UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

VICTOR HENRIQUE RABESQUINE NOGUEIRA

# Validação de métodos de Monte Carlo para avaliação de energia de interação proteína-ligante

São Carlos

2019

VICTOR HENRIQUE RABESQUINE NOGUEIRA

# Validação de métodos de Monte Carlo para avaliação de energia de interação proteína-ligante

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular
Orientador: Prof. Dr. Alessandro Silva Nascimento.

Versão Corrigida

(versão original disponível na Unidade que aloja o Programa)

São Carlos

2019

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

Nogueira, Victor Henrique Rabesquine
Validação de métodos de Monte Carlo para avaliação de energia de interação proteína-ligante / Victor Henrique Rabesquine Nogueira; orientador Alessandro Silva Nascimento - versão corrigida -- São Carlos, 2019.
144 p.

Dissertação (Mestrado - Programa de Pós-Graduação em Física Aplicada Biomolecular) -- Instituto de Física de São Carlos, Universidade de São Paulo, 2019.

1. Monte Carlo. 2. Energia livre de ligação. 3. Interações moleculares. I. Nascimento, Alessandro Silva, orient. II. Título.

À minha família, com muita gratidão e amor por
todo o apoio e compreensão durante toda minha vida.

## AGRADECIMENTOS

Primeiramente, gostaria de dizer que não tenho palavras para agradecer meus pais e à minha irmã por todo o suporte que deram durante todos esses anos, por acreditarem em mim e não me deixarem desistir, mesmo quando essa parecia a melhor opção. Amo muito vocês. Muitíssimo obrigado por tudo!

Também gostaria de agradecer imensamente ao Prof. Dr. Alessandro Silva Nascimento, por ter aceitado ser meu orientador durante todo esse período, por toda a paciência e ajuda com os problemas que apareceram e principalmente por compartilhar seus conhecimentos sobre os mais diversos assuntos. Sugiro a qualquer um que nunca tenha conversado com ele que o faça. Chame-o para um café. Será bastante produtivo e enriquecedor.

Gostaria de agradecer a todo o pessoal que passou pela sala 9 durante essa minha estadia. Em especial à Sthéffany, que além de cozinhar muito bem e me ensinar a usar uma panela de pressão, viveu os mesmo momentos de tensão durante esse tempo e, como dito em *Harry Potter e a Pedra Filosofal, "há coisas que não se pode fazer junto sem acabar gostando um do outro, e derrubar um trasgo montanhês de quase quatro metros de altura é uma dessas coisas"*, no nosso caso, o trasgo montanhês foi o mestrado. À Érika, por me acompanhar durante toda essa jornada, tornando-se uma excelente amiga, além de todos os "cafés com Érika", cheios de conversas profundas, compartilhando suas experiências, sofrendo junto e ajudando a resolver muitos problemas, e, é claro, por ter me apresentado à Casa Aurora. Ao Geraldo, que apesar do curto período, se tornou um amigo bastante querido. À Lívia, por todos os nossos karaokês, pelo portunhol quebrado e por todas as azedinhas.

Um agradecimento especial também à Heloisa e ao João Vitor, que, apesar do pouco tempo de contato, foram responsáveis, juntamente com o prof. Alessandro, por viabilizar meu trabalho.

Gostaria também de agradecer muito ao Murilo, que de longa data vem me ajudando e me acompanhando em todos os momentos, seja durante o surto da madrugada indo comer porcaria e conversar até o surto passar, seja para chamar minha atenção para focar no que eu deveria fazer e sentar junto para estudar. Você é demais!

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Agradeço a CAPES pela bolsa concedida e à FAPESP e ao CNPq pelo financiamento do projeto.

Agradeço ao IFSC pela infraestrutura.

It is important to draw wisdom from many different places.
If you take it from only one place, it becomes rigid and stale.

Uncle Iroh – Avatar: The Last Airbender.

## RESUMO

NOGUEIRA, V. H. R. **Validação de métodos de Monte Carlo para avaliação de energia de interação proteína-ligante**. 2019. 144 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

Os sistemas biológicos macromoleculares são conhecidos por serem sistemas interagentes. Essas interações são fundamentais para processos como comunicação celular, especificidade de reações enzimáticas e regulação da expressão gênica. Os métodos disponíveis atualmente para estimar a afinidade das interações biomoleculares podem ser divididos, basicamente, em dois grupos: métodos rápidos que estimam a energia livre de ligação através de aproximações de campo de força (por exemplo, *docking*); e os métodos que são baseados em *ensembles* de Dinâmica Molecular (DM) para calcular as energias livres de ligação de maneira mais rigorosa, porém, com custo computacional mais elevado. O objetivo deste trabalho é aprimorar e validar um método menos custoso para o cálculo da energia livre de ligação. Para isso, simulações atomísticas de Monte Carlo (MC) dos ligantes no sítio de ligação são usadas para gerar *ensembles* termodinâmicos. Depois disso, as energias livres de ligação são calculadas usando uma combinação de energias e entropias estimadas através de uma estratégia de aproximação de primeira ordem. Dois algoritmos de amostragem foram avaliados no cálculo de energia de ligação. O primeiro algoritmo amostra graus de liberdade de translação e rotação randômicas do centro de massa do ligante no sítio de ligação, além de variações randômicas nos ângulos de torção envolvendo átomos pesados (não hidrogênio). O segundo amostra graus de liberdade rotacional e translacional do centro de massa, além de deslocamentos atômicos individuais para cada átomo do ligante. Além disso, diferentes modelos para calcular as contribuições polares para interação intermolecular foram utilizados. Comparações entre as energias livres de ligação calculadas com baixo custo computacional e as experimentais disponíveis na literatura para o sistema modelo utilizado, lisozima do vírus T4, mostraram uma correlação considerável ($r$=0,64 para $N$=27). Esses dados também apresentaram resultados interessantes quando comparados com outras metodologias, tais como LIE, MM-PBSA e MM-GBSA. Assim, a abordagem utilizada para a determinação das energias de interação mostrou-se eficiente em termos de tempo computacional e

para comparação com dados de energia livre de ligação determinados experimentalmente.

Palavras-chave: Monte Carlo. Energia livre de ligação. Interações moleculares.

## ABSTRACT

NOGUEIRA, V. H. R. **Validation of Monte Carlo methods for evaluating protein-ligand binding free energy**. 2019. 144 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

Macromolecular biological systems are widely known by its interaction properties. Those interactions play fundamental roles in processes such as cellular communication, specificity of enzymatic reactions and regulation of gene expression. The methods currently available to estimate the affinity of biomolecular interactions can be divided basically into two groups: fast methods that estimate the free energy of binding through force field approximations (e.g., docking); and methods that are based on Molecular Dynamics (DM) ensembles to calculate binding free energies more rigorously, however, with higher computational cost. The objective of this work is to improve and validate a less costly method for calculating binding free energy. For this, atomistic Monte Carlo (MC) simulations of ligands at the binding site are used to generate thermodynamic ensembles. Thereafter, the binding free energies are calculated using a combination of energies and entropy estimated through a first-order approximation strategy. Two sampling algorithms were evaluated in the calculation of the binding energy. The first one samples the degrees of freedom from translation and rotation of the center of mass of the binder at the binding site, as well as random variations in the torsion angles involving heavy atoms (non-hydrogen). The second one samples the rotational and translational degrees of freedom of the ligand center of mass, as well as individual atomic displacements for each atom of the ligand. In addition, different models to calculate the polar contributions for intermolecular interaction were used. Comparisons between the binding free energies calculated with low computational cost and the experimental ones available in the literature for the system used, T4 virus lysozyme, resulted in acceptable correlation values ($r$=0.64 for $N$=27). Those data also showed interesting results compared to different methodologies such as LIE, MM-PBSA and MM-GBSA. Therefore, the used approach for determining the binding energies was efficient in terms of computational time and for comparison with free energy data determined experimentally.



Keywords: Monte Carlo. Binding free energy. Molecular interactions.

LISTA DE FIGURAS

Figura 1 - Ligantes das enzimas T4 mutantes, com suas energias livres de ligação, escolhidos para as simulações. ..... 45

Figura 2 - Estrutura 3D e visão de seção da superfície de lisozima mutante de vírus T4. À esquerda, em azul, está o ligante 5-cloro-2-metilfenol, na cavidade, com os resíduos mutantes (L99A/M102Q) representados, interagindo com o ligante. PDB ID: 3HT8. À direita, está o ligante benzeno, na cavidade, com o resíduo mutante (M102E) representado junto com uma molécula de água no sítio de interação, interagindo com o ligante. PDB ID: 3GUJ. ..... 46

Figura 3 - Comparação entre os dados experimentais e os valores obtidos de energia livre de ligação utilizando o módulo de docking do LiBELa. ..... 55

Figura 4 - Histogramas apresentando a distribuição de RMSDs obtidos para a as simulações de cada complexo. Os valores de RMSDs são calculados para o ligante somente e tendo como referência a conformação cristalográfica do ligante. ..... 57

Figura 5 - Avaliação das energias médias de ligação para as simulações de dinâmica molecular para os 30 complexos como função do RMSD do ligante comparado com a posição cristalográfica. ..... 59

Figura 6 - Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno), cujo complexo apresenta apenas a mutação L99A. No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura. ..... 60

Figura 7 - Comparação entre os dados experimentais e os valores obtidos de energia livre de ligação para as simulações de DM par aos 30 complexos simulados. A) Energias obtidas das simulações de DM no vácuo. B) Energias obtidas utilizando o método de LIE, com $\beta$ = 0,5 e $\alpha$ = 0,16. C) Energias obtidas utilizando o método de MM-GBSA. D) Energias obtidas utilizando o método de MM-PBSA. ..... 63

Figura 8 - Histogramas apresentando a distribuição de RMSDs obtidos para a as simulações de cada complexo. Os RMSDs são calculados para o ligante somente e tendo como referência a conformação cristalográfica do ligante. Em azul, vemos a amostragem para as simulações de DM. Em vermelho, está a amostragem para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Coulomb, 250K. Em verde, está a amostragem, como uma distribuição de RMSD, para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Poisson, 250K. ..... 67

Figura 9 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Coulomb, temperatura é 250K e amostragem é sample_torsion..... 70

Figura 10 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Poisson, temperatura é 250K e amostragem é sample_torsion. ..... 72

Figura 11- Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno). No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura. Em marrom, está o perfil gerado por DM; em vermelho o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Coulomb como modelo eletrostático; e em verde o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Poisson como modelo eletrostático. ..... 73

Figura 12 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações..... 75

Figura 13 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações..... 76

Figura 14 - Autocorrelações obtidas para as simulações de MC para todos os alvos, utilizando amostragem do tipo *sample_torsion*, com temperatura de 250K e Coulomb como modelo eletrostático. ..... 79

Figura 15 - Autocorrelações obtidas para as simulações de MC para todos os alvos, utilizando amostragem do tipo *sample_torsion*, com temperatura de 250K e Coulomb como modelo eletrostático. ..... 81

Figura 16 - Convergência das energias livres de ligação para os 30 complexos, em escala logarítmica, para 40 milhões de passos de simulação de MC, usando Coulomb como modelo eletrostático. ..... 83

Figura 17 - Convergência das energias livres de ligação para os 30 complexos, em escala logarítmica, para 40 milhões de passos de simulação de MC, usando Poisson como modelo eletrostático..... 84

Figura 18 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes. ..... 87

Figura 19 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes. ..... 88

Figura 20 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático e 0,5 Å como constante de amortecimento. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes..... 90

Figura 21 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Poisson como modelo eletrostático e 0,5 Å como constante de amortecimento. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes..... 91

Figura 22 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático e $\alpha$ = 0,2 kcal/(mol.$e^{-2}$) e $\beta$ = -0,005 kcal/mol como parâmetros de solvatação. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes. ..... 94

Figura 23 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Poisson como modelo eletrostático e α = 0,2 kcal/(mol.$e^{-2}$) e β = -0,005 kcal/mol como parâmetros de solvatação. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes. ..... 95

Figura 24 - Comparação das correlações entre os valores experimentais e os valores calculados de entropia (-TΔS) e entalpia (ΔH). Valores calculados utilizando Coulomb como modelo eletrostático e 0,5 Å como constante de amortecimento. ..... 97

Figura 25 - Comparação das correlações entre os valores experimentais e os valores calculados de entropia (-TΔS) e entalpia (ΔH). Valores calculados utilizando Poisson como modelo eletrostático e 0,5 Å como constante de amortecimento. ..... 98

Figura 26 - Diedros amostrados para a parte alifática, durante uma simulação sem a utilização de um fato de correção jacobiano, para o ligante 4W55 (N-propilbenzeno). ..... 101

Figura 27 - Resultado das amostragens obtidas nas simulações. Em azul, vemos a amostragem, como uma distribuição de RMSD, para as simulações de DM. Em vermelho, está a amostragem, como uma distribuição de RMSD, para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Poisson, 250K. Em verde, está a amostragem, como um distribuição de RMSD, para as simulações de MC, utilizando *full_flex* e Poisson como modelo eletrostático. ..... 103

Figura 28 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Poisson, temperatura é 250K e amostragem é *full_flex*. ..... 106

Figura 29 - Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno). No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura. Em marrom, está o perfil gerado por DM; em vermelho o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Coulomb como modelo eletrostático, e em verde o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Poisson como modelo eletrostático; e em laranja está representado o perfil gerado por MC a 250K, usando *full_flex* como amostragem e Poisson como modelo eletrostático. ..... 107

Figura 30 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. ..... 109

Figura 31 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes..... 112

Figura 32 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem, considerando a dessolvatação, com $\alpha$ = 0,2 kcal/(mol.$e^{-2}$) e $\beta$ = -0,005 kcal/mol como parâmetros. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes..... 114

Figura 33 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem, considerando a dessolvatação, com $\alpha$ = 0,2 kcal/(mol.$e^{-2}$) e $\beta$ = -0,005 kcal/mol como parâmetros. Em azul estão representados os dados obtidos para ciclodextrina e em vermelho, os dados de lisozima de vírus T4. ..... 120

Figura 34 - Comparação entre as médias das contribuições das energias eletrostáticas utilizando os diferentes modelos. Azul, representa a média para a componente eletrostática utilizando Poisson como modelo para os quatro ligantes indicados e vermelho, representa a média para a componente eletrostática utilizando Coulomb como modelo. Aqui, a temperatura foi de 250K, a solvatação não foi considerada e não foi utilizado um potencial amortecido. ..... 123

Figura 35 - Deslocamento dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa os deslocamentos na coordenada x do espaço, amarelo representa os deslocamentos na coordenada y e marrom, em z..... 135

Figura 36 - Deslocamento dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. Na figura, azul representa os

deslocamentos na coordenada x do espaço, amarelo representa os deslocamentos na coordenada y e marrom, em z. ..... 136

Figura 37 - Rotação dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa rotações em α no espaço rotacional, amarelo representa rotações em β e marrom, as rotações em γ. ..... 137

Figura 38 - Rotação dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa rotações em α no espaço rotacional, amarelo representa rotações em β e marrom, as rotações em γ. ..... 138

Figura 39 - Comparação entre os valores experimentais e calculados das energias livres de ligação, entalpia e entropia para 250K, usando Coulomb como modelo eletrostático, sem considerar solvatação, usando 0,5 Å como constante de amortecimento. Nos gráficos, ΔH é a entalpia e -TΔS é a entropia. ..... 140

Figura 40 - Comparação entre os valores experimentais e calculados das energias livres de ligação, entalpia e entropia para 250K, usando Poisson como modelo eletrostático, sem considerar solvatação, usando 0,5 Å como constante de amortecimento. Nos gráficos, ΔH é a entalpia e -TΔS é a entropia. ..... 142

LISTA DE TABELAS


Tabela 1 - Descrição dos ligantes utilizados neste trabalho, juntamente com seus códigos PDB e energias livres experimentais. T4L representa lisozima de vírus T4. ..... 44

Tabela 2 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando *docking* molecular. ..... 54

Tabela 3 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando como base as simulações de Dinâmica Molecular. ..... 62

Tabela 4 - Tempo gasto para obter as coordenadas por meio da dinâmica molecular e para o cálculo das variações de energia livre a partir das coordenadas geradas. ..... 65

Tabela 5 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos. ..... 86

Tabela 6 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos, utilizando 0,5 Å como a constante de amortecimento no potencial de Lennard-Jones. ..... 89

Tabela 7 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos, utilizando os como parâmetros de solvatação $\alpha$ = 0,2 kcal/(mol.$e^{-2}$) e $\beta$ = -0,005 kcal/mol. ..... 93

Tabela 8 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Coulomb como modelo eletrostático, sem solvatação, com 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas. ..... 96

Tabela 9 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Poisson como modelo eletrostático, sem solvatação, com 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas. ..... 98

Tabela 10 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, utilizando Poisson como modelo eletrostático e *full_flex* como modelo de amostragem. ..... 111

Tabela 11 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, utilizando Poisson como modelo eletrostático, full_flex como modelo de amostragem, utilizando os como parâmetros de solvatação $\alpha$ = 0,2 kcal/(mol.$e^{-2}$) e $\beta$ = -0,005 kcal/mol..... 113

Tabela 12 - Comparação entre os valores de energia livre de ligação calculados por Steinbrecher e colaboradores[32] com os valores calculados pelo nosso programa. ..... 116

Tabela 13 - Comparação entre os valores de energia livre de ligação calculados por Minh e colaboradores[36] com os valores calculados pelo nosso programa. ..... 117

Tabela 14 - Comparação entre os valores de energia livre de ligação calculados por Purisima e colaboradores[70] com os valores calculados pelo nosso programa. ..... 118

Tabela 15 - Tempo gasto para obter as coordenadas por meio da dinâmica molecular e Monte Carlo e para o cálculo das variações de energia livre a partir das coordenadas geradas. ..... 128

Tabela 16 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Coulomb como modelo eletrostático, sem solvatação, usando 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas. ..... 139

Tabela 17 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Poisson como modelo eletrostático, sem solvatação, usando 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas. ..... 141

## LISTA DE ABREVIATURAS E SIGLAS

| | |
|---|---|
| AMBER | Assisted Model Building with Energy Refinement |
| DM | Dinâmica Molecular |
| FEP | Perturbação da Energia Livre (Free Energy Perturbation) |
| GAFF | Generalized AMBER force field |
| IT / TI | Integração Termodinâmica (Thermodynamic Integration) |
| LiBELa | Ligand Binding Energy Landscape |
| LIE | Linear Interaction Energy |
| MC | Monte Carlo |
| MCLiBELa | Monte Carlo-Based Ligand Binding Energy Landscape |
| MCM | Monte Carlo Metropolis |
| MM-GBSA | Molecular Mechanics – Generalized Born/Surface Area |
| MM-PBSA | Molecular Mechanics – Poisson-Boltzmann/Surface Area |
| PB | Poisson-Boltzmann |
| PDB | Protein Data Bank |
| RMSD | Raiz do Desvio Médio Quadrático |
| SF1 | Scoring Function 1 |
| SF2 | Scoring Function 2 |
| SF3 | Scoring Function 3 |

SUMÁRIO


**1 INTRODUÇÃO** ..... 25
**1.1 Interações Moleculares** ..... 26
**1.1.1 Termodinâmica** ..... 26
**1.1.2 Modelos Energéticos** ..... 29
**1.2 Equação de Poisson Boltzmann** ..... 30
**1.3 Dinâmica Molecular** ..... 32
**1.4 Monte Carlo** ..... 33
**1.4.1 Simulação de Monte Carlo** ..... 33
**1.5 Cálculo da Energia Livre de Ligação** ..... 34
**1.6 Entropia** ..... 37
**1.6.1 Aproximação de primeira ordem para entropia** ..... 38
**2 OBJETIVOS** ..... 41
**2.1 Objetivos Gerais** ..... 41
**2.2 Objetivos Específicos** ..... 41
**3 MATERIAIS E MÉTODOS** ..... 43
**3.1 Conjunto de teste** ..... 43
**3.2 Preparação das moléculas para simulação e cálculo das energias** ..... 47
**3.2.1 Dinâmica Molecular** ..... 47
**3.2.2 Monte Carlo** ..... 49
**3.3 Entropia** ..... 51
**4 RESULTADOS** ..... 53
**4.1 Dinâmica Molecular** ..... 56
**4.1.1 Amostragem** ..... 56
**4.1.2 Energia Livre de Ligação para DM** ..... 61
**4.2 Monte Carlo** ..... 65
***4.2.1 Sample Torsion*** ..... 66
**4.2.1.1 Amostragem** ..... 66
**4.2.1.2 Autocorrelação** ..... 77
**4.2.1.3 Convergência** ..... 82
**4.2.1.4 Energia Livre de Ligação para MC** ..... 85
**4.2.1.5 Aproximação de Primeira Ordem** ..... 95
***4.2.2 Full Flex*** ..... 99

**4.2.2.1 Amostragem** ..... 102
**5 DISCUSSÕES** ..... 115
**5.1 Comparação com outros autores** ..... 115
**5.2 Amostragem** ..... 120
**5.3 Efeito do Modelo Eletrostático** ..... 122
**5.4 Entalpia e Entropia de Ligação** ..... 124
**6 CONCLUSÃO** ..... 127
**REFERÊNCIAS** ..... 129
APÊNDICE A – Deslocamentos na Caixa de Simulação ..... 135
APÊNDICE B – Entropias e Entalpias ..... 139
APÊNDICE C – Input MCLiBELa ..... 145

# 1 INTRODUÇÃO

Os sistemas biológicos macromoleculares são bastante conhecidos por suas propriedades interagentes, as quais são essenciais para o funcionamento dos organismos. Podemos citar como exemplo de moléculas interagentes os hormônios, que são produzidos pelas glândulas e que precisam se ligar especificamente a um receptor num ambiente intracelular repleto de outras moléculas a fim de exercerem suas funções. A regulação da transcrição gênica também apresenta exemplos de sistemas interagentes, uma vez que ela está intimamente associada à capacidade do DNA de recrutar complexos proteicos que interagem em porções específicas do ácido nucléico. Para realizar uma catálise necessária à homeostase, uma enzima precisa se associar a seu substrato de forma transiente.[1]

A capacidade de uma proteína se ligar a diferentes ligantes ou mesmo a diferentes proteínas com alta especificidade é essencial para diversos processos biológicos e, portanto, caracterizar as energias dos complexos moleculares bem como as suas estruturas torna-se fundamental para compreendermos a função biológica.[2] Nesse contexto, as interações de alta especificidade são tipicamente caracterizadas pelo reconhecimento de alta afinidade que um componente do sistema tem por outro.[3]

Ao conhecer a afinidade entre os complexos biomoleculares, que será definida na seção a seguir, é possível ter uma descrição mais detalhada do sistema no que tange a sua composição química, interação entre seus componentes, suas estruturas e funções, que são alvos de interesse de setores relacionados à saúde como o *design* molecular, importante para o desenvolvimento de novos fármacos, ajudando também a reduzir os efeitos colaterais de um fármaco e possível resistência a ele.[4]

Entretanto, embora ensaios laboratoriais sejam bastante utilizados para obtenção dessas grandezas, a estimativa computacional ainda é um grande desafio, apesar dos avanços na determinação de estruturas cristalográficas que permitiu um aumento exponencial no número de estruturas depositadas no *Protein Data Bank* (PDB).[3] O gasto computacional e o tempo para realizar estas avaliações tipicamente

aumentam exponencialmente à medida que o número de átomos e dos graus de liberdade do sistema crescem.

Para as biomoléculas, assumindo que todas as interações microscópicas são conhecidas, os métodos de simulação baseados em dinâmica molecular (DM) ou Monte Carlo (MC) podem ser aplicados a fim de gerar um *ensemble* termodinâmico. Contudo, a obtenção da energia livre de interação entre receptor e ligante a partir deste *ensemble* ainda representa um grande desafio.

Dessa forma, para melhor elucidar os processos das interações biomoleculares e compreender seu papel na execução das ferramentas computacionais que auxiliam na obtenção das entidades termodinâmicas, na próxima seção, faremos uma breve revisão sobre a termodinâmica envolvida em tais processos.

## 1.1 Interações Moleculares

### 1.1.1 Termodinâmica

A afinidade de um ligante *A* a um receptor *B* formando um complexo *AB* pode ser descrita no equilíbrio como:

$$A + B \rightleftharpoons AB \tag{1.1}$$

$$K_a = \frac{[AB]}{[A][B]} \tag{1.2}$$

onde $K_a$ é a constante de afinidade, *[A]*, *[B]* e *[AB]* são as concentrações no equilíbrio de ligante, receptor e complexo, respectivamente. A grandeza termodinâmica diretamente associada à afinidade é a variação da energia livre, por mol, $\Delta G$, associada à interação:

$$\Delta G = -RT\ln K_a c_{conc} \tag{1.3}$$

onde R é a constante dos gases ideais, *T* é a temperatura e $c_{conc}$ = 1 mol.L$^{-1}$ é uma constante para correção da unidade. Alternativamente, uma vez que $K_a$ é igual a $1/K_d$ sendo $K_d$ a constante de dissociação, a equação 1.3 pode ser reescrita da forma

$$\Delta G = RT\ln K_d c_{conc} \quad (1.4)$$

Essa variação na energia livre, que como dito anteriormente, está associada à interação entre os componentes do complexo, pode ser expressa em termos da energia livre de *Helmholtz (F)* (equação 1.6) - que melhor descreve sistemas com número de partículas, temperatura e volume constantes (*ensemble* NVT) - ou em termos da energia livre de *Gibbs (G)* - mais apropriada para sistemas que contenham número de partículas, temperatura e pressão constantes (*ensemble* NPT).[3-4] Em termos práticos, ambas as energias têm o mesmo significado físico, o que difere uma da outra são as condições termodinâmicas nas quais elas são obtidas.

Em termos de pressão constante, podemos escrever a variação da energia livre de Gibbs associada a algum processo termodinâmico da seguinte maneira:

$$\Delta G = P\Delta V + \Delta U - T\Delta S \quad (1.5)$$

onde $\Delta$U é a variação da energia interna do sistema interagente, $\Delta$S é a variação de sua entropia, *P* é a pressão, $\Delta V$ é a variação do volume e *T* é a temperatura. Uma vez que a variação de volume $\Delta$V é nula para o caso da energia livre de *Helmholtz*, a variação desta grandeza devida à interação entre os componentes do sistema, pode ser dada por

$$\Delta F_{int} = \Delta U - T\Delta S \quad (1.6)$$

Se considerarmos que a energia interna do sistema U é uma grandeza macroscópica, então ela pode ser determinada pela média sobre os estados microscópicos

$$U = \langle E \rangle = \sum_{i=1}^{N} p_i E_i \quad (1.7)$$

onde $p_i$ é a probabilidade no estado *i*, com energia $E_i$.

Também consideramos a entropia S uma entidade macroscópica que, assim como a energia interna *U*, poderá ser definida com base em uma probabilidade dos estados microscópicos

$$S = \sum_{j=1}^{N} p_j \ln p_j \quad (1.8)$$

Aqui, a entropia é descrita de forma simples como a soma da contribuição de cada estado *j* com a probabilidade de ocorrência $p_j$.

Assim, para o caso geral, onde a variação de volume é negligenciável, podemos escrever a energia livre de Gibbs, ou Helmholtz, como

$$\Delta G \approx \Delta F = \Delta\langle E\rangle - T\Delta S \quad (1.9)$$

Isso significa que para calcular a variação da energia livre de ligação é necessário estimar a diferença de energia entre os estados ligado e não-ligado do sistema bem como suas entropias correspondentes. Entretanto, tais estimativas não são triviais. A dificuldade em realizá-las encontra-se no fato de que os métodos que costumeiramente calculam essas grandezas (dinâmica molecular ou Monte Carlo) têm um caráter dinâmico.[5] Esses métodos permitem que se amostre $N_i$ configurações de um sistema de acordo com a probabilidade de *Boltzmann*, $p_i^B$, a saber,

$$p_i^B = \frac{\exp[-\frac{E_i}{k_B T}]}{Z} \quad (1.10)$$

Na equação (1.10), a probabilidade $p_i^B$ de se observar um estado *i*, de energia $E_i$, tem um peso de *Boltzmann*, onde $k_B$ é a constante de *Boltzmann*, *T* é a temperatura e *Z* é a função de partição do sistema, dada por:

$$Z = \sum_i \exp[-\frac{E_i}{k_B T}] \quad (1.11)$$

Um dos problemas está em calcular a função de partição a partir de uma amostragem finita, embora *Z* seja definida sobre todo o espaço de fase. Além disso, a acurácia na determinação de cada estado interfere diretamente na geração da função de partição, bem como uma amostragem inadequada também pode resultar em problemas para obter a função de partição.

De forma geral, pela equação (1.9), é possível compreender que dois dos desafios associados à estimativa da energia de ligação são: *(i)* a definição precisa da energia de cada estado *i* do sistema, i.e., a definição de $E_i$; e *(ii)* a estimativa da entropia *S* e de sua variação ($\Delta S$) associada à interação. Obviamente, a fim de estimar precisamente as duas grandezas, é necessária a geração de um conjunto de

conformações que represente um conjunto de configurações termicamente acessíveis ao sistema em uma temperatura de referência *T*. Algumas das abordagens empregadas para contornar estes desafios, bem como para a geração destes *ensembles*, ou conjunto de configurações, serão descritas nas seções abaixo.

### 1.1.2 Modelos Energéticos

A definição da energia potencial *U(r)* associada à configuração atômica de uma determinada molécula pode ser obtida a partir do modelo baseado em campo de força.[6] Um campo de força é uma expressão matemática que descreve a energia do sistema em função das coordenadas das partículas desses sistemas.[7] Basicamente, um campo de força é composto por uma forma analítica da energia potencial interatômica, $U(\boldsymbol{r_1}, \boldsymbol{r_2}, \ldots, \boldsymbol{r_N})$, e um conjunto de parâmetros de potenciais e topologias para essa forma. Geralmente, esses parâmetros são obtidos por cálculos de mecânica quântica (semi-empíricos ou *ab initio*) ou mesmo por meio de dados experimentais de difração de nêutrons, raios-x e elétrons ou por meio de espectroscopias como RMN, Raman e infravermelho.[7]

Os principais papéis dos campos de força são descrever razoavelmente as propriedades gerais das moléculas, como a sua geometria, energias conformacionais e estereoisométricas, barreiras torcionais, deformação por torção e energia de interações intermoleculares; acessar a geometria entre moléculas interagentes; e avaliar frequência vibracional e calor de formação.[8]

Existem diversos campos de força disponíveis na literatura e que são amplamente utilizados para realização de simulações computacionais, como OPLS,[9] CHARMM,[10] GROMOS,[11] e AMBER.[12] Uma forma bastante comum de se representar um campo de força para um sistema atômico é:

$$U = \sum_{Ligações} k_b\,(r - r_0)^2 + \sum_{Ângulos} k_\theta(\theta - \theta_0)^2 \qquad (1.12)$$

$$+ \sum_{Torções} k_t(1 + \cos(n\omega - \gamma)) + E_{não-ligado}$$

onde $k_b$, $k_\theta$ e $k_t$ são as constantes elásticas para os termos harmônicos associados a ligações, ângulos e torções, respectivamente. $r_0$ é o comprimento da ligação no

equilíbrio, $\theta_0$ é o ângulo da ligação no equilíbrio, $\omega$ é o ângulo de diedro, $\gamma$ é a fase do ângulo de diedro e $n$ é a periodicidade.

O potencial $E_{não\text{-}ligado}$ corresponde às interações não covalentes entre o ligante e o receptor. Para isso, basicamente são considerados os potenciais de Lennard-Jones típico para a modelagem matemática de moléculas[6] e o potencial eletrostático de Coulomb:

$$E_{não-ligado} = \sum_{i \neq j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \sum_{i \neq j} \frac{C q_i q_j}{\epsilon r_{ij}} \quad (1.13)$$

onde C é uma constante de conversão dimensional, $q_i$ e $q_j$ são as cargas parciais dos átomos $i$ e $j$, $r_{ij}$ é a distância entre os átomos $i$ e $j$, A e B são os parâmetros de Lennard-Jones para os átomos $i$ e $j$, $\epsilon$ é a permissividade relativa.

