



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

Avaliação do Uso de Técnicas de MMGBSA na Triagem Virtual de Bibliotecas de Ligantes para o Desenvolvimento de Fármacos contra *Leishmania amazonensis* e *Leishmania chagasi*.

Gabriel Aires Urquiza de Carvalho

João Pessoa – PB – Brasil
Outubro/2011



UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

DISSERTAÇÃO DE MESTRADO

Avaliação do Uso de Técnicas de MMGBSA na Triagem Virtual de Bibliotecas de Ligantes para o Desenvolvimento de Fármacos contra *Leishmania amazonensis* e *Leishmania chagasi*.

Gabriel Aires Urquiza de Carvalho*

Dissertação apresentada
como requisito para a
obtenção do título de Mestre
em Química
pela Universidade Federal da
Paraíba

Orientador: Prof. Dr. Gerd Bruno da Rocha

***Bolsista CNPq**

**João Pessoa – PB – Brasil
Outubro/2011**

C331a Carvalho, Gabriel Aires Urquiza de.

Avaliação do uso de técnicas de MMGBSA na triagem virtual de bibliotecas de ligantes para o desenvolvimento de fármacos contra leishmania amazonensis e leishmania chagasi / Gabriel Aires Urquiza de Carvalho.-- João Pessoa, 2011.

165f. : il.

Orientador: Gerd Bruno da Rocha

Dissertação (Mestrado) – UFPB/CCEN

1. Química. 2. Leishmania. 3. MMGBSA. 4. MMPBSA.
5. Dinâmica molecular. 6. Triagem virtual.

UFPB/BC

CDU: 54(043)

A meu pai Francisco. Que tantas vezes ensinou, com empolgação, sobre como o mundo é mais belo quando olhamos através das lentes da ciência. Dedico

AGRADECIMENTOS

- **A minha mãe**, por todas as vezes que a acordei no meio da madrugada fazendo ruídos ao ir dormir.
- **Ao Prof. Dr. Gerd Bruno da Rocha**, meu orientador e amigo, que disse algo como “*vamos nessa*” quando eu saí de seis meses de iniciação científica, sem saber o que diabos era um autovetor, e disse a ele “*Gerd, quero ser mestre em química. Vamos?*”. Foi muita coragem da parte dele.
- **Aos Basterds** Ítalo, Higo e Jessé. Meus colegas, amigos, irmãos. Pelas discussões de alto nível, absolutamente bem gabaritadas, sobre assuntos diversos. Tenho certeza que o mestrado não teria sido tão agradável sem a presença destas pessoas absolutamente ímpares (com exceção, talvez, de Higo-Ítalo).
- **Ao Prof. Dr Silmar Andrade do Monte**, pelas conversas quânticas na copa do LQQC, regada a cafezinho, pelas questões impossíveis de tunelamento e, claro, pelos zumbis.
- **À Profa. Dra. Elizete Ventura do Monte**, por todas as vezes que foi atrás de respostas para perguntas vindas sabe-se lá de onde que apareciam no meio das aulas e bagunçavam todo o esquema. Minhas sinceras desculpas se alguém se feriu com uma pergunta perdida.
- **A Isabelle Cardoso Ferreira**, minha namorada, meu amor e, espero, futura esposa. Porque só ela mesmo é quem sabe como é ouvir “*Não dá, tenho coisas do mestrado para fazer*” toda vez que eu jogava terra em nossos planos. Aguentou a parada com tenacidade heroica, nem sequer me esganar (muito).
- **Ao Prof. Dr. Sidney Ramos de Santana**, por fazer o AMBER calcular o que ele supostamente deveria calcular, *como* ele supostamente deveria calcular. Pode parecer fácil dito assim, mas sem ele o trabalho não teria saído. Além disso, agradeço por todas as dicas, pérolas de conhecimento, explicações e sabedoria Jedi. Em especial naqueles dias em que o computador venceu.
- **A todos os amigos do LQQC**. Gessenildo, Felipe, Walkíria, Weruska, Kelson, Júlio Daniel, Lucinês, Carlos, Lucas, Raílton, Arquimedes, Jefferson, Aline, Vanessa Cristina e especialmente a Juliana Monteiro. Pelo apoio, conversas jogadas fora, fofocas na hora do cafezinho e discussões científicas enquanto os cálculos rodavam (e frequentemente caíam).
- **Ao Dr. Carl Sagan**, por ter escrito o livro *O Mundo Assombrado Pelos*

Demônios. No qual com suas palavras tranquilas e seu raciocínio suave porém preciso transpôs suavemente todas as barreiras da minha mente e ligou as luzes que estavam dentro.

- **A Sir Terenced David John “Terry” Pratchett**. Meu escritor favorito. Por seus livros hilários, seu humor sarcástico e sua forma completamente inventiva de fazer rir e ao mesmo tempo ensinar uma forma mais divertida de olhar as coisas.
- **A Toda a Minha Família**. Que me encorajaram e me deram apoio para seguir adiante qualquer que fosse o tamanho do obstáculo.
- **Aos meus irmãos**, em especial. Pelas conversas filosóficas absolutamente gratuitas e vindas do nada, na cozinha.
- **A Toda a Família da Minha Namorada**. Pelas palavras de apoio, as reuniões, festas e comemorações. Por me chamarem a todos os jantares e sempre me perguntarem, sem falta, *“como vai o mestrado?”*.
- **À M^a Magda Cecília Cardoso Ferreira**, em especial. Minha futura sogra. Por dividir comigo os prazeres e percalços de já ter feito esta jornada que estou findando. E também por todas as lasanhas, macarronadas, macaxeiras, cachorros-quentes, e tudo o mais que não posso listar aqui por falta de espaço.
- **Aos Amigos do LABID**, Jocelmo, Talita, Elis, que ainda estão lá e aos amigos de batalhas passadas, Larissa, Lívia, Airlla e Chris. Por todos os passos dado em conjunto, na aurora da minha vida científica.
- **Ao Prof. Dr. Luis Fernando Marques dos Santos**, por ter me iniciado nas sutilezas e detalhes da metodologia da ciência (nem sempre com tanta sutileza assim) e por ser um amigo com quem falar sobre biologia é sempre divertido.
- **A todos os pesquisadores de todas as épocas**. Pois assim como todo ser vivo está conectado à árvore da vida, também cada trabalho científico está conectado um ao outro. Ciência não se faz sozinho.

“Retornar ao ponto de partida não é o mesmo que nunca partir.”

– Sir Terence David John “Terry” Pratchett

Sumário

RESUMO.....	16
ABSTRACT.....	17
1. Introdução.....	18
1.1. Leishmanioses.....	19
1.1.1. Leishmania. sp.....	19
1.1.2. Tratamento.....	22
1.1.3. Projeto Racional de Fármacos e Triagem Virtual.....	23
1.2. Fundamentação Teórica.....	27
1.2.1. Docking Molecular.....	27
1.2.2. Dinâmica Molecular.....	29
1.2.2.1. Ensembles e Funções de Partição.....	30
1.2.2.2. Mecânica Clássica.....	38
1.2.2.3. Mecânica Molecular.....	44
1.2.2.4. Raios de Corte.....	46
1.2.2.5. Fronteiras Periódicas.....	50
1.2.2.6. A Soma de Ewald.....	51
1.2.2.7. O Método PME (Particle-Mesh Ewald).....	55
1.2.2.8. Controle da Simulação.....	56
1.2.2.8.1 Termostatos.....	57
1.2.2.8.1.1. Berendsen.....	57
1.2.2.8.1.2. Andersen.....	58
1.2.2.8.1.3. Langevin.....	59
1.2.2.8.2. Barostatos.....	59
1.2.2.8.2.1. Berendsen.....	65
1.2.2.8.2.2. Andersen.....	65
1.2.2.9. Análise da Simulação.....	66
1.2.2.10. Energia Livre.....	67
1.2.2.10.1. Integração Termodinâmica	69
1.2.2.10.2. Perturbação da Energia Livre	72
1.2.2.10.3. Método MMGBSA.....	80
1.2.2.10.3.1. MMGBSA e Estruturas de Mínimo.....	90
2. Objetivos.....	91
2.1. Geral.....	92

2.2. Específicos.....	92
3. Metodologia.....	93
3.1 Busca e Seleção de Alvos Moleculares.....	94
3.2. Cálculos de Homologia das Enzimas Seleccionadas.....	100
3.3. Geração Automatizada de Ligantes	101
3.4. Docking Molecular.....	102
3.5. Cálculos de Estimativa da Energia Livre.....	104
3.5.1. Preparação das Estruturas dos Complexos Enzima-Ligante.....	105
3.5.2. Dinâmica Molecular.....	107
3.5.3. MMGBSA.....	108
4. Resultados e Discussões.....	110
4.1. Análise dos Modelos por Homologia.....	111
4.2. Ligantes Gerados com o Procedimento Automatizado.....	125
4.3. Redock dos Ligantes Cristalográficos.....	128
4.4. Sobre as Trajetórias de Dinâmica Molecular.....	130
4.5. Avaliação dos Cálculos de Afinidade de Interação.....	131
4.5.1. Correlação nas Estimativas de Afinidade.....	133
4.5.2. Correlação nos Ordenamentos	138
4.5.3. Avaliação do Docking Rígido na Criação dos Complexos.....	148
5. Conclusões e Perspectivas.....	154
5.1. Conclusões.....	155
5.2. Perspectivas.....	156
6. Referências.....	157

Índice de ilustrações

Figura 1: Indivíduo da espécie <i>Lutzomyia longipalpis</i> . http://upload.wikimedia.org/wikipedia/commons/5/59/Lutzomyia_longipalpis-sandfly.jpg	19
Figura 2: Formas parasitárias de <i>Leishmania</i> sp. Retirado de http://www.medicinapratica.com.br/wp-content/uploads/2009/07/leishmaniose.JPG	20
Figura 3: Ciclo evolutivo da Leishmaniose (figura adaptada de http://dpd.cdc.gov/dpdx/HTML/Leishmaniasis.htm).....	21
Figura 4: Tipos de leishmanioses.....	22
Figura 5: Principais fármacos contra leishmanioses.....	23
Figura 6: Interação Enzima-Ligante.....	25
Figura 7: Relação Termodinâmica Estatística/Clássica	30
Figura 8: A aproximação de Stirling.....	34
Figura 9: Ciclo de Integração do Velocity-Verlet.....	42
Figura 10: Esquema da função switch. O potencial só sofre ação suavizante após o raio interno.....	48
Figura 11: Esquema ilustrativo da função shift. Até mesmo dentro do raio o potencial sofre suavização.....	49
Figura 12: Exemplo de um grid.....	49
Figura 13: Função erro erfc.	55
Figura 14: Posicionamento relativo do vetor r com o elemento de área da parede.....	62
Figura 15: Uma transformação qualquer.....	69
Figura 16: Esquema conceitual mostrando uma série de estados acoplados.....	70
Figura 17: Esquema de uma perturbação.....	72
Figura 18: Esquema da quebra de uma perturbação em várias perturbações intermediárias acopladas	75
Figura 19: Cálculo de muitas perturbações.....	76
Figura 20: Formação de uma interação sódio-cloreto em solução.....	77
Figura 21: Formação de uma interação sódio-cloreto em fase gasosa.....	77
Figura 22: Espaço abstrato de estados com energias internas dependentes da cor e da forma do estado.....	78
Figura 23: Transição para a qual desejamos calcular a variação na energia livre.....	78
Figura 24: Ciclo termodinâmico para considerar todas as variações na energia livre decorrentes da transformação de A em B.....	79
Figura 25: Ciclo termodinâmico para calcular a variação da energia livre quando ocorre a formação de uma ligação sódio-cloreto em solução.....	79
Figura 26: Ciclo termodinâmico para calcular a variação da energia livre para uma interação ligante-macromolécula em solução aquosa.....	80

Figura 27: Linearização da função $\sinh$	82
Figura 28: Esquema ilustrativo de uma carga pontual q dentro de uma cavidade em meio a uma dielétrica contínua.....	83
Figura 29: Utilização de um microestado do complexo solvatado para conseguir as estruturas no vácuo.....	87
Figura 30: Esquema mostrando uma trajetórias clássica.....	88
Figura 31: Mesma partícula da figura 30, com o primeiro passo da trajetória levemente modificado por um evento caótico.....	88
Figura 32: Mapa metabólico demonstrando onde estão as enzimas selecionadas. Em azul, a DHDH, em verde a TR e em amarelo a FPPS. Adaptado a partir de http://www.kegg.jp/kegg/atlas/?01100 .	99
Figura 33: Reação catalizada pela FPPS. Adaptada do KEGG. http://www.genome.jp/Fig/reaction/R01658.gif	99
Figura 34: Reação de redução catalizada pela DHDH. Adaptado do KEGG, http://www.genome.jp/Fig/reaction/R01868.gif	100
Figura 35: Reação de redução-oxidação da Tripanotiona. Adaptado do KEGG http://www.genome.jp/Fig/reaction/R03821.gif	100
Figura 36: Excerto dos resultados produzidos pela ferramenta BLAST no portal UniProt para uma busca por sequências homólogas da sequência da subunidade A da enzima de código PDB 2JK6.	101
Figura 37: Grid 50x38x46 da DHDH.....	104
Figura 38: Esquema ilustrando as séries de dinâmicas moleculares.....	108
Figura 39: Alinhamento estrutural da FPPS, amarelo, com seu molde em preto.....	113
Figura 40: Alinhamento estrutural da TR, verde, com seu molde em preto.....	114
Figura 41: Alinhamento estrutural da DHDH, azul, com seu molde em preto.....	114
Figura 42: Risedronato, inibidor da FPPS. Retirado de http://www.pdb.org/pdb/images/RIS_500.gif	116
Figura 43: Orotato, substrato natural da DHDH. Retirado de http://www.pdb.org/pdb/images/ORO_500.gif	116
Figura 44: Alinhamento DHDH. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.....	116
Figura 45: Alinhamento DHDH. Seleção em verde representa aminoácidos a 7Å do ligante cristalográfico.....	117
Figura 46: Alinhamento DHDH. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.....	118
Figura 47: Alinhamento FPPS. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.....	119
Figura 48: Alinhamento FPPS. Seleção em verde representa aminoácidos a 7Å do ligante cristalográfico.....	120
Figura 49: Alinhamento FPPS. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.....	121
Figura 50: Tripanotiona, substrato natural da TR. Retirado de http://www.pdb.org/pdb/images/GCG_500.gif	122

Figura 51: Alinhamento TR. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.....	123
Figura 52: Alinhamento TR. Seleção em verde representa aminoácidos a 7Å do ligante cristalográfico.....	125
Figura 53: Alinhamento TR. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.....	126
Figura 54: Grupos utilizados nos análogos do Orotato, separados por seção.....	128
Figura 55: Grupos utilizados nos análogos da Tripanotiona, separados por seção.....	129
Figura 56: Adutos de Morita Baylis-Hillman.....	130
Figura 57: Redock do Orotato.....	131
Figura 58: Redock do Risedronato.....	131
Figura 59: Redock da Tripanotiona. Cadeias diferentes mostradas em cores diferentes para propósitos de clareza.....	132
Figura 60: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na TR.....	136
Figura 61: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na FPPS.....	136
Figura 62: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na FPPS sem o risedronato.....	137
Figura 63: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na DHDH.....	137
Figura 64: Padrão de quatro ligações de hidrogênio. Sítio ativo em branco, ligações de hidrogênio em ciano.....	138
Figura 65: Padrão de seis ligações de hidrogênio. Sítio ativo em preto, ligações de hidrogênio em ciano.....	139
Figura 66: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da TR.....	144
Figura 67: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da DHDH.....	145
Figura 68: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da FPPS.....	145
Figura 69: Moléculas coincidentes selecionadas pelo min-MMGBSA.....	150
Figura 70: Superposição da tripanotiona cristalográfica (azul) com a tripanotiona posicionada com o docking rígido e minimizada por 1000 passos (vermelho).....	151
Figura 71: Superposição da tripanotiona cristalográfica (azul) com a tripanotiona posicionada com o docking rígido e minimizada por 3000 passos (vermelho).....	152
Figura 72: Superposição da tripanotiona cristalográfica (azul) com a tripanotiona posicionada com o docking rígido e minimizada por 5000 passos (vermelho).....	153

Figura 73: Figura mostrando duas ligações de hidrogênio intramoleculares na tripanotona e uma ligação de hidrogênio intermolecular. Estrutura minimizada aos 1000 passos.....	154
Figura 74: Figura mostrando a distância entre os mesmos átomos participantes na ligação de hidrogênio da figura anterior, na estrutura cristalográfica (em rosa).....	154
Figura 75: Ligação de hidrogênio formada na estrutura minimizada (em cores) que se encontra ausente na estrutura cristalográfica (em rosa).....	155

Índice de Tabelas

Tabela 1: Enzimas selecionadas do PDB por serem alvos moleculares de interesse medicinal e pela carência de estudos teóricos sobre elas.....	96
Tabela 2: Percentual de similaridade das enzimas selecionadas com suas sequências homólogas de <i>L. chagasi</i> , e <i>L. amazonensis</i>	97
Tabela 3: Percentual de aminoácidos de cada enzima em regiões de qualidade dos diedrais.....	112
Tabela 4: Sumário dos dados da homologia das três enzimas.....	115
Tabela 5: Estimativas de afinidade dos ligantes da TR pelos três métodos.....	134
Tabela 6: Estimativas de afinidade dos ligantes da FPPS pelos três métodos.....	134
Tabela 7: Estimativas de afinidade dos ligantes da DHDH pelos três métodos.....	135
Tabela 8: Coeficiente de correlação linear (R^2) dos valores de ΔG -min e AD com os valores do ΔG -multi.....	140
Tabela 9: Distribuição dos créditos feita por cada método para cada ligante da DHDH.....	142
Tabela 10: Distribuição dos créditos feita por cada método para cada ligante da TR.....	143
Tabela 11: Distribuição dos créditos feita por cada método para cada ligante da FPPS.....	144
Tabela 12: Ordenamento dos ligantes da TR seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.....	146
Tabela 13: Ordenamento dos ligantes da DHDH seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.....	147
Tabela 14: Ordenamento dos ligantes da FPPS seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.....	148
Tabela 15: 1/3 superior da tabela de ordenamento da TR. Coincidências com multi-MMGBSA estão em negrito.	148
Tabela 16: 1/3 superior da tabela de ordenamento da FPPS. Coincidências com o multi-MMGBSA estão em negrito.....	149
Tabela 17: 1/3 superior da tabela de ordenamento da DHDH. Coincidências com o multi-MMGBSA estão em negrito.....	149

Índice de Abreviaturas e Notações

LTC.....	Leishmaniose Tegumentar Cutânea
LMC.....	Leishmaniose Mucocutânea
LV.....	Leishmaniose Visceral
PMF.....	Potential Mean Field
NVT.....	Ensemble com Número de Partículas, Temperatura e Volume constante
NPT.....	Ensemble com Número de Partículas, Temperatura e Pressão constantes
NVE.....	Ensemble com Número de Partículas, Energia Total e Volume constantes
NHP.....	Ensemble com Número de Partículas, Entalpia e Pressão constantes
$Q(N,V,T)$	Função de Partição canônica
$\langle A \rangle$	Valor esperado da propriedade A
$\mathbf{R}(t)$	Vetor Posição
$\mathbf{V}(t)$	Vetor Velocidade
$U(\mathbf{R})$	Potencial
$\mathbf{F}(t)$	Vetor Força
AM1-BCC.....	Austin Method 1 – Bond Charges Correction
RESP.....	Restricted Eletrostatic Potential
PBC.....	Periodic Boundary Conditions
PME.....	Particle Mesh Ewald
S_p	Soma de Ewald de Ordem p
erfc.....	Função erro Complementar
erf.....	Função erro
FFT.....	Fast Fourier Transform.
RMSD.....	Root Mean Square Displacement
K_B	Constante de Boltzmann
ΔA	Variação da Energia Livre de Helmholtz
ΔG	Variação da Energia Livre de Gibbs
MMGBSA.....	Molecular Dynamics, Generalized Born, Solvent Acessibility
GB.....	Generalized Born
FPPS.....	Farnesil Pirofosfato Sintase
TR.....	Tripanotona Redutase
DHDH.....	Dihidro-Orotato Desidrogenase

C.E.	Classificação Enzimática
PDB	Protein DataBank
UniProt	Universal Protein Resources
BLAST	Basic Local Alignment Search Tool
gaff	general amber force field
CUDA	Computer Unified Device Architecture
AD	escore do AutoDock 4.2
min-MMGBSA	Método para se obter a variação de Energia livre aplicando-se o MMGBSA em uma estrutura minimizada
multi-MMGBSA	Método para se obter a variação de Energia livre obtida aplicando-se o MMGBSA em configurações retiradas de múltiplas trajetórias.
ΔG -min	Variação da energia livre obtida aplicando-se o método min-MMGBSA
ΔG -multi	Variação da energia livre obtida aplicando-se
$\Delta\Delta G$ -min	Valor obtido ao se subtrair o valor de ΔG -min de um ligante do mesmo valor estimado para o ligante cristalográfico de sua enzima.
$\Delta\Delta G$ -multi	Valor obtido ao se subtrair o valor de ΔG -multi de um ligante do mesmo valor estimado para o ligante cristalográfico de sua enzima.
ΔAD	Valor obtido ao se subtrair o valor de AD de um ligante do mesmo valor estimado para o ligante cristalográfico de sua enzima.
RGB	Red, Green and Blue. Padrão para especificar cores.
QM/MM	Quantum Mechanics / Molecular Mechanics. Método de dinâmica molecular utilizando potenciais híbridos quânticos/clássicos.
LIE	Linear Interaction Energy. Método de estimativa de energia livre semelhante ao MMGBSA, mas não exatamente.
QSAR	Quantitative Structure-Activity Relationship. Técnica quimiométrica para obter relações matemáticas entre a estrutura e as propriedades de uma molécula.

RESUMO

As leishmanioses são um conjunto de doenças provocadas por protozoários do gênero *Leishmania sp.* cujo vetor é o mosquito-palha. Elas se caracterizam por lesões na pele e nas mucosas que podem deformar severamente os portadores, sem matá-los. Contudo, as leishmanioses podem evoluir para formas mais graves que atingem os órgãos internos, particularmente o baço e o fígado. Em seus últimos estágios, as leishmanioses são fatais.

O tratamento das leishmanioses é feito sobretudo com antimoniais pentavalentes (notavelmente o estibogluconato de sódio e o antimoniato de meglumina) como fármacos de primeira escolha assistidos por outros antibióticos (anfotericina B, paromomicina e pentamidina) como agentes terapêuticos de segunda escolha. Estas opções de terapêutica, entretanto, sofrem com diversos problemas de efeitos colaterais severos, alto custo, resistência do parasita ou uma combinação qualquer destes fatores.

Diante deste fato, nós avaliamos métodos de estimativa da variação de energia livre de interação entre ligantes e enzimas de interesse medicinal (Tripanotiona Redutase, Farnesil Pirofosfato Sintase e Diidroorotato Desidrogenase) para as quais não existem trabalhos teóricos desenvolvidos neste sentido, e sugerimos aqueles que possam vir a se tornarem ligantes promissores.

Avaliamos ainda a possibilidade de substituir o *docking* molecular convencional, em estudos de triagem virtual, pelo cálculo de estimativa da energia livre feita pela metodologia MMGBSA (*Molecular Dynamics, Generalized Born, Solvent Accessibility*) aplicada a uma única estrutura minimizada do complexo enzima-ligante solvatado. O papel da tecnologia de cálculos em placas gráficas, considerações sobre economia de tempo, exatidão dos valores de afinidade e da capacidade geral dos métodos de apontar bons ligantes foram analisadas e discutidas.

ABSTRACT

Leishmaniasis are a group of diseases caused by protozoans of the *Leishmania sp.* genus. whose vector are the sandflies. These diseases are characterized by lesions in the skin and mucous tissues which can severely impair and deform the patient, without killing them. Leishmaniasis, however, can evolve to more severe forms and attack internal organs, particularly the spleen and liver. In their final stages of progression, leishmaniasis can be fatal.

The treatment of leishmaniasis is carried out mainly with the pentavalent antimonials (remarkably sodium stibogluconate and meglumine antimoniate) as first choice drugs, assisted by other antibiotics (anfotericin B, paromomycin and pentamidine) as second line therapeutic agents. These therapeutic options, however, suffer with a number of problems with severe side effects, high costs, the appearance of parasite resistant strains or any combination of these factors.

Because of these facts, we have evaluated methods to estimate the free energy variation of interaction between ligands and medically interesting enzymes (Trypanothione Reductase, Pharnesyl Pyrophosphate Synthase and Dihydro Orotate Dehydrogenase) for which there are no theoretical works in this sense. We have also suggested, based on the method that performed better in the tests, those ligands which seem promising.

We have evaluated the possibility of replacing the conventional molecular docking, in studies of virtual screening, with free energy estimation calculations done with the MMGBSA (Molecular Dynamics, Generalized Born, Solvent Accessibility) methodology applied to a single minimized geometry of the solvated ligand-enzyme complex. The role of computation in graphics boards processors, time economy, accuracy of affinity values and the overall capacity of the methods to point out the good ligands were analysed and discussed.

1. Introdução

1.1. Leishmanioses

1.1.1. Leishmania. sp.

Leishmanioses são um conjunto de doenças causadas por protozoários do gênero *Leishmania sp.* São parasitas intracelulares obrigatórios que infectam humanos e alguns outros mamíferos. Os vetores da leishmaniose são dípteros (insetos de duas asas) da família Psychodida do gênero *Lutzomyia* (Figura 8) nas américas, conhecido popularmente como mosquito-palha, e pelo gênero *Phlebotomus* no velho mundo. Estimam-se que cerca de 30 das 500 espécies destes dípteros estejam envolvidos com a transmissão das leishmanioses no mundo (KILLICK-KENDRICK, 1999). As fêmeas da espécie são hematofágicas, sendo as responsáveis pela transmissão das leishmanioses.

Os vetores costumam habitar lugares úmidos e escuros, protegidos do vento, entretanto, a interferência do ser humano em seu ambiente tem forçado as populações de mosquito-palha a expandirem seu *hábitat* e conviverem com o progresso humano, isto muitas vezes causa uma maior exposição de populações de seres humanos ao contato com os vetores (SHAW, 2007).



Figura 1: Indivíduo da espécie *Lutzomyia longipalpis*. http://upload.wikimedia.org/wikipedia/commons/5/59/Lutzomyia_longipalpis-sandfly.jpg

O parasita apresenta duas morfologias distintas: A forma promastigota e a forma amastigota. A forma promastigota é alongada, com núcleo central e possui um flagelo locomotor; a forma amastigota é ovalada, não apresenta flagelos visíveis e é encontrada no interior de macrófagos

(Figura 2).

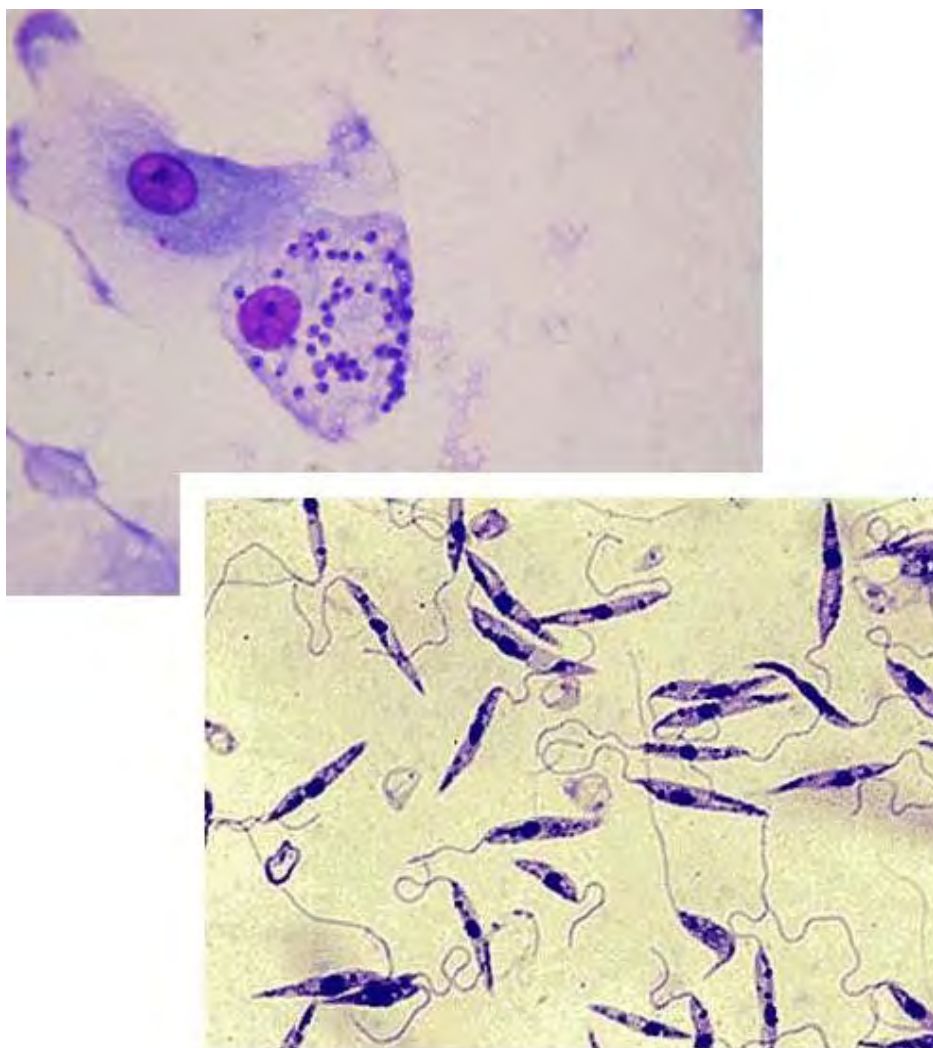


Figura 2: Formas parasitárias de Leishmania sp. Retirado de <http://www.medicinapratica.com.br/wp-content/uploads/2009/07/leishmaniose.JPG>

O vetor infecta-se ao ingerir sangue de um hospedeiro vertebrado cujos macrófagos estejam contaminados com a forma amastigota do parasita. No interior do parasita os macrófagos sofrem lise, liberando os amastigotas. Estes parasitas diferenciam-se em suas formas promastigotas procíclicas e aderem às microvilosidades intestinais do inseto.

Nesta fase, os parasitas multiplicam-se intensamente por divisão binária, assumindo a forma infectante promastigota metacíclica em cerca de alguns dias. Nesta forma, os parasitas não têm a capacidade de realizar divisão binária, mas possuem grande mobilidade. Os promastigotos metacíclicos migram para a probóscide do inseto, e caem na corrente sanguínea do hospedeiro vertebrado através do contato da saliva do inseto com o sangue, durante o repasto sanguíneo (SACKS; KAMHAWI, 2001).

Na corrente sanguínea do hospedeiro os parasitas são então fagocitados pelos macrófagos. No interior dos fagossomos dos macrófagos, as formas infectantes promastigotas diferenciam-se nas formas amastigotas que reproduzem-se por divisão binária até o rompimento da membrana

plasmática do macrófago, causando sua lise e liberando parasitas em formas amastigotas, que por sua vez, são fagocitados por outros macrófagos, recomeçando o ciclo (SACKS;KAMHAWI, 2001) (Figura 3).

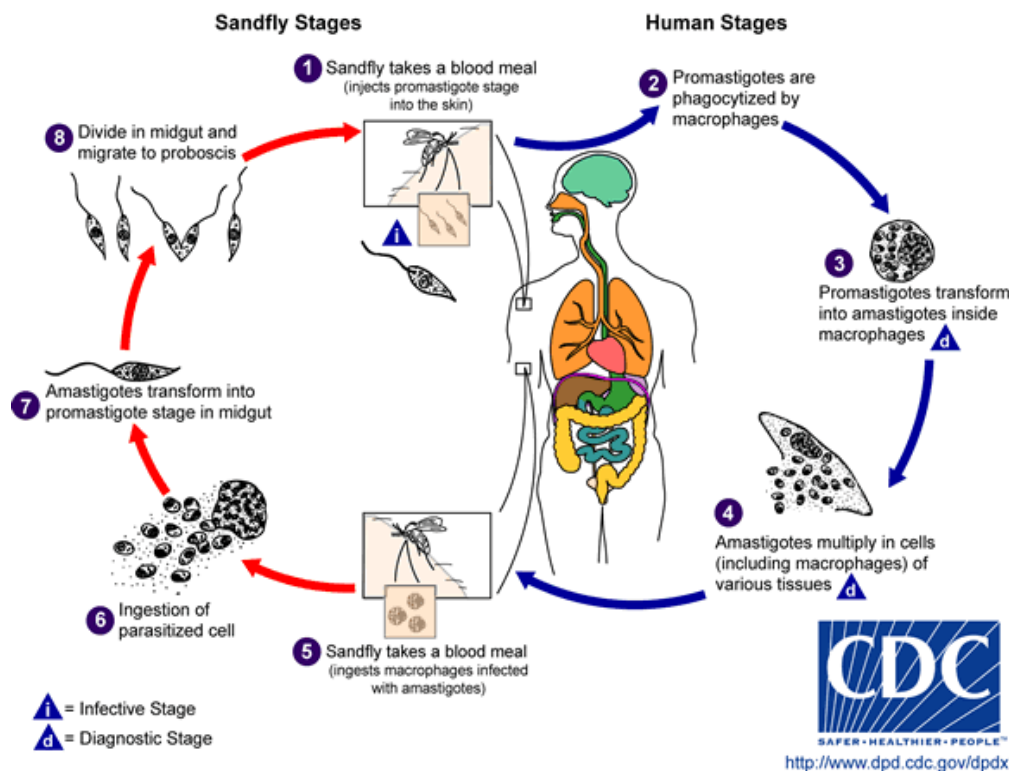


Figura 3: Ciclo evolutivo da Leishmaniose (figura adaptada de <http://dpd.cdc.gov/dpdx/HTML/Leishmaniasis.htm>)

A doença pode se apresentar em até três formas clínicas distintas (Figura 4): a forma tegumentar cutânea (LTC), a forma mucocutânea (LMC) e a forma visceral (LV), assim chamada por causar hipertrofia de órgãos internos, tais como o baço e o fígado. A forma de apresentação clínica depende da espécie do parasita e da imunocompetência do hospedeiro.

No caso das nossas espécies de interesse, a *Leishmania amazonensis*, em geral é associada com as formas cutâneas da doença (LTC e LMC), mas existem relatos de amostras de *Leishmania amazonensis* terem sido isoladas de pacientes de todas as formas clínicas e os parasitas da espécie *Leishmania chagasi* são usualmente associados a casos de leishmaniose visceral (BARRAL et al, 1991).



Figura 4: Tipos de leishmanioses

Segundo dados da Organização Mundial de Saúde, 350 milhões de pessoas, em 88 países, encontram-se em regiões ameaçadas. Além disso, 12 milhões de pessoas estão infectadas, e estima-se que hajam cerca de um a dois milhões de novos casos por ano. Leishmanioses são endêmicas apenas em países subdesenvolvidos e em geral são mais pronunciadas em regiões mais rurais. Este perfil de distribuição contribuiu para tornar as leishmanioses parte do conjunto de doenças negligenciadas.

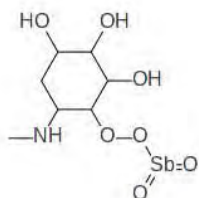
1.1.2. Tratamento

Atualmente, os fármacos de primeira linha disponíveis no mercado para o tratamento desta doenças são os antimoniais pentavalentes: O Estibogluconato de sódio (Pentostam) e o Antimoniato de Meglumina (Glucantime) (figura 5).

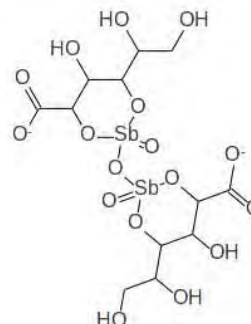
Ambos são tóxicos, apresentam eficácia limitada em seres humanos, atividade teratogênica e efeitos adversos fortes (SANTOS et al, 2008; SUNDAR; CHAKRAVARTY, 2010; ASTELBAUER; WALOCHNIK, 2011). Como a excreção do antimônio na urina é muito rápida, doses altas e intermitentes dos fármacos precisam ser tomadas.

Os efeitos colaterais dos antimoniais pentavalentes incluem artralgia, mialgia, elevação dos níveis de enzima hepatocelular e mudanças nos resultados do eletrocardiograma (ECG) (HERWALDT; BERMAN, 1992; NAVIN et al, 1992). Cardiopatias ocorrem em até 9% dos pacientes e existem casos documentados de mortes decorrentes deste efeitos colateral em particular (HERWALDT; BERMAN, 1992; CHULAY; SPENCER; MUGAMBI, 1985; SUNDAR et al, 1997).

Devido ao desconforto da terapia em virtude das injeções frequentes e os efeitos colaterais fortes, a descontinuação é recorrente em pacientes submetidos aos antimoniais. Devido a este mal uso, formas resistentes do parasita tornaram-se um problema (RATH et al, 2003). A manifestação mais notável deste problema é em Bihar, na Índia, por exemplo, praticamente mais da metade da população não responde aos tratamentos (THAKUR et al, 2001).



GLUCANTIME®
Antimoniato de Meglumina



PENTOSTAM®
Estibogluconato de Sódio

Figura 5: Principais fármacos contra leishmanioses.

Como alternativa ao tratamento com os antimoniais, fármacos já existentes estão sendo testados contra leishmanioses. Pentamidina, Anfotericina B, um antibiótico e um antifúngico, são fármacos de segunda escolha no tratamento das leishmanioses (NATERA et al, 2007; CROFT; BARRET; URBINA, 2005). O Miltefosine, um antineoplásico em uso contra câncer de pele, também já teve documentada sua atividade leishmanicida (SINDERMANN; ENGEL, 2006; CROFT; ENGEL, 2006).

Casos de uso de Anfotericina B no tratamento de Leishmaniose visceral tem mostrado altas taxas de cura (90% a 95%). A maior desvantagem das preparações convencionais de Anfotericina B está nos efeitos adversos fortes que decorrem do seu uso prologado, incluindo febres e calafrios relacionados à infusão, nefrotoxicidade e hipocalcemia, que tornam necessário a aplicação do medicamento em um hospital ou unidade de saúde competente (SUNDAR et al, 2007). Na Índia, mesmo estas preparações são utilizadas rotineiramente em virtude da alta resistências das cepas locais aos antimoniais.

Há, ainda, algumas preparações de Paromomicina aprovadas na Índia para utilização contra leishmanioses (DAVIDSON ;DEN BOER; RITMEIJER, 2009). No entanto, existem cepas resistentes de parasitas utilizando um mecanismo de resistência à Paromomicina, envolvendo a diminuição da absorção do fármaco (MAAROUF et al, 1998). Em virtude da crescente resistência das espécies de parasitas causadoras das leishmanioses, e dos efeitos colaterais fortes, com grandes taxas de descontinuação, é necessário o desenvolvimento inteligente e eficaz de novos fármacos.

1.1.3. Projeto Racional de Fármacos e Triagem Virtual

O processo de desenvolvimento de um medicamento, no entanto, é bastante complexo e

chegar a uma molécula que seja farmacologicamente interessante envolveria pesquisa sobre uma série de propriedades da referida molécula, entre eles testes de farmacocinética, farmacodinâmica, ensaios clínicos e desenvolvimento de tecnologia farmacotécnica para criar formas farmacêuticas adequadas para o consumo em massa, bem como o desenvolvimento de padrões de qualidade.

Com o aparecimento da química combinatória (LAZO; WIPF, 2000; MIERTUS; FASSINA; SENEI, 2000; WINOGRAD; BRAUN, 2001), tornou-se possível sintetizar rapidamente uma grande quantidade de compostos relacionados. Com estas técnicas é possível explorar racionalmente um espaço bem definido de estruturas químicas e observar exatamente qual a relação deste espaço químico com as propriedades das moléculas, inclusive a atividade biológica.

Métodos automatizados de triagem de alto rendimento (*high-throughput screening*) (HANN; OPREA, 2004; EISENSTEIN, 2006; HOWE et al, 2008) e outras técnicas alternativas capazes de testar esta vasta quantidade de compostos surgiram para conseguir processar as vastas bibliotecas produzidas pelos métodos de química combinatória.

Associados a estes métodos experimentais de se explorar este espaço químico produzido pelos métodos de química combinatória, técnicas computacionais capazes de calcular ou prever de alguma forma a atividade biológica de moléculas também foram desenvolvidos.

Métodos deste tipo são utilizados em associação aos métodos experimentais como forma de varrer imensas bibliotecas virtuais em busca de famílias de moléculas promissoras, nos chamados experimentos de triagem virtual (*virtual screening*). Esta associação de métodos computacionais e experimentais faz parte do esquema hierárquico de técnicas normalmente empregadas no processo de descobrimento de novos fármacos (JORGENSEN, 2004).

Muitos destes métodos computacionais baseiam-se nas estruturas dos ligantes e buscam formas eficientes de se extrair informações de suas estruturas (REYMOND et al, 2010; SALUM; ANDRICOPULO, 2009; RIPPHAUSEN; NISIUS; BAJORATH, 2011; LISUREK et al, 2009). Entre estes podemos citar o QSAR (WINKLER, 2001).

No QSAR uma série de descritores geométricos são calculados para as estruturas de interesse e relações numéricas quantitativas são buscadas entre estes descritores numéricos e os valores de atividade das moléculas. Espera-se com isto que esta relação possa ser extrapolada para prever a atividade de uma molécula que não se encontrava no conjunto original utilizado para obter a relação.

Em uma triagem virtual, o que se deseja é uma forma ágil de se calcular certas propriedades para uma biblioteca de moléculas, e então utilizar estes dados para filtrar aqueles grupos de moléculas que, em média, obtiverem melhor resultado nos cálculos. No caso de uma triagem virtual de fármacos, uma propriedade de interesse seria a interação entre os ligantes da biblioteca e um alvo farmacológico específico.

Uma outra abordagem de se realizar uma triagem virtual de fármacos, procura levar em consideração esta propriedade, através de métodos computacionais eficientes de se estimar a afinidade de interação entre a estrutura de uma molécula pequena e a estrutura de um alvo macromolecular. Uma das técnicas mais conhecidas e computacionalmente rápidas para se efetuar este tipo de estimativa é o *docking* molecular (TAYLOR; JEWSBURY; ESSEX, 2002).

Desde sua criação, o *docking* tem sido utilizado para uma variada gama de aplicações em combinação com outros métodos na busca por novos fármacos (DORMAN et al, 2002; SEIFERT, 2009; HUBBARD, 2011). Nesta técnica, um ligante pequeno tem seu espaço conformacional explorado e para cada pose avaliada, a força das interações intermoleculares entre os átomos da pose e a seção do alvo molecular especificada como sítio ativo é calculada.

O cálculo das interações é feito através de funções empíricas, parametrizadas para

reproduzirem dados experimentais. Estas funções paramétricas dos métodos de *docking* não representam uma estimativa formal de variação da energia livre, sendo um índice relativamente arbitrário de afinidade entre o ligante e o alvo.

Formalmente, se quisermos avaliar a interação entre duas moléculas, devemos recorrer à termodinâmica. A grandeza que mede afinidade com que as duas moléculas interagem é essencialmente a constante de equilíbrio da reação mostrada na Figura 6.



Figura 6: Interação Enzima-Ligante

Esta constante de equilíbrio é definida como o quociente da concentração do complexo pelo produto das concentrações dos reagentes (ligante e alvo numa proporção de 1:1), segundo a equação (1)

$$K = \frac{[complexo]}{[ligante] \cdot [alvo]} \quad (1)$$

$$K = e^{-\left(\frac{\Delta G}{RT}\right)} \quad (2)$$

Segundo a equação (2), Essa constante de equilíbrio está associada a uma variação da energia livre de forma que quanto mais negativa a variação da energia, maior a constante de equilíbrio e portanto maior a afinidade.

O problema então pode ser resumido a encontrar o ligante para o qual a variação da energia livre para essa reação seja o mais negativa possível. A metodologia formal de se fazer isto é complexa e passa por cálculos custosos de dinâmica molecular, utilizando os métodos ditos de formalismos exatos, discutidos na fundamentação teórica.

A dinâmica molecular é uma técnica que simula o comportamento físico completo dos átomos do sistema ao longo do tempo e a partir destas simulações, calcula propriedades termodinâmicas macroscópicas utilizando o formalismo da mecânica estatística.

Apesar de não ser a única propriedade calculável com dinâmica moleculares, a variação da energia livre de interação é uma propriedade particularmente complicada de se obter através desta metodologia, sobretudo se a interação calculada ocorre em solução.

Em virtude do esforço computacional envolvido em uma estimativa detalhada da afinidade entre duas moléculas em solução, o *docking* molecular surgiu como uma alternativa para se estimar estas propriedades que fosse rápida o suficiente para ser utilizada para avaliar uma grande quantidade de moléculas em um intervalo de tempo razoável.

Em paralelo ao *docking* molecular, métodos mais rápidos e aproximados para se estimar a variação da energia livre considerando o efeito do solvente foram desenvolvidos. Entre eles encontramos o MMGBSA (SRINIVASAN et al, 1998).

O MMGBSA é um método que utiliza modelos de solvente implícito para calcular as parcelas da variação da energia livre decorrentes da solvatação do sistema, evitando que esta contribuição precise ser calculada diretamente com dinâmica molecular em solvente explícito, resolvendo um ciclo termodinâmico, diminuindo a quantidade de cálculos de dinâmica molecular necessários para se conseguir uma estimativa da variação de energia livre de afinidade entre uma enzima e um ligante.

Este método tem sido utilizado para reavaliar poses produzidas por programas de *docking* molecular (THOMPSON et al, 2007; RASTELLI et al, 2009) com a finalidade de melhorar o ordenamento das moléculas pelo *docking*, potencialmente diminuindo o número de falsos positivos/negativos.

Recentemente, foi sugerido que seria possível realizar estimativas da variação de energia livre de interação utilizando o método MMGBSA para uma única estrutura do complexo enzima-ligante (RASTELLI et al 2010).

A ideia central por trás desta técnica é que quando analisamos o comportamento dinâmico de um sistema, as médias temporais de suas propriedades sobre todo o intervalo de tempo observado são muito semelhantes às propriedades de uma geometria estática que corresponde a um mínimo de energia.

Ao ser utilizado sobre uma única geometria, o MMGBSA torna-se um forte candidato a ser um método de triagem virtual. Sem o empecilho dos cálculos de dinâmica molecular, o MMGBSA sobre uma estrutura minimizada pode ser utilizado para triar uma série de compostos. A metodologia do cálculo em uma única geometria também já foi utilizada para reavaliar poses produzidas por programas de *docking* (GRAVES et al, 2008; HAIDER; BERTRAND; HUBBARD, 2011).

Neste trabalho, analisamos a possibilidade de substituir completamente a avaliação de afinidade via *docking* molecular pela avaliação via MMGBSA a partir de um única estrutura minimizada do complexo enzima-ligante na triagem de pequenos conjuntos de ligantes em enzimas de *Leishmania chagasi* e *Leishmania amazonensis*. Sobretudo quanto à exatidão dos valores, a capacidade dos métodos de ordenarem os ligantes segundo a afinidade e também o tempo de cálculo.

1.2. Fundamentação Teórica

1.2.1. *Docking* Molecular

Numa visão inicial, *docking* molecular é uma técnica computacional desenvolvida para calcular como duas moléculas se encaixam no espaço tridimensional. Em particular, o *docking* surgiu como uma forma de calcular a forma preferencial com que um ligante pequeno acomoda-se no sítio ativo de uma macromolécula biológica e estimar sua afinidade de ligação (YURIEVA; AGOSTINO; RAMSLAND, 2001; BROOIJMANS; KUNTS, 2003).

O *docking* é um procedimento que consiste de duas partes. A busca pelo espaço dos graus de liberdades configuracionais e conformacionais (que envolve o posicionamento do ligante no sítio ativo), e a atribuição de uma pontuação a essas conformações segundo uma função-escore (RAMSLAND, 2001; BROOIJMANS; KUNTS, 2003; WASZKOWICZ; CLARK; GANCIA, 2011). Esta função-escore serve para estimar a afinidade de ligação entre ligantes e macromoléculas, ranqueando assim múltiplos ligantes (WASZKOWICZ; CLARK; GANCIA, 2011), e é a função a ser otimizada pelos algoritmos de *docking* molecular.

A busca por calibrar a razão exatidão/rapidez dos algoritmos de *docking* ao longo dos anos deram origem a uma vasta gama de métodos e implementações (MOITESSIER et al, 2008).

Mais de 60 programas de *docking* são conhecidos (BEGLOV et al, 2009), cada um deles se valendo dos mais variados algoritmos de busca para solucionar este problema de otimização (DESJARLAIS; CUMMINGS; GIBBS, 2007; KONTOYIANNI et al, 2008). Poucos deles concordam entre si sobre como fazer as etapas do *docking*, esse fato pode ser parcialmente resolvido se for utilizada uma função consenso que são construídas a partir de vários algoritmos e/ou programas diferentes.

A variedade surge já na fase de posicionar o ligante no sítio ativo. Pelo menos duas abordagens existem. Na primeira, o ligante é posicionado a partir de conhecimento da configuração do sítio ativo. Nestes, os grupos na superfície do sítio ativo são mapeados e grupos complementares dos ligante são orientados da melhor forma possível para satisfazer estes grupos. O programa DOCK (EWING et al, 2001; MOUSTAKAS et al, 2006), por exemplo, preenche o sítio ativo com esferas que mimetizam propriedades de grupos de átomos do sítio ativo (ex. porção doadora de ligação de hidrogênio, cargas positivas, etc.) e depois as agrupam para revelar as características do sítio ativo.

A segunda abordagem utiliza a pontuação da própria função-escore para decidir onde colocar o ligante. Permite-se que o ligante passe por qualquer ponto da caixa que define o sítio ativo, inclusive os pontos mais distantes. Eventualmente, o programa converge para as poses mais próximas do sítio ativo e com melhor escore (ABAGYAN; TOTROV; KUZNETSOV, 1994; VENKATACHALAM et al, 2003), ou então tratam de eliminar os pontos mais desfavoráveis de posicionamento depois de uma triagem preliminar (FRIESNER et al, 2004). Em geral estes métodos são mais lentos que os primeiros, mas exploram melhor as possibilidades de posicionamento.

Uma vez que o posicionamento do ligante esteja resolvido, a busca pelo espaço conformacional do ligante é o próximo passo. Mais uma vez, dois métodos se mantiveram ao longo do tempo como as formas mais comuns de se fazer isso. O primeiro deles consiste em produzir uma

biblioteca de confôrmeros do ligante posicionado e calcular o escore de cada uma dessas conformações de forma rígida. O FRED, o módulo de *docking* do pacote OpenEye utiliza essa metodologia (VIGERS; RIZZI, 2004).

Essa técnica é muito eficiente porque o *docking* rígido é bastante rápido e nenhuma exploração adicional do espaço conformacional é realizada durante o cálculo do escore. Além disso, como a biblioteca de confôrmeros é externa ao procedimento de pontuação é possível gerar esta biblioteca através de métodos externos ao programa de *docking* em si, da forma como for mais apropriado. Em compensação, é preciso gerar uma quantidade razoável de conformações para garantir que a conformação ótima esteja entre elas (WASZKOWICZ; CLARK; GANCIA, 2011).

