadas cristalográficas do PDB (formato pdb) e adicionar hidrogênios, já que as estruturas estão depositadas sem eles. Para realizar esta tarefa, o programa tem que ter parametrizados todos os tipos de átomos. O programa utilizado para realizar a adição de hidrogênios, foi o Vega (PEDRETTI, VILLA, VISTOLI, 2002, 2003, 2004).

O formato mol2, além das coordenadas x, y e z das moléculas, também informa a hibridização do átomo, pode apresentar ou não as cargas parciais, a conectividade e o tipo de ligação. O programa GOLD aceita as duas possibilidades de formatos: pdb e mol2, mas por uma questão de conveniência, todas as moléculas foram salvas em formato mol2 (arquivo de entrada) e foram obtidas nesse formato nos arquivos de saída do programa GOLD.

O programa GOLD aceita duas possibilidades de formatos: pdb e mol2 (TRIPOS); tanto os receptores utilizados no trabalho R-1vzk e R-2b3e (nomes derivados dos complexos obtidos do PDB, 1vzk e 2b3e), quanto os ligantes foram formatados no formato mol2 ao construir os arquivos input do programa GOLD.

Um exemplo do arquivo em formato mol2:

```

<TRIPOS>MOLECULE
BENZIMIDAZOL
41 44 0 0 0
SMALL
CASTEIGER
USER CHARGES
<TRIPOS>ATOM
      número de átomos
      número de ligações
      coordenadas x,y,z
      hibridização
      cargas parciais
1 C1      10.4960  24.7480  5.1570 C.ar      1 BENZ      0.0200
2 S2      10.8640  23.5850  6.3590 S.2      1 BENZ      -0.0851
3 C3      9.3890   23.8880  7.1520 C.ar      1 BENZ      0.0569
4 C4      8.6300   24.8360  6.5250 C.ar      1 BENZ      -0.0433
5 C5      9.2570   25.3310  5.3830 C.ar      1 BENZ      -0.0463
17 C17     12.0560  26.4610  2.2350 C.ar      1 BENZ      -0.0461
18 C18     13.1270  25.5950  1.9000 C.ar      1 BENZ      0.0667
19 C19     13.2800  24.4340  2.6890 C.ar      1 BENZ      -0.0461
20 C20     12.4270  24.1570  3.7570 C.ar      1 BENZ      -0.0527
21 C21      9.1370  19.6370  13.1840 C.2      1 BENZ      0.3159
22 N22     10.4010  19.4980  13.6000 N.pl3     1 BENZ      -0.0978
23 N23      8.2660  18.8560  13.8360 N.2      1 BENZ      -0.2839
24 C24     13.9750  25.8990  0.8150 C.2      1 BENZ      0.3158
25 N25     14.3680  24.9540  -0.0520 N.pl3     1 BENZ      -0.0978
26 N26     14.4180  27.1250  0.5350 N.2      1 BENZ      -0.2839
27 H       7.6420  25.1690  6.8800 H      1 BENZ      0.0630
28 H       8.8170  26.1020  4.7300 H      1 BENZ      0.0629
41 H       14.0940  23.7290  2.4550 H      1 BENZ      0.0626

<TRIPOS>BOND
      conectividade
      tipo de ligação:
      ar -> aromático
      1 -> simples
      2 -> dupla
1 1 15
2 1 5 ar
3 1 2 ar
4 2 3 ar
5 3 4 ar
6 3 6 1
7 4 5 ar
8 4 27 1
9 5 28 1
10 6 7 ar
11 6 10 ar
12 7 8 ar
13 8 9 ar
14 8 11 ar
32 19 20 ar
33 19 34 1
36 21 23 2
37 22 36 1
38 23 37 1
39 23 38 1
40 24 25 1
41 24 26 2
42 25 39 1
43 26 40 1
44 26 41 1

```

Na seção “@<TRIPOS>ATOM”, seção da estrutura 3D, além das informações de número, nome e coordenadas x, y e z dos átomos nas cinco primeiras colunas, esse tipo de arquivo também contém, na sexta coluna, a hibridização dos átomos. Por exemplo, C.2 significa carbono sp^2 , N.ar, significa nitrogênio que faz parte de um anel aromático. As últimas 3 colunas referem-se ao número de fragmento (resíduo ou base nitrogenada no caso de proteínas e ácidos nucleicos, respectivamente), a última coluna é a das cargas parciais de cada átomo, neste exemplo nulas. Na seção “@<TRIPOS>BOND” são fornecidas as ordens de ligação entre os átomos, no exemplo: simples, **1**, ou dupla, **2**, também existem opções para tripla, **3**, ou aromática, **ar**.

2.2. LIGANTES

Os ligantes estudados foram obtidos do PDB, CSD ou modelados quando a estrutura cristalográfica não estava disponível.

2.2.1. Ligante 1

A estrutura do ligante 1, de nome 2-(5-{4-[amino (imino)metil]fenil}-2-tienil)-1h-benzimidazol-6-carboximida, foi obtida do complexo cristalográfico 1vzk (MALLENA *et al.*, 2004). Esta molécula também foi estudada como ligante no trabalho de Campbell *et al.* (2006), com o código DB818 (Figura 2.3).

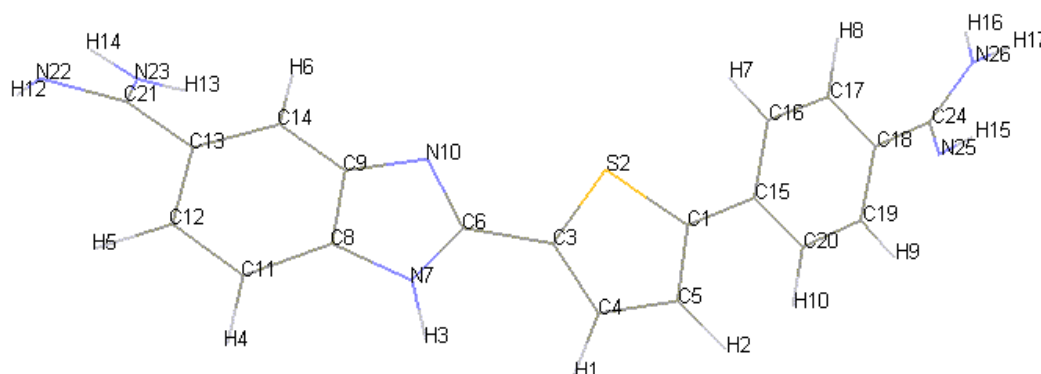


Figura 2.3 Ligante 1 – DB818 - com a numeração dos átomos utilizada neste trabalho.

Este ligante foi extensamente estudado neste trabalho, desde a etapa inicial de estudo complexo cristalográfico de partida, até nas etapas de validação e otimização das opções de cálculo utilizadas posteriormente com os outros ligantes.

2.2.2. Ligante 2

O ligante 2 é um composto de cromo de fórmula geral $C_{25}H_{26}ClCrN_2O_9$, cujo nome é $[Cr(1,2\text{-bis}(\text{naftilidenoamino})\text{etano})(H_2O)_2]$. A estrutura desse ligante foi obtida do CSD, onde está depositada com o código DAZJOE e está apresentada na Figura 2.4. A molécula também foi estudada no trabalho de Vaidyanathan *et al.* (2005).

A escolha desse composto deu-se pelo fato de que, apesar dos compostos de Cr(III) serem considerados não carcinogênicos, este pode danificar o DNA fazendo com que a célula exiba atividade tipo nuclease que é evidência de que enzimas reparadoras de DNA estão sendo ativadas (VAIDYANATHAN *et al.*, 2005).

A molécula DAZJOE é uma base de Schiff, mesma característica de uma série de fármacos largamente utilizados no tratamento de tumores, sendo que sua eficácia aumenta quando complexados com metais (VAIDYANATHAN *et al.*, 2005).

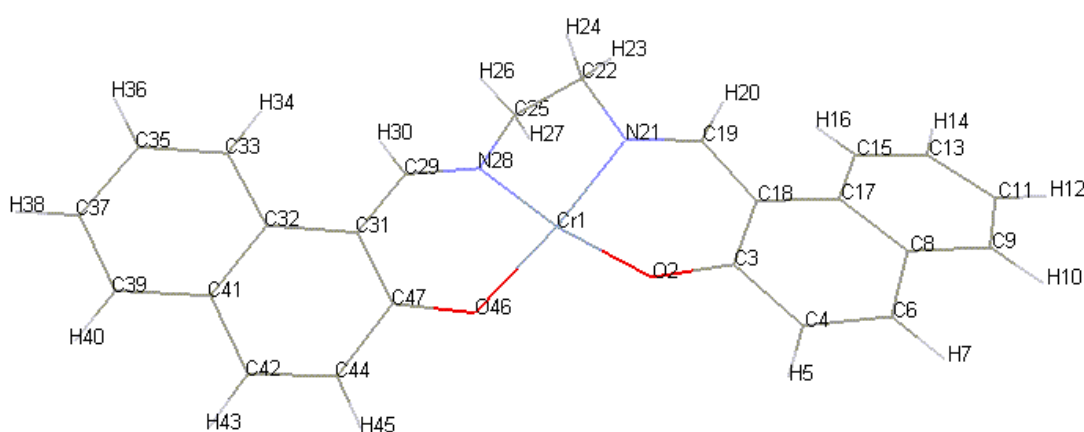


Figura 2.4 Ligante 2 - DAZJOE - com a numeração dos átomos utilizada neste trabalho

2.2.3. Ligante 3

O ligante **3** é o composto de cromo [Cr(2,3-bis{[(2-hidroxi-4-dietilamino) (fenil) (metileno)]amino}2-butenodinitrila) (H₂O)₂], (CR3CN). Esta molécula aparece citada no artigo de Vijayalakshmi *et al.* (2000). Como não se contava com uma estrutura cristalográfica deste ligante, ele foi modelado a partir de DAZJOE. A geometria de CR3CN foi otimizada utilizando o método semi-empírico PM6, implementado no programa MOPAC2007 (Stewart, 2007). A estrutura deste ligante está apresentada na Figura 2.5.

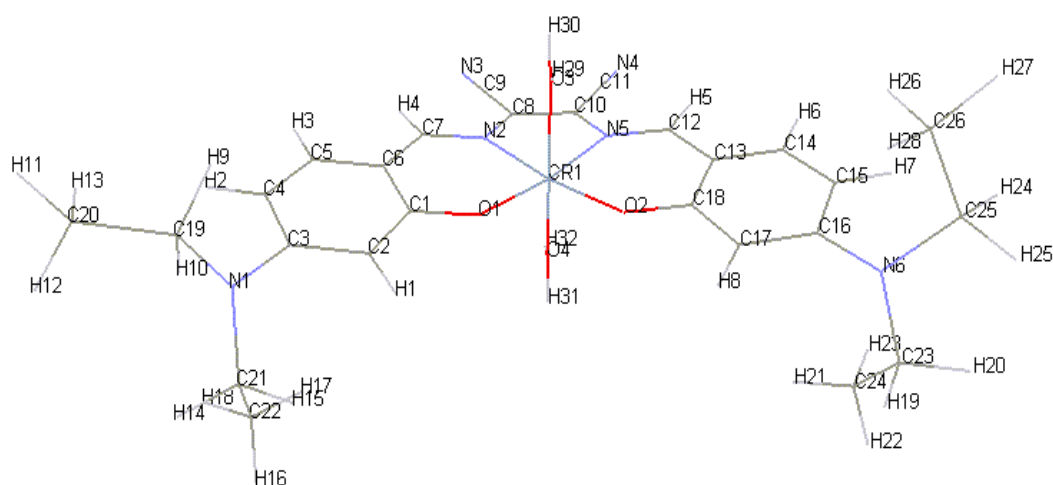


Figura 2.5 Ligante 3 - CR3NC - com a numeração dos átomos utilizada neste trabalho.

2.2.4. Adição de Hidrogênios

ligante 1: **DB818** - Os hidrogênios foram adicionados utilizando o programa Vega

ligante 2: **DAZJOE** – A molécula foi obtida do CSD com os hidrogênios.

ligante 3: **CR3NC** – Os hidrogênios foram adicionados durante a modelagem.

2.3. DOCKING MOLECULAR

2.3.1. O que é docking?

O docking molecular é um método que permite estudar como duas estruturas tridimensionais se ajustam uma à outra, da melhor forma possível. Este método tem sido muito utilizado para estudar como algumas moléculas podem inibir enzimas e planejar fármacos. A mesma idéia tem sido aplicada ao estudo de DNA.

No docking o que se pretende é a formação de um complexo teórico, partindo de duas estruturas tridimensionais obtidas independentemente (Seção 1.7). Em geral, denomina-se a molécula maior receptor e a menor de ligante. Mas, ultimamente também tem sido um método utilizado para estudar como macromoléculas interagem entre si.

Aqui o receptor será uma molécula de DNA (obtida do PDB) e o ligante será uma molécula pequena que pode ser obtida por cristalografia (CSD) e/ou modelagem molecular.

Um esquema mostrando a idéia do docking molecular está apresentada na Figura 2.6.

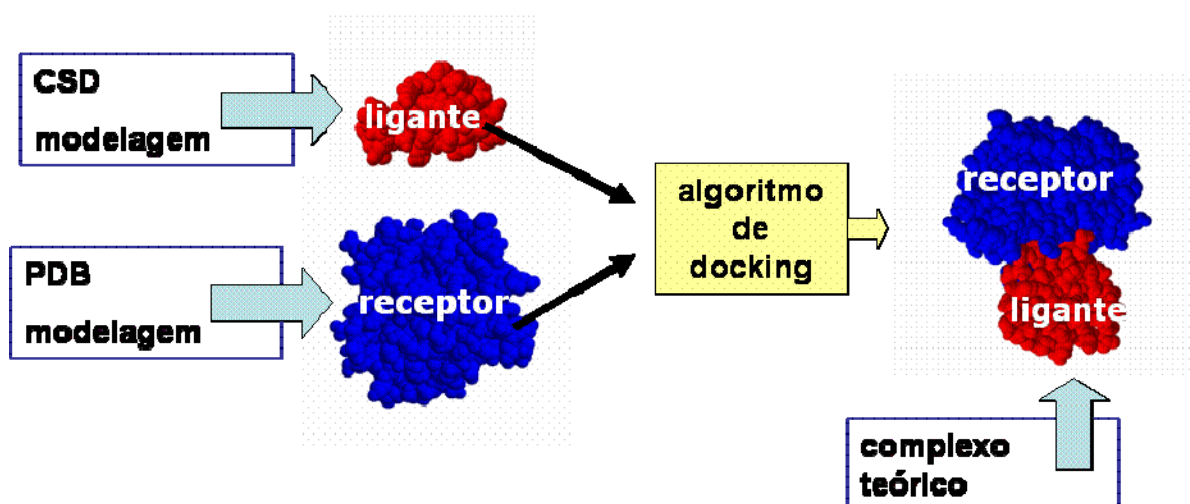


Figura 2.6 Docking: a formação de um complexo tridimensional teórico partindo de duas estruturas tridimensionais

Na Figura 2.6, observa-se que há um algoritmo de docking. Na realidade, no início dos estudos de docking, tratava-se de um método manual utilizando os

programas gráficos de visualização molecular. No caso deste trabalho, o programa GOLD, descrito na Seção 2.3.4, foi utilizado para criar os complexos virtuais que foram analisados posteriormente em tela gráfica.

2.3.2. Redocking ou validação

Normalmente, quando o trabalho de docking tem início, é necessário verificar se todas as escolhas feitas (a macromolécula, o estado da macromolécula, o sítio de ligação, o raio R utilizado nos cálculos para construir a esfera que inclui todos os átomos, parâmetros utilizados no programa de cálculos de docking) são adequadas. Para tal, escolhe-se (quando possível) uma macromolécula do PDB que pertence a um complexo do tipo que se pretende formar. No caso do trabalho presente, a estrutura selecionada foi 1vzk, um complexo DNA-ligante. A molécula foi obtida do PDB. Como primeiro passo transforma-se o complexo cristalográfico (PDB) em duas moléculas isoladas, e tenta-se reconstruir o complexo inicial. Um esquema mostrando os passos do redocking está apresentado na Figura 2.7.

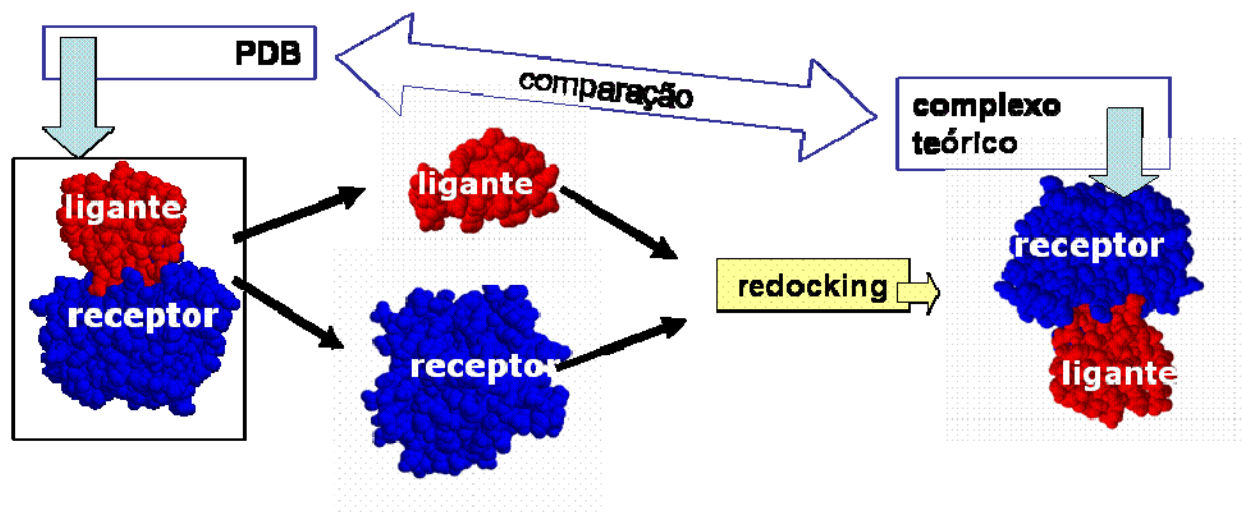


Figura 2.7 Redocking. Inicia-se com um complexo cristalográfico. Com as estruturas tridimensionais das moléculas isoladas, reconstrói-se o complexo virtual e procede-se à comparação com o complexo cristalográfico.