De forma geral, os campos de força de mecânica molecular são bastante apropriados para a definição de energia potencial associada à configuração atômica (U(r)) em simulações como Monte Carlo (MC) e simulações de dinâmica molecular (DM) e são empregadas com sucesso nestes contextos, além do contexto do *docking*, por exemplo.[6]

## 1.2 Equação de Poisson Boltzmann

No contexto de aproximação contínua da matéria, é possível utilizar a equação de Poisson-Boltzmann para o cálculo do potencial eletrostático associado a qualquer molécula. Nesse modelo, moléculas e solvente são tratadas como um meio que reage linear e uniformemente ao campo eletrostático gerado por alguma fonte, que, tipicamente, é uma distribuição de cargas nas próprias moléculas. No geral, todos os átomos pertencentes a uma molécula têm a mesma constante dielétrica, que é baixa, e reagem com a mesma intensidade a um campo elétrico pré-existente.[13]

A equação de Poisson-Boltzmann combina a distribuição de Boltzmann pra uma densidade de carga $\rho$ com uma equação eletrostática, a lei de Gauss, ou equação de Poisson, a qual relaciona a segunda derivada do potencial eletrostático $\Phi$ para a densidade de carga. Assim, para a posição $x$:

$$\nabla \cdot [\epsilon(x)\nabla\Phi(x)] = -4\pi\rho_{soluto}(x) \quad (1.14)$$

onde $\rho_{soluto}(x)$ é densidade de carga do soluto, em função da posição, $\epsilon(x)$ é a função dielétrica dependente da posição e $\Phi(x)$ é o potencial eletrostático dependente de $x$.[6]

Para um solvente polar, a constante dielétrica efetiva cresce à medida que $x$ se distancia do soluto. Essa função dielétrica que depende da posição, a qual é baixa próxima ao soluto e alta para o solvente, permite uma melhor descrição do que a lei de Coulomb, uma vez que ela leva em consideração a diferença na polarizabilidade entre macromoléculas e o solvente.[6]

Na presença de íons em solução, a densidade de carga não fica mais localizada apenas na fronteira soluto/solvente. Portanto, uma melhor aproximação é feita considerando a densidade de carga ao redor do soluto resultante da distribuição de cargas no meio. Assim, essa densidade pode ser escrita como:

$$\rho(x) = \sum_{i=1}^{n_i} q_i c_i \exp\left[-\frac{\tilde{E}_i(x)}{k_B T}\right] \quad (1.15)$$

onde $q_i$ é a carga do íon da espécie $i$, $c_i$ é a concentração do íon $i$ em solução, $k_B$ é constante de *Boltzmann*, $T$ é a temperatura e $\tilde{E}_i(x)$ é a energia necessária para trazer o íons do infinito até a posição $x$.[6]

Em termos práticos, assume-se que $\tilde{E}_i(x)$ é o produto entre a carga e o potencial, isto é, a distribuição de íons é determinada pelo campo eletrostático, de modo que:

$$\tilde{E}_i(x) \approx q_i\Phi(x) \quad (1.16)$$

Combinando as equações (1.14), (1.15) e (1.16), chegamos à equação de *Poisson-Boltzmann* (PB):

$$\nabla \cdot [\epsilon(x)\nabla\Phi(x)] = -4\pi\rho_{soluto}(x) - 4\pi \sum_{i=1}^{n_i} q_i c_i \exp[-\frac{q_i\Phi(x)}{k_B T}] \quad (1.17)$$

Para sistemas pouco carregados, a equação de PB geralmente é linearizada. Expandindo o fator de Boltzmann em Taylor e truncando na primeira ordem, temos

$$\exp\left[-\frac{q_i\Phi(x)}{k_BT}\right] \approx 1 - \frac{q_i\Phi(x)}{k_BT} \quad (1.18)$$

sendo essa versão linearizada válida para quando $q_i\Phi(x) \ll k_BT$. Assumindo uma simetria esférica, temos a forma linearizada da equação PB:

$$\nabla \cdot [\epsilon(r)\nabla\Phi(r)] = \frac{1}{\epsilon}[\rho_{soluto}(r) - \frac{\rho_{solvente}\kappa^2}{4\pi}\Phi(x)] \quad (1.19)$$

onde $\rho_{solvente}$ é a densidade do solvente e κ² é dada por:

$$\kappa^2 = \frac{8\pi e^2 c_s}{\epsilon k_BT} \quad (1.20)$$

sendo $c_s$ é concentração iônica, *e* é a carga do próton e *ϵ* é a constante dielétrica do solvente.[6]

O modelo de Poisson-Boltzmann permite que o solvente aquoso seja tratado como um meio implícito e contínuo, de alta constante dielétrica, enquanto as macromoléculas são consideradas cavidades com baixa constante dielétrica. Além disso, ele modela o sistema no equilíbrio.[14] De forma geral, o emprego do modelo de Poisson-Boltzmann (PB) permite uma melhor estimativa do papel do solvente nas interações polares, bem como pode permitir uma estimativa mais robusta do potencial eletrostático (*φ(x)*) ao estimar um meio dielétrico continuamente variável.

## 1.3 Dinâmica Molecular

As simulações de dinâmica molecular (DM) são técnicas computacionais de geração de *ensembles* que consistem de soluções numéricas, passo-a-passo, para as equações clássicas de movimento (equações de Newton)[15] Para um sistema atômico simples podemos escrever

$$\boldsymbol{f}_i = m_i\ddot{\boldsymbol{r}}_i \quad (1.21)$$

onde $\boldsymbol{f}_i$ é a força resultante, $m_i$ é a massa do átomo e $\ddot{\boldsymbol{r}}_i$ sua aceleração. Em termos de energia potencial *U*, podemos escrever:

$$\boldsymbol{f}_i = -\frac{\partial}{\partial \boldsymbol{r}_i} U \tag{1.22}$$

Todas as forças $\boldsymbol{f}_i$ que atuam sobre os átomos precisam ser calculadas. Estas forças derivam de uma energia potencial $U(\boldsymbol{r}^N)$, onde $\mathbf{r}^N = (r_1, r_2, \ldots, r_N)$ representa um sistema atômico de 3N coordenadas. Assim, temos

$$\boldsymbol{F}(\boldsymbol{r}^N) = -\vec{\nabla} U(\boldsymbol{r}^N) \tag{1.23}$$

Esse cálculo é possível uma vez que os potenciais energéticos são conhecidos, isto é, são determinados a partir dos campos de força. Não existe solução analítica para esse problema, de modo que as equações devem ser resolvidas numericamente por meio de algoritmos como o de *Verlet,*[16] *velocity Verlet,*[17] *leapfrog*[18] e *Beeman,*[19] dentre outros.

A partir de uma simulação de dinâmica molecular é possível gerar configurações para um estado termodinâmico de um sistema de moléculas e, com essas informações, analisar as entidades termodinâmicas de interesse, que para os nossos propósitos, são as energias livres de ligação do sistema.

## 1.4 Monte Carlo

### 1.4.1 Simulação de Monte Carlo

O método de Monte Carlo é baseado na amostragem aleatória de amostras ou parâmetros para analisar sistemas ou processos complexos de se resolver analiticamente, respeitando uma distribuição estatística.[20] Apesar dos primeiros registros do uso desse método datarem dos séculos XVIII e XIX para o cálculo mais preciso da constante π, o seu uso tornou-se muito mais frequente com o advento dos computadores, popularizado próximo à década 50 por von Neumann, Ulam e Metropolis.[21-22]

Em mecânica estatística, para o método de Monte Carlo gerar um conjunto aleatório apropriado de estados de acordo com a distribuição de Boltzmann, costuma-se usar um *processo de Markov*, onde, dado um sistema num estado μ, é possível obter um novo estado ν do sistema. A probabilidade de gerar o estado *ν*

dado *μ* é chamada de *probabilidade de transição* – de μ para ν – *P(μ→ν)*. Essa transição deve ser invariável no tempo e deve depender apenas do estado anterior.[20]

Para avaliar se essa transição pode ou não acontecer, um algoritmo bastante empregado é o algoritmo de Metropolis.[21] Esse algoritmo baseia-se em escolher configurações *i* com uma probabilidade $e^{-E_i/k_BT}$ e ponderá-las uniformemente, onde $E_i$ é a energia na configuração *i*, $k_B$ é a constante de *Boltzmann* e *T* é a temperatura. Por exemplo, um sistema de N partículas numa dada configuração, na qual é possível deslocar cada uma dessas partículas da posição *R até R+αδ*$_i$, onde α é o deslocamento máximo e $\delta_i$ é um número arbitrário entre -1 e 1. A cada mudança de posição a variação de energia do sistema ΔE é recalculada e se ΔE < 0 o novo valor é aceito. Se ΔE > 0, a mudança é permitida com probabilidade $e^{-E/k_BT}$, isto é, sorteando um número aleatório entre 0 e 1. Se esse número for menor que a probabilidade, aceita-se o novo valor, caso contrário, retorna-se ao valor anterior.[21]

A vantagem no Método de Monte Carlo empregado com o critério de Metropolis é que não é necessário determinar a função de partição *Z* para o sistema, uma vez que apenas as probabilidades relativas dos diferentes estados, dependentes da mudança de energia entre eles, é que importam.

## 1.5 Cálculo da Energia Livre de Ligação

Uma vez definidos os métodos de geração de *ensembles* e definida a métrica para o cálculo de energias para cada configuração, i.e., *U(r)*, empregando um campo de força, por exemplo, a etapa seguinte envolve a estimativa da energia livre para o sistema em estudo. Os métodos mais rápidos para a estimativa da energia livre de ligação são os chamados de empíricos ou baseados em conhecimento (estatístico). Eles se baseiam, respectivamente, em funções de energia simples ou na frequência em que ocorrem interações átomo-átomo entre o complexo ligante-receptor (para estruturas conhecidas).[2] Usar uma função simples para energia junto com a falta de amostragem conformacional e a abordagem dada ao solvente faz com que esses métodos sejam bastante rápidos, porém, comprometendo sua acurácia. Nestes casos, a energia de interação é tipicamente modelada a partir de funções de energia baseadas em campos de força de mecânica molecular e a estimativa da energia

(potencial) de interação é baseada numa busca não exaustiva de uma pose, i.e., uma conformação e orientação do ligante que minimize a energia de interação. Um exemplo de aplicação destes métodos pode ser visto no programa LiBELa, dedicado ao *docking* de ligantes, e desenvolvido em nosso grupo.[22]

Além dos métodos empíricos, os métodos conhecidos como *'end-point'*, nos quais os pontos finais do ciclo termodinâmico de interação são simulados, também podem ser utilizados para o cálculo da energia livre de ligação. Um exemplo é o método de LIE (*Linear Interaction Energy)*, o qual considera que a variação de energia livre varia linearmente com os potenciais eletrostáticos e de van der Waals, considerando que a mudança na energia livre se dá quando há a transferência do ligante em solução para o sítio de ligação solvatado[23]. Dessa forma, é necessário fazer simulações do sistema proteína-ligante solvatado e do ligante sozinho em solução.[3] Assim, a energia livre de ligação pode ser escrita como:

$$\Delta G_{ligação} = \beta\left(\langle V^{ele}_{proteína-ligante}\rangle - \langle V^{ele}_{ligante-solvente}\rangle\right) \quad (1.24)$$

$$+ \alpha\left(\langle V^{vdw}_{proteína-ligante}\rangle - \langle V^{vdw}_{ligante-solvente}\rangle\right) + \gamma$$

onde ⟨ ⟩ indica a média do *ensemble*, α e β são coeficientes ponderados para as contribuições apolares e eletrostáticas, respectivamente. $V^{ele}_{proteína-ligante}$ representa a contribuição eletrostática do complexo proteína-ligante, $V^{ele}_{ligante-solvente}$ representa a contribuição eletrostática para o ligante em solução, $V^{vdw}_{proteína-ligante}$ e $V^{vdw}_{ligante-solvente}$ representam os respectivos análogos para a contribuição apolar.

Outro método *'end-point'* é o MM-PBSA (*Molecular Mechanics/Poisson-Boltzmann/Surface Area*), o qual analisa as trajetórias da DM usando uma aproximação de solvente contínuo. A variação de energia é então calculada como:[25-26]

$$\Delta G_{ligação} = G_{comp} - \left(G_{rec} + G_{lig}\right) = \Delta H - T\Delta S \approx \Delta E_{MM} + \Delta G_{sol} - T\Delta S \quad (1.25)$$

$$\Delta G_{sol} = \Delta G^{PB}_{polar} + \Delta G_{apolar} \quad (1.26)$$

$$\Delta G_{apolar} = \gamma \Delta A + b \quad (1.27)$$

Aqui, T é a temperatura do sistema. $\Delta E_{MM}$ engloba as variações de energia interna, eletrostática e de van der Waals. $\Delta E_{interna}$ inclui os termos de energia das ligações intramoleculares, ângulos e torções. $\Delta G_{polar}^{PB}$ é a energia livre de solvatação e é obtida resolvendo a equação de Poisson-Boltzmann, apresentada na seção 1.2. $\Delta G_{apolar}$ é estimada baseada na área de superfície acessível ao solvente, $\Delta A$, $\gamma$ é a tensão superficial e *b* é um termo de correção.

Um método mais rápido que o MM-PBSA, porém, que parte do mesmo princípio, é o MM-GBSA (*Molecular Mechanics Generalized Born*), o qual difere do método anterior no cálculo da energia livre de solvatação, a qual é uma linearização da equação de Poisson-Boltzmann, sendo definida pela relação de Still:[26-29]

$$G_{polar}^{GB} = -\frac{1}{8\pi}\left(\frac{1}{\epsilon_0} - \frac{1}{\epsilon}\right)\frac{q_i q_j}{f_{GB(r_{ij})}} \tag{1.28}$$

$$f_{GB(r_{ij})} = \left(r_{ij}^2 + \alpha_i \alpha_j exp\left[-\frac{r_{ij}^2}{4\alpha_i \alpha_j}\right]\right)^{1/2} \tag{1.29}$$

onde $\epsilon_0$ é a permissividade elétrica no vácuo, $\epsilon$ é a permissividade do meio, $q_i$ e $q_j$ são as cargas dos átomos *i* e *j*, $r_{ij}$ é a distância entre eles, $\alpha_i$ e $\alpha_j$ são os raios efetivos de Born correspondentes, correspondendo ao quanto o átomo penetra na superfície do receptor.

Além dos métodos empíricos e os *'end-points'* existem métodos ainda mais rigorosos, que também se baseiam em campo de força e são conhecidos como métodos de alquimia, baseados na equação de *Zwanzig*[29] para estimar a diferença de energia livre entre dois estados A e B:

$$\Delta G = G_B - G_A = -\beta^{-1}\ln \langle e^{-\beta \Delta V} \rangle_A \tag{1.30}$$

onde β = 1/kT e $\langle\ \rangle_A$ é a média (tanto para MC quanto DM) do *ensemble* para ΔV = $V_B$ - $V_A$ amostrada usando o potencial $V_A$ como referência, $G_A$ é a energia no estado *A* e $G_B$ é a energia no estado *B*.

Um desses métodos é a ***Perturbação da Energia Livre ou Free Energy Perturbation (FEP)***, no qual é possível quebrar o cálculo de energia livre em muitas partes, cada uma envolvendo um intervalo pequeno o bastante entre um estado de

energia e seu subsequente, o que permite que a energia livre seja calculada com grande confiabilidade, por meio de uma somatória.[30]

Outra abordagem é a ***Integração Termodinâmica ou Thermodynamic Integration (TI)***, a qual faz a diferencial da energia potencial ΔV com relação ao intervalo entre um estado e outro. Assim a média do *ensemble* dessa diferencial da variação de energia com respeito ao intervalo é avaliada para vários valores do tal intervalo e então, por meio de integração numérica, é possível calcular a variação de energia ΔG, como mostram Brandsdal e colaboradores e Kollman.[31-32]

Embora FEP e TI possam ter uma acurácia bastante elevada, o gasto computacional também é bastante alto, tipicamente envolvendo dezenas de simulações para amostrar o caminho ao longo de janelas com valores de λ distintos. Em contrapartida, estes métodos representam o estado da arte para o cálculo de energia livre, com os melhores resultados obtidos, usualmente alcançando precisão entre 0,6 – 1 kcal/mol.[32-33]

Uma pergunta neste contexto seria: é possível ter um método relativamente rápido e barato computacionalmente e que preserve o rigor termodinâmico necessário para estimativas precisas da energia de ligação entre receptor e ligante? É exatamente nessa lacuna em que entra o método proposto neste trabalho, o qual, como poderá ser visto mais adiante, estima os componentes entálpicos e entrópicos da energia livre de ligação diretamente, de uma forma mais precisa possível, porém, com o custo computacional menos elevado.

## 1.6 Entropia

A estimativa da entropia ($\Delta S$) é um grande desafio em modelagem de sistemas biológicos.[5] Como apresentado por Edholm e Berendsen,[33] a distribuição multidimensional das variáveis do sistema compromete o cálculo da entropia. Além dos problemas que se tem para avaliar as quantidades termodinâmicas associadas à ligação, amostrar os estados termodinamicamente acessíveis e os microestados relevantes é por si só um problema para os métodos que calculam energia livre.

Uma maneira bastante comum que se utiliza é estimar a entropia configuracional através de uma aproximação quase harmônica, a qual permite fazer essa estimativa baseada em simulações de dinâmica molecular.[34] De maneira geral,

essa abordagem envolve utilizar as simulações de DM, calcular uma matriz de covariância das coordenadas atômicas, interpretar as covariâncias como superfícies harmônicas de energia em vez das próprias superfícies enarmônicas de energia, e calcular as constantes de força efetivas, ou quase harmônicas, nessa nova base. Essas constantes de forças efetivas são utilizadas para o cálculo das entidades termodinâmicas, como a entropia, por exemplo.[35]

No geral, esse tipo de abordagem para o cálculo da entropia não é muito eficiente tanto no conceito quanto na execução, uma vez que ela não é capaz de reproduzir as integrais configuracionais se o perfil da energia for enarmônico ou se apresentar sobreposição de poços de potencial, de modo que ela se torna numericamente instável e computacionalmente cara.[36]

Outra estratégia para a estimativa da entropia é baseada numa aproximação de primeira ordem. Esta estratégia, empregada neste trabalho, será apresentada na seção a seguir.

### 1.6.1 Aproximação de primeira ordem para entropia

Uma aproximação razoável para contornar a dificuldade de estimar as entropias envolve considerar os graus de liberdade gerados como uma distribuição descorrelacionada de modo que a entropia associada a cada grau *i* do sistema pode ser determinada como uma entropia de Shannon para cada grau de liberdade[37]

$$\frac{S_i}{k_B T} = -\int \rho(r_i) \ln(\rho_i(r_i)) dr \tag{1.33}$$

onde $\rho(r_i)$ é a função de probabilidade dada uma configuração $(r_i)$ para um sistema. Nessa aproximação, chamada de “aproximação de primeira ordem”, proposta por Edholm e Berendsen[33] e recentemente avaliada por Killian e colaboladores,[37] a entropia total é dada pela soma de todas $S_i$, isto é:

$$\frac{S}{k_B T} = \sum_i S_i \tag{1.34}$$

Essa abordagem tem um custo menor em termos de poder computacional uma vez que ela se baseia na determinação da função probabilidade de apenas um

grau de liberdade por vez e pode ser estimada diretamente a partir das trajetórias das simulações de DM e MC.

# 2 OBJETIVOS

## 2.1 Objetivos Gerais

Este projeto visou aprimorar e utilizar métodos com baixo custo computacional para realização do cálculo da variação da energia livre de ligação em sistemas macromoleculares interagentes (proteína-ligante).

## 2.2 Objetivos Específicos

Para esse trabalho, os seguintes objetivos específicos também foram considerados:

1. Aprimoramento e validação do método de Monte Carlo Metropolis desenvolvido pelo grupo de Biotecnologia Molecular para a geração de *ensembles* termodinâmicos e comparação com *ensembles* gerados por dinâmica Molecular.
2. Comparação dos modelos eletrostáticos de Coulomb e Poisson-Boltzmann para geração de *ensembles* e para o cálculo das variações de energia livre.
3. Determinação da variação da energia livre de ligação para complexos proteína-ligante do sistema modelo lisozima T4 e comparação dos valores obtidos com resultados experimentais e resultados obtidos por MM-GBSA e MM-PBSA.

# 3 MATERIAIS E MÉTODOS

## 3.1 Conjunto de teste

Para realização dos testes com o método desenvolvido como uma possível ferramenta que calcule a variação da energia livre de interação entre receptor e ligante, um conjunto de estruturas foi selecionado. Esse conjunto consiste em 30 diferentes complexos de lisozima mutante do bacteriófago T4, contendo diferentes ligantes que foram cocristalizados com a enzima.[4,39-43]

O conjunto de complexos escolhidos apresenta receptores que possuem diferentes tipos de mutação em seu sítio de ligação: complexos com apenas a mutação L99A (de leucina, na posição 99, para alanina), que permite a existência de uma cavidade apolar; complexos que, além da L99A, apresentam a mutação M102Q (metionina, na posição 102, para glutamina), que confere característica polar, capaz de fazer ligações de hidrogênio, no sítio de ligação; e um terceiro conjunto que apresenta, além de L99A, a mutação M102E (metionina, na posição 102, para glutamato), que adiciona um resíduo carregado à cavidade.

Tais complexos foram os escolhidos por serem comumente utilizados como forma de avaliar métodos computacionais que analisam energias entre ligantes e receptores. A mutação L99A cria uma cavidade interna com um volume de aproximadamente 150 $Å^3$ e essa cavidade torna-se um alvo de estudo de interação entre proteína-ligante bastante interessante.[40]

Todas as simulações foram realizadas a partir das estruturas cristalográficas disponíveis no PDB, sendo os valores de energia livre retirados da literatura específica. Os códigos PDB das moléculas utilizadas, bem como nome, tipo de mutação que o sítio de interação apresenta e energias livres de ligação, retiradas de trabalhos da literatura que apresentam os valores experimentais, estão apresentados na Tabela 1.

Tabela 1 - Descrição dos ligantes utilizados neste trabalho, juntamente com seus códigos PDB e energias livres experimentais. T4L representa lisozima de vírus T4.

| # | PDB ID | Ligante | Mutação | Energia Livre de Ligação Experimental (kcal/mol) |
|---|---|---|---|---|
| | | **T4L L99A** | | |
| **1** | 182L | Benzofurano | L99A | -5,46 |
| **2** | 183L | Indeno | L99A | -5,13 |
| **3** | 184L | Isobutilbenzeno | L99A | -6,51 |
| **4** | 185L | Indol | L99A | -4,89 |
| **5** | 186L | N-butilbenzeno | L99A | -6,70 |
| **6** | 187L | p-Xileno | L99A | -4,67 |
| **7** | 188L | o-Xileno | L99A | -4,60 |
| **8** | 1NHB | Feniletano | L99A | -5,76 |
| **9** | 3DMX | Benzeno | L99A | -5,35 |
| **10** | 4W52 | Benzeno | L99A | -5,19 |
| **11** | 4W53 | Tolueno | L99A | -5,52 |
| **12** | 4W55 | N-propilbenzeno | L99A | -6,55 |
| **13** | 4W57 | N-butilbenzeno | L99A | -6,70 |
| | | **T4L L99A/M102Q** | | |
| **14** | 1LGW | 2-Fluoroanilina | L99A/M102Q | -5,53 |
| **15** | 1LI2 | Fenol | L99A/M102Q | -5,58 |
| **16** | 1LI3 | 3-Clorofenol | L99A/M102Q | -5,87 |
| **17** | 1LI6 | 5-methilpirrol | L99A/M102Q | -5,25 |
| **18** | 1XEP | Catecol | L99A/M102Q | -4,16 |
| **19** | 2RBO | 2-Nitrotiofeno | L99A/M102Q | -5,52 |
| **20** | 3HT8 | 5-Cloro-2-metilfenol | L99A/M102Q | -5,42 |
| **21** | 3HTB | 2-Propilfenol | L99A/M102Q | -5,31 |
| **22** | 3HU8 | 2-Etoxifenol | L99A/M102Q | -4,02 |
| **23** | 3HUA | 4,5,6,7-tetra-hidroindol | L99A/M102Q | -4,61 |
| **24** | 3HUK | Benzilacetato | L99A/M102Q | -4,48 |
| | | **T4L L99A/M102E** | | |
| **25** | 3GUJ | Benzeno | L99A/M102E | -5,80 |
| **26** | 3GUK | Tolueno | L99A/M102E | -5,30 |
| **27** | 3GUL | Etilbenzeno | L99A/M102E | -5,50 |
| **28** | 3GUM | p-Xileno | L99A/M102E | -4,90 |
| **29** | 3GUN | Anilina | L99A/M102E | -4,00 |
| **30** | 3GUO | Fenol | L99A/M102E | -5,00 |

Fonte: Elaborada pelo autor.

A Figura 1 mostra cada uma das moléculas de ligantes utilizados para as simulações, bem como o código PDB para as estruturas dos complexos e os valores das energias livres de ligação experimentais, determinadas por meio de calorimetria de titulação isotérmica (ITC).

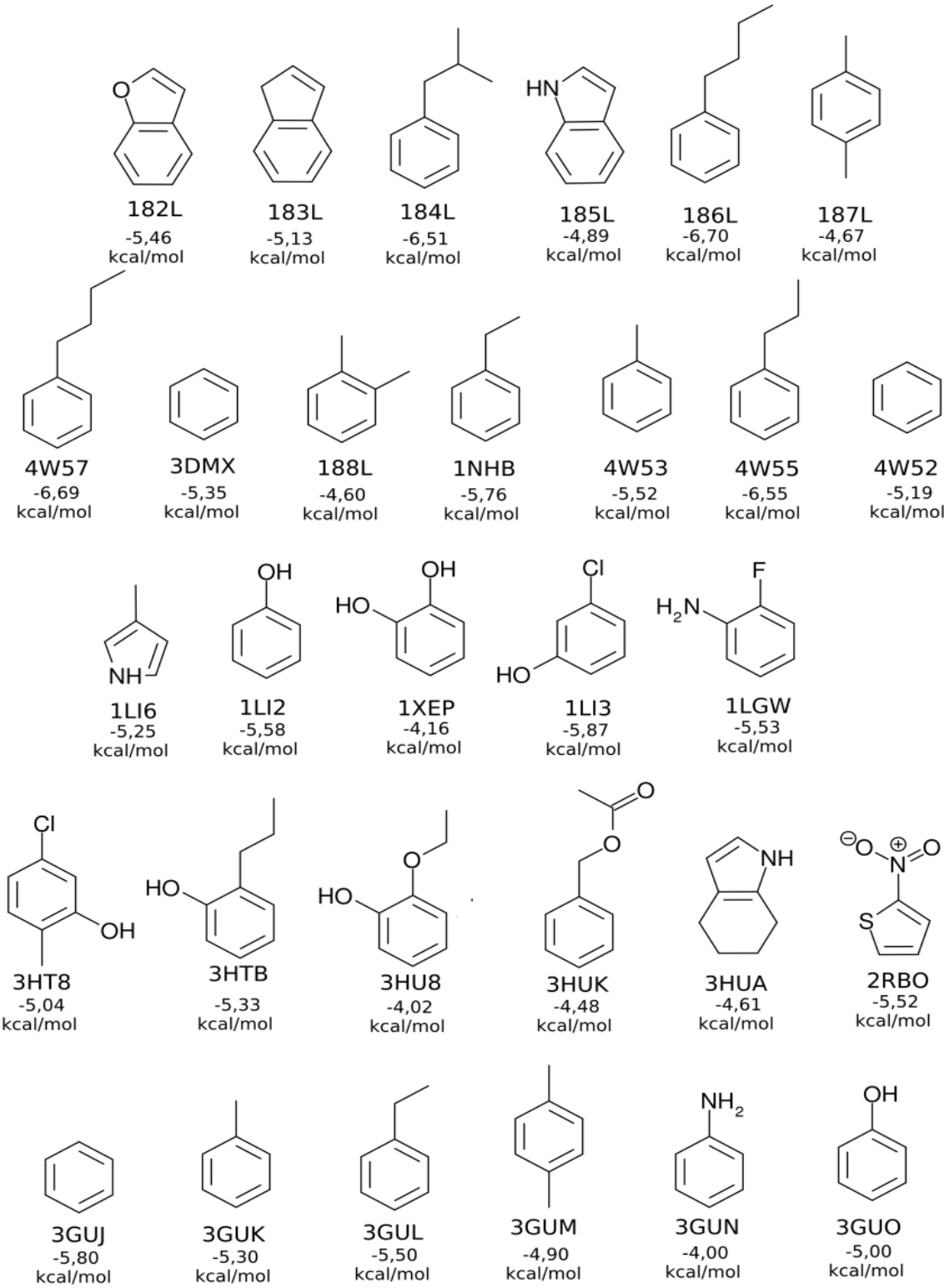



Figura 1 - Ligantes das enzimas T4 mutantes, com suas energias livres de ligação, escolhidos para as simulações.

Fonte: Elaborada pelo autor.

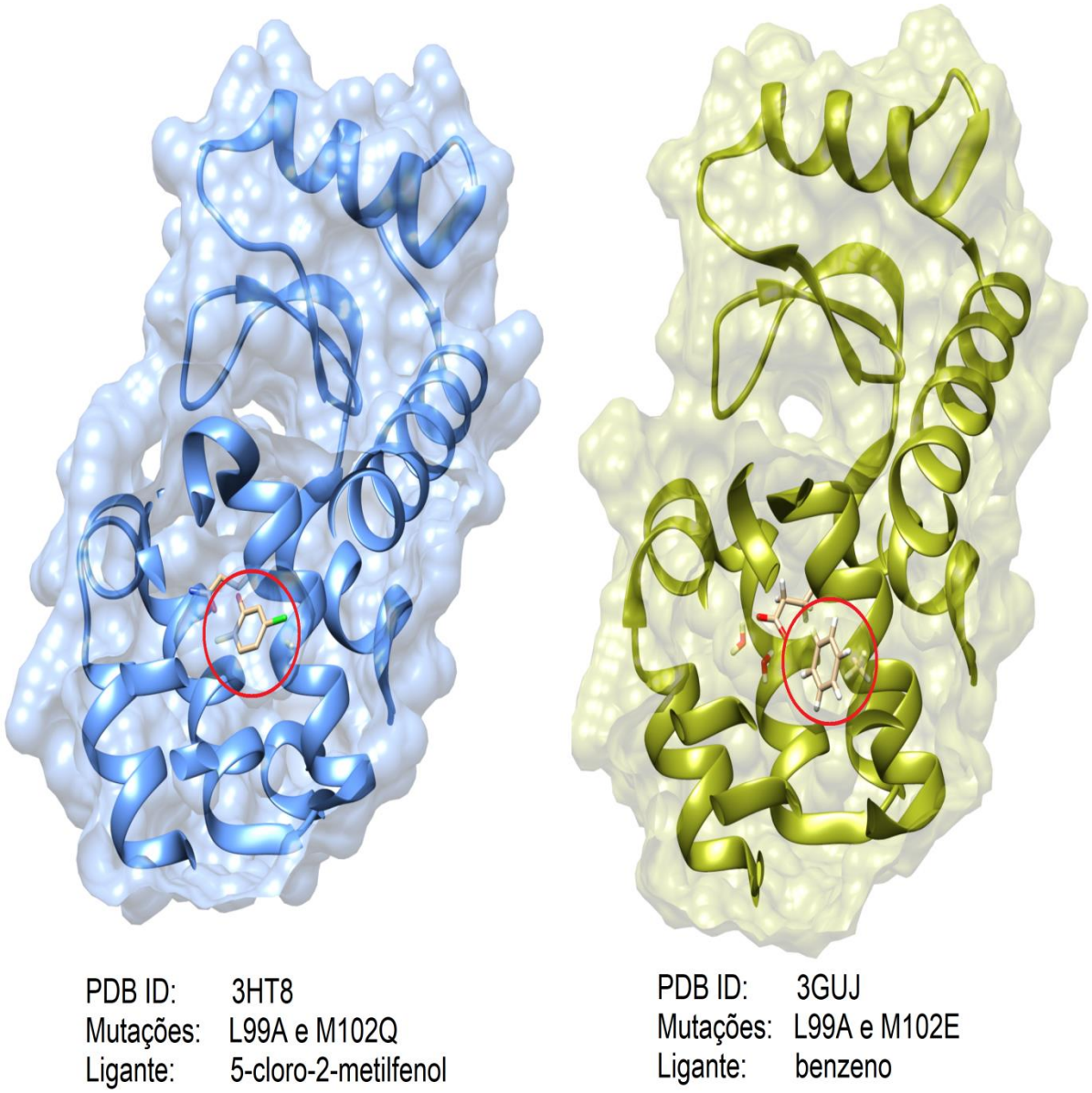



Figura 2 - Estrutura 3D e visão de seção da superfície de lisozima mutante de vírus T4. À esquerda, em azul, está o ligante 5-cloro-2-metilfenol, na cavidade, com os resíduos mutantes (L99A/M102Q) representados, interagindo com o ligante. PDB ID: 3HT8. À direita, está o ligante benzeno, na cavidade, com o resíduo mutante (M102E) representado junto com uma molécula de água no sítio de interação, interagindo com o ligante. PDB ID: 3GUJ.

Fonte: Elaborada pelo autor.

A Figura 2 mostra duas das 30 estruturas de lisozima de vírus T4 utilizadas aqui juntamente com o destaque para o ligante e a região de interação de cada um deles. À esquerda, em azul, é possível ver a estrutura do complexo de código PDB 3HT8, com a cavidade de interação destacada, na qual se encontra o ligante, 5-cloro-2-metil-fenol, e os resíduos mutantes A99 e Q102. À direita, em amarelo, é

possível ver a estrutura do complexo de código PDB 3GUJ, com a cavidade de interação destacada, na qual está presente o ligante, benzeno, juntamente com uma molécula de água mediando a interação. Além disso, o resíduo mutante E102 também está destacado.

## 3.2 Preparação das moléculas para simulação e cálculo das energias

### 3.2.1 Dinâmica Molecular

Todas as estruturas dos complexos foram obtidas a partir das estruturas cristalográficas depositadas no PDB, e, antes de inicializarmos as simulações, as moléculas foram preparadas de modo a padronizá-las. Primeiro, para as simulações de DM, adicionamos os hidrogênios faltantes às moléculas com o *software* ***reduce***.[43] do pacote ***AmberTools16***.[44] Depois, o *software* ***antechamber***, também parte do ***AmberTools16***, foi utilizado para parametrizar as estruturas dos ligantes, auxiliando a gerar os arquivos de topologia e coordenadas, usando **GAFF** (*General Amber Force Field*)[45] como modelo de campo de força e o **AM1-BCC**[46] para as cargas atômicas.

Para os complexos, a parametrização e montagem das estruturas foram feitas com a utilização do *software* **tLEaP**, que também gera arquivos de topologia a partir de coordenadas. Para o receptor, o campo de força utilizado foi o **FF14SB**.[47] O complexo foi então mergulhado numa caixa cúbica, com 12,0 Å de distância entre as extremidades da caixa e a proteína, a qual foi preenchida com moléculas de água do tipo ***TIP3P***. Íons de cloro ($Cl^-$) ou sódio ($Na^+$) foram adicionados a fim de manter os sistemas neutros.