Na segunda técnica de explorar o espaço conformacional, mais comum e mais nova que a primeira, o cálculo do escore é parte integrante da exploração conformacional, e serve de guia para o procedimento de otimização. Programas como o AutoDock (MORRIS et al, 1998) e o GOLD (JONES; WILLET; GLEN, 1995; JONES et al, 1997) utilizam este tipo de técnica.

Dependendo da função-escore, estes algoritmos podem ser mais lentos que os primeiros, mas precisam produzir menos poses para convergirem em soluções ótimas.

Aqui se encontra a principal fraqueza do *docking* molecular. Estudos indicam que os algoritmos de exploração conformacional até conseguem gerar boas conformações para serem analisadas, entretanto, as funções-escore dificilmente conseguem ordená-las adequadamente, gerando uma série de falsos positivos e falsos negativos (DESJARLAIS; CUMMINGS; GIBBS, 2007; WARREN et al, 2006).

Para ser boa, uma função-escore precisa ser capaz de reconhecer uma conformação que tenha grande possibilidade de ser a pose real de interação entre o ligante e a macromolécula e ela deve atribuir um alto valor de *score* para esta pose. O escore deve ser capaz de separar ligantes de não-ligantes e produzir uma sequência ordenada das melhores poses. Para tentar fazer isto, diversas técnicas estão implementadas nos diversos programas de *docking* molecular.

Apesar de muitos métodos e formas de funções-escore existirem, a maior parte delas estão contidas no conjunto das funções-escore empíricas. Estas funções empíricas são criadas baseadas nas informações contidas nas crescentes bases-de-dados de estruturas cristalográficas de proteínas, das quais o Protein Data Bank (o PDB) (BERNSTEIN et al, 1977) é o exemplo mais notável. Estas funções se dividem em duas categorias. Funções baseadas em regressões e funções baseadas em conhecimento.

As funções baseadas em regressão são construídas otimizando-se os parâmetros de formas funcionais conhecidas para que elas relacionem bem as afinidades de ligação experimentais com descritores geométricos extraídos das estruturas cristalográficas dos complexos. O ChemScore (ELDRIDGE et al, 1997) é um exemplo de uma função-*score* baseada em regressão e está implementada no programa GOLD.

O problema deste tipo de função é que sua eficácia está limitada ao conteúdo do conjunto de calibração. Assim, funções diferentes apresentam resultados bem distintos que podem ter a ver com: o algoritmo de calibração, a quantidade de dados utilizados na calibração da função ou à confiabilidade dos dados coletados. No fim das contas, é difícil estimar quando uma função é "bem" calibrada.

Funções baseada em conhecimento também utilizam estruturas cristalográficas de complexos proteína-ligante para serem construídas. Entretanto, ao contrário de relacionar afinidades experimentais a propriedades geométricas do sistema, estas funções coletam os descritores geométricos e assumem que aqueles traços que ocorrem mais vezes provavelmente são os mais favoráveis. Um potencial médio é então construído para conferir altos scores a complexos que

apresentem descritores favoráveis. Um exemplo é a função PMF (MUEFFE; MARTIN, 1999), implementada no programa SciGress Explorer, da Fujitsu (FUJITSU SCIGRESS EXPLORER, 2008).

Devido à rapidez com que realiza suas estimativas (YURIEVA; AGOSTINO; RAMSLAND, 2001), o *docking* tem se tornado o método preferencial usado em triagem virtual tanto em indústrias como na academia (YURIEVA; AGOSTINO; RAMSLAND, 2001; WASZKOWICZ; CLARK; GANCIA, 2011). Apesar dos avanços tecnológicos e conceituais que o *docking* tem passado nestes últimos anos, entretanto, outros problemas ainda prejudicam sua exatidão e põem em dúvida sua confiabilidade.

Em geral, boa parte dos programas de *docking* utilizam como aproximação fundamental em seus cálculos o modelo de chave flexível e fechadura rígida (semi-flexível) (YURIEVA; AGOSTINO; RAMSLAND, 2001). Isto é, os graus de liberdade do alvo molecular não são considerados quando minimiza-se a função-escore.

Entretanto, sabe-se que a interação entre duas moléculas é um processo que afeta ambas as moléculas (YURIEVA; AGOSTINO; RAMSLAND, 2001) e que desconsiderar isto nas aproximações faz com que as conformações geradas pelo *docking* não tenham o peso que realmente deveriam ter na otimização das interações.

A flexibilidade é incluída apenas nos algoritmos mais sofisticados, dos quais o maior exemplo é o Rosetta (DAVIES et al, 2009; DAVIS; BAKER, 2009;). Isso porque a adição da flexibilidade do alvo molecular ao cálculo aumenta exponencialmente o tempo de cálculo e torna o *docking* rapidamente impraticável (YURIEVA; AGOSTINO; RAMSLAND, 2001).

Adicionalmente, o desempenho dos resultados varia amplamente segundo o programa selecionado. Estudos têm chegado à conclusão de que é bastante improvável ordenar confiavelmente um grupo de ligantes em um estudo de triagem virtual com *docking* (LEACH; SHOICHET; PEISHOFF, 2006; HUANG; SHOICHET; IRWIN, 2006).

Devido a essa não consideração dos graus de liberdade do alvo molecular e do solvente, a exatidão dos métodos de *docking* molecular tem se mostrado prejudicada (LEACH; SHOICHET; PEISHOFF, 2006; HUANG; SHOICHET; IRWIN, 2006; YOSHIFUMI, 2010).

1.2.2. Dinâmica Molecular

O conhecimento completo do comportamento de um sistema molecular pode ser obtido, em princípio, resolvendo-se a equação de Schrödinger dependente do tempo (3), que governa a dinâmica de todas as partículas no sistema, tanto elétrons quanto núcleos.

$$\hat{H}\Psi = i\hbar \frac{\partial \Psi}{\partial t} \quad (3)$$

Para conseguir resolver esta equação, entretanto, nós nos valem da aproximação de Born-Oppenheimer, que desacopla o movimento nuclear do movimento dos elétrons. Neste caso, as posições nucleares entram como parâmetros fixos para os quais a equação eletrônica é resolvida. Isto gera uma solução ótima para cada configuração específica de núcleos no espaço.

A forma como a energia da função de onda eletrônica varia segundo as posições dos núcleos descreve uma hiper superfície de dimensão 3N, conhecida como superfície de energia potencial.

Em uma amostra, entretanto, uma molécula não está em uma única configuração o tempo inteiro. Molécula colidem umas com as outras, átomos vibram, ângulos se deformam e elétrons e núcleos interagem o tempo inteiro. Este comportamento, essa *dinâmica* do sistema faz com que em um dado momento, uma molécula esteja em um estado e em um momento seguinte ela esteja em outro.

As transições entre um estado microscópico e outro são, entretanto, muito rápidas. Muito mais rápidas que qualquer medição macroscópica de propriedades termodinâmicas clássicas como entropia, capacidade calorífica e energia livre. Por conta disto, as medições macroscópicas são frequentemente médias tomadas sobre uma grande quantidade de configurações microscópicas, chamadas microestados.

As transições entre um microestado e outro dependem em grande parte da superfície de energia potencial do sistema, mas as propriedades macroscópicas dependem de como as moléculas encontram-se distribuídas entre os vários estados de energia microscópicos durante a medição. A função que descreve a ocupação de cada nível de energia pelas moléculas é chamada *função de partição*.

1.2.2.1. Ensembles e Funções de Partição

As propriedades termodinâmicas estão conectadas às propriedades macroscópicas através de uma função que descreve como as inúmeras moléculas de um sistema macroscópico se repartem entre os níveis de energia acessíveis ao sistema.

A função de partição é a formalização matemática do objetivo final da termodinâmica estatística, que seria a descrição da termodinâmica de um sistema macroscópico a partir de informações microscópicas.

Para conseguir uma expressão para a função de partição vamos partir de um conceito que surgiu nos trabalhos de Boltzmann, Maxwell e Gibbs, e este seria o conceito de *ensemble*. Um ensemble nada mais é que uma coleção de inúmeras cópias do sistema, cada uma em um dos inúmeros microestados com a restrição de que todos os estados microscópicos correspondam à mesma condição macroscópica. Um conceito ilustrado na Figura 7.

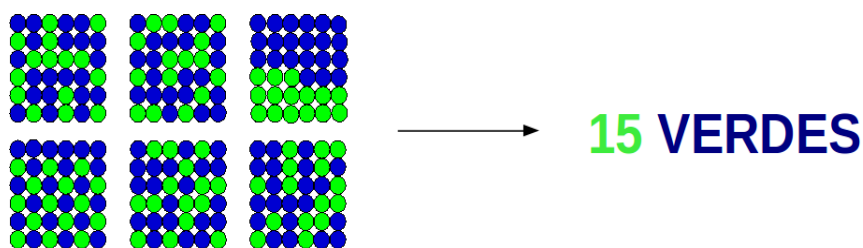


Figura 7: Relação Termodinâmica Estatística/Clássica

Considere, agora, o ensemble de um sistema com o número de partículas N , um volume V e uma temperatura T . Isso é possível de ser visualizado considerando que cada microestado está fechado dentro de um volume V com paredes rígidas, porém diatérmicas, e todos os microestados estão envoltos em um banho térmico de temperatura T de modo que no estado de equilíbrio, todos os microestado do ensemble irão se encontrar na mesma temperatura e depois envolvemos o

ensemble com fronteiras adiabáticas.

Delineado desta forma, um ensemble NVT com A microestados possui $A \cdot N$ moléculas e energia total E . A termodinâmica estatística então postula que a média das propriedades microscópicas sobre o ensemble representa a propriedade termodinâmica. De forma que $\langle E \rangle = E_{\text{termodinamica}}$.

Pelo fato de a temperatura ser constante, a quantidade de microestados em cada valor diferente de energia é assimétrica de modo que teremos que analisar todos os valores de energia possíveis. Pensando em estados discretos de energia, podemos dizer que o estado 1 têm energia ε_1 e existem a_1 microestados com esta energia, no ensemble. Os termos a são chamados de “números de ocupação” pois denota a quantidade de microestados *ocupando* aquele estado particular de energia.

A função que descreve como objetos se *distribuem* em grupos discretos ou contínuos é chamada (pasmem) de *distribuição*. Ou seja, é uma função que relaciona os grupos em questão (no nosso caso, microestados de energia) com os números de ocupação.

A função de partição é a distribuição que relaciona os microestados de energia com seus números de ocupação. Como calcular os valores de energia envolve um procedimento já conhecido, o desafio consiste em encontrar os números de ocupação a_j para o espectro de todos os valores de energia possíveis.

Evidentemente, estes números de ocupação devem obedecer às seguintes restrições:

$$\begin{aligned} A &= \sum_i a_i \\ E &= \sum_i a_i \varepsilon_i(N, V) \end{aligned} \quad (4)$$

Que têm a ver com as características do ensemble que estamos tentando descrever. A configuração dos microestados de energia acessíveis ao sistema dependem profundamente da intensidade do confinamento do sistema e da quantidade de partículas interagentes.

Acontece, no entanto, de existirem inúmeras formas de se distribuir os sistemas do ensemble entre os níveis de energia de modo a satisfazer as condições (4). *A priori*, pelo menos, não há porque considerar que alguma delas é mais correta que as outras.

Vamos definir um vetor $\mathbf{a}$ de tal forma que $\mathbf{a} = (a_1, a_2, a_3, \dots, a_i)$. Ou seja, o vetor $\mathbf{a}$ é a representação de uma distribuição.

A quantidade $W(\mathbf{a})$ de formas diferentes de se distribuir os A sistemas **distinguíveis** de um ensemble entre k microestados de energia de tal forma que haja a_i sistemas no nível de energia $\varepsilon_i(N, V)$ é:

$$W(\mathbf{a}) = \frac{A!}{\prod_k a_k!} \quad (5)$$

A ideia de que os sistemas em questão são distinguíveis entre si é justamente a ideia de ensemble. As propriedades microscópicas dos sistemas dentro de um ensemble são diferentes, de modo que podemos distingui-los uns dos outros, mas todos eles estão associados a propriedades

macroscópicas idênticas.

O número W conta quantas formas diferentes existem de se organizar os sistemas do ensemble sem alterar os números de ocupação. Na matemática, a forma **(5)** de se distribuir e arranjar objetos entre grupos distintos é chamada de distribuição multinomial.

A fração da população total dos microestados que se encontra em um nível i de energia, para uma determinada distribuição, é dada por $P_i^a = \frac{a_i}{A}$.

Perceba que se fecharmos os olhos e escolher um sistema do ensemble aleatoriamente, a probabilidade de tomarmos um sistema que se encontra em um determinado microestado i de energia ε_i é idêntica à fração P_i^a .

Portanto, a probabilidade de tomarmos um sistema no i -ésimo nível de energia considerando-se todas as distribuições é $P_i = \frac{\bar{a}_i}{A}$. Onde $\bar{a}_i$ é a média do coeficiente a_i em todas as distribuições.

As leis da probabilidade define que se considerarmos uma variável aleatória discreta t , que pode assumir N valores $(t_1, t_2, t_3, \dots, t_N)$, com probabilidades de ocorrência de

$$(p(t_1), p(t_2), p(t_3), \dots, p(t_N)) \text{ a média } \bar{t} \text{ desta variável é dada por } \bar{t} = \frac{\sum_i^M t_i p(t_i)}{\sum_i^M p(t_i)}.$$

Neste caso, estamos fazendo a média do coeficiente a_i ao longo de todas as distribuições possíveis. O que quer dizer que a função $p(a_i)$ que estamos procurando é um função que nos diz qual a probabilidade do valor do número de ocupação do i -ésimo nível de energia ser a_i , dentre todos os valores que ele pode ter ao longo de todas as distribuições possíveis.

Vimos que a função $W(a)$ representa o número de formas diferentes de se combinar os diferentes sistemas do ensemble de forma que os coeficientes a_i permaneçam iguais, alterando-se apenas quais sistemas em particular ocupam cada nível.

Dito de outra forma, o número $W(a)$ de uma distribuição diz de quantas formas diferentes podemos organizar os sistemas do ensemble usando esta distribuição. Ao somar todos os W para todas as distribuições possíveis a , teremos a quantidade total de todas as formas possíveis que existem de se organizar os sistemas entre os microestados, sem que as propriedades macroscópicas sejam alteradas.

Segue-se daí que a probabilidade de que um número de ocupação tenha o valor a_i , dentre todos os valores que ele pode ter é, de fato, $p(a_i) = \frac{W[a_i]}{\sum_a W[a]}$ onde a_i é a distribuição onde o número de ocupação do i -ésimo microestado de energia é a_i .

Assim, se considerarmos que o valor a_i depende da distribuição em questão, de modo que $a_i = a_i(a)$, o valor médio $\bar{a}_i$ sobre todas as distribuições possíveis é:

$$\bar{a}_i = \frac{\sum_a a_i(a) W[a]}{\sum_a W[a]}$$

O que implica que:

$$P_i = \frac{1}{A} \frac{\sum_a a_i(a) W[a]}{\sum_a W[a]} \quad (6')$$

É importante conhecer essa probabilidade exatamente porque o *valor esperado* de qualquer propriedade M no ensemble é dada por

$$\langle M \rangle = \sum_i M_i P_i \quad (6).$$

E este valor esperado é a propriedade macroscópica do sistema.

Sabemos que a função W nos dá uma ideia da probabilidade de que os números de ocupação de um sistema tomado ao acaso sejam aqueles de uma distribuição x ou y . O que temos de fazer para encontrar a distribuição com maior influência nos coeficientes é encontrar os números de ocupação para os quais W é máximo. Isto é, encontrar a distribuição mais provável.

Isto pode ser feito encontrando-se os números de ocupação que correspondem a pontos críticos da função W . Um ponto crítico é um ponto onde a derivada primeira é zero.

$$\frac{d}{da_k} W = 0$$

É possível fazer isto diretamente a partir da definição (5), no entanto, é mais conveniente e matematicamente permitido otimizar os número a_k em função do $\ln W$ ao invés da função original. Desta forma $\ln W = \ln(A!) - \sum_k \ln(a_k!)$.

Para simplificar a demonstração, vamos diminuir a quantidade de níveis de energia para 2. Isto facilita a visualização da resolução mas não invalida a extrapolação para mais valores. Sendo assim, nossa expressão para o $\ln W$ diminui para:

$$\ln W = \ln(A!) - \ln(a_1!) + \ln(a_2!)$$

Agora, lembre-se de que estamos assumindo que A e os números de ocupação são muito, muito grandes. O que faz sentido para um sistema macroscópico com quantidades de moléculas da ordem de 1×10^{23} . Para número deste tamanho, é impraticável realizar as operações fatoriais no

interior dos logaritmos.

Para resolver este problema, utilizamos a aproximação proposta pelo matemático escocês James Stirling (Figura 8) de que $\ln N! = N \ln N - N$, onde N é um número muito grande.

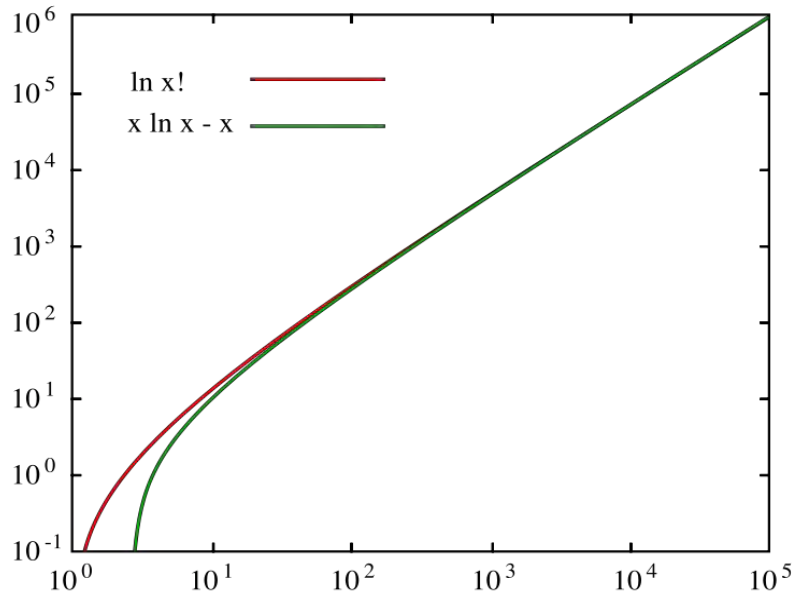


Figura 8: A aproximação de Stirling

Estamos também fazendo uso da relação $a_2 = A - a_1$, que é consequência de estarmos utilizando uma distribuição multinomial. Esta seleção é arbitrária, poderia ter sido feito com o a_2 , também. Fazendo uso desta aproximação, temos que $\ln W = A \ln A - A - a_1 \ln a_1 - a_1 + (A - a_1) \ln (A - a_1) - A - a_1$.

Derivando W em função dos coeficientes temos que Para conseguir uma expressão para a derivada de W em função do coeficiente a_1 temos que:

$$\frac{d}{da_1} \ln W = \ln A - \ln a_1 - \frac{d}{da_1} (A - a_1) \ln (A - a_1) - 2$$

É fácil demonstrar, no entanto, que o termo $\frac{d}{da_1} (A - a_1) \ln (A - a_1)$ é igual a zero. E portanto a expressão final para $\ln W$ é $\frac{d}{da_1} \ln W = \ln A - \ln a_1 - 2$

Para encontrar os pontos críticos desta função procuramos os coeficientes onde

$\frac{d}{da_1} \ln W = 0$. Ao fazer isto, encontramos que os coeficientes que satisfazem esta equação são dados por $a_0 = \frac{A}{k}$.

E vemos que a distribuição multinomial mais provável é aquela onde todos os números de ocupação são iguais a a_0 . Agora que conhecemos um ponto de máximo na superfície de $\ln W$, podemos calcular o valor de W para os pontos próximos utilizando uma série de Taylor (7).

$$\ln W(a) = \ln W(a_0) + \frac{[\ln W]^{(1)}(a_0)}{1!}(a-a_0) + \frac{[\ln W]^{(2)}(a_0)}{2!}(a-a_0)^2 + \dots + \frac{[\ln W]^{(n)}(a_0)}{n!}(a-a_0)^n \quad (7)$$

Mas como nós ainda não temos uma forma de calcular infinitos termos, teremos que nos contentar com uma aproximação razoável, truncando a série em seu terceiro termo (seria uma $T_3(a)$, se fôssemos traçar um paralelo com a nomenclatura da figura).

$$\ln W(a) \approx \ln W(a_0) + \frac{[\ln W]^{(1)}(a_0)}{1!}(a-a_0) + \frac{[\ln W]^{(2)}(a_0)}{2!}(a-a_0)^2 \quad (7')$$

A própria definição de a_0 parte da premissa que a derivada primeira de $\ln W$ é zero nestes pontos, portanto $[\ln W]^{(1)}(a_0) = 0$, de forma que o segundo termo da equação desaparece. Ficamos então com $\ln W(a) \approx \ln W(a_0) + [\ln W]^{(2)}(a_0) \frac{1}{2}(a-a_0)^2$. Onde o termo $[\ln W]^{(2)}(a_0)$ significar o valor da expressão encontrada para $\frac{d^2}{da^2} \ln W(a)$ no ponto a_0 .

Como a derivada segunda de $W(a)$ é $\frac{d^2}{da^2} \ln W(a) = \left(\frac{a-A}{Aa} \right)$, podemos calcular $[\ln W]^{(2)}(a_0)$ considerando que $a_0 = A/k$ e obtemos que $[\ln W]^{(2)}(a) = \left(\frac{1-k}{A} \right)$.

Como estamos utilizando o exemplo representativo de $k=2$, então $[\ln W]^{(2)}(a) = -\frac{1}{A}$. Com isso, nossa série de Taylor para $\ln W$ toma uma forma mais elegante, e através de algumas manipulações algébricas podemos chegar a um jeito simples para se estimar a forma como o próprio W se comporta ao distanciar-se da distribuição ótima:

$$W(a) \approx \frac{A!}{a_0^k} \exp \left[-\frac{1}{2A}(a-a_0)^2 \right] \quad (8)$$

A expressão (8) é uma Gaussiana. A Gaussiana tem a propriedade de ser praticamente zero quando a diferença $a-a_0$ é muito maior que o desvio padrão (da ordem de $A^{1/2}$). Para as escalas a com a qual estamos trabalhando (por volta de 10^{20}), nossa Gaussiana estará bastante concentrada em um intervalo centrada no valor de $a_0 \pm n 10^{10}$, onde n são alguns poucos múltiplos

inteiros.

Nesta escala de quantidade, a Gaussiana plotada seria praticamente zero em qualquer intervalo fora de $a_0 \pm n10^{10}$, e formaria um pico agudo no valor de a_0 . Isto mostra que a distribuição mais provável é também a que mais contribui para descrever o sistema e que esta descrição é tão significativa, que praticamente só precisamos dela para descrevê-lo adequadamente.

Substituindo estas informações na equação (6') significa que $P_i = \frac{a_0^{[i]}(a)}{A}$.

Sendo assim, a probabilidade de que tomando um sistema ao acaso, ele esteja no i -ésimo microestado de energia depende apenas do número de ocupação $a_0^{[i]}$ daquele estado de energia na distribuição mais provável, e não mais de uma média sobre todas as distribuições possíveis.

Esta forma para a distribuição mais provável, no entanto, não considera as condições em (4). O que significa dizer que o valor ótimo, quando calculado levando em consideração as restrições de um ensemble canônico, não será o ponto onde todos os números de ocupação são iguais. Entretanto, a mesma propriedade se aplica e, uma vez conhecendo-se a forma geral a distribuição mais provável para o ensemble, ela também será tão mais provável que as outras que será a única necessária para descrevê-lo.

O que fazemos, então, é minimizar $\ln W$ levando-se em consideração as restrições em (4). A forma de fazer isto é análoga ao procedimento de encontrar o menor valor em uma parábola que também esteja contido em uma reta, por exemplo. A isto se chama *vínculos*.

Existe um procedimento relativamente simples para se minimizar uma função com vínculos, é o chamado método dos Multiplicadores Indeterminados de Lagrange. Os quais, aliás, queremos determinar. Neste método, não minimizamos diretamente $\ln W(a)$, ao invés disso, minimizamos uma função L de tal forma que $L = \ln W - \alpha \sum_i a_i - \beta \sum_i a_i \varepsilon_i(N, V)$, onde os termos em azul e vermelho são os multiplicadores indeterminados de Lagrange. O que se faz é encontrar os valores dos números de ocupação para os quais a derivada de L é zero.

Mais uma vez utilizando a definição (5) juntamente com a relação de Stirling, chegamos à seguinte relação $-\ln(a_0^{(i)}) - \alpha - 1 - \beta \varepsilon_i(N, V) = 0$. Onde $a_0^{(i)}$ significa o número de ocupação do i -ésimo microestado de energia na distribuição mais provável. Utilizando as propriedades de exponencial e logaritmos, é possível concluir que $a_0^{(i)} = \exp[-\alpha'] \exp[-\beta \varepsilon_i(N, V)]$.

Esta expressão descreve os números de ocupação de cada nível de energia na distribuição mais provável. Para encontrar a distribuição das probabilidades P_j de cada nível de energia, fazemos:

$$\begin{aligned} A &= \sum_i a_0^{(i)} = \sum_i \exp[-\alpha'] \exp[-\beta \varepsilon_i(N, V)] \\ A &= \exp[-\alpha'] \sum_i \exp[-\beta \varepsilon_i(N, V)] \\ P_i &= \frac{a_0^{(i)}}{A} = \frac{\exp[-\alpha'] \exp[-\beta \varepsilon_i(N, V)]}{\exp[-\alpha'] \sum_i \exp[-\beta \varepsilon_i(N, V)]} \end{aligned}$$

Ao fazer isto, o multiplicador indeterminado α , incluído em α' , desaparece. O parâmetro β está conectado à temperatura termodinâmica através da constante de Boltzmann

K_B de tal forma que $\beta = 1/K_B T$.

O denominador soma por todas as equações do ensemble canônico e é chamada **função de partição canônica**. O numerador foi chamado de peso ou ponderação de Boltzmann, em homenagem ao físico austríaco Ludwig Boltzmann, um dos pais fundadores da termodinâmica estatística.

$$Q = \rho_{NVT}(\epsilon_i) = \frac{\exp\left[-\frac{1}{K_B T} \epsilon_i(N, V)\right]}{\sum_i \exp\left[-\frac{1}{K_B T} \epsilon_i(N, V)\right]} \quad (9)$$

Se definirmos a função de partição canônica como sendo $Q(N, V, T)$, qualquer propriedade mecânica de um sistema canônico pode ser calculada seguindo a relação (6):

$$\langle P \rangle = \frac{1}{Q} \sum_i p_i \exp\left[-\frac{1}{K_B T} \epsilon_i(N, V)\right] \quad (10)$$

Desde que consigamos calcular a propriedade microscópica p_i .

A forma exata da distribuição depende da temperatura, o que significa dizer que em diferentes temperaturas, a ocupação ótima de cada microestado de energia será diferente.

Em um sistema que está em equilíbrio termodinâmico, as moléculas individuais que compõem o sistema podem mudar de um microestado para outro, mas a quantidade de moléculas que entram em um microestado é a mesma que sai, de modo que não há uma alteração no valor medido da propriedade macroscópica, pois não há variação na população média de moléculas ocupando cada nível de energia.

No equilíbrio termodinâmico, portanto, a função de partição é invariante no tempo. Quando isto acontece, qualquer média temporal da propriedade de interesse tomada sobre um conjunto de microestados do sistema é equivalente ao valor esperado da propriedade no ensemble. E o valor esperado da propriedade no ensemble conecta-se à propriedade macroscópica. Esta é a hipótese ergódica.

Se pudermos simular as transições de microestados que ocorrem durante a progressão temporal de um sistema e guardar informações sobre estas transições, podemos obter a propriedade macroscópica deste sistema. No entanto, o algoritmo de simulação deve ser construído de modo que as probabilidades das transições na simulação sejam consistentes com a distribuição do ensemble, para garantir que a média temporal será feita sobre um somatório que incluirá microestados mais prováveis em maior quantidade.

Se o sistema simulado estiver em equilíbrio termodinâmico e o algoritmo for adequado para amostrar o ensemble, a relação ergódica nos garante que para um tempo muito longo de simulação, a média temporal e a média no ensemble serão idênticas. Isto porque como a distribuição de partículas entre os microestados é invariante, eventualmente o sistema passará, ao longo do tempo, por todas as transições possíveis entre os microestados.

1.2.2.2. Mecânica Clássica

Embora seja teoricamente possível adquirir informações sobre o ensemble de configurações nos estados nucleares resolvendo-se a equação de onda para os núcleos (e um potencial eletrônico paramétrico), esta técnica pode ser aplicada apenas para sistemas muito simples. A solução das equações de onda nucleares rapidamente se torna impraticável à medida que o tamanho do sistema aumenta.

Felizmente, mesmo que seja muito difícil ou até mesmo impossível estudar a dinâmica dos sistemas utilizados neste trabalho aplicando-se o formalismo da mecânica quântica, ainda é possível amostrar as configurações nucleares propagando o sistema no tempo segundo equações de mecânica clássica.

Embora seja apenas uma aproximação da dinâmica nuclear real, simulações de mecânica clássica fornecem descrições bastante exatas em muitos casos, apesar de ignorar alguns fenômenos que são importantes em certas circunstâncias. Por exemplo, o tunelamento no tratamento de partículas muito leves ou a vibração de ponto zero em temperaturas muito baixas.

Por este motivo, a aplicação da mecânica clássica está restrita a sistemas onde o comportamento quântico médio do movimento nuclear possa ser confiavelmente modelado a partir da mecânica clássica. Formalmente, a aproximação clássica descreve bem a dinâmica de um sistema atômico-molecular se o comprimento de onda térmico de De Broglie (11) for muito menor que a distância média de separação entre os núcleos do sistema.

$$\Lambda = \sqrt{\frac{2\pi\hbar^2}{M k_B T}} \quad (11)$$

A mecânica clássica pode ser formalizada definindo-se um Hamiltoniano clássico do sistema, que é uma função das coordenadas de posição e de momento linear (12).

$$H(\mathbf{p}_i, \mathbf{r}_i) = \sum_{i=1}^N \frac{1}{2m_i} \mathbf{p}_i^2 + U(\mathbf{r}_i) \quad (12)$$

Os vetores $\mathbf{p}_i$ representam as coordenadas de momento linear para cada partícula i e os vetores $\mathbf{r}_i$ representam as coordenadas de posição. Ao todo, portanto, o Hamiltoniano clássico é uma função de $6N$ variáveis, onde N é o número de partículas (neste caso, núcleos atômicos).

Em um sistema dinâmico, dizemos que posições e momentos variam no tempo. Descobrir a forma como esta variação ocorre é o problema central da mecânica clássica. Na mecânica de Hamilton, o comportamento dinâmico de um sistema Hamiltoniano pode ser capturado a partir da diferencial total da função Hamiltoniana, considerando-se que as posições e momentos das partículas são função do tempo. Esse resultado é representado pelas equações de movimento de Hamilton (13,14).

$$\dot{\mathbf{p}}_i = -\frac{\partial H}{\partial \mathbf{r}_i} \quad (13)$$

$$\dot{\mathbf{r}}_i = \frac{\partial H}{\partial \mathbf{p}_i} = \frac{\mathbf{p}_i}{m_i} \quad (14)$$

Resolver a dinâmica do sistema envolve encontrar as velocidades e posições das partículas em um tempo posterior a partir de tempos anteriores, isto consiste em resolver estas equações diferenciais de primeira ordem para cada uma das partículas, a cada passo de tempo, e encontrar as posições e velocidades no intervalo de tempo posterior.

Para resultados de alta exatidão, é preferível resolver as duas equações diferenciais de primeira ordem de Hamilton (13,14). Entretanto, a partir da definição clássica de momento linear, é possível chegar, a partir das equações de Hamilton, a uma equação diferencial de segunda ordem (15), que nada mais é que a equação de Newton e relaciona as forças sobre a partícula a uma derivada temporal da posição.

$$\mathbf{F}_i = m_i \mathbf{a} \quad (15)$$

Resolver a equação de Newton é mais eficiente computacionalmente que resolver as duas equações de Hamilton, apesar de ser menos exato, por este motivo normalmente os métodos de dinâmica molecular integram a equação de Newton.

Os algoritmos utilizados para integrar as equações de movimento da qual estamos tratando agora são historicamente conhecidos como algoritmos tipo *Verlet*, em homenagem ao físico francês Loup Verlet, que foi o primeiro a aplicar métodos de integração numérica às equações de movimento em simulações de dinâmica molecular (VERLET, 1967).

Considerando a posição como função do tempo (o indispensável para se pensar em um sistema *dinâmico*) de modo que em um tempo t a posição dos átomos no sistema seja $\mathbf{R}(t)$, a ideia fundamental dos algoritmos tipo Verlet está em escrever duas expansões de Taylor da posição: uma para um passo à frente (16) no tempo e outra para um passo atrás (17). Em ambas, a expansão é truncada no terceiro termo.

$$\mathbf{R}(t+\Delta) = \mathbf{R}(t) + \Delta \dot{\mathbf{R}}(t) + \frac{\Delta^2}{2} \ddot{\mathbf{R}}(t) + O(\Delta^3) \quad (16)$$

$$\mathbf{R}(t-\Delta) = \mathbf{R}(t) - \Delta \dot{\mathbf{R}}(t) + \frac{\Delta^2}{2} \ddot{\mathbf{R}}(t) - O(\Delta^3) \quad (17)$$

Adicionando-se estas duas expressões, conseguimos uma terceira que relaciona as coordenadas no futuro às coordenadas no presente e no passado (18).

$$\mathbf{R}(t+\Delta) \approx 2\mathbf{R}(t) - \mathbf{R}(t-\Delta) + \Delta^2 \mathbf{M}^{-1} \mathbf{F}(t) \quad (18)$$

Subtraindo-se as equações (16) de (17), obtemos uma equação para as velocidades das partículas em um tempo t em função da velocidade que depende das posições um passo atrás e um passo à frente, mas não no passo atual. Perceba que à medida que o passo de tempo torna-se muito pequeno, esta definição aproxima-se da definição de derivada da posição, o que é consistente.

$$\mathbf{V}(t) = \frac{[\mathbf{R}(t+\Delta) - \mathbf{R}(t-\Delta)]}{2\Delta} \quad (19)$$

Apesar de suficientes para a integração das equações de movimento, as equações (16) e (17) não são muito apropriadas. O motivo disto é que utilizando estas equações para propagar o sistema, o cálculo das velocidades está sempre defasado em relação ao das posições. Isto é, para saber $\mathbf{V}(t)$ é preciso antes saber $\mathbf{R}(t+\Delta)$.

Esta característica é levemente inconveniente para as rotinas de saída de dados mas também tem uma complicação adicional de ordem prática: no estado inicial do sistema, as velocidades não podem ser calculadas porque não se tem qualquer informação sobre o estado futuro das posições e portanto é preciso usar uma forma diferente de se calcular os valores no estado inicial.

Uma variação do algoritmo de Verlet convencional, chamado criativamente de *velocity-Verlet*, resolve este problema integrando as velocidades no passo atual utilizando as forças do passo seguinte. Este algoritmo parte das definições de velocidade em (21') e (21).

$$\mathbf{R}(t+\Delta) = \mathbf{R}(t) + \Delta \dot{\mathbf{R}}(t) + \frac{\Delta^2}{2} \ddot{\mathbf{R}}(t) + O(\Delta^3)$$

$$\mathbf{R}(t+\Delta) \approx \mathbf{R}(t) + \Delta \mathbf{V}(t) + \frac{\Delta^2}{2} \mathbf{M}^{-1} \mathbf{F}(t) \quad (20)$$

$$\frac{1}{\Delta} [\mathbf{R}(t+\Delta) - \mathbf{R}(t) - \frac{\Delta^2}{2} \mathbf{M}^{-1} \mathbf{F}(t)] \approx \mathbf{V}(t)$$

$$\mathbf{V}(t) \approx \frac{1}{\Delta} [\mathbf{R}(t+\Delta) - \mathbf{R}(t) - \frac{\Delta^2}{2} \mathbf{M}^{-1} \mathbf{F}(t)]$$

$$\mathbf{V}(t) \approx \frac{\mathbf{R}(t+\Delta) - \mathbf{R}(t)}{\Delta} - \frac{\Delta}{2} \mathbf{M}^{-1} \mathbf{F}(t)$$

$$\mathbf{V}(t) \approx \frac{[\mathbf{R}(t+\Delta) - \mathbf{R}(t)]}{\Delta} - \frac{\Delta}{2} \mathbf{M}^{-1} \mathbf{F}(t) \quad (21')$$

Note que a quantidade $\frac{[\mathbf{R}(t+\Delta) - \mathbf{R}(t)]}{\Delta}$ é uma velocidade segundo a definição (19), essa velocidade considera um passo que é metade do passo de integração original. Portanto podemos

escrever que $V(t+1/2\Delta) = \frac{[R(t+\Delta) - R(t)]}{\Delta}$. Substituindo-se esta definição em (21') segue-se que $V(t) \approx V(t+1/2\Delta) - \frac{\Delta}{2} M^{-1} F(t)$. Que ao ser rearranjada assume a forma da equação (21):

$$-V(t+1/2\Delta) \approx -V(t) - \frac{\Delta}{2} M^{-1} F(t)$$

$$V(t+1/2\Delta) \approx V(t) + \frac{\Delta}{2} M^{-1} F(t)$$

$$V(t+\Delta) \approx V(t+1/2\Delta) + \frac{\Delta}{2} M^{-1} F(t+\Delta)$$

$$V(t+\Delta) - V(t) \approx \frac{\Delta}{2} M^{-1} [F(t+\Delta) + F(t)]$$

$$V(t+\Delta) \approx V(t) + \frac{\Delta}{2} M^{-1} [F(t+\Delta) + F(t)] \quad (21)$$

A equação (21) pode ser utilizada para calcular as velocidades em tempos posteriores a partir do conhecimento das velocidades no estado atual do sistema, dependendo no entanto da informação sobre as forças.

Em sistemas conservativos (onde o potencial depende apenas da posição das partículas), as forças podem ser calculadas a partir do gradiente do potencial.

$$F(t) = -\nabla U[R(t)] \quad (22)$$

Como as posições posteriores são calculadas primeiro, com a equação (20), é possível calcular as forças a partir da equação (22) e então calcular as velocidades no mesmo passo temporal das posições a partir da equação (21), solucionando o problema do algoritmo básico de Verlet.

Note, no entanto, que no passo inicial $t = 0$, as posições são aquelas da geometria de entrada, mas as velocidades iniciais precisam ser atribuídas de forma externa ao algoritmo de integração, pois elas são necessárias para se calcular as próximas posições e portanto as próximas forças, para então encontrar as próximas velocidades.

Também é necessário saber o tamanho do passo de integração, uma vez que todas as quantidades dependem desta informação para serem apropriadamente calculadas a cada iteração do procedimento de integração.

O ciclo de integração então segue-se da seguinte forma, ilustrado na Figura 9:

1. Defina a quantidade e os tipos de átomos, bem como o potencial de interação;

2. Atribua valores iniciais de posições R_0 e velocidades V_0 ;
3. Defina um passo temporal Δ de integração;
4. Determine o número n_t de passos a serem integrados;
5. Calcule as forças F_0 para as posições iniciais;
6. Itere ao longo de n_t passos:
 1. Calcule as novas posições no passo $t+\Delta$ com (20);
 2. Calcule as novas forças no passo $t+\Delta$ com (22);
 3. Calcule as novas velocidades em $t+\Delta$ com (21);
 4. Escreva informações pertinentes sobre a integração do passo t para $t+\Delta$;
 5. Incremente o passo de tempo;
7. Analise os resultados.

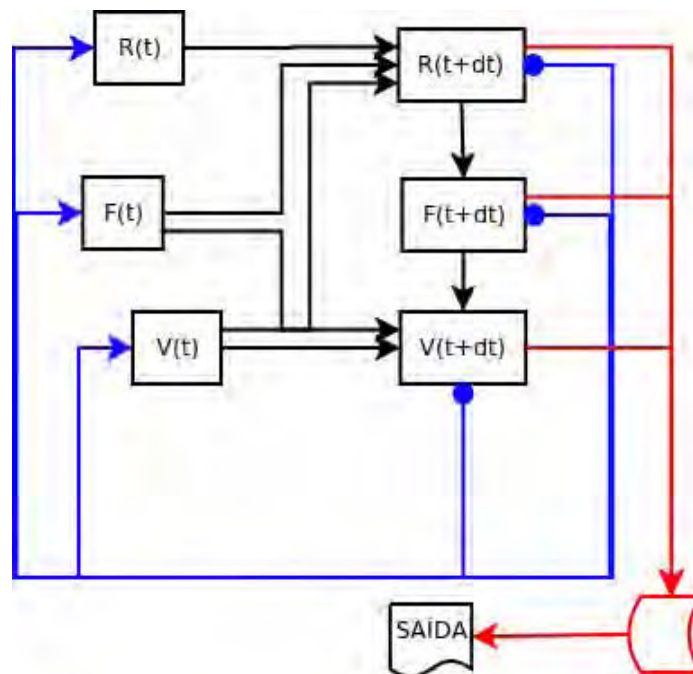


Figura 9: Ciclo de Integração do Velocity-Verlet

As posições iniciais R_0 são obtidas a partir da geometria de entrada. A composição do sistema (número e tipos de átomos) também precisam ser informadas no início do procedimento. Estes parâmetros são utilizados no cálculo do potencial além de outros termos.

Uma forma conveniente de se estabelecer velocidades iniciais V_0 é gerando-as aleatoriamente a partir da distribuição de Maxwell-Boltzmann (23). Este método permite que as velocidades sejam atribuídas de forma que o sistema tenha uma determinada “temperatura instantânea” no começo da simulação.

$$f(v)dv = \sqrt{\frac{m_i}{2\pi K_B T}} \exp\left(-\frac{m_i v^2}{2 K_B T}\right) dv \quad (23)$$

Esta forma de se atribuir velocidades é justificada por um resultado da termodinâmica estatística que relaciona a média da energia cinética K do sistema com sua temperatura macroscópica (24).

$$T = \frac{1}{N_A K_B} \langle K \rangle \quad (24)$$

Analogamente, podemos pensar que uma temperatura instantânea T que está relacionada à energia cinética instantânea da trajetória (25).

$$T = \frac{1}{N_A K_B} K \quad (25)$$

Esta temperatura instantânea pode ser calculada a partir das velocidades instantâneas de cada microestado por onde a dinâmica passa. E portanto, as velocidades iniciais têm influência na temperatura do microestado inicial.

O passo de integração é uma variável de entrada cuja atribuição não é muito trivial. Quanto mais curto o passo, menos tempo “real” a dinâmica será capaz de simular antes de teoricamente faltar espaço para o armazenamento de dados de saída. Entretanto, um passo muito grande compromete o algoritmo de integração porque a expansão de Taylor é uma aproximação local.

Além disso, a “resolução” da dinâmica é determinada pelo tamanho desse passo. Se o passo é muito largo, fenômenos que acontecem em intervalos de tempo inferiores ao tamanho do passo não são adequadamente amostrados do *ensemble*.

Por esse motivo, o passo de integração acaba sendo restrito a intervalos bastante curtos para não prejudicar a exatidão do algoritmo de integração e para amostrar apropriadamente os graus de liberdade mais rápidos do sistema.

Em geral, estes graus de liberdades mais rápidos correspondem a estiramentos de ligações envolvendo hidrogênios, particularmente a ligação carbono-hidrogênio, cuja vibração é cerca de 3000 cm^{-1} que ocorrem na escala de tempo da ordem de femtossegundos (10^{-15} s). Na prática, portanto, o intervalo de integração utilizado em simulações de dinâmica molecular usuais é de 1fs.

O potencial é mais complicado. Dissemos que, segundo a aproximação Born-Oppenheimer, é possível simular o movimento nuclear a partir do conhecimento da superfície potencial derivada da resolução paramétrica da função de onda eletrônica para cada configuração especial nuclear.

Em teoria, seria possível utilizar este potencial *quântico*, por assim dizer, para movimentar os núcleos. Entretanto, o custo computacional de se calcular uma função de onda eletrônica para cada passo de 1fs cresce muito rapidamente com a quantidade e os tipos de átomos no sistema.

Esta lentidão na propagação dos sistema no tempo faz com que cálculos de dinâmica molecular utilizando um potencial deste tipo sejam realizados apenas para sistemas muito pequenos, onde o tempo de convergência do cálculo do potencial não seja um fator tão limitante.

1.2.2.3. Mecânica Molecular

Para diminuir a demanda computacional envolvida no cálculo do potencial, formas diferentes de se calcular esta energia foram propostas. A mais computacionalmente eficiente destas formas utiliza os chamados potenciais *clássicos* ou de mecânica molecular.

Independente do método utilizado para se calcular os potenciais quânticos, o cálculo do potencial utilizando os potenciais empíricos da mecânica molecular é muito mais rápido. Esta diferença de velocidade permite que estes potenciais sejam utilizados para propagar sistemas muito maiores e amostrar muito mais configurações em uma dinâmica molecular.

Os potenciais de mecânica molecular abandonam completamente o conceito de função-de-onda. Neles, os núcleos são vistos como partículas pontuais maciças e a energia é modelada por uma função analítica cujos parâmetros são ajustados de modo que os valores calculados concordem com observáveis termodinâmicos experimentais.

A função analítica mais comumente utilizada nas simulações clássicas é uma soma de várias parcelas de energia (26). Cada parcela representa uma forma diferente pela qual a energia potencial total do sistema varia.

$$U(\mathbf{R}) = U_{\text{ligações}} + U_{\text{ângulos}} + U_{\text{diedrais}} + U_{\text{impróprios}} + U_{\text{eletrostático}} + U_{\text{Lennard-Jones}} \quad (26)$$

Os termos $U_{\text{ligações}} + U_{\text{ângulos}} + U_{\text{diedrais}} + U_{\text{impróprios}}$ referem-se a energias calculadas entre átomos que estão na mesma molécula e são chamados de termos *ligados*. Os termos $U_{\text{eletrostático}} + U_{\text{Lennard-Jones}}$ contabilizam interações entre átomos em moléculas diferentes ou moléculas a mais de um diedral (quatro ligações) de distância um do outro, e são chamados de termos de energia *não-ligados*.

A equação (26) para expressar a energia do sistema pode então ser reescrita de forma que $U(\mathbf{R}) = U_{\text{ligados}} + U_{\text{não-ligados}}$.

A contribuição total da energia das ligações é calculada somando-se parcelas de cada ligação segundo a equação (27). Cada parcela é calculada segundo o potencial harmônico onde os parâmetros k_{lig} e r_0 são a constante de força e a distância de equilíbrio da ligação respectivamente.

$$U_{\text{ligações}} = \sum_{\text{ligações}} \frac{1}{2} k_{\text{lig}} (r - r_0)^2 \quad (27)$$

Os ângulos seguem um potencial harmônico semelhante aos potenciais das ligações (28). Mais uma vez, a contribuição de cada ângulo do sistema é contabilizado.

$$U_{\text{ângulos}} = \sum_{\text{ângulos}} \frac{1}{2} k_{\text{angl}} (\theta - \theta_0)^2 \quad (28)$$

Os ângulos diedrais, ao contrário dos ângulos convencionais não são adequadamente modelados segundo um potencial harmônico. Isto porque um ângulo diedral em geral pode assumir valores entre 0° e 360°. Funções trigonométricas do tipo (29) (ou combinações de funções trigonométricas deste tipo) são utilizadas como forma empírica de um diedral porque são contínuas para todo o intervalo de ângulos possíveis.

$$U_{\text{diedrais}} = \sum_{\text{diedrais}} \frac{1}{2} V_A [1 + \cos(A \phi + \delta)] \quad (29)$$

A energia envolvida em movimentos para fora do plano, uma barreira de energia importante para grupos planares como carbonos sp², são modelados de forma semelhante aos diedrais (30), mas entre átomos que não estão diretamente ligados entre si. Por este motivo, estes ângulos são chamados *diedrais impróprios*.

$$U_{\text{impróprios}} = \sum_{\text{impróprios}} \frac{1}{2} V_B [1 + \cos(B \omega + \delta)] \quad (30)$$

Os parâmetros A , e B são a *fase* da função cosseno utilizada para modelar o comportamento periódico. Os termos V são as constantes de força dos diedrais, e dependem dos termos A e B . Estes termos representam a periodicidade da função cosseno e o δ é a fase do ângulo. Por questões de praticidade, a fase do ângulo normalmente é fixada em 0° ou 180°.

À parte estes termos que contabilizam a contribuição devido à configuração relativa dos átomos ligados entre si, temos ainda um conjunto de termos cujo objetivo é simular interações a longa distância entre quaisquer dois átomos.

A primeira e mais elementar forma de duas partículas carregadas interagirem é seguindo a lei do potencial eletrostático de Coulomb (31).

$$U_{\text{eletrostático}} = C \epsilon^{-1} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (31)$$

Neste potencial, C é uma constante igual a $1/4 \pi \epsilon_0$ que precisa ser utilizada quando queremos o resultado no sistema de unidades MKSA, ϵ é a constante dielétrica do meio onde se encontra o sistema (que será igual a 1 se for no vácuo), r_{ij} é a distância de separação entre quaisquer dois núcleos e os termos q são as cargas pontuais centradas nos núcleos atômicos considerados.

Calcular o potencial de Coulomb desta forma para todo o sistema gera dois problemas. O primeiro ficará mais claro quando falarmos de fronteiras periódicas, o segundo problema é muito mais evidente: como determinar as cargas de cada átomo de forma que a soma de todas elas seja igual à carga total da molécula e as propriedades calculadas tenham boa correlação com os observáveis macroscópicos?

Esse problema deu origem a uma série de algoritmos de atribuição de cargas parciais. Um dos mais conhecidos e utilizados entre eles é o AM1-BCC (JAKALIAN; JACK; BAYLT, 2002), que foi

utilizado neste trabalho para gerar as cargas parciais das moléculas. Este método foi parametrizado para distribuir as cargas de forma a reproduzir a superfície de potencial eletrostático *ab initio* de mais de 2700 moléculas de treino.

Outro modelo de cargas é o RESP. RESP é a sigla em inglês para “Potencial Eletrostático Restrito”. Neste modelo, a superfície de potencial eletrostático calculado na base 6-31G* para uma série de moléculas é parametrizado não apenas para fazer as cargas reproduzirem o potencial eletrostático, mas também para minimizarem uma certa função de restrição (BAYLY et al, 1993). Esta função tem como objetivo diminuir a inexactidão com que as cargas no interior das moléculas é descrita.

Além da interação eletrostática decorrente da lei de Coulomb aplicada às cargas parciais, existe um conjunto de interações que dependem da distribuição da densidade eletrônica em torno dos átomos. Este conjunto complementar de interações é modelada por um potencial empírico descrito pela primeira vez pelo matemático inglês Sir John Edward Lennard-Jones, em 1924 (LENNARD-JONES, 1924) (32).

$$U_{Lennard-Jones} = \sum_i \sum_{j \neq i} \left(\frac{X_{ij}}{r_{ij}^{12}} - \frac{Y_{ij}}{r_{ij}^6} \right) \quad (32)$$

Ao se aproximarem demais, os átomos sobrepõe suas densidades eletrônicas, produzindo uma elevação brusca no potencial. Este tipo de interação decai muito rapidamente com a distância e é representado pelo primeiro termo da diferença.