Depois de obtido o complexo virtual faz-se comparações com a molécula cristalográfica inicial. Se os resultados são compatíveis, pode-se considerar que a metodologia está validada. Caso contrário, são necessárias modificações, como, por exemplo, modificar parâmetros utilizados no programa, escolher outro centro do sítio receptor. As modificações são feitas, os cálculos de docking refeitos, até que se

obtenha um resultado compatível com o complexo original. Isto significa então que a metodologia utilizada está validada e pode-se iniciar os estudos de docking com a molécula receptora e os ligantes que se deseja.

2.3.3. Cross-docking

No experimento de cross-docking, toma-se o complexo A (receptor A – ligante A) e o complexo B (receptor B – Ligante B) e separam-se receptores de ligantes (Figura 2.8). Neste caso, os receptores devem ser a mesma molécula, mas podem apresentar diferente estado conformacional, uma vez que no complexo cristalográfico formado, o receptor pode sofrer mudanças para acomodar um ligante determinado. Toma-se, por exemplo, o ligante A e realiza-se o experimento de docking utilizando o receptor B. Experimentos como esse permitem analisar se determinadas características do receptor e/ou ligante o tornam específico ou não.

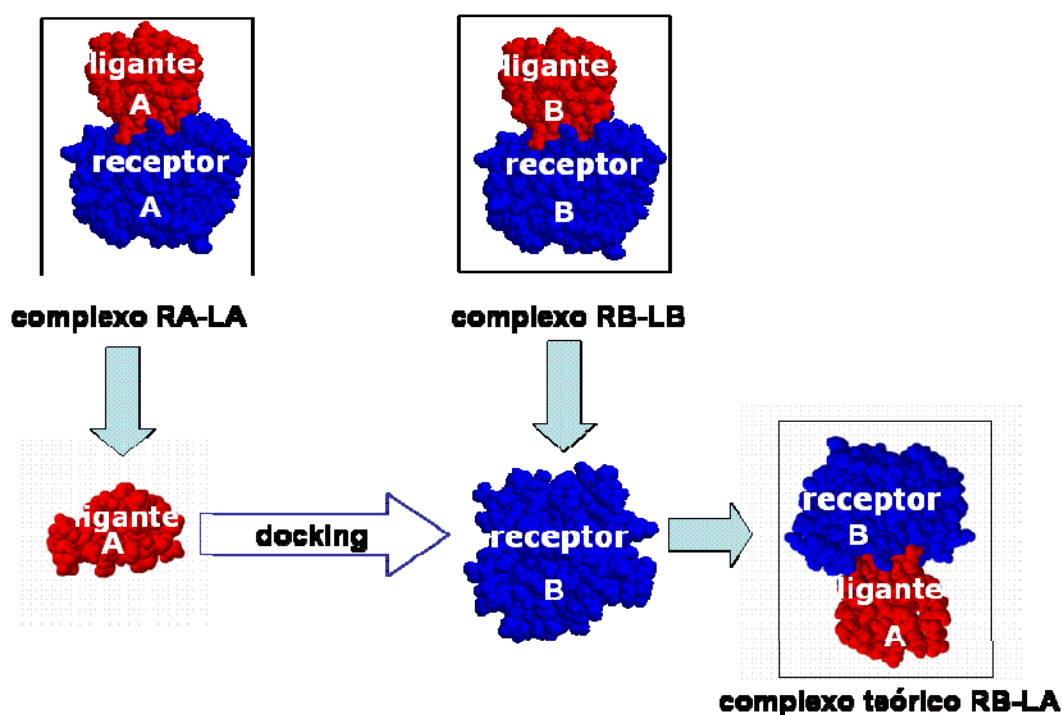


Figura 2.8 Cross-docking

2.3.4. O Método Computacional

Os estudos de docking foram realizados utilizando GOLD 3.1.1 (Genetic Optimization for Ligand Docking) (JONES, WILLETT, GLEN, 1995, JONES *et al.*,

1997). O programa GOLD utiliza um algoritmo genético (Seção 1.8) na busca de uma população de possíveis soluções utilizando operadores genéticos (mutações, *crossovers* e migrações) para obter uma população final, trabalhando com a otimização de uma função *Fitness* pré-definida, o GoldScore ou o ChemScore. O uso desta função *Fitness* permite que o docking seja realizado com flexibilização dos ligantes e das hidroxilas da macromolécula. Desta forma, o programa GOLD opera com um método de ajuste do ligante ao sítio, considerando os aspectos conformacional e de energia do ligante e da macromolécula. A função pré-definida compreende quatro componentes (JONES, WILLETT, GLEN, 1995, JONES *et al.*, 1997, VERDONK *et al.*, 2003):

- (a) energia de ligação de hidrogênio do complexo receptor-ligante;
- (b) energia de ligação de van der Waals;
- (c) energia de ligação de hidrogênio intramolecular do ligante;
- (d) energia de van der Waals interna do ligante

O programa realiza uma seleção interna dos resultados do docking com base na função *Fitness* e no escore escolhido (GoldScore ou ChemScore). Os cálculos de docking em geral estão planejados para obter 10 saídas em cada etapa de cálculo. Porém pode ser alterado para obtenção de mais ou menos saídas. Quando fornece menos saídas que o solicitado, é porque o valor de RMSD – o desvio médio quadrático, do inglês *root mean square deviation*, entre os ligantes é muito pequeno, de forma que o programa considera que as saídas são iguais.

A função escore dentro do programa, opera como apresentado na Figura 2.9. A função escore adotada permite que o processo possa ser interrompido e modificado durante o docking. Se algum usuário quiser, também pode modificar a função escore, construindo novas versões de `libfitfunc_dll.so` (GOLD SCORING-FUNCTION A.P.I. DOCUMENTATION).

(b) ChemScore, que é uma função empírica para estimar a energia livre de ligação do ligante à proteína. Usa termos de contato simples para estimar as contribuições lipofílicas e contribuições metal-ligante, incluindo interações de hidrogênio. Esse escore não faz diferenciação entre os diversos tipos de ligações de hidrogênio e a natureza geométrica da interação e ainda adiciona uma penalidade quando os contatos são muito próximos. ChemScore estima a variação da energia livre que ocorre no ligante quando forma o complexo com a macromolécula, conforme observado na equação abaixo:

$$\Delta G_{\text{binding}} = \Delta G_o + \Delta G_{\text{hbond}} + \Delta G_{\text{metal}} + \Delta G_{\text{lipo}} + \Delta G_{\text{rot}} \dots \quad (2)$$

onde cada componente dessa equação é o produto de um termo dependente da magnitude de uma contribuição física P para a energia livre (por exemplo, ligação de hidrogênio) e um fator de escala determinado por regressão. Por exemplo, $\Delta G_{\text{hbond}} = v P_{\text{hbond}}$, sendo v o fator de escala e P_{hbond} o parâmetro físico. O escore final (Equação 3) é obtido considerando também como penalidades os parâmetros de choques (*clash*) e torções internos, que não favorecem os contatos para o docking. Parâmetros para ligações covalentes também podem ser considerados.

$$\text{ChemScore} = \Delta G_{\text{binding}} + P_{\text{clash}} + C_{\text{internal}} P_{\text{internal}} + (C_{\text{covalent}} P_{\text{covalent}} + P_{\text{constraint}}) \quad (3)$$

Neste trabalho, foi utilizado apenas o GoldScore.

2.4. VISUALIZAÇÃO GRÁFICA

Utiliza-se a visualização gráfica em cada passo do trabalho. Na escolha do receptor, na análise das estruturas cristalográficas envolvidas no trabalho (complexos, receptores e ligantes), na escolha do sítio receptor e na análise dos resultados dos cálculos de docking.

A parte inicial de visualização permitirá que se tome decisões importantes e podem definir ou não o sucesso do experimento *in silico*. Para isso é necessário acostumar-se com os programas gráficos, quais as suas possibilidades e ferramentas. Essa parte inicial pode tomar até cerca de 30% do tempo do projeto como um todo. Normalmente seleciona-se qual a molécula será tomada como

receptor a partir do uso dos bancos de dados (PDB, PDBSum, CSD, NDB, entre outros) descritos na seção 1.6.

Quando já se tomaram decisões acerca do receptor, e dos ligantes, inicia-se o experimento de docking. O primeiro passo, redocking, tem que ter seus resultados cuidadosamente analisados. Quando o redocking permite determinar a validade dos procedimentos e escolhas realizados, passa-se a fazer o chamado “virtual screening”, onde são analisados mais ligantes com o receptor validado no redocking.

Os programas de docking geram muitas saídas. Cabe ao pesquisador através da investigação visual, determinar a validade ou não dos resultados. Em geral, o primeiro passo é o de verificar quantas saídas de um mesmo cálculo são iguais ou diferentes quanto à sua orientação e/ou conformação, separar em grupos por similaridades de orientação, analisar as interações receptor-ligante, verificar quais são as interações, se são desejáveis ou não. A partir disso cabe ao pesquisador decidir qual será a orientação escolhida. Esta parte do trabalho toma cerca de 60% do tempo como um todo.

Os programas gráficos são muitos, alguns são específicos para macromoléculas (proteínas, DNA) outros para moléculas pequenas e outros servem a ambos. Cada um dispõe de ferramentas específicas, mas todos permitem a visualização gráfica em diversas formas, movimentar a molécula (rotação, translação, zoom), alguns permitem modificar ligações (quebrar, ligar), tomar medidas de distâncias e ângulos, determinar os átomos envolvidos nas diferentes ligações e interações, medir os parâmetros das ligações de hidrogênio, bem como calcular as coordenadas dos centróides dos anéis aromáticos e cálculos de RMSD. Alguns deles também permitem salvar as figuras de moléculas apresentadas neste trabalho.

Os programas gráficos utilizados para sistema Windows, todos tipo software livre, foram:

programa “O” (JONES E KJELDGAARD, JONES, 1982);
DS VISUALIZER™;

VEGA ZZ 2.1.0 (PEDRETTI, VILLA, VISTOLI, 2002, 2003, 2004);

MERCURY CSD 1.4.2 - Department of Chemistry of Cambridge University;

Qmol (GANS, SHALLOWAY, 2001).

2.5. EXPERIMENTO *IN SILICO* × DADOS EXPERIMENTAIS

Na realidade, um trabalho totalmente teórico é possível, mas não teria sentido se não pudesse ser comparado com dados experimentais. Em alguns casos é possível que seja feita associação entre grupo de pesquisa teórico – grupo de pesquisa experimental, mas isso nem sempre é possível, ou o que se pretende.

Na indústria de fármacos, a parte experimental pode ser abreviada pelo trabalho teórico.

Hoje, com o Portal de Periódicos da CAPES, e acesso a informações das mais diversas fontes e áreas, pode-se comparar, confrontar os resultados teóricos com os experimentais, permitindo que se valide totalmente o experimento *in silico*, e fazendo com que se torne um modelo para planejamento de futuros experimentos *in silico*, *in vitro*, *in vivo*.

3 RESULTADOS e DISCUSSÃO

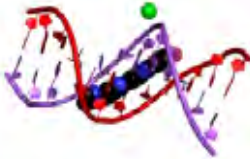
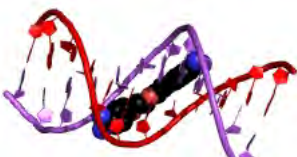
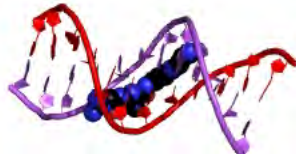
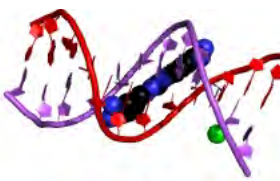
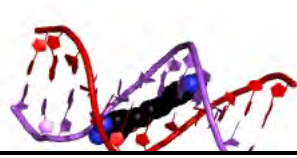
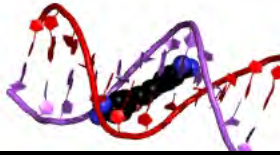
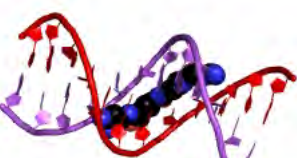
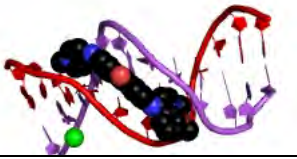
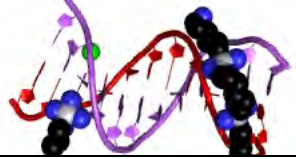
3.1. LEVANTAMENTO NO PDB DAS ESTRUTURAS TRIDIMENSIONAIS DE DNA 5'-CGCGAATTCGCG-3' COMPLEXADAS COM MOLÉCULAS PEQUENAS

A Tabela 3.1 contém informações a respeito de 25 estruturas de DNA 5'-CGCGAATTCGCG-3' complexadas com moléculas pequenas. Nesta tabela são apresentados o código pdb de cada complexo B-DNA-ligante, o nome do ligante, uma figura mostrando o complexo (PDBSum), comentários acerca do dado experimental obtido e a respectiva referência sobre a estrutura tridimensional.

Em destaque na Tabela 3.1, estão as figuras dos complexos que mostram as orientações dos ligantes em relação ao DNA. São 25 estruturas tridimensionais de complexos DNA-moléculas pequenas, e todos se ligam no sulco menor, com uma exceção, de código pdb 2dyw. Todos os complexos são formados por ligações não-covalentes.

Na Tabela 3.2 são apresentados os ligantes relacionados às 25 estruturas dos complexos de DNA 5'-CGCGAATTCGCG-3' contidos na Tabela 3.1.

Tabela 3. 1 Levantamento de complexos de DNA CGCGAATTCGCG e pequenos ligantes

109d	 Número de anéis: 6 (sendo 3 aromáticos) Comprimento: 18,65 Å Bases em contato com o ligante: A18 T19 T20 C21 – A5 A6 T7 T8 C9 (WOOD <i>et al.</i>, 1995)	1fms	 Número de anéis: 5, sendo 1 aromático Comprimento: 22,97 Å Bases em contato com o ligante: A6 T7 T8 C9 – A18 T19 T20 C21 (SIMPSON <i>et al.</i>, 2000)
127d	 Número de anéis: 6 (sendo 3 aromáticos) Comprimento: 19,05 Å Bases em contato com o ligante: A18 T19 T20 C21 – A6 T7 T8 (SRIRAM <i>et al.</i>, 1992)	1ftd	 Número de anéis: 6, sendo 4 aromáticos Comprimento: 25,14 Å Bases em contato com o ligante: A5 A6 T7 T8 C9 G10 – G16 A17 A18 T19 T20 C21 G22 (MANN <i>et al.</i>, 2001)
166d	 Número de anéis: 2 aromáticos Comprimento: 18,98 Å Bases em contato com o ligante: A18 T19 T20 C21 – A5 A6 T7 T8 C9 (NUNN <i>et al.</i>, 1994)	1ley	 Número de anéis: aromáticos Comprimento: 19,75 Å Bases em contato com o ligante: A6 T7 T8 C9 – A17 A18 T19 T20 C21 (GOODSELL <i>et al.</i>, 1995)
1d30	 Número de anéis: 3, sendo 2 aromáticos Comprimento: 13,12 Å Bases em contato com o ligante: A17 A18 T19 T20 C21 – A6 T7 T8 C9 (LARSEN <i>et al.</i>, 1989)	1m6f	 Número de anéis: 2 aromáticos Comprimento: 18,68 Å Bases em contato com o ligante: A6 T7 T8 C9 – A18 T19 T20 C21 G22 (NGUYEN <i>et al.</i>, 2002)
1d64	 Número de anéis: 2 aromáticos Comprimento: 18,20 Å Bases em contato com o ligante: A17 A18 T19 T20 C21 – A5 A6 T7 T8 C9 (EDWARDS <i>et al.</i>, 1992)	1prp	 Número de anéis: 2 anéis aromáticos Comprimento: 16,97 Å Bases em contato com o ligante: A6 T7 T8 C9 – A17 A18 T19 T20 C21 (NGUYEN <i>et al.</i>, 2002)
1d86	 Número de anéis: 2 anéis de 5 átomos. Comprimento: 18,79 Å Bases em contato com o ligante: G4 A5 A6 T7 T8 C9 – A17 A18 T19 T20 C21G22. (SRIRAM <i>et al.</i>, 1992)	1vzk	 Número de anéis: 4, sendo 2 aromáticos Comprimento: 16,83 Å Bases em contato com o ligante: A6 T7 T8 C9 – A17 A18 T19 T20 C21 (MALLENA <i>et al.</i>, 2004)
1zpi	 Número de anéis: 5 aromáticos Comprimento: 21,64 Å Bases em contato com o ligante: G4 C5 A6 T7 T8 C9 – A18 T19 T20 C21 G22 (ADAMS <i>et al.</i>, 2005)	2dyw	 Número de anéis: nenhum Comprimento: 40,17 Å Bases em contato com o ligante: C1 G2 C3 G4 A5 A6 (KOMEDA <i>et al.</i>, 2006)