Para analisar as energias livres de ligação dos sistemas a partir de dinâmica molecular, os 30 ligantes tiveram simulações realizadas para eles livres e para os sistemas complexados. Tais simulações obedeceram a protocolos de minimização, termalização, densidade, equilíbrio e produção, e foram feitas utilizando o pacote de programas *Amber16.*[44]

A minimização foi um processo curto, feita durante 5000 ciclos, com os pesos das restrições posicionais de 2,0 kcal/mol.Å², restrição harmônica de todo o sistema, para cada complexo, com raio de corte não-ligado de 8 Å e volume constante (*ensemble* NVT). A minimização foi seguida pela etapa de termalização durante 50ps, também a volume constante, mas variando a temperatura até ela atingir 300K, com controle de temperatura de Langevin[6], e frequência de colisão de 2 $ps^{-1}$. A próxima etapa foi a de equilíbrio da densidade, que teve duração de 50 ps, também usando as restrições para todos os sistemas. Essa etapa também ocorreu à pressão constante com tempo de relaxação da pressão de 1,0 ps, Langevin para controle de temperatura e frequência de colisão de 2 $ps^{-1}$.

Após essas etapas, as simulações prosseguiram para a fase de equilíbrio dos sistemas, à pressão constante, com temperatura constante de 300K, controlada por Langevin, com raio de corte não-ligado de 8 Å, frequência de colisão de 2 $ps^{-1}$, por 500 ps.

A fase produtiva da simulação de DM foi feita usando as mesmas condições da fase de equilíbrio a fim de evitar problemas nas energias potenciais, portanto, o *ensemble* utilizado também foi o NPT. Ao todo, cada complexo teve sua simulação produtiva de DM realizada por 100 ns. As coordenadas das trajetórias foram salvas a cada 10 ps, a um passo de 2 fs. Todas as etapas das simulações, isto é, minimização, termalização, densidade, equilibração e produção, foram feitas utilizando a metodologia *SHAKE*[48] nos átomos de hidrogênio e utilizaram condições periódicas de contorno.

A análise das energias livres de ligação para as simulações de DM foi feita também utilizando o pacote AmberTools16. Para o cálculo da Energia de Interação Linear (LIE, da sigla em inglês para *Linear Interaction Energy*), a ferramenta *CPPTRAJ*[49] foi a escolhida. Para os cálculos por esse método, utilizamos a constante $\alpha = 0,16$ e a constante $\beta = 0,50$. Para os cálculos das energias no vácuo, MM-GBSA e MM-PBSA nós utilizamos a ferramenta ***PYMDPBSA***, que também faz parte do pacote de ferramentas do AmberTools16[44].

As simulações de DM foram feitas em GPUs NVIDIA GTX 980 ou GTX 780, tanto para a obtenção das trajetórias quanto para o cálculo das energias.

### 3.2.2 Monte Carlo

Todas as estruturas dos complexos obtidas a partir das estruturas cristalográficas depositadas no PDB e utilizadas para a padronização das moléculas para as simulações de DM também foram utilizadas aqui, entretanto, sua padronização foi diferente. Para realizar as parametrizações, utilizamos o *software* UCSF-Chimera[50], com sua ferramenta *DockPrep*, que parametriza o ligante e o receptor utilizando os campos de força **FF14SB** e **GAFF**, para o receptor e o ligante, respectivamente. A atribuição e correção das cargas atômicas é feita com **AM1-BCC**.

Para as simulações de Monte Carlo utilizou-se uma versão modificada do programa *LiBELa* (Ligand Binding Energy Landscape)[22] (disponível em: https://github.com/alessandronascimento/LiBELa)□, o qual foi desenvolvido pelo nosso grupo em linguagem C++, inicialmente como um programa de *docking* molecular e que teve a adição do método de Monte Carlo posteriormente.[52-53] Ao utilizar o programa, o usuário pode definir a temperatura, a quantidade de passos que ele quer utilizar, o campo de força para os parâmetros do ligante, o modelo eletrostático, o modelo de amostragem, os tamanhos dos deslocamentos a cada passo, entre outros parâmetros.

Para aumentar a velocidade com que os cálculos são feitos, o receptor é considerado rígido e os potenciais de ligação são pré-calculados em um malha (*grid*) tridimensional.[54-55] Em seguida o programa procede à amostragem para o cálculo das energias livre de interação com os ligantes de duas maneiras distintas, que o usuário define: uma que será denominada *sample_torsion* e outra que será denominada por *full_flex*:

1) O algoritmo para *sample_torsion* envolve uma translação randômica do centro de massa do ligante dentro do sítio de interação do receptor seguida por uma rotação também aleatória e, por fim, uma variação também aleatória em cada ângulo de torção envolvendo os átomos pesados, isto é, átomos que não são hidrogênio, é introduzida para as

moléculas que possuem ângulos de torção. O conjunto de 6 + $n_{torções}$ variáveis define um passo.

2) O algoritmo para *full_flex* também faz a rotação e a translação aleatórias seguidas por um deslocamento atômico aleatório para cada átomo que compõe a molécula. Os critérios de aceitação e a temperatura do sistema são variáveis importantes para que não se observe a deformação da molécula.

A energia do sistema, composta pelas energias de interação e as energias internas (calculadas utilizando a biblioteca OpenBabel),[55] é então testada pelo critério de Metropolis, a cada passo, independentemente do tipo de amostragem.

Antes de cada simulação produtiva, o sistema foi equilibrado por $10^6$ passos de MC e, depois disso, a simulação foi feita por 40 milhões de passos de MC. A translação máxima a cada passo para cada direção para *sample_torsion* foi 0,5 Å, enquanto a rotação máxima foi de 2 graus a cada passo e a torção máxima foi de 1 grau, a cada passo. Para *full_flex*, deslocamentos máximos de 0,02 Å a 0,05 Å foram validados. Um arquivo típico de *input* pode ser encontrado no APÊNDICE C – Input MCLiBELa. Para ambos os modelos de amostragem, simulações dos ligantes livres e dos complexos foram feitas.

Durante os cálculos, o *LiBELa* guarda as energias médias das simulações de Monte Carlo que passaram pelo critério de Metropolis. Essa energia média e seu desvio padrão são usados para o cálculo das energias livres assumindo que $U = E_{interna} + E_{não-ligado} = \langle E \rangle$ , onde

$$E_{não-ligado} = \sum_{i \neq j} \left( \frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) + \sum_{i \neq j} q_i \Phi(r_i) \qquad (3.1)$$

$q_i$ é a carga parcial do átomo *i*, $r_{ij}$ é a distância entre os átomos *i* e *j*, A e B são os parâmetros de Lennard-Jones para os átomos *i* e *j* e $\Phi(r_i)$ é o potencial eletrostático, podendo ser Coulomb ou Poisson-Boltzmann.

Temos então $\Delta U = \langle E \rangle_{complexo} - \langle E \rangle_{ligante}$, isto é, a energia macroscópica no *ensemble* canônico é dada como a média dos microestados amostrados e a energia de ligação $\Delta U$, é calculada como a diferença entre a energia do complexo ligado ($E_{complexo}$) e o ligante livre ($E_{ligante}$).

Para cada complexo ligante-receptor e para o ligante livre foram feitas 10 simulações com diferentes sementes randômicas para estimar as energias livres de ligação e os erros associados, obtidos a partir dos desvios padrão do conjunto de 10 simulações. A consistência das correlações entre os dados calculados e os dados experimentais foi avaliada usando o método de ***Bootstraping***. Assumiu-se que os dados experimentais tinham distribuições gaussianas de erros, com σ=0.5 kcal/mol. Mudando aleatoriamente as energias livres determinadas experimentalmente usando uma distribuição gaussiana de números,[56]□ um milhão de 'novos' conjuntos de dados foram gerados e o coeficiente de correlação entre os dados calculados e os novos conjuntos de dados foram avaliados.[57] A média do coeficiente de correlação será apresentada aqui juntamente com seu desvio padrão.

Todas as simulações apresentadas aqui foram feitas para a temperatura de 250K, utilizando dois modelos para o cálculo do termo eletrostático da energia de interação – Coulomb e Poisson-Boltzmann (que aqui chamaremos de Poisson) – e as duas formas de amostragem, em caixas cúbicas de lado 30,0 Å e com espaçamento no *grid* de 0,4 Å, resultando em uma caixa cúbica com 75 x 75x 75 pontos na malha do *grid*.

O potencial eletrostático empregando o modelo de PB foi calculado empregando o programa Delphi, o qual calcula os potenciais eletrostáticos dos sistemas e de seus entornos, usando uma solução de diferencial finita para a equação de PB.[58] Manuais do programa e seu download podem ser encontrados na página http://compbio.clemson.edu/delphi.

Cada uma das simulações de MC foi feita em um único núcleo, utilizando processadores Xeon E5640 2.67GHz.

## 3.3 Entropia

Para estimar as entropias, assumimos cada grau de liberdade *r* ($r_1$, $r_2$, $r_3$, ..., $r_n$) desacoplado e, para esse propósito, nós calculamos a entropia de cada grau de liberdade como uma entropia de Shannon, onde *r* representa uma configuração de translação (*x, y, z*), rotação da molécula (ângulos de Euler *α, β, γ*) e torção (*n*):

$$\frac{s_r}{k_B T} = -\int \rho(r_i) \ln \rho(r) dr_i \quad (3.2)$$

onde *ρ(r)* é a função probabilidade do sistema dada uma configuração *r*, é estimada a partir de uma função densidade de probabilidade de dimensão 1 para cada grau de liberdade. A entropia total do sistema é então dada por

$$S = s_x + s_y + s_z + s_\alpha + s_\beta + s_\gamma + \sum_n^{N_{rot}} s_n \quad (3.3)$$

onde $N_{rot}$ representa o número de ligações torcionáveis entre os átomos pesados, do ligante. O cálculo da entropia de ligação $\Delta S$ é feito de maneira igual à $\Delta U$: são feitas simulações do complexo e do ligante livre e então tomamos a diferença entre eles.

Cada integral da equação (3.2) é resolvida por meio do método de contagem de colunas,[52] o qual divide em regiões com larguras pré-estabelecidas os valores das variáveis dependentes. Um histograma é criado com os pontos contabilizados dentro das regiões, de modo que a integral se resolve ao somar a área do histograma criado.[20, 59]

Depois de todos esses processos, as energias livres de ligação são expressas da forma:

$$\Delta F \approx \Delta G = \Delta U - T\Delta S \quad (3.4)$$

# 4 RESULTADOS

Nosso principal objetivo aqui é calcular as energias livres de ligação entre os complexos escolhidos, utilizando diferentes métodos. Entretanto, apenas o cálculo por si só não trazer resultados significativos, isto porque a escolha de um modelo que faça uma amostragem relativamente boa e que possibilite uma melhor aproximação ao resultado experimental, ou pelo menos ao perfil energético das interações, é essencial para obter tal grandeza.

O método mais simples para se estimar a energia de interação entre um ligante e um receptor é o método do *docking* molecular. Sucintamente, o método consiste em encontrar uma conformação viável do ligante que minimize a energia de interação com seu respectivo receptor. Como uma primeira abordagem, estimamos as energias de interação para os complexos listados na Tabela 2, tomando as próprias estruturas cristalográficas como coordenadas atômicas.

Os valores das energias de interação obtidas por *docking* molecular, utilizando o módulo de *docking* do *MCLiBELa*, podem ser observados na tabela a seguir na coluna $E_{calc}$, utilizando o modelo descrito pelas equações (1.12) e (1.13), sem considerar o potencial eletrostático.

Tabela 2 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando *docking* molecular.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $E_{calc}$ (kcal/mol) |
|---|---|---|
| **182L** | -5,46 | -19,80 |
| **183L** | -5,13 | -19,47 |
| **184L** | -6,51 | -21,24 |
| **185L** | -4,89 | -21,32 |
| **186L** | -6,70 | -21,23 |
| **187L** | -4,67 | -15,91 |
| **188L** | -4,60 | -16,28 |
| **1NHB** | -5,76 | -17,06 |
| **3DMX** | -5,35 | -13,96 |
| **4W52** | -5,19 | -14,02 |
| **4W53** | -5,52 | -16,70 |
| **4W55** | -6,55 | -21,09 |
| **4W57** | -6,69 | -22,30 |
| **1LGW** | -5,53 | -24,42 |
| **1LI2** | -5,58 | -15,77 |
| **1LI3** | -5,87 | -23,46 |
| **1LI6** | -5,25 | -16,55 |
| **1XEP** | -4,16 | -4,05 |
| **2RBO** | -5,52 | -20,94 |
| **3HT8** | -5,04 | -23,07 |
| **3HTB** | -5,33 | -25,75 |
| **3HU8** | -4,02 | -22,41 |
| **3HUA** | -4,61 | -26,75 |
| **3HUK** | -4,48 | -22,23 |
| **3GUJ** | -5,80 | -15,12 |
| **3GUK** | -5,30 | -16,50 |
| **3GUL** | -5,50 | -19,96 |
| **3GUM** | -4,90 | -18,43 |
| **3GUN** | -4,00 | -24,76 |
| **3GUO** | -5,00 | -22,28 |
| **Correlação de Pearson** | | **0,13** |

Fonte: Elaborada pelo autor.

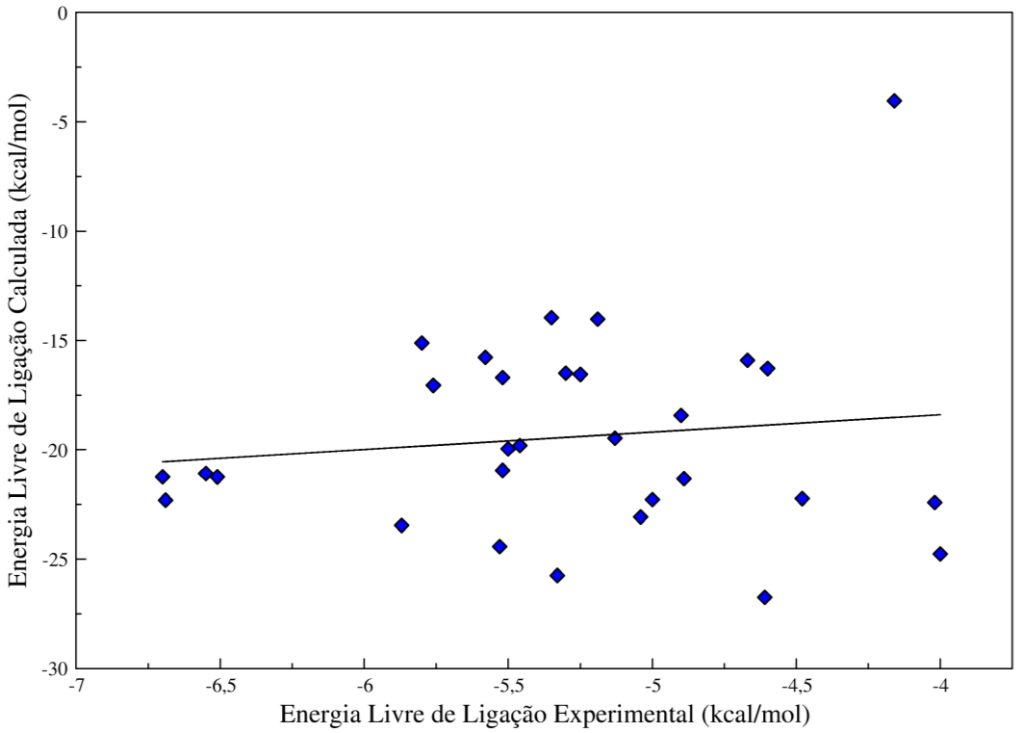



Figura 3 - Comparação entre os dados experimentais e os valores obtidos de energia livre de ligação utilizando o módulo de docking do LiBELa.

Fonte: Elaborada pelo autor.

A Figura 3 compara os valores de energia livre de ligação experimental e os valores calculados por *docking*. A correlação de Pearson, indicada na Tabela 2, foi bastante baixa para o modelo utilizado, o que era esperado, dado que não há uma otimização das posições do ligante que gere melhores configurações e, consequentemente, melhores energias. Mesmo excluindo o ponto mais distante dos demais (1XEP, cuja estrutura cristalográfica apresenta duas conformações distintas para o ligante no sítio de interação), a correlação ainda permanece baixa (-0,08), indicando que o modelo energético baseado em uma amostragem limitada é ineficiente.

Embora tenha sido possível obter os valores das energias para os complexos sem fazer otimizações nas configurações, métodos confiáveis de cálculo de energia livre de interação utilizam meios de determinar *ensembles* bons o bastante que representem essas interações. Dessa forma, para gerar nossas configurações e calcular as energias livres de ligação, utilizamos as simulações de MC e de DM.

Os métodos de dinâmica molecular estão entre os mais empregados para a geração de *ensembles* termodinâmicos visando o cálculo de energias livres de ligação entre proteína e ligantes. Desta forma, optamos por avaliar inicialmente o conjunto de dados de complexos da lisozima do vírus T4 com diferentes ligantes empregando a técnica de DM a fim de ter um controle positivo para a geração de microestados e estimativas da energia livre de ligação.

## 4.1 Dinâmica Molecular

A seguir, mostramos os resultados obtidos para a dinâmica molecular utilizando a metodologia descrita anteriormente.

### 4.1.1 Amostragem

As trajetórias das simulações de DM foram salvas em arquivos binários (em formato NETCDF), de modo que a visualização delas pudessem ser feitas utilizando ferramentas como o VMD[60] e o UCSF Chimera.[50]

O primeiro critério de avaliação da amostragem de conformações do ligante foi análise da raiz do desvio médio quadrático, ou RMSD. A figura a seguir mostra como os sistemas se comportaram durante as amostragens dos RMSDs.

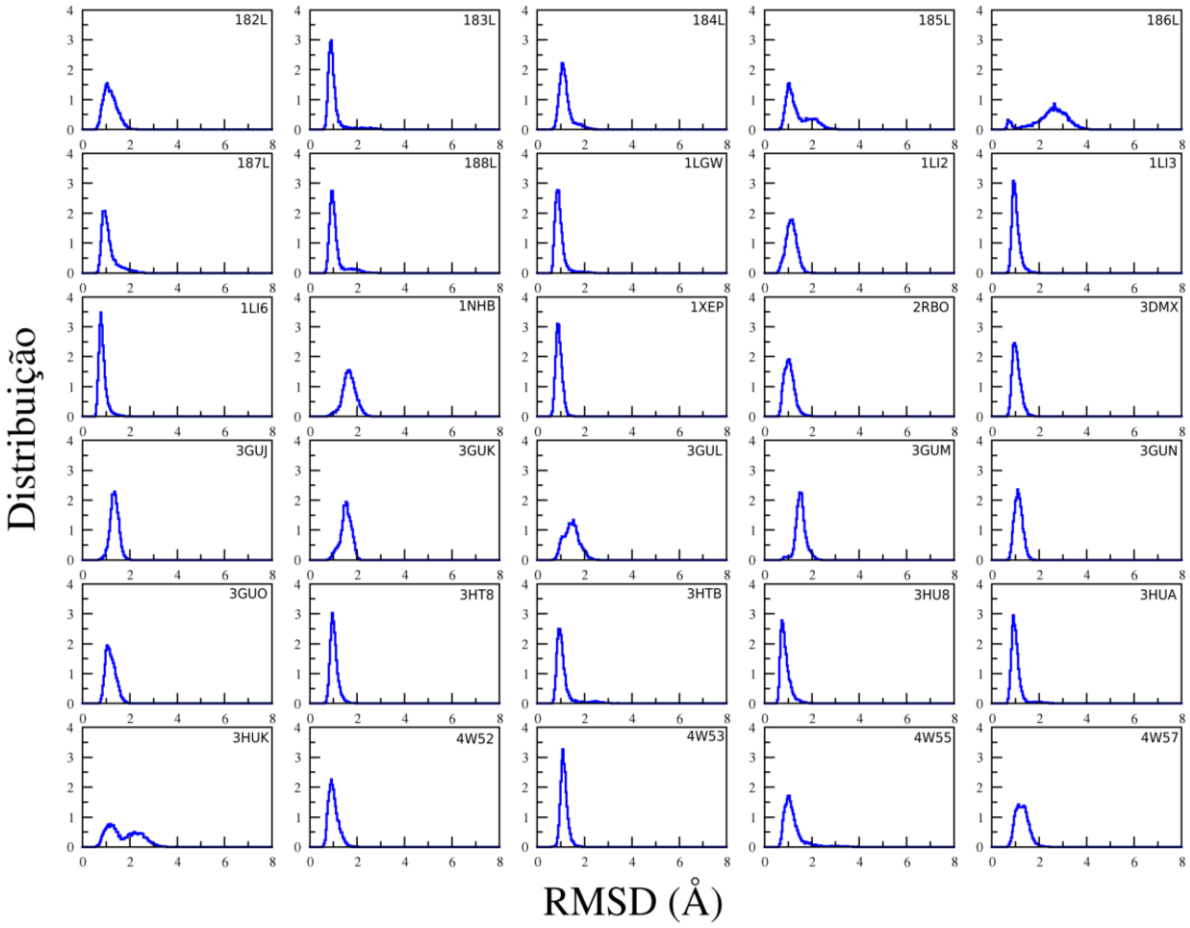
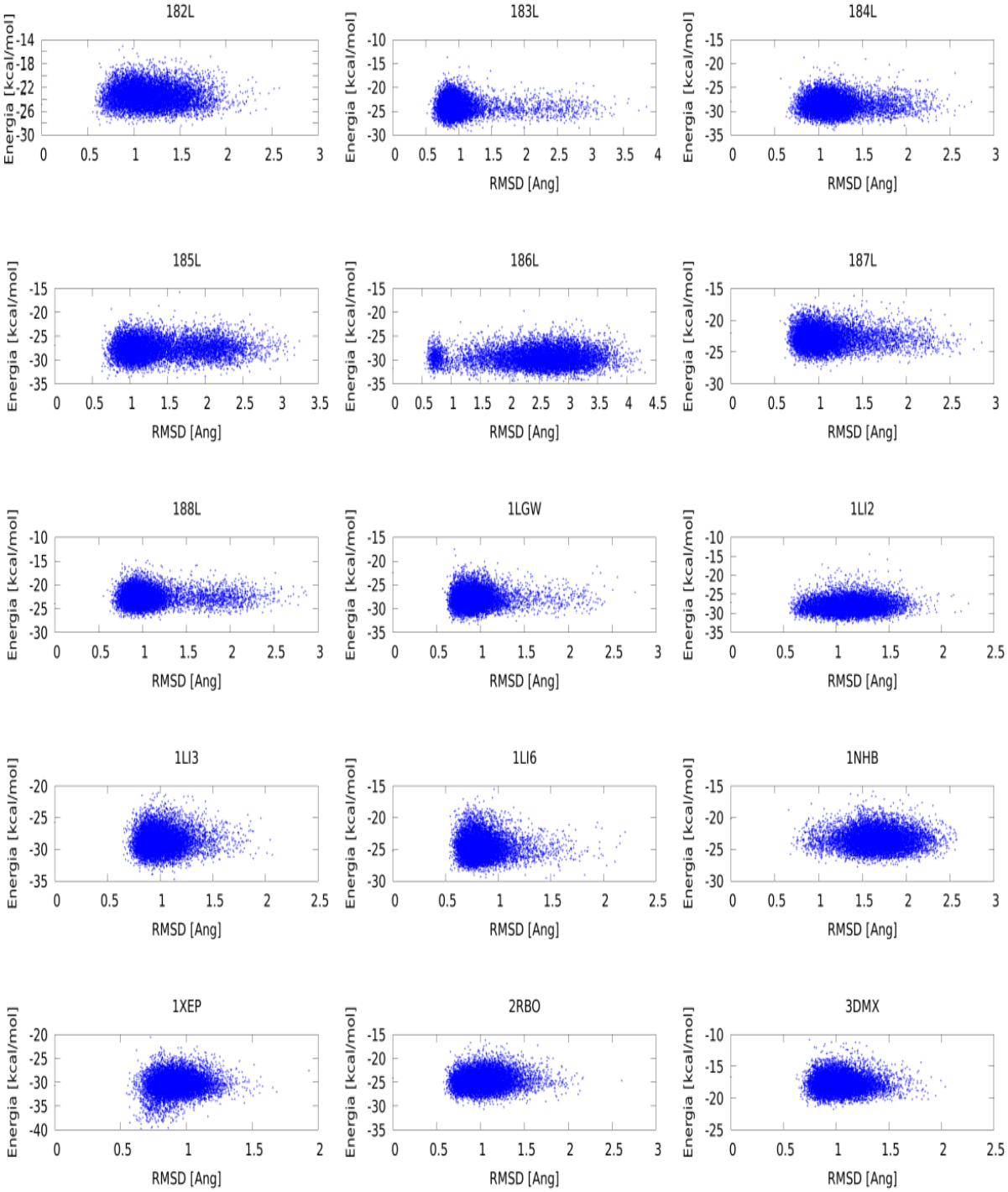



Figura 4 - Histogramas apresentando a distribuição de RMSDs obtidos para a as simulações de cada complexo. Os valores de RMSDs são calculados para o ligante somente e tendo como referência a conformação cristalográfica do ligante.

Fonte: Elaborada pelo autor.

A Figura 4 mostra que as moléculas dos ligantes, no geral, não se distanciaram muito da configuração apresentada pela estrutura cristalográfica, de modo a indicar que a exploração do espaço de configurações é limitada a um sub-espaço próximo da configuração inicial do sistema, como esperado.

Outro critério de comparação bastante útil foi analisar o quanto cada um dos ligantes se desvia da posição cristalográfica inicial, isto é, a mudança, durante as simulações de DM, nos RMSDs de acordo com as energias de interação, calculadas no vácuo com o programa *AmberEnergy*++ (disponível em: https://github.com/alessandronascimento/amberenergy). A Figura 5 mostra o perfil de energia resultante assumido pelos complexos durante as simulações e os estados de ocupação de RMSD.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
Energia [kcal/mol]
RMSD [Ang]

(continua)

(continuação)

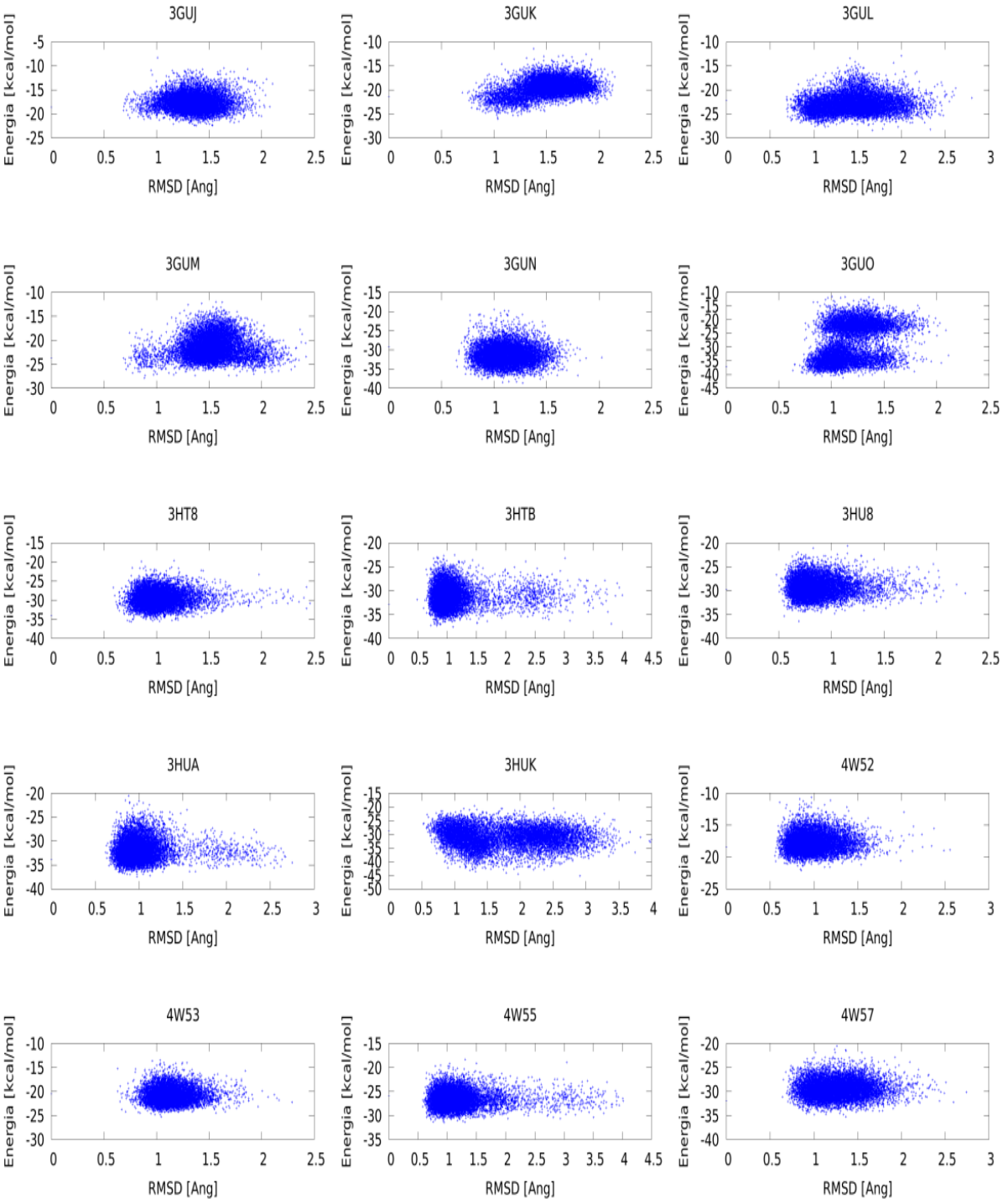



Figura 5 - Avaliação das energias médias de ligação para as simulações de dinâmica molecular para os 30 complexos como função do RMSD do ligante comparado com a posição cristalográfica.

Fonte: Elaborada pelo autor.

Da Figura 5 é possível ver que boa parte dos ligantes concentrou suas amostragens em torno de 1 Å, apesar de apresentarem uma distribuição de pontos também para valores maiores de RMSD, com energias de ligação variando entre -40 kcal/mol e -15 kcal/mol. Essa distribuição de pontos para valores maiores era esperada durante as simulações dado que os sistemas são tidos como dinâmicos e os ligantes podem se movimentar dentro dos sítios de ligação, ao contrário da posição cristalográfica.

Os ligantes que apresentam ligações que formam ângulos de torção, no geral, apresentam valores de RMSD variando com a energia livre de ligação maiores que a média, como também era esperado. Assim, para termos mais um parâmetro de comparação com os demais métodos, os ângulos de diedro para os sete ligantes que apresentam torções, 184L, 186L, 3THB, 3HU8, 3HUK, 4W55 e 4W57, fizemos uma análise dos ângulos de torção, utilizando a ferramenta *CPPTRAJ*.[49]

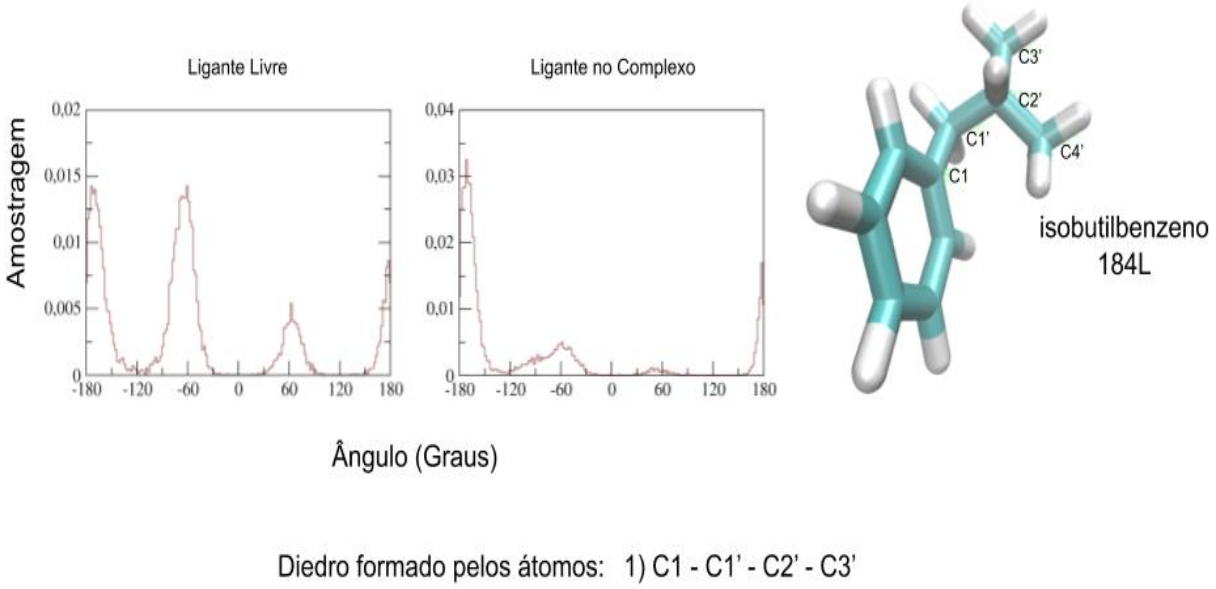



Figura 6 - Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno), cujo complexo apresenta apenas a mutação L99A. No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura.

Fonte: Elaborada pelo autor.