O segundo termo representa as interações de Van der Waals, decorrentes da polarização instantânea ou permanente da densidade eletrônica em torno dos átomos de uma molécula.

Os termos X_{ij} e Y_{ij} são parâmetros empíricos calculados a partir da profundidade do poço de energia potencial ϵ_{ij} e são definidos em termos das distâncias de equilíbrio σ_{ij} entre os átomos de tal forma que:

$$U_{Lennard-Jones} = \sum_i \sum_{j \neq i} \epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (33)$$

Os parâmetros ϵ_{ij} e σ_{ij} , apesar de dependerem formalmente de dois átomos são determinados a partir de parâmetros atômicos, pelas seguintes regras simples de combinação:

$$\epsilon_{ij} = \sqrt{\epsilon_{ii} \epsilon_{jj}} \quad \text{e} \quad \sigma_{ij} = \sqrt{\sigma_{ii} \sigma_{jj}} \quad .$$

Os termos que correspondem a contribuições não-ligadas não são calculados para átomos que estão ligados. Isto porque nesta situação, como as distâncias são muito pequenas, as interações seriam muito grandes e seriam adicionadas às contribuições envolvendo os termos ligados. Isto produz contribuições anormalmente altas de energia e não se correlacionam bem com valores de observáveis experimentais.

1.2.2.4. Raios de Corte

Um dos problemas a respeito das interações entre átomos não ligados é que elas somam por todos os pares de átomos possíveis (com a exceção dos átomos ligados entre si) e portanto o tempo de cálculo exato destas interações cresce com o quadrado do número de átomos.

Evidentemente, isto torna o cálculo muito lento, até mesmo para sistemas não tão grandes. Considerando que em um cálculo de dinâmica molecular este cálculo precisa ser feito muitas vezes ao longo da simulação, esta não é uma forma apropriada de calcular este potencial.

Em particular, as interações eletrostáticas são mais problemáticas porque decaem muito lentamente com a distância, o que quer dizer que muitos pares de átomos precisam ser considerados na equação (31) para obter um aumento sensível na precisão do cálculo.

Uma das formas de se acelerar o cálculo das equações (31) e (32) para um sistema é fingir que ela será resolvida para um sistema menor, isto significa não calcular a interação entre alguns pares de átomos. A forma escolhida de fazer isto é considerar as interações de um átomo apenas com aqueles que estiverem a uma certa distância ou mais próximos.

A forma descrita acima torna o cálculo linearmente proporcional ao número de átomo e não quadraticamente proporcional, como o cálculo completo. Como o cálculo estabelece um raio limite para o cálculo da interação, este método é chamado método dos *raios de corte*.

O método dos raios de corte funciona bem para permitir o cálculo da energia eletrostática para sistemas muito grandes, mas como já citado anteriormente, a natureza a longo alcance do potencial eletrostático exige que para conseguir uma boa energia aproximada, muitos termos do somatório em (31) sejam considerados. A sabedoria está em encontrar um raio de corte que consiga satisfazer ambas as necessidades: exatidão suficiente e menor tempo de cálculo.

Zerar o potencial que esteja além do raio de corte, entretanto, não é a única forma de se aplicar um raio de corte, nem é também muito elegante. O que acontece é que a energia potencial aumenta e diminui de forma abrupta à medida que os átomos entram e saem uns dos raios de corte dos outros.

Em muitos casos, os pequenos saltos para cima, provocados por átomos entrando nos raios onde eles não estavam antes, são compensados por pequenos saltos para baixo provocados por átomos deixando os raios de corte onde antes se encontravam, de modo que a variação brusca provocada por este modelo de raio de corte não compromete a exatidão da simulação.

Em alguns casos, no entanto, pode acontecer de a conservação da energia total do sistema ser prejudicada por este tipo de raio de corte. Por este motivo, outros métodos mais sofisticados aplicam funções que “suavizam” o potencial até que ele chegue a zero de forma contínua.

Utilizar uma função “suavizante” $S(r)$ significa multiplicar esta função pela função do potencial. A forma mais simples da função S é uma função igual a 1 quando a distância de interação é igual ou menor que o raio de corte, e 0 quando ocorre o contrário. Esta é a função da qual temos falado até agora.

Outras funções S mais sofisticadas, no entanto, foram desenvolvidas. Uma função deste tipo bastante popular é a função *interruptor (switch)* (34). Esta função S utiliza dois raios de corte, um raio de corte mais interno r_i e outro mais externo r_e . Para qualquer distância abaixo de r_i o cálculo é realizado normalmente e a função S é dada como sendo igual a 1.

$$S(r) = \frac{(r_i^2 - r^2)^2 (r_e^2 + 2r^2 - 3r_i^2)}{(r_e^2 - r_i^2)^3} \quad (34)$$

Quando a distância ultrapassa este raio mais interno, entretanto, a suavização entra em cena e o potencial entre os raios de corte interno e externos sofre uma suavização. Quando potencial cruza o raio externo, a função $S(r)$ é desligada e automaticamente dada como 0 (Figura 10).

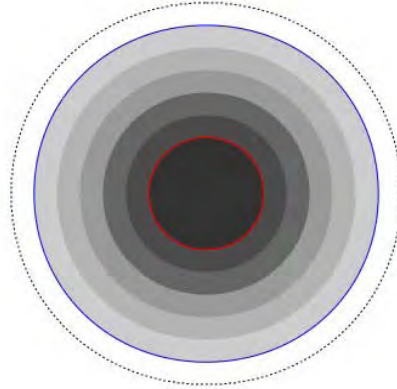


Figura 10: Esquema da função switch. O potencial só sofre ação suavizante após o raio interno

A função é construída de modo que $S(r_e)$ também seja igual a zero, de modo que a suavização seja contínua.

Outra função que suavemente decresce o potencial até o final é a função *deslocamento* (35). Assim chamada porque desloca os valores do potencial para próximo de zero quando a distância de separação se aproxima do raio de corte e retorna zero para qualquer valor de distância acima dele.

$$S(r) = \left[1 - \left(\frac{r}{r_c} \right)^2 \right]^2 \quad (35)$$

Esta função, assim como a função *switch* é contínua, mas diferentemente das duas primeiras, esta modifica o cálculo do potencial para *quaisquer* valores de distância (Figura 11), mesmo aqueles dentro o raio de corte.

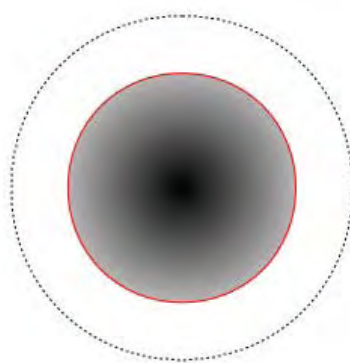


Figura 11: Esquema ilustrativo da função shift. Até mesmo dentro do raio o potencial sofre suavização

Aparentemente, não há nenhum ganho em desempenho computacional, afinal, o papel dos raios de corte era justamente evitar o cálculo das distâncias e funções entre átomos todas as iterações e agora temos que calcular funções “raio de corte” além das próprias funções potenciais.

O ganho computacional, no entanto, vem do fato de que o cálculo de quais átomos estão a quais distâncias não é repetido todas as vezes, e também não é feito átomo por átomo. O que se faz é dividir a caixa de simulação de uma forma parecida com que os geógrafos dividem a Terra com seus meridianos e paralelos (Figura 50).

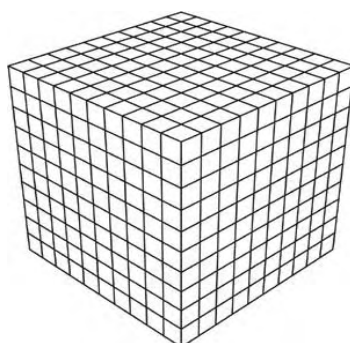


Figura 12: Exemplo de um grid

Chamados este conjunto de “paralelos” e “meridianos” que dividem a caixa de simulação de *grid*, a palavra em inglês para grade, na realidade, o que nos interessa desta grade não são as linhas completas, mas apenas os pontos de intersecção entre cada linha. Vamos chamar estes pontos de intersecção de *vértices*.

Fazendo isto, mapeamos cada átomo do sistema no vértice mais próximo (é como se estivéssemos dando uma aproximação de sua coordenada verdadeira) e utilizamos as distâncias entre os vértices para decidir quais distâncias reais e interações serão calculadas e quais não serão.

O método não é contraproducente porque a distância de separação entre quaisquer dois

vértices vizinhos é constante e idêntica de modo que a distância entre eles pode ser calculada muito mais rapidamente que a distância arbitrária que pode haver entre dois átomos. Desta forma, o tempo de cálculo depende unicamente da quantidade de pontos no *grid*.

Além do mais, o cálculo do *grid* não é feito todas as vezes que o potencial é calculado porque os átomos não se movem para longe de seus vértices mapeados tão rapidamente assim. Apenas quando o *grid* atual não descreve mais a distribuição dos átomos no espaço é que um novo é calculado.

1.2.2.5. Fronteiras Periódicas

Um problema que surge quando se tenta resolver a dinâmica de um sistema atômico-molecular está na descrição das fronteiras do sistema.

Tome, por exemplo, uma simulação biomolecular típica: uma biomolécula solvatada por uma grande quantidade de moléculas de água. O que ocorre quando se chega à fronteira do sistema?

A menos que o interesse esteja em simular uma única molécula de proteína solvatada por um único e microscópico agregado de moléculas de água, esta simulação não se aproxima nem um pouco do sistema real.

A forma mais comum (e a que vamos usar) para tratar este problema é utilizando-se as chamadas Condições Periódicas de Fronteira (PBC, do inglês, *Periodic Boundary Conditions*). Na técnica PBC, o sistema é imaginado como estando dentro de uma caixa de simulação e esta caixa se repete para todas as direções.

Isto quer dizer que se tivermos uma partícula do sistema original em uma posição $\mathbf{r}$, então temos na verdade um conjunto infinito de réplicas desta tal partícula (36) de forma que a posição de cada réplica dessas é dada por:

$$\mathbf{r}_{l,m,n} = \mathbf{r} + l\mathbf{a} + m\mathbf{b} + n\mathbf{c} \quad (36)$$

Onde l, m e n são escalares inteiros e os vetores $\mathbf{a}$, $\mathbf{b}$ e $\mathbf{c}$ são as coordenadas das arestas da caixa de simulação. Esta equação nos diz que a posição da réplica é a posição da partícula original mais o vetor de translação que separa a caixa da réplica da caixa original.

Todas estas partículas se movem simultaneamente e, devido à relação que guardam entre si, apenas a posição “original” $\mathbf{r}$ de cada partícula precisa ser armazenada durante a execução do algoritmo. O grande trunfo da PBC é permitir que cada partícula da caixa interaja com as partículas próximas das caixas vizinhas, além de interagir com as partículas de sua própria caixa. Desta forma não se criam superfícies livres em volta do sistema, porque sempre haverá interações com as caixas próximas.

Se para um sistema muito grande já era caro calcular as interações sobre todos os pares de partículas, podemos ter certeza de que o método PBC certamente ampliaria ainda mais esse cenário. Imaginemos que agora, além de nos preocuparmos com um sistema, temos que tomar interações entre réplicas deste sistema.

Uma saída desenvolvida para contornar este problema é a chamada *convenção da mínima imagem*, segundo a qual uma partícula i só interage com a imagem de uma partícula j que esteja

mais próxima, e com nenhuma outra.

A convenção da mínima imagem reduz bastante a quantidade de cálculos de interação realizados e pode ser justificada por uma lógica bastante simples: suponha que tenhamos uma partícula i dentro da caixa de simulação de aresta igual a $2r_c$. Suponha ainda que esta partícula i esteja interagindo com duas imagens da partícula j e sejam elas $j1$ e $j2$.

O vetor posição de cada imagem de j está separado da posição original de j por um vetor translação cujo menor comprimento é $2r_c$, que é o tamanho do lado da caixa. Significa dizer que duas imagens estão separadas uma da outra por uma distância maior que esta.

Para i poder interagir com $j1$ e $j2$, ele teria que estar a uma distância r_c de ambas. Mas a distância entre quaisquer duas imagens deve ser mais que $2r_c$, portanto a possibilidade de que i esteja a r_c , ou menos, de ambas é necessariamente falsa.

Por este motivo, i pode interagir apenas com a imagem de j que estiver mais próxima, e com nenhuma outra, assumindo que o raio de corte seja igual ou menor que o comprimento L da aresta da caixa, isto é (37).

$$r_c \leq \frac{1}{2}L \quad (37)$$

Trabalhar nos limites da convenção da mínima imagem significa que não será necessário calcular várias interações entre partículas e imagens em caixas vizinhas, entretanto, também significa que as avaliações de potenciais não-ligados serão mais grosseiras que um sistema periódico tratado por um método de maior resolução, isto porque os raios de corte precisam ter valores relativamente pequenos, calculando apenas as interações mais locais.

1.2.2.6. A Soma de Ewald

A importância dos métodos de cálculo utilizando raios de corte é bastante evidente. Entretanto, existem formas mais sofisticadas de se calcular com exatidão as interações não-ligadas de um sistema periódico. Um destes métodos, e o mais utilizado, é o método da Soma de Ewald e seus derivados (como o método PME).

A Soma de Ewald é um método originalmente proposto pelo físico americano John Peter Ewald em 1921 como uma forma de se calcular a energia eletrostática de cristais iônicos, sistemas naturalmente periódicos. O objetivo inicial de Ewald era que este método o ajudasse em seu campo de pesquisa, a cristalografia de raios-X, entretanto algoritmos eficientes tornaram possível, desde 1990, a utilização da Soma de Ewald no cálculo de energias eletrostáticas de sistemas moleculares e macromoleculares.

Considere um sistema periódico com o potencial de interação da forma $\lambda_i \lambda_j / r_{ij}^p$. Esta forma geral inclui tanto o potencial eletrostático quanto as duas parcelas do potencial de Lennard-Jones. Em um sistema periódico usando um potencial deste tipo, energia total será a soma de todas as interações possíveis entre partículas dentro da caixa de simulação e entre partículas em caixas diferentes (38).

$$S_p = \frac{1}{2} \sum_n \sum_i \sum_j \frac{\lambda_i \lambda_j}{|\mathbf{r}_i - \mathbf{r}_j + \mathbf{t}_n|^p} \quad (38)$$

Onde a soma no índice $\mathbf{n}$, i e j indicam o somatório sobre todas as caixas, os átomos na caixa $\mathbf{n}$ e os átomos da caixa central respectivamente, com $\mathbf{n}=\mathbf{0}$ representando a caixa central. O vetor $\mathbf{t}_n$ indica a translação que distancia as caixas interagentes. As auto-interações ($\mathbf{n}=\mathbf{0}$ e $i = j$) não são consideradas por serem divergentes.

Para entender a soma de Ewald, considere a seguinte identidade:

$$\frac{1}{r} = \frac{\phi(r)}{r} + \frac{1-\phi(r)}{r} \quad (39)$$

Ela é verdadeira para qualquer função. Entretanto, considere de forma particular o que acontece quando a função $\phi(r)$ toma o valor 1 para $r=0$ e decai rapidamente à medida que o r aumenta.

Nestas condições, o segundo termo $[1-\phi(r)]/r$ desaparece para valores de r muito pequenos e o primeiro termo $\phi(r)/r$ desaparece para valores de r muito grandes, de modo que o termo $1/r$ pode ser descrito como sendo uma soma de uma parcela que descreve os valores de r pequenos e outra que descreve a identidade em valores grandes de r .

Vamos dizer que quando uma função $\phi(r)$ obedece àquelas condições (ser 1 quando $r=0$ e decair rapidamente com o valor de r), ela é uma função de convergência de Ewald.

Se multiplicarmos os dois lados da relação (39) por, $\frac{\lambda_i \lambda_j}{2}$ teremos:

$$\frac{1}{2} \frac{\lambda_i \lambda_j}{r} = \frac{1}{2} \left[\frac{\lambda_i \lambda_j \phi(r)}{r} + \frac{\lambda_i \lambda_j [1-\phi(r)]}{r} \right]$$

Perceba que o lado esquerdo desta equação é igual a um dos termos do somatório triplo do potencial (31) para um sistema periódico, de modo que:

$$\frac{1}{2} \frac{\lambda_i \lambda_j}{r_{nij}^p} = \frac{1}{2} \left[\frac{\lambda_i \lambda_j \phi(r)}{r_{nij}^p} + \frac{\lambda_i \lambda_j [1-\phi(r)]}{r_{nij}^p} \right]$$

onde $r_{nij}^p = |\mathbf{r}_i - \mathbf{r}_j + \mathbf{t}_n|^p$. O somatório S_p pode então ser reescrito da seguinte forma:

$$S_p = \frac{1}{2} \sum_n \sum_i \sum_j \frac{\lambda_i \lambda_j \phi(r_{nij}^p)}{r_{nij}^p} + \frac{\lambda_i \lambda_j [1-\phi(r_{nij}^p)]}{r_{nij}^p} \quad (40)$$

Agora considere que a função $\phi(r)$ na equação (40) é uma função de convergência de Ewald. Isto é, ela é igual a 1 quando $r_{nij}^p=0$ e decai rapidamente para 0 à medida que r_{nij}^p aumenta. Ao fazer isto, podemos separar o somatório S_p em uma soma de dois termos (41). Um

dos termos converge para descrever o potencial quando as distâncias são curtas **(41-1)** e o outro termo descreve o potencial quando as distâncias são longas **(41-2)**. Esta soma **(41)** é a chamada Soma de Ewald.

$$S_p = \sigma_p^1 + \sigma_p^2 \quad (41)$$

$$\sigma_p^1 = \sum_{nij} \frac{\lambda_i \lambda_j \phi(r_{nij})}{r_{nij}^p} \quad (41-1) \quad \sigma_p^2 = \sum_{nij} \frac{\lambda_i \lambda_j (1 - \phi(r_{nij}))}{r_{nij}^p} \quad (41-2)$$

O termo das interações de curto alcance e sua convergência pode ser alcançada em um tempo computacional apreciável utilizando-se os procedimentos convencionais já descritos aqui, com bons raios de corte.

O segundo termo, das interações a longa distância, é mais complicado. Ele não converge muito rapidamente utilizando-se os procedimentos convencionais de cálculo de interações. Felizmente como a função $\phi(r)$ a soma do segundo termo converge muito rapidamente no espaço recíproco. Para converter esta soma do espaço real para o espaço recíproco, utilizamos a Transformada de Fourier.

A transformada de Fourier é uma operação integral que toma qualquer função periódica e a decompõe em uma soma de funções seno e cosseno. É um processo análogo a decompor um acorde musical em suas notas constituintes.

Usando a relação de Euler temos que $e^{-iA\theta} = \cos(A\theta) - i\sin(A\theta)$ o termo A representa o período da função. O período das funções seno/cosseno é $A = n2\pi$, onde n pode ter qualquer valor. A frequência ν destas funções é igual a $1/A$. O problema então consiste em encontrar a combinação particular de senos e cossenos de diferentes frequências que mimetiza a fonte de calor complexa. Ou seja, encontrar uma equação que descreve as *amplitudes* de cada frequência na combinação.

A transformada de Fourier pode ser utilizada para decompor qualquer função periódica $f(x)$ em suas amplitudes constituintes. Esta função relaciona as amplitudes da soma a cada frequência. Se imaginarmos que a frequência ν , assim como A , pode assumir qualquer valor. Então a função $\hat{f}(\nu)$ é contínua e pode ser obtida resolvendo-se a equação **(42)**.

$$\hat{f}(\nu) = \int_{-\infty}^{\infty} f(x) \cdot e^{-2\pi x\nu} dx \quad (42)$$

E dizemos que $\hat{f}(\nu)$ é a Transformada de Fourier de $f(x)$. Inicialmente, a transformada de Fourier foi empregada para analisar sinais ondulatórios propagando-se no tempo e ver suas frequências constituintes. Por este motivo, o domínio onde x é definido é chamado de domínio temporal e o domínio onde ν é definido é chamado de domínio das frequências.

Ewald percebeu que, mesmo que os termos de longo alcance não tenham uma convergência muito rápida no espaço cartesiano, no espaço onde é definida sua Transformada de Fourier (o chamado espaço recíproco), essa soma converge rapidamente.

A eficiência com que esta ideia é posta em prática, entretanto, depende absolutamente da forma escolhida para a função $\phi(r)$. A forma da equação (43), em particular, satisfaz os critérios (NIJBOER; DE WETTE, 1957).

$$\phi(r) = \frac{\Gamma(p/2, \kappa^2 r^2)}{\Gamma(p/2)} \quad (43)$$

onde κ é um parâmetro arbitrário e as funções Γ da forma $\Gamma(x)$ e $\Gamma(x, y)$ são as funções-gama incompleta e completa respectivamente da forma:

$$\Gamma(x) = \int_0^\infty t^{x-1} \exp(-t) dt$$

$$\Gamma(x, y) = \int_y^\infty t^{x-1} \exp(-t) dt$$

Para $p > 3$, a soma S_p é absolutamente convergente. Significa dizer que a soma da energia de interação da forma como mostrado na equação (38) é convergente não importa os valores de λ nem a ordem em que os termos são calculados.

Usando a função $\phi(r)$ da forma (43), aplicada na equação (41), obtemos a seguinte equação para $p = 6$.

$$\begin{aligned} S_6 = & \frac{1}{2} \sum_{nij} \left[1 + (\kappa r_{nij})^2 + \frac{(\kappa r_{nij})^4}{2} \right] \exp[-(\kappa r_{nij})^2] \\ & + \frac{(\sqrt{\pi} \kappa)^3}{6V} \sum_k \left\{ [1 - 2(k/2\kappa)^2] \exp[-(k/2\kappa)^2] + 2(k/2\kappa)^3 \sqrt{\pi} \operatorname{erfc}(k/2\kappa) \right\} \\ & \times \left\{ \left[\sum_i \lambda_i \cos(\mathbf{k}^T \mathbf{r}_i) \right]^2 + \left[\sum_i \lambda_i \sin(\mathbf{k}^T \mathbf{r}_i) \right]^2 \right\} \\ & - \frac{\kappa^3}{12} \sum_i \lambda_i^2 \end{aligned} \quad (44)$$

Onde V é o volume da caixa periódica e a função erfc é a função-erro complementar. A função erfc é definida em termos da função erro erf , mostrada na Figura 13. A função erfc é então definida como sendo $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$, assim é fácil ver como a função erfc converge à medida que o x aumenta (Figura 13).

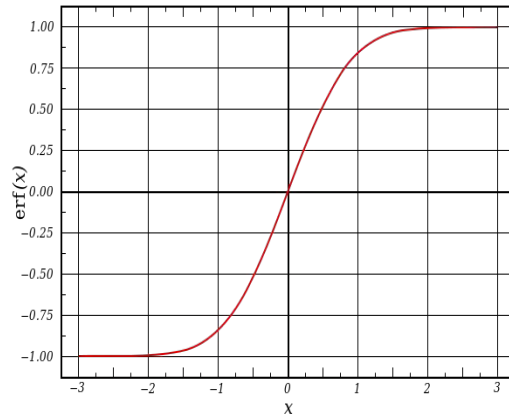


Figura 13: Função erro $erfc$.

Perceba que os termos da segunda parcela desta soma são somas de senos e cossenos cujas frequências são somadas sobre um conjunto de vetores $\mathbf{k}$. Estes valores são obtidos realizando a transformada de Fourier no segundo termo da equação (40) e por isso são chamados de vetores do espaço recíproco.

Esta soma irá convergir rapidamente à medida que $k=|\mathbf{k}|$ aumenta devido aos termos exponenciais (que são Gaussianas) e envolvendo a função erro complementar. Esta expressão pode ser utilizada para calcular o termo atrativo do potencial de Lennard-Jones.

Somas S_p onde $p \leq 3$ são apenas condicionalmente convergentes, o que significa dizer que o valor final da soma (e se ele irá convergir ou não) depende da forma como a soma é realizada. Em particular, a soma nunca irá convergir a menos que a condição $\sum_i \lambda_i = 0$ seja satisfeita. Para as interações eletrostáticas, isto quer dizer que a soma de todas as cargas na caixa de simulação deve ser zero.

Um detalhe importante quando utilizando as expressões da Soma de Ewald para calcular as energias de interação é que a soma contabiliza todas elas. Mas isto não é verdade para os campos-de-força, que explicitamente excluem interações 1-2, 1-3 e às vezes 1-4 entre átomos ligados, que são calculadas segundo os termos de energia ligados. Portanto eles devem ser calculados da forma convencional e depois subtraídos da energia encontrada pela soma de Ewald.

1.2.2.7. O Método PME (Particle-Mesh Ewald)

O método da Soma de Ewald é um resultado teórico da física e da matemática, obtido antes que os computadores estivessem prontos para simular a dinâmica de sistemas físicos. Apesar de suas equações poderem ser aplicadas diretamente, o método de Ewald sofreu substanciais modificações ao longo do tempo para ganhar em eficiência (o topo de eficiência que a Soma de Ewald consegue alcançar é da ordem de $N^{3/2}$).

Um dos métodos derivados da Soma de Ewald é o método PME (*Particle-Mesh Ewald* algo como “Ewald Partícula-Malha”). Este método foi primeiro proposto por Darden, York, e Pedersen (1993) para tratar as somas de Ewald em sistemas muito grandes (no caso específico do artigo, o

exemplo são cristais iônicos).

Neste método, a transformada de Fourier da Soma de Ewald são substituídas por transformadas discretas. Transformadas discretas de Fourier são transformadas integrais que lidam com coleções de pontos (ao invés de funções contínuas) e calculam transformadas igualmente descontínuas.

Para tornar o problema contínuo com o qual temos trabalhado até o momento em um problema discreto, o método PME cria um *grid* de espaçamento finito que inclui toda a caixa de simulação e então mapeia a densidade de cargas (uma função contínua) neste *grid* por métodos de interpolação (tornando-a discreta). Um procedimento análogo ao que fizemos com os raios de corte, exceto que naquele caso a posição dos átomos já era discreta e neste caso, a densidade de carga é contínua.

O *grid* é necessário para tornar todas as transformadas de Fourier contínuas em transformadas discretas, que podem ser avaliadas através de métodos FFT (*Fast Fourier Transform*) (COOLEY; TUKEY, 1965) desenvolvidos para realizar transformadas de Fourier discretas com maior eficiência que o método convencional, utilizando a definição da transformada. A soma de convergência recíproca é então avaliada utilizando o *grid* (a “Malha” do método) e a soma de convergência real é avaliada normalmente, utilizando raios de corte (a “Partícula” do método).

A eficiência exata deste algoritmo depende da forma como a densidade de cargas varia no sistema. Se a densidade variar de forma suave, o cálculo será rápido porque a soma no espaço de Fourier convergirá muito rapidamente, do contrário, o cálculo será mais lento.

Por este motivo, o PME, assim como a soma de Ewald é utilizado para acelerar o cálculo apenas das interações de alcance mais longo, cuja variação da densidade de carga obedece às restrições do método. Desta forma, o *grid* não precisa ser recalculado todas as vezes e, ainda por cima, a soma no espaço recíproco converge rapidamente.

1.2.2.8. Controle da Simulação

As trajetórias de dinâmica molecular só têm o significado do qual fala a hipótese ergódica quando feitas com sistemas em equilíbrio. Entretanto, existem muitas condições de equilíbrio para um sistema. Temos o equilíbrio térmico, químico, mecânico... Cada um deles associado a diferentes propriedades como temperatura, potencial químico e pressão.

As condições precisas do equilíbrio em que se encontra o sistema que se deseja simular determinam os microestados acessíveis e a probabilidade de transições de um microestado para outro de forma que cada conjunto de propriedades *conservadas* durante a trajetória em equilíbrio delimita um ensemble estatístico diferente de microestados.

Os ensembles mais importantes são o *microcanônico* NVE (onde o que se conserva é o número de partículas, o volume e a energia), o *canônico* NVT e o *isotérmico-isobárico* NPT. Ensembles onde outros tipos de quantidades se conservam são utilizados em menor proporção.

O ensemble mais comum e de onde partiu toda a formulação da mecânica estatística, o NVE, descreve sistemas onde o número de partículas, o volume e a energia interna se conservam. Apesar de útil na formulação deste campo de estudo, este ensemble não descreve sistemas de grande interesse químico, pois os sistemas normalmente estudados dificilmente têm sua energia interna conservada. Muito mais interesse seria, por exemplo, estudar o comportamento microscópico de um sistema NPT (um béquer aberto) ou NVT (o interior de um calorímetro).

1.2.2.8.1 Termostatos

Para poder calcular trajetórias nestes ensembles, primeiro precisamos formalizar métodos de se controlar a pressão e a temperatura. Os primeiros algoritmos desenvolvidos para se controlar estes parâmetros do sistema foram formalizados por H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola e J. R. Haak, em 1984 (BERENDSEN et al, 1984) no que ficou conhecido como Termostato de Berendsen.

Apesar de não gerar resultados consistentes com o ensemble canônico para sistemas pequenos, o termostato de Berendsen se aproxima bastante dos resultados deste ensemble para sistemas da ordem de centenas a milhares de átomos ou mais, que é precisamente a ordem de grandeza que se deseja.

Ainda que não fosse o caso, entretanto, o termostato de Berendsen é uma boa ferramenta para se compreender alguns aspectos teóricos do controle das simulações, então vale a pena analisá-lo.

Além do termostato de Berendsen, apresentamos os termostatos de Andersen, por sua engenhosa simplicidade e o de Langevin pela sua ampla aplicabilidade em campos que vão além do controle da temperatura de uma simulação.

1.2.2.8.1.1. Berendsen

A ideia geral é que o sistema não está isolado. Na termodinâmica, isto tem um significado muito preciso. Por “o sistema não estar isolado” queremos dizer que suas fronteiras não são adiabáticas e portanto permitem a troca de energia com as vizinhanças, que neste caso é imaginada como um banho de temperatura constante.

As fronteiras diatérmicas e a temperatura constante do banho externo, é fácil perceber que todas as vezes que o sistema estiver fora de equilíbrio térmico com suas vizinhanças, sua energia interna irá variar devido ao calor.

Até então temos trabalhado com a ideia de um ensemble em um sistema cuja energia permanece constante, mas a temperatura flutuava. Agora, se desenvolvermos um algoritmo para descrever o *acoplamento* entre as vizinhanças e o sistema, podemos usar a variação da energia para controlar a temperatura.

Como já vimos na equação (24), existe um resultado da termodinâmica estatística que relaciona a média da energia cinética à temperatura macroscópica do sistema. Por este motivo, Berendsen e seus colaboradores propuseram um algoritmo de termostato que altera as velocidades na tentativa de controlar a temperatura (44).

$$\dot{V} = M^{-1} F + \frac{1}{2\tau_T} \left(\frac{T_0}{T} - 1 \right) V \quad (45)$$

O termo T_0 refere-se à temperatura desejada de equilíbrio, que é a temperatura do banho

externo, T é a temperatura instantânea do sistema, definida na equação (25) e τ_T é uma constante de acoplamento.

Esta é na verdade uma variação da equação de Newton convencional com termos adicionais. Estes termos adicionais colocados na equação interferem na aceleração das partículas (e portanto nas velocidades), gerando uma força friccional que é negativa, diminuindo as acelerações, quando o banho está a uma temperatura inferior à do sistema, e positiva quando ocorre o contrário.

Esta aceleração ou desaceleração das partículas, em geral, perturba a energia cinética para mais ou para menos, interferindo assim na temperatura.

A forma como esta fricção interfere na aceleração depende da constante de acoplamento. Se ela for muito grande, o sistema se acopla apenas fracamente às vizinhanças e as acelerações são apenas suavemente amortecidas, reduzindo a temperatura ao longo de uma curva muito longa de tempo. Se, por outro lado, a constante de acoplamento é pequena, as diferenças de temperatura produzem fortes perturbações na energia cinética do sistema.

O termostato de Berendsen, como já dito, só amostra o ensemble NVT aproximadamente para sistemas muito grandes. Alguns vezes o que se faz é equilibrar o sistema com o termostato de Berendsen e amostrar os valores com o ensemble NVE, sem o acoplamento com a temperatura.

1.2.2.8.1.2. Andersen

Como já citado anteriormente, entretanto, o termostato de Berendsen não garante que a trajetória será amostrada do ensemble canônico, com resultados apenas aproximado para sistemas muito grandes.

O termostato mais simples capaz de amostrar rigorosamente uma trajetória no ensemble NVT talvez seja o termostato de Hans C. Andersen (ANDERSEN, 1980). Neste termostato, o controle da temperatura é feito através de colisões entre os átomos do sistema e partículas friccionais virtuais. Na prática isto quer dizer trocar os momentos lineares de alguns átomos, fazendo isso seguindo a distribuição de Boltzmann para a temperatura desejada (46).

$$\rho(\mathbf{p}) \propto \exp\left[-\frac{\mathbf{p}^2}{2m K_B T}\right] \quad (46)$$

A equação (46) relaciona os valores do vetor do momento linear com uma probabilidade de ele ser selecionado. As colisões são descritas como um processo aleatório discreto (Poisson) de forma que a probabilidade de uma colisão acontecer em um Δt é descrita pela equação (47).

$$P(\Delta t, \tau) = 1 - \exp\left[-\frac{\Delta t}{\tau}\right] \quad (47)$$

O parâmetro τ representa o intervalo de tempo médio em que as colisões acontecem. O acoplamento mais forte possível substitui todos os momentos de todos os átomos, substituindo-os

por valores tomados da distribuição de Boltzmann (46).

Formais mais suaves de acoplamento envolvem substituir os momentos de apenas alguns átomos, ou corrigir os momentos dos átomos com uma mistura dos momentos p_B retirados da distribuição de Boltzmann com os momentos originais p_0 (48).

$$p \leftarrow \sqrt{1 - \alpha^2} p_0 + \alpha p_B \quad (48)$$

Onde o parâmetro alfa é uma estimativa da magnitude das colisões dos átomos com as partículas virtuais.

1.2.2.8.1.3. Langevin

O primeiro algoritmo utilizando o termostato de Langevin foi proposto por Gary S. Grest e Kurt Kremer, em 1986 como uma forma geral e eficiente de simular polímeros (GREST; KREMER, 1986). O método foi batizado em homenagem ao físico francês Paul Langevin.

O termostato de Langevin parte de uma ideia similar à de Andersen. Assim como no termostato de Andersen, o termostato de Langevin também utiliza partículas virtuais capazes de transferir momento linear aos átomos, mas onde o primeiro as utilizava intermitentemente, o último assume que o sistema está mergulhado em um mar de pequenas partículas “térmicas” fictícias movendo-se caoticamente.

À medida que os átomos se deslocam neste mar caótico de partículas virtuais, as pequenas colisões aleatórias perturbam a movimentação “regular” dos átomos do sistema, calculada segundo a integração das equações de movimento. Assim como o termostato de Andersen, a magnitude da interação do sistema com as partículas virtuais pode ser ajustada segundo um parâmetro.

O formalismo do termostato de Langevin, as equações de movimento são modificadas. O cálculo das forças sofre uma ligeira mudança para incluir as modificações aleatórias devido às colisões com as partículas virtuais (49).

$$M \ddot{\mathbf{R}}(t) = -\nabla U(\mathbf{R}) - M \Gamma \dot{\mathbf{R}}(t) + \mathbf{W}(t) \quad (49)$$

Onde o parâmetro Γ é um coeficiente de atrito que indica o quão amortecidas estão as forças e velocidades originais e $\mathbf{W}$ é um vetor com forças geradas aleatoriamente tomadas a partir de uma distribuição Gaussiana.

Em simulações utilizando o termostato de Langevin com um gama tal que não haja aceleração média no sistema, é possível simular movimento Browniano.

1.2.2.8.2. Barostatos

Ao contrário da ligação até mesmo intuitiva da temperatura com a agitação térmica representa pela energia cinética do sistema, a pressão não possui um equivalente microscópico tão intuitivo, por assim dizer. Antes de entrarmos diretamente no problema, considere o seguinte exemplo prático do conceito do qual vamos partir.

Suponha que temos um carro andando em um estacionamento, em busca de uma vaga. Se ele andar por lá durante um longo tempo, a velocidade média dele será zero porque não houve modificação na distância percorrida *resultante*.

Abstraindo o exemplo, trazemos para sistemas moleculares. Não faz sentido pensar em pressão sem pensar em volume e se estamos falando de volume então o sistema encontra-se enclausurado em um espaço de dimensões finitas. Significa dizer que a função $r(t)$ é **ligada** apenas àqueles valores que se encontram dentro deste volume, para quaisquer valores de t .

Da estatística de variáveis contínuas, temos que a média de uma função $f(x)$ é definida como sendo o valor médio da função ao longo de todo o domínio:

$$\langle f(x) \rangle = \left(\frac{1}{b-a} \right) \int_a^b f(x) dx \quad (50)$$

Usando a relação (50) podemos estabelecer que a média de uma velocidade $\dot{r}(t)$ se modificando no tempo é definida por:

$$\left\langle \frac{d\mathbf{r}}{dt} \right\rangle = \langle \dot{\mathbf{r}}(t) \rangle = \left(\frac{1}{t_2 - t_1} \right) \int_{t_1}^{t_2} \left(\frac{d\mathbf{r}}{dt} \right) dt = \frac{\mathbf{r}(t_2) - \mathbf{r}(t_1)}{t_2 - t_1} \quad (51)$$

Como os valores de $\mathbf{r}$ têm um limite máximo e mínimo, o numerador da função pode atingir, no máximo, um valor k finito. O denominador, por outro lado, pode incluir um intervalo de tempo tão grande quanto se deseje. Isto quer dizer que $\left| \frac{\mathbf{r}(t_2) - \mathbf{r}(t_1)}{t_2 - t_1} \right| < \frac{k}{t_2 - t_1}$.

À medida que o intervalo de tempo integrado cresce, o valor da média temporal da derivada de $\mathbf{r}(t)$ aproxima-se assintoticamente de zero. Esse exemplo foi feito com posições no tempo, mas é de fato geral. A média assintótica da derivada de qualquer função ligada é zero.

Se tanto a função $\mathbf{r}(t)$ quanto a função $\dot{\mathbf{r}}(t)$ são funções ligadas em t , então seu produto também o é, de modo que:

$$\left\langle \frac{d}{dt} (\mathbf{r} \cdot \dot{\mathbf{r}}) \right\rangle = \langle \mathbf{r} \cdot \ddot{\mathbf{r}} + |\dot{\mathbf{r}}|^2 \rangle = \langle \mathbf{r} \cdot \ddot{\mathbf{r}} \rangle + \langle |\dot{\mathbf{r}}|^2 \rangle = 0$$

Como todos os termos desta igualdade são iguais entre si e iguais a zero, a equação (52) é verdadeir.:

$$\langle \mathbf{r} \cdot \ddot{\mathbf{r}} \rangle + \langle |\dot{\mathbf{r}}|^2 \rangle = 0 \quad (52)$$

Multiplicando-se esta equação por uma massa m , temos:

$$\langle \mathbf{r} \cdot m \cdot \ddot{\mathbf{r}} \rangle + \langle m \cdot |\dot{\mathbf{r}}|^2 \rangle = 0$$

Da equação de Newton temos que $\mathbf{F} = m \cdot \ddot{\mathbf{r}}$ e a definição clássica de energia cinética diz que $K = \frac{m \cdot |\dot{\mathbf{r}}|^2}{2}$. Levando em consideração todos os termos, temos que:

$$\langle \mathbf{r} \cdot \mathbf{F} \rangle + 2 \langle K \rangle = 0 \quad (53)$$

Para um conjunto de N partículas, com posições ligadas $\mathbf{r}_i$ sujeitas a forças ligadas $\mathbf{F}_i$ com energia cinética total igual a K , a equação (53) torna-se:

$$\left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle + 2 \langle K \rangle = 0$$

Que rearranjada, toma a seguinte forma:

$$\left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i \right\rangle = -2 \langle K \rangle \quad (54)$$

O termo $O = \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_i$ é chamado de virial do sistema. Este termo foi primeiro descrito pelo físico alemão Rudolf Julius Emanuel Clausius, em 1870. Para um sistema de posições e forças restringidas, o virial permite o cálculo de sua energia cinética, mesmo para estados fora do equilíbrio.

Estamos assumindo que este sistema preso em um volume V está exercendo uma pressão P sobre as paredes do recipiente. O que significa que uma pequena área ΔA da parede exerce uma contrapressão de magnitude $-P \Delta A$ para dentro do recipiente.

Se pegarmos uma área muito pequena da parede, *infinitesimal* na verdade, podemos dizer que a contrapressão $-P \mathbf{n} dA$ (onde $\mathbf{n}$ é o vetor normal à superfície da parede) exercida pelo elemento de área infinitesimal da superfície atua nas partículas *próximas*, combatendo a pressão exercida por elas. Chamamos esta parte do virial de virial externo, W . A parte do virial que contabiliza interações partícula-partícula nós chamamos de virial interno, denotado por X . De forma que o virial total O é igual à soma do virial externo com o virial interno.

$$O = W + X \quad (55)$$

Como dA é muito pequeno, a contrapressão agindo sobre as partículas é muito pequena, o que quer dizer que as posições $\mathbf{r}_i$ das partículas próximas, onde o virial externo atua, é aproximadamente $\mathbf{r}$. Onde $\mathbf{r}$ representa um vetor posição partindo da origem e chegando na região dA das paredes (Figura 14).

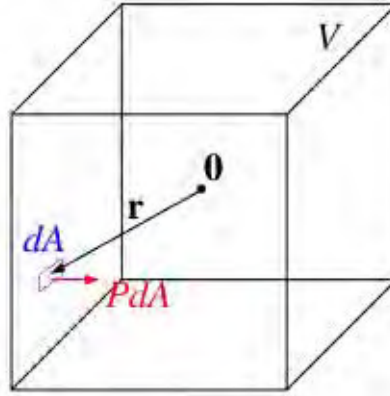


Figura 14: Posicionamento relativo do vetor $\mathbf{r}$ com o elemento de área da parede

Utilizando-se desta aproximação, podemos escrever o virial externo para aquela pequena porção de parede da seguinte forma:

$$\left\langle \sum_{i=1}^N \mathbf{r}_i \cdot \mathbf{F}_{P,i} \right\rangle = \left\langle \mathbf{r} \sum_{i=1}^N \mathbf{F}_{P,i} \right\rangle = -P \mathbf{r} \cdot \mathbf{n} dA$$

A forma do virial externo pode então ser expressa integrando-se sobre toda a superfície fechada que envolve o volume V . Esta operação é denotada pela expressão (56).

$$W = -P \oint_A (\mathbf{r} \cdot \mathbf{n}) dA \quad (56)$$

Aqui faremos uso de uma relação matemática chamada de teorema da divergência. O teorema da divergência diz o seguinte: suponha eu tenho um volume V , subconjunto de um espaço $\mathbb{R}^n$ (onde aqui o $n=3$ para um espaço tridimensional), que é compacto e possui uma superfície S suave em todas as suas partes como fronteira. Para qualquer vetor continuamente diferenciável $\mathbf{F}$ definido nas vizinhanças de V temos que:

$$\iiint_V (\nabla \cdot \mathbf{F}) dV = \oint_S (\mathbf{F} \cdot \mathbf{n}) dS \quad (57)$$

O lado esquerdo desta equação representa a integral sobre todo o volume V e o lado direito refere-se à integral sobre a fronteira S de V . $\nabla \cdot \mathbf{F}$ É o *divergente* do campo vetorial $\mathbf{F}$.

O teorema da divergência formaliza a noção de que o fluxo resultante de um campo vetorial para dentro ou para fora de um volume V qualquer delimitado por uma superfície suave e fechada S tem que ser igual à resultante de $\mathbf{F}$ passando através desta superfície. Isto faz sentido porque qualquer vetor que entre em V deve cruzar S , uma vez que S delimita V e é uma superfície fechada.

Quando o vetor em questão é a posição média das partículas $\mathbf{r}$, o vetor resultante de $\mathbf{r}$ passando por uma superfície A , quando multiplicado pela pressão, é o virial externo. Assim, aplicando-se a relação (56) na equação (57) vemos o seguinte:

$$-P \iiint_V (\nabla \cdot \mathbf{r}) dV = -P \oint_A (\mathbf{r} \cdot \mathbf{n}) dA$$

Que nos leva à conclusão de que $W = -P \iiint_V (\nabla \cdot \mathbf{r}) dV$. Onde o termo $\nabla \cdot \mathbf{r}$ é a divergência do campo vetorial $\mathbf{r}(t)$. Para um sistema de coordenadas cartesianas x, y, z no espaço euclidiano, a divergência de um campo vetorial continuamente diferenciável escrito da forma $\mathbf{r} = U\mathbf{i} + V\mathbf{j} + W\mathbf{k}$ é igual à função escalar:

$$\nabla \cdot \mathbf{r} = \frac{\partial U}{\partial x} + \frac{\partial V}{\partial y} + \frac{\partial W}{\partial z}$$

Para um sistema onde a origem cartesiana coincide com a origem da base $\{\mathbf{i}, \mathbf{j}, \mathbf{k}\}$ do vetor $\mathbf{r}$, a divergência é constante e igual a 3. O que leva à conclusão de que:

$$W = -3PV \quad (58)$$

Juntando as equações (54), (55) e (58), temos:

$$P = \frac{2}{3V} \left[\langle K \rangle + \frac{1}{2} X \right] \quad (59)$$

Como já explicado, o virial interno X é um termo virial dependente de forças entre partículas e sua expressão é equivalente àquela do virial comum.

É conveniente expressar este termo em uma forma que não dependa da origem das coordenadas do sistema. Isto pode ser feito expressando $\mathbf{F}_i$ como a soma das forças $\mathbf{F}_{ij}$ atuando no átomo i devido ao átomo j

$$\Xi = \sum_i \mathbf{r}_i \cdot \mathbf{F}_i = \sum_i \sum_{j \neq i} \mathbf{r}_i \cdot \mathbf{F}_{ij} = \frac{1}{2} \sum_i \sum_{j \neq i} \mathbf{r}_i \cdot \mathbf{F}_{ij} + \mathbf{r}_j \cdot \mathbf{F}_{ji}$$

Aplicamos a terceira lei de Newton para reverter os índices $F_{ij} = -F_{ji}$ e ficamos com:

$$X = \sum_i \sum_{j>i} \mathbf{r}_{ij} \cdot \mathbf{F}_{ij} \quad (60)$$

Quando a pressão não é isotrópica (isto é, quando ela não age da mesma forma ao longo de todos os pontos da superfície), a pressão do sistema é descrita pela tensão Π , que é uma medida da intensidade com que as forças internas de um corpo deformável agem sobre este corpo.

$$\Pi = \begin{pmatrix} \Pi_{xx} & \Pi_{xy} & \Pi_{xz} \\ \Pi_{yx} & \Pi_{yy} & \Pi_{yz} \\ \Pi_{zx} & \Pi_{zy} & \Pi_{zz} \end{pmatrix}$$

Onde o elemento Π_{xy} , por exemplo, é a pressão exercida por uma força na direção y agindo em um setor da superfície com vetor normal na direção x de acordo com a seguinte relação:

$$\Pi_{lm} = \frac{2}{3V} \sum_i \left[m_i (\dot{x}_l^i - \bar{x}_l) (\dot{x}_m^i - \bar{x}_m) + \frac{1}{2} \sum_{j>i} (x_l^i - x_l^j) \cdot F_m^{ij} \right] \quad (61)$$

Onde os índices l e m representam as componentes dos vetores de posição e velocidade e os índices i e j representam os átomos do sistema. A força F_m^{ij} denota a componente m da força entre os átomos i e j .

Isto pode ser pensado da seguinte forma: O elemento Π_{11} representa uma força na direção x dividido por um plano perpendicular a esta mesma direção. Considerando-se dois volumes separados por ambos os planos, este elemento é a soma de todas as interações entre pares de átomos de ambos os lados.

Apenas em sistemas isotrópicos, onde não há deformações no formato da caixa, pois a pressão é igual em todas as direções, é que os termos fora da diagonal da matriz de tensão são zero (pois não há cisalhamento), e $P = \Pi_{xx} = \Pi_{yy} = \Pi_{zz}$. Nestes casos a pressão pode ser expressa como a média escalar dos componentes normais à superfície (61).

$$P = \frac{Tr(\Pi)}{3} \quad (62)$$

A ideia central dos algoritmos discutidos abaixo é modificar os valores calculados para o virial interno X ajustando as distâncias e as forças entre as partículas.

1.2.2.8.2.1. Berendsen

No barostato de Berendsen, o controle da pressão segue princípios semelhantes àqueles descritos para a temperatura no termostato de mesmo nome, interferindo não apenas nos cálculos da posição, mas também do volume do sistema.

$$\dot{R} = V - \frac{\beta}{3\tau_P} (P_0 - P) R \quad (63)$$

$$\dot{V} = -\frac{\beta}{\tau_P} (P_0 - P) V \quad (64)$$

Aqui, os termos P_0 , P e τ_P tomam papéis análogos ao T_0 , T e ao τ_T do controle da temperatura imaginando um “banho isobárico” com o qual o sistema tenta entrar em equilíbrio mecânico. A constante β é a *compressibilidade isotérmica* do sistema. A razão β/τ_P representa a constante de acoplamento do sistema com um “banho de pressão constante”. Nestas equações, V denota o volume do sistema, enquanto $\dot{V}$ denota o vetor velocidades do sistema.

Quando utilizamos este algoritmo para corrigir as posições e velocidades do sistema, o tamanho da caixa é corrigido por um fator η de tal forma que:

$$\eta = 1 - \frac{\Delta t}{\tau_P} \beta (P_0 - P) \quad (65).$$

Em uma correção isotrópica, a caixa não muda de formato com o tempo, isto nos permite calcular a pressão da forma descrita na equação (62).

1.2.2.8.2.2. Andersen

Ao contrário do barostato de Berendsen, que segue uma lógica similar à do termostato, o barostato de Andersen é um método desenvolvido sobre bases não análogas ao termostato proposto pelo mesmo Andersen.

O barostato de Andersen envolve o acoplamento do sistema com uma variável externa V , o volume da caixa de simulação. Este acoplamento imita o comportamento de um pistão agindo sobre o sistema. O pistão possui uma massa Q e suas energias potencial $U_V = P_0 V$

e cinética $K_V = \frac{1}{2} Q \dot{V}^2$ associadas.

U_V é a energia potencial oriunda de uma pressão P_0 externa agindo sobre o pistão. As coordenadas e velocidades são dadas em termos de coordenadas corrigidas r' .

$$\mathbf{r}_i = V^{1/3} \mathbf{r}'_i \quad (66), \quad \dot{\mathbf{r}}_i = V^{1/3} \dot{\mathbf{r}}'_i \quad (67)$$

Com isso chegamos às equações de movimento relacionando

$$\ddot{\mathbf{r}}'_i = \frac{\mathbf{F}_i}{m_i V^{1/3}} - \frac{2}{3} \frac{\dot{\mathbf{r}}'_i \dot{V}}{V} \quad (68) \text{ e } \ddot{V} = \frac{P - P_0}{Q} \quad (69).$$

Onde a pressão instantânea P do sistema e as forças $\mathbf{F}_i$ são calculados utilizando os valores não corrigidos de momentos e posições. O Hamiltoniano (70) do sistema é conservado e é igual à entalpia do sistema, mais um termo $(1/2)K_B T$ associado com o movimento do pistão, ou seja, com a flutuação do volume.

$$H = K + U + K_V + U_V \quad (70)$$

As equações de movimento (70) produzem um ensemble isobárico-isoentálpico NHP. Para utilizar este barostato para amostrar o ensemble NPT é preciso combiná-lo com um dos termostatos.