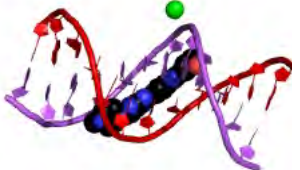
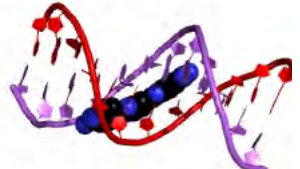
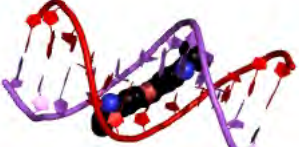
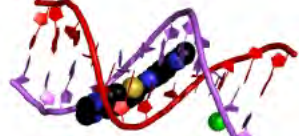
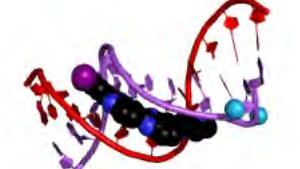
227d	 Número de anéis: 3, sendo 2 aromáticos Comprimento: 14,71 Å Bases em contato com o ligante: A6 T7 T8 C9 – A18 T19 T20 C21 (LAUGHTON <i>et al.</i>, 1996)	302d	 Número de anéis: 6, sendo 3 aromáticos Comprimento: 19,22 Å Bases em contato com o ligante: A5 A6 T7 T8 C9 – A18 T19 T20 C21 (CLARK <i>et al.</i>, 1996)
289d	 Número de anéis: 3, sendo 2 aromáticos. Comprimento: 19,12 Å Bases em contato com o ligante: A6 T7 T8 C9 G10 – A18 T19 T20 (TRENT <i>et al.</i>, 1996)	311d	 Número de anéis: 5, sendo 3 aromáticos Comprimento: 18,66 Å Bases em contato com o ligante: A6 T7 T8 C9 G10 – A17 A18 T19 T20 C21 (CLARK <i>et al.</i>, 1997)
298d	 Número de anéis: 3, sendo 2 aromáticos. Comprimento: 18,25 Å Bases em contato com o ligante: A6 T7 T8 C9 – A17 A18 T19 T20 C21 (TRENT <i>et al.</i>, 1996)	328d	 Número de anéis: 5 anéis aromáticos Comprimento: 20,85 Å Bases em contato com o ligante: A5 A6 T7 T8 C9 – A17 A18 T19 T20 C21 (SQUIRE <i>et al.</i>, 1997)
2b3e	 Número de anéis: 6, sendo 2 aromáticos. Comprimento: 19,45 Å Bases em contato com o ligante: T7 T8 C9 – A17 A18 T19 T20 C21 (CAMPBELL <i>et al.</i>, 2006)	360d	 Número de anéis: 3, sendo 2 aromáticos Comprimento: 19,23 Å Bases em contato com o ligante: A6 T7 T8 C9 – A17 A18 T19 T20 C21 (GUERRI <i>et al.</i>, 1998)
2dbe	 Número de anéis: 2 aromáticos Bases em contato com o ligante: A6 T7 T8 C9 – A18 T19 T20 (BROWN <i>et al.</i>, 1990)	442D	 Número de anéis: 6, sendo 4 aromáticos Comprimento: 20,64 Å Bases em contato com o ligante: A5 A6 T7 T8 C9 – A18 T19 T20 C21 G22 (SQUIRE <i>et al.</i>, 2000)
8bna	 Número de anéis: 6, sendo 3 aromáticos Comprimento: 20,12 Å Bases em contato com o ligante: G4 A5 A6 T7 T8 C9 – A17 A18 T19 T20 C21 (PJURA <i>et al.</i>, 1987)		

Tabela 3. 2 Ligantes dos complexos DNA-pequenas moléculas apresentados na Tabela 3.1

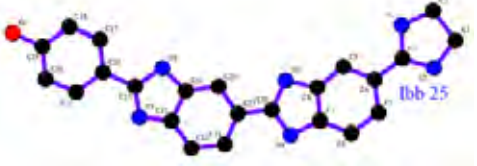
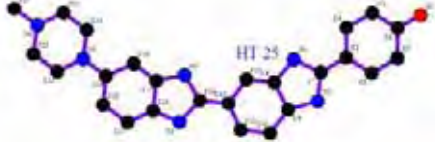
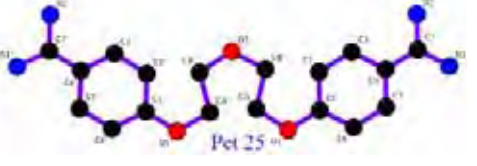
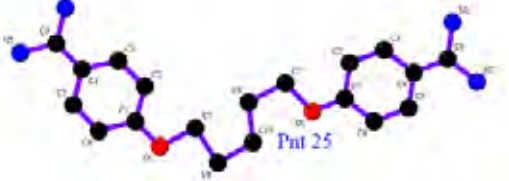
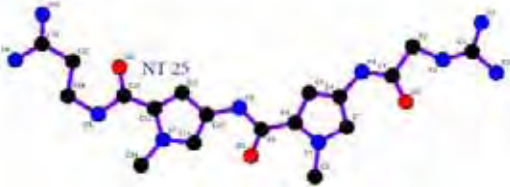
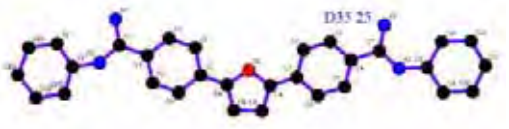
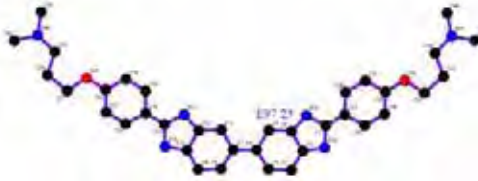
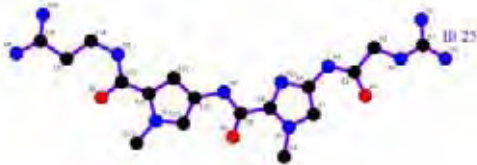
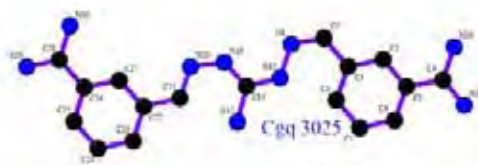
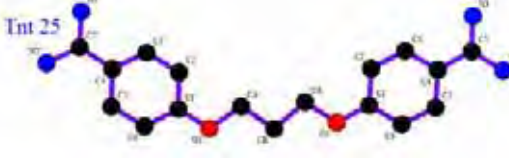
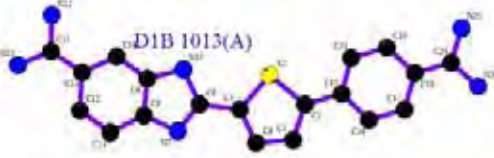
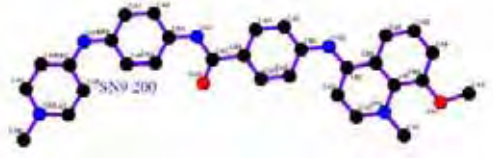
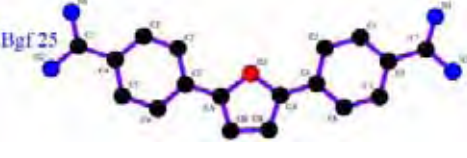
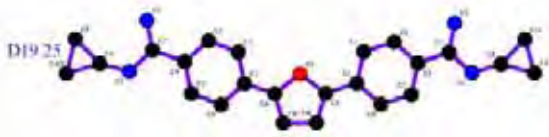
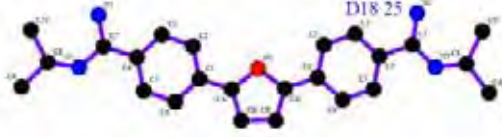
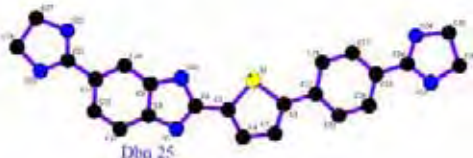
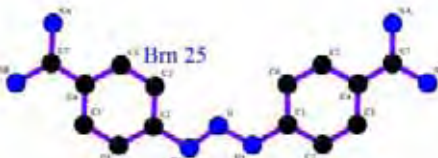
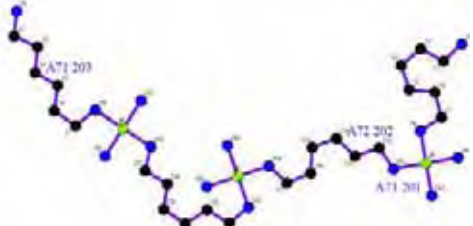
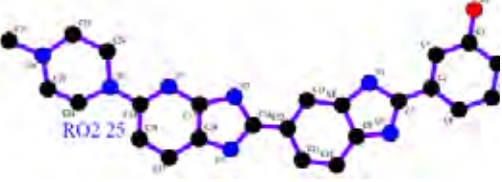
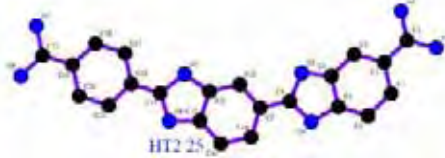
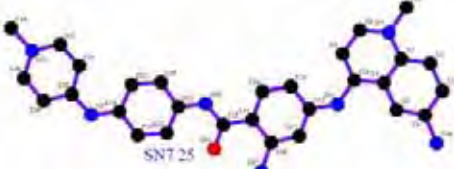
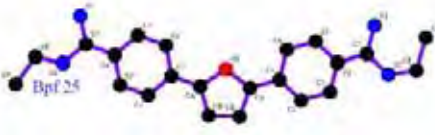
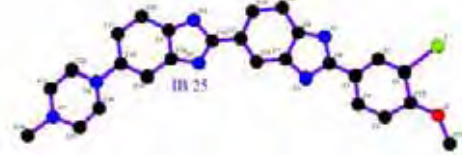
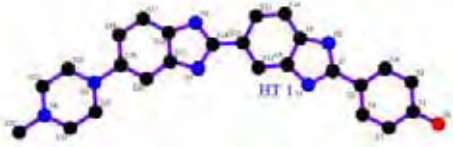
109d		127d	
166d		1d30	
1d64		1d86	
1fms		1ftd	
1ley		1m6f	
1prp		1vzk	
1zpi		227d	

Tabela 3.2 (continuação)

289d 	 298d
2b3e 	 2dbe
2dyw 	 302d
311d 	 328d
360d 	 442d
8bna 	

As moléculas dos ligantes apresentados na Tabela 3.2 contêm principalmente átomos de C, N, O e H exceto os ligantes pertencentes às estruturas de código pdb 1vzk e 2b3e com um átomo de S, a de 2dyw com 3 átomos de Pt e, finalmente, a 442d com um átomo de I.

Observando a geometria destas moléculas, percebem-se conformações similares, possivelmente por estarem acomodadas no sulco menor da mesma seqüência de DNA. A maior diferença é observada na molécula 2dyw, que é a única formada exclusivamente por uma cadeia alifática, enquanto que todas as outras moléculas apresentam pelo menos dois anéis não saturados.

As estruturas escolhidas para serem estudadas foram 1vzk e 2b3e, já que ambas possuem S na estrutura dos ligantes e os compostos estudados contêm crômio com coordenação similar à de S. Mais ainda, deve ser destacado que uma das moléculas de crômio possui uma forma geométrica similar às dos ligantes apresentados na Tabela 3.2 (DAZJOE) e a outra, uma forma diferente (CR3CN).

A escolha de 1vzk, conforme foi observado no decorrer do trabalho permitiu a realização de estudos de docking tanto no sulco menor como no sulco maior. Um levantamento similar ao realizado neste trabalho, foi descrito no trabalho de Evans e Neidle (2006), em que analisam resultados de docking somente no sulco menor, testando outros programas computacionais (Dock e AutoDock) e analisaram 28 complexos DNA-pequenas moléculas selecionados do PDB, verificando que a estrutura duplex de DNA do complexo 1vzk foi a mais representativa do conjunto de 28 receptores.

3.2. ANÁLISE DAS ESTRUTURAS CRISTALOGRAFICAS DOS COMPLEXOS 1vzk E 2b3e EXTRAÍDAS DO PDB

As duas estruturas estudadas são complexos constituídos por B-DNA de duas cadeias d(C-G-C-G-A-A-T-T-C-G-C-G)₂. Na Figura 3.1 observa-se o pareamento das bases desta seqüência, a mesma para 1vzk e 2b3e.

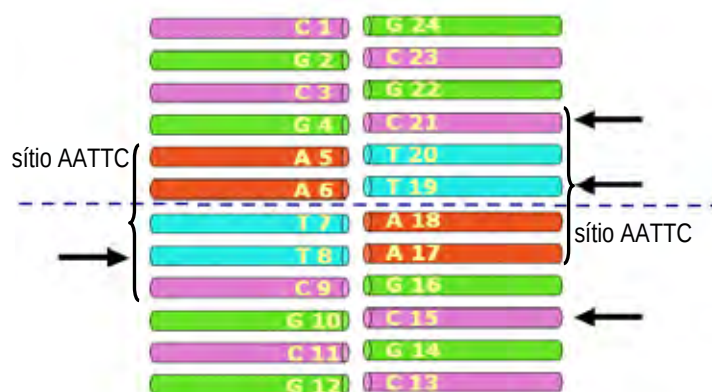


Figura 3. 1 O dodecâmero d(C-G-C-G-A-A-T-T-C-G-C-G)₂ das estruturas 1vzk e 2b3e. Os centros dos cálculos são mostrados com as setas indicativas.

As representações das estruturas obtidas do PDB 1vzk e 2b3e são apresentadas na Figura 3.2. As duas foram obtidas por difração de raio X. A primeira foi selecionada com base no levantamento feito e descrito na Seção 3.1, e a outra para fins de comparação. A estrutura 1vzk (MALLENA *et al.*, 2004) foi utilizada nos estudos de redocking e docking. A estrutura 2b3e (CAMPBELL *et al.*, 2006) foi utilizada nos estudos de cross-docking. As duas estruturas cristalográficas apresentam inúmeras semelhanças. Em ambos os casos os ligantes estão dentro do sulco menor junto à sequência AATTC (no caso A17 até C21). A mesma sequência aparece repetida na outra cadeia (A5 a C9), como mostrado na Figura 3.1. A nomenclatura conferida aos compostos aparece no trabalho de Campbell *et al.* (2006).

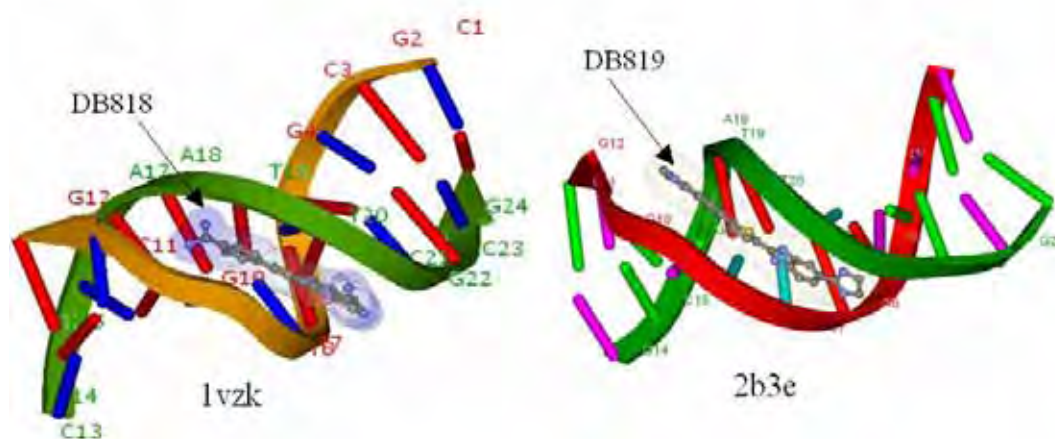


Figura 3. 2 Estruturas tridimensionais 1vzk e 2b3e, com seus respectivos ligantes DB818 e DB819, dentro do sulco menor.

Na Figura 3.3, observa-se o mesmo tipo de representação, mas observando-se o ligante a partir de uma perspectiva diferente.

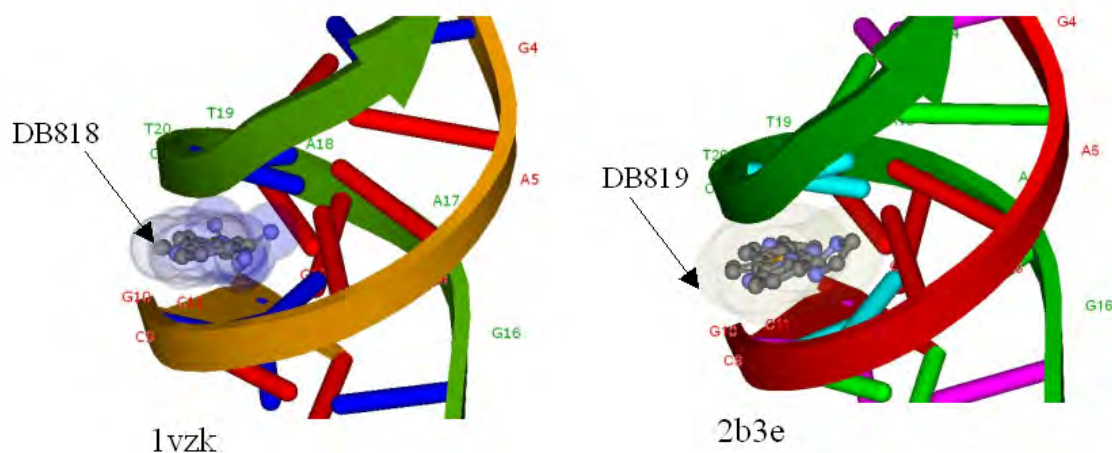


Figura 3. 3 Estruturas tridimensionais 1vzk e 2b3e, e seus respectivos ligantes DB818 e DB819 dentro do sulco menor.