A Figura 6 apresenta os gráficos, normalizados, das distribuições de ângulos de diedros formados pelos átomos C1 – C1’ – C2’ – C3’ indicados na molécula, obtidas a partir das análises das trajetórias das simulações de DM para o ligante 184L livre e complexado. Para o ligante livre é possível ver que as distribuições formadas pelos átomos indicados se concentraram em torno de -180º, -60º, 60º e 180º. Entretanto, para o ligante complexado a distribuição apresentada se concentra mais em torno de -180º e 180º, com distribuições menos acentuadas em torno de -60º e 60º.

Para todos os outros ligantes com ângulos de torção, os resultados obtidos foram similares, isto é, as amostragens dos ângulos de diedro se concentram em torno de ângulos específicos, variando pouco de acordo com o ligante e suas ligações torcionáveis.

### 4.1.2 Energia Livre de Ligação para DM

As simulações de DM, feitas como foi descrito anteriormente para os 30 complexos ligante-receptor envolvendo a lisozima de T4, tiveram as energias livre de ligação calculadas das quatro maneiras diferentes: MM-GBSA, MM-PBSA, LIE e no vácuo. Uma comparação entre os valores dessas energias livres de ligação obtidas computacionalmente e os valores experimentais disponíveis na literatura pode ser encontrada na tabela bem como nas figuras a seguir. As correlações entre os dados, apresentadas nas últimas linhas, foram obtidas a partir da correlação de Pearson e da média pelo método de *bootstraping*.

Tabela 3 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando como base as simulações de Dinâmica Molecular.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $E_{LIE}$ (kcal/mol) | $E_{vácuo}$ (kcal/mol) | $E_{MM-GBSA}$ (kcal/mol) | $E_{MM-PBSA}$ (kcal/mol) |
|---|---|---|---|---|---|
| **182L** | -5,46 | -3,64 ± 0,02 | -23,0 ± 0,6 | -17,1 ± 0,4 | -14,1 ± 0,7 |
| **183L** | -5,13 | -3,76 ± 0,02 | -24,6 ± 0,5 | -19,0 ± 0,4 | -16,9 ± 0,6 |
| **184L** | -6,51 | -4,23 ± 0,02 | -30,3 ± 0,5 | -25,2 ± 0,4 | -22,5 ± 0,8 |
| **185L** | -4,89 | -4,93 ± 0,02 | -26,3 ± 0,6 | -17,9 ± 0,4 | -15,4 ± 0,7 |
| **186L** | -6,70 | -4,85 ± 0,01 | -31,5 ± 0,3 | -26,0 ± 0,3 | -24,9 ± 0,4 |
| **187L** | -4,67 | -3,12 ± 0,02 | -23,6 ± 0,5 | -19,4 ± 0,4 | -16,2 ± 0,7 |
| **188L** | -4,60 | -2,99 ± 0,02 | -22,8 ± 0,6 | -18,6 ± 0,5 | -15,1 ± 0,8 |
| **1NHB** | -5,76 | -3,42 ± 0,02 | -24,4 ± 0,5 | -20,0 ± 0,4 | -17,2 ± 0,7 |
| **3DMX** | -5,35 | -2,30 ± 0,01 | -17,4 ± 0,5 | -13,4 ± 0,4 | -10,5 ± 0,6 |
| **4W52** | -5,19 | -2,44 ± 0,01 | -17,6 ± 0,5 | -13,6 ± 0,4 | -11,1 ± 0,6 |
| **4W53** | -5,52 | -2,88 ± 0,02 | -20,7 ± 0,5 | -16,5 ± 0,4 | -13,7 ± 0,7 |
| **4W55** | -6,55 | -3,33 ± 0,02 | -26,3 ± 0,8 | -21,7 ± 0,7 | -17 ± 1 |
| **4W57** | -6,69 | -4,61 ± 0,01 | -31,1 ± 0,4 | -25,7 ± 0,4 | -24,2 ± 0,7 |
| **1LGW** | -5,53 | -5,68 ± 0,03 | -27,4 ± 0,8 | -17,3 ± 0,5 | -13,7 ± 0,8 |
| **1LI2** | -5,58 | -5,66 ± 0,04 | -26,8 ± 0,9 | -12,5 ± 0,4 | -8,9 ± 0,6 |
| **1LI3** | -5,87 | -6,85 ± 0,04 | -29,6 ± 0,8 | -15,9 ± 0,5 | -11,8 ± 0,7 |
| **1LI6** | -5,25 | -3,32 ± 0,03 | -21 ± 1 | -13,3 ± 0,6 | -8,3 ± 0,8 |
| **1XEP** | -4,16 | -8,90 ± 0,03 | -31,5 ± 0,5 | -12,5 ± 0,2 | -10,3 ± 0,3 |
| **2RBO** | -5,52 | -5,09 ± 0,02 | -24,36 ± 0,06 | -18,82 ± 0,04 | -14,9 ± 0,6 |
| **3HT8** | -5,04 | -6,95 ± 0,03 | -31,2 ± 0,8 | -17,9 ± 0,5 | -14,2 ± 0,7 |
| **3HTB** | -5,33 | -7,17 ± 0,02 | -34,2 ± 0,4 | -20,9 ± 0,3 | -18,7 ± 0,4 |
| **3HU8** | -4,02 | -5,53 ± 0,02 | -28,81 ± 0,06 | -18,88 ± 0,04 | -14,0 ± 0,5 |
| **3HUA** | -4,61 | -6,01 ± 0,03 | -31,4 ± 0,8 | -21,9 ± 0,6 | -17,8 ± 0,9 |
| **3HUK** | -4,48 | -6,04 ± 0,03 | -31,8 ± 0,8 | -22,3 ± 0,6 | -14,3 ± 0,7 |
| **3GUJ** | -5,80 | -2,31 ± 0,02 | -17,36 ± 0,06 | -12,61 ± 0,04 | -8,28 ± 0,05 |
| **3GUK** | -5,30 | -2,65 ± 0,02 | -19,98 ± 0,06 | -15,21 ± 0,05 | -11,29 ± 0,06 |
| **3GUL** | -5,50 | -3,23 ± 0,01 | -23,80 ± 0,05 | -19,23 ± 0,04 | -15,62 ± 0,06 |
| **3GUM** | -4,90 | -2,77 ± 0,02 | -21,84 ± 0,07 | -17,41 ± 0,06 | -12,05 ± 0,07 |
| **3GUN** | -4,00 | -7,61 ± 0,03 | -30,02 ± 0,09 | -15,38 ± 0,05 | -12,10 ± 0,06 |
| **3GUO** | -5,00 | -5,92 ± 0,05 | -14,80 ± 0,08 | -12,19 ± 0,07 | -6,92 ± 0,06 |
| **Correlação de Pearson** | | **-0,33** | **0,003** | **0,41** | **0,48** |
| **Correlação por média de Bootstrap** | | **-0,29 ± 0,02** | **0,08 ± 0,02** | **0,47 ± 0,02** | **0,53 ± 0,02** |

Fonte: Elaborada pelo autor.

Comparações gráficas dos valores experimentais das energias livres de ligação com os valores estimados utilizando os dados de DM podem ser encontradas na figura a seguir.

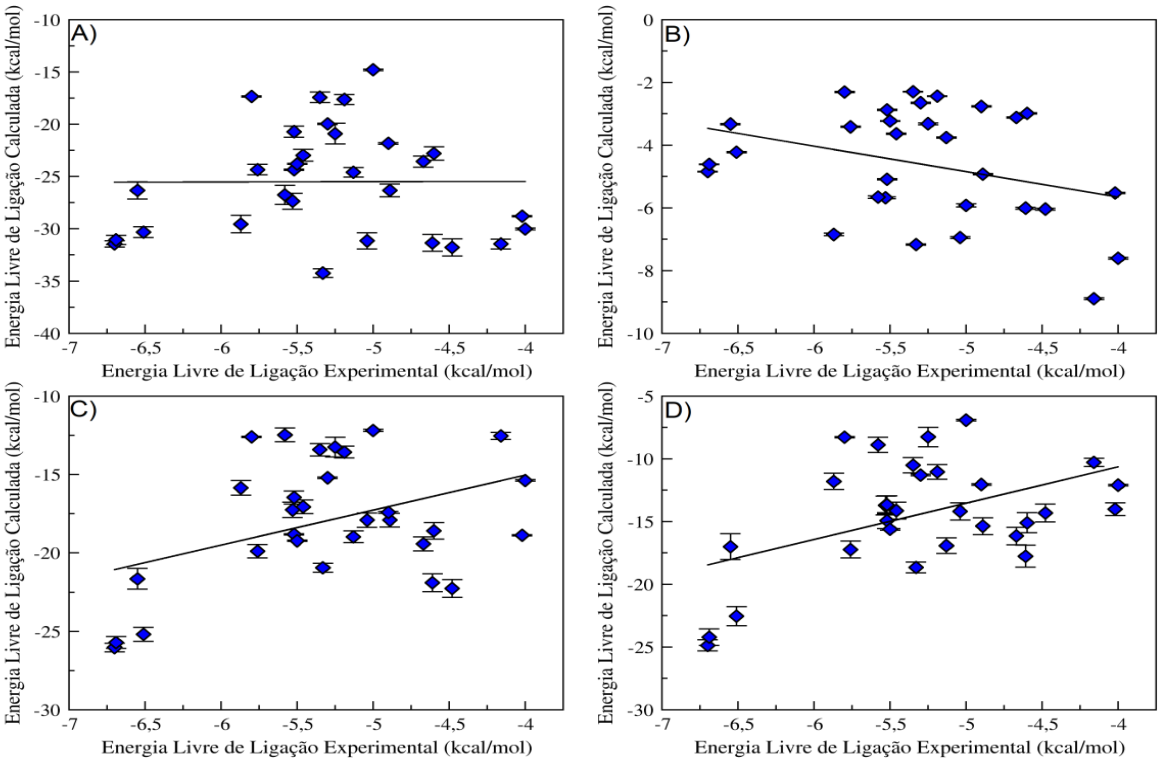



Figura 7 - Comparação entre os dados experimentais e os valores obtidos de energia livre de ligação para as simulações de DM par aos 30 complexos simulados. A) Energias obtidas das simulações de DM no vácuo. B) Energias obtidas utilizando o método de LIE, com β = 0,5 e α = 0,16. C) Energias obtidas utilizando o método de MM-GBSA. D) Energias obtidas utilizando o método de MM-PBSA.

Fonte: Elaborada pelo autor.

Para as simulações de dinâmica molecular, os cálculos de energia livre de interação no vácuo são menos custosos quando comparado com métodos como MM-GBSA e MM-PBSA, computacionalmente falando. Ao observarmos seus resultados, na Tabela 3 e Figura 7 A), é possível ver que esse modelo de cálculo, quando analisamos os 30 ligantes juntos, a correlação com as energias experimentais são bastante baixas, isto é, 0,003 quando analisamos a correlação de Pearson e 0,08 ± 0,02 por meio da média por *bootstrapping*, com desvio médio absoluto de 4,36 kcal/mol (N=30).

Ao olharmos para os valores de energia calculados, é possível ver que todos eles foram superestimados, isto é, assumiram valores muito mais negativos e favoráveis do que aqueles determinados experimentalmente.

O método de Energia de Interação Linear (da sigla em inglês, LIE), que usou como constante multiplicativa da contribuição eletrostática β = 0,5 e como constante multiplicativa da contribuição de van der Waals α = 0,16, apresentou os resultados mais baixos para a correlação entre os dados experimentais e computacionais. Para o conjunto com 30 ligantes, na Tabela 3 e Figura 7 B), o valor obtido para a correlação de Pearson foi de -0,29, enquanto que utilizando a média por *bootstrapping*, o valor de correlação obtido foi de -0,29 ± 0,02, com desvio médio absoluto 1,51 kcal/mol (N=30).

O método de *Molecular Mechanics Generalized - Born Surface Area* (MM-GBSA) teve um custo computacional mais elevado comparado com os anteriores, entretanto, isso resultou em uma correlação também maior. Para o conjunto contendo os 30 ligantes, a correlação de Pearson entre as energias livre de ligação experimentais e calculadas, como pode ser visto na Tabela 3 e o comportamento observado na Figura 7 C), foi de 0,41, Usando a média por *bootstrapping*, a correlação foi de 0,47 ± 0,02, com um desvio médio absoluto de 3,09 kcal/mol.

Para esse método, os valores de energia livre de ligação calculados também foram superestimados, assumindo valores de energia mais favoráveis que aqueles disponíveis na literatura.

O último método utilizado para calcular as energias livres de ligação para os dados da dinâmica molecular foi o método de *Molecular Mechanics - Poisson-Boltzmann Surface Area* (MM-PBSA). Dentre todas as maneiras de calcular as energias usando DM, essa foi a que apresentou os melhores resultados, porém, o custo computacional disso foi bastante elevado. A correlação de Pearson entre os valores experimentais e os calculados, apresentados na Tabela 3 e o comportamento observado na Figura 7 D), foi de 0,48. Utilizando a média por bootstrapping, a correlação foi 0,53 ± 0,02, com desvio médio absoluto de 3,28 kcal/mol.

Assim como os métodos anteriores, a utilização do método de MM-PBSA também resultou em energias livres de ligação superestimadas quando comparadas com as energias determinadas experimentalmente.

Entre as abordagens utilizadas para o cálculo das energias livres de ligação para as simulações de MD, LIE se mostrou a pior delas, com correlações muito baixas tanto para os 30 ligantes quanto para os conjuntos menores, porém, em termos de custo computacional, ele é o mais barato. O cálculo das energias livres no vácuo foi a segunda pior estimativa e segunda menos custosa em termos computacionais, porém, seus resultados não foram nada satisfatórios quando comparados com MM-GBSA e MM-PBSA. Esses últimos ocupam o segundo e o primeiro melhores lugares, respectivamente, em termos de maior correlação com os dados experimentais, porém também ocupam as posições de maior custo computacional, sendo MM-PBSA mais custoso que MM-GBSA. Além disso, os valores das energias de ligação foram superestimados, com relação aos valores absolutos, para todas as quatro abordagens utilizadas para obtê-las. É válido lembrar que todas as etapas que envolveram as simulações de DM, isto é, obtenção das trajetórias e cálculo das energias, foram realizadas em GPUs NVIDIA GTX 980 e/ou GTX 780.

Tabela 4 - Tempo gasto para obter as coordenadas por meio da dinâmica molecular e para o cálculo das variações de energia livre a partir das coordenadas geradas.

| Método | Simulação | Tempo Gasto | Processamento |
|---|---|---|---|
| DM | Coordenadas | ~ 15h | 1x GTX 980 |
| DM | Energia – LIE | < 5 min | 1x Xeon E5640 |
| DM | Energia – Vácuo | ~ 3h | 1x GTX 980 |
| DM | Energia – MM-GBSA | ~ 7h | 1x GTX 980 |
| DM | Energia – MM-PBSA | ~ 12h | 1x GTX 980 |

Fonte: Elaborada pelo autor.

## 4.2 Monte Carlo

Após testes preliminares, um espectro de temperaturas de MC adequado foi testado em ampla amostragem de conformação dos ligantes, porém, focando na região próxima ao mínimo de energia, isto é, à estrutura cristalográfica. Para os resultados obtidas, mostraremos um conjunto de simulações de MC à temperatura

de 250K, usando modelos eletrostáticos diferentes e métodos de amostragem distintos.

As trajetórias das simulações de MC foram salvas em SYBYL MOL2, de modo que a visualização de ambas as simulações pudesse ser feita utilizando ferramentas como o VMD[60] e o UCSF-Chimera.[50]

### *4.2.1 Sample Torsion*

Um dos métodos utilizados para gerar as configurações durante as simulações de MC foi o que chamamos aqui de *sample_torsion*, o qual se baseia em fazer uma translação aleatória do centro de massa do ligante dentro do sítio de interação do receptor, seguida por uma rotação também aleatória e, por fim, uma torção aleatória envolvendo os átomos pesados do ligante para as moléculas que possuem ângulos de torção.

Além disso, dois diferentes tipos de modelos eletrostáticos foram utilizados aqui para gerar as trajetórias: Coulomb e Poisson. A seguir, apresentaremos os resultados obtidos para essas simulações.

#### 4.2.1.1 Amostragem

Uma vez gerados os dados de MC, foi possível fazer uma comparação entre ambos os modelos eletrostáticos e as amostragens geradas por DM. As imagens a seguir mostram essa comparação.

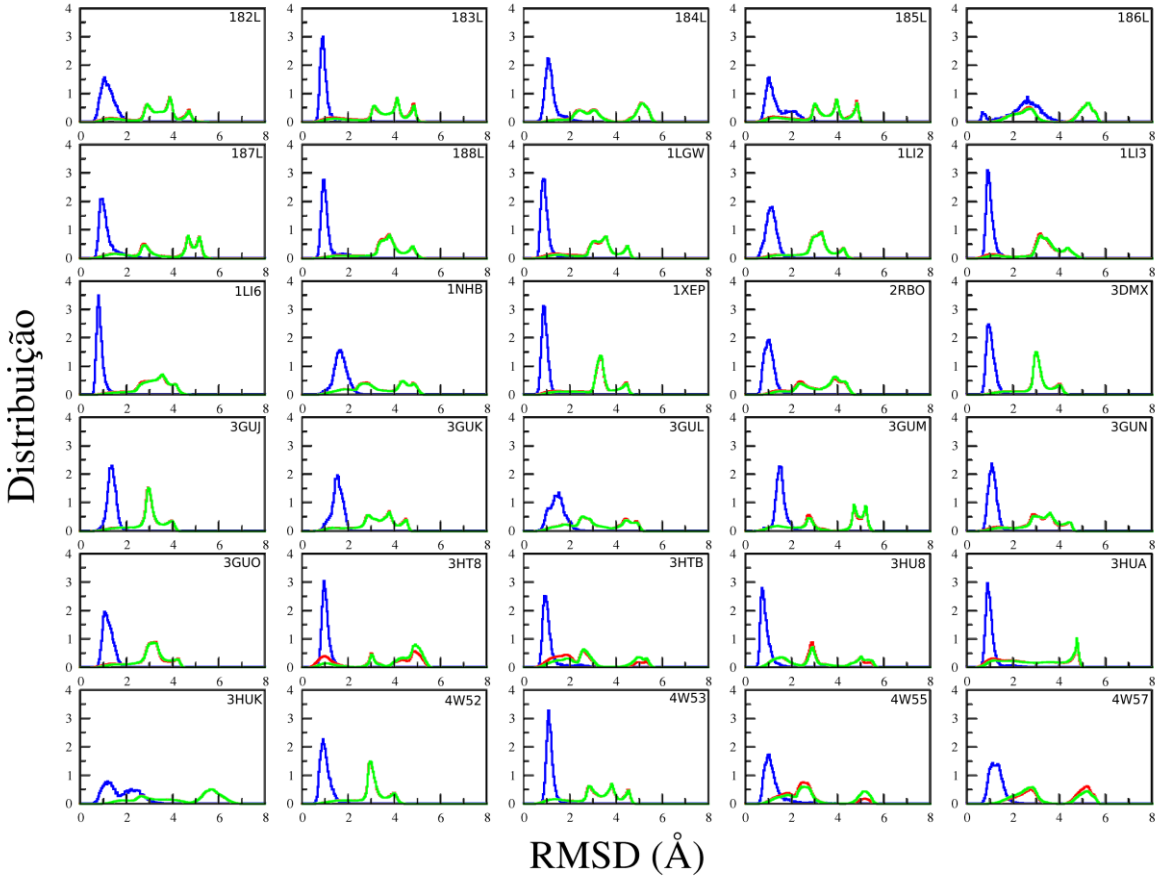


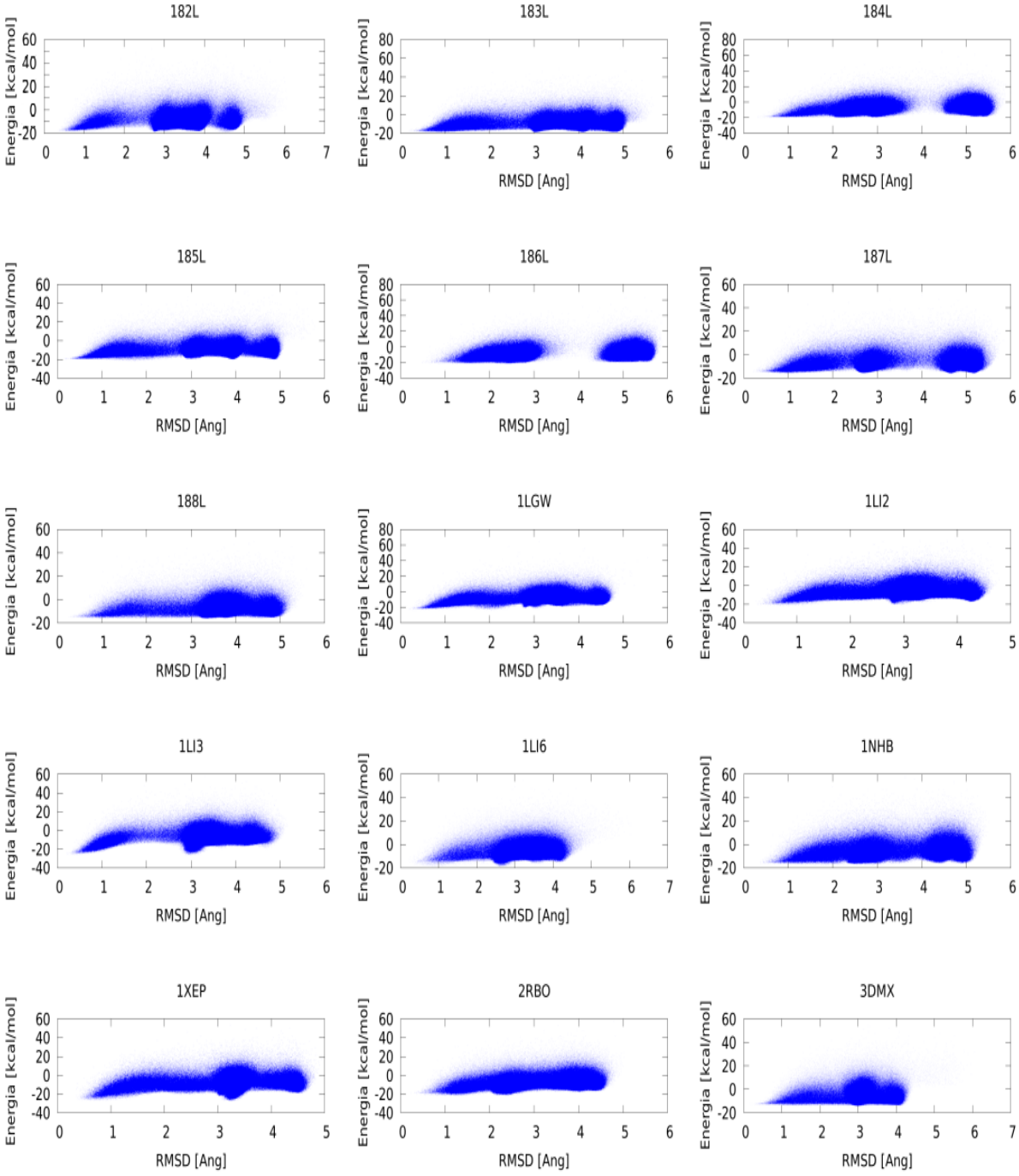

Figura 8 - Histogramas apresentando a distribuição de RMSDs obtidos para a as simulações de cada complexo. Os RMSDs são calculados para o ligante somente e tendo como referência a conformação cristalográfica do ligante. Em azul, vemos a amostragem para as simulações de DM. Em vermelho, está a amostragem para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Coulomb, 250K. Em verde, está a amostragem, como uma distribuição de RMSD, para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Poisson, 250K.

Fonte: Elaborada pelo autor.

Ao analisarmos a Figura 8 é possível ver uma nítida diferença entre as simulações de DM e MC: no geral, os desvios com relação às conformações das estruturas cristalográficas para os ligantes na maior parte dos complexos são relativamente maiores para MC do que para DM, apresentando alguns picos distintos na distribuição de RMSD. Entretanto, além desses RMSDs maiores, a maioria das simulações de MC amostrou também configurações nas faixas de RMSD apresentadas por DM.

Ao compararmos as amostragens geradas pelas simulações de MC por meio de modelos eletrostáticos distintos é possível perceber que tal mudança não resultou em grande diferença entre os estados de RMSD acessados.

Comparando os dados da amostragem por MC e as conformações dos ligantes nas estruturas cristalográficas, as conformações amostradas apresentaram-se dentro de um desvio de 6 Å com relação às conformações experimentais (RMSD para todos os átomos, como mostrado na figura a seguir, com as distribuições de RMSD de acordo com os estados das energias amostradas para os complexos analisados, com diferentes modelos eletrostáticos.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
Energia [kcal/mol]
RMSD [Ang]

(continua)

(continuação)

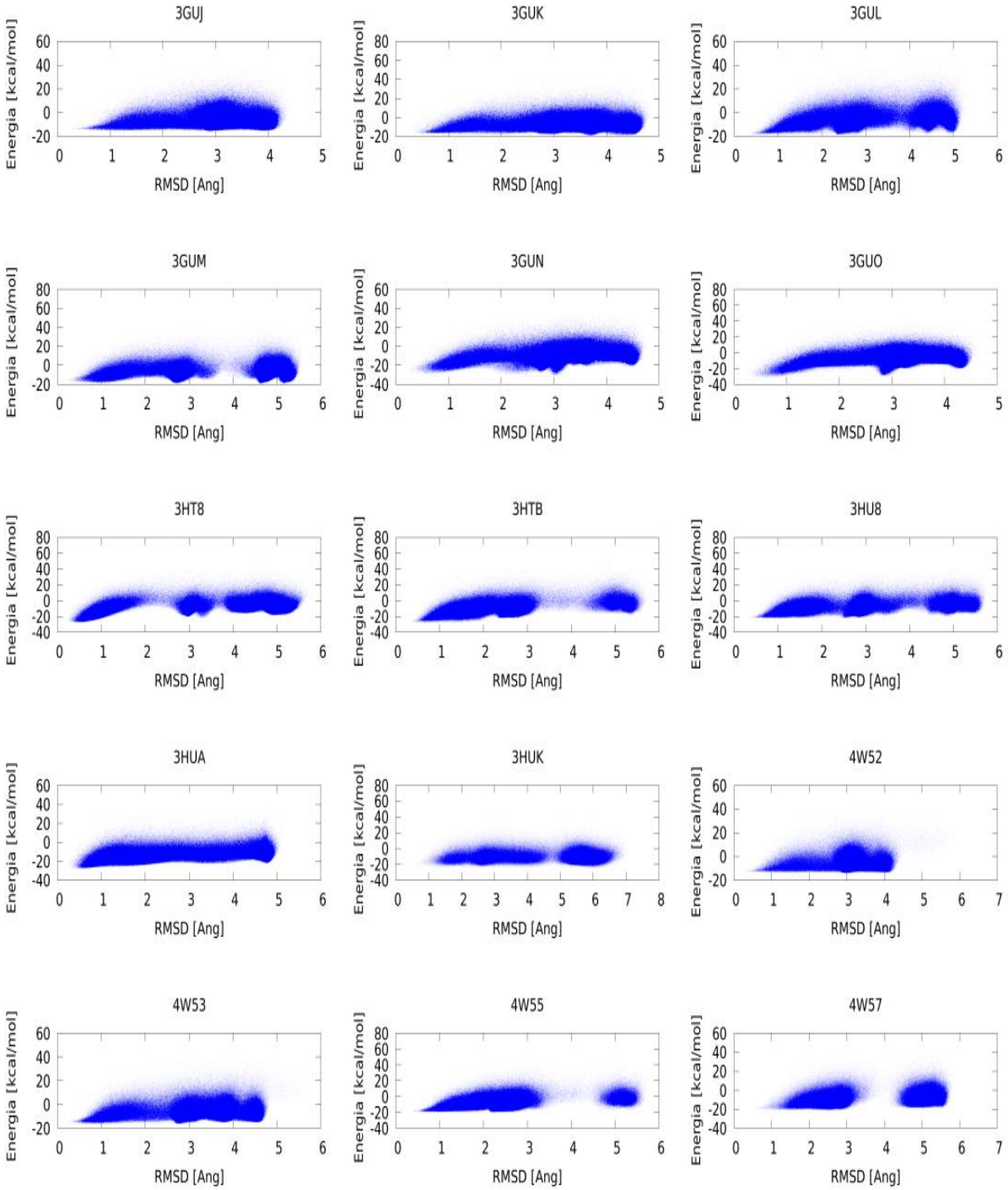
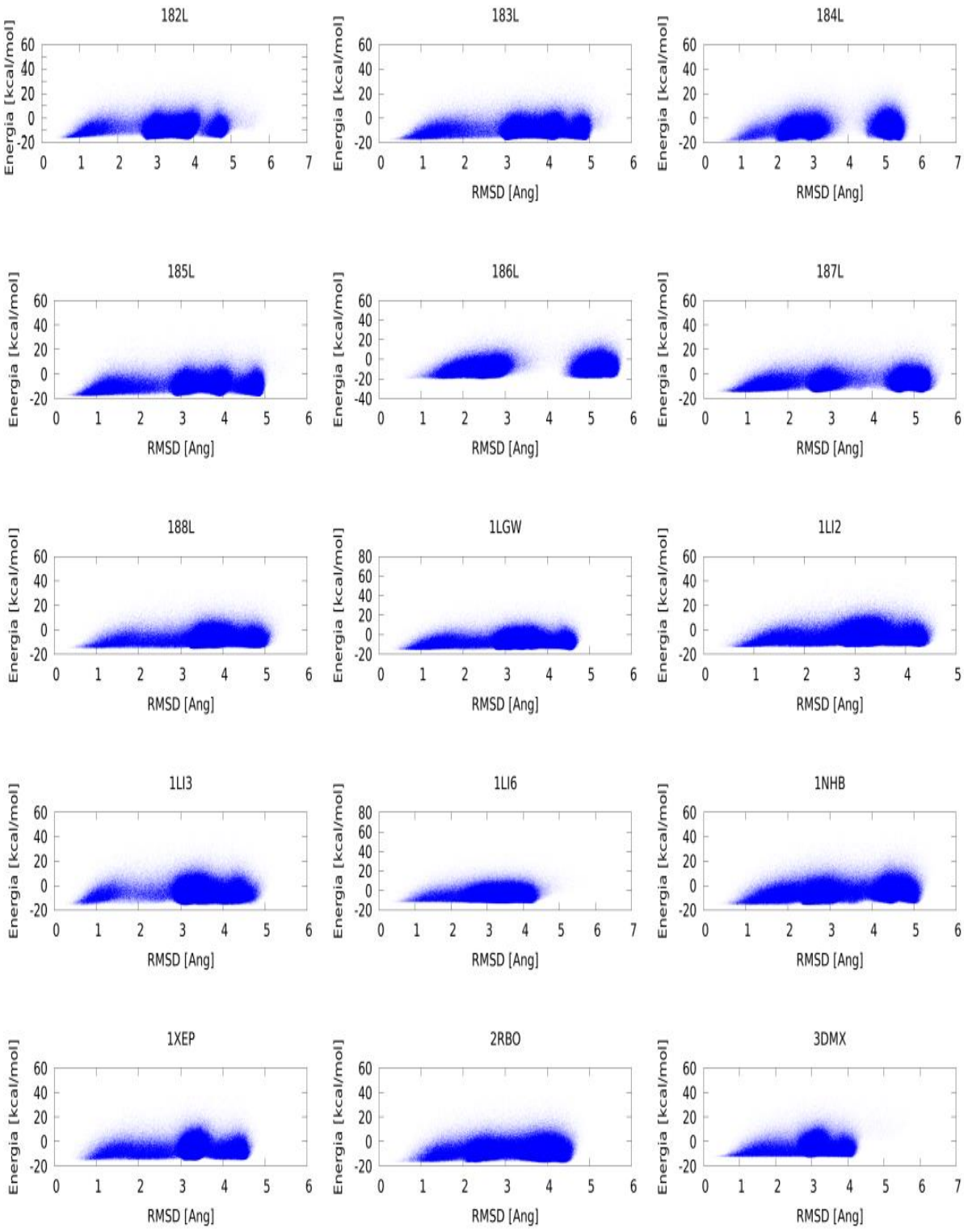



Figura 9 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Coulomb, temperatura é 250K e amostragem é sample_torsion.