A massa Q do pistão é um parâmetro de entrada. Uma massa muito pequena produz oscilações muito rápida no volume que não são eficientemente compensadas pelo movimento molecular. Massas maiores, ao contrário, produzem suaves ajustes no volume (e na pressão). Uma massa grande demais, entretanto, regressa às equações de dinâmica molecular para o ensemble NVE original.

1.2.2.9. Análise da Simulação

Os métodos de simulação discutidos anteriormente utilizados para propagar o sistema no tempo de forma a amostrar a função de partição e produzir uma trajetória de dinâmica molecular com significado físico.

Antes de prosseguir, entretanto, é necessário ser bem específico sobre o que queremos dizer com uma “trajetória de dinâmica molecular”. Queremos dizer que se trata de um conjunto de microestados (configurações microscópicas de posições e velocidades dos átomos) do sistema, organizados em sequência. Como estes microestados foram conseguidos sondando-se o ensemble de equilíbrio com equações de dinâmica, então trata-se de uma *série temporal* de microestados.

Calculando-se uma dada propriedade microscópica χ para cada microestado desta série, obtemos uma série temporal desta propriedade. Com esta série temporal é possível extrair informações estatísticas sobre a tal propriedade. Segundo a hipótese ergódica, estas estatísticas feitas sobre uma série temporal equilibrada e suficientemente longa têm a mesma representatividade que as estatísticas tomadas sobre o ensemble de onde estes microestados foram amostrados.

Duas das análises estatísticas mais simples que podemos retirar destas séries temporais são *médias* (71) e *flutuações* (72). A importância destas quantidades é que quando elas são calculadas para séries temporais de certas propriedades microscópicas, servem como uma estimativa de observáveis termodinâmicos.

$$\langle X \rangle = \frac{1}{N} \sum_{n=1}^N X_n \quad (71)$$

$$\langle \sigma_X^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2 \quad (72)$$

E evolução temporal das médias das propriedades de temperatura, pressão e energia total, por exemplo, ajudam no monitoramento do equilíbrio da simulação. Podemos calcular a média da temperatura utilizando a descrição microscópica desta, que se relaciona com a energia cinética segundo a equação (25).

$$T = \frac{1}{N_A K_B T} K$$

Onde o valor $\langle T \rangle$ ao longo da trajetória representa a temperatura do sistema que o ensemble amostrado pela trajetória representa. A evolução temporal da temperatura ajuda no acompanhamento da qualidade dos termostatos empregados.

A própria energia cinética pode ser amostrada sobre a trajetória. Como estamos utilizando um campo de força clássico, o cálculo da energia cinética instantânea para um dado microestado é relativamente direto.

$$K = \frac{1}{2} \sum_i m_i \dot{r}_i^2 \quad (73)$$

Este valor pode ser combinado, juntamente com o valor instantâneo para o potencial $U(\mathbf{R})$ para calcular a energia total de cada passo temporal: $E = K + U$.

Normalmente, as flutuações das propriedades são utilizadas para mensurar a estabilidade relativa de uma propriedade. No entanto, algumas propriedades podem ser ativamente calculadas a partir das flutuações de algumas grandezas. A capacidade calorífica a volume constante C_v , por exemplo, tem relação com a flutuação da energia termodinâmica do ensemble canônico (74).

$$\sigma_E^2 = K_B T^2 C_v \quad (74)$$

O acompanhamento temporal destas propriedades permite não apenas observar quando o sistema atinge o equilíbrio como também verificar quebras neste equilíbrio. Esta ferramenta é útil para avaliar instabilidades numéricas nos algoritmos bem como para verificar mudanças espontâneas no estado de equilíbrio, provocadas por fenômenos inerentes da física com que o sistema é simulado.

1.2.2.10. Energia Livre

Importante como possam ser as propriedades anteriores, a propriedade de maior interesse para este trabalho, que é o objeto de análise fundamental do estudo proposto aqui, é a energia livre. No entanto, enquanto as propriedades anteriores dizem respeito a um sistema, apenas, a energia livre é medida como a diferença entre dois desses.

No caso do estado macroscópico do sistema ser descrito pelo ensemble canônico, a energia livre associada é a energia livre de Helmholtz, primeiro formalizada pelo físico e fisiologista alemão Hermann von Helmholtz. Que pode ser convenientemente calculada a partir da função de partição (75).

$$A = -K_B T \ln Q \quad (75)$$

Para calcular a energia de Helmholtz para uma simulação, portanto, pegamos a expressão para a função de partição Q e resolvemos a expressão acima.

$$Q(N, V, T) = \sum_i^k \exp \left[-\frac{1}{K_B T} \varepsilon_i(N, V) \right] \quad (76)$$

Multiplicando a equação (72) no numerador e no denominador por $\exp \left[\frac{1}{K_B T} - \varepsilon_i(N, V) \varepsilon_i(N, V) \right]$ e utilizando a equação (10) podemos concluir que:

$$Q = \left\langle \exp \left[\frac{\varepsilon_i}{K_B T} \right] \right\rangle^{-1} \quad (77)$$

E com isso, temos uma fórmula bastante conveniente para calcular a energia livre total do sistema a partir de alguma média sobre a trajetória substituindo a equação (77) na equação (71) para obter a equação (78).

$$A = -K_B T \ln \left\langle \exp \left[\frac{\varepsilon_i}{K_B T} \right] \right\rangle^{-1} \quad (78)$$

Com esta expressão nós poderíamos conseguir uma trajetória, calcular a média daquele termo em cada passo da simulação, tirar a média e usar a equação (78) para conseguir um resultado para a energia livre. Poderíamos então calcular a energia livre para um sistema e depois para outro e tomar a diferença para obter a variação.

A raiz do problema está no expoente negativo da função de partição. Da forma como este termo é calculado, os microestados que terão maior peso na média da função de partição serão justamente aqueles de *maior* energia, por causa desse expoente. Entretanto, o algoritmo da dinâmica molecular amostra preferencialmente os estados de mais baixa energia.

Por este motivo, é preciso realizar um tempo absurdamente longo de simulação, para garantir que os microestados de maiores energia (e que são amostrados de vez em nunca pelo algoritmo de simulação) sejam amostrado adequadamente, revelando o espaço de fase mais improvável do ensemble e permitindo o cálculo da energia livre.

Felizmente, esta não é a única forma de se calcular a energia livre de um processo. A energia

livre também pode ser calculada como o trabalho *reversível* necessário para transformar o sistema de um estado em outro.

Para calcular a energia livre desta maneira, vários algoritmos foram desenvolvidos. Dois dos mais importantes são discutidos a seguir, juntamente com um método novo e aproximado para se obter estimativas da energia livre.

1.2.2.10.1. Integração Termodinâmica

Considere dois sistemas, 1 e 2. Cada um deles pode ser simulado separadamente, cada um deles possui um estado de equilíbrio por si só. No entanto, na natureza, verificamos que o estado 1 muitas vezes torna-se espontaneamente o estado 2 (Figura 15).

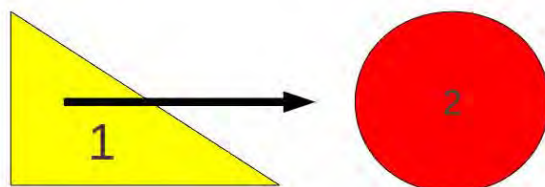


Figura 15: Uma transformação qualquer.

Para monitorar a variação da energia livre associada a esta transformação, precisamos sondar os estágios intermediários. A técnica da Integração Termodinâmica *acopla* os dois sistemas de modo que haja uma forma de transformar a descrição do sistema 1 na descrição do sistema 2, para que seja possível amostrar os estágios intermediários entre os dois. No caso os potenciais estão acoplados:

$$U(\mathbf{P}, \lambda) = (1 - \lambda)U_1 + \lambda U_2 \quad (79)$$

Este parâmetro λ é o parâmetro de acoplamento. Ele deve estar conectado à forma pela qual um sistema torna-se outro. A forma exata como o parâmetro é utilizado para transformar um sistema em outro (isto é, o caminho empregado na transformação) depende do problema em questão e daquilo que é computacionalmente viável.

Por exemplo, uma forma de utilizar o acoplamento é a seguinte: imagine que temos dois sistemas, um possui átomos diferentes do outro. O algoritmo de acoplamento poderia lentamente mudar os parâmetros dos átomos em 1 para os átomos em 2 em função de λ de modo que em $\lambda=0$ a simulação utiliza os parâmetros de 1 e em $\lambda=1$ a simulação utiliza os parâmetros de 2. E assim com todos os valores intermediários.

Com isso, à medida que o parâmetro de acoplamento aumenta ou diminui, o sistema simulado se parece mais ou menos com um dos extremos (Figura 16).

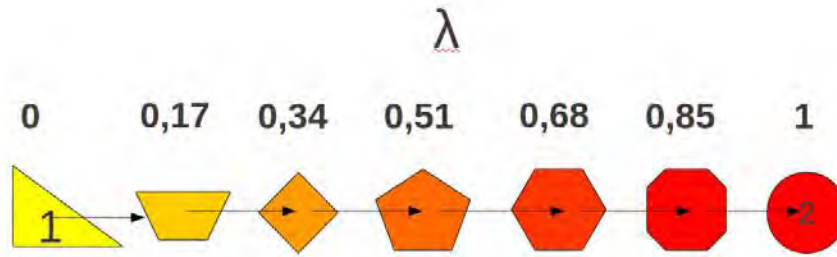


Figura 16: Esquema conceitual mostrando uma série de estados acoplados

Utilizando um potencial acoplado, teremos uma função de partição acoplada. Para calcular efetivamente a energia livre, temos que calcular o trabalho necessário para transformar *reversivelmente* o sistema de 1 para 2.

Na termodinâmica, reversivelmente significa dizer infinitesimalmente. Os valores de λ são variados de forma muito suave, infinitesimal na verdade, assim como os valores de A . A integral sob a curva formada desta variação infinitesimal é o valor da energia livre.

$$\Delta A_{1 \rightarrow 2} = \int_{\lambda_1}^{\lambda_2} \frac{dA(\lambda)}{d\lambda} d\lambda \quad (80)$$

Para calcular esta integral definida, primeiro vamos encontrar a forma indefinida.

$$\begin{aligned} A(\lambda) &= -K_B T \ln Q(\lambda) \\ \frac{dA(\lambda)}{d\lambda} &= -K_B T \frac{d \ln Q(\lambda)}{d\lambda} \\ \frac{d \ln Q(\lambda)}{d\lambda} &= \frac{1}{Q} \frac{dQ(\lambda)}{d\lambda} \\ \frac{dA(\lambda)}{d\lambda} &= -K_B T \frac{1}{Q} \frac{dQ(\lambda)}{d\lambda} \quad (81) \end{aligned}$$

Note que se o Hamiltoniano está acoplado, a função de partição também está, pois ela depende da energia do Hamiltoniano calculada para cada microestado de energia ϵ_i , calculada para um potencial U . Logo, substituindo este valor da derivada de Q na expressão (81), temos:

$$\begin{aligned}\frac{dA(\lambda)}{d\lambda} &= -K_B T \frac{1}{Q} \frac{dQ(\lambda)}{d\lambda} \\ \frac{dA(\lambda)}{d\lambda} &= \frac{1}{Q} \sum_i^k \exp\left[-\frac{1}{K_B T} \epsilon_i(\lambda)\right] \cdot \frac{d\epsilon_i(\lambda)}{d\lambda} \\ \frac{dA(\lambda)}{d\lambda} &= \frac{1}{Q} \sum_i^k \left(\frac{d\epsilon_i(\lambda)}{d\lambda}\right) \exp\left[-\frac{1}{K_B T} \epsilon_i(\lambda)\right]\end{aligned}$$

Na última expressão percebemos que esta expressão representa o valor esperado de $d\epsilon_i(\lambda)/d\lambda$ no ensemble de microestados postos sobre a curva de acoplamento. O que quer dizer que:

$$\frac{dA(\lambda)}{d\lambda} = \left\langle \frac{d\epsilon_i(\lambda)}{d\lambda} \right\rangle_\lambda$$

Que por sua vez implica que $\int \frac{dA(\lambda)}{d\lambda} = \int \left\langle \frac{d\epsilon_i(\lambda)}{d\lambda} \right\rangle_\lambda$. Com isso, chegamos à seguinte conclusão de que:

$$\Delta A_{1 \rightarrow 2} = \int_{\lambda_1}^{\lambda_2} \left\langle \frac{d\epsilon_i(\lambda)}{d\lambda} \right\rangle_\lambda d\lambda \quad (82)$$

Ao contrário dos valores absolutos de A , a derivada da energia do sistema pode ser prontamente calculada para os microestados que surgem na trajetória à medida que o parâmetro de acoplamento toma valores diferentes.

Esta curva é integrada numericamente a partir de uma tabela de valores de ΔA para cada valor de $\Delta \lambda$. Para que a integração numérica funcione, é preciso que as variações $\Delta \lambda$ sejam suficientemente pequenas, respeitando o princípio da reversibilidade, simulando da melhor maneira possível uma variação infinitesimal.

Esta curva pode ser obtida basicamente de duas formas diferentes. A primeira forma consiste em calcular uma única trajetória e nela mesma os valores de acoplamento são lentamente modificados, um infinitésimo de cada vez. Para cada incremento de acoplamento, esperamos o sistema equilibrar no novo estado e calculamos o valor esperado de $d\epsilon_i(\lambda)/d\lambda$ sobre os passos da trajetória que compreendem este valor de λ .

A principal desvantagem deste método é amostragem insuficiente. Mais conveniente seria escolher valores diferentes de acoplamento e calcular uma trajetória para cada um deles, obtendo os valores esperados de $d\epsilon_i(\lambda)/d\lambda$ para cada estado intermediário e depois realizando a integração da curva.

Qualquer que seja a forma, em ambos os casos a convergência só acontece quando $A(\lambda)$ é uma curva suave. E a suavidade (ou sua ausência) da função $A(\lambda)$ depende da forma como o acoplamento entre os dois sistemas é realizado. Isto ocorre porque, apesar de quimicamente falando a energia livre ser uma função de estado, na prática isto não acontece, pelo menos numericamente

falando em simulações computacionais. Devido a imprecisões numéricas, existem diferenças entre os caminhos escolhidos, o que reflete na forma de $A(\lambda)$ e consequentemente na capacidade de convergência do método.

Existem ainda outros esquemas de acoplamento que não são lineares, como o mostrado na equação (79) (RESAT; MEZEI, 1993), mas estes métodos em geral são mais custosos, difíceis de implementar e ainda mais de convergir.

1.2.2.10.2. Perturbação da Energia Livre

A teoria da perturbação é um poderoso formalismo da matemática aplicada, utilizada para descrever sistemas não-convencionais em termos de sistemas mais triviais. Quer seja aplicada em química quântica, física relativística ou engenharia aeroespacial, a ideia básica simples e elegante da teoria da perturbação se mantém inalterada.

A teoria da perturbação descreve o comportamento de um estado “alvo” em termos da descrição de um estado “de referência”. Ela faz isso descrevendo o estado alvo como sendo resultado de uma *perturbação* agindo sobre o estado de referência (Figura 17).

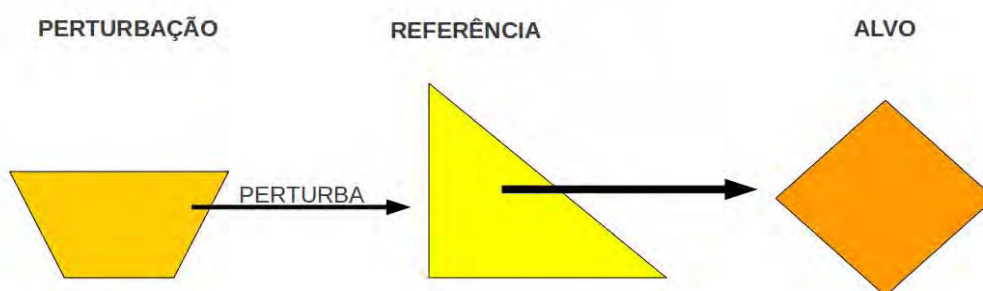


Figura 17: Esquema de uma perturbação

Um bom exemplo para ilustrar como isto acontece é justamente aquela que foi provavelmente a primeira aplicação da teoria da perturbação na física: descrever como as órbitas dos planetas se deformam com a presença de outros corpos celestes maciços no sistema além da estrela em torno do qual eles orbitam.

Este problema foi resolvido considerando-se a órbita elíptica descrita por um planeta em volta apenas de sua estrela como sistema de referência e adicionando as outras massas com uma perturbações gravitacionais adicionais.

Utilizando o mesmo princípio, podemos descrever um sistema com um potencial U_1 em função de um potencial U_0 de referência e uma perturbação ΔU que descreve todos os termos que existem em U_1 mas não em U_0 . Por exemplo, para avaliar a energia livre de solvatação do sódio por água, a perturbação incluiria a descrição de todos os termos envolvendo interações sódio-água.

$$U_1 = U_0 + \Delta U \quad (83)$$

Aqui retiramos a conclusão de que $\Delta U = U_1 - U_0$. Isto significa que se pudéssemos calcular uma energia $\epsilon^{(1)}$ para um potencial 1 e outra para um potencial 0, a diferença de energia entre as duas vai dá precisamente a perturbação $\Delta \epsilon$ na energia. Isto será crucial adiante porque precisaremos calcular esta quantidade.

Quando escrevemos um potencial desta forma, no entanto, considerando que um potencial pode ser escrito como uma correção perturbativa em outro potencial, obtemos uma expressão para o cálculo da energia livre envolvida na transformação de um estado em outro em termos de uma média sobre uma trajetória do estado de referência. Vamos ver como isto é feito.

Mais uma vez partimos da equação (71), já usada por nós anteriormente em outras deduções.

$$A = -K_B T \ln Q$$

A partir dela temos que:

$$\Delta A_{0 \rightarrow 1} = A_1 - A_0 = -K_B T \ln \left(\frac{Q_1}{Q_0} \right)$$

E também sabemos que:

$$Q = \sum_i^k \exp \left[-\frac{1}{K_B T} \epsilon_i \right]$$

A energia ϵ_i de cada microestado da dinâmica é calculada a partir de um potencial. Como temos dois potenciais, vamos denotar a energia calculada pelo potencial 1 como sendo $\epsilon_i^{(1)}$ e $\epsilon_i^{(0)}$ como sendo a energia calculada com o potencial 0 de tal modo que:

$$Q_j = \sum_i^k \exp \left[-\frac{1}{K_B T} \epsilon_i^{(j)} \right]$$

Neste caso:

$$\Delta A_{0 \rightarrow 1} = -K_B T \ln \left(\frac{\sum_i^k \exp \left[-\frac{1}{K_B T} \epsilon_i^{(1)} \right]}{\sum_i^k \exp \left[-\frac{1}{K_B T} \epsilon_i^{(0)} \right]} \right)$$

Então somamos no expoente do numerador o termo $\varepsilon_i^{(0)} - \varepsilon_i^{(0)}$. Uma vez que esta soma é zero, podemos fazer isto sem prejudicar a igualdade.

$$\left(\frac{Q_2}{Q_1}\right) = \left(\frac{\sum_i^k \exp\left[\frac{1}{K_B T} - \varepsilon_i^{(1)} + \varepsilon_i^{(0)} - \varepsilon_i^{(0)}\right]}{\sum_i^k \exp\left[-\frac{1}{K_B T} \varepsilon_i^{(0)}\right]} \right)$$

$$\left(\frac{Q_2}{Q_1}\right) = \left(\frac{\sum_i^k \exp\left[\frac{1}{K_B T} - (\varepsilon_i^{(1)} - \varepsilon_i^{(0)}) - \varepsilon_i^{(0)}\right]}{\sum_i^k \exp\left[-\frac{1}{K_B T} \varepsilon_i^{(0)}\right]} \right)$$

Definimos que $(\varepsilon_i^{(1)} - \varepsilon_i^{(0)}) = \Delta \varepsilon_i$, onde o $\Delta \varepsilon_i$ é a diferença de energia do microestado i quando calculada pelos potenciais 1 e 0. Separamos os dois termos do expoente no numerador segundo a propriedade da exponencial $e^{(a-b)} = e^a e^{(-b)}$ e ficamos com:

$$\left(\frac{Q_2}{Q_1}\right) = \left(\frac{\sum_i^k \exp\left[-\frac{1}{K_B T} \Delta \varepsilon_i\right] \exp\left[-\frac{1}{K_B T} \varepsilon_i^{(0)}\right]}{\sum_i^k \exp\left[-\frac{1}{K_B T} \varepsilon_i^{(0)}\right]} \right)$$

Aqui percebemos mais uma vez a semelhança com a equação (10) e chegamos à formulação do quociente das duas funções de partição em termos do valor esperado de uma propriedade possível de ser calculada para cada microestado ao longo de uma trajetória.

$$\left(\frac{Q_2}{Q_1}\right) = \left\langle \exp\left[-\frac{1}{K_B T} \Delta \varepsilon_i\right] \right\rangle_0 \quad (84)$$

onde o subscrito 0 indica a média é tomada sobre um ensemble do sistema no potencial 0. Com isso, a equação para o cálculo do ΔA torna-se:

$$\Delta A_{0 \rightarrow 1} = -K_B T \ln \left\langle \exp\left[-\frac{1}{K_B T} \Delta \varepsilon_i\right] \right\rangle_0 \quad (85)$$

A equação (85) é a equação fundamental da teoria da perturbação. Ela diz quando o potencial de um estado pode ser escrito como o potencial perturbado de outro estado em termos da teoria da perturbação, a variação na energia livre devido à transformação do estado de referência para o

estado perturbado pode ser calculada gerando-se uma trajetória para o sistema de referência e tirando a média daquela exponencial sobre esta trajetória.

Um ponto importante sobre esta equação é que se um potencial for muito diferente do outro, a energia da perturbação será muito grande e o valor da exponencial será muito pequeno e a convergência da média na simulação será muito lenta, ridiculamente lenta porque dependerá da amostragem de estados muito improváveis.

Por este motivo, já de saída os dois estados considerados precisam ser bastante parecidos entre si, ou o potencial perturbado (83) não será suficiente para descrever exatamente de que forma o potencial de um estado diferencia-se do potencial de outro.

Uma forma de contornar esta limitação é realizar um acoplamento parecido com aquele utilizado no método da integração termodinâmica.

$$U(\eta) = (1 - \eta)U_0 + \eta U_1 \quad (86)$$

É basicamente a mesma técnica que foi utilizada no método da integração termodinâmica, com exceção de que aqui o potencial 1 é uma correção perturbativa do potencial 0 e não potenciais independentes.

São definidos então vários estados intermediários de transformação (Figura 18) onde o potencial de cada um deles depende do acoplamento entre o potencial 0 e o potencial 1 (que por sua vez é definido como uma perturbação em 0).

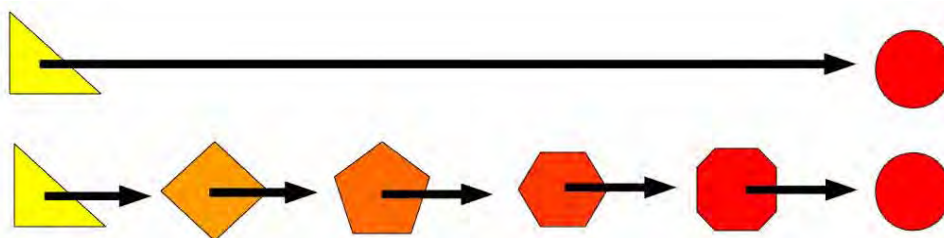


Figura 18: Esquema da quebra de uma perturbação em várias perturbações intermediárias acopladas

E como fizemos para a perturbação dos estados finais, podemos calcular a variação da energia livre para cada um dos pequenos passos, desde que os estados intermediários sondados guardem suficiente semelhança entre si.

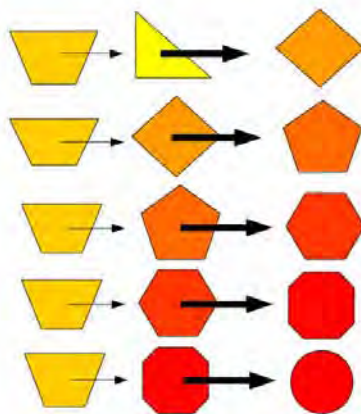


Figura 19: Cálculo de muitas perturbações

Para quaisquer dois estados intermediários i e j , temos:

$$\delta A_{i \rightarrow j} = A_j - A_i = -K_B T \ln \left\langle \exp \left[-\frac{1}{K_B T} \Delta \varepsilon_i(\eta) \right] \right\rangle_i \quad (87)$$

onde $\Delta \varepsilon_i(\eta) = U(\eta_j) - U(\eta_i)$.

Neste caso, a variação total da energia livre será o somatório das parcelas encontradas com a equação (87) para cada transição intermediária.

$$\Delta A_{0 \rightarrow 1} = \sum_i \sum_{j=i+1} \delta A_{i \rightarrow j} \quad (88)$$

Pela mesma relação (85) para o processo $0 \rightarrow 1$, o cálculo para a variação da energia livre no processo $1 \rightarrow 0$ pode ser realizado. A diferença entre os dois chama-se *histerese*, e é uma boa medida do quão exato foi o cálculo para a variação da energia livre. Em uma perfeita simulação da situação real, a variação da energia livre no processo inverso deveria ter a mesma magnitude da variação da energia livre no processo direto, mas com o sinal invertido.

A principal vantagem destes métodos de cálculo de energia livre é que o *acoplamento* entre os dois sistemas não precisa mimetizar um evento físico real. Por exemplo, se desejarmos medir a variação na energia livre de solvatação do íon sódio, podemos simplesmente definir uma forma de perturbação que inclua todas as interação envolvendo moléculas de solvente.

O estado de referência seria o sódio no vácuo e à medida que o acoplamento vai mudando o sistema as interações envolvendo moléculas de solvente são lentamente fortalecidas até o ponto onde elas assumem sua força plena, passando por estágios intermediários de “meias-ligações” que não correspondem a uma situação física real.

Similarmente, se desejarmos comparar as energias de solvatação entre um íon de sódio e outro de potássio, por exemplo, posso definir uma perturbação nos parâmetros eletrostático e de Lennard-Jones de modo que quando o sódio sofre esta perturbação, seus parâmetros tornam-se os do potássio. Ao fazer isto, o acoplamento lentamente muda o íon sódio para um potássio passando por

estágios intermediários que podem não ter qualquer significado físico.

O que acontece, no entanto, se quisermos medir a energia de interação entre um íon de sódio e outro de cloreto em solução aquosa? É uma estimativa interessante de se fazer porque tem a ver com propriedades de soluções eletrolíticas, que é um sistema experimentalmente simples de se trabalhar, o quer dizer que as estimativas seriam de ordem bastante prática.

Em princípio, poderíamos pensar que ao calcular esta variação, poderíamos montar um sistema como aquele mostrado na Figura 20, onde a perturbação seria a força da interação sódio-cloreto.

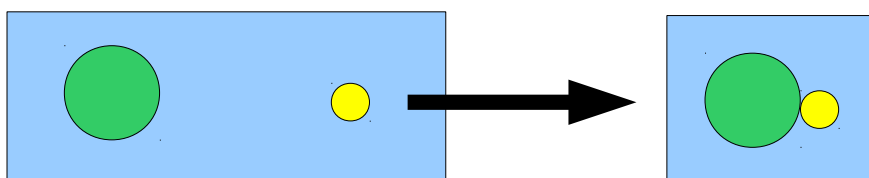


Figura 20: Formação de uma interação sódio-cloreto em solução

A questão é que a variação de energia medida para esta transformação teria um valor muito mais alto que o convencional porque as configurações seriam amostradas de trajetórias contabilizando uma boa parcela de interações das moléculas de solvente e os íons e das moléculas do solvente entre si. Estaríamos então, desprezando outra energia envolvida no processo. Acontece que se você retirar o solvente e calcular apenas a interação entre os íons, como mostrado na Figura 21, a sua estimativa também será errada, porque não levará em consideração as interações com o solvente que os íons precisam vencer para conseguirem formar uma ligação plena. Desta forma, também, a estimativa da energia livre para este processo seria irreal.

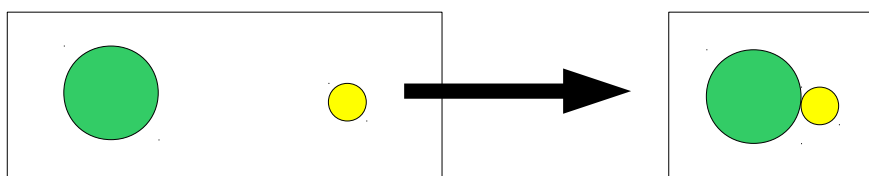
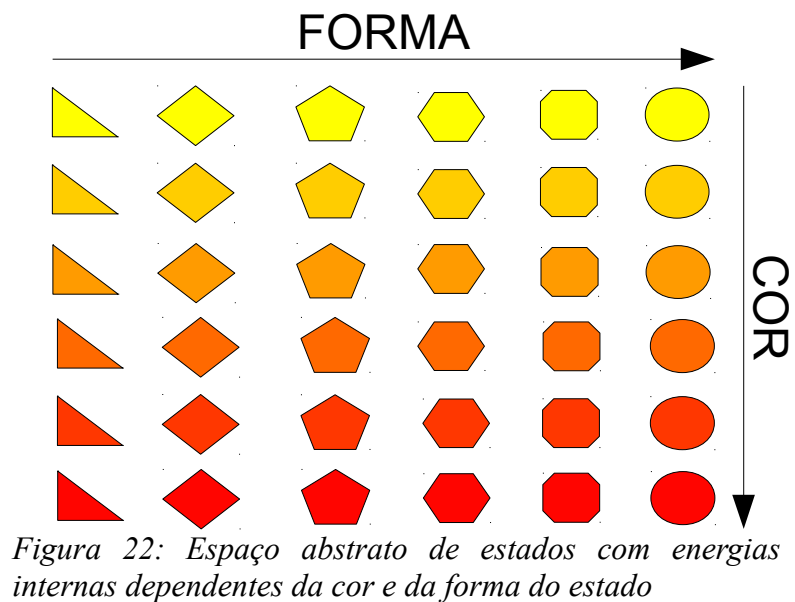


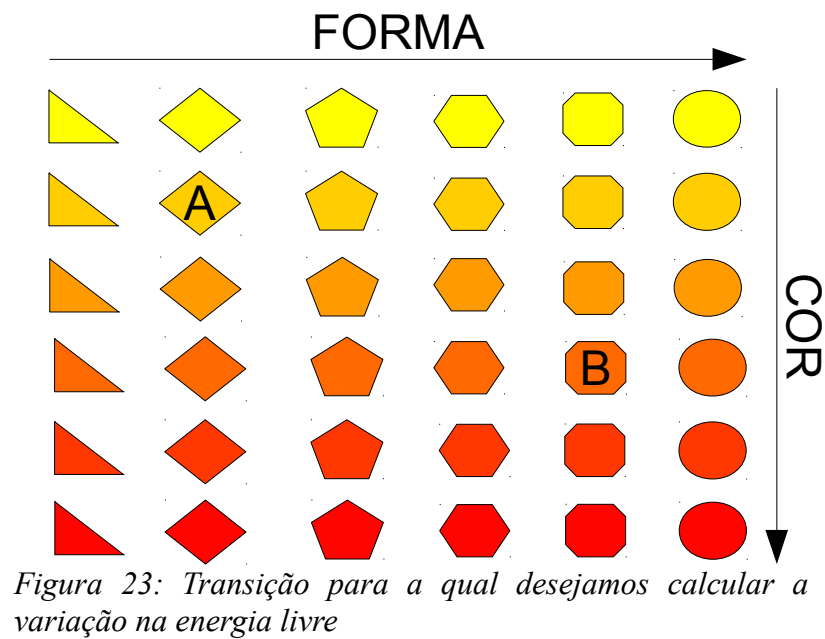
Figura 21: Formação de uma interação sódio-cloreto em fase gasosa

A alternativa é utilizar um *ciclo termodinâmico*. Uma ferramenta conceitual que nos auxilia a contabilizar todas as contribuições de energia livre envolvidas em transformações mais complexas ou quando mais de uma transformação se processa simultaneamente.

O primeiro passo é entender quais são as transformações simultâneas que estão acontecendo. Neste caso, o dessolvatamento dos íons e o surgimento de uma interação entre eles. Para propósitos didáticos, vemos na Figura 22 um diagrama que mostra vários estados de um sistema ilustrativo, onde a mudança de cor e forma do sistema são transformações onde há variação da energia livre.



Para calcular a variação de energia livre entre quais quer dois pontos da figura, precisaríamos montar um ciclo termodinâmico que contabilizasse tanto as transformações de cor quanto de forma. Considere, por exemplo, os sistemas A e B da Figura 23.



Então podemos montar um caminho cíclico que passe por ambos os estados de interesse de forma que cada transformação seja contabilizada uma de cada vez Figura 24.

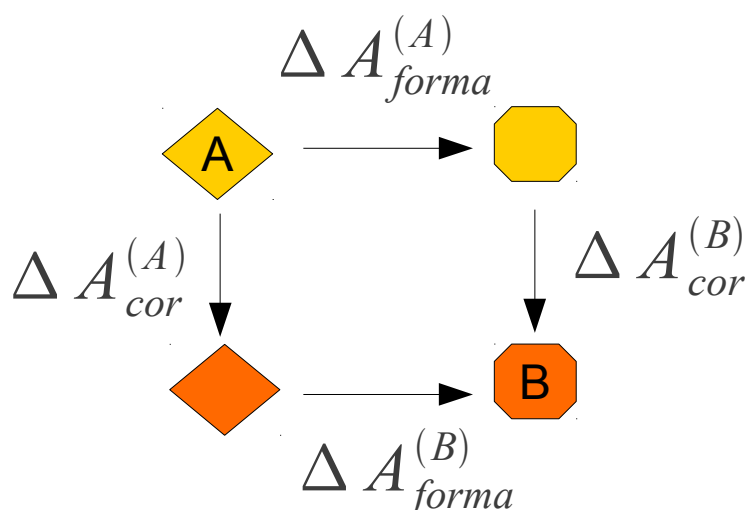


Figura 24: Ciclo termodinâmico para considerar todas as variações na energia livre decorrentes da transformação de A em B

Como a energia livre é função de estado, o ciclo tem a propriedade de que $\Delta A_{forma}^{(A)} + \Delta A_{cor}^{(B)} - \Delta A_{forma}^{(B)} - \Delta A_{cor}^{(A)} = 0$. De forma análoga, podemos montar um ciclo termodinâmico para calcular a energia dos nossos íons em solução de forma que contabilizemos realmente todas as transformações e consigamos uma boa estimativa para a nossa energia de interação.

Com o ciclo da Figura 25 podemos efetivamente ver que a variação da energia livre que acontece quando os dois íons interagem

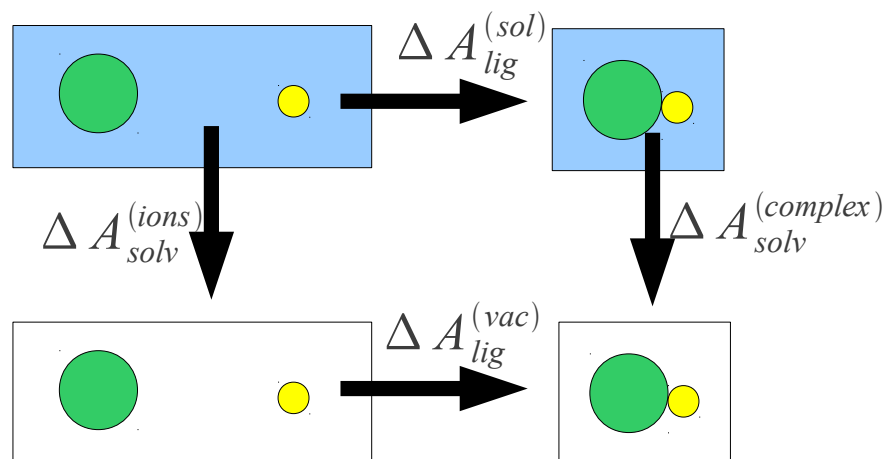


Figura 25: Ciclo termodinâmico para calcular a variação da energia livre quando ocorre a formação de uma ligação sódio-cloreto em solução

Com isso verificamos que a variação da energia livre para a formação da ligação entre os íons em solução é $\Delta A_{lig}^{(sol)} = +\Delta A_{lig}^{(vac)} - \Delta A_{solv}^{(complex)} + \Delta A_{solv}^{(ions)}$.

Para estimar adequadamente esta energia de interação, seria preciso então calcular todas estas variações de energia livre para considerar (mas não exageradamente) os efeitos decorrentes do fato de o sistema estar solvatado. Como pode-se perceber trata-se de uma quantidade imensa de cálculos de trajetórias.

Imagine agora, que ao invés de dois íons, temos uma enzima e um ligante. Uma enzima com milhares de átomos, interagindo com um ligante *em solução* e com contra-íons a sua volta. Trata-se justamente do sistema em que estamos interessados em simular. Para calcular esta energia livre de interação em solução, precisamos resolver o ciclo termodinâmico da Figura 26.

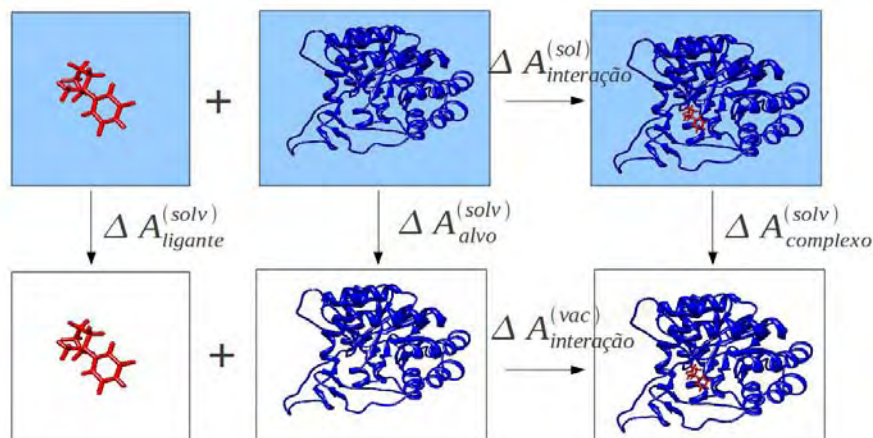


Figura 26: Ciclo termodinâmico para calcular a variação da energia livre para uma interação ligante-macromolécula em solução aquosa

De forma que $\Delta A_{\text{interação}}^{(\text{sol})} - \Delta A_{\text{interação}}^{(\text{vac})} - \Delta A_{\text{complexo}}^{(\text{solv})} + \Delta A_{\text{alvo}}^{(\text{solv})} + \Delta A_{\text{ligante}}^{(\text{solv})} = 0$. Isto é, aquelas contribuições contribuindo *contra* o sentido do ciclo são negativas e aquelas a favor são positivas de modo que o caminho indo somado ao caminho voltando seja zero. Com isso, chegamos à seguinte equação:

$$\Delta A_{\text{interação}}^{(\text{sol})} = \Delta A_{\text{interação}}^{(\text{vac})} + \Delta A_{\text{complexo}}^{(\text{solv})} - (\Delta A_{\text{alvo}}^{(\text{solv})} + \Delta A_{\text{ligante}}^{(\text{solv})}) \quad (89)$$

O cálculo deste ciclo termodinâmico para estimar confiavelmente a verdadeira variação da energia livre em um ensemble canônico é muito custoso. Por este motivo, métodos alternativos de avaliar a afinidade entre duas moléculas, particularmente entre ligantes e macromoléculas, apareceram. Entre eles, o *docking* molecular citado anteriormente.

1.2.2.10.3. Método MMGBSA

Acuradas e exatas como sejam, no entanto, estas técnicas canônicas de cálculo de energia livre não são rápidas de se realizar nem tampouco simples de se preparar. Muita fineza e intuição química levam o processo adiante com a escolha dos intermediários de acoplamento e análises para saber qual técnica é válida e em que ocasião.

Se o que se deseja são resultados que não sejam muito caros computacionalmente, técnicas deste tipo não são muito viáveis para uma triagem virtual. Na verdade, a análise finamente detida e reveladora por trás destes métodos é o oposto inverso da análise rápida, geral e de grande escopo que se deseja das técnicas de triagem virtual.

Por outro lado, os métodos de *docking* molecular, criados para preencher este nicho na técnica de triagem virtual desviaram-se de tal modo do significado físico das medidas de afinidade que procuram estimar que perderam, no processo de tornarem-se rápidos, exatidão.

E nestes últimos dez anos, recuperar esta exatidão sem perder sua estimada rapidez tem sido o grande desafio das implementações de *docking* molecular. Um desafio, diga-se de passagem, que vem avançando apenas lentamente, um nanojoule de cada vez.

Neste meio termo, técnicas aproximadas para se resolver a equação (89) têm aparecido na tentativa de tornar criar métodos de estimativa de energia livre ao mesmo tempo simples de reproduzir efetivamente mas exatos e rápidos o suficiente para serem utilizados com bibliotecas contendo dezenas de moléculas.

Um desses métodos de estimativa mais populares é o chamado MMGBSA (*Molecular Dynamics, Generalized Born and Solvent Accessibility*), proposto por Srinivasan et al (1998), para estudar a estabilidade termodinâmica de hélices de ácidos nucleicos com diferentes composições e tamanhos. Desde então, este método vem sendo utilizado com boa exatidão na estimativas da energia de interação entre pequenos ligantes e alvos macromoleculares.

No método MMGBSA, cada um dos termos da equação (89) é calculado de forma aproximada utilizando-se microestados retirados de uma única trajetória, em contraste com os métodos convencionais, que precisam sondar múltiplas trajetórias.

Normalmente, precisaríamos de *pelo menos* quatro trajetórias de dinâmica molecular para conseguir calcular cada uma das variações de energia da qual depende a quantidade que se deseja estimar. Para evitar fazer isto, o modelo faz amplo uso das chamadas *técnicas de solvente implícito* para encurtar a avaliação das contribuições do solvente pelo ciclo.

A técnica do solvente implícito basicamente modela o sistema como se o soluto estivesse posicionado dentro de uma cavidade aberta em um meio contínuo. Este meio contínuo é o chamado solvente implícito e é descrito como um contínuo polarizável que age sobre as cargas do soluto e é, em retorno, polarizado por elas.

Este comportamento complicado é descrito resolvendo-se a equação diferencial de Poisson-Boltzmann para o potencial ϕ .

$$\nabla[\epsilon \nabla \phi] - \epsilon \kappa^2 \sinh \phi + 4 \pi \rho = 0 \quad (90)$$

onde ϵ é a constante dielétrica do contínuo, ρ é a densidade de carga do sistema e κ é um parâmetro de Debye-Hückel e contabiliza a concentração de carga iônica em volta de uma partícula e seu cálculo depende da concentração iônica C .

$$\kappa^2 = \left(\frac{2000 \cdot F^2}{\epsilon_0 \epsilon_r R T} \right) \cdot C \quad (91)$$

Onde F é a constante de Faraday, ϵ_0 é a permissividade elétrica do vácuo, ϵ_r é a

constante dielétrica da solução em torno da partícula de posição r . R é a constante dos gases ideais e T a temperatura termodinâmica.

Para os casos de uma estrutura em um mínimo (ou próximo de um) de energia, onde o potencial individual sobre cada carga não é muito grande, e com o modelo de cargas pontuais, o termo do seno hiperbólico pode ser linearizado (Figura 27)

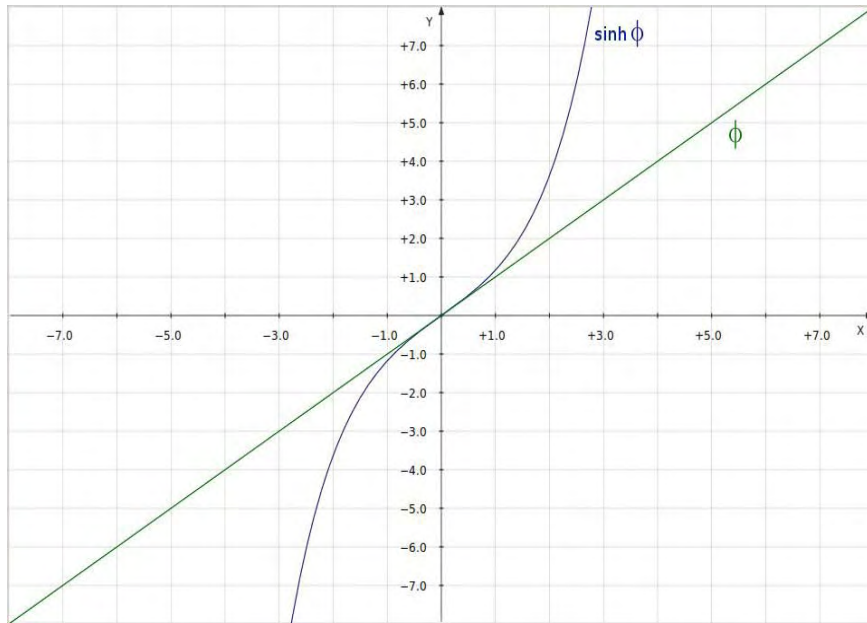


Figura 27: Linearização da função $\sinh$.

De forma que a equação de Poisson-Boltzmann torna-se:

$$\nabla[\epsilon \nabla \phi] - \epsilon \kappa^2 \phi + 4\pi\rho = 0 \quad (92)$$

Quando o potencial é encontrado resolvendo-se a equação (92), a energia A_ϵ para um conjunto de n cargas pontuais q_i em posições r_i interagindo com um contínuo de solvente de constante dielétrica ϵ é:

$$A_\epsilon = \sum_i^n q_i \phi(r_i) \quad (93)$$

Neste caso, podemos calcular a variação de energia provocada pela mudança entre dois solventes como sendo:

$$\Delta A_\epsilon = \sum_i^n q_i [\phi_2(r_i) - \phi_1(r_i)] \quad (94)$$

onde ϕ_1 é o potencial calculado para a dielétrica ϵ_1 e ϕ_2 a dielétrica calculada para ϵ_2 . Se um dos meios for o vácuo ($\epsilon=1$), a variação de energia ΔA_ϵ representa uma variação na energia devido às interações eletrostáticas entre soluto e solvente.

A resolução das equações (91) e (92), apesar de apresentar bons resultados, não podem ser resolvidas exatamente para sistemas de geometria complexa. Em função disto, sua resolução deve ser numérica e não é muito rápida de ser realizada.

Por este motivo, modelos de solvente implícito raramente (PRABHU; ZHU; SHARP, 2004; WILLIAMS; HALL, 1999) são utilizados em simulações computacionais, onde o andamento da simulação significaria um cálculo desses a cada passo de integração.

Para simplificar a solução, um modelo foi proposto baseado nos estudos do físico e matemático alemão Max Born sobre a solvatação de íons.

O raciocínio é mais ou menos o seguinte. Imagine um íon como uma cavidade esférica de raio a com nada em seu interior exceto uma carga pontual q . Tudo isto envolto por um solvente de constante dielétrica ϵ (Figura 28)

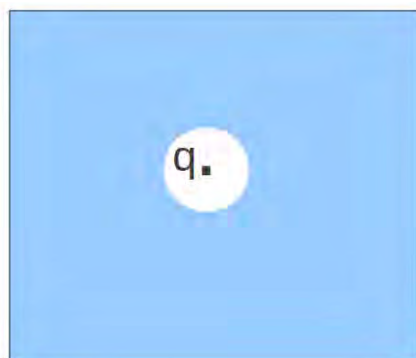


Figura 28: Esquema ilustrativo de uma carga pontual q dentro de uma cavidade em meio a uma dielétrica contínua

Por causa da simetria esférica do potencial eletrostático, resolver a equação de Poisson-Boltzmann para este sistema é um problema de uma única coordenada, a distância r a partir da carga pontual q . Quando resolvemos a equação, obtemos duas expressões analíticas para o potencial:

$$\phi_{r \geq a} = \frac{q}{4\pi\epsilon r} \quad (95-1) \quad \text{e} \quad \phi_{r < a} = \frac{q}{4\pi\epsilon r} - \frac{q}{4\pi a} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \quad (95-2)$$

Perceba que o potencial dentro da esfera é uma soma de dois termos. O primeiro é o potencial eletrostático de Coulomb convencional que existiria caso não houvesse a fronteira dielétrica em $r=a$.

O segundo termo se deve à interação da carga pontual com a dielétrica contínua fora da esfera e representa a energia potencial de interação entre a distribuição de carga no interior da esfera e no

exterior. Como já mencionamos, esta é a parcela da energia de solvatação que se deve unicamente às interações eletrostáticas.

Se chamarmos o potencial (95-1) de ϕ_0 e o potencial (95-2) de ϕ , podemos dizer que:

$$\phi = \phi_0 - \frac{q}{4\pi a} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \quad (96)$$

De forma que $\Delta\phi = \phi - \phi_0 = -\frac{q}{4\pi a} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right)$. Esta diferença entre os dois potenciais é chamado *campo de reação*, denotado por ϕ_{reac} .

Em princípio, segundo a equação (94), a energia de interação da carga pontual com o campo reacional seria $A_\epsilon = q\phi_{\text{reac}}$. Mas a carga q interagindo com este potencial também é responsável por gerá-lo de modo que contribuições serão contadas duas vezes. De modo que:

$$A_\epsilon = \frac{1}{2} q \phi_{\text{reac}} = -\frac{q^2}{8\pi a} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \quad (97)$$

A equação (97) é a *Expressão de Born* da solvatação eletrostática polarizável de íons em solução. Seguindo o mesmo princípio, nós imaginamos várias partículas carregadas na mesma dielétrica.

Generalizando a expressão de Born, tomamos cada um dos íons, calculando a energia para cada par de cargas utilizando a equação (97) podemos estimar a energia das interações entre uma molécula de soluto de geometria arbitrária e uma dielétrica contínua que a envolve.

$$A_\epsilon = \frac{1}{8\pi} \left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon} \right) \sum_i^N \sum_{j \neq i}^{N-1} \frac{q_i q_j}{f_{\text{GB}}(r_{ij}, a_i, a_j)} \quad (98)$$

Este é o tal do método generalizado de Born (*Generalized Born*) ou GB (STILL et al, 1990). A função f é uma função suave que generaliza o conceito do raio de Born (o raio a da resolução para um único íon). E sua forma mais utilizada é:

$$f(r_{ij}, a_i, a_j) = \left[r_{ij} + a_i a_j \exp\left(-\frac{r_{ij}^2}{4a_i a_j}\right) \right]^{\frac{1}{2}} \quad (98')$$

Os raios de Born a_i são parâmetros empíricos, numericamente otimizados. A forma exata da função e dos raios de Born determinam qual dos inúmeros métodos de solvente implícito tipo GB se está utilizando.

Utilizar a aproximação de Born acelera incrivelmente a velocidade de cálculo das energias eletrostáticas de solvatação, entretanto, esta não é a única contribuição de energia decorrente da solvatação de compostos.