Um ponto importante no que tange à formação do complexo é analisar as interações receptor-ligante. Na Figura 3.4 são apresentadas as ligações de hidrogênio entre o ligante DB818 e o receptor R-1vzk. Há três ligações de hidrogênio:

átomo do ligante DB818	átomo do receptor 1vzk	distância (Å)
N22	O2/T20	2,93
N25	O2/C9	3,17
N10	O2/T19	2,68

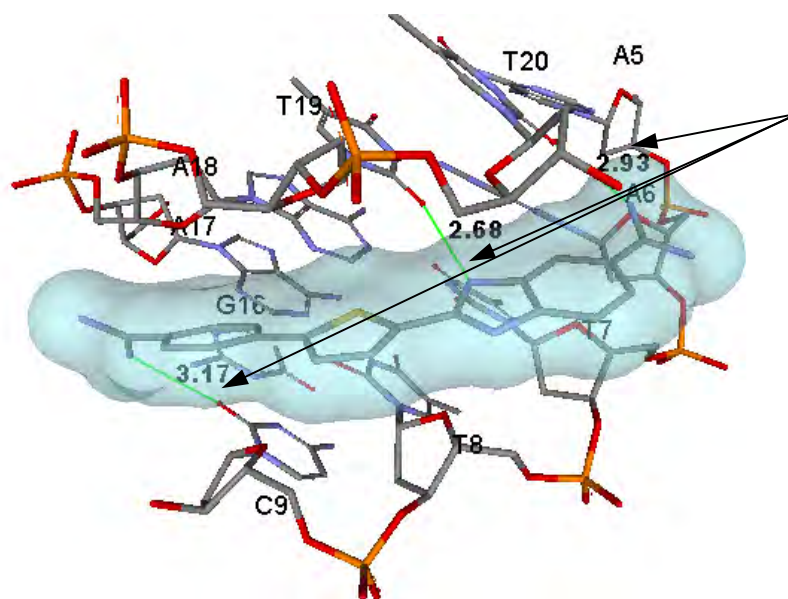


Figura 3. 4 O ligante DB818 dentro do sulco menor do receptor 1vzk, mostrando as ligações de hidrogênio.

Quando se faz a mesma análise para a estrutura cristalográfica 2b3e, obtém-se:

átomo do ligante DB819	átomo do receptor 2b3e	distância (Å)
N22	O2/T20	2,89
N10	O2/T19	2,83

e a Figura 3.5 mostra o posicionamento do ligante DB819 dentro do sulco.

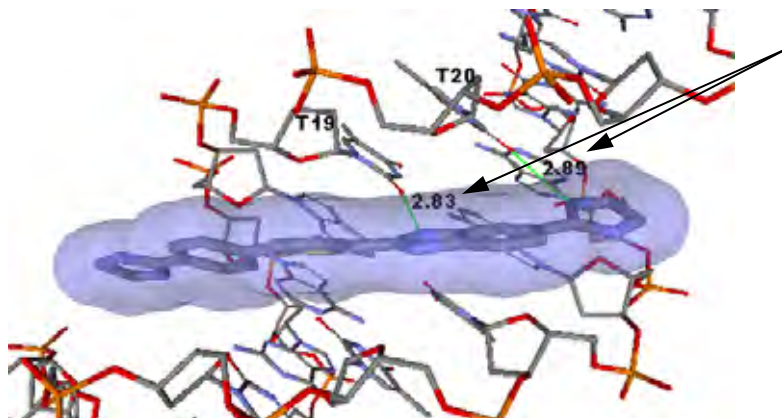


Figura 3. 5 O ligante DB819 dentro do sulco menor do receptor 2b3e, mostrando as ligações de hidrogênio.

Finalmente, ao se comparar os ligantes DB818 (1vzk) e DB819 (2b3e) observa-se que o ligante DB819 apresenta dois imidazóis em lugar dos grupos amidino em DB818. (Figura 3.6).

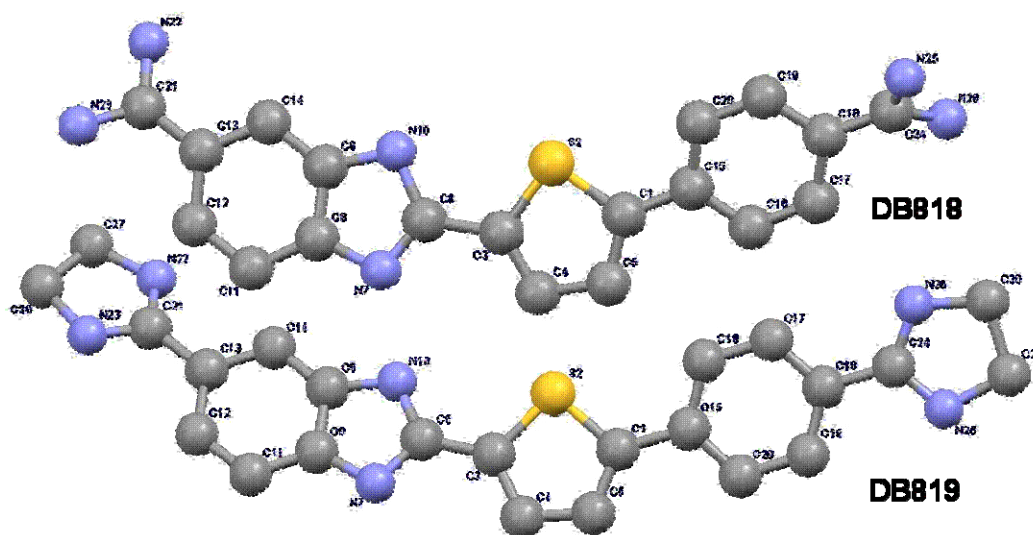


Figura 3. 6 Os ligantes DB818 e DB819.

3.3. ESTUDOS DO MODO DE LIGAÇÃO NO SULCO MENOR DO DNA.

Neste ponto se inicia o relato dos resultados obtidos a partir dos cálculos de docking realizados utilizando o programa GOLD, adotando a opção GoldScore para a função Fitness. Em todos os passos seguiu-se os procedimentos de visualização gráfica para a tomada de decisões quanto à seleção da orientação e/ou conformação do ligante dentro do sítio receptor.

3.3.1. Resultados do redocking de DB818 no sulco menor de R-1vzk

Da análise das interações entre DB818 e DNA podem ser extraídas muitas informações úteis no processo de compreensão do modo de ligação dos compostos em estudo no DNA. Como visto na Seção 3.2, o composto DB818 liga-se ao sulco menor do DNA fazendo interações com uma sequência de pares de bases descrita como AATTC. Este sulco é um importante alvo para a atuação de enzimas de controle e transcrição do DNA e tem sido um alvo atrativo para o desenvolvimento de agentes sintéticos [EVANS, NEIDLE, 2006], na proposta de novas terapias [CHAIRES, 1998, MUNDE *et al.*, 2007] e em estudos que visam no reconhecimento seletivo do DNA [MUNDE *et al.*, 2007]. Como pode ser observado no levantamento realizado no PDB o sulco menor também é o principal alvo dos estudos cristalográficos disponíveis até esta data.

Primeiramente foi feito o *redocking* (Seção 2.3.2), cálculos de docking nos quais o alvo do receptor e o ligante usado derivam de resultados experimentais, no nosso caso, o mesmo complexo cristalográfico. A fundamental importância deste tipo de estudo é que eles permitem além dos estudos estruturais para a formação dos complexos e afinidade do ligante pelo sítio receptor, validar as opções de cálculo e a capacidade do programa usado de fornecer resultados realísticos e confiáveis. Como foi visto, a partir do levantamento descrito na Seção 3.1, a estrutura cristalográfica selecionada foi a 1vzk. Como descrito na Seção 3.2, na análise das estruturas cristalográficas, o complexo cristalográfico 1vzk é constituído de uma molécula de DNA (a qual será chamada de R-1vzk), e o ligante do complexo que passará a ser designado como DB818.

Na linguagem usada, *solução* representa cada estrutura do ligante que o cálculo de docking ajustou no sítio receptor, um dos possíveis complexos modelados e selecionados pelo programa GOLD segundo o valor de escore calculado. *Orientação* representa a posição no sítio que agrupa uma ou mais soluções com RMSD baixo e superposição gráfica que permitem afirmar que formam um *grupo* de soluções equivalentes. Um número alto de soluções em um grupo será considerado como tendo certo peso estatístico que favorece a orientação por eles definida, e uma delas será selecionada como representante da orientação. As duas primeiras definições referem-se aos complexos; o termo *conformação* fica reservado para a estrutura tridimensional adotada pelo ligante, independentemente de estar ligado ou não.

Após cada cálculo de docking, conforme o procedimento descrito no Capítulo 2 de Metodologia, foi determinado no input do programa para que fossem fornecidas 10 soluções. Foi usado um mecanismo de controle interno do programa de forma tal que se para 3 soluções de maior escore encontradas durante a otimização fosse detectado um RMSD menor que 1,5Å, automaticamente se encerraria o cálculo. Por esta razão alguns cálculos têm um número de soluções menor que 10. Os resultados dos cálculos serão apresentados em forma de tabelas de 5 colunas sendo elas:

- 1) Solução complexo modelado.
- 2) Fitness escore de ajuste calculado
- 3) S(hb_ext) escore das ligações de hidrogênio entre o ligante e o receptor
- 4) S(vdw_ext) escore das forças de interação de van der Waals entre o ligante e o receptor.
- 5) S(int) escore de penalidade devido à tensão interna do ligante ao se adaptar ao receptor.

Uma solução foi selecionada em cada cálculo como representante da orientação do seu grupo, sendo que para uma orientação fosse selecionada, avaliou-se suas interações com os nucleotídeos do DNA, seu escore de ajuste e o número de soluções obtidas com essa orientação. No caso do redocking, a seleção da orientação foi feita antes da comparação com o resultado com a estrutura cristalográfica.

Nos subitens seguintes serão apresentados os resultados de um estudo de varredura do sulco menor de R-1vzk usando para isto 4 pontos onde centrar a procura do sítio de ligação:

- (a) T19 (leia-se o átomo O2 da base nucleotídica timina 19)
- (b) C15 (leia-se o átomo O2 da base nucleotídica citosina C15)
- (c) T8 (leia-se o átomo O2 da base nucleotídica timina 8)
- (d) C21 (leia-se o átomo O2 da base nucleotídica citosina C21)

O O2, presente nas bases nucleotídicas pirimidínicas (C e T), foi o átomo escolhido porque ele tem um posicionamento central dentro do sulco menor. As bases escolhidas respondem a uma separação crescente do centro da seqüência simétrica que responde a: primeiro par de bases T19, segundo par T8, terceiro par C15, quarto par T21. Com esta escolha, de um lado e do outro do plano de simetria, ao usar um raio de busca de 25Å, todo o sulco menor foi explorado neste estudo.

Como em todos os casos o centro foi o átomo O2, quando estiver assinalado que o centro foi a base T19, isto significará, neste trabalho, que o centro foi em O2 de T19.

3.3.1.a. Redocking do ligante DB818 no receptor R-1vzk, centrado em T19, com raio 10Å

Foram obtidos três grupos de orientações:

grupo 1: soluções 1 e 4

grupo 2: soluções 2, 3, 5, 6, 7, 8 e 9

grupo 3: somente a solução 10

Na Figura 3.7 é apresentada a superposição, no sítio receptor, da orientação da solução selecionada do grupo 1 com a do grupo 2; o grupo 3 não é mostrado, pois, sendo os átomos da extremidade flexível (C21 e próximos) perpendiculares aos grupos 1 e 2, não é possível a visualização em uma figura bidimensional.

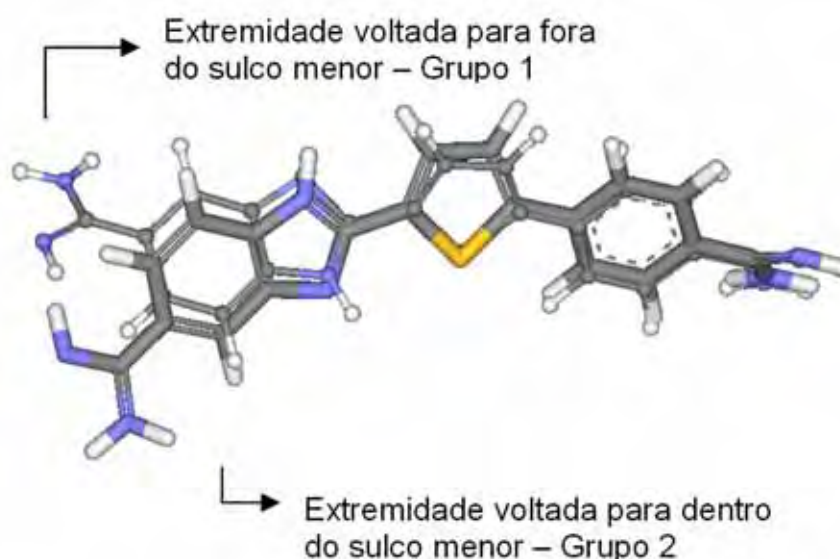


Figura 3. 7 Duas orientações representam os dois grupos obtidos no redocking em torno do centro O2/T19.

Por serem muito pequenos e de escores levemente mais baixos os grupos 1 e 3 foram descartados. A Tabela 3.3 mostra os resultados obtidos, destacando-se em negrito a solução escolhida como mais favorável do cálculo. Observa-se que justamente este grupo de maior número de soluções é aquele que tem uma superposição gráfica em toda a estrutura, ou seja, a mesma conformação.

Tabela 3. 3 Resultados do redocking para o cálculo com o centro T19, raio 10 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	78,89	10,43	51,82	-2,80
2	80,80	3,48	57,32	-1,48
3	78,99	4,00	55,51	-1,34
4	79,37	10,35	51,48	-1,76
5	80,23	4,00	56,73	-1,78
6	80,53	4,00	56,59	-1,28
7	79,92	4,00	55,97	-1,03
8	80,34	3,82	56,90	-1,72
9	79,07	4,00	56,09	-2,05
10	77,55	7,85	51,95	-1,72

Este cálculo inicial, o primeiro dos trabalhos, foi realizado com uma versão anterior do programa GOLD, diferentemente do resto do trabalho que foi realizado com a última versão disponível. Como poderá ser percebido logo adiante, a nova versão estabelece valores de penalidade de tensão interna do ligante (quinta coluna

das tabelas – S(int)) maiores que os obtidos com a versão anterior, ainda que sejam usados os mesmos *inputs* do cálculo inicial. O termo S(int) refere-se escore de penalidade devido à tensão interna do ligante ao se adaptar ao receptor. Os outros termos do escore não apresentam variação significativa entre ambas as versões do programa GOLD, permitindo assim recorrer a um fator de correção com o fim de homogeneizar os valores na Tabela 3.9 de análise geral.

Na Figura 3.8 é apresentada a superposição, no sulco menor, da solução selecionada com a estrutura cristalográfica do ligante. A coerência estrutural do resultado foi evidente e alentadora na continuação dos estudos.

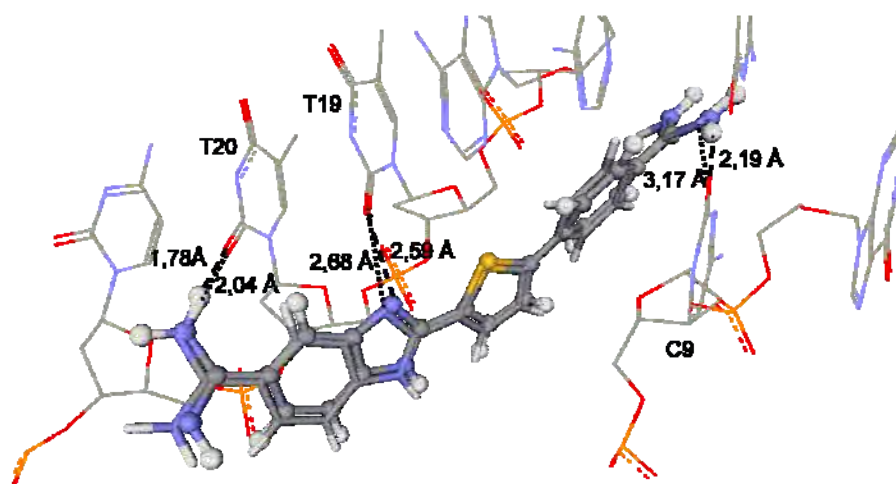


Figura 3. 8 Superposição da estrutura cristalográfica do complexo DNA-DB818 (*stick*) com a orientação selecionada no redocking (*ball-stick*). As interações análogas entre essas estruturas são apresentadas.

3.3.1.b. Redocking do ligante DB818 no receptor R-1vzk, centrado em T19, com raio 25Å

Para verificar se o resultado anterior não representava um mínimo local, mas sim um mínimo global de energia, o raio para exploração das orientações foi aumentado para 25 Å. Os resultados obtidos são apresentados na Tabela 3.4.

Tabela 3. 4 Resultados do redocking para o cálculo com centro T19, raio 25 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	74,88	3,98	56,98	-7,45
2	72,06	5,89	52,85	-6,50
3	73,66	10,72	51,69	-8,13
4	74,69	3,87	56,33	-6,63
5	72,24	2,75	55,19	-6,40
6	71,82	1,95	55,56	-6,52
7	74,69	2,60	56,50	-5,60

Nesta etapa, as soluções foram bastante diferentes umas das outras, sendo que somente 1, 4 e 7 foram consideradas equivalentes. Estas três soluções são as de maior escore de Fitness, e com um RMSD entre elas menor de 1,5 Å, razão pela qual o cálculo forneceu só 7 soluções. A orientação deste grupo correspondeu à orientação cristalográfica.