Fonte: Elaborada pelo autor.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
Energia [kcal/mol]
RMSD [Ang]

(continua)

(continuação)

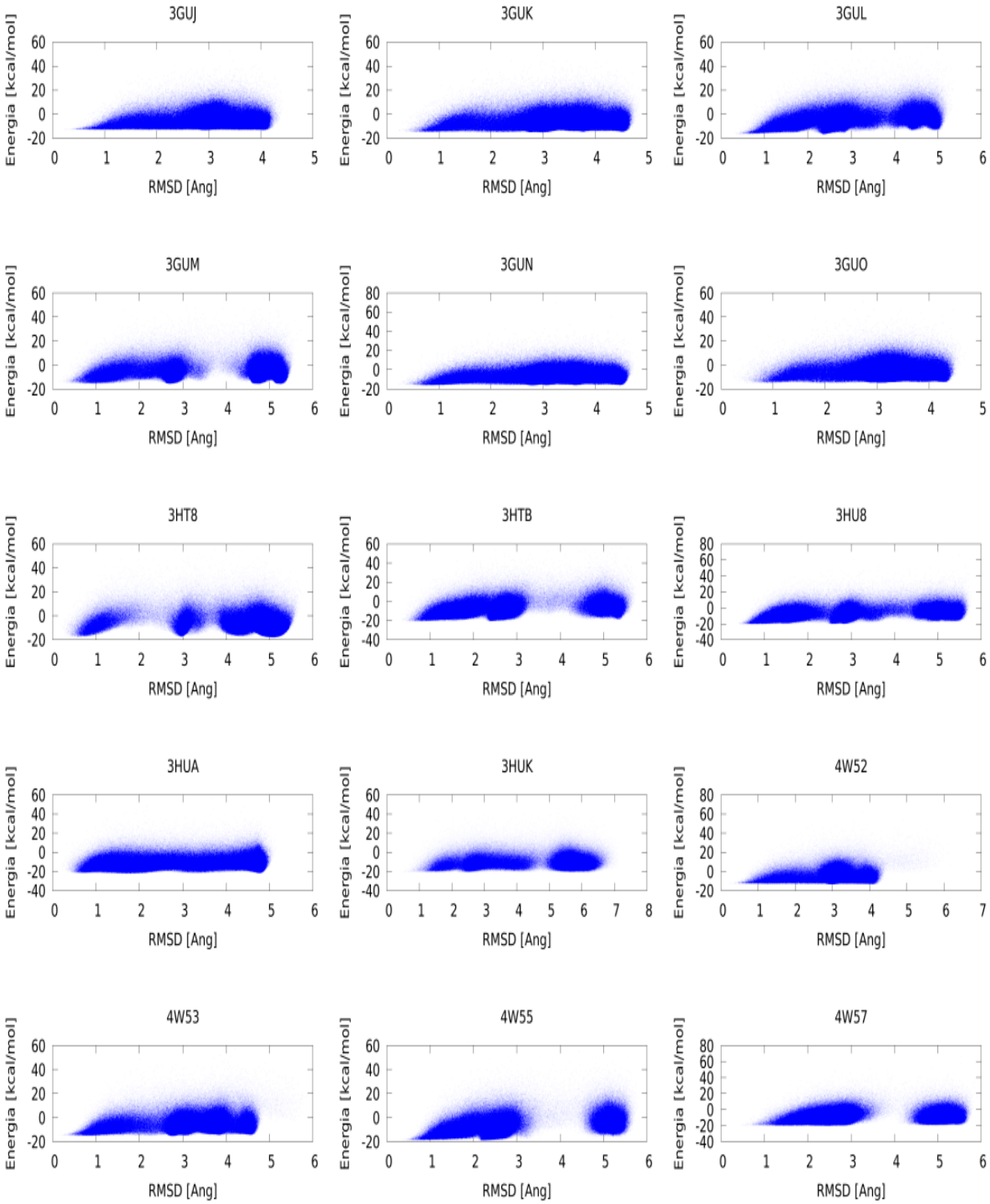



Figura 10 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Poisson, temperatura é 250K e amostragem é sample_torsion.

Fonte: Elaborada pelo autor.

Ao compararmos as amostragens apresentadas nas Figura 9 e Figura **10** é possível ver que os perfis assumidos pelas distribuições de RMSD com relação às energias livres de ligação médias não foram muito diferentes um do outro, isto é, os estados de RMSD acessados foram, essencialmente, os mesmos. Entretanto, se focarmos na energia é possível ver que a utilização do modelo eletrostático de Coulomb permitiu que os ligantes acessassem estados energeticamente mais favoráveis, isto é, energias mais negativas. Isso se deve ao fato de o modelo de Coulomb, que é menos robusto que o modelo de Poisson, tende a superestimar a componente eletrostática da energia livre de ligação quando comparado com o modelo de PB.

Assim como para as simulações de DM, uma análise dos ângulos de diedro formados pelos átomos pesados dos sete ligantes que apresentam ligações torcionáveis também foi feita utilizando a ferramenta *CPPTRAJ*. Os resultados são os mostrados na Figura 11.

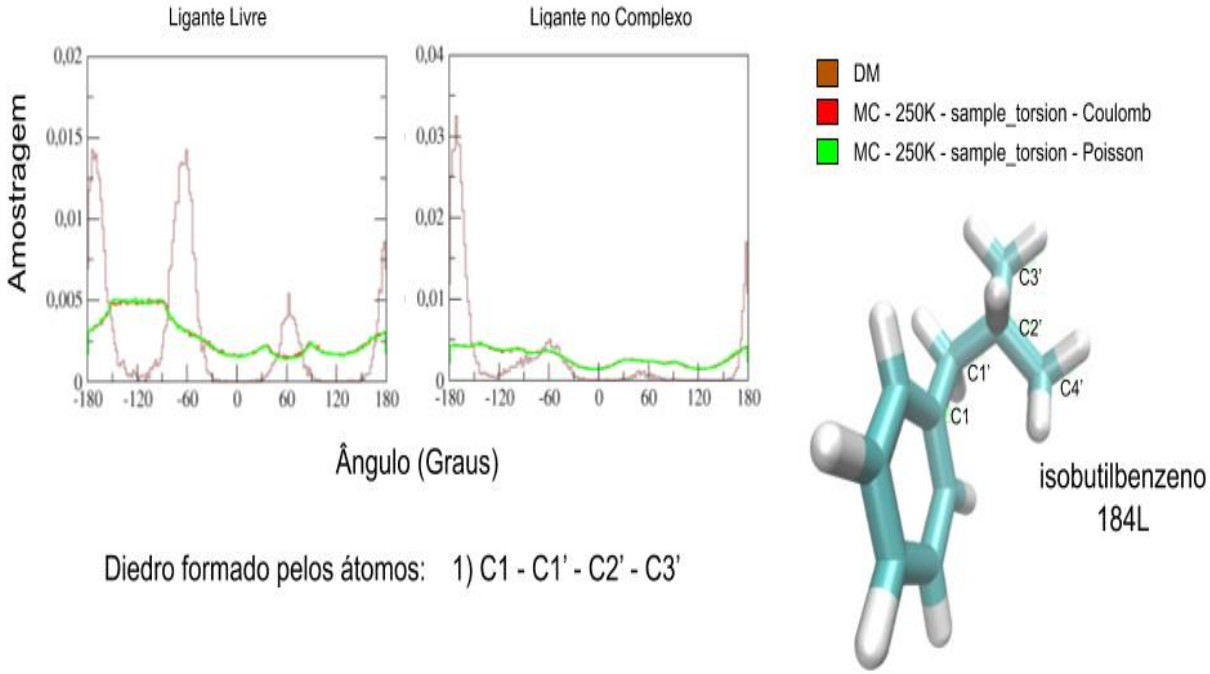



Figura 11 - Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno). No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura. Em marrom, está o perfil gerado por DM; em vermelho o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Coulomb como modelo eletrostático; e em verde o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Poisson como modelo eletrostático.

Fonte: Elaborada pelo autor.

A Figura 11 apresenta os gráficos, normalizados, das distribuições de ângulos de diedros formados pelos átomos C1 – C1’ – C2’ – C3’ indicados na molécula, obtidas a partir das análises das trajetórias das simulações de DM, em marrom, e de MC, em vermelho para as simulações a 250K, utilizando amostragem por *sample_torsion* e Coulomb como modelo eletrostático, e em verde para simulações que usaram o modelo eletrostático de Poisson, do ligante 184L livre e complexado. Nela é possível ver que todos os conjuntos de parâmetros usados para gerar as simulações de MC apresentaram resultados bastante semelhantes entre si para a amostragem dos ângulos de diedro e, além disso, elas também apresentaram perfis semelhantes àqueles gerados pelas simulações de DM.

No entanto, quando comparamos os resultados de ambos os métodos, é possível ver que a distribuição de ângulos para as simulações de Monte Carlo se dá de forma mais homogênea, ao passo que, para a dinâmica molecular, existem distribuições de base estreita em torno de alguns ângulos mais acentuados. Além disso, MC amostrou diedros que não foram amostrados por DM. A modificação do modelo eletrostático não influenciou a amostragem dos ângulos de diedro para esse ligante.

O perfil apresentado para os demais ligantes com ligações torcionáveis foi bastante similar ao da Figura 11, com os ângulos de diedros amostrados concentrados ao redor de ângulos específicos.

Para avaliar a geração de configurações dos ligantes que ocupassem todo o espaço configuracional disponível, a amostragem dos deslocamentos desses ligantes dentro da caixa cúbica de simulação e as rotações apresentadas por eles são mostradas nas seguintes imagens.

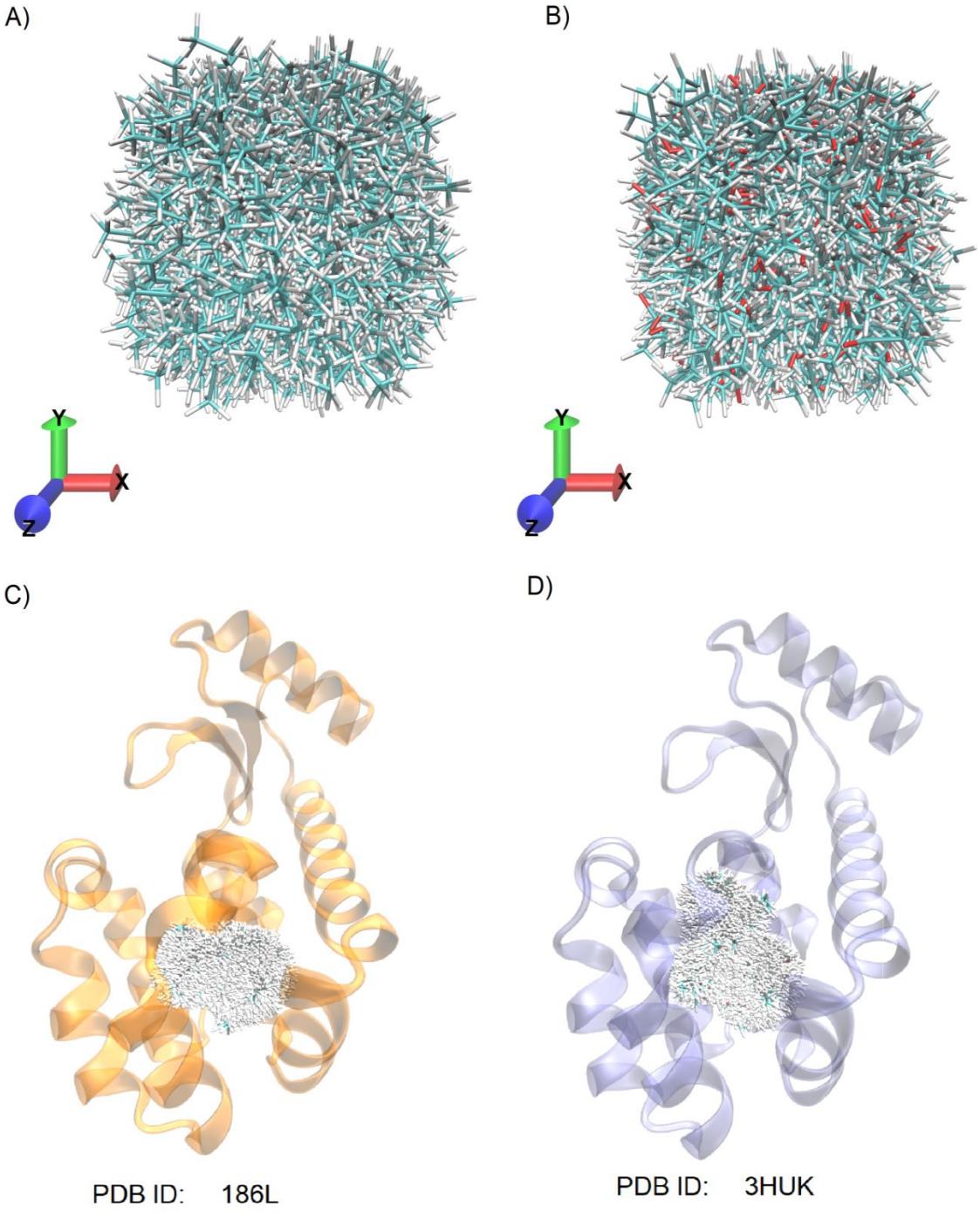



Figura 12 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações.

Fonte: Elaborada pelo autor.

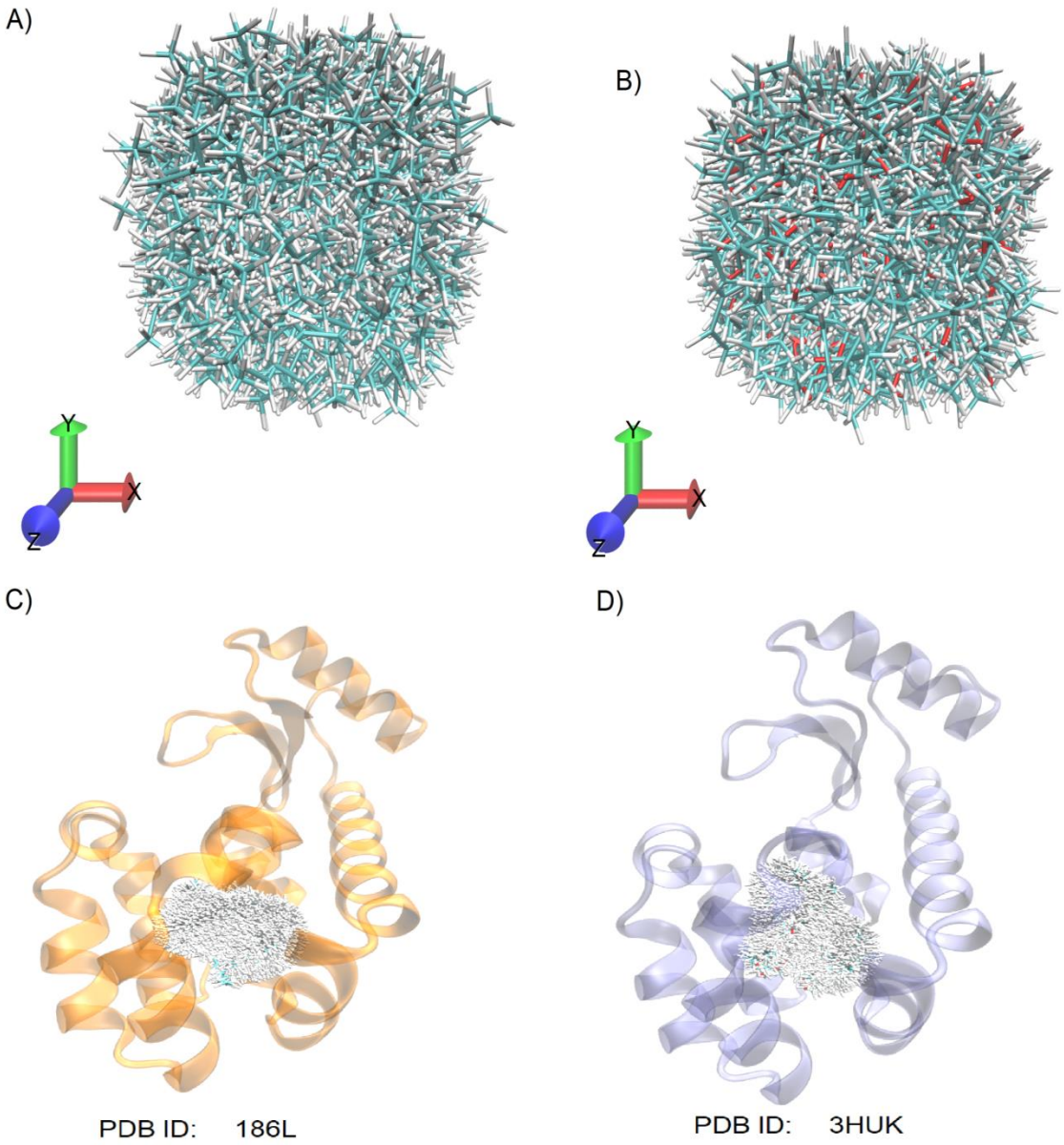



Figura 13 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações.

Fonte: Elaborada pelo autor.

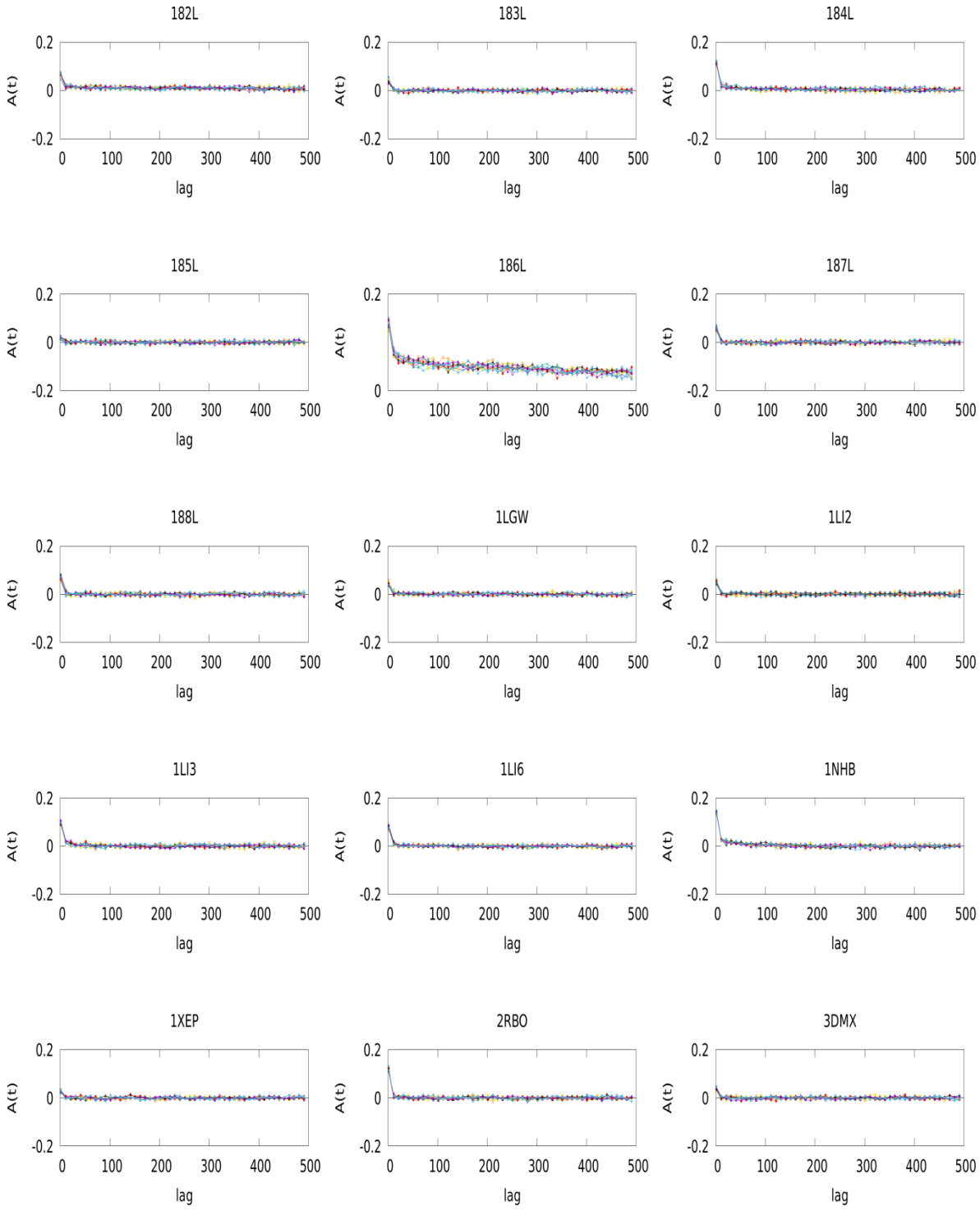
As Figura 12 e Figura **13** apresentam as configurações acessadas pelos ligantes durantes as simulações. Ao compararmos tais figuras vemos que houve a ocupação de todo o espaço de conformações em ambas as situações, tanto para os ligantes livres, que ocuparam toda a caixa de simulação determinada, quanto para os ligantes nos complexo, que ocuparam todo o sítio de interação.

O comportamento dos ligantes durante as simulações, isto é, os deslocamentos no espaço Cartesiano e para os ângulos de Euler, componente a componente, para todos complexos também foi analisado – resultados podem ser visto no **APÊNDICE A – Deslocamentos na Caixa de Simulação** – e mostrou que, no geral, todos os ligantes amostram em todo o espaço de conformações.

#### 4.2.1.2 Autocorrelação

O tempo de correlação é a medida de quão demorada é a transição de um sistema de um estado para outro significativamente diferente do primeiro.[20] Entretanto, embora esse conceito seja aplicado em simulações de DM, o tempo não é um parâmetro considerado para simulações de MC, de modo que uma entidade equivalente seria a correlação entre os passos da simulação,[61] isto é, uma média sobre as configurações geradas de modo que cada estado *i* não deve possuir dependência direta, ou 'memória', do estado *i-2*. Assim, o sistema evoluirá de maneira aleatória de modo a não guardar informações de estados anteriores que possam enviesar os resultados.

A fim de avaliar a correlação entre os passos aceitos durante as simulações de MC e, consequentemente, a eficiência na geração de configurações que não dependessem de todas as anteriores, nós usamos uma função de autocorrelação.[61] As figuras a seguir mostram os resultados obtidos para as análises das autocorrelações dos sistemas.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
A(t)
lag
0.2
0
-0.2
0
100
200
300
400
500

(continua)

(continuação)

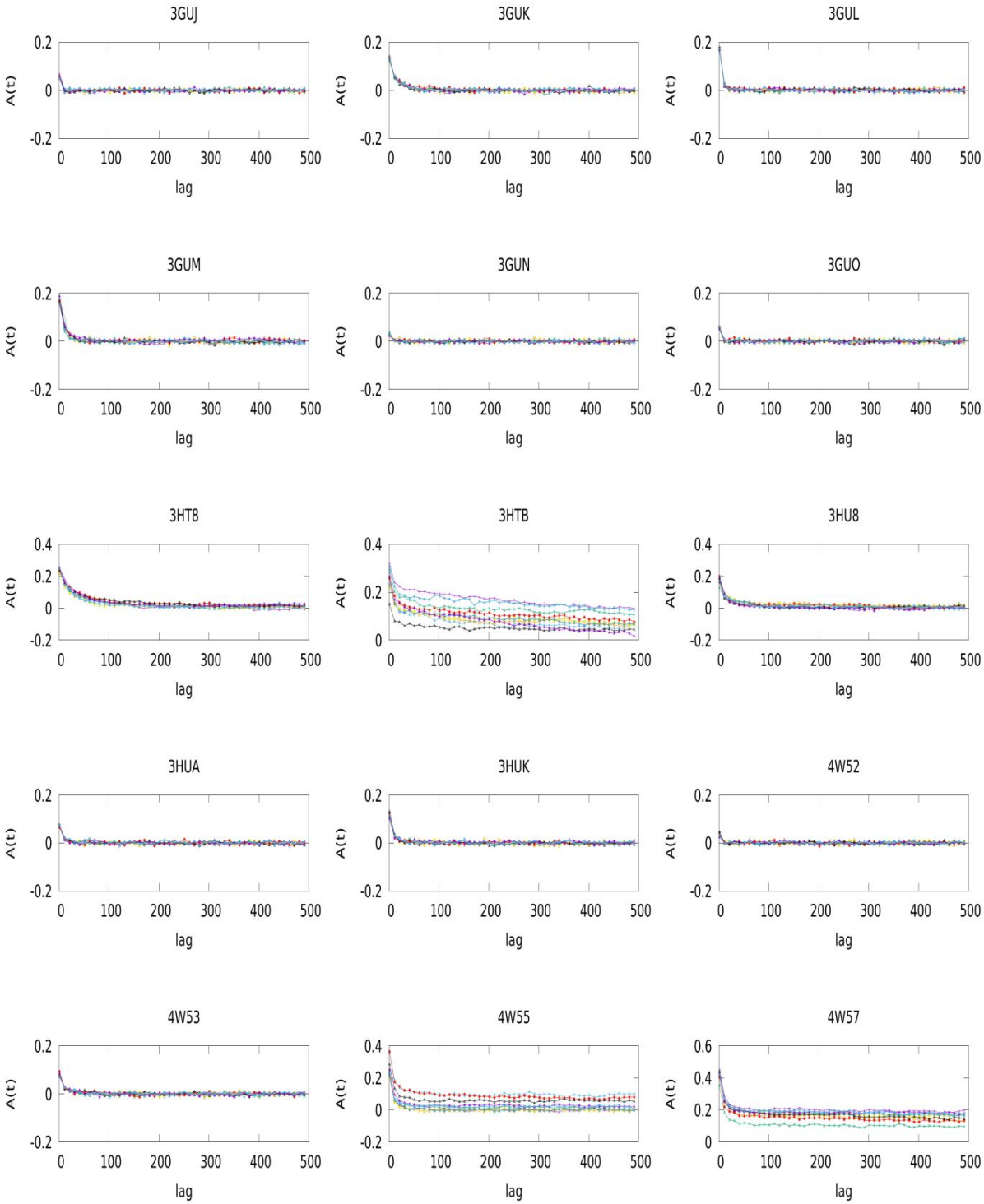
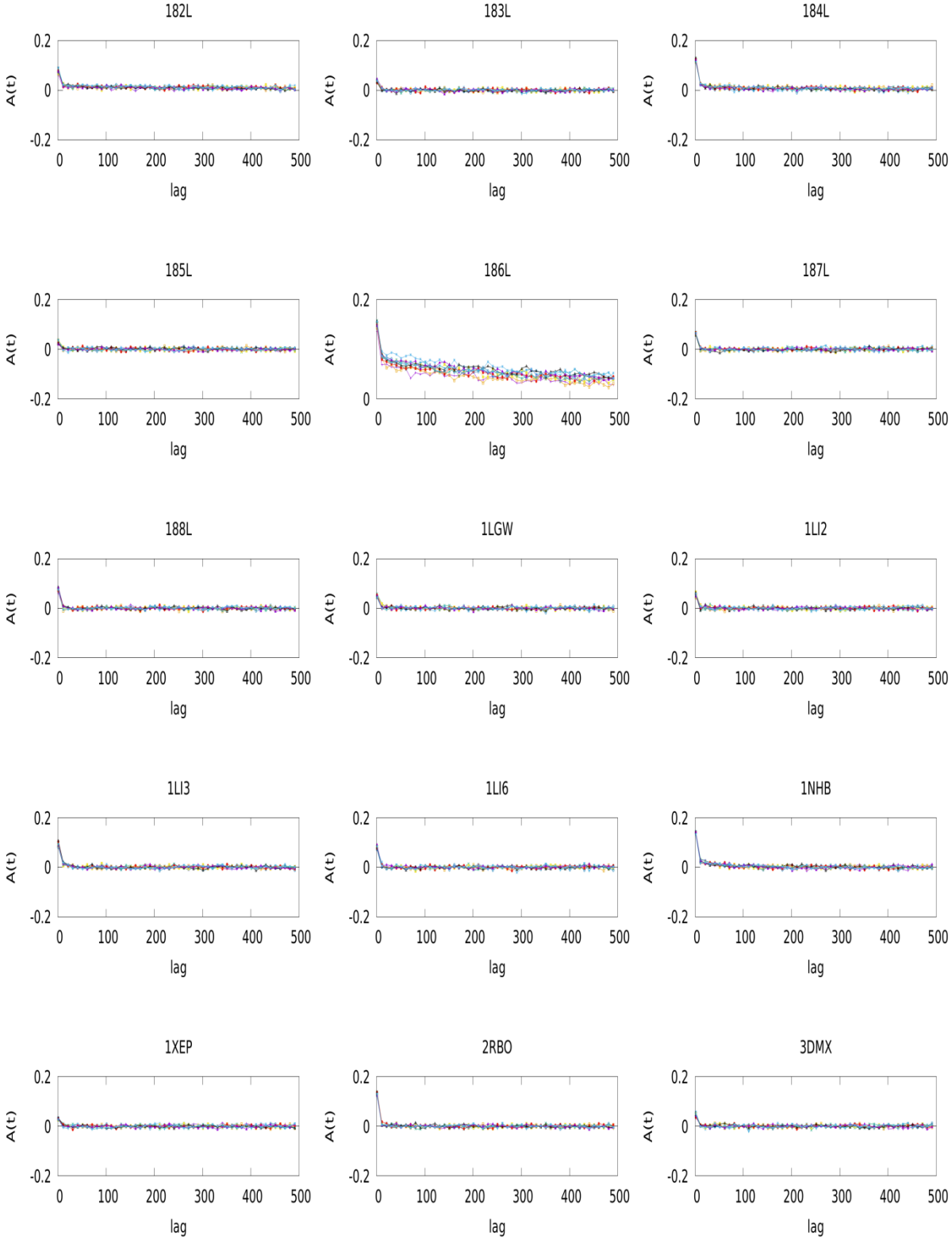



Figura 14 - Autocorrelações obtidas para as simulações de MC para todos os alvos, utilizando amostragem do tipo *sample_torsion*, com temperatura de 250K e Coulomb como modelo eletrostático.

Fonte: Elaborada pelo autor.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
A(t)
lag
0.2
0
-0.2
0
100
200
300
400
500

(continua)

(continuação)

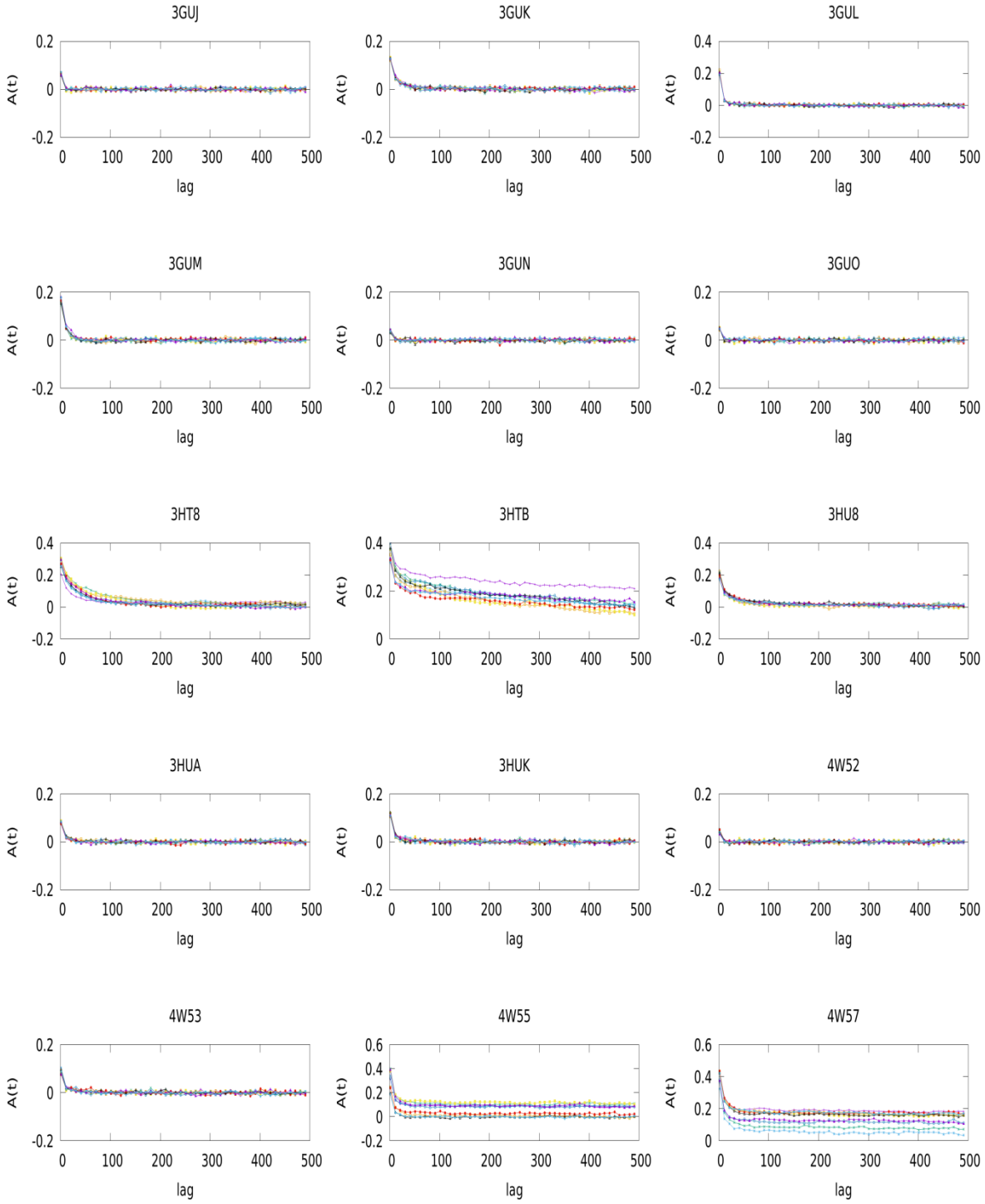



Figura 15 - Autocorrelações obtidas para as simulações de MC para todos os alvos, utilizando amostragem do tipo *sample_torsion*, com temperatura de 250K e Coulomb como modelo eletrostático.

Fonte: Elaborada pelo autor.

Nas Figura 14 e Figura **15** é possível perceber que quase todos os alvos convergem rapidamente para zero, mostrando que a geração de configurações do sistema está de fato descorrelacionada, e fomentando a validade de tal geração. Os ligantes mais discrepantes foram 3HTB, 4W55 e 4W57; todos os três apresentam ligações torcionáveis, o que poderia ser a fonte das diferenças. Entretanto, se olharmos para os valores da autocorrelação apresentados nas Figura 14 e Figura **15**, é possível ver que eles não estão distantes do zero, de modo que a descorrelação entre uma configuração e outra também é válida para esses casos.