A adição de solutos em um solvente perturba a rede dinâmica de interações intermoleculares

que se tece entre todas as moléculas do solvente. A presença de uma molécula de soluto abre um lugar entre as moléculas de solvente, separando-as e rompendo interações intermoleculares.

As moléculas imediatamente em volta do soluto não podem realizar interações através da molécula de soluto, isto restringe a direção em que as interações podem ser feitas, quando comparado em relação às moléculas de um solvente puro, diminuindo a entropia do sistema.

Esta contribuição é sempre positiva para a energia interna do sistema solvatado e pode ser parcialmente compensada quando a estrutura do soluto também é capaz de compensar pelas interações rompidas formando ela mesma interações equivalentes com as moléculas de soluto circundante.

Portanto, a variação na energia de solvatação se constitui de uma parcela que contabiliza pelas interações polares entre o solvente e o soluto, calculada segundo a equação (98).

$$\Delta A^{(solv)} = \Delta A_{\epsilon} + A_{np} \quad (99)$$

Onde o termo A_{np} se refere à interferência entrópica da adição de soluto. Empiricamente, é sabido que a perturbação entrópica do solvente é proporcional à sua área superficial. Por este motivo, o termo “não-polar” é estimado a partir da área acessível ao solvente sobre cada um dos n átomos da molécula de soluto a partir da equação (100).

$$A_{np} = \sum_i^n \gamma_i \sigma_i \quad (100)$$

σ_i é a área superficial acessível ao solvente sobre o átomo i e o parâmetro gama é ajustado empiricamente para os vários tipos de átomos.

Utilizando este formalismo, podemos calcular a variação na energia livre decorrente da solvatação de um sistema em relação à fase gasosa como sendo:

$$\Delta A^{(solv)} = A_{\epsilon} - A_{\epsilon_0} + A_{np} \quad (101)$$

Em um sistema sem a presença do solvente, o cálculo da interação entre duas únicas moléculas torna-se muito mais simples e direto. Podemos até mesmo utilizar a expressão básica da energia livre de Helmholtz como ponto de partida, para um sistema canônico.

$$\Delta A = \Delta E - T \Delta S \quad (102)$$

Para um sistema em fase gasosa, a energia termodinâmica E depende apenas do potencial interno das moléculas. Isto quer dizer que uma boa estimativa de ΔE para uma determinada geometria seria, por exemplo, a diferença entre as energias calculadas quanticamente, através de métodos de função de onda.

Mas para um sistema grande como uma proteína, isso poderia levar dias ou até semanas. A alternativa é utilizar as energias calculadas pelos potenciais empíricos de mecânica molecular.

Como já estamos utilizando uma forma aproximada de se calcular a energia, o faremos da melhor forma possível, contabilizando todas as interações sem usar raios de corte para pegar com a maior exatidão possível a energia do sistema.

O termo da variação da entropia diz respeito à perda de graus de liberdade decorrentes da formação de uma interação. Onde antes haviam duas moléculas que podiam acessar quaisquer graus de liberdade rotacionais, translacionais ou vibracionais, agora existe um complexo onde a movimentação das moléculas é mais limitado.

Esta estimativa pode ser obtida a partir de cálculos de modos normais para os dois sistemas. Assim, para um complexo macromolecular com um alvo e um ligante, a energia de interação do complexo pode ser estimada como sendo:

$$\begin{aligned}\Delta A_{interação}^{(vac)} &= (E_{complexo}^{MM} - E_{alvo}^{MM}) - (E_{complexo}^{MM} - E_{ligante}^{MM}) + T \Delta S_{NORMODS} \\ \Delta A_{interação}^{(vac)} &= E_{complexo}^{MM} - E_{alvo}^{MM} - E_{complexo}^{MM} + E_{ligante}^{MM} + T \Delta S_{NORMODS} \\ \Delta A_{interação}^{(vac)} &= -E_{alvo}^{MM} + E_{ligante}^{MM} + T \Delta S_{NORMODS}\end{aligned}$$

Ou seja:

$$\Delta A_{interação}^{(vac)} = E_{ligante}^{MM} - E_{alvo}^{MM} + T \Delta S_{NORMODS} \quad (103)$$

Unindo a avaliação da equação (100) com potenciais de Mecânica Molecular [resultando na equação (102)] para calcular a variação da energia livre procedente da interação das moléculas, o método *Generalized Born* para calcular a variação da energia livre procedente das interações polares entre solutos e solventes, e o método da acessibilidade do solvente (do inglês, *Solvent Accessibility*) para estimar as contribuições entrópicas da solvatação, temos um método para avaliar cada um dos termos da equação (89). Este é o método MMGBSA.

Agora o que fazemos é substituir as equações (103) e (101) na equação (89) e temos que:

$$\begin{aligned}\Delta A_{interação}^{(sol)} &= \Delta A_{interação}^{(vac)} + \Delta A_{complexo}^{(sol)} - (\Delta A_{alvo}^{(sol)} + \Delta A_{ligante}^{(sol)}) \\ \Delta A_{complexo}^{(sol)} &= A_{(\epsilon=80)}^{(complexo)} - A_{\epsilon_0}^{(complexo)} - A_{np}^{(complexo)} \\ \Delta A_{alvo}^{(sol)} &= A_{(\epsilon=80)}^{(alvo)} - A_{\epsilon_0}^{(alvo)} - A_{np}^{(alvo)} \\ \Delta A_{ligante}^{(sol)} &= A_{(\epsilon=80)}^{(ligante)} - A_{\epsilon_0}^{(ligante)} - A_{np}^{(ligante)} \\ \Delta A_{interação}^{(vac)} &= E_{ligante}^{MM} - E_{alvo}^{MM} + T \Delta S_{NORMODS}\end{aligned} \quad (104)$$

E aqui temos a forma básica do método MMGBSA. O que essas equações nos dizem é que para calcular a energia de interação não precisamos fazer várias simulações de perturbações ou integrar sobre várias trajetórias de intermediários de acoplamento. Basta que se calcule as energias utilizando campos de força empíricos, área superficial acessível ao solvente, modelos de solvente implícito e uma forma de calcular os modos normais dos sistemas.

Isto pode ser feito para cada passo de trajetórias, uma do ligante, uma do alvo e uma do

complexo, sem precisar simular solvatações com perturbações ou métodos mais sofisticados. Tomando-se o valor esperado de cada propriedade sobre as trajetórias em questão, obtêm-se uma estimativa da variação da energia livre que ocorre no sistema quando estas duas moléculas interagem em solução.

$$\begin{aligned}
\Delta A_{interação}^{(sol)} &= \Delta A_{interação}^{(vac)} + \Delta A_{complexo}^{(sol)} - (\Delta A_{alvo}^{(sol)} + \Delta A_{ligante}^{(sol)}) \\
\Delta A_{complexo}^{(sol)} &= \langle A_{(\epsilon=80)}^{(complexo)} - A_{\epsilon_0}^{(complexo)} - A_{np}^{(complexo)} \rangle \\
\Delta A_{alvo}^{(sol)} &= \langle A_{(\epsilon=80)}^{(alvo)} - A_{\epsilon_0}^{(alvo)} - A_{np}^{(alvo)} \rangle \\
\Delta A_{ligante}^{(sol)} &= \langle A_{(\epsilon=80)}^{(ligante)} - A_{\epsilon_0}^{(ligante)} - A_{np}^{(ligante)} \rangle \\
\Delta A_{interação}^{(vac)} &= \langle E_{ligante}^{MM} \rangle - \langle E_{alvo}^{MM} \rangle + T \langle \Delta S_{NORMODS} \rangle
\end{aligned} \tag{105}$$

Esta variação pode então ser utilizada para medir constantes de equilíbrio e afinidade (como já discutido anteriormente). Apesar disso, o método ainda depende de mais de uma trajetória, mas ao contrário dos outros métodos apresentados ele não *precisa*, por construção, depender delas.

Na verdade, se pararmos para pensar, a única coisa da qual o método realmente precisa é uma coleção de estruturas amostradas de um ensemble canônico sobre a qual retirar as médias em **(105)**. A forma tradicional de fazer isto é, evidentemente, a partir de trajetórias de dinâmicas moleculares no ensemble NVT.

Mas, em princípio, estas estruturas poderiam ser retiradas da mesma trajetória. Poderíamos, por exemplo, produzir uma única trajetória do complexo solvatado e de alguns passos retirar estruturas para o ligante e o alvo (Figura 29), e realizar as estimativas de energia de solvatação e energia de interação no vácuo sobre estas estruturas “recortadas”.

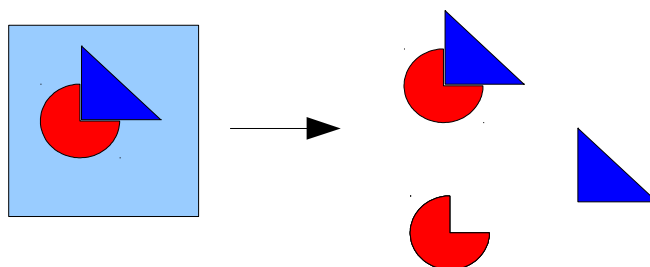


Figura 29: Utilização de um microestado do complexo solvatado para conseguir as estruturas no vácuo.

O maior cuidado que se deve tomar ao se calcular a variação da energia livre desta forma é escolher passos na dinâmica que estejam o mais *descorrelacionados* possível. Isto é, queremos tomar passos que não tenham parâmetros muito parecidos. Se os resultados se parecerem demais, o fato de que estamos utilizando uma única trajetória torna-se aparente na nossa estimativa, por isto fingimos uma amostragem “independente” (que seria com trajetórias separadas) utilizando passos descorrelacionado.

Na verdade, o conceito de correlação em dinâmica molecular é mais profundo que simples

similaridade e tem a ver com o caos. Para explicar melhor isto, vamos pensar em um exemplo bem simples mas muito esclarecedor. Imagine uma partícula descrevendo uma trajetória no espaço (Figura 30).

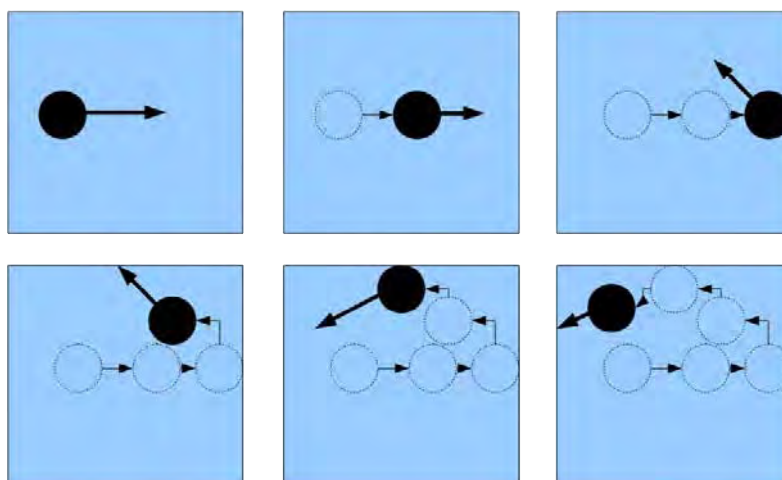


Figura 30: Esquema mostrando uma trajetórias clássica.

Agora imagine esta mesma partícula, exceto que do primeiro para o segundo passo, o cálculo da velocidade sofre um pequeno desvio imprevisível e aleatório. Vamos dizer que este desvio é um *evento caótico*. Agora dê uma olhada na Figura 31 e compare-a com Figura 30.

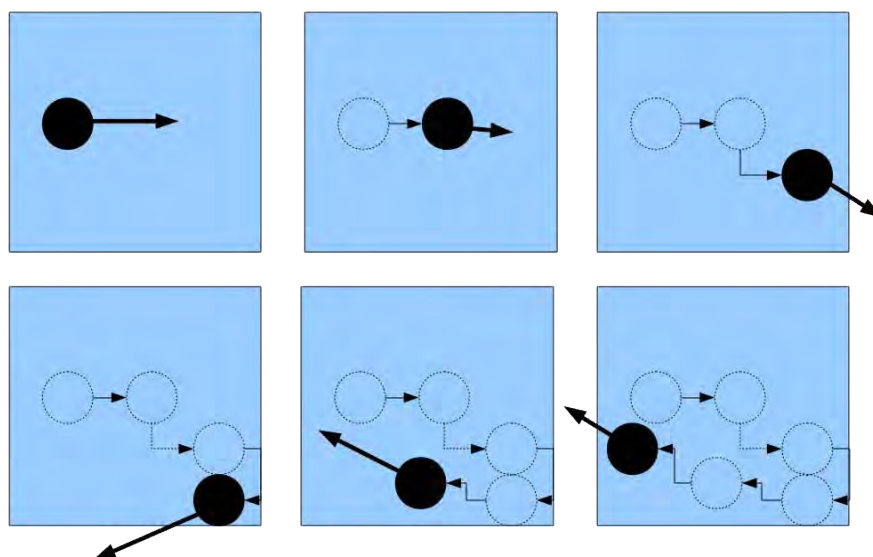


Figura 31: Mesma partícula da figura 30, com o primeiro passo da trajetória levemente modificado por um evento caótico

Apesar de ser apenas um exemplo ilustrativo e não um resultado formal de um cálculo, é fácil entender que, mesmo que seja possível obter de forma determinística novas posições a partir de posições antigas, basta um pequeno desvio neste procedimento numérico para que os resultados de dois cálculos independentes sejam bastante diferentes. Esta é uma das manifestações do chamado

caos determinístico.

Mas cálculos numéricos contêm erros. É impossível realizar um cálculo com precisão total, por exemplo. Você não pode ter infinitas casas decimais para um valor de posição ou momento linear e eventualmente, arredondamentos e erros da máquina ocorrerão. Estes erros são imprevisíveis e aleatórios por natureza. Normalmente, isto não é problema, mas quando cálculos são realizados em sistemas caóticos (como dinâmicas moleculares), isto faz toda a diferença.

Em geral, um sistema caótico é extraordinariamente sensível a variações em seus parâmetros. Para um sistema deste tipo, mesmo variações ridiculamente pequenas causam perturbações que se propagam por todo o sistema e culminam em resultados caóticos.

Em virtude disto, as trajetórias de dinâmica molecular não são prontamente reproduzíveis. Isto é, dado um dos passos de uma dinâmica é possível calcular deterministicamente novas velocidades, forças e posições, mas não se pode ter garantia absoluta que estes valores serão os mesmos para vários cálculos independentes.

Esta “garantia de reprodutibilidade” é o que se chama de *correlação*. Em uma trajetória, dois passos estão fortemente correlacionados se houver uma alta possibilidade de chegarmos à geometria exata do passo mais recente propagando-se o passo mais antigo.

Esta correlação diminui à medida que os passos tornam-se cada vez mais afastados no tempo e as divergências caóticas acumulam-se. Até que fica impossível prever o passo mais recente a partir do mais antigo. Fica impossível dizer se o passo mais antigo é *causa* do passo mais recente. A causalidade das transições é o que define uma trajetória.

Sabemos que uma trajetória de dinâmica molecular é uma representação de um ensemble. Se eu tomar um passo a cada vinte, este subconjunto ainda será uma representação do ensemble, assim como qualquer equipartição desta trajetória, na verdade, também o será (a partir de um certo tamanho, pois à medida em que o conjunto torna-se menor, a descrição do ensemble que ele é capaz de fornecer deteriora-se).

Tomando-se vários pontos descorrelacionados da trajetória, assumindo-se que o subconjunto de pontos seja amplo o suficiente, obtemos um conjunto de microestados que foram amostradas de um ensemble mas que praticamente não possuem relação de causalidade entre si. Ou seja, não constituem uma dinâmica molecular propriamente dita, mas são uma amostragem do ensemble.

Se tomarmos estes passos descorrelacionados de um complexo ligante-alvo molecular e retirarmos deles também as trajetórias do ligante e do alvo, teremos três amostragens do ensemble canônico diferentes. Pois não há causalidade suficiente entre as estruturas para que seja possível traçar a origem de cada uma delas como sendo parte de uma única dinâmica.

É sobre este conjunto de microestados descorrelacionados que as médias das equações (103) são tomadas. De forma que agora justifica-se utilizar uma única dinâmica para se amostrar os microestados tanto do complexo quanto do ligante e do alvo.

Utilizar mais de uma trajetória de onde tomar os microestados descorrelacionados aprimora a estabilidade estatística das estimativas e diminui um pouco o desvio (GENHEDEN; RYDE, 2010) das propriedades. É uma forma mais robusta de se realizar o cálculo mas retorna à questão de se utilizar várias trajetórias.

Utilizado desta forma, o MMGBSA pode ser feito com trajetórias mais curtas, uma vez que a variabilidade é conseguida em virtude do caos inerente aos cálculos de dinâmica, de modo que cada trajetória provavelmente amostrará melhor partes diferentes do ensemble.

1.2.2.10.3.1. MMGBSA e Estruturas de Mínimo

Por mais diferentes que sejam todas estas trajetórias, entretanto, uma vez que os cálculos são feitos a partir de uma estrutura minimizada, é razoável supor que a maior contribuição para as médias em (105) se devem a microestados de mínimo ou próximos do mínimo.

Considerando este fato, Rastelli et al (2010) sugeriram que seria possível conseguir uma boa estimativa da energia livre realizando um cálculo de MMGBSA em um único microestado bem minimizado .

A estimativa da energia feita desta forma fornece um valor com mais significado físico que o *docking* molecular, mas a metodologia é rápida o suficiente para ser utilizada em uma possível triagem virtual de ligantes, embora talvez não para a mesma quantidade de ligantes usualmente realizado com o *docking*. Comparações entre o MMGBSA convencional e MMGBSA aplicado a uma única trajetória têm sido feitos, mas utilizando métodos diferentes de cálculos de energia de polarização (o modelo de Poisson-Boltzmann) (KUHN et al, 2004).

A estimativa acurada da variação de energia livre é uma área sutil e desafiadora de se compreender plenamente. Para maiores informações sobre termodinâmica estatística, simulações computacionais e cálculos de energia livre, é encorajada a leitura cuidadosa das referências (FIELD, 2007; CHIPHOT; PHORILLE, 2007; MCQUARIE, 1976; ERCOLESSI, 1997; VESELY, 2005; ABRAMS, 2009).

2. Objetivos

2.1. Geral

- Comparar o desempenho do método min-MMGBSA em relação ao método de *docking* molecular na triagem virtual de pequenas coleções de ligantes para uma enzima de *Leishmania chagasi* e duas de *Leishmania amazonensis* para as quais não existem estudos teóricos.

2.2. Específicos

- Selecionar enzimas de *Leishmania sp.* que sejam alvos de poucos estudos, sobretudo teóricos, mas que sejam de interesse medicinal;
- Obter as estruturas tridimensionais para estas enzimas a partir de técnicas de modelagem por homologia;
- Obter, de forma sistemática e reprodutível, coleções ligantes para cada enzimas;
- Avaliar as vantagens e desvantagens de se utilizar o método min-MMGBSA em detrimento do *docking* molecular na triagem virtual destes pequenos conjuntos de ligantes;

3. Metodologia

3.1 Busca e Seleção de Alvos Moleculares

Existem vários alvos moleculares possíveis de serem explorados por estratégias terapêuticas. Cada um deles responsável por uma parte da fisiologia do parasita. Dentre estes, entretanto, os mais tratáveis por simulações computacionais e aqueles cujos ambientes químicos em solução são melhor simulados através das técnicas mostradas aqui, são as enzimas.

Seres vivos são sistemas bastante sensíveis que dependem do equilíbrio da concentração de várias substâncias para manterem sua homeostase. A biossíntese e a degradação destas substâncias precisa ser finamente balanceada e reguladas em escalas de tempo muito pequenas.

Para realizar este controle tão sutil, seres vivos empregam enzimas. As enzimas são, em sua maioria, proteínas que funcionam como catalisadores bastante eficientes e específicos. As moléculas sobre as quais as enzimas agem são chamadas substratos da enzima e as moléculas resultantes da catálise são os produtos.

Enzimas formam, nas células vivas, várias redes interconectadas de forma que muitas delas têm como substratos os produtos das reações catalisadas por outras. Esta malha metabólica é a ferramenta que os seres vivos utilizam para aproveitar matéria e energia do ambiente e alimentarem o próprio estado vital.

Pela própria natureza do metabolismo, perturbar um passo deste fluxo de matéria e energia significa perturbar toda esta teia metabólica de enzimas interconectadas. As malhas metabólicas têm, entretanto, uma plasticidade imensa, ativando e encontrando rotas alternativas quando alguma das enzimas está comprometida.

Em alguns casos, entretanto, a função de uma enzima importante a ponto de seu comprometimento produzir danos irreversíveis e impossíveis de serem compensados por outras formas. Estes pontos importantes da malha metabólica são particularmente interessantes no desenvolvimento de antiparasitários e/ou antibióticos.

Quando lidamos com micro-organismos, utilizamos nosso conhecimento do metabolismo para encontrar os pontos fracos no metabolismo destes seres e inibir suas enzimas mais importantes de modo a levar as células infectantes à morte.

Uma dificuldade deste procedimento se encontra no fato de que todos os seres vivos estão unidos entre si por algum ancestral comum, por mais remoto que seja. Em virtude deste fato, é provável que dois organismos, até mesmo muito diferentes entre si, possuam enzimas de grande similaridade de forma que inibidores destas enzimas podem atuar em ambas as espécies.

Quando as duas espécies comparadas são os seres humanos e um parasita, a preocupação é encontrar alvos em pontos sensíveis da biologia do parasita que não se intersectem com os do hospedeiro.

Fora a *Leishmania chagasi*, que é causadora de leishmaniose visceral, a outra espécie mais presente no país, e responsável pela leishmaniose cutânea e mucocutânea é a *Leishmania amazonensis*. Como a *Leishmania amazonensis* é pouco estudada, em virtude da baixa letalidade de sua contaminação, decidimos nos aprofundar mais nesta espécie. Decidimos, como já comentado antes, focar nossos estudos nestas duas espécies de parasitas causadores das leishmanioses.

Para utilizarmos dinâmica molecular para estudar enzimas de interesse destas espécies, precisamos saber quais são os alvos e ter acesso a suas geometrias. Precisamos saber o tipo e a posição dos átomos e esta não é uma tarefa direta de ser realizada para enzimas.

A geometria das enzimas (proteínas, em geral) não pode ser criada totalmente em computador. Aliás, esse é um dos grandes problemas da biologia molecular e diz respeito à previsão da estrutura tridimensional de uma proteína. Por este motivo, foi preciso obter dados sobre a estrutura das proteínas a partir de dados experimentais.

Informações deste tipo estão disponíveis em bancos de dados cristalográficos contendo estruturas químicas resolvidas por uma mistura de técnicas experimentais e computacionais. O mais conhecido banco de dados de estruturas de proteínas do mundo é o já citado *Protein Data Bank*, o PDB (BERNSTEIN et al, 1977).

Ocorre que não existem estruturas de *Leishmania chagasi* nem de *Leishmania amazonensis* no PDB. Decidimos então partir de enzimas cristalizadas para outras espécies de *Leishmania sp.* E obter estruturas de partida para alguns alvos de interesse, que posteriormente pudessem ser utilizadas para se conseguir as estruturas dos alvos homólogos nas espécies de interesse, através de métodos de modelagem por homologia.

Para o método da modelagem por homologia funcionar, é necessário que haja alguma similaridade entre as sequências dos moldes (enzimas que têm estruturas resolvidas) e dos modelos (sequências para as quais queremos calcular uma estrutura tridimensional, a partir do molde). Portanto, as estruturas de partida que temos que obter do PDB precisam ser enzimas de interesse medicinal e com alguma similaridade, em sua sequência primária, com seus análogos em *Leishmania chagasi* e *Leishmania amazonensis*.

Para chegar a enzimas que atendam a estas características, partimos do conjunto de todas enzimas do PDB relacionadas às espécies de *Leishmania sp.* de maior participação no banco de dados do PDB, são elas: *Leishmania major* e *Leishmania infantum*. Destas estruturas selecionamos um subconjunto daquelas que julgamos serem alvos medicinais interessantes e que não tivessem trabalhos em química teórica publicados sobre elas, na ocasião da seleção.

Fizemos isto baseado nos artigos de inclusão das estruturas no banco de dados. Este subconjunto se compôs de doze enzimas (Tabela 1).

Tabela 1: Enzimas selecionadas do PDB por serem alvos moleculares de interesse medicinal e pela carência de estudos teóricos sobre elas. Inclui % de Identidade e os E-values dos alinhamentos das enzimas tabeladas com seus homólogos nas espécies de interesse.

Código PDB	Nome	Via Metabólica/Função	<i>L. chagasi</i>	<i>L. amazonensis</i>	E-Value
1EZR	Nucleosídeo Hidrolase	Hidrólise de Nucleosídeos	94,00%	-	1×10^{-160}
3GZ3	Diidroorotato Desidrogenase	Biossíntese de Pirimidinas	-	93,00%	1×10^{-167}
3OGZ	UDPG Pirofosforilase	Biossíntese de Polissacarídeos	-	-	-
3LJN	Anquirina	Acoplamento de Proteínas Integrais de Membrana com o	-	-	-

		citoesqueleto			
2VOB	Tripanotiona Sintase	Controle do Stress Oxidativo	-	97,00%	0,0
2C21	Glioxalase I	Controle do Stress Oxidativo	-	-	-
2JK6	Tripanotiona Redutase	Controle do Stress Oxidativo	-	95,00%	0,0
2HFS	Mevalonato Cinase	Biossíntese de Esteróis	-	-	-
1YHL*	Farnesil Pirofosfato Sintase	Biossíntese de Esteróis	60,00%	-	1×10^{-128}
3L4D	Esterol 14- α Demetilase	Biossíntese de Esteróis	-	97,00%	0,0
2HQJ	Ciclofilina	Enovelamento de Proteínas	-	-	-
1LML	Leishmanolisina	Protease de Membrana	85,00%	-	0,0

A **1YHL**, marcada em detalhe não foi obtida a partir das enzimas de *Leishmania infantum* nem de *Leishmania major*. Esta enzima é de *Trypanosoma cruzi*. Sua inclusão no banco-de-dados de enzimas é motivada por outros dados.

Em 2010, nosso grupo de química medicinal publicou um estudo detalhando dados microbiológicos e de docking molecular sobre o composto 3-Hydroxy-2-methylene-3-(4-nitrophenyl)propanenitrile, um aduto de Morita-Bayllis-Hillman, sugerindo este composto como um possível inibidor desta enzima (SANDER et al, 2010).

Pouco tempo depois, ainda neste mesmo ano, nosso grupo lançou um outro artigo com dados experimentais documentando propriedades antiparasitária de uma série de Adutos de Morita-Bayllis-Hillman contra parasitas das espécies *Leishmania chagasi* e *Leishmania amazonensis* (JUNIOR et al, 2010). Estes dados inspiraram a inclusão da **1YHL** a este subconjunto de alvos de interesse.

Em seguida, para cada uma destas doze enzimas, buscamos no banco de dados de sequências do UniProt (THE UNIPROT CONSORTIUM, 2011), usando o algoritmo de busca BLAST (ATLSCHUL et al, 1990), por sequências semelhantes a estas que pertencessem a *Leishmania chagasi* e *Leishmania amazonensis*. Como as enzimas com mais de uma subunidade tinham todas elas iguais, o BLAST foi feito apenas com uma das subunidades. Estes dados de similaridade também estão sumarizados na Tabela 1.

Das doze, sobraram a **1EZR**, **3GZ3**, **2VOB**, **2JK6**, **1YHL**, **3L4D** e **1LML**. **2VOB** e **2JK6** estão na mesma via metabólica, assim como a **1YHL** e a **3L4D**. Para que as enzimas estudadas envolvessem a gama mais abrangente possível de vias metabólicas, decidimos que seria prudente escolher apenas uma dessas.

As enzimas restantes foram então novamente analisadas pela disponibilidade de ligantes no sítio ativo (para podermos encontrar o sítio ativo) e pela presença de íons de difícil tratamento pelos campos de força dos pacotes de dinâmica molecular e aminoácidos não-padronizados.

A **1EZR**, **3L4D**, e a **1LML** foram desclassificadas na seleção por conterem íons para os quais não há parâmetros de qualidade nos campos de força usados pelos programas utilizados: Ca^{2+} , Fe^{2+} e Zn^{2+} respectivamente. A **2VOB** foi descartada porque não foi cristalizada com nenhum ligante que possa apontar o sítio ativo, nem possui qualquer estrutura homóloga que tenha sido cristalizada com um ligante e que pudesse ser utilizada para realizar esta inferência.

Restaram **3GZ3**, **2JK6** e **1YHL**. Que naturalmente foram as escolhidas para o trabalho. Na Figura 32 podemos ver que as enzimas estão bastante distantes umas das outras, no mapa metabólico, estando em vias bem desconectadas, o que nos dá segurança de estar selecionando alvos o mais variados possível. Além disso, nenhum destes alvos, durante o BLAST de suas sequências mostrou ter similaridade significativa com qualquer enzima humana.

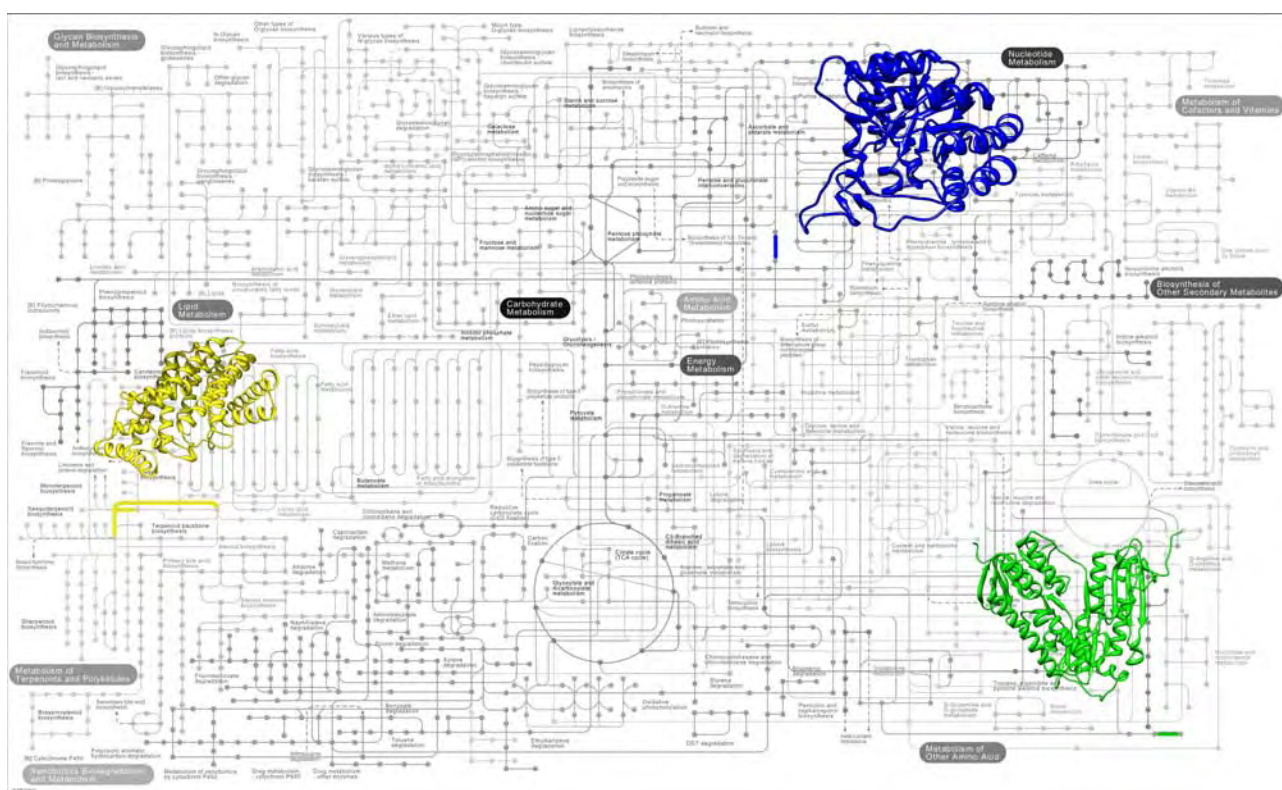


Figura 32: Mapa metabólico demonstrando onde estão as enzimas selecionadas. Em azul, a DHDH, em verde a TR e em amarelo a FPPS. Adaptado a partir de <http://www.kegg.jp/kegg/atlas/?01100>

A Farnesil Pirofosfato Sintetase (FPPS, C.E. **2.5.1.10**) é uma enzima da via do ácido mevalônico envolvida na biossíntese de Farnesil Pirofosfato, um importante precursor de compostos esteroidais (Figura 33). Sua importância metabólica encontra-se no fato de que parasitas tripanossomatídeos dependem de hormônios esteroidais endógenos e nos processos de modificações pós-traducionais em proteínas, especialmente prenilação (PEREZ-SALA, 2007). Inibidores da biossíntese de esteróides já foram demonstrados como sendo fármacos potentes (URBINA, 1997; URBINA et al, 2002).

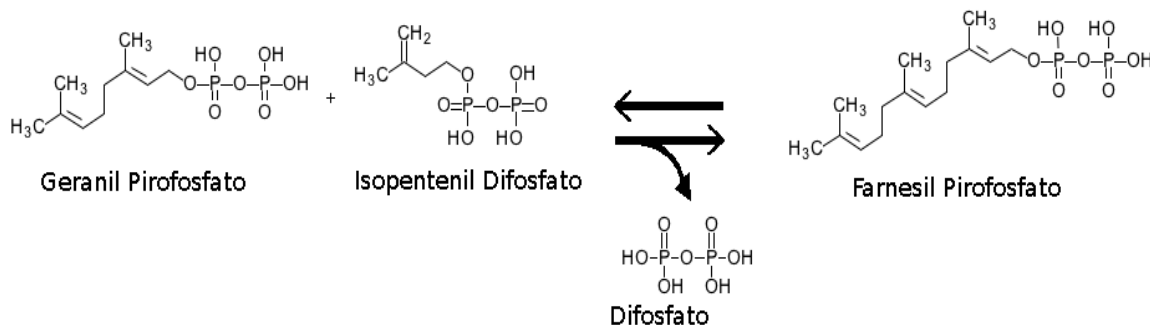


Figura 33: Reação catalizada pela FPPS. Adaptada do KEGG. <http://www.genome.jp/Fig/reaction/R01658.gif>

A Diidroorotato Desidrogenase (DHDH, C.E. **1.3.5.2**) é a responsável pela redução de diidroorotato a orotato (Figura 34), que por sua vez é um dos precursores para a biossíntese de pirimidinas. A Uracila, uma pirimidina que ocorre apenas em RNA, faz parte de um mecanismo de checagem de integridade celular, inibindo uma proteína chaperona chamada p53.

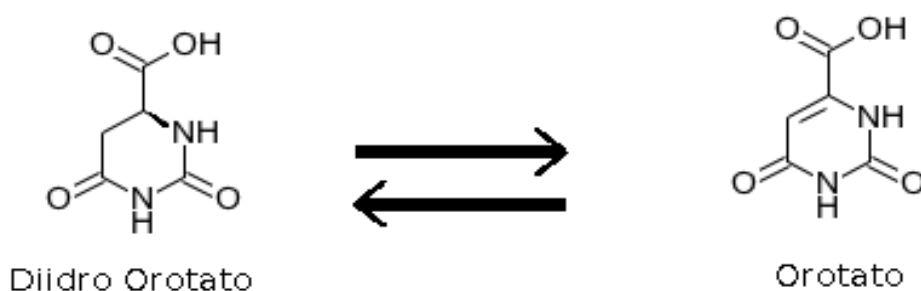


Figura 34: Reação de redução catalizada pela DHDH. Adaptado do KEGG, <http://www.genome.jp/Fig/reaction/R01868.gif>

Quando ativa, esta proteína suspende o ciclo celular e freia a proliferação das células. Uma suspensão prolongada do ciclo celular leva a célula à morte. Fármacos que exploram este mecanismo de ação já apresentaram resultados satisfatórios em *Plasmodium falciparum*, um dos parasitas causadores da Malária (BAIOCCO et al, 2011).

O terceiro alvo molecular estudado neste trabalho é a Tripanotiona Redutase (TR, C.E. **1.8.1.12**). Esta enzima reduz a tripanotiona dissulfeto a tripanotiona ditiol (Figura 35). Isto é importante porque a tripanotiona ditiol oxida-se, reduzindo radicais livres e outros compostos reativos a formas mais estáveis, protegendo os delicados componentes celulares contra stress oxidativo, em especial as membranas fosfolipídicas e os ácidos nucleicos.

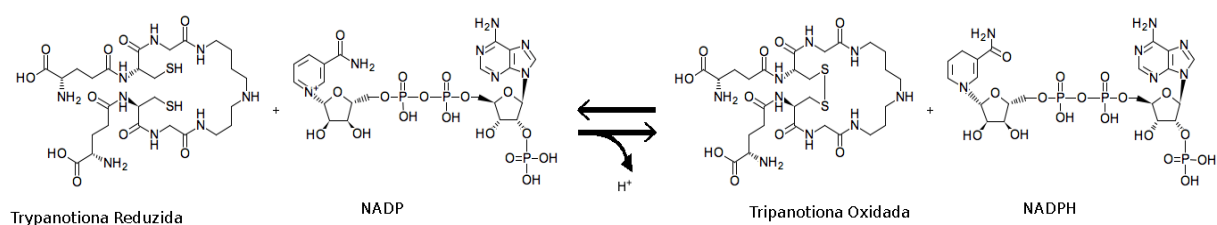


Figura 35: Reação de redução-oxidação da Tripanotiona. Adaptado do KEGG <http://www.genome.jp/Fig/reaction/R03821.gif>

Esta é a opção mais forte de alvo molecular porque células humanas não contêm esta enzima, mas uma enzima diferente de função análoga chamada de glutathione redutase (C.E. 1.8.1.7), que cataliza a redução de uma molécula bem menor chamada glutathione.

A TR e a glutathione redutase possuem funções semelhantes em organismos diferentes. Ao verificar pelo BLAST, no Uniprot, por sequências homólogas da TR (**2JK6**), vemos que a glutathione redutase mais próxima tem menos de 45% de homologia com ela (Figura 36), e pertence ao organismo *Thalassiosira pseudonana*, um fitoplâncton eucariótico.

Em vista destas informações, parece seguro supor que inibidores da TR têm grandes possibilidades de serem exclusivos da TR, o que os torna uma classe de ligantes interessante na busca de antiparasitários seletivos.

Trypanothione reductase, putative	Trypanosoma cruzi	492	66.0%
Trypanothione reductase, putative	Trypanosoma cruzi	492	66.0%
Trypanothione reductase	Trypanosoma cruzi	453	66.0%
Trypanothione reductase	Trypanosoma cruzi marinkellei	430	67.0%
Trypanothione reductase	Trypanosoma cruzi	430	66.0%
Trypanothione reductase	Trypanosoma cruzi	430	66.0%
Trypanothione reductase	Trypanosoma cruzi	430	65.0%
Trypanothione reductase	Trypanosoma cruzi	430	65.0%
Trypanothione reductase	Trypanosoma cruzi	430	65.0%
Trypanothione reductase	Trypanosoma cruzi	430	65.0%
Trypanothione reductase	Trypanosoma cruzi	430	65.0%
Trypanothione reductase	Trypanosoma vespertilionis	430	65.0%
Trypanothione reductase, putative	Trypanosoma cruzi	388	65.0%
Glutathione reductase, gro-2	Thalassiosira pseudonana (Marine diatom)	530	42.0%
Glutathione-disulfide reductase	Puniceispirillum marinum (strain IMCC1322)	450	43.0%
Putative uncharacterized protein	Selaginella moellendorffii (Spikemoss)	485	44.0%
Glutathione reductase	Brucella microti (strain CCM 4915)	464	41.0%
Putative uncharacterized protein	Selaginella moellendorffii (Spikemoss)	485	44.0%
Glutathione-disulfide reductase	Brucella ceti M490/95/1	464	41.0%
Glutathione reductase	Brucella sp. F5/99	464	41.0%
Glutathione-disulfide reductase	Brucella ceti B1/94	464	41.0%
Glutathione-disulfide reductase	Brucella ceti str. Cudo	486	41.0%
Glutathione reductase	Brucella melitensis	483	41.0%
Gor, glutathione reductase	Brucella abortus	464	41.0%
Pyridine nucleotide-disulphide oxidoreductase...	Brucella abortus (strain 2308)	464	41.0%

Figura 36: Excerto dos resultados produzidos pela ferramenta BLAST no portal UniProt para uma busca por sequências homólogas da sequência da subunidade A da enzima de código PDB 2JK6

3.2. Cálculos de Homologia das Enzimas Seleccionadas

Como não existiam estruturas cristalográficas para as enzimas seleccionadas nas espécies interesse, técnicas de modelagem por homologia de estrutura foram empregadas. O objetivo destas técnicas é calcular uma possível estrutura tridimensional para uma sequência baseada na estrutura tridimensional de outra sequência que guarde alguma similaridade com a primeira.

A modelagem por homologia parte de duas ou mais sequências e de ao menos uma estrutura tridimensional tomada como molde, e calcula uma estrutura tridimensional teórica para uma das sequências utilizadas, chamada modelo. De uma forma mais precisa, o algoritmo de homologia tenta prever os valores dos ângulos diedrais da cadeia principal da estrutura do modelo a partir de informações contidas na estrutura do molde.

Para fazer isto, o programa que nós utilizamos, o Modeller v9.9 (SALI; BLUNDELL, 1993), se vale de uma função de distribuição de probabilidade condicional calculada a partir de um grande conjunto de proteínas do PDB. Esta função associa uma probabilidade do valor de um ângulo diedral estar contido em um determinado intervalo quando outras condições são especificadas, por exemplo, o tipo de aminoácido e os valores dos ângulos diedrais em outros pontos da cadeia.

Para produzir um mapa dos pontos onde as duas sequências são diferentes e onde elas são iguais, utilizamos os algoritmos de alinhamento. As sequências do molde e do alvo foram alinhadas, utilizando o algoritmo ClustalX (LARKIN et al, 2007).

O alinhamento foi utilizado pelo programa de homologia de forma que onde as duas proteínas eram idênticas, os valores dos ângulos diedrais do molde foram utilizados diretamente no modelo, onde as duas sequências discordaram (e nas adjacências destes pontos), os ângulos diedrais foram calculados a partir das probabilidades condicionais.

As sequências, então, não podem ser muito diferentes entre si, pois para que a leitura correta da função de distribuição de probabilidade seja feita, quanto mais informação for conhecida sobre a estrutura do modelo, melhor. Desta forma menos informação precisa ser inferidas dos cálculos das probabilidades e mais exatamente esta inferência será efetuada.

Para a TR e a DHDH, que são homodímeros, a homologia foi feita apenas para uma única subunidade. Esta estrutura foi utilizada para os cálculos de afinidade, e não o dímero completo. Em seu artigo de 2009, Genheden e Ryde (2010), demonstraram não há diferença nos resultados dos cálculos de energia livre com seu método MMGBSA das múltiplas trajetórias para as subunidades do tetrâmero que eles utilizaram em seu estudo, mostrando que elas podem ser tratadas independentemente.

3.3. Geração Automatizada de Ligantes

Para produzir uma biblioteca para ser utilizada na triagem virtual para estas enzimas, utilizamos o programa BROOD v. 1.0, do pacote Open Eye (OPEN EYE BROOD, 2006).

A ideia foi realizar substituições em uma molécula, produzindo um banco de dados de diferentes derivados, com diferentes substituintes.

O BROOD é um pacote que contém um enorme banco de dados de fragmentos moleculares com diversas propriedades já calculadas. Entre estas propriedades temos peso molecular, resultados em modelos de farmacocinética, facilidade de obtenção sintética, quantidade de átomos pesados, e muitos outros.

O banco de dados é grande demais para ser utilizado integralmente, são mais de quatro milhões de fragmentos, então o usuário precisa restringir algumas propriedades da busca. Por exemplo, pedindo que o programa utilize apenas os fragmentos abaixo de um determinado peso molecular, ou apenas entre aquelas que passaram em testes farmacocinéticos. Dos vários índices disponíveis para se restringir o banco de dados, os mais importantes são Peso Molecular, Testes Farmacocinéticos e a Disponibilidade Sintética.

Em 1997, Cristopher A. Lipinski e seus colaboradores foram os primeiros a demonstrar que fármacos possuem propriedades físico-químicas e estruturais que estão restritas a determinados valores, delimitando um espaço de moléculas “semelhantes a fármacos” (*drug-like*) no espaço de todas as moléculas possíveis (LIPINSKI et al, 1997).

Este foi o início das chamadas “regras dos cinco” de Lipinski. O nome surgiu a partir das cinco propriedades físico-químicas e estruturais para as quais Lipinski e seus colaboradores

estabeleceram parâmetros claros para esta fronteira do espaço de moléculas semelhantes a fármacos: ClogP, MlogP, peso molecular e a quantidade de doadores e aceptores de ligações de hidrogênio.

O BROOD contém vários fragmentos que não obedecem a regra dos cinco, mas ele possui a opção de levar em consideração apenas aqueles que obedecem. Neste trabalho, utilizamos apenas os fragmentos contidos no espaço *drug-like*. Entretanto, mesmo assim a quantidade de fragmentos considerados no cálculo ainda seria muito grande. Para excluir mais fragmentos do cálculo e torná-lo factível, foi necessário estabelecer restrições.

Uma dessas restrições foi a exigência de que o índice de *Synthetic Availability* (Disponibilidade Sintética) dos fragmentos incluídos no cálculo fosse acima dos 80%.

Além disto, um ajuste manual das propriedades foi realizado de forma a deixar o banco de dados com até 500 mil fragmentos. O ajuste manual foi realizado em duas etapas. Primeiro o peso molecular foi cortado de 10 em 10 por cento. Depois, ao atingir a metade do peso molecular original, se a quantidade de fragmentos não estivesse satisfatória, alterava-se em 1 a quantidade de átomos pesados até que fosse o caso.

Depois de restringido o banco de dados, o cálculo dos análogos foi realizado de modo a tomar os fragmentos que se assemelhassem com o ligante quanto à superfície eletrostática e à forma geométrica. Isto é, o programa tomou a molécula de entrada e substituiu a região designada por cada um dos fragmentos-teste, obtendo propriedades para uma série de análogos da molécula base.

Em seguida foram selecionados, dos análogos produzidos, os 100 resultados mais próximos em questão geométrica e eletrostática, acionando a opção de que os ligantes fossem agrupados em *clusters* de similaridade.

Tomamos um representante de cada *cluster* e ordenamos estes representantes segundo a disponibilidade sintética. Esta lista final de representantes dos *clusters* chamamos de base.

O BROOD não permite a modificação simultânea de duas regiões da molécula de entrada. Para contornar este problema, dividimos os ligantes cristalográficos em regiões e os variamos de forma independente, produzindo uma base para cada região. Ao final, tínhamos uma base para cada região dos ligantes.

As moléculas em uma base diferem entre si apenas em uma única região, e são idênticas à molécula de entrada em todos o resto da molécula. O procedimento então seria tomar de cada membros, o grupo que ele possui nesta região variável, e realizar uma combinação completa de todos os grupos de cada base com todos os grupos de cada uma das outras base para produzir um conjunto de ligantes.

O problema é que cada seção da molécula deu origem a uma base com 12 a 15 moléculas, o que significa dizer que para uma única enzima, cujo ligante cristalográfico tivesse sido dividido em duas regiões, teríamos que avaliar entre 144 e 255 moléculas. Isto geraria cerca de 2.550 dinâmicas só para uma enzima. Este tipo de explosão combinatória torna infactível a utilização das combinações completas.

Para reduzir este problema, selecionamos apenas os membros de cada base com melhor acessibilidade sintética e utilizamos seus grupos variáveis para gerar as combinações.

3.4. Docking Molecular

O *docking* molecular foi realizado utilizando o programa Autodock 4.2. Entretanto, o

procedimento envolveu toda uma preparação antes que o *docking* pudesse ser realizado. Esta preparação foi com os programas do pacote AutoDock Tools (SANNER, 1999).

Antes que qualquer coisa possa ser feita, os átomos do ligante precisam receber um *tipo de átomo* do Autodock. Este tipo de átomo enquadra o átomo como tendo determinadas propriedades específicas, por exemplo, um carbono sp^3 , um oxigênio carregado ou um nitrogênio de amina quaternária.

O tipo de átomo determina vários parâmetros atômicos e de ligação que fazem parte do cálculo de afinidade do Autodock, de uma forma muito semelhante (e às vezes até mesmo idêntica) à mecânica molecular e seus campos de força empíricos. Além disto, é preciso fazer o cálculo das cargas parciais para todos os átomos. O Autodock disponibiliza um algoritmo para fazer isto utilizando o método de Gasteiger e Marsili (1978) e os tipos de átomos foram determinados pelo Autodock, também, seguindo uma heurística interna própria.

O sítio ativo e o espaço onde os cálculos de afinidade e interação foram realizados foi feito enquadrando-se o sítio ativo em um *grid* (Figura 37), e ajustando-se seu tamanho de modo a não ficar tão grande a ponto de inviabilizar o cálculo mas também não tão pequeno que ligantes um pouco maiores não pudessem se acomodar adequadamente, produzindo assim escores pouco realistas.

O *grid* foi definido a partir de ligante. Isto quer dizer que seu centro foi colocado exatamente no centro geométrico da posição dos ligantes cristalográficos, de forma a capturar todo o sítio ativo. Dependendo do tamanho do substrato natural, o tamanho do *grid* variou.

Para a DHDH o *grid* foi de 40x38x46, para a FPPS foi de 40x40x40 e para a TR foi de 60x60x60. Os tamanhos foram ajustados manualmente de forma a conter todo ligante cristalográfico utilizado no posicionamento do *grid* e mais um espaço extra para permitir a melhor acomodação dos ligantes durante o cálculo.

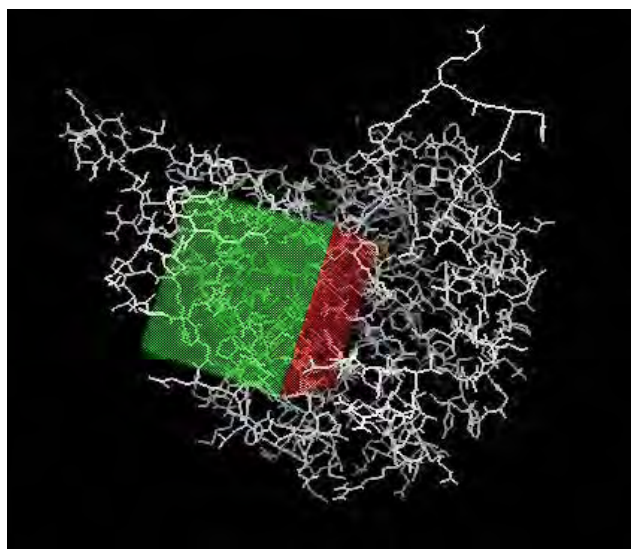


Figura 37: Grid 50x38x46 da DHDH.

Para cada ponto deste *grid*, vários parâmetros foram calculados, principalmente o potencial elétrico, e a tendência a ser doador ou aceptor de ligações de hidrogênio. Este *grid* foi utilizado para realizar o *Docking Molecular*, ao invés do cálculo diretamente com os átomos. Os motivos disto ser feito desta forma são os mesmos discutidos para as dinâmicas moleculares. O cálculo com o *grid* é mais rápido e mais econômico que o cálculo diretamente com os átomos. No Autodock, o *grid* foi gerado por um programa à parte chamado Autogrid.