Neste cálculo, repetição do anterior, usando os mesmos *inputs*, agora com raio 25 Å, e usando a nova versão do programa GOLD observa-se que os escores da quinta coluna, a correspondente à penalidade derivada da tensão interna do ligante, tem maior peso na equação de escore, sendo eles os responsáveis da mudança do escore de fitness total. Com fins de comparação, foi calculado o RMSD entre a solução selecionada no cálculo com a versão antiga e a solução selecionada com a nova versão do GOLD. O valor de RMSD obtido foi igual a 0,16Å, e apresentou uma ótima superposição gráfica entre eles. Considerando estes fatos pode-se então considerar estes dois resultados como equivalentes. Calculando a diferença entre os escores totais obtém-se um valor de 5,92 unidades, enquanto que a diferença entre os escores das penalidades de tensão interna é de 5,97 unidades. Desta forma, pode-se acrescentar ao primeiro cálculo um fator de correção estimado, igual ao valor médio arredondado, ou seja, 5,95 unidades. Sendo assim, o valor corrigido para o escore de penalidade fica em -7,43 fornecendo um escore total de 74,85.

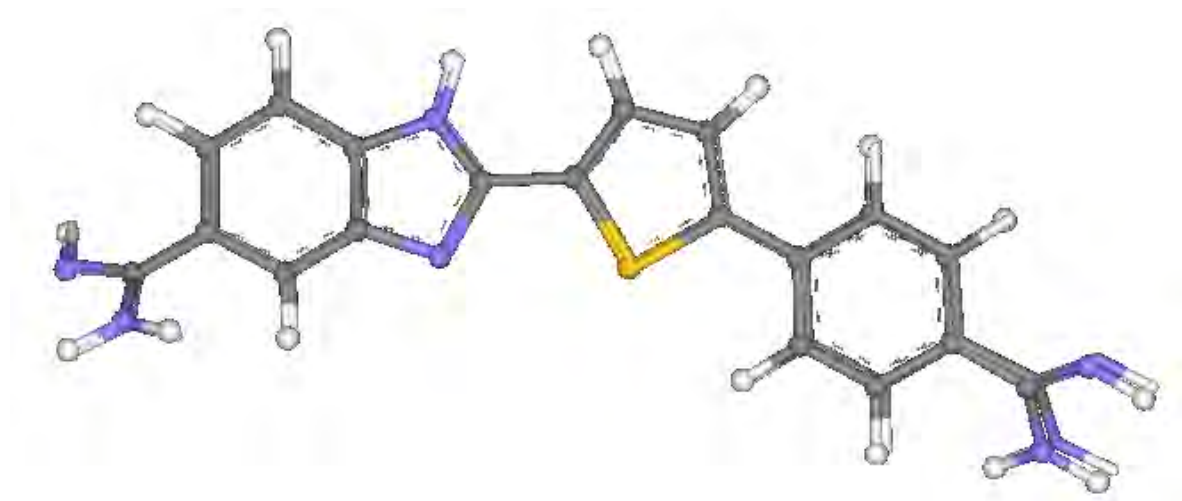


Figura 3. 9 Sobreposição da solução obtida em cálculos com a versão anterior do GOLD e a solução obtida utilizando-se a última versão. Um RMSD de 0,16 Å foi encontrado, evidenciando a compatibilidade das duas versões.

3.3.1.c. Redocking do ligante DB818 no receptor R-1vzk, centrado em C15, com raio 25Å

Os grupos identificados através de visualização gráfica foram:

grupo 1: soluções 1, 6 e 9;

grupo 2: soluções 3 e 7;

sem grupo: as demais soluções foram bastante diferentes umas das outras e não puderam ser agrupadas.

Os resultados são apresentados na Tabela 3.5.

Tabela 3. 5 Resultados do redocking para o cálculo com o centro C15, raio 25 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	73,90	3,70	55,51	-6,13
2	72,69	8,17	52,69	-7,94
3	75,05	3,54	57,30	-7,28
4	75,01	10,67	53,97	-9,87
5	74,63	11,52	53,00	-9,78
6	76,10	2,65	57,49	-5,60
7	75,17	4,00	56,65	-6,72
8	74,78	11,72	53,00	-9,82
9	74,64	3,66	56,07	-6,12
10	73,90	11,57	51,68	-8,36

A solução 6 possui o maior escore, porém a sua superposição gráfica com a estrutura cristalográfica não é boa. No entanto, observa-se que a solução 7 tem o

segundo maior escore e a solução 3 o terceiro, e ainda apresentam um maior numero de ligações de hidrogênio com o DNA, isto é, mais orientada. Ao observar a conformação tridimensional de 7, percebe-se que a mesma é quase plana, enquanto a estrutura tridimensional da solução 6 contém um anel aromático perpendicular ao restante da molécula (Figura 3.10), deixando esse grupo um pouco fora do sulco menor. Sendo assim, decidiu-se, mesmo com um escore alto, considerar a mesma como um falso positivo.

Este cálculo teve C15 como ponto da varredura com raio de 25Å. Dessa forma constitui-se no centro mais distante do sítio cristalográfico (T19) e se encontra na região do sítio simétrico (T7). Neste cálculo, onde, ainda que com menor repetição e escores, foram encontradas duas soluções (3 e 7) superpostas com o ligante original no sítio cristalográfico.

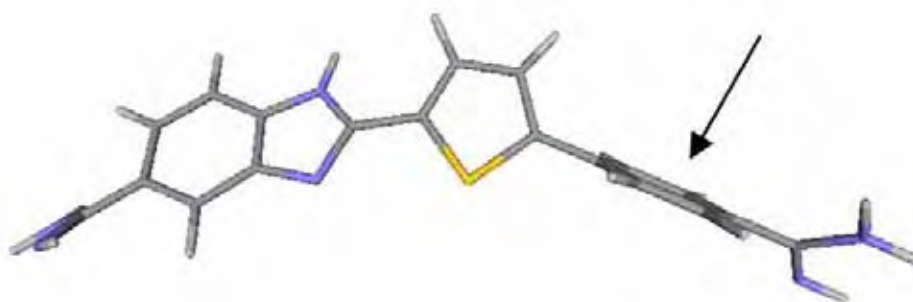


Figura 3. 10 Na orientação 6 o anel aromático é perpendicular ao restante da molécula, o que dificulta sua entrada no sulco.

3.3.1.d. Redocking do ligante DB818 no receptor R-1vzk, centrado em T8, com raio 25Å

Foram obtidos dois grupos de conformações:

grupo 1: soluções 1 e 10;
 grupo 2: soluções 2, 3, 5, 6, 7, e 9;
 sem grupo: as demais soluções foram bastante diferentes umas das outras e não puderam ser agrupadas.

Os resultados estão mostrados na Tabela 3.6. De maneira semelhante a quando os cálculos foram realizados em torno de T19, os dois grupos formados neste cálculo são semelhantes aos mostrados na Figura 3.7, sendo o grupo menor com o grupo amidino voltado para fora do sulco e o grupo maior com o amidino voltado para dentro.

Tabela 3. 6 Resultados do redocking para o cálculo com o centro T8, raio 25 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	74,72	10,72	53,79	-9,94
2	75,66	3,95	57,69	-7,61
3	74,92	3,90	56,28	-6,37
4	70,97	7,80	51,87	-8,16
5	74,71	3,71	57,27	-7,74
6	74,66	2,75	56,80	-6,19
7	74,45	3,31	57,62	-8,09
8	71,72	2,28	55,75	-7,22
9	73,80	3,06	56,70	-7,22
10	74,32	9,06	52,95	-7,54

Como aconteceu nos cálculos prévios a estrutura escolhida neste cálculo concordou com a estrutura cristalográfica.

3.3.1.e. Redocking do ligante DB818 no receptor R-1vzk, centrado em C21, com raio 25Å

Foram obtidos 3 grupos:

grupo 1: soluções 1, 6, 7 e 9;

grupo 2: soluções 2, 5 e 10;

grupo 3: soluções 3, 4 e 8.

Não foi dada preferência a nenhum grupo, visto que o número de moléculas em cada um não variava consideravelmente. Sendo assim, escolheu-se a orientação com melhor fitness, como mostra a Tabela 3.7.

Tabela 3. 7 Resultados do redocking para o cálculo com o centro C21, raio 25 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	71,91	7,10	52,60	-7,52
2	75,22	4,00	56,30	-6,22
3	73,95	10,19	53,08	-9,22
4	73,70	10,92	52,60	-9,53
5	74,65	3,09	57,33	-7,26
6	72,26	7,92	52,54	-7,90
7	71,95	7,51	54,00	-9,81
8	75,22	11,36	53,04	-9,07
9	72,45	7,93	52,85	-8,15
10	74,26	2,65	56,65	-6,29

A solução selecionada, como aconteceu nos cálculos anteriores também foi coerente com a estrutura cristalográfica. O fato de haver um menor número de soluções no grupo da orientação escolhida pode ser considerado como dependente da escolha de centrar a procura dos complexos, ainda que do mesmo lado da dupla hélice que o sítio cristalográfico, a uma distância de dois pares de bases do ponto mais central do ligante, ou seja, centrado em T19.

3.3.1.f. Redocking do modelo-2 de DB818 no receptor R-1vzk, centrado em T19, com raio 25Å

Um último cálculo de redocking foi realizado agora tentando observar se variam os resultados quando o ligante é considerado como duplamente protonado nas extremidades, ou seja, os 4 nitrogênios dos extremos na forma NH_2 e a carga (+) deslocalizada. O redocking do novo ligante, o Modelo-2 de DB818 (**M2**), foi também centrado no T19 e com um raio de 25Å.

Foram obtidos três grupos de orientações:

grupo 1: soluções 3 e 7;

grupo 2: soluções 2, 4, 5, 6, 8, 9 e 10;

grupo 3: solução 1.

A solução de número 8 (Figura 3.11) foi selecionada por possuir o maior escore do grupo majoritário e o segundo maior do total, conforme mostra a Tabela 3.8.

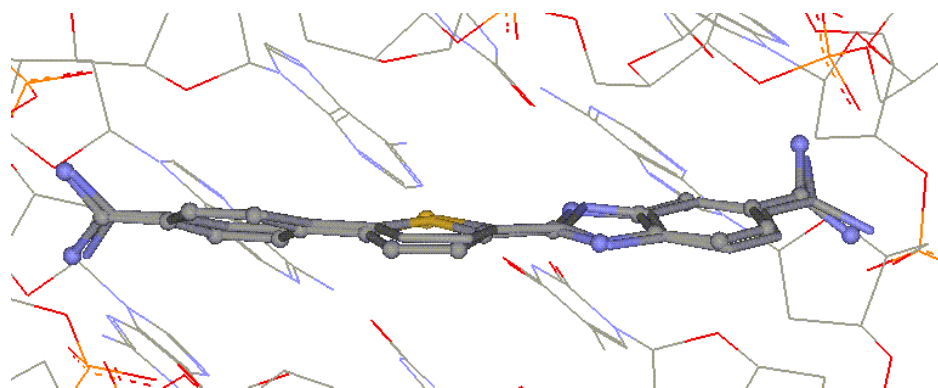


Figura 3. 11 Observa-se que o resultado do cálculo de redocking (*ball-stick*) está em concordância com a estrutura cristalográfica (*stick*). No entanto observa-se uma aproximação do N32 do O4 do açúcar.

As soluções 3 e 7 formam um grupo minoritário e estão deslocados do sulco menor, ainda que 3 tenha o maior escore; por isto foram considerados como falsos positivos.

Tabela 3. 8 Resultados do redocking para o cálculo com o centro C21 raio 25 Å (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Orientação	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	74,27	5,00	56,02	-7,75
2	74,96	4,19	56,47	-6,87
3	78,31	10,51	55,59	-8,65
4	76,34	4,08	58,02	-7,52
5	74,43	5,48	54,86	-6,48
6	76,34	4,42	57,20	-6,74
7	76,13	11,48	53,47	-8,87
8	76,51	3,81	57,89	-6,90
9	75,70	4,49	57,00	-7,71
10	74,96	3,14	57,81	-7,66

Não foram detectadas grandes mudanças nos resultados quando o ligante M2 se encontra duplamente protonado, Os escores obtidos, embora um pouco maiores do que os do caso sem a dupla protonação estão na faixa dos valores obtidos para o modelo inicial. O perfil das soluções obtidas e a facilidade/dificuldade de selecionar o complexo mais favorável parece ser o mesmo usando o ligante sem protonar ou duplamente protonado, dado que na equação de ajuste do GoldScore, as cargas parciais não são consideradas mas sim as ligações de hidrogênio. Em ambos casos deste estudo só é possível obter uma ligação de hidrogênio dado que o segundo hidrogênio fica na direção da saída do sulco menor e os resultados do cálculo mostram existir a ligação de hidrogênio em ambos modelos.

3.3.1.g. Comparação dos resultados de redocking

Com a finalidade de conseguir uma avaliação geral dos procedimentos e resultados obtidos no estudo de varredura do sulco menor de R-1vzk com o ligante cristalográfico DB818, neste item serão comparados paralelamente todos os cálculos de redocking realizados. Para fazer isto, para cada cálculo, foi escolhida uma solução como representante da orientação selecionada. Os critérios de seleção, explicados nos itens anteriores, acabaram confirmando como majoritária e detectável que a orientação obtida é coerente com a cristalográfica.

Na Tabela 3.9 estão compilados os seis resultados selecionados, com o número da solução, bem como seu escore de Fitness.

Tabela 3. 9 Solução selecionada dos cálculos de redocking e valor de escore de Fitness.

raio (Å)	solução	centro	Fitness
10	2	T19	74.85 ^a
25	1	T19	74,88
25	7	C15	75,17
25	2	T8	75,66
25	2	C21	75,22
25	8 ^b	T19 ^b	76,51

^a valor corrigido; ^b ligante **M2**.

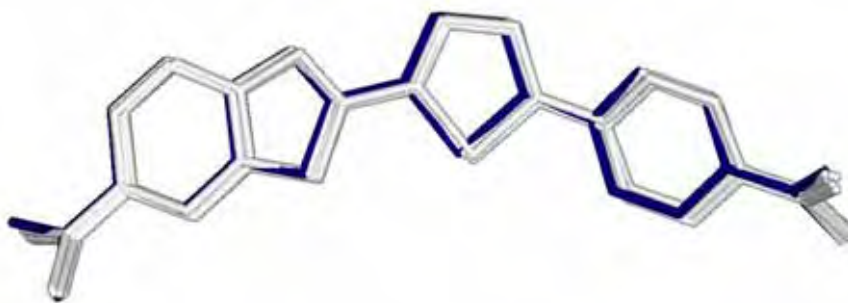


Figura 3. 12 Superposição da estrutura cristalográfica do ligante DB818 (em preto, ao fundo) com as 6 soluções selecionadas nos 6 cálculos de redocking (em branco).

Na Figura 3.12 é apresentada a superposição da estrutura cristalográfica do ligante DB818 com as 6 soluções selecionadas nos cálculos de redocking; como fica evidente a superposição foi, em todos os casos, muito boa. Pode-se então afirmar que, mesmo com alguma dificuldade de detecção nos cálculos em que o centro de procura do sítio foi mais distante da posição do ligante no complexo cristalográfico, a

consistência da metodologia para encontrar uma solução compatível e os critérios de seleção utilizados no trabalho de análise permitem fazer uma construção e seleção de complexos moleculares teóricos realísticos.

Os resultados do redocking deixam claro que há grande afinidade do ligante DB818 pela sequência AATTC. A estrutura tridimensional correspondente a esta sequência, está em uma conformação adaptada para receber o ligante, e seguramente será escolhida em cálculos em que serão usados como modelo a estrutura do DNA R-1vzk. Observa-se que, não importa de onde se iniciem os cálculos, a molécula sempre vai para esse sítio. Mesmo assim, os cálculos centrados em T19 são os de menor ambigüidade, e por esta razão este foi o ponto escolhido como centro do sítio de ligação nos demais cálculos.

Dos dois ligantes usados (DB818 e o modelo M2), M2 parece ter uma adaptação levemente melhor ao sítio e foi o escolhido para os estudos de cross-docking do item seguinte.

Deve ficar claro, também, que nem sempre o resultado apontado como o melhor pelo programa é realmente o que ocorre na natureza, por exemplo, no caso da simulação da Seção 3.3.1.c, onde a solução apontada pelo programa como de maior escore de Fitness era a 6 (deslocada em relação à cristalográfica), mesmo assim, a segunda de maior escore, a 7, era coerente com os outros resultados. É necessário lembrar que nesse caso, a busca foi feita no ponto mais distante do centro do ligante cristalográfico, e mesmo assim foi possível detectar a orientação de interesse. Nesta situação uma verificação da quantidade de ligações de hidrogênio foi fundamental.

Mais um comentário é necessário sobre um cálculo que foi executado para testar as capacidades do programa utilizado, agora usando a própria estrutura cristalográfica do ligante com um raio de 8Å para centrar a procura dos complexos. Neste caso, o mais direcionado, o resultado foi unânime à orientação cristalográfica, com somente 3 soluções com RMSD menor a 1.5Å (neste caso o cálculo pára ao cumprir as condições de controle interno da coerência de resultados).

Após o estudo completo de redocking e a análise dos resultados em seu

conjunto pode-se afirmar que a metodologia e critérios de seleção utilizados, são adequados e permitem a detecção de resultados realísticos e confiáveis. Estes resultados também representam uma base sólida para o restante do trabalho, onde, o local e a forma de se ligar ao DNA, não tem uma estrutura de referência, mas é possível contar com dados experimentais (não estruturais), indícios e/ou hipóteses de ligação que permitem avaliar e/ou propor complexos realísticos que permitam guiar futuros estudos experimentais e teóricos nesta área de pesquisa tão importante e ainda tão inexplorada.