A utilização de diferentes tipos de modelos eletrostáticos não influenciou na descorrelação entre as configurações dos sistemas.

#### 4.2.1.3 Convergência

Para avaliar se as amostragens resultantes das simulações de MC foram apropriadas e atingiram o estado de equilíbrio, nós analisamos a convergência para as energias livres de ligação calculadas. A fim de obtê-las, testes com diferentes quantidades de passos de MC foram realizados para os 30 complexos. As figuras a seguir mostram os resultados de convergência para os sistemas, utilizando a temperatura de 250K, amostragem do tipo *sample_torsion* para ambos os modelos eletrostáticos.

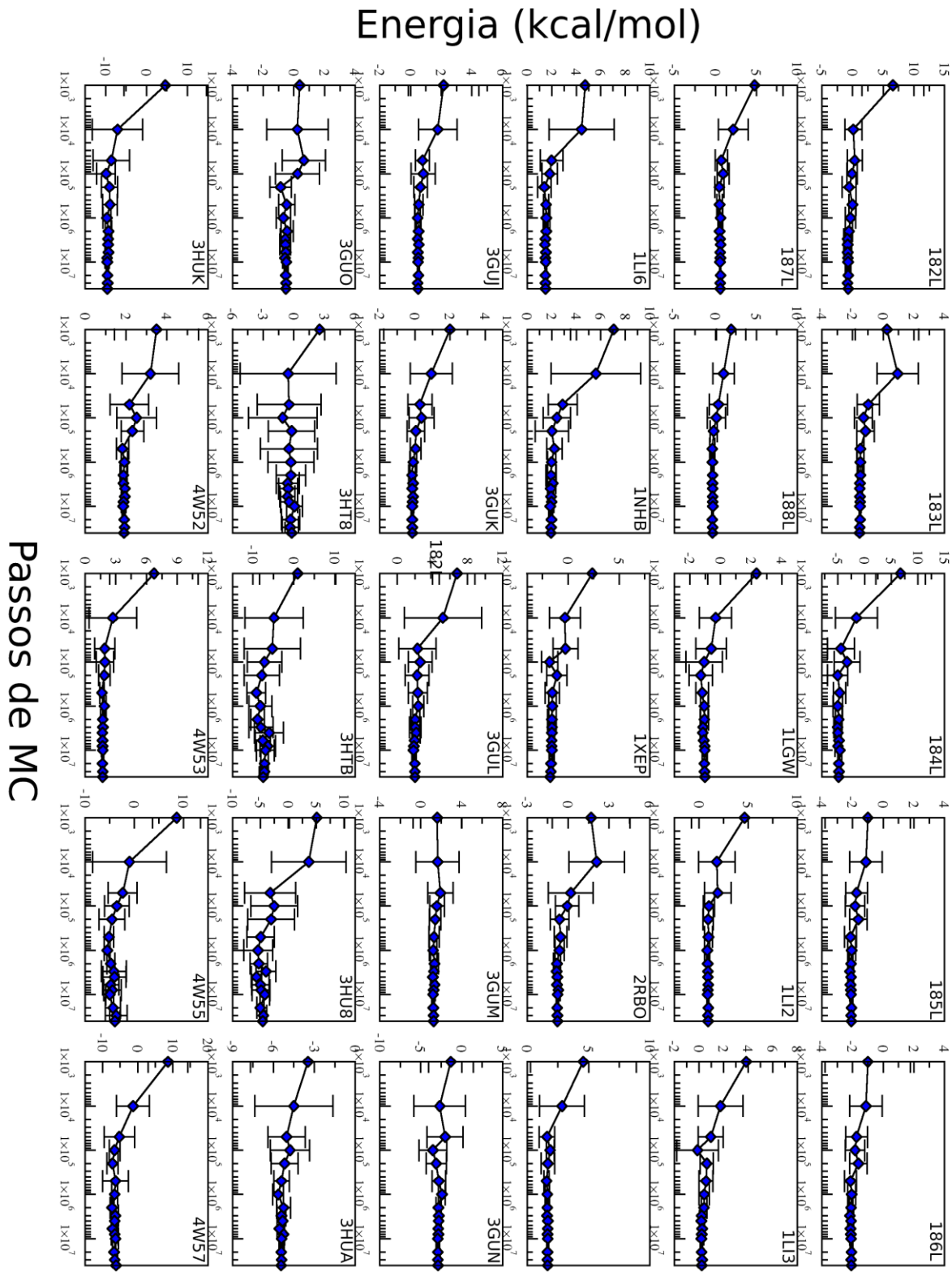



Figura 16 - Convergência das energias livres de ligação para os 30 complexos, em escala logarítmica, para 40 milhões de passos de simulação de MC, usando Coulomb como modelo eletrostático.

Fonte: Elaborada pelo autor.

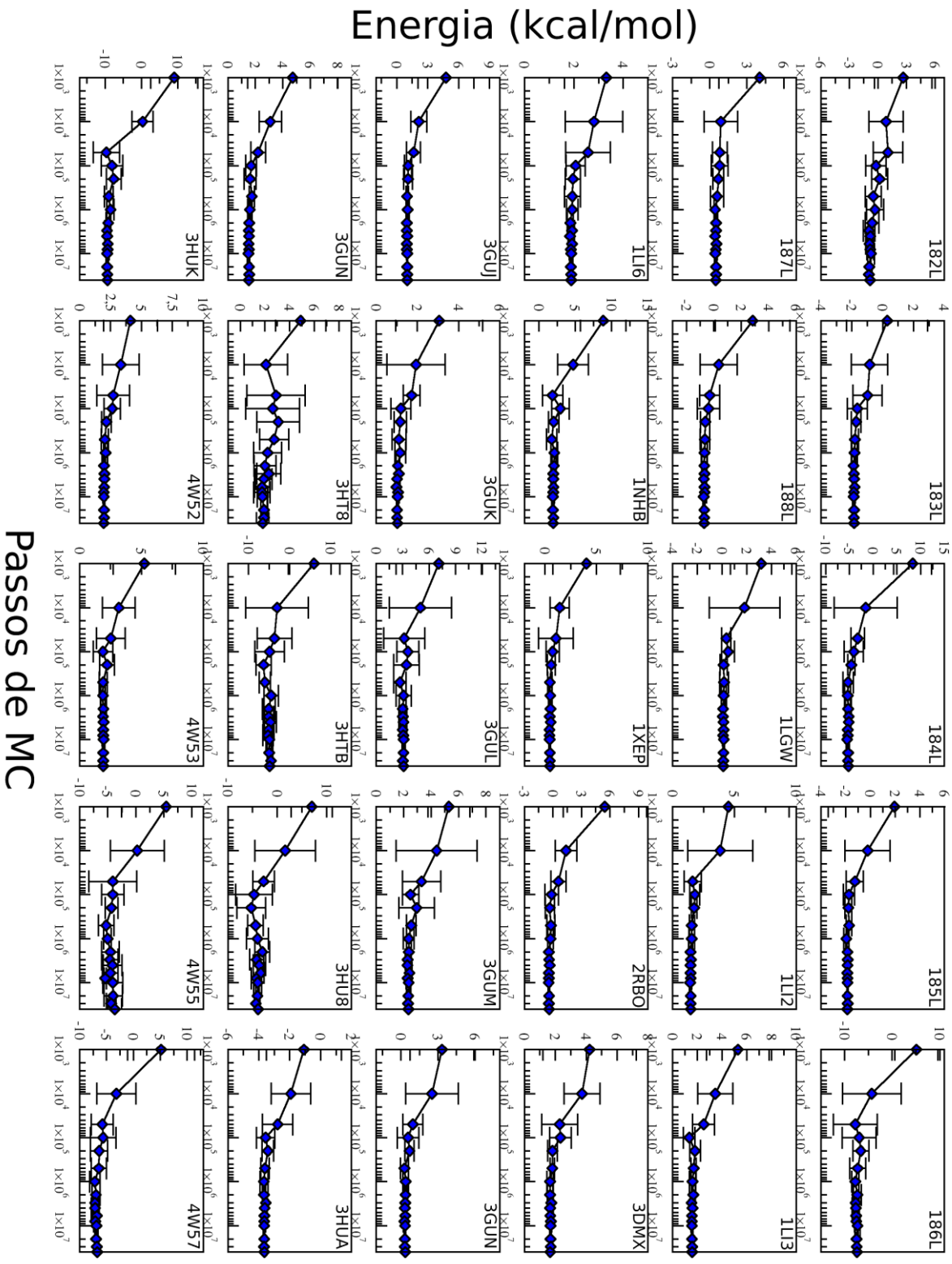



Figura 17 - Convergência das energias livres de ligação para os 30 complexos, em escala logarítmica, para 40 milhões de passos de simulação de MC, usando Poisson como modelo eletrostático.

Fonte: Elaborada pelo autor.

Como mostrado acima, a maior parte dos complexos estudados apresentaram convergência nos cálculos de energia livre de ligação após, aproximadamente, 5 milhões de passos, de modo que definimos $4x10^7$ passos de MC como um valor apropriado para realizar todas as simulações e prosseguir com as análises.

A utilização de modelos eletrostáticos diferentes resultou em valores distintos de energia livre de ligação calculados, entretanto, o comportamento de convergência da energia em, aproximadamente, 5 milhões de passos se manteve.

#### 4.2.1.4 Energia Livre de Ligação para MC

Os cálculos de MC foram feitos a temperatura de Monte Carlo de 250K, para os 30 complexos ligante-receptor. Uma comparação entre os valores determinados experimentalmente e as energias livres calculadas pode ser vista na Tabela 5. As energias experimentais foram obtidas a partir de trabalhos publicados[39-43] e um trabalho de revisão recente feito por Mobley e Gilson[4].

Tabela 5 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $\Delta E_{MC}^{Coulomb}$ (kcal/mol) | $\Delta E_{MC}^{Poisson}$ (kcal/mol) |
|---|---|---|---|
| **182L** | -5,46 | -3,6 ± 0,2 | -3,6 ± 0,1 |
| **183L** | -5,13 | -4,33 ± 0,03 | -4,65 ± 0,03 |
| **184L** | -6,51 | -7,7 ± 0,3 | -8,1 ± 0,2 |
| **185L** | -4,89 | -4,87 ± 0,03 | -4,63 ± 0,02 |
| **186L** | -6,70 | -9,9 ± 0,2 | -10,2 ± 0,2 |
| **187L** | -4,67 | -2,20 ± 0,06 | -0,2 ± 0,7 |
| **188L** | -4,60 | -3,13 ± 0,04 | -3,50 ± 0,04 |
| **1NHB** | -5,76 | -0,82 ± 0,06 | -0,82 ± 0,05 |
| **3DMX** | -5,35 | -1,13 ± 0,04 | -1,12 ± 0,03 |
| **4W52** | -5,19 | -0,86 ± 0,04 | -0,84 ± 0,03 |
| **4W53** | -5,52 | -1,06 ± 0,03 | -0,87 ± 0,04 |
| **4W55** | -6,55 | -7 ± 1 | -7 ± 1 |
| **4W57** | -6,69 | -9,1 ± 0,3 | -9,7 ± 0,2 |
| **1LGW** | -5,53 | -3,83 ± 0,07 | -2,68 ± 0,02 |
| **1LI2** | -5,58 | -1,86 ± 0,03 | -1,35 ± 0,03 |
| **1LI3** | -5,87 | -2,54 ± 0,09 | -1,22 ± 0,04 |
| **1LI6** | -5,25 | -1,32 ± 0,03 | -0,89 ± 0,01 |
| **1XEP** | -4,16 | -4,5 ± 0,2 | -2,31 ± 0,03 |
| **2RBO** | -5,52 | -3,59 ± 0,06 | -3,20 ± 0,03 |
| **3HT8** | -5,04 | -2,9 ± 0,9 | -0,9 ± 0,3 |
| **3HTB** | -5,33 | -10,3 ± 0,8 | -7,7 ± 0,5 |
| **3HU8** | -4,02 | -7,4 ± 0,5 | -6,8 ± 0,3 |
| **3HUA** | -4,61 | -8,19 ± 0,05 | -6,44 ± 0,04 |
| **3HUK** | -4,48 | -12,3 ± 0,1 | -12,4 ± 0,1 |
| **3GUJ** | -5,80 | -2,29 ± 0,02 | -1,77 ± 0,01 |
| **3GUK** | -5,30 | -2,92 ± 0,02 | -1,75 ± 0,02 |
| **3GUL** | -5,50 | -0,8 ± 0,1 | 0,2 ± 0,1 |
| **3GUM** | -4,90 | -1,55 ± 0,09 | -0,47 ± 0,06 |
| **3GUN** | -4,00 | -5,66 ± 0,07 | -2,46 ± 0,03 |
| **3GUO** | -5,00 | -3,27 ± 0,04 | -1,26 ± 0,03 |
| **Correlação de Pearson** | | **0,08** | **0,22** |
| **Correlação por média de Bootstrap** | | **0,16 ± 0,02** | **0,30 ± 0,02** |

Fonte: Elaborada pelo autor.

Ao compararmos as energias livres de ligação calculadas para os 30 complexos com os valores de energia livre experimentais é possível ver que, independentemente do conjunto de parâmetros utilizados para gerar tais energias o valor da correlação entre os dados é bastante baixo. Entretanto, se retirarmos três dos ligantes, 3HU8, 3HUA e 3HUK podemos ver uma tendência maior das energias a ter uma melhor correlação com os dados experimentais. As figuras a seguir mostram esses resultados.

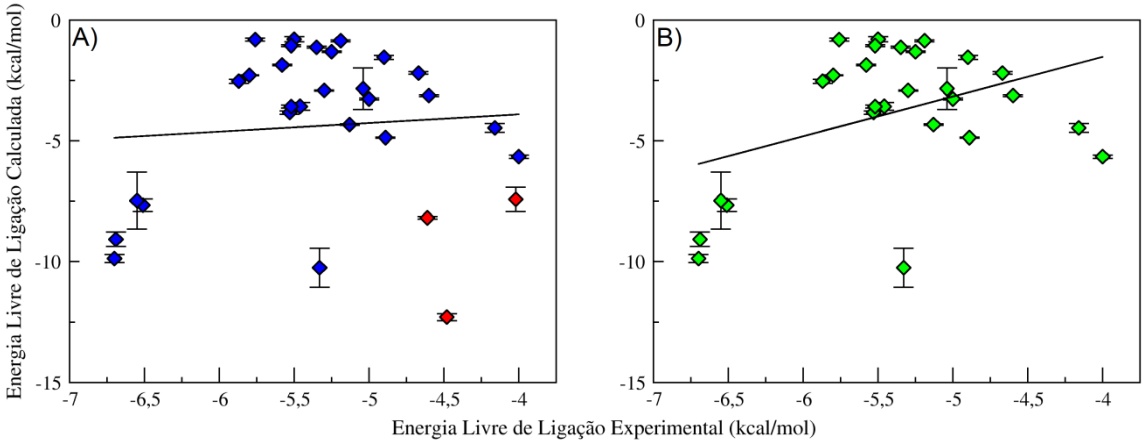



Figura 18 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Utilizando a temperatura de 250K, com Coulomb como modelo eletrostático, obtivemos uma correlação de Pearson de 0,08, enquanto que usando a média pro *bootstrapping* a correlação for de 0,16 ± 0,02 para o conjunto de 30 ligantes (Figura 18 A), com o desvio absoluto médio de 2,61 kcal/mol (N=30). Entretanto, se removermos alguns dos ligantes e olharmos para a Figura 18 B), que possui 27 deles, vemos que a correlação de Pearson aumenta para 0,40, enquanto que usando a média por *bootstrapping* q correlação sobe para 0,46 ± 0,02, com desvio médio absoluto de 2,50 kcal/mol (N=27). Entretanto, é notável a diferença com relação aos valores absolutos. Esse resultado deve-se ao fato de que o modelo utilizado superestima as contribuições apolares das interações bem como a contribuição eletrostática.

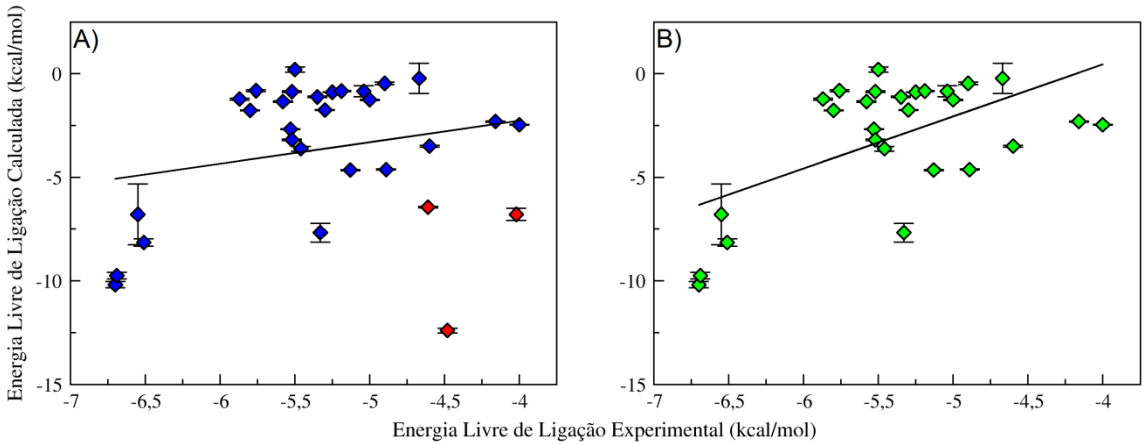



Figura 19 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Ao mudarmos o modelo eletrostático para Poisson observamos um aumento na correlação entre os valores de energia livre de interação experimental e os calculados para os 30 complexos, chegando a 0,22 (para correlação de Pearson) e, por meio da média por *bootstrapping* a correlação foi de 0,30 ± 0,2, com desvio absoluto médio de 2,75 kcal/mol (N=30). Novamente, se retirarmos alguns dos ligantes da análise, Figura 19 B), a correlação aumenta, chegando a 0,58, na correlação de Pearson, e por média de *bootstrapping* chegou a 0,61 ± 0,02, com desvio absoluto médio de 2,33 kcal/mol (N=27). Assim como para o caso anterior, os valores absolutos das energias experimentais não foram obtidos, apenas o perfil de correlação entre elas.

Ao compararmos os modelos utilizados para o cálculo das energias livres de interação fica bastante claro que o método possui limitações, principalmente por não ser capaz de acertar os valores absolutos. Porém, é nítida a melhora nos resultados ao utilizar Poisson como modelo eletrostático, o que era de se esperar, uma vez que esse modelo é mais robusto.

Ainda sim, considerando todos os 30 ligantes, a correlação com os dados experimentais se mostrou baixa, de modo a sugerir que uma mudança deva ser aplicada ao modelo para que possamos melhorar nossos cálculos.

Uma alternativa é utilizar um parâmetro de amortecimento no potencial de Lennard-Jones, para diminuir a intensidade das energias nas regiões de distância interatômicas tendendo a zero, compensando a rigidez do receptor e permitindo que o sistema acesse estados de energia não disponíveis anteriormente,[62] fazendo a componente da energia não-ligada assumir a forma:

$$E_{não-ligado} = \sum_{i \neq j} \left( \frac{A_{ij}}{(r_{ij}+\delta_{LJ})^{12}} - \frac{B_{ij}}{(r_{ij}+\delta_{LJ})^{6}} \right) + \sum_{i \neq j} q_i \Phi(r_i) \qquad (4.1)$$

$q_i$ é a carga do átomo *i*, $r_{ij}$ é a distância entre os átomos *i* e *j*, A e B são os parâmetros de Lennard-Jones para os átomos *i* e *j*, $\Phi(r_i)$ é o potencial eletrostático, (Coulomb ou PB) e $\delta_{LJ}$ é a constante de amortecimento. Temos assim:

Tabela 6 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos, utilizando 0,5 Å como a constante de amortecimento no potencial de Lennard-Jones.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $\Delta E_{MC}^{Coulomb}$ (kcal/mol) | $\Delta E_{MC}^{Poisson}$ (kcal/mol) |
|---|---|---|---|
| **182L** | -5,46 | -3,5 ± 0,2 | -3,5 ± 0,2 |
| **183L** | -5,13 | -4,28 ± 0,04 | -4,59 ± 0,03 |
| **184L** | -6,51 | -7,7 ± 0,2 | -8,2 ± 0,1 |
| **185L** | -4,89 | -4,82 ± 0,06 | -4,60 ± 0,03 |
| **186L** | -6,70 | -10,0 ± 0,2 | -10,2 ± 0,2 |
| **187L** | -4,67 | -2,13 ± 0,03 | -2,25 ± 0,03 |
| **188L** | -4,60 | -3,09 ± 0,05 | -3,44 ± 0,04 |
| **1NHB** | -5,76 | -0,81 ± 0,08 | -0,78 ± 0,06 |
| **3DMX** | -5,35 | -1,09 ± 0,02 | -1,09 ± 0,02 |
| **4W52** | -5,19 | -0,8 ± 0,3 | -0,81 ± 0,02 |
| **4W53** | -5,52 | -1,0 ± 0,3 | -0,84 ± 0,04 |
| **4W55** | -6,55 | -7 ± 1 | -7 ± 1 |
| **4W57** | -6,69 | -9,0 ± 0,5 | -9,7 ± 0,2 |
| **1LGW** | -5,53 | -3,77 ± 0,08 | -2,66 ± 0,03 |
| **1LI2** | -5,58 | -1,85 ± 0,04 | -1,31 ± 0,03 |
| **1LI3** | -5,87 | -2,51 ± 0,05 | -1,17 ± 0,04 |
| **1LI6** | -5,25 | -1,25 ± 0,03 | -0,85 ± 0,03 |
| **1XEP** | -4,16 | -4,4 ± 0,2 | -2,27 ± 0,05 |
| **2RBO** | -5,52 | -3,55 ± 0,08 | -3,17 ± 0,04 |
| **3HT8** | -5,04 | -2,9 ± 0,7 | -0,9 ± 0,2 |
| **3HTB** | -5,33 | -9 ± 1 | -7,6 ± 0,4 |
| **3HU8** | -4,02 | -7,4 ± 0,07 | -6,9 ± 0,3 |
| **3HUA** | -4,61 | -8,14 ± 0,05 | -6,38 ± 0,02 |
| **3HUK** | -4,48 | -12,4 ± 0,2 | -12,22 ± 0,09 |
| **3GUJ** | -5,80 | -2,25 ± 0,03 | -1,72 ± 0,02 |
| **3GUK** | -5,30 | -2,85 ± 0,03 | -1,70 ± 0,02 |
| **3GUL** | -5,50 | -0,74 ± 0,09 | 0,3 ± 0,2 |
| **3GUM** | -4,90 | -1,46 ± 0,09 | -0,34 ± 0,07 |
| **3GUN** | -4,00 | -5,61 ± 0,09 | -2,41 ± 0,03 |
| **3GUO** | -5,00 | -3,26 ± 0,08 | -1,21 ± 0,03 |
| **Correlação de Pearson** | | **0,07** | **0,21** |
| **Correlação por média de Bootstrap** | | **0,15 ± 0,02** | **0,29 ± 0,02** |

Fonte: Elaborada pelo autor.

Novamente, se olharmos para as correlações para os 30 complexos vemos valores bastante baixos, porém, assim como no caso anterior, ao retirarmos os mesmos três ligantes, observamos uma melhora na tendência de correlação entre as energias experimentais e as calculadas. A seguir estão representados os gráficos.

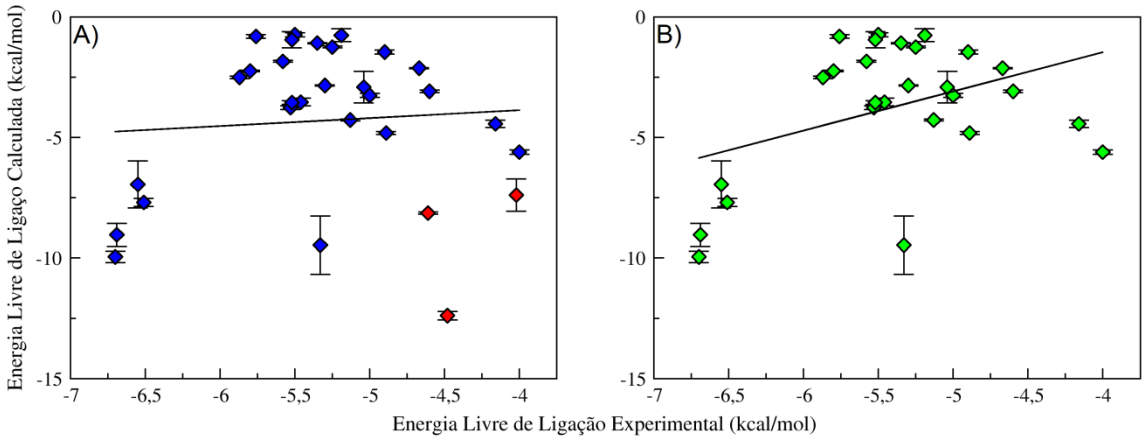



Figura 20 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático e 0,5 Å como constante de amortecimento. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Ao utilizar 0,5 Å como constante de amortecimento e Coulomb como modelo eletrostático obtivemos uma correlação de Pearson de 0,07 para o conjunto dos 30 ligantes, e 0,15 ± 0,02 por meio da média por bootstrapping, com desvio médio absoluto de 2,57 kcal/mol (N=30). Ao removermos os três ligantes, a correlação de Pearson sobe para 0,40, enquanto a correlação média por bootstrapping vai para 0,47 ± 0,02, com desvio médio absoluto de 2,30 kcal/mol (N=27). Aqui, como nos casos anteriores, os valores absolutos das energias não foram obtidos, apenas o perfil correspondente.

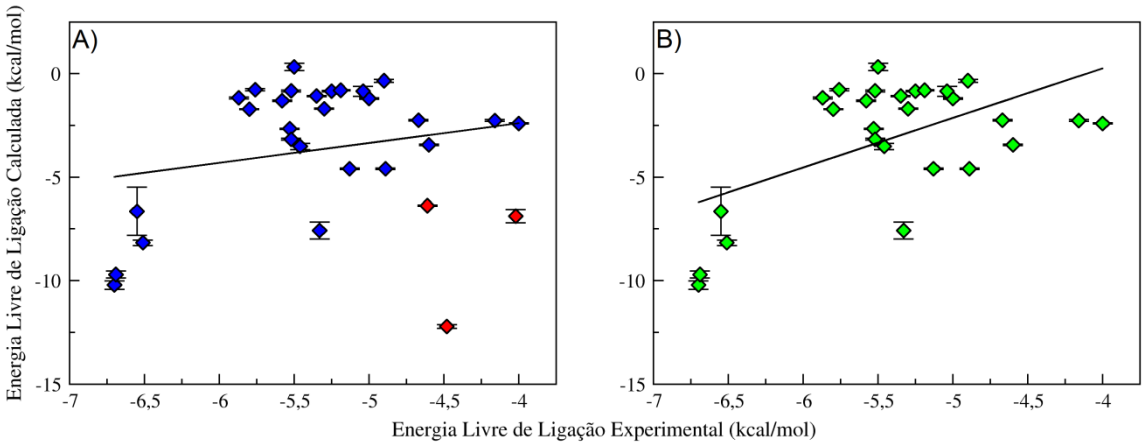



Figura 21 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Poisson como modelo eletrostático e 0,5 Å como constante de amortecimento. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Utilizando Poisson como modelo eletrostático e 0,5 Å como constante de amortecimento, a correlação de Pearson entre os dados experimentais e os calculados foi de 0,21, enquanto a correlação por média de *bootstrapping* foi de 0,29 ± 0,02, com desvio médio absoluto de 2,71 kcal/mol (N=30). Ao retirarmos os três ligantes, como nos casos anteriores, os valores de correlação entre os dados sobem, com a correlação de Pearson chegando a 0,56 enquanto a correlação por média de bootstrapping atinge 0,61 ± 0,02, com desvio médio absoluto de 2,47 kcal/mol.

Aqui, utilizar Poisson com modelo eletrostático também resultou numa melhora significativa na correlação entre os dados experimentais e aqueles obtidos a partir das simulações de MC. Entretanto, o amortecimento não apresentou resultados muito melhores do que aqueles obtidos sem a utilização de tal estratégia. Dessa forma, uma melhora nas simulações ainda era necessária.

Para os modelos anteriores, o termo correspondente à solvatação era levado em conta para os cálculos das energias livres de interação. Uma vez que os sistemas biológicos macromoleculares e interagentes costumam estar imersos em água, utilizar um modelo que leve em consideração o solvente, e que tem apresentado resultados interessantes,[63-65] se mostra uma alternativa para ser

adicionada aos nossos cálculos. Assim, utilizando o modelo proposto por Stouten[63] e Verkhivker[64], o potencial não ligado assume a seguinte forma:

$$E_{não-ligado} = \sum_{i \neq j} \left( \frac{A_{ij}}{(r_{ij})^{12}} - \frac{B_{ij}}{(r_{ij})^{6}} \right) + \sum_{i \neq j} q_i \Phi(r_i) \qquad (4.2)$$

$$+ \sum_{i}^{lig} \sum_{j}^{rec} S_i X_j + S_j X_i$$

$q_i$ é a carga do átomo *i*, $r_{ij}$ é a distância entre os átomos *i* e *j*, A e B são os parâmetros de Lennard-Jones para os átomos *i* e *j*, $\Phi(r_i)$ é o potencial eletrostático, (Coulomb ou PB), $S_i$ é o parâmetro de solvatação atômica e é proporcional ao carga $q_i$, ($S_i = \alpha q_i^2 + \beta$, com α e β ajustáveis), $X_j$ é o volume de dessolvatação, dado por

$$X_j = \left(\frac{4}{3}\pi\right) a^3 \frac{\exp\left[-r_{ij}^2/(2\sigma^2)\right]}{\sigma^3} \qquad (4.3)$$

onde *a* é o raio atômico do átomo *j*, e σ = 3,5 Å. Os resultados utilizando esse modelo podem ser vistos a seguir.

Tabela 7 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, para os dois modelos eletrostáticos, utilizando os como parâmetros de solvatação α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $\Delta E_{MC}^{Coulomb}$ (kcal/mol) | $\Delta E_{MC}^{Poisson}$ (kcal/mol) |
|---|---|---|---|
| **182L** | -5,46 | -0,7 ± 0,1 | -0,7 ± 0,2 |
| **183L** | -5,13 | -2,02 ± 0,03 | -2,34 ± 0,02 |
| **184L** | -6,51 | -5,5 ± 0,2 | -5,8 ± 0,1 |
| **185L** | -4,89 | -1,65 ± 0,03 | -1,45 ± 0,02 |
| **186L** | -6,70 | -7,5 ± 0,2 | -7,8 ± 0,2 |
| **187L** | -4,67 | -0,21 ± 0,04 | -0,32 ± 0,04 |
| **188L** | -4,60 | -1,20 ± 0,03 | -1,59 ± 0,03 |
| **1NHB** | -5,76 | 1,34 ± 0,09 | 1,33 ± 0,07 |
| **3DMX** | -5,35 | 0,95 ± 0,03 | 0,97 ± 0,03 |
| **4W52** | -5,19 | 1,21 ± 0,03 | 1,27 ± 0,03 |
| **4W53** | -5,52 | 1,08 ± 0,04 | 1,25 ± 0,03 |
| **4W55** | -6,55 | -4,2 ± 0,9 | -4 ± 1 |
| **4W57** | -6,69 | -6,7 ± 0,4 | -7,3 ± 0,1 |
| **1LGW** | -5,53 | 2,6 ± 00,1 | 3,64 ± 0,03 |
| **1LI2** | -5,58 | 2,61 ± 0,02 | 3,14 ± 0,03 |
| **1LI3** | -5,87 | 2,03 ± 0,07 | 3,37 ± 0,03 |
| **1LI6** | -5,25 | 1,28 ± 0,02 | 1,69 ± 0,03 |
| **1XEP** | -4,16 | 1,4 ± 0,1 | 3,63 ± 0,03 |
| **2RBO** | -5,52 | 0,44 ± 0,05 | 0,79 ± 0,04 |
| **3HT8** | -5,04 | 2,1 ± 0,9 | 3,9 ± 0,3 |
| **3HTB** | -5,33 | -5,3 ± 0,9 | -2,7 ± 0,5 |
| **3HU8** | -4,02 | -2,1 ± 0,6 | -1,7 ± 0,5 |
| **3HUA** | -4,61 | -4,05 ± 0,09 | -2,28 ± 0,03 |
| **3HUK** | -4,48 | -5,6 ± 0,2 | -5,5 ± 0,2 |
| **3GUJ** | -5,80 | 0,54 ± 0,03 | 1,05 ± 0,02 |
| **3GUK** | -5,30 | -0,17 ± 0,05 | 0,99 ± 0,02 |
| **3GUL** | -5,50 | 2,0 ± 0,2 | 3,0 ± 0,1 |
| **3GUM** | -4,90 | 0,98 ± 0,06 | 2,10 ± 0,05 |
| **3GUN** | -4,00 | 1,04 ± 0,09 | 4,27 ± 0,02 |
| **3GUO** | -5,00 | 1,5 ± 0,1 | 3,54 ± 0,03 |
| **Correlação de Pearson** | | **0,33** | **0,42** |
| **Correlação por média de Bootstrap** | | **0,40 ± 0,02** | **0,48 ± 0,02** |

Fonte: Elaborada pelo autor.