O Autodock trabalha, sobretudo, com rotações de ângulos diedrais para gerar as conformações a serem avaliadas. Para fazer isto, determinamos uma *raiz* ou *núcleo* rotacional de cada ligante, que corresponde ao átomo cujas ligações tenham a menor mobilidade possível. A partir deste núcleo o Autodock estabeleceu a mobilidade de cada ligação e criou uma representação interna da molécula, que ele utilizou para gerar as conformações. As informações de carga do ligante e de sua raiz e torsões foram todos salvos em um arquivo tipo PDBQT.

As conformações dos ligantes para o *docking* foram geradas utilizando-se um algoritmo genético Lamarckiano (MORRIS et al, 1998). 100 estruturas foram geradas, cada uma delas produzida por um cálculo realizado sobre uma população de 500 indivíduos. Cada cálculo foi efetuado por até 27.000 gerações ou 2 milhões e 500 mil avaliações de afinidade, o que ocorresse primeiro. Utilizou-se a taxa de mutação de 0,02 e a de cruzamentos foi 0,8. Todos os cálculos de *docking* foram feitos com o alvo molecular rígido e o ligante flexível.

Após os cálculos de *docking*, agrupamos as conformações baseadas no valor de escore, em *clusters* de conformações semelhantes. Em seguida tomamos a estrutura de melhor (mais negativo) escore dentro do *cluster* de maior número de conformações. Incidentalmente, este *cluster* sempre era o de mais baixa energia, de modo que no fim de tudo, as estruturas cujos valores de afinidade foram tomados para análise foram aquelas de mais baixo escore.

3.5. Cálculos de Estimativa da Energia Livre

A variação da energia livre é uma grandeza macroscópica, propriedade de grandes quantidades de partículas. Sua estimativa a partir de informações microscópicas envolvendo átomos e moléculas só pode ser feita a partir do emprego do formalismo da termodinâmica estatística. Como vimos, os algoritmos que empregam o formalismo da termodinâmica estatística são os algoritmos de simulação molecular (dinâmica molecular e o método de Monte Carlo).

O *docking* molecular não é um destes métodos, o que significa que os escores do *docking* não são estimativas propriamente ditas de variações da energia livre. Os cálculos de *Docking* molecular são, na verdade, aproximações empíricas baseadas em parâmetros energéticos e geométricos de dados experimentais de complexos.

Neste trabalho, analisamos a possibilidade de substituir o *docking* molecular com a técnica do MMGBSA sobre uma única minimização nas etapas de análise *in silico* de triagem molecular em pequenos grupos de ligantes. Como não dispomos de valores experimentais da variação da energia livre para estas interações ligante-enzima, foi necessário comparar estes dois métodos com alguma estimativa mais exata de energia livre.

O parâmetro de comparação, tomado como a estimativa mais exata da energia livre de que dispomos é o já mencionado método do MMGBSA sobre múltiplas trajetórias, como proposto por Genheden E Ryde (2010). Trabalhos utilizando o MMGB/PBSA em geral têm demonstrado bons resultados (CASE; GOHLKE, 2004; STOICA; SADIG; COVENEY, 2008; GUODONG et al, 2010).

Em particular, trabalhos utilizando especificamente o método de MMGBSA das múltiplas trajetórias têm obtido correlações muito boas com valores experimentais e no ordenamento de ligantes, inclusive obtendo correlações experimentais na faixa de métodos tidos como exatos, como o método da Perturbação da Energia Livre (SADIQ et al, 2010; GUIMARÃES, 2011).

Por causa destes resultados, o método das múltiplas trajetórias foi tomado por nós como uma estimativa suficientemente exata da energia livre mas também suficientemente rápido para que

pudesse ser feito para todos os ligantes no decorrer do trabalho. Para calcular as dinâmicas moleculares utilizamos o pacote AMBER 11 (CASE et al, 2010).

Antes de efetuar as dinâmicas, realizamos uma etapa de preparação, pois as estruturas em formato PDB (dos alvos moleculares) e Sybyl mol2 (ligantes) não são utilizadas pelo *pmemd* (executável do pacote AMBER 11 que roda as dinâmicas implementando o método PME para o cálculo das interações não-ligadas).

3.5.1. Preparação das Estruturas dos Complexos Enzima-Ligante

O *pmemd* precisa, como entrada, de um arquivo de *topologia* e um arquivo de *coordenadas* além dos parâmetros da simulação (temperatura, acoplamentos, vínculos, números de passos, informações de *output*, entre outras coisas).

Um arquivo de topologia descreve a forma como os átomos estão conectados entre si, bem como os tipos dos átomos e das ligações e ângulos, de forma que seja possível para o programa conhecer os parâmetros do campo de força para cada átomo. Sem esta informação, o programa não é capaz de calcular as energias e o procedimento não acontece. As coordenadas descrevem as posições de cada átomo no espaço. Ambos podem ser gerados, no pacote AMBER, a partir de um programa chamado *Leap*.

Os parâmetros do campo de força utilizados pelos átomos do ligantes são determinados segundo seu tipo de átomo. Os tipos de átomos dos vinte aminoácidos principais são padronizadas neste meio de estudo e portanto é seguro supor que o *Leap* não vai errar na geração da topologia de uma proteína pura. O mesmo não pode ser dito, entretanto, de um sistema envolvendo moléculas menos canônicas, como um ligante sozinho ou um complexo de uma proteína com um ligante. Há, portanto, necessidade de se atribuir tipos de átomos corretos para os átomos dos ligantes.

Além disto, o cálculo das cargas parciais da proteína é automático, bem como a adição dos hidrogênios que não são detectados pela cristalografia de raios-X e precisam ser adicionados posteriormente, de forma computacional.

No caso da molécula do ligante, as posições dos hidrogênios, cargas parciais e tipos de átomo ficam indefinidos. Para resolver o primeiro problema, o AMBER dispõe do programa *reduce*, para o segundo utilizamos o *antechamber* e para o terceiro e último temos o *parmchk*.

O uso do *reduce* é bastante direto. O programa recebe um arquivo PDB e escreve um arquivo PDB de saída idêntico ao de entrada, mas com os hidrogênios. Ele é capaz de fazer isto até mesmo para PDBs que não sejam de proteínas. Mais complicado é o uso do *antechamber*. Neste, foi preciso definir a carga formal da molécula (calculado no UCSF Chimera 2.0 (PETTERSEN et al, 2004) e decidir um algoritmo para o cálculo das cargas parciais, pelo método AM1-BCC (JAKALIAN; JACK; BAYLY, 2002), bem como um nome para a molécula (um conjunto de letras, em geral três, que identifica os átomos do ligante entre os átomos do sistema).

A saída do *antechamber* é um novo arquivo, com as cargas calculadas e o nome de molécula estabelecido. O programa também é capaz de converter um tipo de arquivo em outro. Esta propriedade foi utilizada para conseguir o arquivo mol2 dos ligantes retirados dos complexos cristalográficos PDB das enzimas.

O *parmchk*, por fim, cria um arquivo de extensão do campo de força e inclui nele, parâmetros para os átomos cujo tipo ele não seja capaz de determinar a partir do Campo de Força AMBER de Uso Geral (*General Amber Force Field*, gaff), que é um campo de força utilizado pelo AMBER para conter tipos de átomos de moléculas gerais (não necessariamente aminoácidos,

nucleotídeos, glicídios e água). Este arquivo de extensão é carregado no Leap antes de se produzir as topologias.

Com os hidrogênios, cargas e tipos de átomos bem definidos para os ligantes, o *Leap* pode produzir suas topologias e coordenadas. Nós utilizamos este procedimento em três passos para todos os ligantes. Para cada um deles foram adicionados os hidrogênios, calculadas as cargas e definidos os tipos de átomos e extensões do *gaff*.

A segunda etapa foi a de produzir as topologias e coordenadas. Para o nosso estudo, precisamos ter pelo menos quatro topologias, uma para a proteína sozinha no vácuo, uma para o ligante sozinho no vácuo, uma do complexo sozinho no vácuo e uma do complexo solvatado em água, que será utilizada para produzir as trajetórias.

O posicionamento do ligante no sítio ativo para produzir as estruturas dos complexos foi feito utilizando-se o programa de Autodock 4.2, mas com um algoritmo muito mais simples e rápido que o *docking* utilizado para estimar afinidade de ligação. Apenas 10 poses foram geradas e aquela de menor energia foi escolhida para servir de estrutura de partida. Cada uma das poses foi produzida a partir da propagação, pelo algoritmo genético do Autodock, de uma população de 100 indivíduos.

Cada população foi propagada por até 27.000 gerações ou através de 250 mil avaliações de afinidade, o que ocorresse primeiro. Utilizou-se a taxa de mutação de 0,02 e a de cruzamentos foi 0,8. Todos os cálculos de *docking* foram feitos com o alvo molecular rígido e o ligante flexível.

Neste ponto, parâmetros importantes devem ser observados. O primeiro deles, são os raio de Born. São os raios utilizados nos cálculos de GB, utilizados pela equação (98). Assim como cargas e campos de força, existem vários esquemas para se atribuir um raio de Born a um átomo. Neste trabalho, todos os cálculos foram feitos com o método *mbondi*.

O segundo parâmetro a ser observado é a geração da topologia solvatada. Como visto na seção 1.2.2.6. sobre a Soma de Ewald, as somas S_p para $p \leq 3$ que contabilizam as interações não-ligadas de decaimento lento, como a componente dispersiva do potencial de Lennard-Jones e as interações eletrostáticas, só convergem se a carga total da caixa de simulação for zero.

Todavia, a carga total de proteínas em geral não é zero. Elas têm vários aminoácidos cuja cadeia lateral apresenta grupos fracamente ácidos e/ou básicos cujo estado de protonação, e consequentemente sua carga, depende do pH. O valor de pH onde a carga total de uma proteína é zero é chamado *ponto isoelétrico* e este valor varia de proteína para proteína mas pode acontecer de ser o mesmo em duas proteínas distintas.

Os ligantes também não são, em geral, eletricamente neutros, de modo que o complexo proteína-ligante quase nunca será naturalmente neutro. O que se faz, então, é adicionar, à caixa de simulação, íons de carga oposta à carga total do complexo, os chamados *contra-íons*, para neutralizar o sistema. Somente então o complexo macromolecular proteína-ligante é solvatado e então sua topologia e as coordenadas de seus átomos são salvos em arquivos *.top* e *.crd* respectivamente.

Os complexos foram solvatados em uma caixa em formato octaédrico truncado, utilizando o modelo de água TIP3P. A utilização desta caixa muda um pouco o formalismo das fronteiras periódicas e da convenção da mínima imagem, mas o princípio permanece o mesmo. A maior vantagem deste tipo de caixa é que a quantidade de moléculas de água utilizadas no procedimento de solvatação é reduzido significativamente, o que significa dizer menos cálculos e mais eficiência.

Uma vez que todos os complexos estivessem preparados, a geometria foi otimizada utilizando o programa *pmemd* do pacote AMBER. Foram 5000 passos de minimização, com raio de

corte de 12 Å dos quais 500 passos utilizaram o algoritmo de otimização conhecido por *steepest descent* e os outros 4500 passos seguiram o algoritmo dos gradientes conjugados.

Estas estruturas minimizadas foram utilizadas tanto para o cálculo do MMGBSA minimizado quanto como ponto de partida para as múltiplas trajetórias.

3.5.2. Dinâmica Molecular

A partir das estruturas minimizadas, realizamos duas etapas de dinâmica molecular. A primeira, com o objetivo de aquecer o sistema da temperatura de mínimo até a temperatura de 300K. A segunda etapa, mais longa, com o objetivo de amostrar e *produzir* os microestados da temperatura correta a serem utilizados nos cálculos de MMGBSA.

Como estamos utilizando o método de Genheden E Ryde (2010), nós não produzimos uma única série de dinâmicas para cada ligante, mas dez. Em cada série, as velocidades iniciais do sistema foram selecionadas de forma aleatória a partir da distribuição na equação (23) de modo que cada série de dinâmicas sondasse uma área diferente do espaço de fase e ao fazer isso, gerasse dez dinâmicas de produção distintas. (Figura 38).

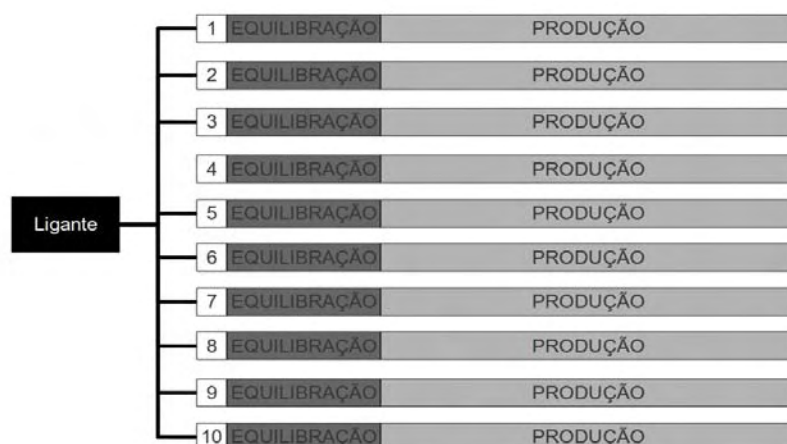


Figura 38: Esquema ilustrando as séries de dinâmicas moleculares

No artigo de referência do método das múltiplas trajetórias, uma série de tempos de equilíbrio e produção são testados para verificar em que ponto a convergência do desvio padrão dos valores calculados da variação da energia livre é atingida. Em particular, se a dinâmica de produção for mais longa que 150ps, 40ps de equilíbrio são suficiente para atingir a convergência.

Cada dinâmica de equilíbrio levou 25.000 passos e cada dinâmica de produção levou 250.000 passos. Entretanto, para conseguir amostrar mais tempo mesmo com dinâmicas tão curtas (em virtude da quantidade de séries por fazer: 690), utilizamos uma técnica para congelar o estiramento das ligações envolvendo hidrogênio. Com estas ligações congeladas, não foi preciso calcular suas energias, mas apenas atribuir uma energia de equilíbrio.

Ignorar estes graus de liberdade, no entanto, interfere nas propriedades termodinâmicas

estimadas a partir da trajetória. Por este motivo, estabelecer vínculos sobre estas ligações requer algoritmos específicos.

No AMBER o algoritmo implementado para realizar este tipo de cálculo é o SHAKE, um dos primeiros algoritmos a fazer isto de forma consistente, desenvolvido por Ryckaert, Ciccotti e Berendsen (1977). Com estas ligações congeladas, pudemos utilizar um passo de 2fs e assim conseguir sondar o espaço de fase por intervalos de tempo mais longos, considerando microestados mais distintos entre si. O parâmetro Γ do termostato de Langevin foi acertado de modo que $\ln(\Gamma)=2,0$. Com passos de 2fs, as dinâmicas de equilíbrio têm 50ps de duração e as de produção 500ps. Suficiente para atingir a convergência estatística.

As dinâmicas foram realizadas a pressão e temperatura constantes, para simular o interior de uma célula viva. Utilizamos o termostato de Langevin e o barostato de Berendsen para equilibrar o sistema e sondar o *ensemble* NPT. A pressão foi controlada de forma isotrópica.

Uma consequência disto é que a variação da energia livre amostrada pelas simulações não será da energia de Helmholtz, mas da energia de Gibbs, o ΔG de interação. E a energia termodinâmica calculada pelos campos de força será uma estimativa da Entalpia H do sistema, ao invés da Energia Interna U como era o caso no *ensemble* NVT.

Todos os cálculos foram realizados utilizando a tecnologia CUDA (*Computer Unified Device Architecture*) da Nvidia, de forma que a maior parte do cálculo foi realizada utilizando GPUs (*Graphical Processing Units*, unidades de processamento gráfico das placas de vídeo) ao invés de CPUs. Isto acelerou bastante o cálculo mas também gerou problemas.

Um dos maiores problemas foi a instabilidade do algoritmo de precisão dupla. O problema consistia no fato de que, aleatoriamente, as dinâmicas que estavam sendo calculadas em precisão dupla simplesmente caíam ou paravam. Em virtude deste fato, muito mais dinâmicas que o esperado foram realizadas para conseguir a amostragem devida.

Utilizamos o AMBER 11 com *bugfixes* de 1 ao 14 instalados e efetuamos cálculos nas placas Nvidia GTX 480 e Nvidia GTX 580. Em ambas as placas o problema continuou mesmo após a instalações de todos os *bugfixes* disponíveis até então. Apenas quando mudamos para um algoritmo misto de precisão simples e dupla é que os problemas cessaram.

O *pmemd* escreveu na trajetória e no arquivo de saída uma vez a cada 500 passos, produzindo trajetórias por volta de 1.0 Gigabyte, variando um pouco de enzima para enzima. Com as trajetórias de produção prontas, o próximo passo foi utilizar a metodologia de Genheden e Ryde (2010) para calcular a variação da energia livre.

3.5.3. MMGBSA

Após a equilíbrio nas dinâmicas de produção, a amostragem dos passos foi feito de 5 em 5, seguindo o intervalo de amostragem de 5ps da metodologia de estimativa da variação da energia livre utilizando múltiplas trajetórias. Os cálculos de MMGBSA foram realizados em um campo de solvente implícito simulando uma solução salina a 0,15 M.

Apesar de formalmente o cálculo da variação da energia livre nesta técnica passar pelo cálculo de entropia a partir dos modos normais do sistema, segundo a equação (103), na prática este cálculo não foi feito.

Os cálculos de modos normais são bastante demorados e custosos computacionalmente. O emprego deste tipo de cálculo inviabiliza completamente os cálculos de triagem virtual, que são o foco deste nosso estudo. Entretanto, existe um motivo mais importante.

Foi mostrado que os cálculos de entropia diminuem a correlação dos valores preditos de afinidades com os valores experimentais, quando o cálculo é feito com poucos microestados amostrados das trajetórias (RASTELLI et al, 2009; HOU; WANG; WANG, 2011). Como as limitações de tempo de cálculo são importantes, incluir estes cálculos no procedimento não é encorajado.

Excetuando-se o fato de que apenas um único microestado foi utilizado no cálculo da variação da energia livre, os cálculos de MMGBSA para a estrutura minimizada foram feitos da mesma forma.

4. Resultados e Discussões

4.1. Análise dos Modelos por Homologia

Quando tomamos um grande conjunto de proteínas, verificamos que os aminoácidos preferem algumas conformações e outras, não. Isto quer dizer que para existem algumas combinações particulares de valores de ângulos diedrais que ocorrem com mais frequência que outras nas cadeias polipeptídicas de proteínas cristalográficas.

Presumivelmente, isto acontece devido à otimização das interações químicas e redução do impedimento estérico das cadeias laterais dos aminoácidos. Quando um aminoácido encontra-se em alguma conformação que sua estereoquímica dificulta assumir, isto significa que há uma determinada tensão na cadeia.

Uma cadeia tensionada não pode estar na conformação nativa, pois a conformação nativa de uma proteína é uma conformação notoriamente estável. Este dado simples porém importante pode nos auxiliar a avaliar a qualidade das estruturas obtidas pela modelagem por homologia.

Baseado nestas conformações características de cada aminoácido, pode-se calcular quais os intervalos de valores de ângulos diedrais favorecidos para cada aminoácido e descobrir quais os valores mais favorecidos. O PROCHECK (LAKOWSKI et al, 1993) faz exatamente isto. Ele estima qual percentual de aminoácidos de uma cadeia estão em intervalos diedrais favoráveis. Estas estimativas foram obtidas a partir do estudo estatístico de um conjunto de 118 proteínas representativas do PDB.

Na Tabela 2 vemos um sumário dos dados produzidos pelo PROCHECK. O percentual de aminoácidos em suas respectivas regiões diedrais.

Tabela 2: Percentual de aminoácidos de cada enzima em regiões de qualidade dos diedrais.

Regiões Diedrais	DHDH	TR	FPFS
Regiões Mais Favorecidas	92,5%	93%	97,3%
Regiões Permitidas	7,1%	6,5%	2,7%
Regiões Generosamente Permitidas	0,2%	0,2%	0%
Regiões Não Permitidas	0,2%	0,2%	0%

As “regiões” do PROCHECK representam intervalos de valores de ângulos diedrais. As Regiões Mais Favorecidas representam aquele intervalo de valores que mais ocorrem em conformações nativas por aquele aminoácido.

Regiões Permitidas e Generosamente Permitidas representam intervalos de valores que vários tipos diferentes de aminoácidos podem assumir em diferentes condições, o que significa dizer que estes valores não representam, necessariamente, um aumento na tensão da cadeia.

As Regiões Não Permitidas representam intervalos de valores de ângulos diedrais desfavoráveis para a estereoquímica do tipo de aminoácido sendo analisado.

Como podemos ver, todas as proteínas possuem mais de 90% dos aminoácidos nas regiões mais favorecidas pela sua estereoquímica. Em cada uma delas, entretanto, a cadeia ainda apresenta alguma margem de tensão devido ao pequeno percentual de aminoácidos em regiões não

permitidas. Esta tensão nos diedrais da cadeia, pode ser relaxada com uma minimização de energia com mecânica molecular ou uma equilibração utilizando-se dinâmica molecular.

Mesmo a FPPS, que era a enzima cuja homologia entre o molde eu modelo foi a menor das três (60%), apresentou um percentual alto de aminoácidos assumindo valores diedrais dentro da faixa dos mais favorecidos. Os alinhamentos dos modelos com os moldes apontam que realmente o modelo produzido não possui distorções em relação ao molde (Figuras 39, 40 e 41).



Figura 39: Alinhamento estrutural da FPPS, amarelo, com seu molde em preto.

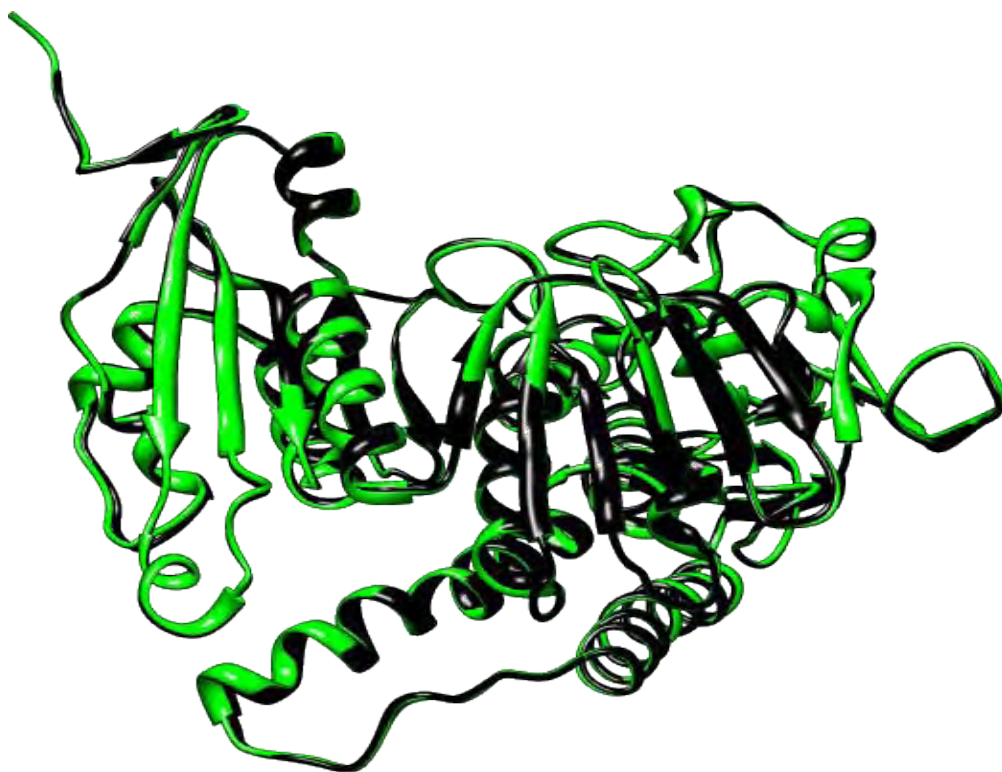


Figura 40: Alinhamento estrutural da TR, verde, com seu molde em preto.



Figura 41: Alinhamento estrutural da DHDH, azul, com seu molde em preto.

Todas as enzimas parecem coincidir particularmente bem com seus moldes. É importante notar, entretanto, o desvio da extremidade carbóxi-terminal da DHDH em relação ao seu molde. Isto ocorreu em virtude da diferença de comprimento entre as sequências, de forma que o modeller não conseguiu realizar a inferência adequada naquela parte da estrutura. No entanto, esta região de menor qualidade não se encontrava próxima do sítio ativo.

Para quantificar as diferenças entre as geometrias dos moldes e dos modelos, utilizamos o método do RMSD (*Root Mean Square Deviation*), uma medida que vai de 0.0 (total superposição) a qualquer número positivo. Quanto maior o RMSD, maior a discordância entre os átomos das cadeias principais das estruturas. Um RMSD de 0,2, por exemplo, significa dizer que, em média, os átomos das cadeias principais das duas estruturas estão deslocados cerca de 0,2 Å uns dos outros, em qualquer direção. Os valores de RMSD da DHDH, TR e FPPS estão colocados na Tabela 3.

Tabela 3: Sumário dos dados da homologia das três enzimas.

	RMSD (Å)	Similaridade	Molde	Modelo
DHDH	0,367	93%	<i>L. major</i>	<i>L. amazonensis</i>
TR	0,251	95%	<i>L. infantum</i>	<i>L. amazonensis</i>
FPPS	0,380	60%	<i>T. cruzi</i>	<i>L. chagasi</i>

Os dados apontam que os modelos de homologia apresentam boa qualidade estrutural. Não se nota nenhum desvio grosseiro, nenhuma incongruência ou irregularidade estrutural foi detectada. Além disto, os dados de RMSD e do PROCHEK apresentam parâmetros numéricos para nos convencer de que os modelos estão bons o suficiente para serem utilizados nas simulações.

É notável que o DHDH não tenha sido o maior RMSD. Sabendo que sua extremidade carbóxi-terminal ficou em uma geometria bastante distinta da do molde, em virtude da diferença de comprimento das sequências, podemos supor que as concordâncias entre todos os outros pontos das estruturas foi muito baixo, compensando o desvio da extremidade carbóxi-terminal.

Após as homologias, os ligantes cristalográficos dos moldes da DHDH e da FPPS foram transferidos para os modelos. O molde da FPPS contém o risedronato (Figura 42), um conhecido inibidor desta enzima (GABELLI et al, 2006) e o molde da DHDH contém o Orotato (Figura 43), seu substrato natural. Para verificar a validade do procedimento de transferência, alinhámos as estruturas e as sequências de cada modelo com seu respectivo molde. Este procedimento foi feito com o software UCSF Chimera 2.0 (PETTERSEN et al, 2004).

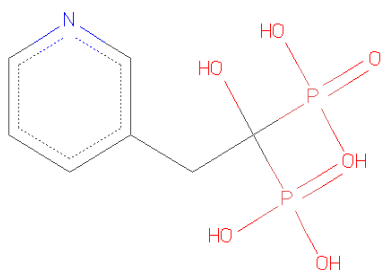


Figura 42: Risedronato, inibidor da FPPS.
Retirado de
http://www.pdb.org/pdb/images/RIS_500.gif

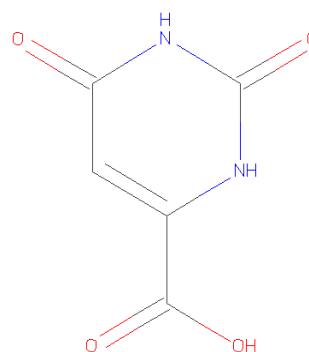


Figura 43: Orotato, substrato natural da DHDH.
Retirado de
http://www.pdb.org/pdb/images/ORO_500.gif

Ainda utilizando o UCSF Chimera, selecionamos os aminoácidos que estivessem em um raio de 5Å, 7Å e 9Å de distância do ligante, tanto no molde como no modelo, e comparamos as duas seleções para cada distância. A seleção foi feita para os aminoácidos que tivessem pelo menos um átomo dentro do raio considerado. A seguir, vemos as seleções de 5Å da DHDH (Figura 44).



Figura 44: Alinhamento DHDH. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.

Como é possível ver, praticamente todo o sítio ativo é idêntico, com exceção de um único aminoácido, selecionado no modelo mas não selecionado no molde. Como as duas estruturas não

são idênticas, os átomos não são perfeitamente superponíveis, e portanto alguns aminoácidos que são do sítio ativo ficam levemente fora do raio arbitrário de 5Å. Quando consideramos um raio de 5,2Å, por exemplo, aquela Serina que não foi selecionada no molde, passa a ser.

Este fenômeno não aparece na seleção aos 7Å (Figura 45), mas ocorre várias vezes na seleção dos 9Å (Figura 46).



Figura 45: Alinhamento DHDH. Seleção em verde representa aminoácidos a 7Å do ligante cristalográfico.



Figura 46: Alinhamento DHDH. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.

Estas falhas de seleção no raio de 9Å, entretanto, não são facilmente solucionadas com um pequeno aumento no raio de corte, como no caso da seleção do raio de 5Å. Quando aumentamos o raio para 9.2Å, percebemos que algumas falhas de seleção são corrigidas, mas outras aparecem. E este fenômeno torna-se crescentemente mais problemático à medida que aumentamos o raio de seleção, sugerindo que estas falhas são discordâncias verdadeiras dos sítios, e não equívocos produzidos por pequenos desvios causados pelas diferenças entre as estruturas comparadas.

Os alinhamentos da FPPS colocados abaixo (Figuras 47, 48 e 49 respectivamente para 5Å, 7Å e 9Å) se comportaram de maneira levemente diferente, talvez refletindo a menor homologia que as sequências do molde e do modelo guardam uma com a outra.



Figura 47: Alinhamento FPPS. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.



Figura 48: Alinhamento FPPS. Seleção em verde representa aminoácidos a 7Å do ligante cristalográfico.



Figura 49: Alinhamento FPPS. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.

Diferentemente dos alinhamentos da DHDH, mesmo as falhas de seleção do raio de 5Å não puderam ser facilmente corrigidas, sugerindo que talvez elas sejam discordâncias genuínas. O aumento do raio de fato corrigiu as seleções que aparecem em raios menores, mas também produziram mais falhas de seleção, um comportamento semelhante ao raio de 9Å da DHDH.

Com ou sem falhas de seleção, entretanto, é visível que uma parcela significativa dos aminoácidos selecionados em cada cadeia (contando com as falhas corrigidas) concordam entre si (são idênticos ou do mesmo tipo, que no alinhamento são representados pela mesma cor da letra). Utilizamos este fato para justificar a transferência dos ligantes dos moldes para os modelos. Ao menos para a DHDH e a FPPS.

A TR é um caso menos trivial porque seu molde (PDB **2JK6**) não possui ligantes cristalográficos. A tripanotona (Figura 50), substrato natural desta enzima, encontra-se cristalizada apenas na tripanotona redutase de *Trypanosoma cruzi* (PDB **1BZL**).

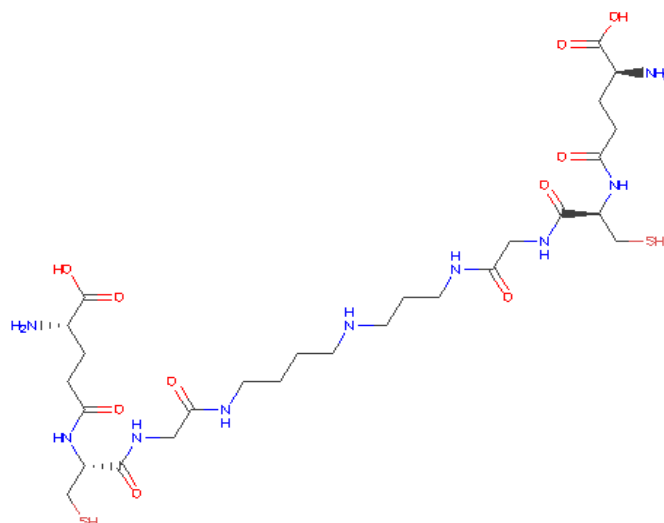


Figura 50: Tripanotiona, substrato natural da TR. Retirado de http://www.pdb.org/pdb/images/GCG_500.gif

Por isso precisamos comparar de forma mais cuidadosa os sítios ativos de cada enzima. Analisemos então, os alinhamentos da TR, mostrados abaixo (Figuras 52, 53 e 54 respectivamente para 5Å, 7Å e 9Å) :



Figura 51: Alinhamento TR. Seleção em verde representa aminoácidos a 5Å do ligante cristalográfico.





Figura 53: Alinhamento TR. Seleção em verde representa aminoácidos a 9Å do ligante cristalográfico.

A análise dos alinhamentos da TR possuem um pouco mais de pontos de discordância entre as três sequências mas, de modo geral, os resíduos que fazem parte do sítio ativo são conservados nas três sequências.

Para delimitar realmente o sítio ativo, decidimos verificar a partir de que raio de seleção os resíduos catalíticos documentados na literatura destas duas enzimas estavam presentes. Para fazer isto, utilizamos os dados do banco de dados CSA (*Catalytic Site Atlas*) (PORTER; BARTLETT; THORNTON, 2004).

Segundo o CSA, os resíduos catalíticos da FPPS **1YHL** são a Arg107 e a Phe246. Os números dos resíduos nas figuras dos alinhamentos podem ser vistos à esquerda de cada linha de sequência. E as linhas, por sua vez, estão separadas por espaços em branco em blocos de dez aminoácidos cada. Cada linha possui no máximo cinco blocos, totalizando cinquenta aminoácidos,

Ao verificarmos os alinhamentos da FPPS, verificamos que os resíduos Arg107 e a Phe246 encontram-se ambos conservados nas duas enzimas, mas ambos são selecionados apenas a partir do raio de 7Å. Detalhamento na análise mostram que eles são selecionados, na verdade a partir de 5.9Å, sugerindo que talvez 6Å seja um bom raio de delimitação do sítio ativo.

Para a TR, é necessário analisar tanto a **1BLZ** quanto a **2JK6**. Segundo o CSA, a **2JK6** possui quatro resíduos catalíticos em seu sítio ativo. Cys53, Cys57, His461 e Glu466. Pelo alinhamento percebemos que todos estes aminoácidos encontram-se conservados por todas as três enzimas. Como aconteceu com a FPPS, estes quatro aminoácidos só se encontram dentro da seleção do sítio ativo quando incluímos aminoácidos a até 6Å do sítio ativo. Seleções de aminoácidos produzidas com valores de corte menores começam a não incluir todos os resíduos catalíticos.

Para a **1BZL** o CSA lista um aminoácido catalítico a mais, a Lys61. Este aminoácido encontra-se conservado nas três sequências, mas não é possível selecioná-lo mesmo a um raio de 9Å da tripanotona cristalográfica. A partir do FAD cristalizado na **1BZL**, no entanto, é possível selecioná-lo até mesmo em raios de seleção tão pequenos quanto 3Å, nas três enzimas. Ao que parece, o resíduo Lys61 deve interagir com o FAD, durante a catálise, ao invés da tripanotona diretamente.

Não existem informações sobre a **3GZ3** no CSA, de modo que a mesma análise não pôde ser feita para ela. Entretanto, como todos os aminoácidos incluídos na seleção são iguais a partir dos 7Å, e a identidade entre as sequências é também muito grande, decidimos que seria razoável assumir que a homologia correu bem para a **3GZ3** também.

Com isso realizamos a transferência dos ligantes (e íons) do sítio ativo dos moldes para seus respectivos modelos a fim de utilizá-los tanto como referência nos cálculos de energia livre como marcadores de sítio ativo para os cálculos de *docking*.

4.2. Ligantes Gerados com o Procedimento Automatizado

Na metodologia explicamos que para reduzir o problema do número muito grande de combinações entre as bases, selecionamos apenas os membros de cada base com melhor acessibilidade sintética e utilizamos seus grupos variáveis para gerar as combinações.

O Orotato foi dividido em duas regiões (Figura 54). A primeira teve 15 grupos e a segunda 6, com um total de 90 moléculas só para a DHDH. Como este número ficaria inviável, tomamos as quantidades que, respeitando as proporções de cada seção, gerariam quantidades tratáveis de moléculas. Para o Orotato, foram os 8 grupos com melhor acessibilidade sintética para a seção R1 e os 3 melhores na seção R2 (Figura 54).

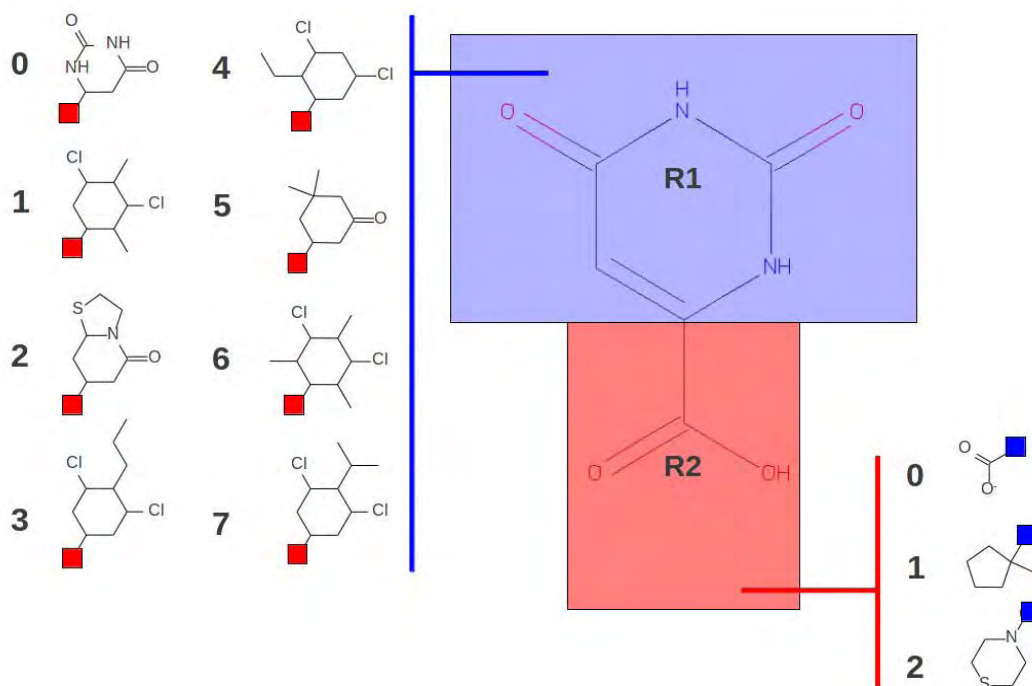


Figura 54: Grupos utilizados nos análogos do Orotato, separados por seção.

Cada grupo recebeu um código numérico de modo que a molécula **O32** seria a molécula resultante da substituição, no Orotato, do grupo 3 na posição R1 e do grupo 2 na posição R2. O código **O00** refere-se o próprio Orotato. Um código semelhante foi pensado para a Tripanotiona. Exceto que a Tripanotiona, por ser uma molécula grande, foi separada em três seções e as extremidades, como são simétricas, utilizaram os mesmo grupos (Figura 55).

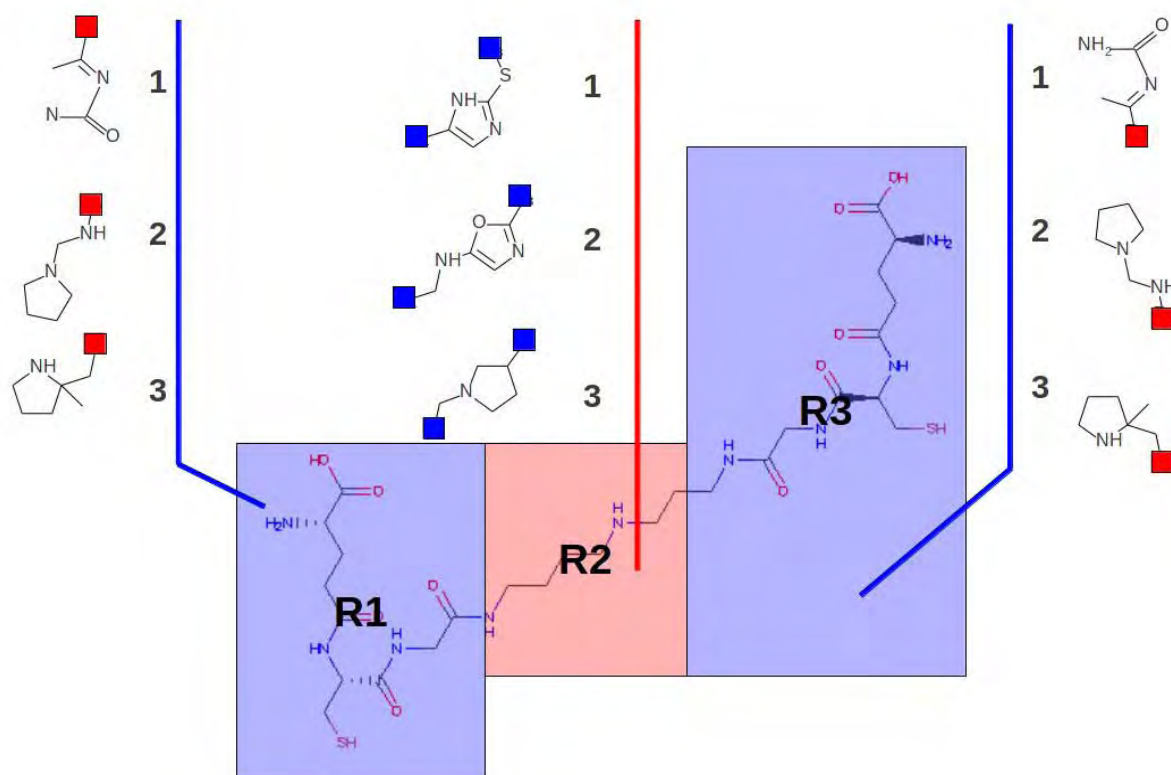


Figura 55: Grupos utilizados nos análogos da Tripanotiona, separados por seção.

Curiosamente, os fragmentos da tripanotiona, a molécula de entrada, não estavam entre os melhores quando ordenamos os membros da base segundo a acessibilidade sintética. Isto quer dizer que segundo o BROOD, moléculas contendo os fragmentos da tripanotiona não são fáceis de serem obtidos sinteticamente. Por este motivo, a numeração começa em 1 no código de combinações dos grupos da Tripanotiona, mas o código da própria tripanotiona é **T000**.

Para a FPPS, entretanto, não utilizamos este método de geração de ligantes. Como já mencionado anteriormente, a FPPS tem recebido atenção do nosso grupo de química medicinal há algum tempo, sobretudo como alvo em séries de adutos de Morita-Baylis-Hillman produzidos pela reação de acrilonitrila e/ou metil acrilato com aldeídos comerciais.

Em 2010, nosso grupo publicou dados de IC_{50} para 16 adutos de Morita-Baylis-Hillman contra *Leishmania chagasi* e *Leishmania amazonensis* (JUNIOR et al, 2010). Estes compostos estão mostrados na Figura 56.

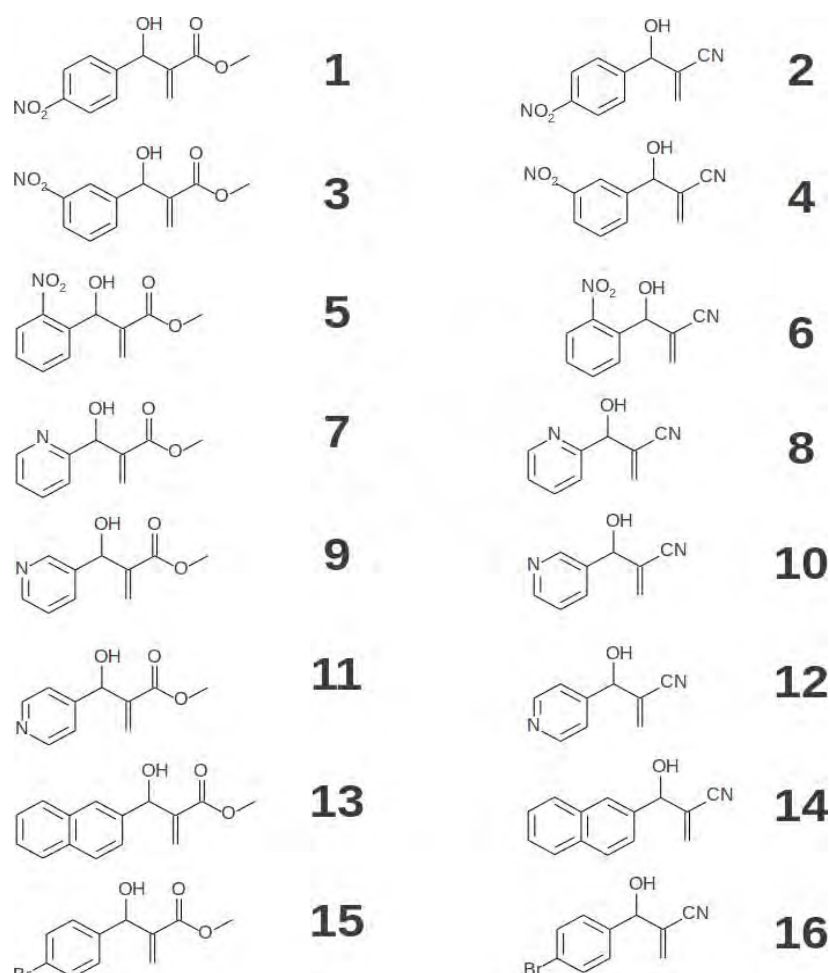


Figura 56: Adutos de Morita Baylis-Hillman.

Em virtude destes dados experimentais referentes a *Leishmania amazonensis* e de nossos estudos prévios apontando a FPPS como possível alvo de atuação de um destes adutos (SANDER et al, 2010), surgiu o interesse de utilizar estas moléculas, ao invés de moléculas geradas automaticamente pela metodologia utilizando o BROOD.

A numeração utilizada no código é a mesmo do artigo de referência, com os adutos tendo código de **M1** a **M16** sendo os ímpares nitrilas e os pares metil-ésteres. O Risedronato recebeu o código **M00**.

Ao todo foram 69 ligantes testados. Os 17 da FPPS, 24 da DHDH e 27 da TR. Cada um deles teve a variação da energia livre de interação com suas respectivas enzimas avaliada pelo escore do Autodock 4.2 e o MMGBSA sobre a estrutura minimizada ou sobre múltiplas trajetórias.

4.3. Redock dos Ligantes Cristalográficos

Para verificar a qualidade dos *docking* moleculares, realizamos o *redocking* dos ligantes cristalográficos e verificamos o quanto a pose produzida pelo *redocking* aproximou-se do posicionamento cristalográfico experimental.

Considerando que o *docking* não movimentava átomos do sítio ativo, é de se esperar que a coincidência seja alta, porque supomos que para aquela posição particular dos átomos do cristal, o ligante deve estar na melhor posição possível.

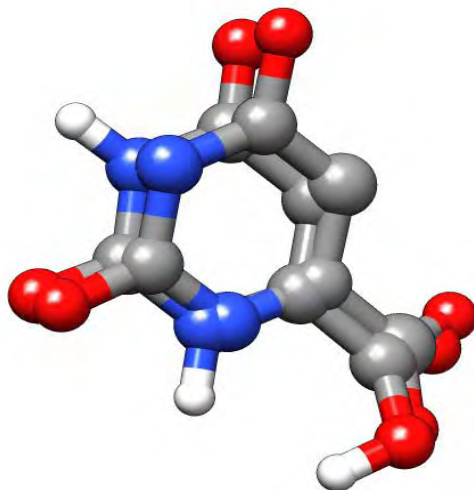


Figura 57: Redock do Orotato. RMSD 0.133

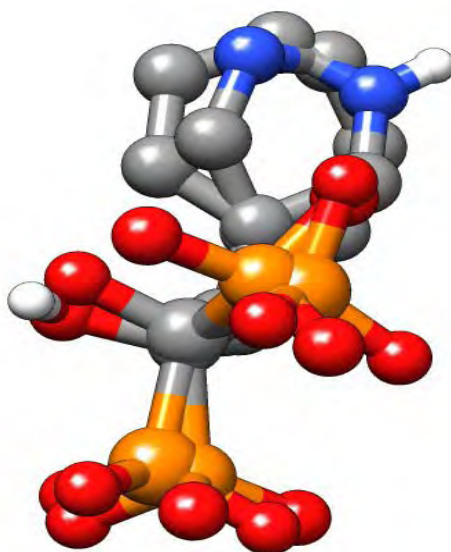


Figura 58: Redock do Risedronato. RMSD 0.312

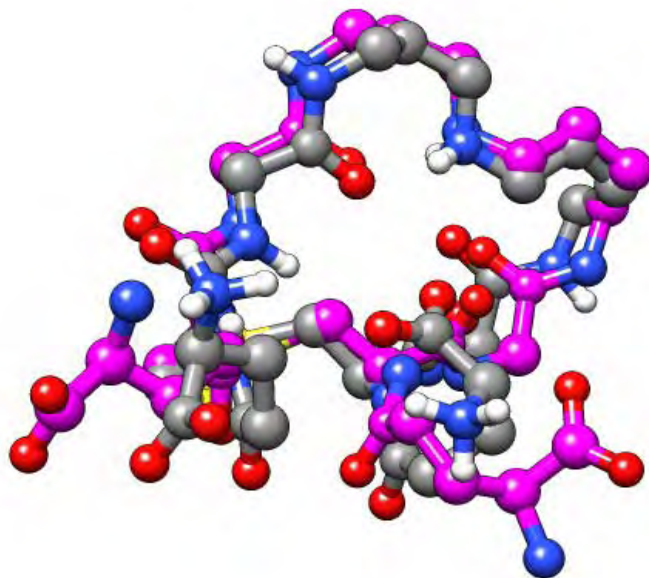


Figura 59: Redock da Tripanotiona. Cadeias diferentes mostradas em cores diferentes para propósitos de clareza. RMSD 0.450

De modo geral, o *docking* molecular saiu-se bem no *redocking* de todos os ligantes, com a possível exceção da tripanotiona. Talvez devido à grande quantidade de ângulos torcionais da molécula otimizados pelo algoritmo genético do Autodock, a solução não tenha convergido apropriadamente.

Aparentemente, o procedimento de *docking* escolhido reproduz bem a pose cristalográfica dos ligantes a menos que a quantidade de graus de liberdade comece a se tornar muito grande. Ainda assim, a orientação geral da tripanotiona no sítio ativo coincidiu com a pose cristalográfica e a região de maior discordância entre o *redock* e a pose cristalográfica foram as extremidades da molécula.

4.4. Sobre as Trajetórias de Dinâmica Molecular

Em seu artigo de 2009, descrevendo o método do MMGBSA de múltiplas trajetórias, Samuel Genheden e Ulf Ryde verificaram que mesmo uma dinâmica molecular longa de 10ns não era suficiente para amostrar adequadamente o espaço de fase e produzir estimativas da variação da energia livre que sejam estatisticamente convergentes.