3.3.2. Cross-docking

A idéia do cross-docking está descrita na Seção 2.3.3. Neste trabalho, o cross-docking foi realizado usando o sítio receptor da estrutura do complexo 2b3e com o ligante DB818 (o ligante cristalográfico de 2b3e, DB819, é um composto derivado de DB818, só que mais volumoso). Pelo fato da sequência de DNA ser simétrica, isto fornece duas variantes a mais sobre o sítio receptor.

Como forma de testar as capacidades de trabalho da metodologia escolhida, independentemente de usar um sítio e um ligante que já tem certo ajuste estrutural (pelo fato de derivarem de um mesmo complexo cristalográfico), nesta seção serão descritos os resultados de um estudo paralelo de cross-docking onde o ligante **M2** foi ligado à estrutura de DNA do complexo cristalográfico 2b3e (de sequência idêntica à anteriormente usada).

Por tratar-se de uma sequência de pares de bases simétrica, na verdade a estrutura apresenta dois sítios possíveis de sequência AATTC, uma em cada fita, A6-A7-T8-T9-C10 e A17-A18-T19-T20-C21. Usando dois pontos onde centrar a procura dos complexos, O2/T8 e seu simétrico O2/T20, consegue-se fazer um teste duplo em mais duas variantes da sequência onde o composto DB818 se liga.

3.3.2.a. Cross-docking do modelo M2 no receptor 2b3e centrado em T8 com raio 25 Å

Neste caso foram obtidos três grupos de conformações:

grupo 1: soluções 3 e 4;
 grupo 2: soluções 1, 2, 5, 7 e 8;
 grupo 3: solução 6, 9 e 10.

Sendo o grupo 2 o mais numeroso, selecionou-se deste grupo a solução 8, a que tem maior escore nesse grupo como mostra a Tabela 3.10. Esta orientação situa-se no sulco menor e faz interações com as bases AATT do sulco menor, mas levemente deslocada do sítio esperado. No grupo 1, mesmo sendo minoritário, ambas soluções são as que tem os segundo e terceiro maior escore de ajuste e elas são coerentes com a localização esperada, ou seja, a seqüência A₆A₇T₈T₉C₁₀, sendo destas duas a solução 4 a selecionada (Figura 3.13).

Tabela 3. 10 Resultados do cross-docking para o cálculo centrado em T8 (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	67,25	1,20	52,84	-6,60
2	67,70	0,99	53,75	-7,20
3	67,96	2,96	52,27	-6,87
4	68,41	1,91	52,32	-6,81
5	67,24	0,20	52,56	-6,60
6	67,66	6,06	49,24	-6,11
7	66,58	2,49	51,31	-6,46
8	68,55	2,95	52,31	-6,32
9	67,69	5,74	49,54	-6,17
10	67,38	2,52	52,02	-6,66

Se este fosse um caso no qual não se tem informação estrutural, o correto seria não descartar nenhuma das duas orientações em destaque e fazer um estudo mais aprofundado e um refinamento nos cálculos para assim tentar definir o sítio de maior possibilidade de ser o real. Segundo os critérios de seleção que foram definidos, nesta etapa da análise não há argumentos para fazer a seleção de um deles; o fato de se ter certeza sobre o ponto esperado, neste caso torna possível que se selecione como correto o grupo 1, minoritário, mas com os dois maiores escores.

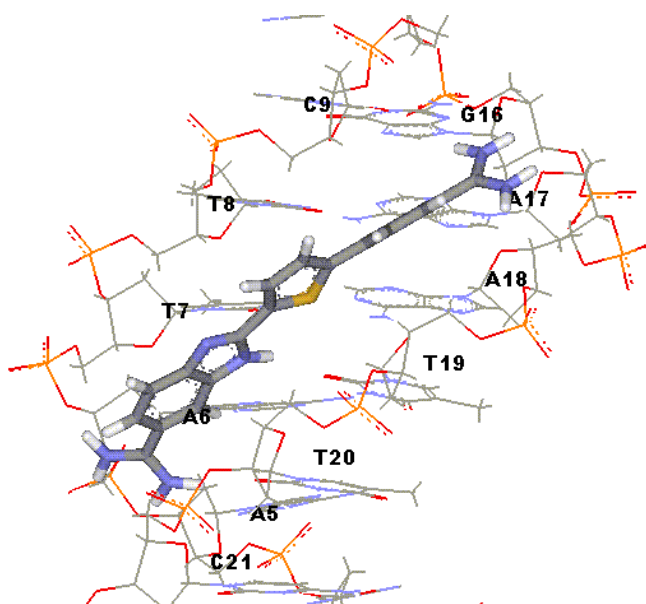


Figura 3. 13 Orientação 4 do ligante no sítio do receptor R-2b3e resultante do cross-docking no centro T8 em 2b3e. O ligante não fica totalmente plano, mas acomoda-se no sulco menor entre as bases AATTC.

3.3.2.b. Cross-docking do ligante modelo-M2 no receptor R-2b3e centrado em T20 com raio 25 Å

Essa nova série de cálculos foi efetuada para verificar a afinidade do ligante pelo sulco menor mesmo partindo-se de outro centro. Este outro centro é o equivalente ao primeiro, mas no sítio simétrico na outra cadeia de DNA. Foram obtidos dois grupos de soluções:

grupo 1: soluções 1, 4, 8 e 9;

grupo 2: soluções 2, 3, 5, 6, 7 e 10.

Em particular, neste cálculo a orientação 7 fica posicionada na seqüência A6-A7-T8-T9-C10 enquanto o resto do seu grupo fica na seqüência A17-A18-T19-T20-C21. Foi escolhida a orientação 7 por pertencer a um grupo maior e ter maior escore (Tabela 3.11). Esta estrutura ficou de acordo com a cristalográfica e superposta com as duas soluções do grupo 1 do cálculo centrado em O2/T8.

Tabela 3. 11 Resultados do cross-docking para o cálculo centrado em T20 (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	68,37	6,84	49,12	-6,00
2	65,93	2,53	50,33	-5,81
3	65,78	3,03	49,87	-5,83
4	65,74	5,01	50,50	-8,71
5	68,81	2,22	54,97	-9,00
6	69,13	3,59	52,28	-6,34
7	71,72	4,71	53,58	-6,66
8	67,93	5,35	49,99	-6,15
9	68,32	6,30	49,49	-6,04
10	65,37	3,57	49,19	-5,84

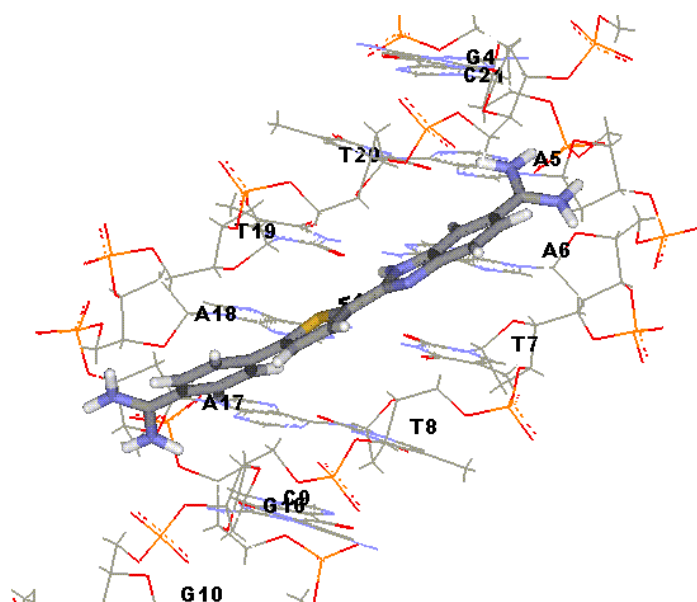


Figura 3. 14 Orientação 7 do ligante no sítio do receptor R-2b3e resultante do cross-docking no centro T20 em 2b3e. Melhor resultado comparando-se com a simulação em T8.

Considerando uma discussão geral dos cálculos de cross-docking pode-se dizer que mesmo tendo certa ambigüidade em um dos cálculos, o resultado “verdadeiro” foi modelado e detectado. Como mencionado anteriormente, em casos onde não há referência estrutural do complexo, seria muito útil realizar um refinamento da busca do sítio de ligação variando-se o raio como feito aqui, usando como referência as orientações selecionadas em primeira instância. Este procedimento permitiria definir qual das soluções seria a mais provável de acontecer na natureza.

3.3.3. Docking do ligante DAZJOE no receptor R-1vzk

O ligante DAZJOE é um ligante orgânico simétrico formado por 7 anéis, seis deles de 6 átomos (3 anéis a cada lado) e um de 5 átomos que une ambas as metades. Este composto tem no centro um átomo de Cr(III) que se coordena equatorialmente com dois O e dois N com uma estrutura de 7 anéis quase coplanares (Figura 2.4). Adicionalmente, no monocristal do ligante o átomo de Cr(III) está coordenado com duas águas cristalográficas. Vaidyanathan *et al.* (2005) estudaram este composto como possível ligante do DNA e seus resultados confirmam a ligação desta molécula ao DNA (em uma sequência inespecífica). O sítio molecular de ligação ainda não é conhecido e os autores apresentam indícios de que seria possível intercalar entre pares de bases. Frente a este panorama de incerteza, os trabalhos aqui realizados focaram-se em primeiramente, imaginar uma situação inicial na qual só são dois os possíveis sítios de ligação ao DNA, o sulco menor (**sm**) e o sulco maior (**SM**) dado que a intercalação só é possível em uma estrutura do DNA distorcida que tem entre dois pares de bases espaço suficiente para acolher uma outra molécula, em geral plana, como apresentado nas Tabelas 3.1 e 3.2.

Para chegar nesta situação qualquer ligante primeiramente deve se encaixar em um dos dois sulcos. Imaginando essa situação foi considerado que as águas cristalográficas coordenadas com o Cr(III) seguramente deveriam sair na hora da ligação. Por ser a forma global de DAZJOE quase planar e com uma leve curvatura foi considerado, a priori, e em função do observado no levantamento dos complexos cristalográficos disponíveis no PDB, como mais provável local de ligação o sulco menor **sm**. Este é o sítio que é observado como amplamente favorável para estruturas moleculares destas características.

Desta forma foram realizados cálculos de docking da DAZJOE nos centros T8 e T19, usando também um raio de 25Å, o que garante que mesmo centrado no **sm** também o sulco maior será explorado nestes estudos.

3.3.3.a. Docking do ligante DAZJOE no receptor R-1vzk centrado em T8 com raio 25 Å

Os resultados do docking forneceram um resultado unânime de somente três soluções, com um RMSD menor a 1,5Å. Os resultados são mostrados na Tabela 3.12. A solução selecionada foi a 3, devido seu maior escore.

Tabela 3. 12 Resultados de docking do ligante DAZJOE no receptor R-1vzk centrado em T8 (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	66,96	0,00	48,70	0,00
2	66,57	0,00	48,41	0,00
3	67,64	0,00	49,20	0,00

Todas as soluções situam-se no sulco menor, na região da sequência ATTC. Observa-se na Tabela 3.12 que as forças de van der Waals tem maior peso na formação dos complexos, determinando a escolha do sulco menor. Observar que a Tabela 3.12 apresenta valores de S(vdw_ext) - coluna 4 - diferentes dos valores da coluna 1, porque os valores da coluna 4 são multiplicados pelo valor empírico 1,375 que aparece na Função Fitness. Por tratar-se de um ligante sem rotação possível não tem penalidade de tensão interna. Pode-se dizer que, neste caso, o cálculo é de docking rígido. Adicionalmente, por não haver substituintes na estrutura DAZJOE, também não aparece valor de escore de ligações de hidrogênio, responsáveis por orientar e estabilizar os complexos moleculares. Desta forma, a ligação é não-específica, dirigida pelas forças de van der Waals.

3.3.3.b. Docking do ligante DAZJOE no receptor R-1vzk centrado em T19 com raio 25 Å

Mais uma vez, os resultados do docking forneceram somente três soluções de orientação idêntica. Semelhantemente ao cálculo anterior, todas se situam no sulco menor sm. Os resultados são mostrados na Tabela 3.13. A solução selecionada foi a 2, devido seu maior escore.

Tabela 3. 13 Resultados de docking do ligante DAZJOE no receptor R-1vzk centrado em T19 (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	67,64	0,00	49,20	0,00
2	67,96	0,00	49,43	0,00
3	67,75	0,00	49,27	0,00

3.3.3.c. Comparação dos resultados de docking centrado em T8 e T19 e Análise das interações.

Ambos os cálculos, mesmo centrados em pontos distantes da estrutura do DNA, apresentaram resultados unânimes e equivalentes. A sobreposição da solução 3 do cálculo da Seção 3.3.3.a e a solução 2 do cálculo 3.3.3.b apresenta um RMSD de apenas 0,71 Å. Estes fatos sugerem que dentro das possíveis situações, esta orientação se destaca claramente do resto. Pode-se então concluir que os resultados sugerem que o sítio de ligação é o sulco menor sm, como inicialmente projetado.

A Tabela 3.14 mostra as interações DNA-ligante encontradas no complexo formado. Algumas delas também podem ser observadas na Figura 3.15. As interações ocorrem entre as bases ATTC do sulco menor.

Tabela 3. 14 Interações entre do ligante DAZJOE e o receptor R-1vzk

Átomo do Ligante	Átomo/Resíduo do DNA	Distância (Å)
C32	H4*/T20	2,25
H14	C5*/G10	2,73
H24	O3*/T19	2,81
N21	H4*/T19	2,32
O46	H4*/T8	2,40
Centróide 32,33,35,37,39,41	H4*/T20	2,34

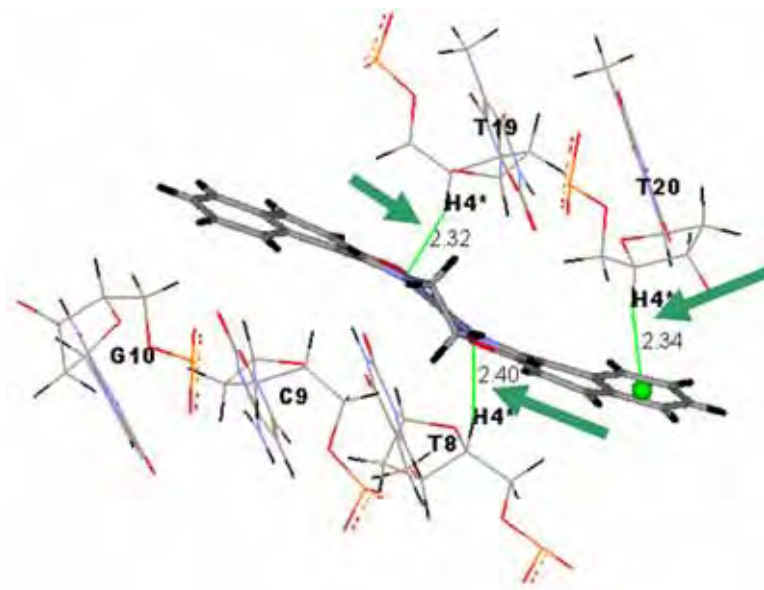


Figura 3. 15 Interações entre os nucleotídeos T8, T19 e T20 e do receptor R-1vzk e o ligante DAZJOE. Estas interações estão descritas na Tabela 3.14.

Podemos dizer que, neste caso, a falta de ambigüidade nos cálculos seguramente está relacionada com o fato de o ligante ser rígido e de não ter grupos doadores ou aceptores de ligação de hidrogênio, e que nos cálculos de docking o programa procurou inespecificamente o local onde as forças de van der Waals tivessem seu valor máximo. Mesmo assim, esta estrutura pareceu interessante e altamente promissora como estrutura líder de modelagem molecular, onde substituintes inseridos nos seus anéis poderiam ter uma participação importante na seletividade do sítio de ligação.

No item seguinte os estudos de docking envolvem um ligante derivado desta estrutura, mais volumoso e com evidências experimentais de que o sítio de ligação neste caso seria o sulco maior SM.

3.4. ESTUDOS DE LIGAÇÃO NO SULCO MAIOR DO DNA

Vijayalakshmi *et al.* (2000) publicaram um trabalho sobre a ligação ao DNA de um composto orgânico que na região central tem um átomo de Cr(III) coordenado equatorialmente com dois átomos de O e dois de N. O composto também apresenta duas águas de coordenação nas posições apicais. Este composto, chamado de CR3CN, é estruturalmente relacionado com DAZJOE, tendo no anel de 5 átomos

central, formado pela sequência C-N-Cr(III)-N-C, dois grupos nitrila ligados em ambos carbonos. Neste caso, os anéis de 6 carbonos que apresenta a estrutura DAZJOE nos extremos foram substituídos por um grupo $N(Et)_2$ em cada extremo. Os autores do trabalho reportaram que, segundo os seus resultados experimentais, o sítio de ligação deste composto seria o sulco maior **SM**; por esta razão, usando a mesma metodologia que nos estudos anteriores, mas agora centrando a procura no sulco maior **SM**, foram realizados cálculos de *docking* entre este ligante e ambas as estruturas de DNA estudadas (R-1vzk e R-2b3e). Para este ligante, foi necessário manter as águas de coordenação devido à presença de muitas de moléculas de água nesta região no meio biológico; além do mais, o átomo de Cr(III) não está envolvido em interações com o DNA, mas só cumprindo uma função geométrico/estrutural no ligante. Esta afirmação deriva dos resultados de Vijayalakshmi *et al.* (2000), onde eles reportam que este ligante após se ligar ao DNA, presumivelmente no sulco maior **SM**, provoca a quebra das cadeias do DNA através do mecanismo apresentado no esquema da Figura 3.16. Segundo os autores, os nitrogênios se ligariam aos fosfatos fazendo com que eles se separassem do $O3^*$.