Como pode ser visto na Tabela 7, ao adicionarmos o termo correspondente à solvatação aparece uma melhora significativa na correlação entre os dados experimentais e os calculados comparando os 30 complexos. Entretanto, apesar de ser melhor que os outros modelos de MC apresentados aqui, os valores de correlação ainda foram baixos. Se retirarmos os mesmos três pontos que retiramos anteriormente, teremos os resultados que são apresentados nas figuras a seguir.

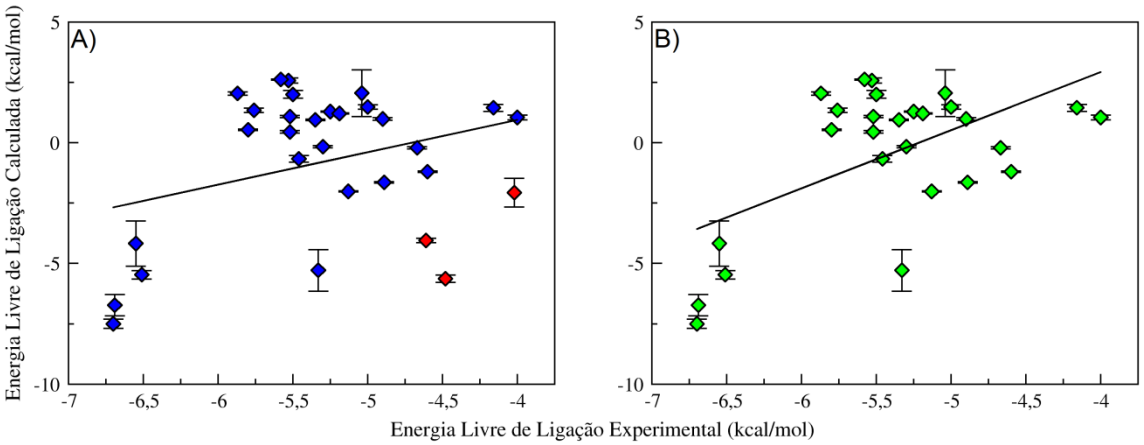



Figura 22 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Coulomb como modelo eletrostático e α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol como parâmetros de solvatação. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Utilizando Coulomb como modelo eletrostático, α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol como parâmetros de solvatação a correlação de Pearson obtida para os 30 ligantes complexados foi de 0,33, enquanto que pela média de *bootstrapping* a correlação foi de 0,40 ± 0,02, com desvio médio absoluto de 2,47 kcal/mol (N=30). Ao retirarmos os ligantes 3HU8, 3HUA e 3HUK a correlação de Pearson sobe para 0,57, enquanto por meio da média de *bootstrapping* o valor vai para 0,62 ± 0,02, com desvio médio absoluto de 2,32 kcal/mol (N=27).

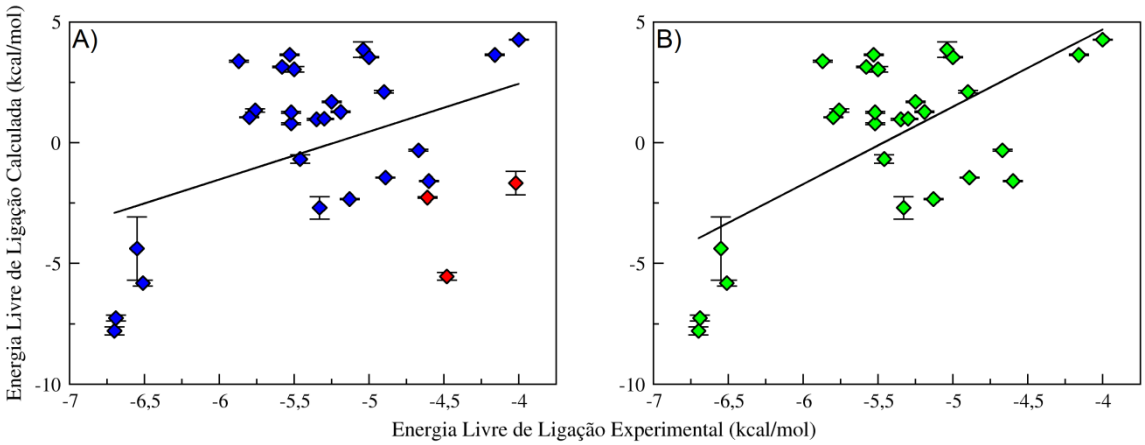



Figura 23 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando Poisson como modelo eletrostático e α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol como parâmetros de solvatação. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Observando a Figura 23 é possível ver que utilizando Poisson como modelo eletrostático e os mesmos valores indicados anteriormente para os parâmetros de solvatação, a correlação de Pearson obtida para os 30 ligantes foi de 0,42, enquanto que por meio da média de *bootstrapping* o valor foi de 0,48 ± 0,02, com desvio médio absoluto de 2,81 kcal/mol (N=30). Se retirarmos os mesmos três ligantes, indicados em vermelho no gráfico, a correlação entre os dados calculados e os experimentais sobe para 0,64, ao passo que por meio da média de *bootstrapping* a correlação sobe para 0,68 ± 0,02, com desvio médio absoluto de 2,72 kcal/mol (N=27).

Assim como todos os casos anteriores, adicionar os termos de solvatação à energia livre de ligação não resultou nos valores absolutos, porém, o perfil de tendência a se correlacionar também foi observado.

#### 4.2.1.5 Aproximação de Primeira Ordem

O cálculo da entropia para os ligantes pode ser feito por meio do método de aproximação de primeira ordem descrito anteriormente. Para a análise desse

método, apenas as simulações do tipo *sample_torsion* tiveram seus valores de entropia configuracional calculados, além disso, apenas aqueles receptores que possuem somente a mutação L99A e os que possuem as mutações L99A e M102 E foram considerados (isto é, para apenas L99A: 182L, 183L, 184L, 185L, 186L, 187L, 188L, 1NHB, 3DMX, 4W52, 4W53, 4W55 e 4W57; e para L99A/M102E: 3GUJ, 3GUK, 3GUL, 3GUM, 3GUN 3GUO), cujos valores de entalpia e entropia foram obtidos a partir de calorimetria de titulação isotérmica, por Morton e seus colaboradores[39] no primeiro caso e Liu e colaboradores no segundo.[66]

A Tabela 8, apresenta os valores de energias, entalpias e entropias experimentais e os valores calculados por meio das simulações de Monte Carlo, com suas respectivas correlações, utilizando a temperatura de 250K, Coulomb como modelo eletrostático, sem levar em consideração a solvatação, mas adicionando a constante de amortecimento de 0,5 Å.

Tabela 8 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Coulomb como modelo eletrostático, sem solvatação, com 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta H_{exp}$ (kcal/mol) | $-T\Delta S_{exp}$ (kcal/mol) | $\Delta G_{calc}$ (kcal/mol) | $\Delta H_{calc}$ (kcal/mol) | $-T\Delta S_{calc}$ (kcal/mol) |
|---|---|---|---|---|---|---|
| **182L** | -5,46 | -8,04 | 2,58 | -3,53 | -5,38 | 1,85 |
| **183L** | -5,13 | -8,31 | 3,18 | -4,28 | -6,19 | 1,91 |
| **184L** | -6,51 | -7,09 | 0,56 | -7,70 | -9,57 | 1,87 |
| **185L** | -4,89 | -11,23 | 6,34 | -4,82 | -6,80 | 1,98 |
| **186L** | -6,70 | -8,06 | 1,36 | -9,96 | -11,92 | 1,96 |
| **187L** | -4,67 | -6,97 | 2,30 | -2,13 | -3,73 | 1,60 |
| **188L** | -4,90 | -8,45 | 3,85 | -3,09 | -4,64 | 1,55 |
| **1NHB** | -5,76 | -6,76 | 1,00 | -0,81 | -2,50 | 1,70 |
| **3DMX** | -5,35 | -6,32 | 1,13 | -1,09 | -2,78 | 1,69 |
| **4W52** | -5,19 | -6,32 | 1,13 | -0,76 | -2,25 | 1,49 |
| **4W53** | -5,52 | -6,53 | 1,10 | -0,95 | -2,64 | 1,69 |
| **4W55** | -6,55 | -9,97 | 3,42 | -6,95 | -8,89 | 1,94 |
| **4W57** | -6,69 | -8,06 | 1,36 | -9,04 | -11,03 | 1,99 |
| **3GUJ** | -5,80 | -6,00 | 0,19 | -2,25 | -4,10 | 1,85 |
| **3GUK** | -5,30 | -5,30 | 0,03 | -2,85 | -4,85 | 2,00 |
| **3GUL** | -5,50 | -5,00 | -0,50 | -0,74 | -2,69 | 1,95 |
| **3GUM** | -4,90 | -3,70 | -1,16 | -1,46 | -3,32 | 1,86 |
| **3GUN** | -4,00 | -8,10 | 4,10 | -5,61 | -7,59 | 1,97 |
| **3GUO** | -5,00 | -9,20 | 4,20 | -3,26 | -5,20 | 1,94 |
| **Correlação de Pearson** | | | | 0,54 | 0,53 | 0,09 |
| **Correlação por média de Bootstrap** | | | | **0,54 ± 0,03** | **0,50 ± 0,01** | **0,13 ± 0,01** |

Fonte: Elaborada pelo autor.

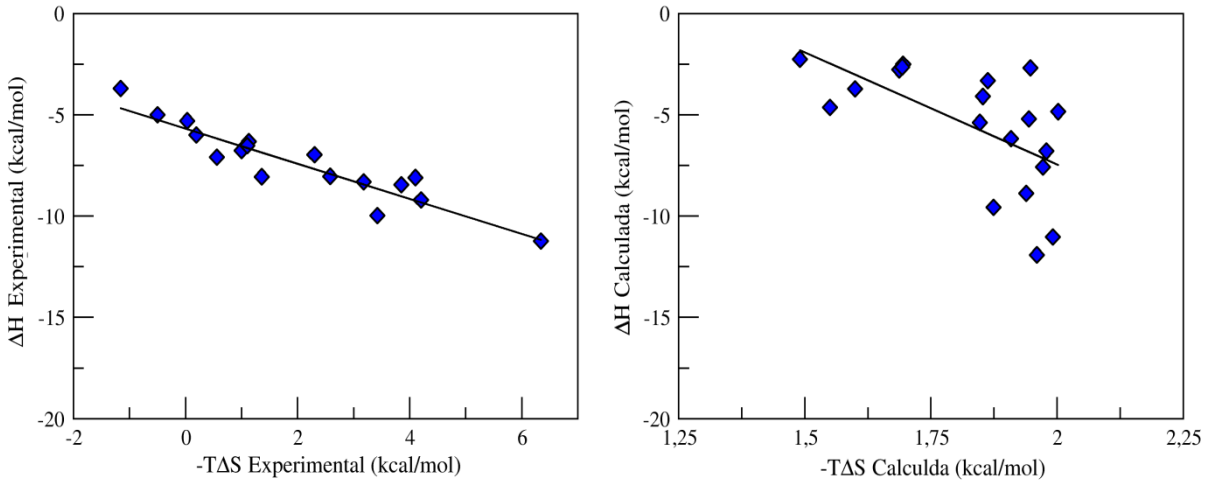



Figura 24 - Comparação das correlações entre os valores experimentais e os valores calculados de entropia (-TΔS) e entalpia (ΔH). Valores calculados utilizando Coulomb como modelo eletrostático e 0,5 Å como constante de amortecimento.

Fonte: Elaborada pelo autor.

A Figura 24 A) compara os valores de entalpia com os valores de entropia obtidos experimentalmente. A correlação de Pearson entre essas grandezas foi de -0,92. Dessa forma, o que esperaríamos ver na correlação entre os valores calculados seria um perfil parecido a esse. Assim, a Figura 24 B) apresenta os resultados obtidos na qual, o valor da correlação foi de r = -0,60, o que não é tão bom quanto a comparação entre os dados experimentais, mas é consideravelmente boa, apresentando o mesmo perfil.

Da Tabela 8 é possível ver que a contribuição entálpica é muito mais considerável na composição das energias livres de ligação do que a contribuição entrópica. Além disso, os valores estimados não são os mesmos dos valores absolutos.

A Tabela 9, apresenta os valores de energias, entalpias e entropias experimentais e os valores calculados, com suas respectivas correlações, Poisson como modelo eletrostático, 250K como temperatura e também sem levar em consideração a solvatação, mas adicionando a constante de amortecimento de 0,5Å.

Tabela 9 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Poisson como modelo eletrostático, sem solvatação, com 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta H_{exp}$ (kcal/mol) | $-T\Delta S_{exp}$ (kcal/mol) | $\Delta G_{calc}$ (kcal/mol) | $\Delta H_{calc}$ (kcal/mol) | $-T\Delta S_{calc}$ (kcal/mol) |
|---|---|---|---|---|---|---|
| **182L** | -5,46 | -8,04 | 2,58 | -3,52 | -5,34 | 1,82 |
| **183L** | -5,13 | -8,31 | 3,18 | -4,59 | -6,49 | 1,90 |
| **184L** | -6,51 | -7,09 | 0,56 | -8,17 | -10,04 | 1,87 |
| **185L** | -4,89 | -11,23 | 6,34 | -4,60 | -6,57 | 1,97 |
| **186L** | -6,70 | -8,06 | 1,36 | -10,22 | -12,16 | 1,94 |
| **187L** | -4,67 | -6,97 | 2,30 | -2,25 | -3,84 | 1,59 |
| **188L** | -4,90 | -8,45 | 3,85 | -3,44 | -4,98 | 1,54 |
| **1NHB** | -5,76 | -6,76 | 1,00 | -0,78 | -2,46 | 1,68 |
| **3DMX** | -5,35 | -6,32 | 1,13 | -1,09 | -2,76 | 1,68 |
| **4W52** | -5,19 | -6,32 | 1,13 | -0,81 | -2,45 | 1,65 |
| **4W53** | -5,52 | -6,53 | 1,10 | -0,84 | -2,71 | 1,87 |
| **4W55** | -6,55 | -9,97 | 3,42 | -6,65 | -8,55 | 1,90 |
| **4W57** | -6,69 | -8,06 | 1,36 | -9,70 | -11,70 | 1,99 |
| **3GUJ** | -5,80 | -6,00 | 0,19 | -1,72 | -3,56 | 1,84 |
| **3GUK** | -5,30 | -5,30 | 0,03 | -1,70 | -3,69 | 2,00 |
| **3GUL** | -5,50 | -5,00 | -0,50 | 0,32 | -1,60 | 1,92 |
| **3GUM** | -4,90 | -3,70 | -1,16 | -0,34 | -2,19 | 1,84 |
| **3GUN** | -4,00 | -8,10 | 4,10 | -2,41 | -4,36 | 1,95 |
| **3GUO** | -5,00 | -9,20 | 4,20 | -1,21 | -3,14 | 1,93 |
| **Correlação de Pearson** | | | | 0,66 | 0,51 | 0,04 |
| **Correlação por média de Bootstrap** | | | | **0,66 ± 0,02** | **0,45 ± 0,01** | **0,08 ± 0,01** |

Fonte: Elaborada pelo autor.

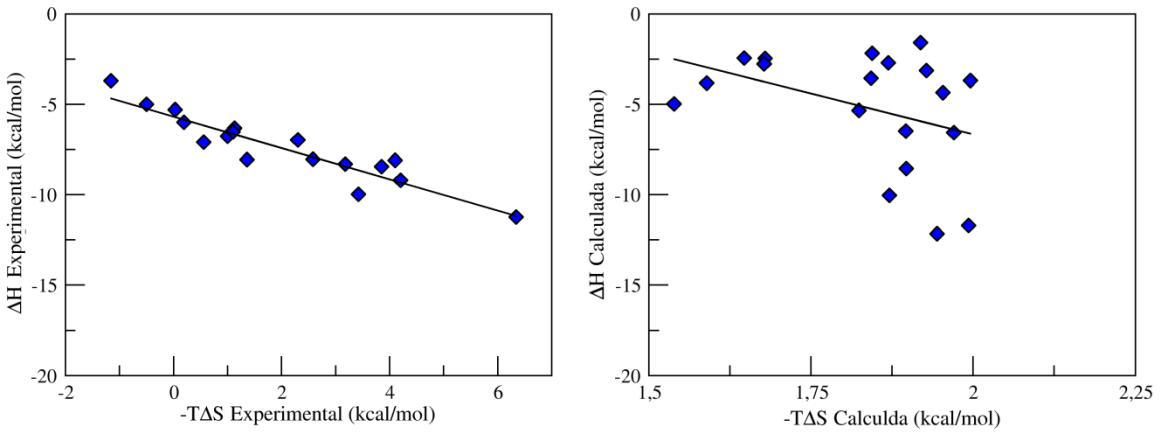



Figura 25 - Comparação das correlações entre os valores experimentais e os valores calculados de entropia ($-T\Delta S$) e entalpia ($\Delta H$). Valores calculados utilizando Poisson como modelo eletrostático e 0,5 Å como constante de amortecimento.

Fonte: Elaborada pelo autor.

A Figura 25 mostra novamente a correlação entre os valores experimentais de entalpia e entropia (r = -0,92) e apresenta ao lado uma comparação entre os valores calculados utilizando Poisson como modelo eletrostático. A correlação obtida nesse caso foi de -0,39, indicando que para essa situação, o modelo anterior foi melhor.

Da Tabela 9 também fica claro que a contribuição da entalpia para compor a energia livre de ligação é maior do que a contribuição da entropia e, novamente, os valores estimados aqui não acertaram os valores experimentais.

### *4.2.2 Full Flex*

Embora os resultados anteriores tenham se mostrado relativamente satisfatórios, em especial quando retiramos três dos 30 ligantes analisados, o método utilizado para gerar as configurações geradas pelas simulações não segue propriamente o rigor de uma amostragem de Monte Carlo com o algoritmo de Metropolis. Isso acontece devido à dificuldade em se fazer simulações no chamado 'espaço de diedros', o qual requer o uso de um termo de correção Jacobiano para satisfazer o princípio da reversibilidade e amostrar ângulos de diedro de forma não enviesada.[67] O exemplo a seguir demostra esta dificuldade.

Consideremos, como exemplo, um sistema composto por uma molécula diatômica, com átomos A e B, numa caixa 2D de largura *L*, cuja função de partição seja dada por

$$Z = \frac{1}{\lambda_A^2 \lambda_B^2} \int_0^L dx^A \int_0^L dy^A \int_0^L dx^B \int_0^L dy^B \, e^{-U(r)/k_B T} \qquad (4.4)$$

onde $\lambda^A$ é o comprimento de onda térmico para o átomo A, $\lambda^B$ para o átomo B, $x^A$, $x^B$, $y^A$ e $y^B$ são as coordenadas cartesianas para ambos os átomos, *T* é a temperatura do sistema, $k_B$ é a constante de Boltzmann. $U(r) = k(r - r_0)^2/2$, com *r* sendo a separação entre os átomos, $r_0$ é o comprimento de equilíbrio da ligação. $r_0 \ll L$ e $\lambda \ll L$. A dificuldade em resolver a equação (4.4) deve-se ao fato de não ser muito claro como fazer as integrações em *x* e *y*, dado que *U(r)* também depende de *x* e *y*, uma vez que *r* = *r(x,y)*.[68]

A alternativa é usar as coordenadas internas, ou coordenadas generalizadas, que são, basicamente, coordenadas relativas, em oposição às coordenadas absolutas.[68]

Ao usarmos coordenadas internas é possível mudar nossas variáveis de integração para que elas correspondam à energia potencial. Assim, a configuração cartesiana, isto é, $x^A$, $x^B$, $y^A$ e $y^B$, pode ser descrita como um $x^A$, $y^A$, $r$ e $\Phi$, onde $\Phi$ é um ângulo dado à orientação da molécula baseado em qualquer valor de referência escolhido. Aqui escolheremos A como tal.

Apesar de parecer simples, não é trivial mudar para coordenadas internas na função de partição, uma vez que é necessário preservar o sentido físico da função de partição, o que implicará em preservar seu valor matemático. Porém, se só integrarmos as variáveis ($x^A$, $y^A$, $r$, $\Phi$) estaremos cometendo um erro, porque o significado de uma função de partição do tipo (4.4) diz que todas as possíveis configurações cartesianas, isto é, todos os valores de $x^A$, $x^B$, $y^A$, $y^B$, são examinados numa mesma base e, então, pesados por um fator de Boltzmann. Se pensarmos em avaliar a integral amostrando todo o espaço, podemos pegar um número igual de pontos para todo pequeno “volume” num espaço de configurações quadridimensional e avaliar o fator de Boltzmann a cada ponto.[68]

A sutileza vem de examinarmos todas as possíveis posições de B dado um A fixo. Os valores cartesianos de $x^B$, $y^B$ devem ser conferidos numa base uniforme (antes de pesar pelo fator de Boltzmann). Mas isso não equivale a amostrar uniformemente $r$ e $\Phi$. O motivo disso é que se amostrarmos as coordenadas cartesianas de B para um A fixo, amostraríamos, necessariamente, mais valores distantes, isto é, para $r$ maiores. Por outro lado, se amostrarmos uniformemente em $r$, então, por definição, conseguimos o mesmo número de configurações para cada $r$. E essa é a origem do fator de correção Jacobiano para mudança de coordenadas.[68]

Exemplos de correções por Jacobianos para os ângulos de diedro podem sem encontrados no trabalho de Herschbach e colaboradores.[69]

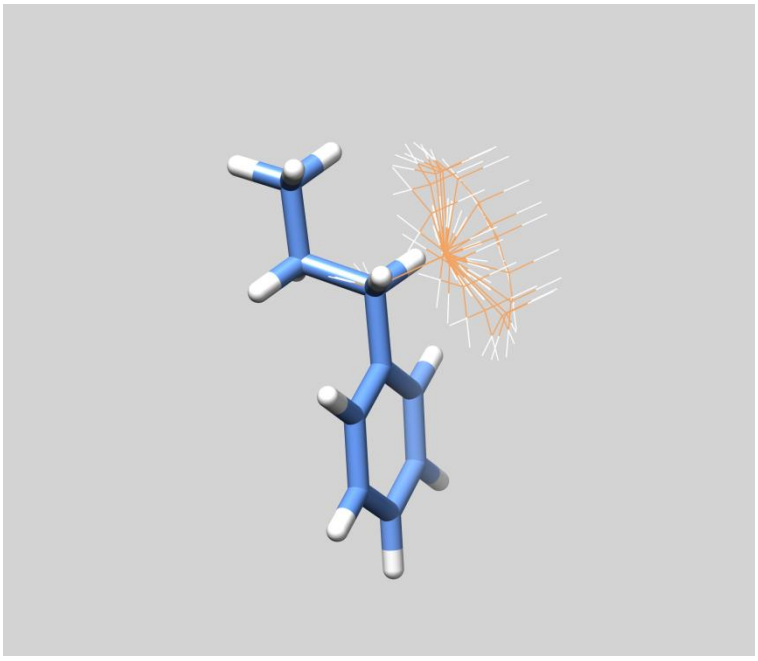


Figura 26 - Diedros amostrados para a parte alifática, durante uma simulação sem a utilização de um fato de correção jacobiano, para o ligante 4W55 (N-propilbenzeno).

Fonte: Elaborada pelo autor.

A Figura 26 apresenta os diedros gerados para o ligante 4W55, sem utilizar o fator de correção jacobiano para a mudança de coordenadas. Fica bastante claro, ao observar essa figura, que todo o espaço de diedros não é amostrando durante uma simulação utilizando *sample*_torsion.

Nas simulações realizadas neste projeto, diversos ligantes possuem ângulos de torção e, para estes, a amostragem em todo o espaço de configurações não estava acontecendo, isto é, ao gerarmos as configurações para os nossos ligantes com ligações torcionáveis, apenas uma parcela era amostrada de modo que uma correção no modelo deveria ser feita.

Como foi descrito na seção 3 (MATERIAIS E MÉTODOS), a amostragem utilizando o algoritmo *full_flex* tenta contornar esse problema, fazendo com que os átomos envolvidos nas ligações torcionáveis possam acessar qualquer lugar do espaço de configurações. Tal modelo de amostragem ainda está em desenvolvimento e ainda carece de aprimoramentos, porém, a seguir, apresentamos os primeiros resultados obtidos utilizando esse método.

Uma vez que para todos os resultados apresentados até então utilizando MC a utilização de Poisson como modelo eletrostático resultou em melhores correlações, aqui mostraremos apenas os resultados para esse modelo.

#### 4.2.2.1 Amostragem

Assim para *sample_torsion*, uma vez geradas as trajetórias por meio das simulações de MC *full_flex*, uma comparação com DM e *sample_torsion* pode ser feita. As imagens a seguir mostram essa comparação.

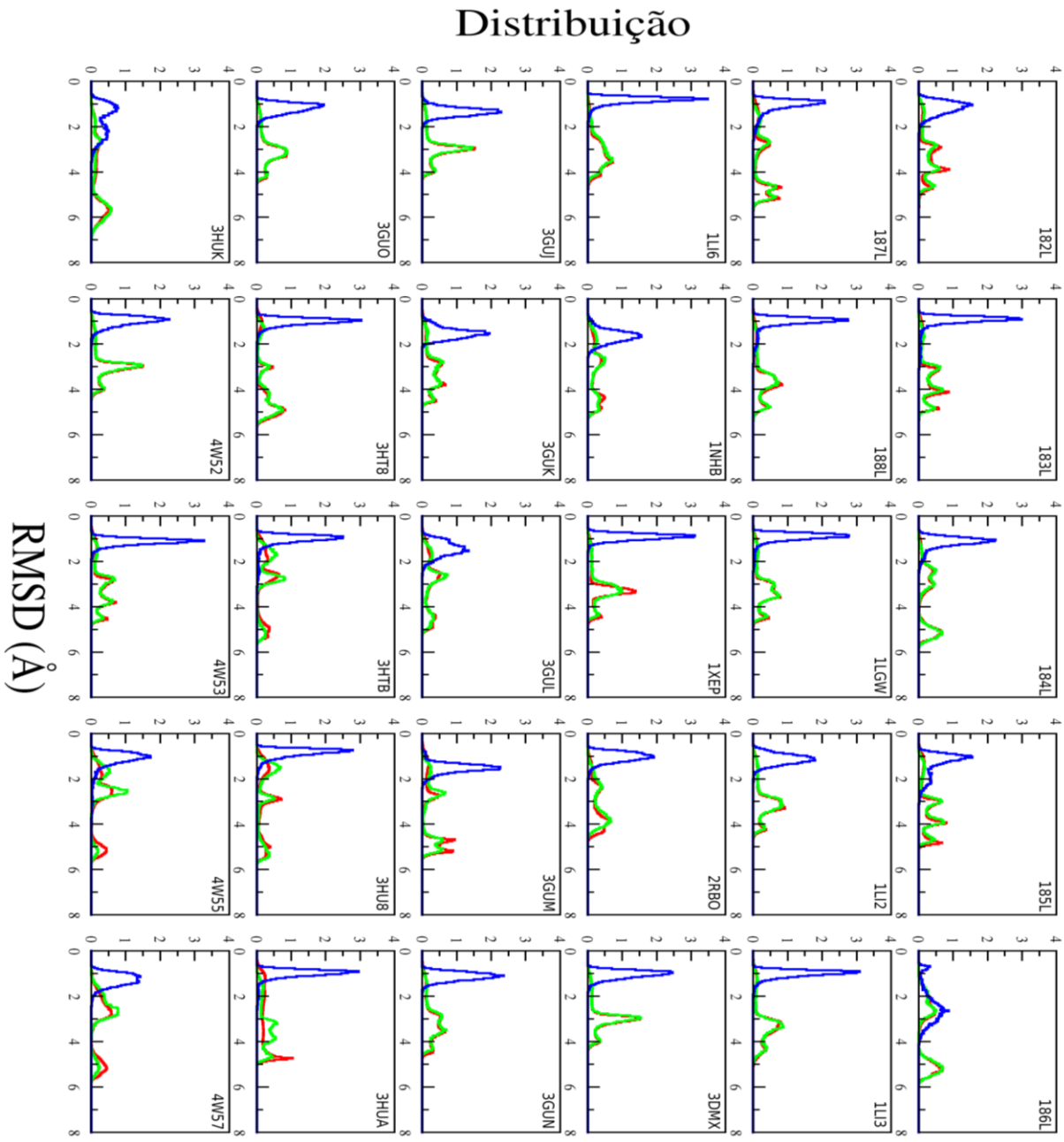



Figura 27 - Resultado das amostragens obtidas nas simulações. Em azul, vemos a amostragem, como uma distribuição de RMSD, para as simulações de DM. Em vermelho, está a amostragem, como uma distribuição de RMSD, para as simulações de MC, para o conjunto de parâmetros *sample_torsion*, Poisson, 250K. Em verde, está a amostragem, como um distribuição de RMSD, para as simulações de MC, utilizando *full_flex* e Poisson como modelo eletrostático.

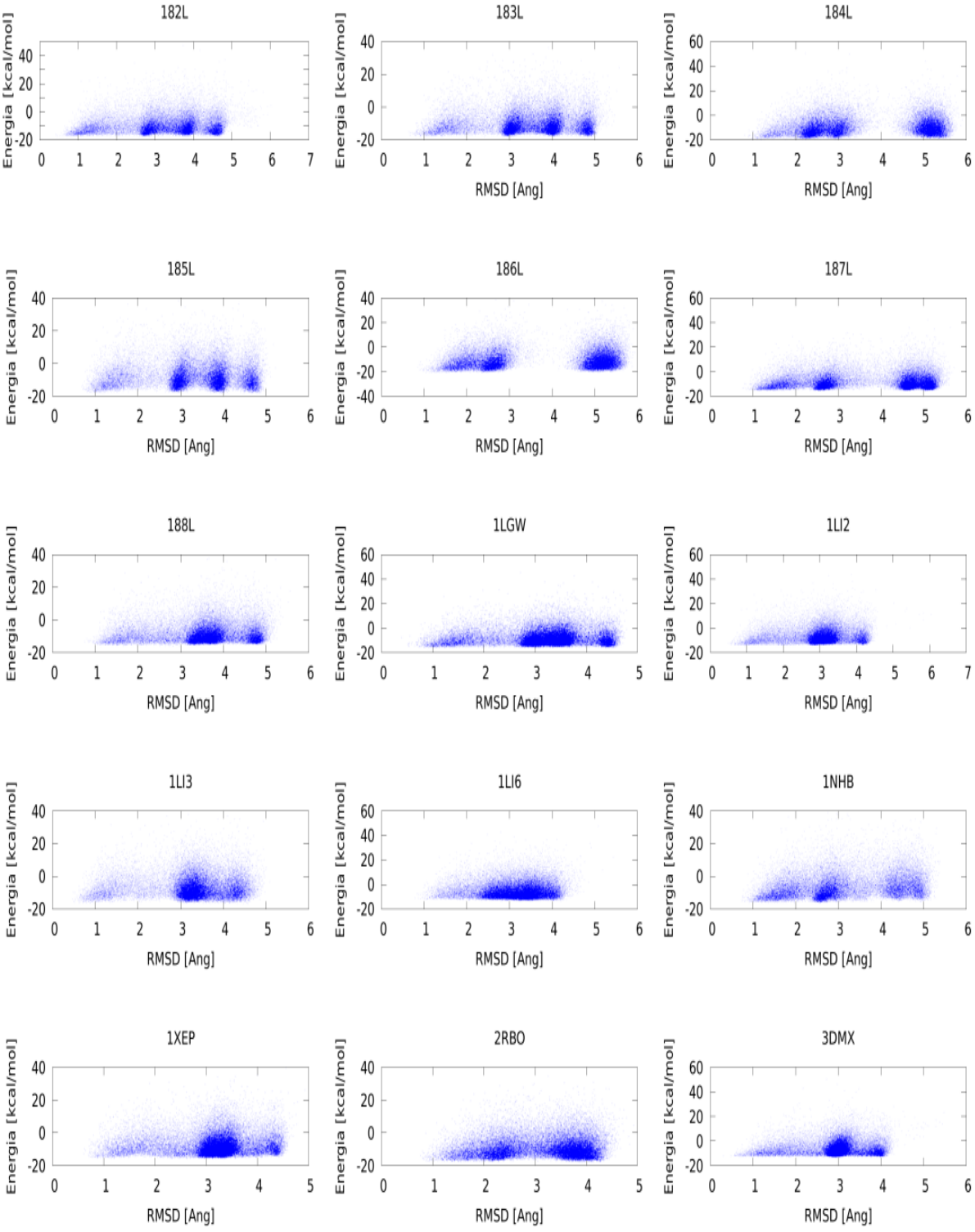
Fonte: Elaborada pelo autor.

Assim como descrito na seção 4.2.1.1, aqui, para amostragem do tipo *full_flex*, também existe uma nítida diferença com relação ao resultado para DM. Para esse novo modelo, os desvios com relação à estrutura cristalográfica também

foram maiores quando comparamos com a dinâmica molecular, indicando que as simulações visitaram uma quantidade maior de conformações. Entretanto, também é possível ver na Figura 27 que, além dos RMDS maiores, as simulações de MC *full_flex* amostraram as conformações amostradas pela DM.

Se comparamos *sample_torsion* com *full_flex* vemos que ambos amostraram praticamente as mesmas distribuições de RMSD ao longo das simulações.

Assim como no caso anterior, ao compararmos os dados da amostragem por MC *full_*flex e as conformações dos ligantes nas estruturas cristalográficas, vemos que as conformações amostradas se apresentaram dentro de um desvio de 6 Å com relação às conformações experimentais (RMSD) para todos os átomos de acordo com as energias amostradas para o sistema. A figura a seguir mostra o perfil de energia assumido pelos complexos durante as simulações e os estados de ocupação de RMSD.