Em seu artigo eles apontam que um tempo trinta vezes maior seria necessário para que as estimativas de energia livre fossem suficientemente confiáveis e reproduzíveis, com erro-padrão dentro da faixa estatisticamente aceitável. 300ns é um intervalo de tempo muito longo para ser simulado para propósitos gerais.

Levando em consideração que o problema é a amostragem ruim do espaço de fase, eles argumentaram, um conjunto de dinâmicas moleculares deveria convergir mais rapidamente. E de fato os estudos dos dois mostram que realizar estimativas de energia livre para várias trajetórias é computacionalmente mais factível e eficiente que uma única dinâmica longa.

Segundo os dados de Genheden e Ryde (2010), para atingir esta convergência estatística, o tempo de equilíbrio determina um intervalo de tempo mínimo a ser amostrado pelas dinâmicas de produção. Isto é incomum no sentido que, em geral, nas dinâmicas moleculares, o tempo de produção é relativamente independente do tempo de equilíbrio. Isto é, uma vez que o sistema esteja equilibrado, qualquer intervalo de produção pode ser amostrado.

Tanto o valor da energia livre quanto o desvio estatístico convergem à medida que os tempos de equilíbrio e produção aumentam, mas é possível observar, nos dados de Genheden e Ryde (2010), pontos onde até mesmo com equilíbrios considerados muito curtos (40ps ou 50ps, por exemplo) certa convergência estatística do erro e do valor da estimativa de energia livre podem ser obtidos, desde que uma trajetória de produção mais longa seja empregada, a uma taxa de amostragem de cerca de 1 passo a cada 5ps.

Nós utilizamos 50ps de equilíbrio o que, segundo o artigo de referência do método, exige uma trajetória de produção de, no mínimo, 250ps para garantir a convergência dos erros e das estimativas. No nosso trabalho, utilizamos 500ps de produção.

Esta dependência entre convergência e a duração relativa entre a equilíbrio e produção indica que uma propriedade importante no MMGBSA, mais que um longo tempo de equilíbrio, parece ser uma boa amostragem, de geometrias suficientemente descorrelacionadas entre si.

E o que é uma boa amostragem depende do tempo de equilíbrio. Com uma maior equilíbrio, menos cálculos de energia é preciso realizar para conseguir resultados convergentes, do contrário, o espaço de fase precisa ser sondado de forma mais completa. Aparentemente é mais importante, em um cálculo de MMGBSA, que o espaço de fase seja amostrado de forma ampla que um longo tempo de equilíbrio.

4.5. Avaliação dos Cálculos de Afinidade de Interação

A afinidade entre ligantes e alvos moleculares foi calculada segundo três métodos: escore do Autodock (AD), MMGBSA com múltiplas trajetórias (multi-MMGBSA) e o MMGBSA com uma única estrutura minimizada (min-MMGBSA). Como já citado antes, as estruturas minimizadas do min-MMGBSA foram produzidas por um cálculo de *docking* mais simples que o cálculo feito para produzir bons escores, utilizados nas estimativas de afinidade. Os resultados estão mostrados nas Tabelas 4, 5 e 6.

Tabela 4: Estimativas de afinidade dos ligantes da TR pelos três métodos.

	ΔG -multi	ΔG -min	AD
T111	-19,0152	-27,5638	-4,19
T112	-23,1372	-24,1496	-6,13
T113	-25,6612	-26,1287	-4,92
T121	-8,0542	-10,0559	-5,19
T122	-23,8795	-34,3863	-6,68
T123	-25,8208	-41,3027	-5,64
T131	-20,9369	-30,0623	-4,31
T132	-20,0634	-33,2026	-5,68
T133	-19,9333	-23,2951	-6,64
T211	-17,5263	-29,8861	-5,03
T212	-31,7005	-48,9209	-5,74
T213	-34,7521	-54,6230	-6,29
T221	-39,2026	-50,4960	-7,22
T222	-34,9487	-62,6378	-5,87
T223	-33,7704	-56,2869	-6,92
T231	-23,2109	-37,2517	-5,81
T232	-24,3304	-40,1954	-5,17
T233	-33,7497	-50,7771	-7,27
T311	-24,1332	-36,0955	-6,09
T312	-29,9347	-40,9492	-5,77
T313	-34,485	-43,9311	-6,4
T321	-36,3352	-45,9055	-6,61
T322	-43,3735	-56,8759	-7,5
T323	-34,3741	-48,5372	-7,83
T331	-27,7014	-37,8225	-6,18
T332	-24,8678	-41,6475	-5,67
T333	-27,4218	-41,3130	-6,74
T000	-35,3216	-41,4471	-6,29

Tabela 5: Estimativas de afinidade dos ligantes da FPPS pelos três métodos.

	ΔG -multi	ΔG -min	AD
M01	-36,1088	-43,2517	-5,8
M02	-36,9893	-36,7461	-5,92
M03	-52,8181	-44,6788	-6,04
M04	-36,0335	-34,7318	-6,6
M05	-40,8227	-39,7964	-6,49
M06	-32,9763	-34,8067	-6,45
M07	-37,3763	-31,7127	-4,13
M08	-28,6611	-25,1298	-4,09
M09	-40,0627	-33,4373	-4,2
M10	-32,0080	-25,8295	-4,82
M11	-32,0220	-31,0042	-4,68
M12	-32,7480	-25,8124	-4,67
M13	-50,4308	-39,0865	-5,94
M14	-44,7350	-37,3751	-5,88
M15	-47,2414	-36,484	-5,12
M16	-31,8276	-30,6342	-5,33
M00	-250,4593	-252,4803	-8,43

Tabela 6: Estimativas de afinidade dos ligantes da DHDH pelos três métodos.

	ΔG-multi	ΔG-min	AD
O01	-24,5685	-29,3333	-5,42
O02	-28,8427	-28,6421	-5,28
O10	-33,6849	-29,0273	-5,98
O11	-33,1428	-28,6999	-5,61
O12	-37,2339	-38,1488	-5,34
O20	-28,0353	-41,3192	-5,97
O21	-36,2777	-33,8107	-5,57
O22	-35,9978	-34,0555	-5,08
O30	-30,5036	-27,3894	-6,04
O31	-30,7803	-32,1776	-6,13
O32	-32,5957	-30,5235	-5,63
O40	-32,6660	-31,7358	-5,69
O41	-25,7686	-24,2138	-5,85
O42	-33,7327	-33,2205	-5,16
O50	-31,9548	-34,0666	-6,46
O51	-24,0542	-28,5398	-5,79
O52	-29,0646	-33,8450	-5,48
O60	-36,8676	-34,2073	-5,78
O61	-27,7078	-29,3835	-6,33
O62	-37,4800	-37,6486	-5,53
O70	-32,0013	-33,8761	-6,06
O71	-25,5655	-27,2719	-5,85
O72	-33,6554	-30,5483	-5,64
O00	-22,4070	-24,4126	-3,56

Para analisar estes conjuntos de valores, a principal técnica empregada foi a regressão linear pelo método dos mínimos quadrados. Este método consiste em encontrar a reta que está o menos distante possível de todos os pontos encontrando os parâmetros da reta que minimizam a soma dos quadrados das distâncias de cada ponto à reta.

Utilizamos o coeficiente R^2 para quantificar a correlação linear entre os dados produzidos pelo multi-MMGBSA com os dados produzidos pelo min-MMGBSA e o Autodock. Estas correlações foram feitas para diferentes conjuntos de dados com o objetivo de avaliar características diversas do poder preditivo das técnicas comparadas.

4.5.1. Correlação nas Estimativas de Afinidade

Cada um dos três métodos fornecem valores numéricos que representam suas estimativas da afinidade que cada ligante tem com sua enzima. Ao correlacionar estes valores queremos saber se um método concorda com outro sobre qual molécula possui melhor afinidade e qual a magnitude desta afinidade. Os resultados das correlações estão mostrados abaixo nas Figuras 60, 61 e 63.

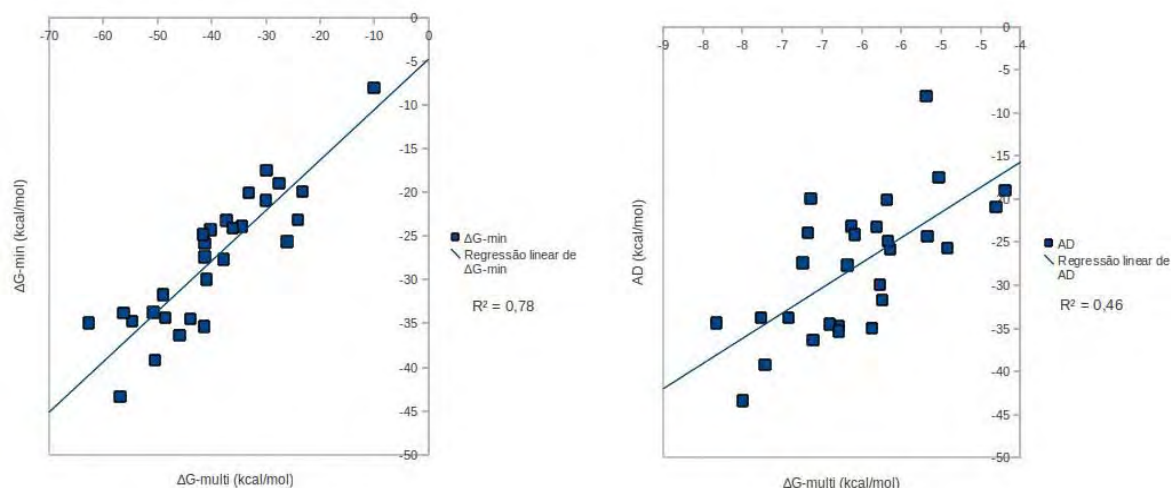


Figura 60: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na TR.

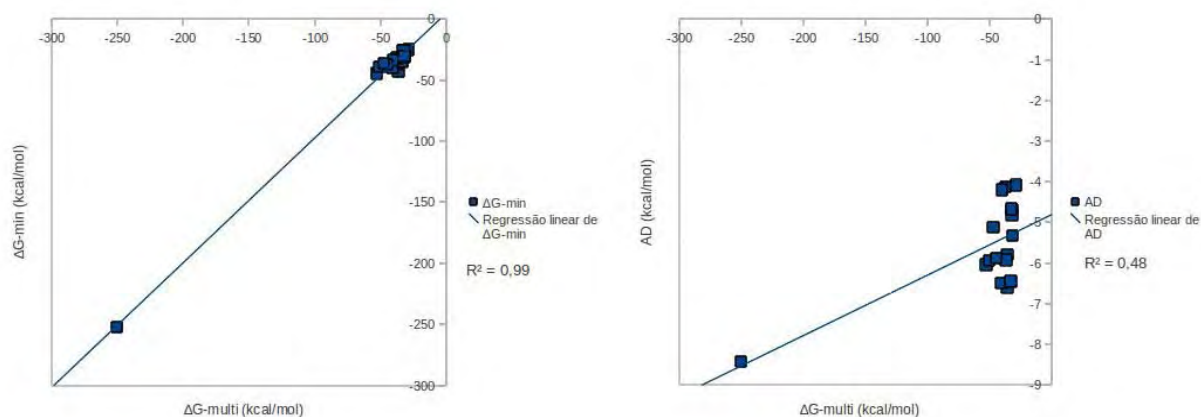


Figura 61: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na FPPS.

No gráfico podemos ver que, em ambos os casos, existe um ponto que se encontra muito mais baixo que os outros. Nisso, todos os três métodos concordam. Este ponto é o **M00**, o risedronato, um inibidor da FPPS.

Em particular, os valores estimados pelo multi-MMGBSA e pelo min-MMGBSA são muito próximos um do outro, levando o R^2 para um valor muito próximo de 1,0. No entanto, pelo gráfico nos vemos que a incidência dos pontos sobre a reta não sugere um valor tão próximo do perfeito, assim.

O *docking* também acertou ao colocar o valor do **M00** mais abaixo que o resto dos ligantes da série. A correlação entre o *docking* e o multi-MMGBSA para o Risedronato também é muito alta.

É importante notar, entretanto, que se o **M00** for tão melhor que os outros ligantes em interagir com a FPPS, mesmo um método capaz de discriminar apenas moléculas com afinidade muito distintas seria capaz de posicioná-lo adequadamente no gráfico e conseguir uma boa correlação com o multi-MMGBSA.

Assim, as correlações seriam todos aumentados artificialmente para ambos os porque ambos

automaticamente seriam capazes de colocar o risedronato em uma posição razoável em relação aos outros ligantes, em virtude de sua afinidade tão distinta. Para ver realmente qual a correlação entre os métodos é preciso remover a contribuição do **M00** aos coeficientes de correlação (Figura 62).

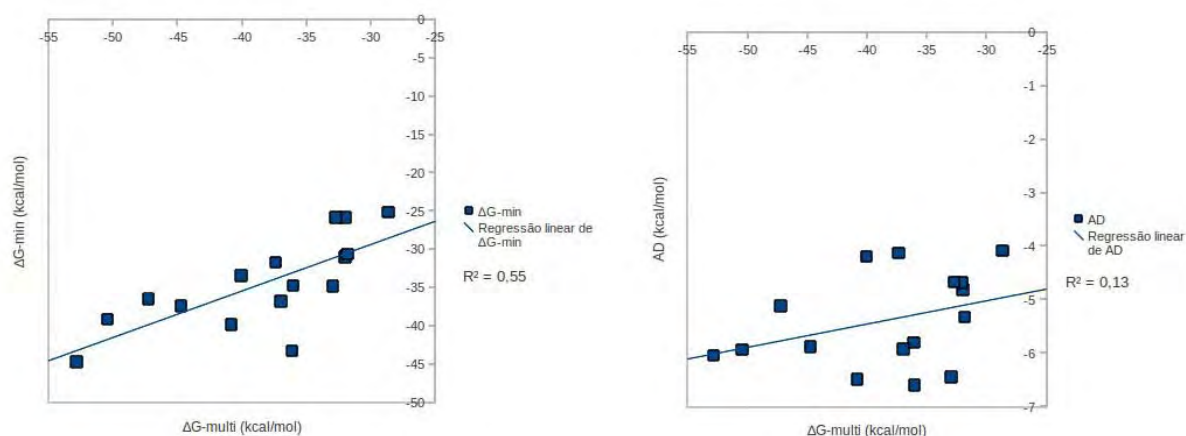


Figura 62: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na FPPS sem o risedronato.

Quando retiramos o **M00** percebemos a real correlação entre os métodos. Mesmo com esta significativa redução nos índices, o min-MMGBSA produz estimativas de afinidade mais correlacionadas com o multi-MMGBSA que os escores do Autodock, que praticamente não apresentam correlação significativa com o multi-MMGBSA.

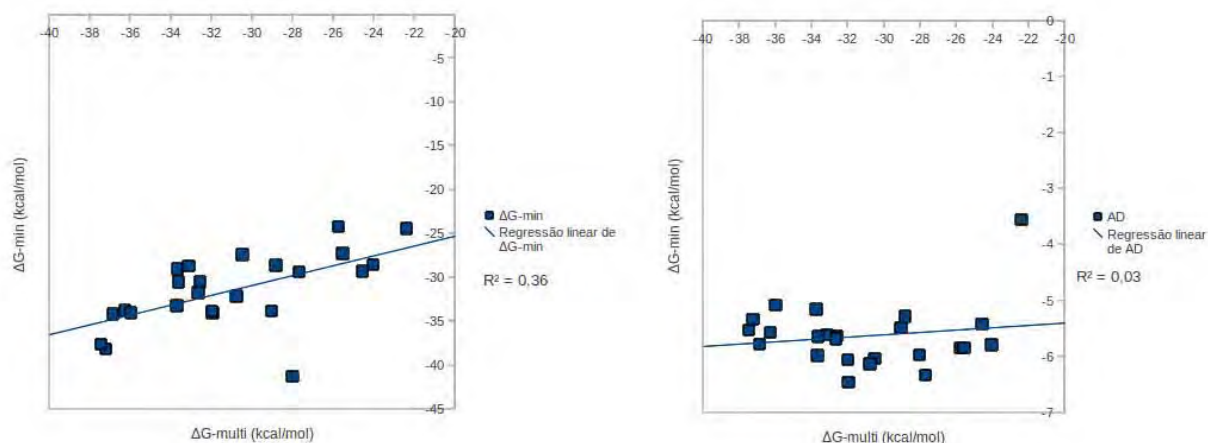


Figura 63: Correlação linear entre os dados de min-MMGBSA e Autodock, da esquerda para a direita, com os dados de multi-MMGBSA na DHDH.

Nos cálculos de correlação da DHDH ocorre um fenômeno inverso ao que ocorreu com a FPPS. Um dos ligantes possui uma distância da reta de regressão que é anormalmente grande (ou seja, os métodos discordam de forma mais enfática que o normal em relação à série) e isto penaliza os coeficientes de correlação, que se não incluíssem este ponto, seriam melhores.

No entanto, diferentemente do **M00**, o min-MMGBSA e o Autodock não concordam sobre

qual molécula é a anomalia. No gráfico do min-MMGBSA, o ponto longe da reta é o **O20**. No Autodock é o **O00**, o próprio substrato natural da DHDH.

Ao contrário do que ocorre na FPPS, não há uma razão bem definida para justificar a remoção destas moléculas do conjunto. Aparentemente, esta é uma discordância genuína entre os métodos e representa informação sobre as relações que existem entre os métodos.

Uma possível razão para o multi-MMGBSA e o min-MMGBSA discordarem seria o fato de haver mais de um mínimo de energia próximos e separados por barreiras de energia pequenas o suficiente para serem atravessadas pelas trajetórias de dinâmica molecular, de forma que a amostragem do multi-MMGBSA provavelmente incluiria estruturas próximas de ambos os mínimos, violando a hipótese inicial do min-MMGBSA de que a população de configurações amostrada provavelmente representa geometrias próximas de um único ponto mínimo de energia.

Para confirmar esta hipótese, tomamos as trajetórias utilizadas pelo multi-MMGBSA e amostramos 10 passos de cada uma das dez trajetórias. De forma geral, as estruturas estão organizadas em torno de dois “padrões” de estrutura. Um destes padrões, levados em consideração pelo multi-MMGBSA, forma 4 ligações de hidrogênio com o sítio ativo. O outro padrão forma 6 ligações de hidrogênio.

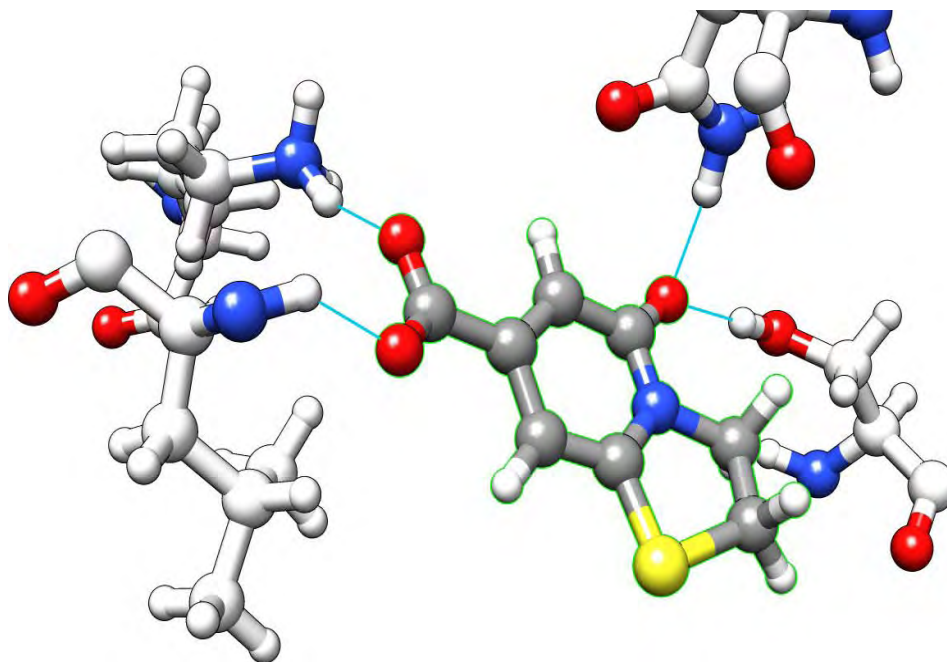


Figura 64: Padrão de quatro ligações de hidrogênio. Sítio ativo em branco, ligações de hidrogênio em ciano.

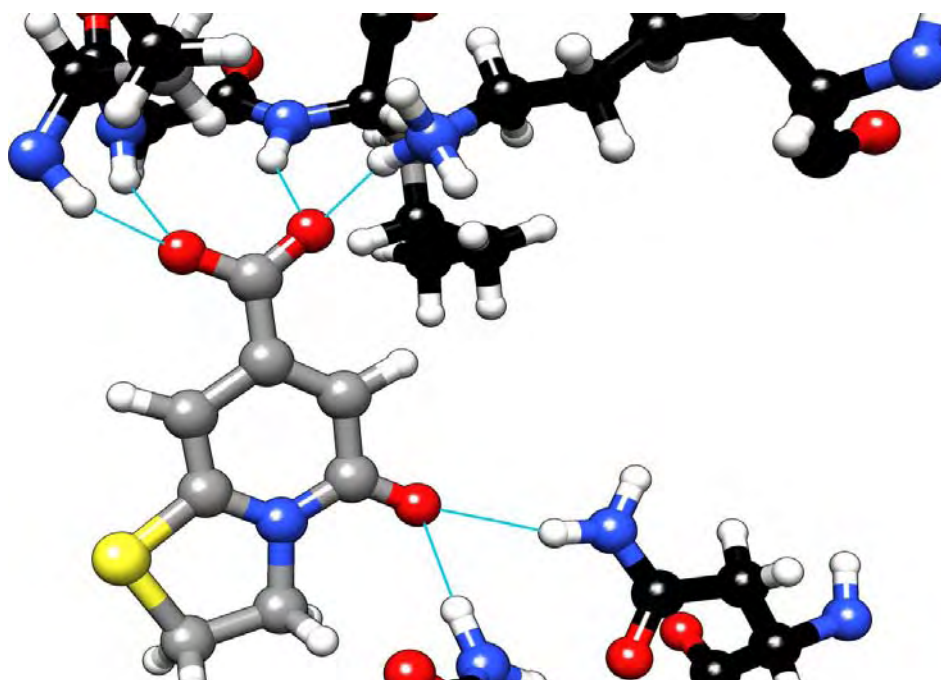


Figura 65: Padrão de seis ligações de hidrogênio. Sítio ativo em preto, ligações de hidrogênio em ciano.

A minimização também consegue encontrar uma estrutura semelhante ao padrão número dois, que também forma seis ligações de hidrogênio. Esta diversidade de geometrias em torno das quais as trajetórias amostram, ao invés de apenas uma, pode ser o motivo da discordância entre o multi-MMGBSA e o min-MMGBSA no cálculo da estimativa da variação da energia livre.

Explicar a discordância do Autodock é mais complicado. Poderíamos sugerir que o algoritmo genético não foi capaz de realizar uma amostragem refinada o suficiente para conseguir encontrar a pose natural. Entretanto, o *redock* mostra que os resultados do *docking* foram apreciáveis, do ponto de vista geométrico.

Além do mais, o Orotato não tem muitos graus de liberdade torcionais (apenas um, na verdade), otimizado pelo *docking*, de modo que não seria plausível supor que esta discordância é causa estritamente pela não convergência do *docking* molecular.

Existe ainda a possibilidade de que existem discrepâncias muito grandes entre os parâmetros utilizados pelo Autodock e pelo campo de força empregado nos cálculos de multi-MMGBSA. No entanto, o Orotato é formado por átomos que aparecem nos outros ligantes, onde não há uma discordância tão expressiva entre os dois métodos.

Uma outra hipótese seria a de que a interação do Orotato com a DHDH dependeria de consideráveis mudanças conformacionais da enzima, ao acomodá-lo no sítio ativo. Como o *docking* não leva em consideração a variação nestes graus de liberdade ao calcular a afinidade, mas o multi-MMGBSA leva, esta seria uma possível fonte de discordância entre os dois métodos.

Em todos os gráficos, o min-MMGBSA obteve um coeficiente de correlação melhor que o Autodock, entretanto, tanto nos gráficos da FPPS quanto nos gráficos da DHDH houveram alguns pontos bastante discrepantes em relação ao resto dos dados.

4.5.2. Correlação nos Ordenamentos

Na seção anterior nós analisamos as correlações existente entre os métodos de se realizar estimativas da afinidade de interação entre os ligantes e seus respectivos alvos moleculares. Estes resultados aparecem na Tabela 7. Os valores da FPPS estão apresentados sem o **M00**.

Tabela 7: Coeficiente de correlação linear (R^2) dos valores de ΔG -min e AD com os valores do ΔG -multi.

Enzima	R^2 (ΔG -multi x ΔG -min)	R^2 (ΔG -multi x AD)
DHDH	0,33	0,03
TR	0,78	0,46
FPPS	0,55	0,13

Pelos valores da Tabela 7 e verificando os gráficos da seção anterior percebemos que não parece haver uma preferência de nenhum dos métodos de teste por coincidir seus valores com o método de referência, o que sugere que nenhum deles concorda com o multi-MMGBSA acerca valores de afinidade para os ligantes ou sobre a magnitude destes valores.

Ainda que, de modo geral, nenhum dos valores sejam fortemente correlacionados com o método de referência, com exceção do min-MMGBSA dos ligantes da TR, é visível que a correlação do *docking* sempre é pior.

A acentuada variação na correlação linear dos valores de min-MMGBSA em relação ao método de referência multi-MMGBSA demonstra que esta técnica aproximada, assim como o *docking* molecular não possui uma exatidão muito previsível. A variação de enzima para enzima é marcante. Esta variação na predição de valores de afinidade é documentada também em estudos utilizando sistemas com dados experimentais disponíveis para comparação (KUHN et al, 2004).

Entretanto, quando falamos de triagens virtuais estamos implicitamente considerando que existe uma grande quantidade de ligantes e que deseja-se ver tendências ao invés de valores exatos. É necessário que o procedimento agrupe todo o banco de dados em “famílias” de moléculas e que aponte qual destas famílias é mais favorável.

Uma forma de reduzir o conjunto de candidatos a fármacos removendo do conjunto aquelas moléculas que têm maior probabilidade de não serem satisfatórias. Este é um procedimento comparativo. Não é preciso estimar a energia livre para fazer isto, é necessário apenas que o método tenha uma forma de comparar duas moléculas e decidir qual delas é mais satisfatória.

Para ilustrar este ponto, considere o seguinte procedimento hipotético: imagine que um programa de triagem virtual qualquer monta uma abstração de um sítio ativo (um *grid* ou algo semelhante) e testa vários ligantes neste sítio ativo. Para cada ligante ele identifica os três melhores pontos de conexão com o sítio ativo e dá uma nota de 0 a 255. O resultado final do resultado é um conjunto de três valores indo de 0 a 255 para cada ligante.

Estes valores podem ser pensados como se fossem cores RGB. No padrão RGB, as cores são expressas como sendo trios de valores. O primeiro valor representa a quantidade de vermelho, o segundo o verde e o terceiro azul. *Red, Green and Blue*. Os valores vão de 0 a 255 de modo que <255,255,255> representa o branco e <0,0,0> representa o preto.

Poderíamos, em princípio, parametrizar um algoritmo para realizar um procedimento desta natureza, de uma forma semelhante ao realizado pelo *docking*. Utilizando vários complexos e modelos de interações para parametrizar funções de escore arbitrárias. Se as funções fossem bem elaborada, o algoritmo colocaria os melhores ligantes no branco e os piores no preto.

Este algoritmo poderia, em tese, ser utilizado para realizar uma triagem virtual. Afinal, ele separa os ligantes em grupos favoráveis (os brancos ou próximos do branco) e desfavoráveis (os pretos ou próximos dos pretos). A única coisa para a qual este método não é apropriado é para dar uma estimativa, qualquer que seja, de “energia”, mas ele não precisa fazer exatamente isto para cumprir seu papel.

A falta de correlação linear dos valores de docking com os valores do multi-MMGBSA só nos revelam o quão inapropriado ele é para calcular a variação da energia livre, mas não diz muito sobre sua capacidade em varrer uma biblioteca de ligantes. O mesmo raciocínio, aliás, aplica-se ao min-MMGBSA. Se o que se deseja saber é se o min-MMGBSA é melhor que o *docking* em um procedimento de triagem virtual, então precisamos proceder de outra forma.

Para avaliar a capacidade dos métodos em discriminar entre ligantes favoráveis e desfavoráveis, transformamos os valores absolutos (kcal/mol) em valores comparativos. Estes dados foram produzidos atribuindo “créditos” arbitrários às moléculas. Este procedimento serviu para produzir dados que não tivessem correlação direta com os valores absolutos de afinidade. Um tipo de dado parecido com as cores discutidas anteriormente no procedimento hipotético.

Nós comparamos cada ligantes de um grupo com todos os outros ligantes daquele mesmo grupo. Cada vez que o ligante em questão fosse mais favorável que o outro com o qual estava sendo comparado, receberia um crédito. Assim, estes créditos não teriam significado físico direto, seriam números adimensionais, mas representariam o quanto um método decidiu quais moléculas são melhores ligantes que quais outras moléculas.

O resultado final foi uma distribuição de créditos diferente para cada método. As tabelas 8,9 e 10 mostram as distribuições de créditos obtidas ao comparar os ligantes de todas as enzimas segundo ΔG -multi, ΔG -min e o escore do Autodock.

Tabela 8: Distribuição dos créditos feita por cada método para cada ligante da DHDH.

	ΔG -multi	ΔG -min	AD
O01	2	8	5
O02	7	5	3
O10	17	7	18
O11	15	6	9
O12	22	22	4
O20	6	23	17
O21	20	15	8
O22	19	18	1
O30	9	3	19
O31	10	13	21
O32	13	10	10
O40	14	12	12
O41	4	0	16
O42	18	14	2
O50	11	19	23
O51	1	4	14
O52	8	16	6
O60	21	20	13
O61	5	9	22
O62	23	21	7
O70	12	17	20
O71	3	2	16
O72	16	11	11
O99	0	1	0

Tabela 9: Distribuição dos créditos feita por cada método para cada ligante da TR.

	$\Delta\Delta G\text{-multi}$	$\Delta\Delta G\text{-min}$	ΔAD
T111	2	4	0
T112	6	2	14
T113	12	3	2
T121	0	0	5
T122	8	8	21
T123	13	14	6
T131	5	6	1
T132	4	7	8
T133	3	1	20
T211	1	5	3
T212	17	21	9
T213	22	24	17
T221	26	22	24
T222	23	27	12
T223	19	25	23
T231	7	10	11
T232	10	12	4
T233	18	23	25
T311	9	9	13
T312	16	13	10
T313	21	18	18
T321	25	19	19
T322	27	26	26
T323	20	20	27
T331	15	11	15
T332	11	17	7
T333	14	15	22
T000	24	16	17

Tabela 10: Distribuição dos créditos feita por cada método para cada ligante da FPPS.

	$\Delta\Delta G$ -multi	$\Delta\Delta G$ -min	ΔAD
M01	7	14	8
M02	8	10	10
M03	15	15	12
M04	6	7	15
M05	11	13	14
M06	5	8	13
M07	9	5	1
M08	0	0	0
M09	10	6	2
M10	2	2	5
M11	3	4	4
M12	4	1	3
M13	14	12	11
M14	12	11	9
M15	13	9	6
M16	1	3	7
M00	16	16	16

A correlação entre as distribuições de créditos obtidas pelo ΔG -multi com as distribuições obtidas pelos outros métodos mostram a capacidade do min-MMGBSA e do Autodock de reproduzir o ordenamento do multi-MMGBSA.

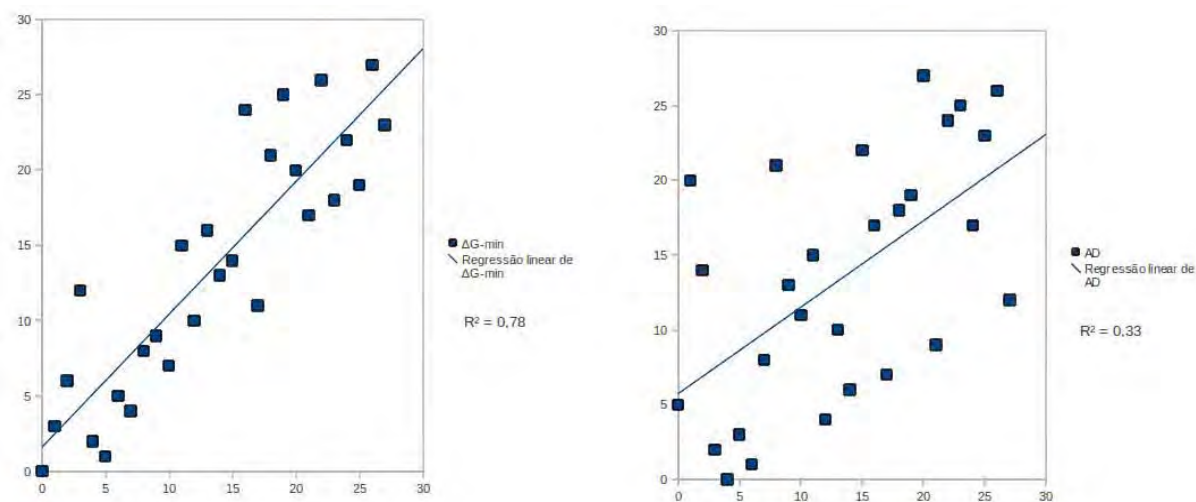


Figura 66: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da TR.

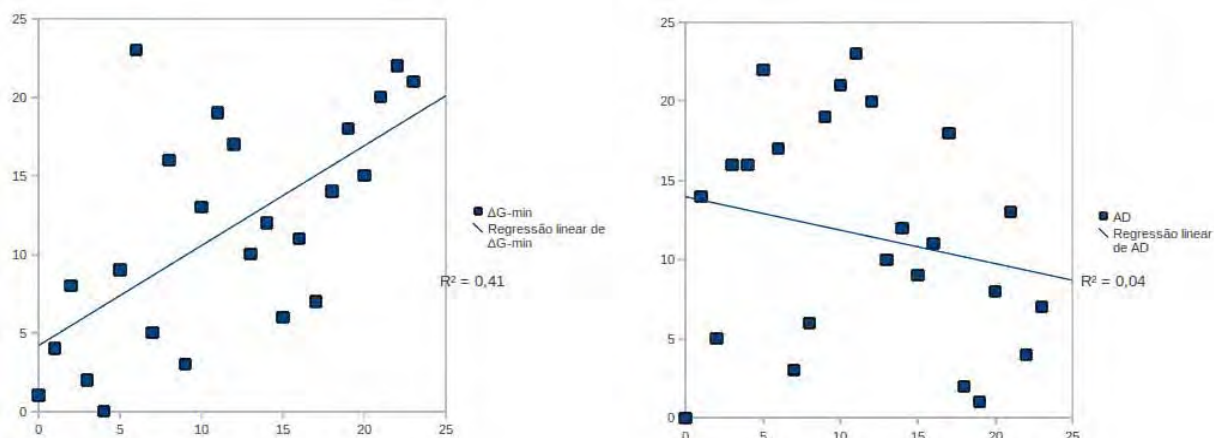


Figura 67: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da DHDH.

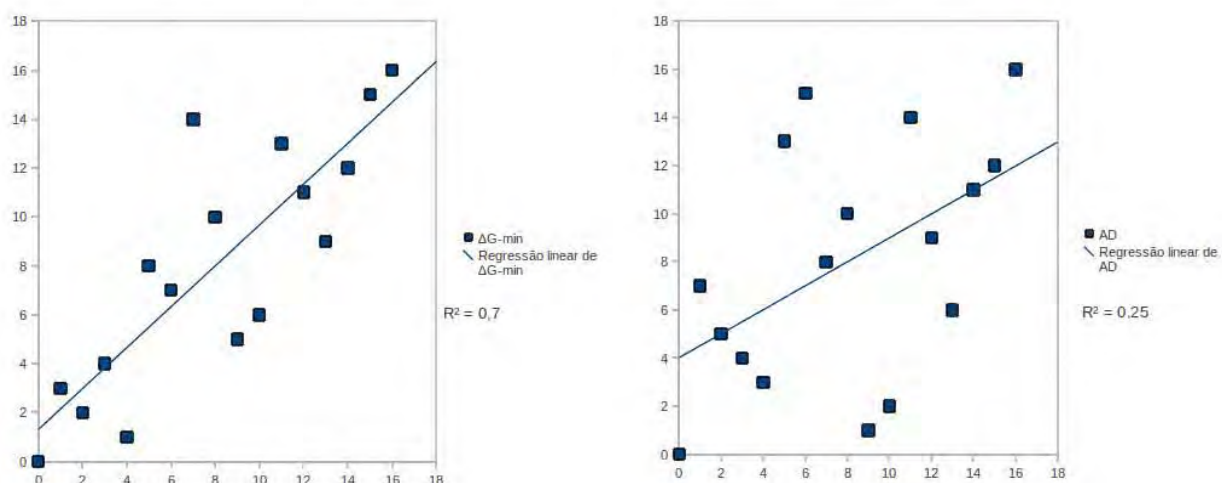


Figura 68: Correlação linear das distribuições de créditos produzidas pelo min-MMGBSA e Autodock, da esquerda para direita, com a distribuição produzida pelo multi-MMGBSA. Dados da FPPS.

Ao avaliar a capacidade de ordenamento dos métodos, ao invés de sua capacidade calcular valores de afinidade, percebemos que os coeficientes de correlação aumentam, de modo geral. Ainda assim, no entanto, os valores obtidos pelo min-MMGBSA correlacionam-se melhor com os valores do multi-MMGBSA.

Para selecionar, dos grupos de moléculas especificados, aquelas mais promissoras segundo as estimativas de afinidade, ordenamos as moléculas segundo a distribuição de créditos de cada método. Os dados estão mostrados nas tabelas 11, 12 e 13 para TR, DHDH e FPPS respectivamente.

Tabela 11: Ordenamento dos ligantes da TR seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.

ΔG -multi	ΔG -min	AD
T222	T322	T323
T322	T221	T322
T223	T321	T233
T213	T000	T221
T233	T222	T223
T221	T213	T333
T212	T313	T122
T323	T323	T133
T321	T223	T321
T313	T233	T313
T332	T212	T213
T000	T312	T000
T333	T331	T331
T123	T333	T112
T312	T123	T311
T232	T113	T222
T331	T332	T231
T231	T232	T312
T311	T311	T212
T122	T122	T132
T132	T231	T332
T131	T112	T123
T211	T131	T121
T111	T132	T232
T113	T133	T211
T112	T111	T113
T133	T211	T131
T121	T121	T111

Tabela 12: Ordenamento dos ligantes da DHDH seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.

ΔG -multi	ΔG -min	AD
O41	O42	O30
O22	O22	O61
O20	O41	O40
O10	O20	O71
O72	O30	O11
O31	O72	O60
O60	O71	O42
O51	O02	O50
O21	O10	O01
O70	O31	O52
O32	O40	O20
O71	O70	O70
O30	O51	O51
O40	O32	O32
O11	O61	O21
O02	O62	O10
O12	O60	O41
O42	O21	O02
O61	O12	O62
O50	O52	O22
O01	O11	O12
O62	O01	O31
O52	O00	O72
O00	O50	O00

Quando utilizamos a avaliação comparativa, os métodos em geral se saem melhores. O **O20**, por exemplo, que era um ponto discrepante entre o multi-MMGBSA e o min-MMGBSA aparece em colocações muito próximas ao ser ordenado por cada método, ocupando a terceira posição na escala do multi-MMGBSA e a quarta posição na escala do min-MMGBSA. O *docking*, por sua vez, coloca o ligante natural, o **O00**, que era seu ponto discrepante, exatamente na mesma posição que o multi-MMGBSA.

Tabela 13: Ordenamento dos ligantes da FPPS seguindo as distribuições de créditos obtidas para cada método pelo procedimento de comparação.

ΔG -multi	ΔG -min	AD
M00	M00	M00
M03	M03	M04
M13	M01	M05
M15	M05	M06
M14	M13	M03
M05	M14	M13
M09	M02	M02
M07	M15	M14
M02	M06	M01
M01	M04	M16
M04	M09	M15
M06	M07	M10
M12	M11	M11
M11	M16	M12
M10	M10	M09
M16	M12	M07
M08	M08	M08

Em seguida, tomamos aquelas moléculas que estivessem no 1/3 superior do ordenamento, aquelas que deveriam ser as melhores segundo cada método, e comparamos quais, dentre as moléculas do multi-MMGBSA, estavam também entre as melhores min-MMGBSA e do *docking*. Estes dados estão mostrados nas tabelas 14, 15 e 16.

Tabela 14: 1/3 superior da tabela de ordenamento da TR. Coincidências com multi-MMGBSA estão em negrito.

ΔG -multi	ΔG -min	AD
T222	T322	T323
T322	T221	T322
T223	T321	T233
T213	T000	T221
T233	T222	T223
T221	T213	T333
T212	T313	T122
T323	T323	T133
T321	T223	T321

Se tomássemos todos os coincidentes do min-MMGBSA ou do Autodock e recalculássemos suas afinidades utilizando o multi-MMGBSA, o ordenamento resultante colocaria o **T322** no topo da lista. Nisto, tanto o min-MMGBSA quanto o multi-MMGBSA concordam.

Tabela 15: 1/3 superior da tabela de ordenamento da FPPS. Coincidências com o multi-MMGBSA estão em negrito.

ΔG -multi	ΔG -min	AD
M00	M00	M00
M03	M03	M04
M13	M01	M05
M15	M05	M06
M14	M13	M03
M05	M14	M13

Nos ligantes da FPPS, tanto o min-MMGBSA quanto o Autodock concordam com o multi-MMGBSA em incluir o **M03** entre os melhores. Um reordenamento, segundo o multi-MMGBSA das coincidências nos dois métodos apontaria o **M03** como o ligante mais favorável.

Tabela 16: 1/3 superior da tabela de ordenamento da DHDH. Coincidências com o multi-MMGBSA estão em negrito.

ΔG -multi	ΔG -min	AD
O41	O42	O30
O22	O22	O61
O20	O41	O40
O10	O20	O71
O72	O30	O11
O31	O72	O60
O60	O71	O42
O51	O02	O50

Para os ligantes da DHDH, entretanto, os resultados do reordenamento das coincidências não produz resultados iguais para quaisquer dos dois métodos. O Autodock, entretanto, só aponta uma coincidência, o **O60**. O **O22**, por outro lado, aparece no ordenamento do multi-MGBSA, e nas coincidências entre min-MMGBSA e multi-MMGBSA.

O **O22** é também a molécula coincidente entre min-MMGBSA e multi-MMGBSA de melhor posição no ordenamento feito com o multi-MMGBSA, sugerindo que ao selecionar as

coincidências do min-MMGBSA como conjunto de testes para métodos mais exatos, existe uma grande probabilidade de uma boa molécula passar pelo crivo.

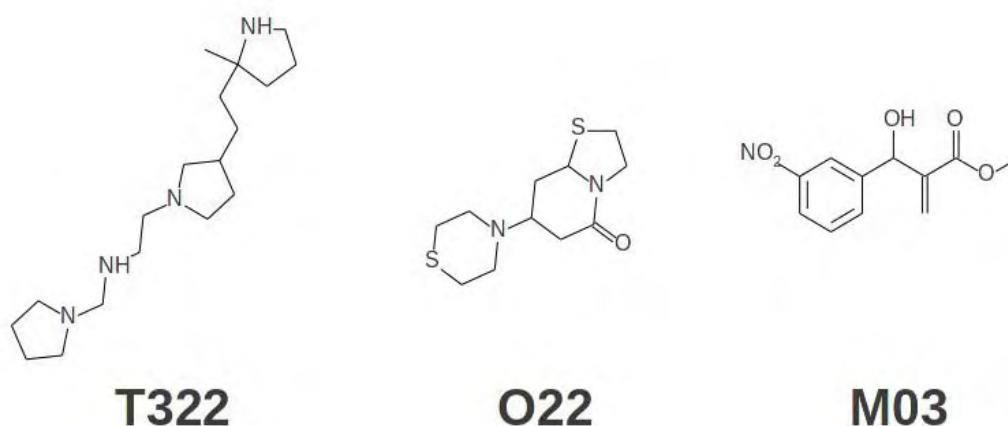


Figura 69: Moléculas coincidentes selecionadas pelo min-MMGBSA.

Ao compararmos com o multi-MMGBSA, o ordenamento do min-MMGBSA saiu-se melhor que o do Autodock. O min-MMGBSA também foi melhor que o Autodock para encontrar moléculas classificadas pelo multi-MMGBSA como sendo bons ligantes.

Para produzir um complexo a ser avaliado pelo min-MMGBSA, no entanto, é preciso realizar um *docking* molecular preliminar. Neste trabalho, este *docking* preliminar para montar os complexos a serem utilizados tanto no min-MMGBSA quanto no multi-MMGBSA foi mais simples e rápido que o *docking* realizado para estimar afinidade. O objetivo destes *dockings* preliminares foi apenas posicionar o ligante no sítio ativo.

Mesmo partindo de um complexo produzido por um *docking* mais simples e rápido que aquele utilizado para estimar a afinidade, o min-MMGBSA conseguiu resultados comparáveis, senão melhores. Utilizar o min-MMGBSA pode ser uma vantagem, se a economia de tempo for comparável à do *docking*.

4.5.3. Avaliação do *Docking* Rígido na Criação dos Complexos

Para obter os valores de min-MMGBSA, tivemos que construir os complexos solvatados, minimizar suas geometrias e efetuar os cálculos de min-MMGBSA. Os complexos foram obtidos pelo procedimento descrito anteriormente e levaram cerca de 10 minutos. cálculos de min-MMGBSA duraram menos de 10 segundos cada um. O procedimento de minimização, entretanto, foi o limitante. Com cinco mil passos de minimização, o procedimento levou cerca de 13 minutos nos processadores das placas gráficas.

O tempo médio de cálculo do Autodock para obter os dados de afinidade foi cerca de 20 minutos. Somando o tempo total de procedimento, o *docking* ainda é mais rápido que o min-

MMGBSA em cerca 3 minutos por molécula. Se considerarmos todas as 69 moléculas, são cerca de 3 horas e 27 minutos de economia da parte do *docking*. Para um conjunto maior de moléculas este intervalo de tempo provavelmente seria maior.

Para conseguir diminuir esta diferença de tempo, verificamos se um *docking* rígido seguido de uma minimização seria capaz de colocar a molécula de tripanotiona na orientação correta no sítio ativo. Escolhemos a tripanotiona porque dos ligantes que utilizamos é aquela que tem maior quantidade de graus de liberdade, supostamente acrescentando mais dificuldade ao sucesso do *docking* rígido. Além disso, a tripanotiona possui estrutura cristalográfica determinada experimentalmente, o que quer dizer que podemos comparar a orientação do resultado do *docking* rígido com a orientação da tripanotiona na cristalografia.

O *docking* rígido foi realizado em populações de 100 indivíduos. Dez cálculos foram feitos por 27000 geração, com taxas de mutação e cruzamento iguais a 0.02 e 0.8 respectivamente. Das dez estruturas produzidas, aquela de menor energia foi tomada para ser minimizada. O complexo contendo a tripanotiona posicionada pelo *docking* rígido foi solvatado e fizemos três minimizações de 1000, 3000 e 5000 passos.

Após as minimizações, comparamos a geometria da tripanotiona no sítio ativo para ver se a orientação da molécula estava correta em relação à orientação cristalográfica. A orientação é particularmente importante porque seria muito difícil a minimização rotacionar a molécula toda no sítio ativo, em busca da melhor geometria. Abaixo, nas Figura 70, 71 e 72 estão mostradas as superposições das tripanotionas minimizadas em relação à cristalográfica.

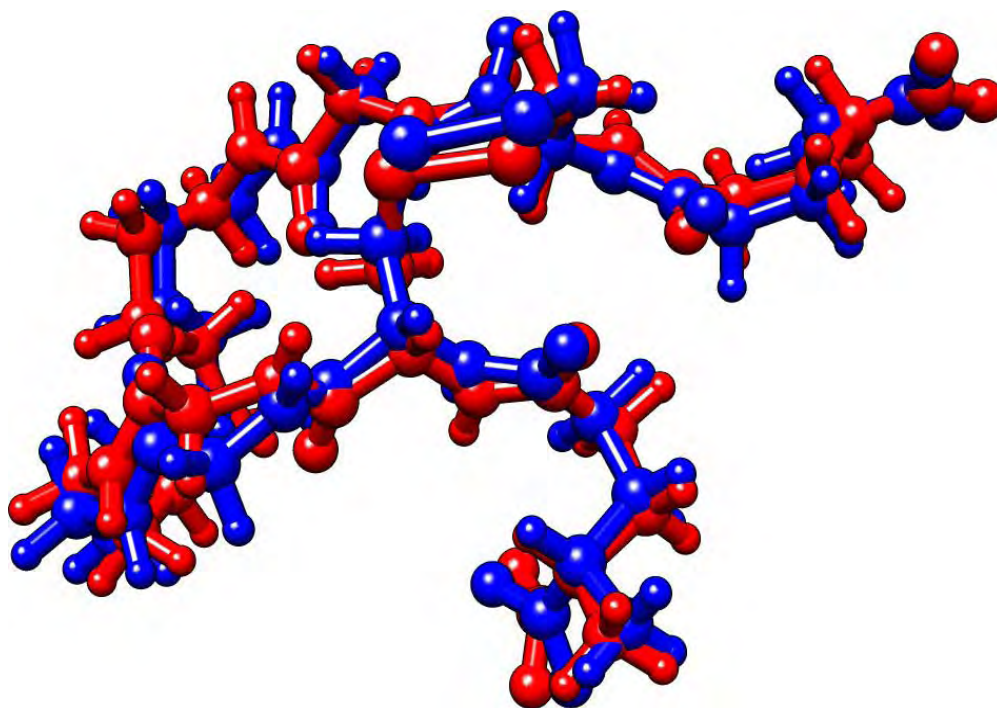


Figura 70: Superposição da tripanotiona cristalográfica (azul) com a tripanotiona posicionada com o *docking* rígido e minimizada por 1000 passos (vermelho).