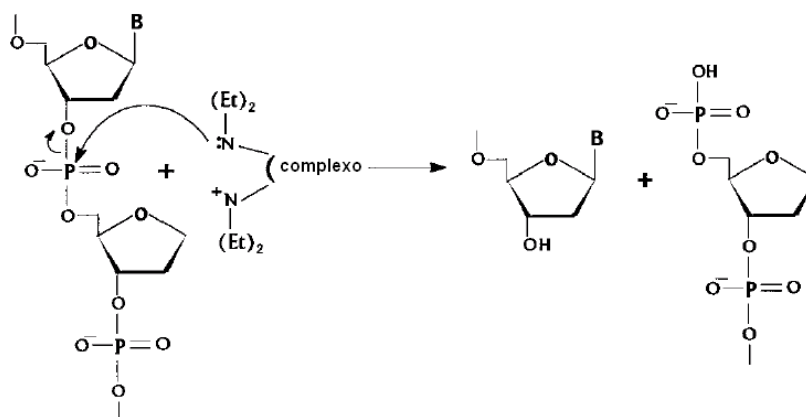


Figura 3. 16 Mecanismo de quebra do DNA pelo composto Cr3NC sugerido por Vaidyanathan *et al.* (2000)

3.4.1. Docking do ligante CR3CN no receptor R-1vzk

Para este cálculo, foi escolhido como centro dos cálculos de *docking* um ponto do espaço aproximadamente no centro do sulco maior **SM**, dado que nessa região não existem átomos do DNA. Do ponto de vista da sequência do DNA, o ponto se encontra aproximadamente no centro de uma das duas metades. Os resultados do

docking geraram nove orientações divididas em dois grupos:

grupo 1: soluções 1, 2, 3, 4, 6 e 9;

grupo 2: soluções 5, 7 e 8.

Enquanto o primeiro grupo interage com ambas as cadeias do DNA o segundo interage somente com uma delas. A Figura 3.17 mostra os grupos de orientações.

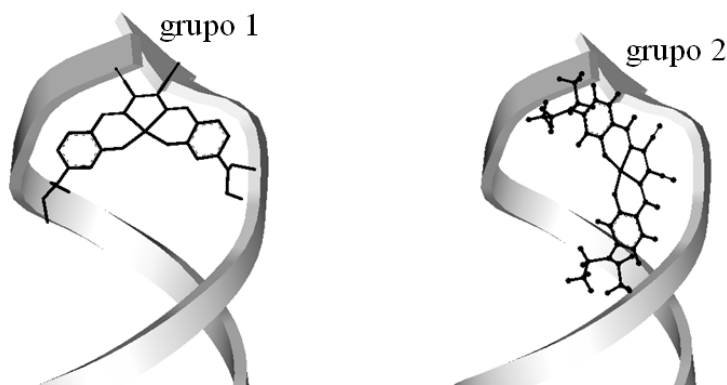


Figura 3. 17 Orientações representativas dos grupos 1 e 2 das soluções do docking do ligante CR3NC no receptor R-1vzk

A solução mais pontuada é a 7, porém faz parte de um grupo minoritário e não foi selecionada. Dentro do grupo 1, a solução melhor pontuada é a solução 6, e foi selecionada como a mais favorável. Na Tabela 3.15 são apresentados os resultados obtidos. Observando a orientação no SM do grupo 1, e levando em conta as informações do artigo de referência ao respeito da reação que ocorre após a ligação, foi considerado que está orientação é a mais compatível com os fatos experimentais.

Tabela 3. 15 Resultados de docking do ligante CR3CN no receptor R-1vzk (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	30,09	0,77	30,71	-12,90
2	28,00	0,73	29,34	-13,07
3	31,42	0,72	31,94	-13,22
4	27,87	0,53	29,31	-12,95
5	30,38	1,09	30,97	-13,29
6	32,04	0,80	31,64	-12,25
7	28,02	1,78	27,94	-12,17
8	28,88	3,79	27,69	-12,99
9	31,93	0,82	31,87	-12,72

Nesta orientação ambos nitrogênios reagentes do ligante ficam bem orientados e a distâncias favoráveis dos fosfatos do DNA. Dentre as diversas interações que ocorrem, aquelas em torno dos oxigênios dos fosfatos merecem destaque na Tabela 3.16.

Tabela 3.16 Interações entre a orientação 6 do ligante CR3CN no receptor R-1vzk

Átomo do Ligante	Átomo do DNA	Distância (Å)
H12	O2P/G14	2,40
H13	O2P/G14	2,97
C21	O2P/G14	2,89
H14	O2P/G14	2,55
C22	O2P/G14	2,90
H17	O2P/G14	2,21
N6	P/T8	5,12
H26	O2P/T7	2,87
C15	O2P/T8	2,77
H7	O2P/T8	1,82
H23	O2P/T8	2,72

3.4.2. Docking do ligante CR3CN no receptor R-2b3e

Neste segundo cálculo de *docking*, agora na estrutura R-2b3e, os resultados do docking forneceram apenas três orientações (Figura 3.18), todas encostadas em uma das faces do sulco maior, interagindo com uma das cadeias. Esta orientação é consistente com a obtida para o grupo 2 no cálculo anterior. Os resultados dos cálculos estão apresentados na Tabela 3.17.

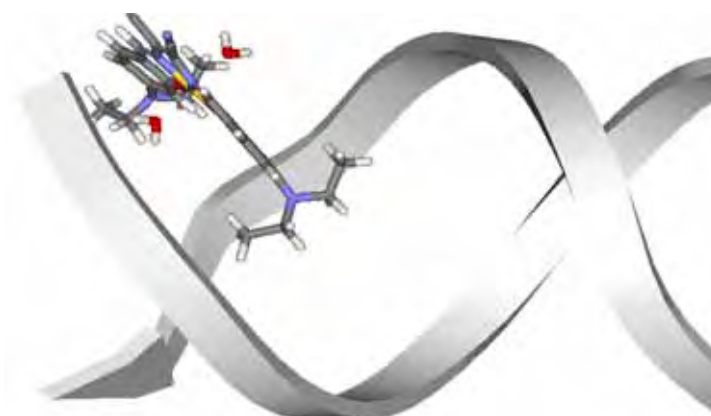


Figura 3. 18 Três orientações coincidentes do ligante CR3NC no receptor R-2b3e. A interação ocorre no sulco maior, porém há contato maior com uma das faces.

Tabela 3. 17 Resultados de docking do ligante CR3CN no receptor R-2b3e (valores em kcal/mol). Em negrito, a orientação escolhida como mais favorável.

Solução	Fitness	S(hb_ext)	S(vdw_ext)	S(int)
1	36,61	2,13	36,20	-12,30
2	36,82	1,43	34,65	-12,25
3	37,62	2,07	35,92	-13,84

Nessa situação, há poucas interações e uma molécula de água faz uma ponte entre o O2 do fosfato da G14 e o ligante. Um ponto que mereceu a nossa atenção foi que neste caso não foi possível obter nenhuma orientação compatível com a selecionada anteriormente, além do mais que agora o resultado foi unânime a favor da orientação do grupo 2. Tentando entender este fato foi realizada uma análise de distâncias para caracterizar a abertura do sulco maior. Foi observado então, que quando se compara o sulco maior de R-1vzk e de R-2b3e, R-1vzk apresenta uma abertura 2 Å maior que a de R-2b3e, apesar do RMSD entre 1vzk e 2b3e ser de apenas 0,7 Å (Figura 3.19). Levando em conta que CR3CN tem um comprimento de aproximadamente 17 Å comprova-se que o espaço requerido para obter uma orientação do tipo da selecionada anteriormente não é suficiente na estrutura 2b3e.

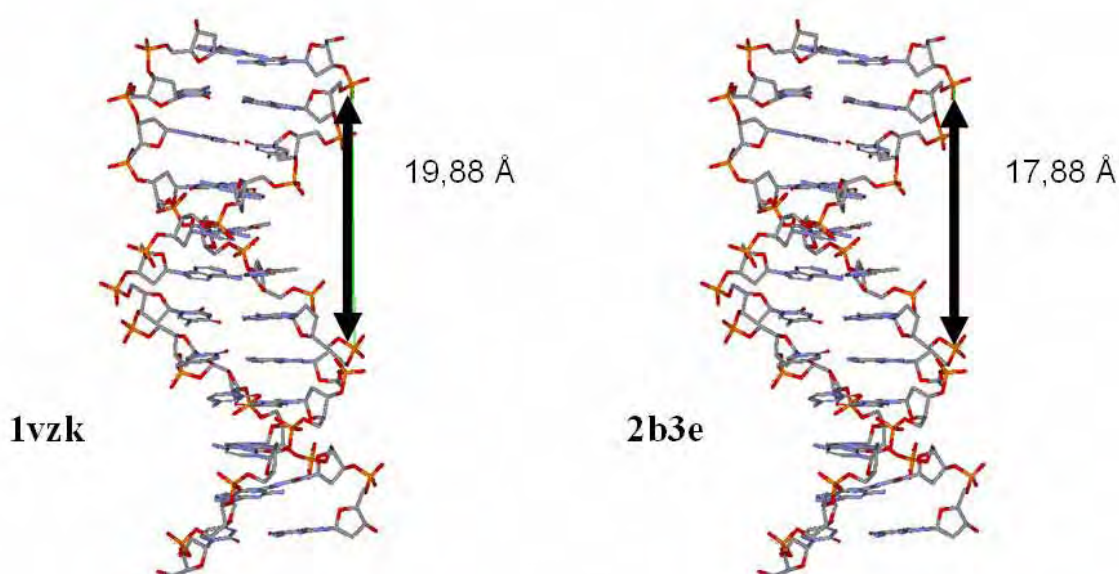


Figura 3. 19 Medida da abertura do sulco maior das estruturas dos receptores R-1vzk e R-2b3e.

Finalmente, considerando a totalidade dos resultados apresentados neste capítulo, pode-se sugerir que há subsídios para considerar que o receptor 1vzk é, de fato, representativo das moléculas de B-DNA formados pelo dodecâmero d(CGCGAATTCGCG)₂ e que R-1vzk pode ser utilizado como receptor para futuros cálculos de docking. Para os estudos no sulco maior de ligantes do tipo de CR3CN, a estrutura 2be3 não parece ser muito adequada, sendo 1vzk um bom modelo para o trabalho. Poderia ainda ser feita uma busca para verificar a existência de alguma outra estrutura de DNA onde o sulco maior apresentasse uma abertura maior que a observada em R-1vzk, dado que os resultados de *docking* no sulco maior sugerem uma dependência forte do grau de abertura dele, ou seja, de quão distorcido o DNA se encontra.

4 CONCLUSÕES

Em recente busca realizada no Portal de Periódicos da CAPES, utilizando (docking + GOLD program + DNA), foram obtidos apenas dois artigos publicados. Em parte, por que o uso do programa GOLD para estudos de docking no DNA não é uma tarefa fácil. Diferentemente do que ocorre com as proteínas, onde as diferenças significativas de cargas e formas entre os aminoácidos conferem aos sítios características próprias, no DNA as semelhanças das bases nucleotídicas, no interior dos sulcos, faz que o panorama seja pouco diferenciado. Então, o primeiro ponto interessante foi o de conseguir trabalhar com o programa GOLD tendo como sistema de análise o DNA. Verificou-se que o programa consegue gerar soluções realísticas, pois a reprodução da estrutura cristalográfica no redocking foi obtida com muito sucesso. O programa consegue, desde que acertadas as condições de entrada, gerar respostas satisfatórias. Conseguiu-se neste trabalho, modificando os centros dos cálculos validar a metodologia, e destacar um centro como o mais promissor para os estudos de docking (T19 em 1vzk).

Outro destaque foi o levantamento sobre as estruturas cristalográficas com a mesma seqüência de bases, moléculas na conformação B-DNA e ligantes. Com esse levantamento foi possível observar que a maioria dos ligantes planares ou quase planares com pelo menos dois anéis aromáticos, muitos dos quais têm sido utilizados como medicamentos, ligam-se ao sulco menor do DNA. O único composto alifático resultante na busca não se liga ao sulco menor.

Também como resultado da busca, foi obtido que das 25 estruturas somente 4 apresentaram átomos diferentes de C, N, H, O. Os ligantes de 1vzk e de 2b3e, apresentaram 1 átomos de S, a de 2dyw 3 átomos de Pt e, finalmente, a 442d, um átomo de I. Neste trabalho, além de ter utilizado os compostos de S, também foram estudados os compostos de crômio, sendo um deles quase planar e outro não. O composto quase planar comporta-se como as demais moléculas e consegue ligar-se ao sulco menor. Observa-se também que a preferência pela seqüência de nucleotídeos AATTC, depende mais da conformação geométrica do ligante do que dos átomos presentes na molécula. Os resultados aqui obtidos

indicam também que a forma da molécula pode ser muito mais determinante para o modo de ligação na formação dos complexos no sulco menor que os próprios átomos das moléculas envolvidas.

A partir dos resultados do trabalho, foi possível verificar que a molécula escolhida como molécula de trabalho (R-1vzk) é uma molécula representativa para a sequência d(CGCGAATTCGCG)₂.

Foram realizados cálculos de docking, redocking, cross-docking e em todos os pontos foi feita uma análise gráfica dos inputs e resultados dos cálculos realizados.

Após o estudo completo de redocking e a análise dos resultados em seu conjunto pode-se afirmar que a metodologia e critérios de seleção utilizados, são adequados e permitem a detecção de resultados realísticos e confiáveis. Estes resultados também representam uma base sólida para trabalhos teóricos, ou seja, naqueles casos em que o local e a forma de se ligar ao DNA não têm uma estrutura de referência, mas conta com dados experimentais (não estruturais), indícios e/ou hipóteses de ligação. Desta forma, é possível avaliar e/ou propor complexos realísticos que guiem futuros estudos experimentais e teóricos nesta área de pesquisa tão importante e ainda tão inexplorada.

A análise dos ligantes estudados permitiu verificar que a molécula DAZJOE, um composto de cromo pode se ligar ao sulco menor como outros compostos que atuam ou são candidatos a fármacos e que tem um modo de ligação similar ao da molécula DB818. Já outro composto de cromo, CR3CN se comportou de forma diferente, e estudos realizados no sulco maior permitiram sugerir um complexo molecular compatível com as evidências experimentais, ligando-se ao sulco maior.

5 REFERÊNCIAS

- ACOT, P. (2003). **A dupla revolução da dupla hélice**. Ciência e Ambiente: Santa Maria - UFSM, 26, 7-16.
- ADAMS, A., LEONG, C., DENNY, W. A., GUSS, J. M. (2005). **Structures of two minor-groove-binding quinolinium quaternary salts complexed with d(CGCGAATTCGCG)(2) at 1.6 and 1.8 Angstrom resolution**. Acta Crystallogr. D. Biol. Crystallogr., 61, 1348 – 1353.
- ALLEN, F.H. (2002). **The Cambridge Structural Database: a quarter of a million crystal structures and rising**. Acta Cryst., B58, 380-388. \
- ANNAMALA, M. K., INAMPUDI, K. K., LALITHA GURUPRASAD, L. (2007). **Docking of phosphonate and trehalose analog inhibitors into *M. tuberculosis* mycolyltransferase Ag85C: Comparison of the two scoring fitness functions GoldScore and ChemScore, in the GOLD software**. Bioinformation, 1, 339–350.
- BAGCHI, D, STOHS, S. J., DOWNS, B. W., BAGCHI, W. M., PREUSS, H. G. (2002). **Cytotoxicity and oxidative mechanisms of different forms of chromium**. Toxicology, 180, 5-22.
- BERA, R., SAHOO, B.K., GHOSH, K.S., DASGUPTA, S. (2008). **Studies on the interaction of isoxazolcurcumin with calf thymus DNA**. Int. J. Biol. Macromol., 42, 14–21.
- BERMAN, H. M. (2008). **The Protein Data Bank: a historical perspective**. Acta Cryst. A64, 88-95.
- BERMAN, H. M., OLSON, W. K., BEVERIDGE, D. L., WESTBROOK, J., GELBIN, A., DEMENY, T., HSIEH, S.-H., SRINIVASAN, A. R., SCHNEIDER, B. (1992). **The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids**. Biophys. J., 63, 751-759.
- BROOIJMANS, N., KUNTZ, I. D. (2003). **Molecular recognition and docking algorithms**. Annu. Rev. Biophys. Biomol. Struct. 32, 335–373.
- BROWN, C.L., BUSHELL, G., WHITEHOUSE, M.W., AGRAWAL, D. S., TUPE, S. G., PAKNIKAR, K. M., TIEKINK, E. R. T. (2007). **Nanogoldpharmaceutics: (i) the use of colloidal gold to treat experimentally-induced arthritis in rat models; (ii) characterization of the gold in Swarna bhasma, a microparticulate used in traditional indian medicine**. Gold Bulletin, 40, 245-250.
- BROWN, D. G., SANDERSON, M. R., SKELLY, J. V., JENKINS, T. C., BROWN, T., GARMAN, E., STUART, D. I., NEIDLE, S. (1990). **Crystal structure of a berenil-dodecanucleotide complex: the role of water in sequence-specific ligand binding**. EMBO J., 9, 1329-1334.
- CALLADINE, C. R.; DREW, H. R.; LUISI, B. F.; TRAVERS, A. A. (2004). **Understanding DNA**. 3a. ed., Elsevier Ltd., Amsterdam – The Netherlands.
- CAMPBELL, N. H., EVANS, D. A., LEE, M. P., PARKINSON, G. N., NEIDLE S. (2006). **Targeting the DNA minor groove with fused ring dicationic compounds: comparison of in silico screening and a high-resolution crystal structure**. Bioorg. Med. Chem. Lett., 16, 15-19.
- CAMPBELL, N. H., EVANS, D.A., LEE, M.P.H., PARKINSON, G. N., NEIDLE, S. (2006). **Targeting the DNA minor groove with fused ring dicationic compounds: Comparison of in silico screening and a high-resolution crystal structure**. Bioorg. Med. Chem. Letters, 16, 15–19.
- CAVALIERE, E. (2004). **Utilizing a Novel Technique for Analyzing the Atomic Structure of DNA Crystals**. oregonstate.edu/dept/biochem/hhmi/undergradresearch/ 2004/cavaliere.ppt.