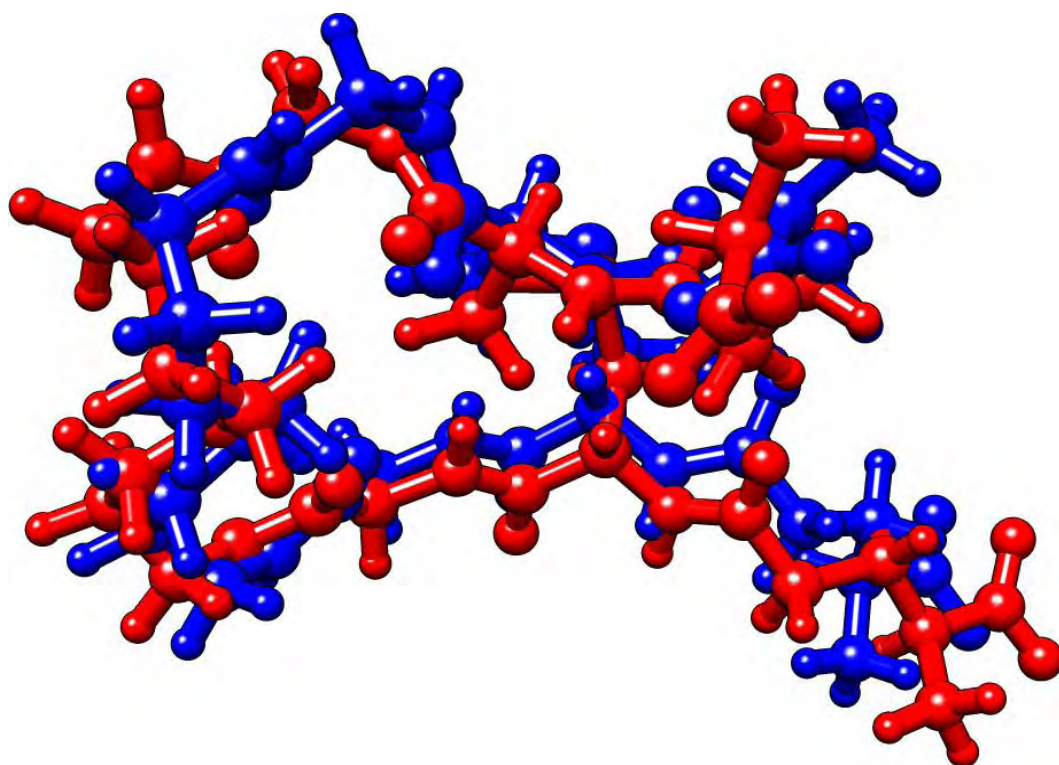


Figura 71: Superposição da tripanotiona cristalográfica (azul) com a tripanotiona posicionada com o docking rígido e minimizada por 3000 passos (vermelho).

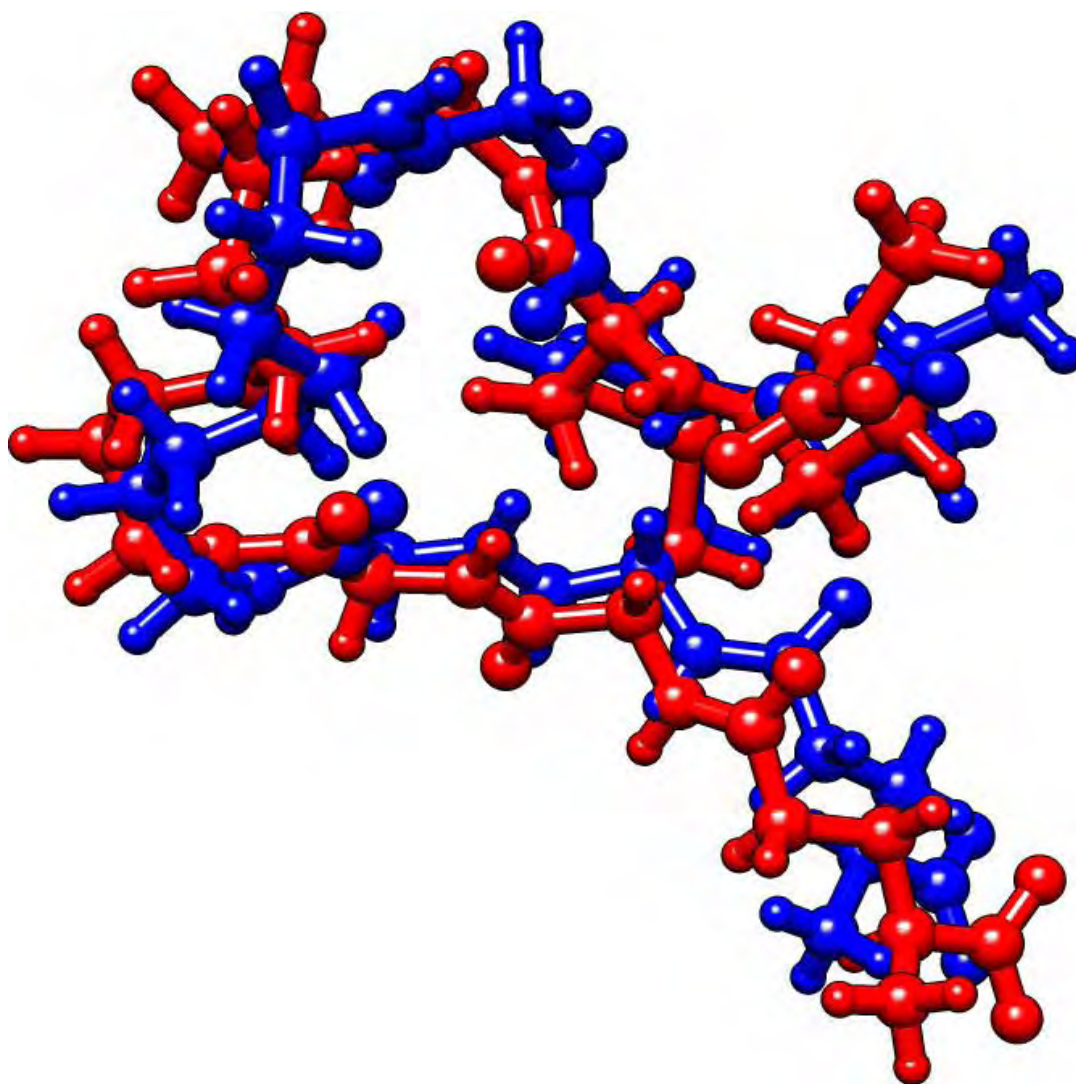


Figura 72: Superposição da tripanotona cristalográfica (azul) com a tripanotona posicionada com o docking rígido e minimizada por 5000 passos (vermelho).

É possível ver que apesar de haver alguns desencontros, a orientação da molécula está correta. Estas “imperfeições” na reprodução exata da estrutura cristalográfica, entretanto, são esperada.

Toda estrutura do sistema, na verdade, passa por pequenas modificações nas posições dos seus átomos durante a minimização. Para quantificar estas mudanças, tiramos o RMSD de cada estrutura minimizada em relação à cristalográfica. O resultado foi 0.340, 0.587 e 0.690 para 1000, 3000 e 5000 passos de minimização respectivamente. Estas modificações surgem da acomodação da estrutura em um mínimo de energia.

Esta acomodação também causa o surgimento de algumas ligações de hidrogênio entre o ligante e o sítio ativo da enzima. Com 1000 passos de minimização, surgem duas interações de hidrogênio entre átomos do ligante e átomos do sítio ativo.

Nas Figuras 73 e 74 verificamos a presença de uma ligação de hidrogênio na estrutura minimizada aos 1000 passos que na estrutura cristalográfica é substituída por uma distância de 3.964Å.

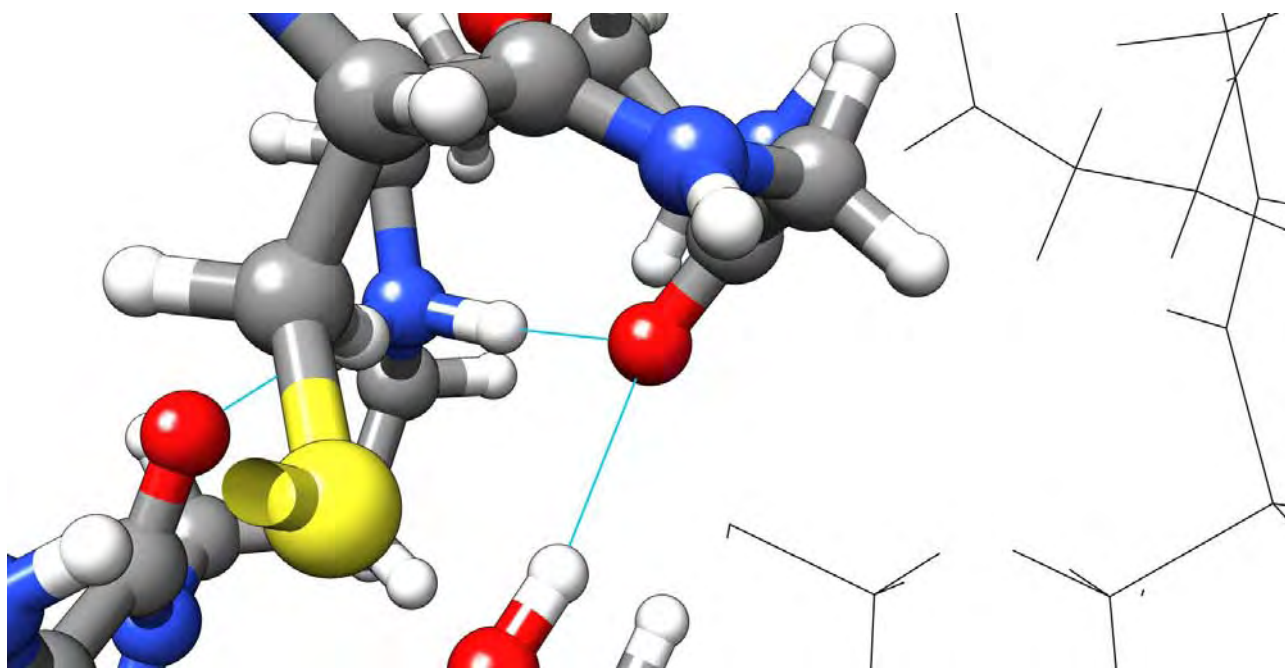


Figura 73: Figura mostrando duas ligações de hidrogênio intramoleculares na tripanotona e uma ligação de hidrogênio intermolecular. Estrutura minimizada aos 1000 passos.

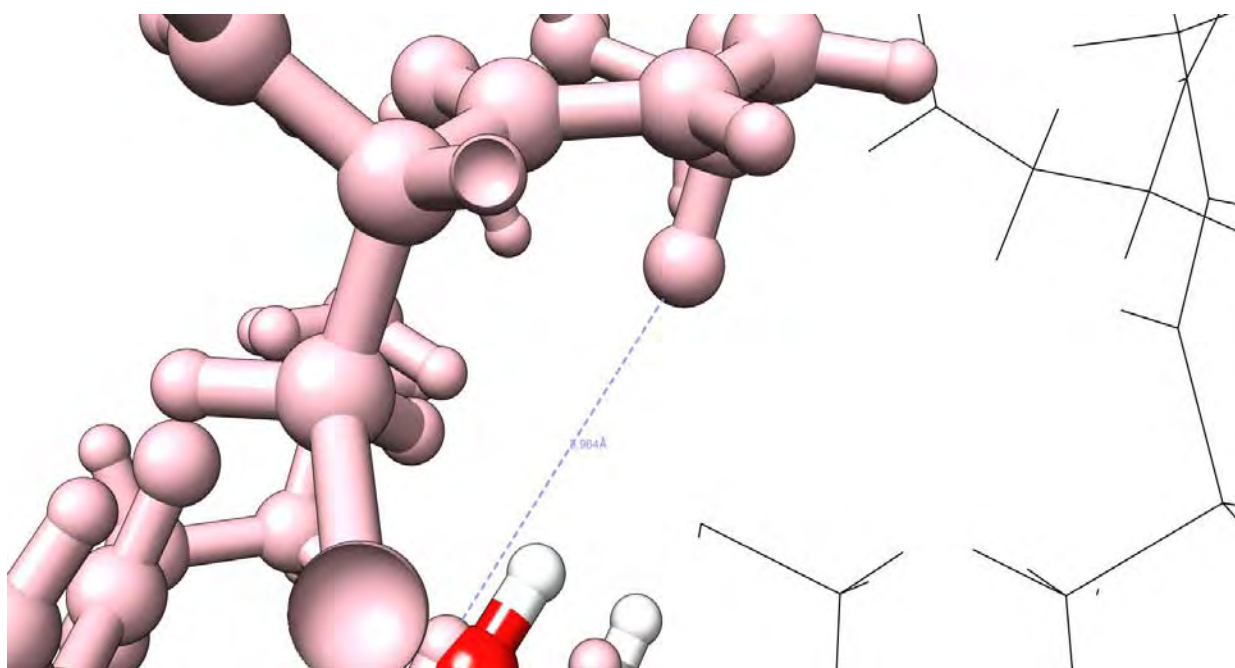


Figura 74: Figura mostrando a distância entre os mesmos átomos participantes na ligação de hidrogênio da figura anterior, na estrutura cristalográfica (em rosa).

Abaixo, na Figura 75, vemos mais uma ligação de hidrogênio intermolecular entre o ligante e os aminoácidos do sítio ativo que aparece na estrutura minimizada com 1000 passos mas que na estrutura cristalográfica é substituída por uma distância de 2.479Å entre os átomos participantes.

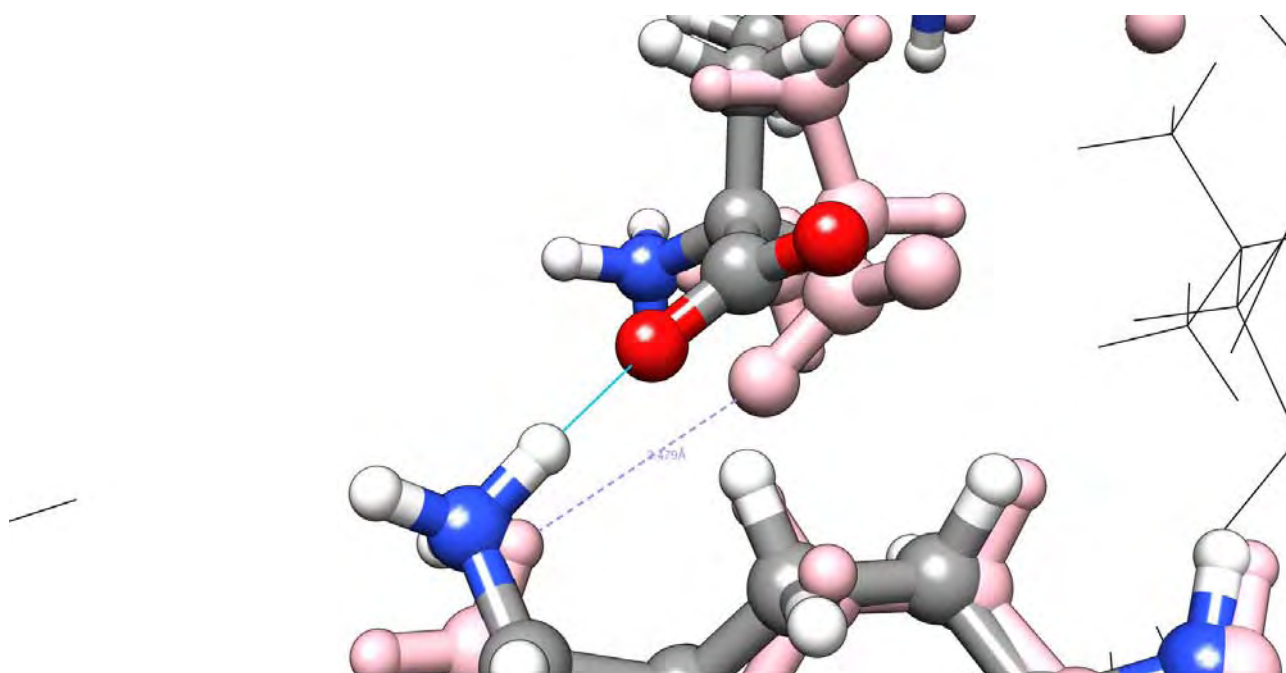


Figura 75: Ligação de hidrogênio formada na estrutura minimizada (em cores) que se encontra ausente na estrutura cristalográfica (em rosa).

Ao aumentar a quantidade de passos para 3000, outras duas ligações de hidrogênio aparecem na estrutura minimizada. Aos 5000 passos, no entanto, não há o surgimento de novas ligações. A contagem destas ligações foi feita com o UCSF Chimera 2.0 (PETTERSEN et al, 2004).

Estas ligações que surgem quando a estrutura se acomoda são levadas em consideração pelo min-MMGBSA porque o complexo utilizado pelo min-MMGBSA é minimizado. O *docking*, por outro lado, não modifica o ambiente químico do sítio ativo no qual o ligante se acomoda. Apenas os graus de liberdade dos ligantes foram considerados nos nossos cálculos e mesmo assim o tempo é bastante parecido.

O *docking* rígido da tripanotona durou 3,02 segundos, diminuindo o tempo total de cálculo do procedimento min-MMGBSA de 23 minutos para cerca de 13 minutos por molécula, superando em 10 minutos por molécula o tempo do *docking*. Note que este tempo de cálculo do min-MMGBSA já considera algum grau de acomodação do ligante no sítio ativo, bem como das moléculas de solvente. Para considerar graus de liberdade do sítio ativo, o *docking* molecular demoraria ainda mais tempo.

Uma outra opção para tornar o min-MMGBSA mais rápido que o *docking* feito neste trabalho seria utilizar 3000 passos de minimização, ao invés de 5000. A minimização com 3000 passos tomou apenas 7 minutos e 48 segundos, tornando o min-MMGBSA cerca de 2 minutos por molécula mais rápido que o *docking*, mesmo sem mudar alterar em nada o procedimento para produzir os complexos para a análise do min-MMGBSA.

Além disso, todos os métodos de minimização reproduziram a estrutura a estrutura cristalográfica da tripanotona com um erro consideravelmente menor que o *redock*, com menos tempo de computação.

5. Conclusões e Perspectivas

5.1. Conclusões

A estratégia de se estimar energia livre do complexo ligante-enzima min-MMGBSA se saiu melhor que o *docking* em todas as comparações. Não apenas ele foi capaz de ordenar melhor os ligantes como também deu uma estimativa de energia livre pautada no formalismo da termodinâmica estatística, e por isso teoricamente mais bem fundamentado que os escores do *docking*.

Mesmo nos experimentos que testaram a capacidade discriminatória, ao invés da exatidão, o *docking* se saiu pior o que realça que o min-MMGBSA é melhor que o *docking* em fazer o que o *docking* faz: varrer bancos de dados.

Uma diferença importante entre os dois métodos está no tempo de processamento. O *docking* é significativamente mais rápido que o min-MMGBSA e até pouco tempo, o min-MMGBSA não seria rápido o suficiente para participar em triagens virtuais. O surgimento da tecnologia Nvidia CUDA, entretanto, mudou isto.

Tomando o multi-MMGBSA, por exemplo. Apesar de ter uma correlação experimental comparável a métodos de Perturbação da Energia Livre (GUIMARÃES, 2011), este método leva cerca de 20 dias para varrer menos de 25 ligantes, mesmo sem os cálculos de entropia. Em comparação, o min-MMGBSA leva menos de 10 horas para realizar a mesma estimativa, utilizando o nosso procedimento de *docking* flexível e 5000 passos de minimização ou seja, é mais de 20 vezes mais rápido.

Sem a possibilidade de utilizar placas de vídeo para paralelizar os cálculos de minimização, o min-MMGBSA perde boa parte da competitividade que ele tem quando comparado ao *docking*. Os tempos de cálculo aumentam bastante.

Esta tecnologia, entretanto, também afeta o *docking* molecular, aumentando significativamente a quantidade de moléculas que este método é capaz de analisar eficientemente quando implementado em hardware e software de ponta. O *docking* hiperparalelo utilizando a tecnologia CUDA seria, em tese, capaz de efetuar muito mais cálculos do que seria possível realizar com o min-MMGBSA, empregando a mesma tecnologia CUDA nas minimizações.

O motivo disto é que o *docking* é um procedimento bastante rápido. E esta é a maior força do *docking* na triagem virtual de bancos de dados com milhares, dezenas de milhares ou centenas de milhares de compostos. Dimensões nas quais o min-MMGBSA não consegue competir em termos de rapidez. Nestes casos, o min-MMGBSA pode ser utilizado para complementar e refinar dados de *docking* molecular que já precisaram passar por um *docking* com propósitos de ordenação.

Mas naquelas aplicações onde o que se deseja é uma estimativa rápida e de melhor exatidão de 50 ou 100 compostos, o min-MMGBSA com *docking* rígido e 5000 passos de minimização seria capaz de fornecer em menos de um dia, enquanto o *docking* levaria por volta de 1 dia, 9 horas, 19 minutos e 59 segundos. Reduzindo para 3000 ou 4000 passos de minimização, o ganho em rapidez é ainda maior.

Muito embora o *docking* hiperparalelo ainda fosse teoricamente mais rápido, a qualidade dos resultados do min-MMGBSA seria maior, tanto em relação à estimativa das afinidades e o ordenamento dos ligantes quanto em relação à geometria do sítio ativo. Isto sugere que deve haver um tamanho tal de coleções de ligantes a serem triadas onde é preferível não utilizar o *docking* para ordenar as moléculas, mas sim o min-MMGBSA.

Da mesma forma que existe um tamanho tal de coleção onde é preferível aumentar a exatidão para métodos mais robustos, em detrimento do min-MMGBSA, sacrificando um pouco de desempenho em relação à rapidez.

É razoável supor também que métodos de exatidão semelhantes ao MMGBSA, como o LIE (*Linear Interaction Energy*) (AQVIST; MEDINA; SAMUELSSON, 1994), possam ser empregados no lugar do MMGBSA, mas um estudo mais aprofundado sobre isto deveria ser feito para estabelecer o ganho computacional.

Com um conjunto contendo entre 50 e 100 moléculas, por exemplo, pode-se empregar a estratégia do MMGBSA em uma única estrutura minimizada, proposta por Rastelli et al (2010)[1], utilizando o *docking* rígido ou um procedimento rápido de geração de complexos. É uma metodologia mais rápida que o *docking* convencional, e fornece melhores resultados que o *docking* convencional acerca do ordenamento relativo.

5.2. Perspectivas

Pretendemos testar outros softwares que se proponham a produzir bibliotecas virtuais de ligantes e testá-los para conseguir encontrar um que seja satisfatório. Um bom candidato, até o momento, é o LigandFit (VENKATACHALAM et al, 2003).

Existe interesse, também, estudar certos aspectos do *docking* que não foram diretamente estudados, por exemplo, o *docking* com receptor flexível. Determinar o ganho em exatidão em relação à perda de velocidade para diferentes metodologias de se incluir a flexibilidade do receptor no *docking* molecular ajudariam a esclarecer exatamente que tipo de vantagem (ou falta de) o min-MMGBSA poderia ter sobre estes métodos. Também seria interessante compara mais de uma função *score* e mais de um programa de *docking*.

Um estudo de particular importância seria determinar exatamente que impacto a tecnologia CUDA tem nas capacidades discriminatórias do *docking* molecular (em grandes bancos de dados).

Em paralelo com tudo isto, parece igualmente importante selecionar métodos adequados de tratar e manusear grandes quantidades de dados. Encontrar formas inteligentes de automatizar o processo pode acabar sendo tão importante quanto o próprio procedimento.

De modo especial, gostaríamos de incluir na avaliação o AutoDock Vina (TROTT e OLSON, 2010) segundo o rendimento das avaliações, sobretudo acerca do *docking* flexível.

6. Referências

1. ABAGYAN, R.; TOTROV, M.; KUZNETSOV, D. ICM – A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comp Chem*, n. 15, p. 448-506, 1994.
2. ABRAMS, C. CHE 800-002: Molecular Simulation Spring 0304. 2009. Disponível em: <<http://www.pages.drexel.edu/~cfa22/msim/msim.html>>. Acessado em: Junho e Julho de 2011.
3. ALTSCHUL, S.F.; GISH, W.; MILLER, W.; MYERS, E.W.; LIPMAN, D.J. Basic Local Alignment Search tool. *J Mol Biol*, n. 215, p. 403-410, 1990.
4. ANDERSEN, H.C. Molecular dynamics and constant pressure and/or temperature. *J Chem Phys*, n. 72, p. 2384-2393, 1980.
5. AQVIST, J.; MEDINA, C.; SAMUELSSON, J.E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng*, n. 7, p. 385-391, 1994.
6. ASTELBAUER, F.; WALOCHNIK, J. Antiprotozoal compounds: state of the art and new developments. *Int J Antimicrob Ag*, n. 38, p. 118-124, 2011.
7. BAIOTTO, P.; ILARI, A.; CECIT P.; ORSINI, S.; GRAMICCIA, M.; DI MUCCIO, T.; COLOTTI, G. Inhibitory Effect of Silver nanoparticles on Trypanothione Reductase Activity and *Leishmania infantum* proliferation. *ACS Med Chem Lett*, n. 2, p. 230-233, 2011.
8. BARRAL, A.; PERDAL-SAMPAIO, D, GRIMALDI JR, G.; MOMEN, H.; MCMAHON-PRATT, D.; DE JESUS, A.R.; ALMEIDA, R.; BADARO, R.; BARRAL-NETTO, M.; CARVALHO, E.M.; JOHNSON, W.D. Leishmaniasis in Bahia, Brazil: Evidence that *Leishmania amazonensis* produces a wide spectrum of clinical diseases. *Am J Trop Med Hyg*, n. 44, p. 536-546, 1991.
9. BAYLY, C.I.; CIEPLAK, P.; CORNELL, W.; KOLLMAN, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J Phys Chem*, n. 97, p. 10269-10280, 1993.
10. BERENDSEN, H.J.C.; POSTMA, J.P.M.; VAN GUNSTEREN, W.F.; DINOLA, A.; HAAK, J.R. Molecular dynamics with coupling to an external bath. *J Chem Phys*, n. 81, p. 3684-3690, 1984.
11. BERNSTEIN, F.C.; KOETZLE, T.F.; WILLIAMS, G.J.; MEYER JR, E.E.; BRICE, M.D.; RODGERS, J.R.; KENNARD, °; SHIMANOUCI, T.; TASUMI, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J Mol Biol*, n. 112, p. 535-542, 1977.
12. BEGLOV, D.; LEE, C.J.; DEBIASIO, A.; KOZAKOV, D.; BRENKE, R.; VAJDA, S.; BEGLOVA, N. Structural Insights into Recognition of β 2-Glycoprotein I by the lipoprotein receptors. *Proteins*, n. 77, p. 940-949, 2009.
13. BROOIJMANS, N; KUNTS, D.I. Molecular Recognition and Docking Algorithm. *Annu Rev Biophys*, n. 32, p. 335-373, 2003.
14. CASE, D.A.; GOHLKE, H. Converging Free Energy Estimates: MM-PB(GB)SA Studies on the Protein-Protein Complex Ras-Raf. *J Comput Chem*, n. 25, p. 238-250, 2004.
15. CASE, D.A.; DARDEN, T.A.; CHEATHAM III, T.E.; SIMMERLING, C.L.; WANG, J.; DUKE, R.E.; LUO, R.; WALKER, R.C.; ZHANG, W.; MERZ, K.M.; ROBERTS, B.; WANG, B.; HAYIK, S.; ROITBERG, A.; SEABRA, G.; KOLOSSVAI, I.; WONG, K.F.; PAESANI, F.; VANICEK, J.; LIU, J.; WU, X.; BROZELL, S.R.; STEINBRECHER, T.;

- GOHLKE, H.; CAI, Q.; YE, X.; WANG, J.; HSIEH, M.-J.; CUI, G.; ROE, D.R.; MATHEWS, D.H.; SEETIN, M.G.; SAGUI, C.; BABIN, V.; LUCHKO, T.; GUSAROV, S.; KOVALENKO, A.; KOLLMAN, P.A. AMBER Suite. Versão 11: University of California, 2010.
16. CHIPHOT, C.; PHORILLE, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*. Ed. 1, 2007. 535 p. ISBN 3540384472.
 17. CHULAY, J.D.; SPENCER, H.C.; MUGAMBI, M. Electrocardiographic changes during treatment of leishmaniasis with pentavalent antimony (sodium stibogluconate). *Am J Trop Med Hyg*, n. 34, p. 792-799, 1985.
 18. COOLEY, J.W.; TUKEY, J.W. An Algorithm for the Machine Calculation of Complex Fourier Series. *Math Comp*, n. 19, p. 297-301, 1965.
 19. CROFT, S.L.; BARRET, M.P.; URBINA, J.A. Chemotherapy of trypanosomiasis and leishmaniasis. *Trends Parasitol*, n. 21, p. 508-512, 2005.
 20. CROFT, S.L.; ENGEL, J. Miltefosine – Discovery of the antileishmanial activity of phospholipid derivatives. *Trans R Soc Trop Med Hyg*, n. 100, p. 4-8, 2006.
 21. DARDEN, T.; YORK, D.; PEDERSEN, L. Particle Mesh Ewald: An N·log(N) method for Ewald Sums in Large Systems. *J Chem Phys*, n. 98, p. 10089-10093, 1993.
 22. DAVIDSON, R.N.; DEN BOER, M.; RITMEIJER, K. Paromomycin. *Trans R Soc Trop Med Hyg*, n. 103, p. 653-660, 2009.
 23. DAVIS, I.W.; BAKER, D. RosettaLigand Docking with Full Ligand and Receptor Flexibility. *J Mol Biol*, n. 385, p. 381-392, 2009.
 24. DAVIS, I.W.; RAHA, K.; HEAD, M.S.; BAKER, D. Blind Docking of Pharmaceutically Relevant Compounds Using RosettaLigand. *Protein Sci* n. 18, p. 1998-2002, 2009.
 25. DESJARLAIS, R.L.; CUMMINGS, M.D.; GIBBS, A.C. Virtual Docking: How are we doing and how can we improve? *Drug Des Discov*, n. 3, p. 81-103, 2007.
 26. DORMAN, T.N.; MCGOVERN, S.L.; WITHERBEE, B.J.; KASTEN, T.P.; KURUMBAIL, R.; STALLINGS, W.C.; CONNOLLY, D.T.; SHOICHET, B.K. Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B. *J Med Chem*, n. 23, p. 2213-2221, 2002.
 27. EISENSTEIN, M. Microarrays: Quality control, *Nature*, n. 442, p. 1067-1070, 2006.
 28. ELDRIDGE, M.D.; MURRAY, C.W.; AUTON, T.R.; PAOLINI, G.V.; MEE, R.P. Empirical Scoring Functions: I. The Development of a Fast Empirical Scoring Function to Estimate the Binding Affinity of Ligands in Receptor Complexes. *J Comput-Aided Mol Dev*, n. 11, p. 425-445, 1997.
 29. ERCOLESSI, F. A Molecular Dynamics Primer. 1997. Disponível em: <http://www.fisica.uniud.it/~ercolessi/md/md/>. Acessado em: Junho e Julho de 2011.
 30. EWING, T.J.A.; MAKINO, S.; SKILLMAN, A.G.; KUNTZ, I.D. DOCK 4.0: Search Strategies for Automated Molecular Docking of Flexible Molecule Databases. *J Comput Aided Mol Des*, n. 15, p. 411-428, 2001.
 31. FRIESNER, R.A.; BANKS, J.L.; MURPHY, R.B.; HALGREN, T.A.; KLICIC, J.J. MAINZ, D.T.; REPASKY, M.P.; KNOLL, E.H.; SHELLEY, M.; PERRY, J.K.; SHAW, D.E.; FRANCIS, P.; SHENKIN, P.S. Glide: A New Approach for Rapid Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem*, n. 47, p. 1739-1749, 2004.

32. FIELD, M.J. *A Practical Introduction to The Simulation of Molecular Systems*. Ed. 2, 2007. 344 p. ISBN 9780521852524.
33. FUJITSU Scigress Explorer. Versão 7.7: Fujitsu Corporation, 2008.
34. GABELLI, S.B.; MCLELLAN, J.S.; MONTALVETTI, A.; OLDEFIELD, E.; DOCAMPO, R.; AMZEL, L.M. Structure and mechanism of the farnesyl diphosphate synthase from *Trypanosoma cruzi*: Implications for drug design. *Proteins*, n. 62, p. 80-88, 2006.
35. GASTEIGER, J.; MARSILI, M. A new model for calculating atomic charges in molecules. *Tetrahedron Letters*, n. 19, p. 3181-3184, 1978.
36. GENHEDEN, S.; RYDE, U. How to Obtain Statistically Converged MM/GBSA Results. *J Comp Chem*, n. 31, p. 837-846, 2010.
37. GRAVES, A.P.; SHIVAKUMAR, D.M.; BOYCE, S.E.; JACOBSON, M.P.; CASE, D.A.; SHOICHET, B.K. Rescoring Docking Hit Lists for Model Cavity Sites: Predictions and Experimental Testing. *J Mol Biol*, n. 377, p. 914-934, 2008.
38. GRETT, G.S.; KREMER, K. Molecular dynamics simulation for polymers in the presence of a heat bath. *Phys Rev A*, n. 35, p. 3628-3631, 1986.
39. GUIMARÃES, C.R.W. A Direct comparison of the MM-GB/SA Scoring Procedure and Free-Energy Perturbation Calculations Using Carbonic Anhydrase as a Test Case: Strengths and Pitfalls of Each Approach. *J Chem Theory Comput*, n. 7, p. 2296-2306, 2011.
40. GUIMARÃES, C.R.W. A Direct Comparison of the MM-GB/SA Scoring Procedure and Free-Energy Perturbation Calculations Using Carbonic Anhydrase as a Test Case: Strengths and Pitfalls of Each Approach. *J Chem Theory and Comp*, n. 7, p. 2296-2306, 2011.
41. GUODONG, H.; DUNYOU, W.; XINGUO, L.; QINGGANG, Z. A computational analysis of the binding model of MDM2 with inhibitors. *J Comput Aided Mol Des*, n. 24, p. 687-697, 2010.
42. HAIDER, M.K.; BERTRAND, H.O.; HUBBARD, R.E Predicting Fragment Binding Poses Using a Combined MCSS MM-GBSA Approach. *J Chem Inf Model*, n. 51, p.1092-1105, 2011.
43. HANN, M.M.; OPREA, T.I. Pursuing the leadlikeness concept in pharmaceutical research. *Curr Opin Chem Biol*, n. 8, p. 255-263, 2004.
44. HERWALDT, B.L.; BERMAN J.D.; Recommendations for treating leishmaniasis with sodium stibogluconate (Pentostam) and Review of Pertinent Clinical Studies. *Am J Trop Med Hyg*, n. 46, p. 296-306, 1992.
45. HOWE, D.; COSTANZO, M.; FEY, P.; GOJOBORI, T.; HANNICK, L.; HIDE, W.; HILL D.P.; KANIA, R.; SCHAEFFER, M.; ST PIERRE, S.; TWIGGER, S.; WHITE, O.; RHEE, S.Y. Big data: The future of biocuration. *Nature*, n. 455, p. 47-50, 2008.
46. HOU, T.; WANG, J.; LI, Y.; WANG, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *J Chem Inf Model*, n. 51, p. 69-82, 2011.
47. HUANG, N.; SHOICHET, B.K.; IRWIN, J.J. Benchmarking Sets for Molecular Docking. *J Med Chem*, n. 49, p. 6789-6801, 2006.
48. HUBBARD, R.E. Structure-based drug discovery and protein targets in the CNS. *Neuropharmacology*, n. 60, p. 7-23, 2011.
49. JAKALIAN, A.; JACK, D.B.; BAYLY, C.I. Fast, efficient generation of high-quality atomic

- charges. AM1-BCC model: II. Parametrization and Validation. *J Comput Chem*, n. 23, p. 1623-1641, 2002.
50. JONES, G.; WILLET, P.; GLEN, R.C. Molecular Recognition of a Receptor Site Using a Genetic Algorithm With a Description of Desolvation, *J Mol Biol*, n. 245, p. 43-53, 1995.
51. JONES, G.; WILLET, P.; GLEN, R.C.; LEACH, A.R.; TAYLOR, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J Mol Biol*, n. 267, p. 727-748, 1997.
52. JORGENSEN, W.L. The Many Roles of Computation in Drug Discovery. *Science*, n. 303, p. 1813-1818, 2004.
53. JUNIOR, C.G.L.; DE ASSIS, P.A.C.; SILVA, F.P.L.; SOUSA, S.C.O.; DE ANDRADE, N.G.; BARBOSA, T.P.; NERÍS, P.L.N.; SEGUNDO, L.V.G.; ANJOS, I.C.; CARVALHO, G.A.U.; ROCHA, G.B.; OLIVEIRA, M.R.; VASCONCELLOS, M.L.A.A. Efficient synthesis of 16 aromatic Morita-Baylis-Hillman adducts: Biological evaluation on *Leishmania amazonensis* and *Leishmania chagasi*. *Bioorg Chem*, n. 38, p. 279-284, 2010.
54. KILLICK-KENDRICK, R. The biology and control of phlebotomine sandflies. *Clin Dermatol*, n. 17, p. 279-289, 1999.
55. KONTOYIANNI, M.; MADHAV, P.; SUCHANEK, E.; SEIBEL, W. Theoretical and Practical Considerations in Virtual Screening: A Beaten Field? *Curr Med Chem*, n. 15, p. 107-116, 2008.
56. KUHN, B.; GERBER, P.; SCHULZ-GASCH, T.; STAHL, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J Med Chem*, n. 48, p. 4040-4048, 2004.
57. LARKIN M.A.; BLACKSHIELDS, G.; BROWN, N.P.; CHENNA R.; MCGETTIGAN, P.A.; MCWILLIAM, H.; VALENTIN, F.; WALLACE, I.M.; WILM A.; LOPEZ, R.; THOMPSON, J.D.; GIBSON, T.J.; HIGGINS, D.G. Clustal W and Clustal X version 2.0. *Bioinformatics*, n. 21, p. 2947-2948, 2007.
58. LASKOWSKI, R.A.; MACARTHURS, M.W.; MOSS, D.S.; THORNTON, J.M. PROCHECK – A Program to Check the Stereochemical Quality of Protein Structures. *J App Cryst*, n. 26, p. 283-291, 1993.
59. LAZO, J.S.; WIPF, P. Combinatorial Chemistry and Contemporary Pharmacology. *J Pharm Expr Ther*, n. 293, p. 705-709, 2000.
60. LEACH, A.R.; SHOICHET, B.K.; PEISHOFF, C.E. Prediction of Protein-Ligand Interactions, Docking and Scoring: Successes and Gaps. *J Med Chem*, n. 49, p. 5851-5855, 2006.
61. LENNARD-JONES, J.E. On the Determination of Molecular Fields. *Proc R Soc Lond A*, n. 106, p. 463-477, 1924.
62. LIPINSKI, C.A.; LOMBARDO, F.; DOMINY, B.W.; PAUL, J.F. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv*, n. 23, p. 3-25, 1997.
63. LISUREK, M.; RUPP, B.; WICHARD, J.; NEUENSCHWANDER, M.; VIN KRIES, J.P.; FRANK, R.; RADEMANN, J.; KÜHNE, R. Design of chemical libraries with potentially bioactive molecules applying a maximum common substructure concept. *Mol Divers*, n. 14, p. 401-408, 2009.
64. MAAROUF, M.; ADELINE M.T.; SOLIGNAC, M.; VAUTRIN, D.; ROBERT-GERO M. Development and characterization of paromomycin-resistant *Leishmania donovani*

- promastigotes. *Parasite*, n. 5, p. 167-173, 1998.
65. MCQUARIE, D.A. *Statistical Mechanics*. Ed. 1, 1976.
 66. MIERTUS, S.; FASSINA, G.; SENEĆI, P.F. Concepts of Combinatorial Chemistry and Combinatorial Technologies. *Chem Listy*, n. 94, p. 1104-1110, 2000.
 67. MOITESSIER, N.; ENGLEBIENNE, P.; LEE, D.; LAWANDI, J.; CORBEIL, C.R. Towards the Developments of Universal, Fast and Highly Accurate Docking/Scoring Methods: A Long Way to Go. *Brit J Pharmacol*, n. 153, p. 7-26, 2008.
 68. MORRIS, G.M.; GOODSSELL, D.S.; HALLIDAY, R.S.; HUEY, R.; HART, W.E.; BELLEW, R.K.; OLSON, A.J. Automated docking using Lamarckian Genetic Algorithm and an Empirical Binding Free Energy Function. *J Comp Chem*, n. 19, p. 1639-1662, 1998.
 69. MOUSTAKAS, D.T.; LANG, P.T.; PEGG, S.; PETTERSEN, E.; KUNTZ, I.D.; BROOIJMANS, N.; RIZZO, R.C. Development and Validation of a Modular, Extensible Docking Program: DOCK 5. *J Comput Aided Mol Des*, n. 20, p. 601-619, 2006.
 70. MUEFFE, I.; MARTIN, Y.C. A General Fast Scoring Function for Protein-ligand Interactions: A Simplified Potential Approach. *J Med Chem*, n. 42, p. 791-804, 1999.
 71. NATERA S.; MACHICA C.; PADRON-NIEVES M.; ROMERO A.; DIAZ E.; PONTE-SUCRE A. Leishmania spp: Proficiency of drug-resistan parasites. *Int J Antimicrob Ag*, n. 29, p. 637-642, 2007.
 72. NAVIN, T.R.; ARANA. B.A.; ARANA, F.E.; BERMAN, J.D.; CHAJÓ, J.F. Placebo-controlled clinical trial of sodium stibogluconate (Pentostam) versus ketoconazole for treating cutaneous leishmaniasis in Guatemala. *J Infect Dis*, n. 165, p. 528-534, 1992.
 73. NIJBOER, B.R.A.; DE WETTE, F.W. On the calculation of lattice sums. *Physica*, n. 23, p. 309-321, 1957.
 74. OPEN EYE BROOD. Versão 1.0: Open Eye Scientific Software, 2006.
 75. PEREZ-SALA D. Protein isoprenylation in biology and disease: general overview and perspectives from studies with genetically engineered animals. *Frontier Biosci*, n. 12, p. 4456-4472, 2007.
 76. PETTERSEN, E.F.; GODDARD, T.D.; HUANG, C.C.; COUCH, G.S.; GREENBLATT, D.M.; MENG, E.C.; FERRIN, T.E. UCSF Chimera – A Visualization System for Exploratory Research and Analysis. *J Comp Chem*, n. 25, p. 1605-1612, 2004.
 77. PORTER, C.T.; BARTLETT, G.J.; THORNTON, J.M. The Catalytic Site Atlas: A resource of catalytic sites and residues identified in enzymes using structural data. *Nucl Acids Res*, n. 32, p. 129-133, 2004.
 78. PRABHU, N.V.; ZHU, P.; SHARP, K.A. Implementation and testing of stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson-Boltzmann method. *J Comp Chem*, n. 25, p. 2049-2064, 2004.
 79. RASTELLI, G.; DEGLIESPOSTI G.; DEL RIO A.; SGOBBA, M. Binding Estimation after Refinement, a New Automated Procedure for the Refinement and Rescoring of Docked Ligands in Virtual Screening. *Chem Biol Drug Des*, n. 73, p. 283-286, 2009.
 80. RASTELLI, G.; DEL RIO, A.; DEGLIESPOSTI, G.; SGOBBA, M. Fast and Accurate Prediction of Binding Free Energies Using MM-PBSA and MM-GBSA. *J Comp Chem*, n. 31, p. 797-810, 2010.
 81. RATH, S.; TRIVELIN, L.A.; IMBRUNITO, T.R.; TOMAZELA, D.M.; DE JESÚS M.N.;

- MARZAL, P.C. Antimonias Empregados no Tratamento da Leishmaniose: Estado da Arte. *Química Nova*, n. 26, p. 550-555, 2003.
82. RESAT, H.; MEZEI, M. Studies on free energy calculations using a polynomial path. I. Thermodynamic Integration. *J Chem Phys*, n. 99, p. 6052-6061, 1993.
83. REYMOND, J.L.; VAN DEURSE, R.; BLUM, L.C.; RUDDIGKEIT, L. Chemical space as a source for new drugs. *Med Chem Comm*, n. 1, p. 30-38, 2010.
84. RIPPHAUSEN, P.; NISIUS, B.; BAJORATH, J. State-of-the-art in ligand-based virtual screening. *Drug Discov Today*, n. 16, p. 372-376, 2011.
85. RYCKAERT, J.P.; CICCOTTI, G.; BERENDSEN, H.J.C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkenes. *J Comp Phys*, n. 23, p. 327-341, 1977.
86. SACKS, D; KAMHAWI, S. Molecular aspects of parasite-vector and vector-host interactions in leishmaniasis. *Annu Rev Microbiol*, n. 55, p. 453-483, 2001.
87. SADIQ, S.K.; WRIGHT, D.W.; KENWAY, O.A.; COVENEY, P.V. Accurate Ensemble Molecular Dynamics Binding Free Energy Ranking of Multidrug-Resistant HIV-1 Proteases. *J Chem Inf Model*, n. 50, p. 890-905, 2010.
88. SALI, A.; BLUNDELL, T.L. Comparative protein modelling by Satisfaction of Spatial Restraints. *J Mol Biol*, n. 234, p. 779-815, 1993.
89. SALUM, L.B.; ANDRICOPULO, A.D. Fragment-based QSAR: perspectives in drug design. *Mol Divers*, n. 13, p. 277-285, 2009.
90. SANDER, J.M.; BORGES, A.R.; JUNIOR, C.G.L.; SILVAR, F.P.L.; CARVALHO G.A.U.; ROCHA, G.B.; VASCONCELLOS, M.L.A.A.; FIGUEIREDO, R.C.B.Q. 3-Hydroxy-2-methylene-3-(4-nitrophenylpropanenitrile): A new highly active compound against epimastigote and trypomastigote form of *Trypanosoma cruzi*. *Bioorg Chem*, n. 38, p. 190-195, 2010.
91. SANNER, M.F. Python: A Programming Language for Software Integration and Development. *J. Mol. Graphics Mod*, n. 17, p. 57-61, 1999.
92. SANTOS, D.O.; COUTINHO, C.E.R.; MADEIRA, M.F.; BOTTINO, C.G.; VIEIRA R.T.; NASCIMENTO, S.B.; BERNARDINO, A.; BOURGUIGNON, S.C.; CORTE-REAL. S.; PINHO, R.T. Leishmaniasis Treatment – A Challenge that Remains: A Review. *Parasitol Res*, n. 103, p. 1-10, 2008.
93. SEIFERT, M.H.J. Targeted scoring functions for virtual screening. *Drug Discov Today*, n. 14, p. 562-570, 2009.
94. SHAW, J. The leishmaniasis – survival and expansion in a changing world. *Mem Inst Oswaldo Cruz*, n. 102, p. 541-547, 2007.
95. SINDERMAN, H.; ENGEL, J. Development of miltefosine as an oral treatment for leishmaniasis. *Trans R Soc Trop Med Hyg*, n. 100, p. 17-20, 2006.
96. SRINIVASAN, J.; CHEATHAM, T.E.; CIEPLAK, P.; KOLLMAN, P.A.; CASE, D.A. Continuum Solvent Studies of the Stability of DNA, RNA and Phosphoramidate-DNA Helices. *J Am Chem Soc*, n. 120, p. 9401-9409, 1998.
97. STILL, W.C.; TEMPCZYK, A.; HAWLEY, R.C.; HENDRICKSON, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*, n. 112, p. 6127-6129, 1990.

98. STOICA, I.; SADIG, S.K.; COVENEY, P.V. Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases. *J Am Chem Soc*, n. 130, p. 2639-2648, 2008.
99. SUNDAR, S.; SINGH, V.P.; SHARMA, S.; MAKHARIA, M.K.; MURRAY, H.W. Response to Interferon- γ plus antimony in indian visceral leishmaniasis. *J Infect Dis*, n. 176, p. 1117-1119, 1997.
100. SUNDAR S.; CHAKRAVARTY, J.; RAI V.K.; AGRAWAL N.; SINGH S.P.; CHAUHAN V.; MURRAY H.W. Amphotericin B treatment for Indian visceral leishmaniasis: response to 15 daily versus alternate-day infusions. *Clin Infect Dis*, n 45, p. 556-561, 2007.
101. SUNDAR, S.; CHAKRAVARTY, J. Antimony Toxicity. *Int J Environ Res Public Health*, n 7, p. 4267-4277, 2010.
102. TAYLOR, R.D.; JEWSBURY, P.J.; ESSEX, J.W. A review of protein-small molecule docking methods. *J Comput-Aided Mol Des*, n. 16, p. 151-166, 2002.
103. THAKUR, C.P.; DEDET, J.P.; NARAIN, S.; PRATLONG F. Leishmania species, drug unresponsiveness and visceral leishmaniasis in Bihar, India. *Trans R Soc Trop Med Hyg*, n 95, p. 187-189, 2001.
104. THOMPSON, D.C.; HUMBLET, C.; JOSEPH-MCCARTHY D. Investigation of MM-PBSA Rescoring of Docking Poses. *J chem Inf Model*, n. 48, p. 1081-1091, 2007.
105. TROTT, O.; OLSON, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comp. Chem*, n. 31, p. 455-461, 2010.
106. UNIPROT CONSORTIUM, Ongoing and Future Developments at the Universal Protein Resource. *Nucleic Acid Res*, n. 39, p. 214-219, 2011.
107. URBINA, J.A. Lipid biosynthesis pathways as chemotherapeutic targets in kinetoplastid parasites. *Parasitology*, n. 114, p. 91-99, 1997.
108. URBINA, J.A.; CONCEPCION, J.L.; RANGEL, S.; VISBAL, G.; LIRA, R. Squalene synthase as a chemotherapeutic target in *Trypanosoma cruzi* and *Leishmania mexicana*. *Mol Biochem Parasitol*, n. 125, p. 35-45, 2002.
109. VENKATACHALAM, C.M.; JIANG, X.; OLDFIEDL, T.; WALDMAN M. LigandFit: A Novel Method for The Shape-Directed Rapid Docking of Ligands to Protein Active Sites. *J Mol Graph Model*, n. 21, p. 289-307, 2003.
110. VERLET, L. Computer Experimentas on Classical Fluids. *Phys Rev*, n. 159, p. 98-103, 1967.
111. VESELY, F.J. Statistica Physics. 2005. Disponível em http://homepage.univie.ac.at/franz.vesely/sp_english/sp/sp.html>. Acessado em: Junho e Julho de 2011.
112. VIGERS, G.P.; RIZZI, J.P. Multiple Active Site Corrections for Docking and Virtual Screening. *J Med Chem*, n. 47, p. 80-89, 2004.
113. WARREN, G.L.; ANDREWS, C.W.; CAPELLI, A.M.; CLARKE, B.; LALONDE, J.; LAMBERT, M.H.; LINDVALL, M.; NEVINS, N.; SEMUS, S.F.; SENGER, S.; TEDESCO G.; WALL, I.D.; WOOLVEN, J.M.; PEISHOFF, C.E.; HEAD, M.S. A Critical Assessment of Docking Programs and Scoring Functions. *J Med Chem*, n. 49, p. 5912-5931, 2006.
114. WASZKOWYCZ, B.; CLARK, D.E.; GANCIA, E. Outstanding challenges in protein-ligand docking and structure-based virtual screening. *WIREs Comput Mol Sci*, n. 1, p. 229-

259, 2011.

115. WILLIAMS, D.J.; HALL, K.B. Unrestrained stochastic dynamics simulations of the UUCG tetraloop using an implicit solvation model. *Biophys J*, n. 76, p. 3192-3205, 1999.
116. WINKLER, D.A. The role of quantitative structure-activity relationship (QSAR) in biomolecular discovery. *Brief in Bioinf*, n. 3, p. 73-86, 2001.
117. WINOGRAD, N.; BRAUN, R.M. Imaging Mass Spectrometry and Combinatorial Chemistry. *Spectroscopy*, n. 16, p. 14-16, 2001.
118. YOSHIFUMI, F. Structural Ensemble in Computational Drug Screening. *Expert Opin on Drug Des & Tox*, n. 7, p. 835-849, 2010.
119. YURIEVA, E.; AGOSTINO, M.; RAMSLAND, P.A. Challenges and Advances in Computational Docking: 2009 in Review. *J Mol Recognit*, n. 24, p. 149-164, 2001.