- CHAIRES, J. B. (2006). **A thermodynamic signature for drug–DNA binding mode**, Archives of Biochemistry and Biophysics, 453, 26–31.
- CHEN, Q., SHAFER, R. H., KUNTZ, I. D. (1997). **Structure-Based Discovery of Ligands Targeted to the RNA Double Helix**, Biochemistry, 36, 11402-11407.
- CHOWDHURY, T., JAMIESON, E. R. (2006). **C4' sugar oxidation of deoxyribonucleotide triphosphates by chromium(V) complexes**, Mutation Res., 610, 66–73.
- CLARK, G. R., BOYKIN, D. W., CZARNY, A., NEIDLE, S. (1997). **Structure of a bis-amidinium derivative of hoechst 33258 complexed to dodecanucleotide d(CGCGAATTCGCG)₂: the role of hydrogen bonding in minor groove drug-DNA recognition**. Nucleic Acids Res., 25, 1510-1515.
- CLARK, G. R., SQUIRE, C. J., GRAY, E. J., LEUPIN, W., NEIDLE, S. (1996). **Designer DNA-binding drugs: the crystal structure of a meta-hydroxy analogue of Hoechst 33258 bound to d(CGCGAATTCGCG)₂**. Nucleic Acids Res., 24, 4882-4889.
- CLARK, M., CRAMER, R.D., VAN OPDENBOSCH, N. (1989). **Validation of the General Purpose Tripos 5.2 Force Field**. J. Comp. Chem., 10, 982-1012.
- CSD – Cambridge Structural Database: <http://www.ccdc.cam.ac.uk/products/csd/>
- DS VISUALIZER™ - Accelrys Software Inc. Discovery Studio Visualizer 2.0
- EDWARDS, K. J., JENKINS T. C., NEIDLE, S. (1992). **Crystal structure of a pentamidine-oligonucleotide complex: implications for DNA-binding properties**. Biochem., 11, 7104-7109.
- ELKIN, L. O. (2003). **Rosalind Franklin and the Double Helix**, Physics Today, 56, 42-49.
- ERMONDI, G., CARON, G. (2006). **Recognition forces in ligand-protein complexes: Blending information from different sources**. Biochem. Pharmacol., 72, 1633-1645.
- EVANS, D. A., NEIDLE, S. (2006). **Virtual Screening of DNA Minor Groove Binders**. J. Med. Chem., 49, 4232-4238.
- FERREIRA, A. D. Q. (2002). **O Impacto do Crômio nos Sistemas Biológicos**. Quim. Nova, 25, 572-578.
- GANS, J., SHALLOWAY, D. (2001). **Qmol: A program for molecular visualization on Windows based PCs**. J. Mol. Graph. Model., 19, 557-559.
- GOLD SCORING-FUNCTION A.P.I. DOCUMENTATION - http://www.ccdc.cam.ac.uk/support/documentation/gold/3_2/api/index.html
- GOODSELL, D. S., NG, H. L., KOPKA, M. L., LOWN, J. W., DICKERSON, R. E. (1995). **Structure of a dicationic monoimidazole lexitropsin bound to DNA**. Biochem., 26, 16654-16661.
- GUERRI, A., SIMPSON, I. J., NEIDLE, S. (1998). **Visualisation of extensive water ribbons and networks in a DNA minor-groove drug complex**. Nucleic Acids Res., 26, 2873-2878.
- HA, S. C., LOWENHAUPT, K., RICH, A., KIM, Y-G, KIM, K.K. (2005). **Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases**. Nature, 437, 1183-1186.
- HUMAN GENOME PROJECT - <http://www.ornl.gov>, consultado em 22/10/2007.

- JONES, G., WILLETT, P., GLEN, R. C., LEACH, A. R. , TAYLOR, R. (1997). **Development and Validation of a Genetic Algorithm for Flexible Docking**. J. Mol. Biol., 267, 727-748.
- JONES, G., WILLETT, P., GLEN, R.C. (1995). **Molecular Recognition of Receptor Sites using a Genetic Algorithm with a Description of Desolvation**. J. Mol. Biol., 245, 43-53.
- JONES, T.A. (1982). **Computational Crystallography**, Clarendon Press, Oxford, Editor D. Sayre., 303-317.
- JONES, T.A., KJELDGAARD, M. **"Manual for O"**, Uppsala, Uppsala University.
<http://xray.bmc.uu.se/alwyn/>
- KOMEDA, S., MOULAEI, T., WOODS, K. K., CHIKUMA, M., FARRELL, N. P., WILLIAMS, L. D (2006). **A third mode of DNA binding: Phosphate clamps by a polynuclear platinum complex**. J. Am. Chem. Soc., 128, 16092-160103.
- LANKAS, F., SPONER, J., HOBZA, P., LANGOWSKI, J. (2000). **Sequence-dependent Elastic Properties of DNA**. J. Mol. Biol., 299, 695-709.
- LARSEN, T.A., GOODSSELL, D.S., CASCIO D., GRZESKOWIAK K., DICKERSON, R.E. (1989). **The structure of DAPI bound to DNA**. J. Biomol. Struct. Dyn., 7, 477- 491.
- LASKOWSKI, R. A, CHISTYAKOV, V.V., THORNTON, J.M. (2005). **PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids**. Nucleic Acids Res., 33, Database issue D266-D268.
- LASKOWSKI, R. A. (2007). **Enhancing the functional annotation of PDB structures in PDBsum using key figures extracted from the literature**. Bioinformatics, 23, 1824-1827.
- LAUGHTON, C. A., TANIOUS, F., NUNN, C. M. , BOYKIN, D. W., WILSON, W. D., NEIDLE S. (1996). **A crystallographic and spectroscopic study of the complex between d(CGCGAATTCGCG)₂ and 2,5-bis(4-guanyphenyl)furan, an analogue of berenil. Structural origins of enhanced DNA-binding affinity**. Biochem., 35, 5655-5661.
- LI, H, LAI, C. S. , WU, J. , HO, P. C. , DE VOS, D., TIEKINK, E. R. T. (2007). **Cytotoxicity, Qualitative Structure-Activity Relationship (QSAR), and anti-tumor activity of bismuth dithiocarbamate complexes**. J. Inorg. Biochem., 101, 809-816.
- MADDOX, B. (2002). **Rosalind Franklin: The Dark Lady of DNA**. HarperCollins Ed., New York.
- MALLENA, S., LEE, M. P. H., BAILLY, C., NEIDLE, S., KUMAR, A., BOYKIN, D. W., WILSON, W.D. (2004). **Thiophene-Based Diamidine Forms a "Super" AT Binding Minor Groove Agent**. J. Am. Chem. Soc., 126, 13659-13669.
- MANN, J., BARON, A., OPOKU-BOAHEN, Y., JOHANSSON, E., PARKINSON, G., KELLAND, L. R., NEIDLE, S. (2001). **A new class of symmetric bisbenzimidazole-based DNA minor groove-binding agents showing antitumor activity**. J. Med. Chem., 18, 138-144.
- MAZZER, P.A., MAURMANN, L., BOSE, R. N. (2007). **Mechanisms of DNA damage and insight into mutations by chromium(VI) in the presence of glutathione**. J. Inorg. Biochem., 101, 44-55.
- MÉNDEZ, P.S, GUEDES, R.C., DOS SANTOS, D.J.V.A., ERIKSSON, L.A. (2007). **Theoretical prediction of binding modes and hot sequences for allopsoralen–DNA interaction**. Chem. Phys. Letters, 450, 127–131.
- MERCURY CSD - <http://www.ccdc.cam.ac.uk/products/mercury/>

MIRANDA, M.N. (2000). **Algoritmos Genéticos: Fundamentos e Aplicações**. Grupo de Teleinformática e Automação – UFRJ -<http://www.gta.ufrj.br/~marcio/genetic.html>.

MITCHELL, M. (1996). **An Introduction to Genetic Algorithms**. MIT Press, Cambridge.

MOPAC 2007, James J. P. Stewart. (2007). Stewart Computational Chemistry, Colorado Springs, CO, USA, <http://OpenMOPAC.net>

MUNDE, M., ISMAIL, M.A., PEIXOTO, R.A.P., J. COLLAR, C.J., LIU, Y., HU, L., LE, M.-H., LANSIAUX, A.I., BAILLY, C., BOYKIN, D.W., WILSON, D. (2007). **Design of DNA Minor Groove Binding Diamidines That Recognize GC Base Pair Sequences: A Dimeric-Hinge Interaction Motif**. J. Am. Chem. Soc., 129, 13732-13743.

NDB - **Nucleic Acid Database**: <http://ndbserver.rutgers.edu/>

NGUYEN, B., LEE, M. P. H., HAMELBERG, D., JOUBERT, A., BAILLY, C., BRUN, R., NEIDLE, S., WILSON, W. D. (2002). **Strong Binding in the DNA Minor Groove by an Aromatic Diamidine with a Shape That Does Not Match the Curvature of the Groove**. J. Am. Chem. Soc., 124, 13680-13681.

NUNN, C. M., JENKINS, T. C., NEIDLE, S. (1994). **Crystal Structure of γ -Oxapentamidine Complexed with d(CGCGAATTCGCG)₂. The Effects of Drug Structural Change on DNA Minor-Groove Recognition**. Eur. J. Biochem., 3, 953-961.

PBS Foundation: <http://www.pbs.org/wgbh/nova/photo51/>

PDB - Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description. Version 2.3 (1998). <http://www.wwpdb.org/documentation/format2.3-0108-us.pdf>.

PDB - **Protein Data Bank**: <http://www.rcsb.org/pdb/home/home.do>

PDBSum: <http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/>

PEDRETTI, A., VILLA, L., VISTOLI, G. (2002). **Vega: A Versatile Program to Convert, Handle And Visualize Molecular Structure On Windows-Based PCs**. J. Mol. Graph., 21, 47-49.

PEDRETTI, A., VILLA, L., VISTOLI, G. (2003). **Atom-Type Description Language: A Universal Language To Recognize Atom Types Implemented In The Vega Program**. Theor. Chem. Acc., 109, 229-32.

PEDRETTI, A., VILLA, L., VISTOLI, G. (2004). **Vega - An Open Platform To Develop Chemo-Bio-Informatics Applications, Using Plug-In Architecture And Script Programming**. J. Computer-Aided Mol. Design, 18, 167-173.

PJURA, P. E., GRZESKOWIAK, K., DICKERSON, R. E. (1987). **Binding of Hoechst 33258 to the minor groove of B-DNA**. J. Mol. Biol., 197, 257-271.

QMOL - <http://www.dnastar.com/products/qmol/index.html>

RHODES, G. (2006). **Crystallography Made Crystal Clear - A Guide for Users of Macromolecular Models**. 3a. ed., Elsevier/Academic Press, New York.

- SARMA, M., MAZUMDER, S., GHOSH, D., ROY, A., DUTHIE, A., TIEKINK, E.R.T. (2007). **Synthesis, spectroscopic characterization and biocidal activity of some diorganotin(IV) complexes of salicylaldehyde thiosemicarbazones and related ligands. Molecular and supramolecular structures of $[R_2Sn(OArCH=N-N=CSNH_2)]$ where $R = Me, Ph$ and $Ar = -C_6H_4, -C_6H_3(5-Br)$ and $C_6H_3(5-Cl)$, and of $[Me_2Sn\{OC_6H_3(5-Br)CH=N-N=CSNH_2\}OH_2]$.** *Appl. Organomet. Chem.*, 21, 890-905.
- SEMENZA, G. (2003). **Fifty Years Ago DNA: the double helix.** *FEBS Letters*, 544, 1-3.
- SIMONI, R.D., HILL, R. L., VAUGHAN, M. (2002). **The Structure of Nucleic Acids and Many Other Natural Products: Phoebus Aaron Levene.** *J. Biol. Chem.*, 277, p. e11.
- SIMPSON, I. J., LEE, M., KUMAR, A., BOYKIN, D. W., NEIDLE, S. (2000). **DNA minor groove interactions and the biological activity of 2,5-bis-[4-(N-alkylamidino)phenyl] furans.** *Bioorg. Med. Chem. Lett.*, 10, 2593-2597.
- SPEETJENS, J. K., COLLINS, R.A., VINCENT, J. B. , WOSKI, S. A. (1999). **The Nutritional Supplement Chromium(III) Tris(piccolinate) Cleaves DNA.** *Chem. Res. Toxicol.*, 12, 483 -487.
- SPEETJENS, J. K., PARAND, A., CROWDER, M. W., VINCENT, J. B., WOSKIA, S. A. (1999). **Low-molecular-weight chromium-binding substance and biomimetic $[Cr_3O(O_2CCH_2CH_3)_6(H_2O)_3]^+$ do not cleave DNA under physiologically-relevant conditions.** *Polyedron*, 18, 2617-2624.
- SQUIRE, C. J., BAKER, L. J., CLARK, G. R., MARTIN, R. F., WHITE, J. (2000). **Structures of m-iodo Hoechst-DNA complexes in crystals with reduced solvent content: implications for minor groove binder drug design.** *Nucleic Acids Res.*, 2000 28, 1252-1258.
- SQUIRE, C. J., CLARK, G. R., DENNY, W. A. (1997). **Minor groove binding of a bis-quaternary ammonium compound: the crystal structure of SN 7167 bound to d(CGCGAATTCGCG)₂.** *Nucleic Acids Res.*, 25, 4072-4078.
- SRIRAM, M., VAN DER MAREL, G. A., ROELEN, H. L. P. F., VAN BOOM, J.H., WANG, A. H. J. (1992). **Conformation of B-DNA containing O6-ethyl-G-C base pairs stabilized by minor groove binding drugs: molecular structure of d(CGC[e6G]AATTCGCG complexed with Hoechst 33258 or Hoechst 33342.** *EMBO J.*, 11, 225-232.
- SRIRAM, M., VAN DER MAREL, G. A., ROELEN, H. L. P. F., VAN BOOM, J.H., WANG, A. H. J. (1992). **Structural consequences of a carcinogenic alkylation lesion on DNA: effect of O6-ethylguanine on the molecular structure of the d(CGC[e6G]AATTCGCG)-netropsin complex.** *Biochem.*, 31, 11823-11834.
- STEWART J. J. P. (2007). **Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements.** *J. Mol. Modeling* 13, 1173-1213.
- SUGDEN, K. D., WETTERHAHN, K. E. (1996) **EPR Evidence for Chromium(V) Binding to Phosphate and Pyrophosphate: Implications for Chromium(V)-DNA Interactions.** *Inorg. Chem.* 1996, 35, 3727-3728.
- TRENT, J. O., CLARK, G. R., KUMAR, A., WILSON, W. D., BOYKIN, D. W., HALL, J. E., TIDWELL, R. R., BLAGBURN, B. L., NEIDLE, S. (1996). **Targeting the Minor Groove of DNA: Crystal Structures of Two Complexes between Furan Derivatives of Berenil and the DNA Dodecamer d(CGCGAATTCGCG)₂.** *J. Med. Chem.*, 39, 4554 -4562.
- TRIPOS - <http://www.tripos.com/data/support/mol2.pdf>.

VAIDYANATHAN, V.G., WEYHERMULLER, T., NAIR, B.U., SUBRAMANIAN, J. (2005). DNA damage induced by a chromium(III) Schiff base complex is reversible under physiological condition. *J. Inorg. Biochem.*, 99, 2248–2255.

VAN GARDEREN, C.J., VAN HOUTE, L. P.A. (1994). **The solution structure of a DNA duplex containing the *cis* -Pt (NH₃)₂ [d (-GT G-) -N7(G), N7(G)] adduct, as determined with high-field NMR and molecular mechanics/dynamics.** *Eur. J. Biochem.*, 225, 1169-1179.

VEGA ZZ 2.1.0 for Windows - <http://www.ddl.unimi.it>

VERDONK, M.L., COLE, J.C., HARTSHORN, M.J., MURRAY, C.W., TAYLOR, R.D. (2003). **Improved Protein–Ligand Docking Using GOLD.** *Proteins*, 52,609–623.

VIJAYALAKSHMI, R., KANTHIMATHI, M., SUBRAMANIAN, V., NAIR, B. U. (2000). **DNA Cleavage by a Chromium(III) Complex.** *Biochem. Biophys. Res. Commun.*, 271, 731–734.

VOET, D.; VOET, J. G.; PRATT, C. W. (2000). **Fundamentos de Bioquímica** 1^a. edição, Editora Artmed, Porto Alegre.

WOOD, A. A., NUNN, C. M., CZARNY, A., BOYKIN, D. W., NEIDLE, S. (1995). **Variability in DNA minor groove width recognised by ligand binding: the crystal structure of a bis-benzimidazole compound bound to the DNA duplex d(CGCGAATTCGCG)₂.** *Nucleic Acids Res.*, 25, 3678–3684.

ZHAO, C., XIA, C., MAO, Q., FÖRSTERLING, H., DEROSE, E., ANTHOLINE, W. SUBCZYNSKI, W. K., PETERING, D. H. (2002). **Structures of HO₂-Co(III)bleomycin A₂ bound to d(GAGCTC)₂ and d(GGAAGCTTCC)₂: structure–reactivity relationships of Co and Fe bleomycins.** *J. Inorg. Biochem.*, 91, 259-268.