182L
183L
184L
185L
186L
187L
188L
1LGW
1LI2
1LI3
1LI6
1NHB
1XEP
2RBO
3DMX
Energia [kcal/mol]
RMSD [Ang]

(continua)

(continuação)

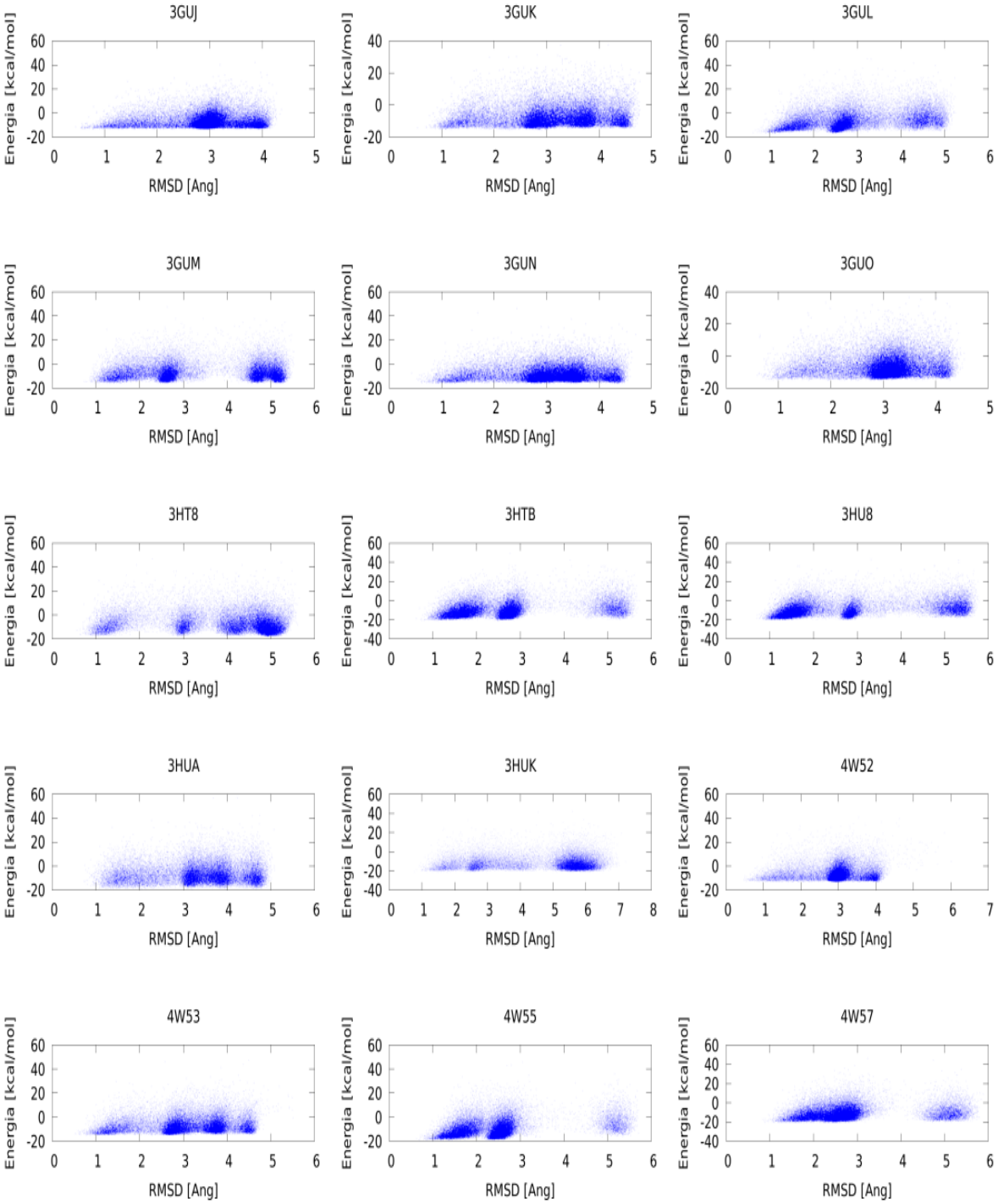



Figura 28 - Avaliação das energias livres de ligação médias para as simulações de MC para os 30 complexos como uma função do RMSD do ligante comparado com a posição cristalográfica. Aqui, o modelo eletrostático é Poisson, temperatura é 250K e amostragem é *full_flex*.

Fonte: Elaborada pelo autor.

Na Figura 28 é possível ver que os estados de energia acessados durante as simulações utilizando full_flex como modelo de amostragem e Poisson como modelo eletrostático se assemelham muito com os perfis apresentados por esse mesmo modelo eletrostático, porém, utilizando sample_torsion. A diferença se dá na distribuição de pontos apresentados nos gráficos, indicando que ao utilizarmos *full_flex*, os ligantes acessam estados de energia, variados com o RMSD, bastante distintos, sem apresentar as 'faixas contínuas' que vimos com *sample_torsion*, uma vez que agora as conformações são amostradas de maneira mais precisa.

Assim como para DM e para o outro modelo de amostragem por MC, também fizemos uma análise dos ângulos de diedro formados pelos átomos pesados dos ligantes que possuíam ligações torcionáveis. A seguir apresentamos o resultado obtido para um dos ligantes.

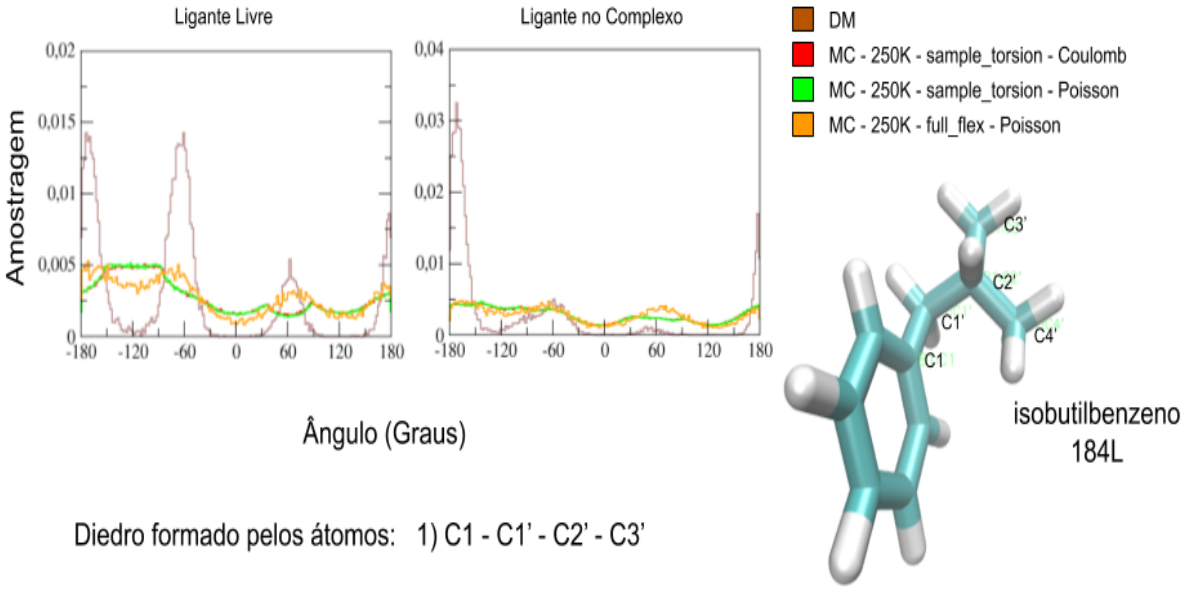



Figura 29 - Análise do perfil de ângulos de diedro para o ligante 184L (isobutilbenzeno). No gráfico, 1), indicado acima dele, representa o ângulo de diedro formado pelos átomos indicados na parte inferior da figura, como mostrado na molécula e significa que os gráficos naquela coluna apresentam resultados para esse ângulo. O número 2), indicado acima da coluna de gráficos à direita indica que os gráficos nessa coluna apresentam resultados para os ângulos de diedro correspondentes com a parte inferior da figura. Em marrom, está o perfil gerado por DM; em vermelho o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Coulomb como modelo eletrostático, e em verde o perfil gerado por MC a 250K, usando amostragem por *sample_torsion* e Poisson como modelo eletrostático; e em laranja está representado o perfil gerado por MC a 250K, usando *full_flex* como amostragem e Poisson como modelo eletrostático.

Fonte: Elaborada pelo autor.

A Figura 29 apresenta os gráficos normalizados com as distribuições de ângulos de diedro formado pelos átomos C1 – C1’ – C2’ – C3’ indicados na molécula, obtidas a partir das análises das trajetórias das simulações de DM e MC, com ambos os modelos de amostragem e modelos eletrostáticos, como indicado na legenda da figura. Aqui, também é possível ver que o perfil assumido pelas simulações de MC com *full_flex* é bastante semelhante ao apresentado pela DM, entretanto, as distribuições geradas por MC aparecem de maneira menos acentuadas do que as obtidas por meio da dinâmica.

Se comparamos os dois modelos de amostragem por MC entre si é possível ver que os perfis que eles apresentam são bastante semelhantes, com *full_flex* se aproximando mais do perfil apresentado por DM.

O perfil apresentado para todos os outros seis ligantes com ligações torcionáveis foi bastante simular ao apresentado pela Figura 29, isto é, as amostragens dos ângulos de diedro também se concentraram ao redor de ângulos específicos, com pouca variação de acordo com cada ligante e suas ligações torcionáveis, sendo sempre menos acentuadas que os resultados apresentados por DM.

Assim como para o sistema amostrado por *sample_torsion*, ao analisar toda a trajetória dos ligantes nos complexos é possível ver que eles acessaram configurações que varrem todo o espaço de configurações acessíveis, como pode ser visto na Figura 30.

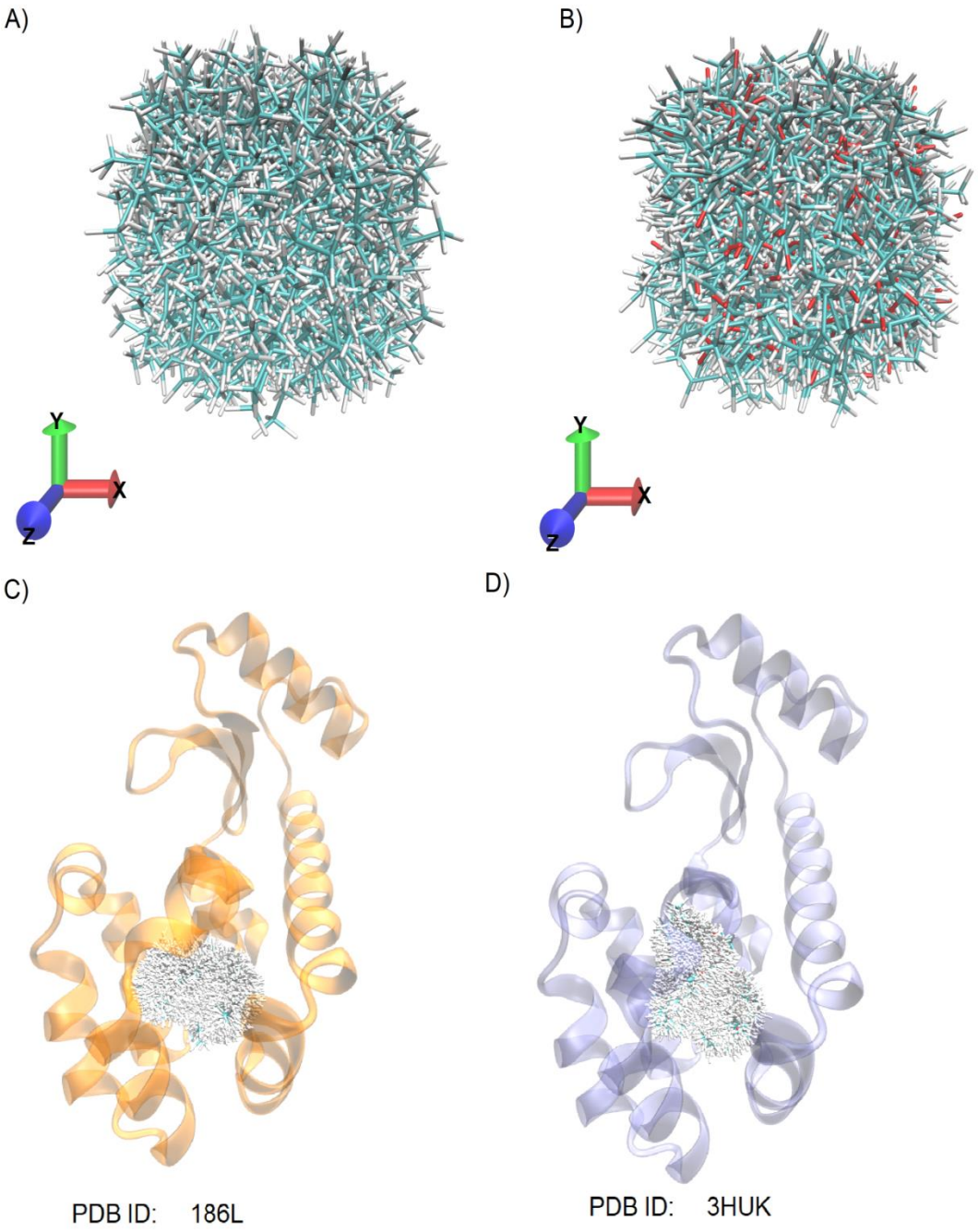



Figura 30 - Configurações amostradas dos ligantes livres e complexados durante as simulações de MC, a 250K, utilizando *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. A) apresenta o ligante 186L (n-butilbenzeno) livre, acessando todo o espaço de conformações na caixa de simulação. B) apresenta o ligante 3HUK (benzilacetato) livre, acessando todo o espaço de conformações na caixa de simulação. C) apresenta o ligante 186L, cujo receptor possui apenas a mutação L99A, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações. D) apresenta o ligante 3HUK, cujo receptor apresenta as mutações L99A e M102Q, no complexo, ocupando todo o espaço do sítio de ligação durante as simulações.

Fonte: Elaborada pelo autor.

Na Figura 30 é possível ver o comportamento de dois dos 30 ligantes no complexo durante as simulações, cujo receptor possui apenas a mutação L99A, 186L, e outro cujo receptor possui as mutações L99A e M102Q, 3HUK. Eles representam o comportamento típico de um ligante durante as simulações. Nessa imagem é possível ver que o espaço de conformações disponível para o ligante foi totalmente preenchido pelos movimentos de rotação, translação e torção ao longo das trajetórias, tanto para os ligantes livres, que preencheram toda a caixa de simulação, quanto para os ligantes complexados, que também preencheram todo o espaço disponível nos sítios de interação.

#### 4.2.2.2 Energia Livre de Ligação para MC

Assim como para *sample_torsion*, os cálculos da energia livre de ligação foram feitos a uma temperatura de 250K, para os 30 complexos ligante-receptor, sem usar amortecimento ou solvatação. As energias obtidas foram comparadas com os valores determinados experimentalmente disponíveis na literatura.[4,38-42] Como os resultados utilizando Poisson com o modelo eletrostático já se mostraram melhores, aqui não apresentaremos energias obtidas a partir de Coulomb.

Tabela 10 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, utilizando Poisson como modelo eletrostático e *full_flex* como modelo de amostragem.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $\Delta E_{MC\text{-}full_flex}^{Poisson}$ (kcal/mol) |
|---|---|---|
| **182L** | -5,46 | -18,3 ± 0,2 |
| **183L** | -5,13 | -19,2 ± 0,1 |
| **184L** | -6,51 | -21,0 ± 0,2 |
| **185L** | -4,89 | -19,56 ± 0,09 |
| **186L** | -6,70 | -22,3 ± 0,1 |
| **187L** | -4,67 | -17,5 ± 0,1 |
| **188L** | -4,60 | -17,6 ± 0,1 |
| **1NHB** | -5,76 | -18,2 ± 0,2 |
| **3DMX** | -5,35 | -15,27 ± 0,08 |
| **4W52** | -5,19 | -15,5 ± 0,2 |
| **4W53** | -5,52 | -17,3 ± 0,1 |
| **4W55** | -6,55 | -20,6 ± 0,4 |
| **4W57** | -6,69 | -21,9 ± 0,4 |
| **1LGW** | -5,53 | -18,0 ± 0,1 |
| **1LI2** | -5,58 | -15,89 ± 0,09 |
| **1LI3** | -5,87 | -18,1 ± 0,2 |
| **1LI6** | -5,25 | -16,29 ± 0,05 |
| **1XEP** | -4,16 | -18,4 ± 0,1 |
| **2RBO** | -5,52 | -17,84 ± 0,09 |
| **3HT8** | -5,04 | -19,39 ± 0,08 |
| **3HTB** | -5,33 | -22,55 ± 0,07 |
| **3HU8** | -4,02 | -21,7 ± 0,1 |
| **3HUA** | -4,61 | -22,54 ± 0,09 |
| **3HUK** | -4,48 | -22,2 ± 0,1 |
| **3GUJ** | -5,80 | -15,5 ± 0,1 |
| **3GUK** | -5,30 | -17,22 ± 0,06 |
| **3GUL** | -5,50 | -17,77 ± 0,08 |
| **3GUM** | -4,90 | -17,3 ± 0,1 |
| **3GUN** | -4,00 | -17,4 ± 0,2 |
| **3GUO** | -5,00 | -16,58 ± 0,09 |
| **Correlação de Pearson** | | **0,12** |
| **Correlação por média de Bootstrap** | | **0,20 ± 0,02** |

Fonte: Elaborada pelo autor.

Para esse caso, como para os anteriores, se olharmos para as correlações dos 30 complexos de uma vez vemos valores bastante baixos. Porém, assim como antes, se retiramos três dos ligantes, notamos uma melhora na correlação com as energias experimentais. As imagens a seguir ilustram esses resultados.

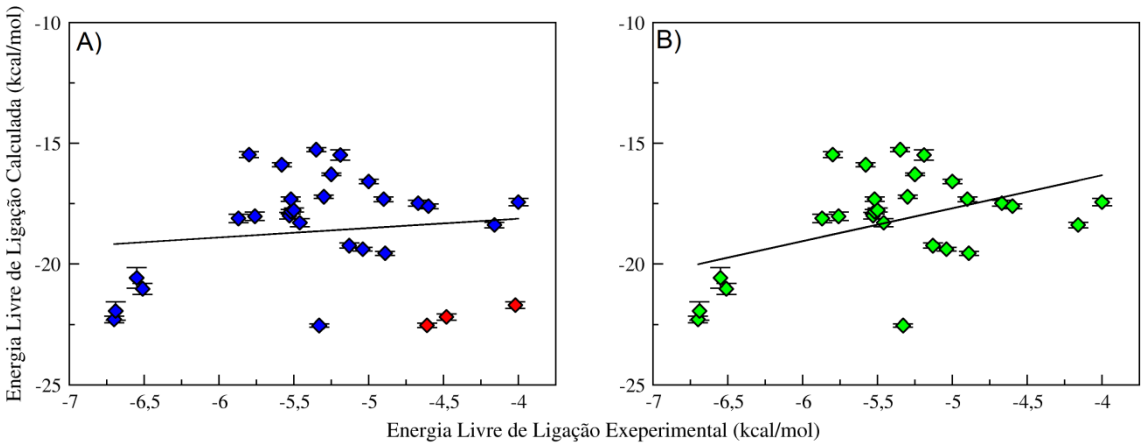



Figura 31 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Na Figura 31 e na Tabela 10 vemos que a correlação entre os valores de energia livre de ligação experimentais e calculados para os 30 complexos foi bastante baixa, sendo a correlação de Pearson 0,12 e por *bootstrapping* 0,20 ± 0,02, com desvio médio absoluto de 1,87 kcal/mol (N=30). Se retiramos os três ligantes destacados na imagem, a correlação de Pearson vai para 0,46, ao passo que por meio de *bootstrapping* a correlação vai para 0,51 ± 0,02, com desvio de 1,55 kcal/mol (N=27).

Fazendo a correção para o solvente, apresentada na equação (4.2), esperava-se que os valores de energia livre de ligação calculados mostrassem uma melhora com relação a não usar tal abordagem. Esses resultados podem ser vistos na Tabela 11.

Tabela 11 - Comparação entre os dados experimentais e os dados calculados para os 30 complexos analisados utilizando as simulações de Monte Carlo, à 250K, utilizando Poisson como modelo eletrostático, full_flex como modelo de amostragem, utilizando os como parâmetros de solvatação α = 0,2 kcal/(mol.$e^{-2}$) e β = -0,005 kcal/mol.

| PDB ID | Energia Livre de Ligação Experimental (kcal/mol) | $\Delta E_{MC\text{-}full_flex}^{Poisson}$ (kcal/mol) |
|---|---|---|
| **182L** | -5,46 | -14,11 ± 0,09 |
| **183L** | -5,13 | -15,5 ± 0,1 |
| **184L** | -6,51 | -17,2 ± 0,2 |
| **185L** | -4,89 | -14,3 ± 0,1 |
| **186L** | -6,70 | -17,82 ± 0,09 |
| **187L** | -4,67 | -14,2 ± 0,1 |
| **188L** | -4,60 | -14,5 ± 0,1 |
| **1NHB** | -5,76 | -14,6 ± 0,1 |
| **3DMX** | -5,35 | -12,1 ± 0,1 |
| **4W52** | -5,19 | -12,0 ± 0,1 |
| **4W53** | -5,52 | -13,71 ± 0,07 |
| **4W55** | -6,55 | -16,2 ± 0,3 |
| **4W57** | -6,69 | -17,3 ± 0,2 |
| **1LGW** | -5,53 | -9,4 ± 0,1 |
| **1LI2** | -5,58 | -10,07 ± 0,09 |
| **1LI3** | -5,87 | -11,75 ± 0,09 |
| **1LI6** | -5,25 | -11,62 ± 0,07 |
| **1XEP** | -4,16 | -9,60 ± 0,09 |
| **2RBO** | -5,52 | -12,3 ± 0,1 |
| **3HT8** | -5,04 | -13,0 ± 0,2 |
| **3HTB** | -5,33 | -15,2 ± 0,2 |
| **3HU8** | -4,02 | -13,7 ± 0,2 |
| **3HUA** | -4,61 | -15,45 ± 0,07 |
| **3HUK** | -4,48 | -11,9 ± 0,1 |
| **3GUJ** | -5,80 | -11,28 ± 0,09 |
| **3GUK** | -5,30 | -12,9 ± 0,1 |
| **3GUL** | -5,50 | -13,42 ± 0,09 |
| **3GUM** | -4,90 | -13,3 ± 0,1 |
| **3GUN** | -4,00 | -7,9 ± 0,1 |
| **3GUO** | -5,00 | -9,88 ± 0,09 |
| **Correlação de Pearson** | | **0,51** |
| **Correlação por média de Bootstrap** | | **0,56 ± 0,02** |

Fonte: Elaborada pelo autor.

Nesse caso, ao considerarmos a solvatação, a correlação dos 30 ligantes calculados com os valores experimentais é consideravelmente maior do que os resultados obtidos até agora. Novamente, retirando os três mesmos ligantes de antes a correlação com os dados experimentais aumenta, como pode ser visto na figura a seguir.

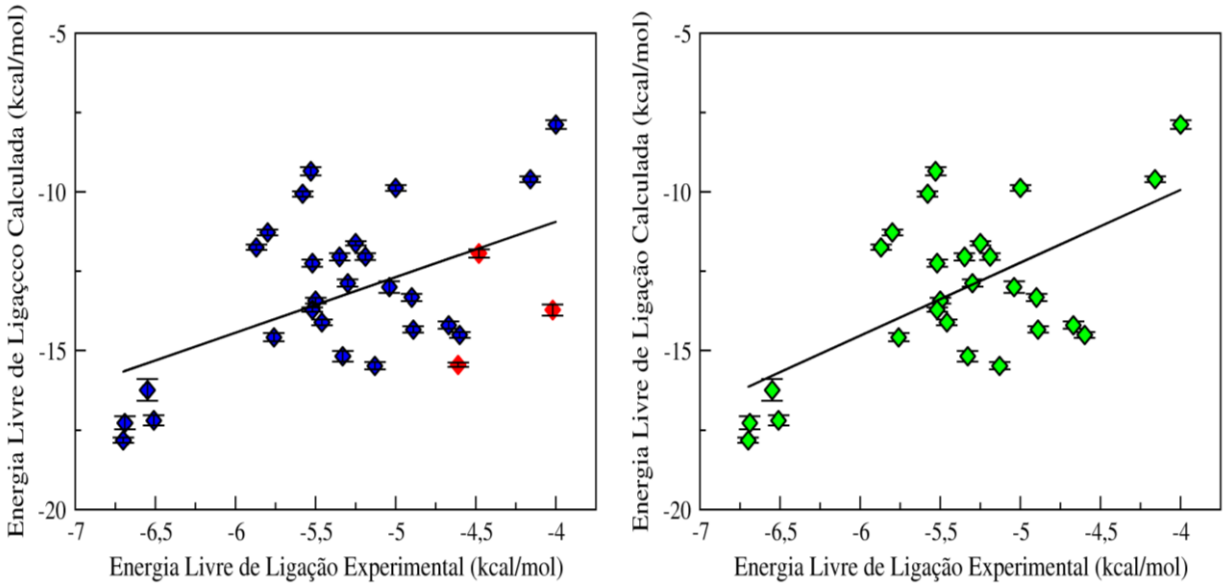



Figura 32 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem, considerando a dessolvatação, com α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol como parâmetros. (A) contém todos os 30 complexos simulados, com os dois ligantes mais discrepantes destacados em vermelho (3HU8, 3HUA e 3HUK). (B) Contém os 27 ligantes restantes.

Fonte: Elaborada pelo autor.

Na Tabela 11 e na Figura 32 vemos que a correlação entre os valores de experimentais e calculados para os 30 complexos foi bastante elevada ao compararmos com os outros resultados, sendo a correlação de Pearson 0,51 e por *bootstrapping,* 0,56 ± 0,2, com desvio médio absoluto de 1,87 kcal/mol (N=30). Retirando os três ligantes de sempre, temos que a correlação aumenta, indo para 0,62, por Pearson, ou 0,62, e por *bootstrapping*, 0,65 ± 0,2, com desvio médio absoluto de 2,02 kcal/mol (N=27).

Assim como para os casos anteriores, utilizar esse modelo de amostragem não resultou em acerto dos valores absolutos das energias, apenas o perfil de tendência a se correlacionar com os dados experimentais.

# 5 DISCUSSÕES

Ao analisarmos todos os resultados apresentados fica claro que para todas as simulações de MC que realizamos os valores calculados das energias livres de ligação não foram os mesmos quando comparados com os dados experimentais. Entretanto, os gráficos mostram uma tendência característica do perfil termodinâmico, o que significa que, em termos de valores absolutos, nosso método ainda não é capaz de acertar, porém, o perfil das energias pode ser observado a partir dos dados gerados.

## 5.1 Comparação com outros autores

Uma vez que nosso método se mostrou minimamente funcional, isto é, ele foi capaz de predizer o perfil termodinâmico para um conjunto de 27 ligantes com uma correlação boa, isto é, r = 0,65 ± 0,02, é natural que façamos uma comparação com outros autores que abordam problemas semelhantes. A seguir, é possível ver tal comparação.

Para representar as energias nas tabelas foram utilizadas siglas, de modo que o sub índice SF3 significa "*Scoring Function 3"*, que representa as energias calculadas sem considerar a solvatação nem amortecimento, baseado na equação (3.1); *SF2* significa *"Scoring Function 2"*, que representa as energias calculadas considerando a solvatação mas não o amortecimento – baseado na equação (4.2) – e *SF1* significa *"Scoring Function 1"*, que representa as energias calculadas utilizando o potencial amortecido, mas não a solvatação, baseado na equação (4.1). *C* representa a utilização de Coulomb como modelo eletrostático e *P*, Poisson.

Assim, nas Tabela 12 a Tabela **13**Tabela **14** $\Delta G_{exp}$ representa a energia livre de ligação experimental, $\Delta G_{SF3\text{-}C}$ representa os valores calculados pelo nosso método, com Coulomb como modelo eletrostático, sem utilizar amortecimento ou solvatação. $\Delta G_{SF3\text{-}P}$ representa o mesmo que o anterior, porém, com Poisson como modelo eletrostático. $\Delta G_{SF2\text{-}C}$ representa as energias calculadas utilizando Coulomb como modelo eletrostático e considerando a solvatação, sem usar amortecimento. $\Delta G_{SF2\text{-}P}$

representa o mesmo que o anterior, porém, com Poisson como modelo eletrostático. $\Delta G_{SF1\text{-}C}$ representa as energias calculadas pelo nosso método utilizando Coulomb como modelo eletrostático, usando o potencial amortecido, mas sem usar solvatação. $\Delta G_{SF1\text{-}P}$ representa o mesmo que o anterior, mas com Poisson como modelo eletrostático.

Tabela 12 - Comparação entre os valores de energia livre de ligação calculados por Steinbrecher e colaboradores[32] com os valores calculados pelo nosso programa.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta G_{Steinb}$ (kcal/mol) | $\Delta G_{SF3\text{-}C}$ (kcal/mol) | $\Delta G_{SF3\text{-}P}$ (kcal/mol) | $\Delta G_{SF2\text{-}C}$ (kcal/mol) | $\Delta G_{SF2\text{-}P}$ (kcal/mol) | $\Delta G_{SF1\text{-}C}$ (kcal/mol) | $\Delta G_{SF1\text{-}P}$ (kcal/mol) |
|---|---|---|---|---|---|---|---|---|
| **184L** | -6,51 | -6,86 | -7,66 | -8,15 | -5,47 | -5,82 | -7,70 | -8,17 |
| **186L** | -6,70 | -7,18 | -9,87 | -10,18 | -7,50 | -7,79 | -9,96 | -10,22 |
| **187L** | -4,67 | -4,20 | -2,20 | -0,23 | -0,21 | -0,32 | -2,13 | -2,25 |
| **188L** | -4,60 | -5,13 | -3,13 | -3,50 | -1,20 | -1,59 | -3,09 | -3,44 |
| **1NHB** | -5,76 | -5,50 | -0,82 | -0,82 | 1,34 | 1,33 | -0,81 | -0,78 |
| **3DMX** | -5,35 | -4,28 | -1,13 | -1,12 | 0,95 | 0,97 | -1,09 | -1,09 |
| **4W53** | -5,52 | -4,93 | -1,06 | -0,87 | 1,08 | 1,25 | -0,95 | -0,84 |
| **4W55** | -6,55 | -6,57 | -7,48 | -6,79 | -4,18 | -4,38 | -6,95 | -6,65 |
| **Correlação** | | 0,89 | 0,76 | 0,79 | 0,73 | 0,71 | 0,73 | 0,73 |

Fonte: Elaborada pelo autor.

Na Tabela 12, comparamos os valores de energia livre de ligação obtidos a partir do nosso método com os valores obtidos por Thomas Steinbrecher e colaboradores,[32] indicada por $\Delta G_{Steinb}$. Nesse trabalho eles validam um protocolo de FEP, o FEP+, baseado em simulações de dinâmica molecular e que considera flexibilidade conformacional e os efeitos da entropia para calcular as afinidades de ligação de diversos compostos.

Ao compararmos os resultados que nós obtivemos com os resultados reportados por eles é possível ver que nossa maior correlação, para esse conjunto de dados, foi de 0,79, utilizando Poisson como modelo eletrostático, sem utilizar amortecimento ou solvatação, enquanto eles obtiveram 0,89 de correlação para esses dados. Essa diferença pode ser justificada por Steinbrecher e seus colaboradores terem usado um modelo mais robusto, porém, com custo computacional bastante elevado. Nesse, Steinbrecher menciona que todos os cálculos reportados por eles foram obtidos depois de 24h de simulações em 30 GPUs.

Se levarmos em conta que nossos sistemas foram simulados em CPUs e que o tempo médio para a obtenção dos valores de energia livre de ligação para 10 réplicas de 40 milhões de passos cada, para os 30 complexos, foi de 4h então podemos afirmar que existe competitividade do MCLiBELa frente aos métodos

desenvolvidos por outros autores, uma vez que os resultados obtidos foram bastante satisfatórios e com poder de processamento menor.

A Tabela 13 apresenta os nossos resultados comparados com os resultados obtidos por David Minh e seus colaboradores.[36] Em seu trabalho, eles utilizam modificações em MM-GBSA e MM-PBSA para realizar o cálculo das energias livres de ligação, além de adicionarem um estimador para a entropia a partir de uma expansão da função para o cálculo das energias livres e utilizar os termos de ordens mais altas dessa expansão como os componentes da energia livre, como a entropia, por exemplo.

Tabela 13 - Comparação entre os valores de energia livre de ligação calculados por Minh e colaboradores[36] com os valores calculados pelo nosso programa.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta G_{Minh}$ (kcal/mol) | $\Delta G_{SF3\text{-}C}$ (kcal/mol) | $\Delta G_{SF3\text{-}P}$ (kcal/mol) | $\Delta G_{SF2\text{-}C}$ (kcal/mol) | $\Delta G_{SF2\text{-}P}$ (kcal/mol) | $\Delta G_{SF1\text{-}C}$ (kcal/mol) | $\Delta G_{SF1\text{-}P}$ (kcal/mol) |
|---|---|---|---|---|---|---|---|---|
| **182L** | -5,46 | -7,04 | -3,58 | -3,62 | -0,67 | -0,68 | -3,53 | -3,52 |
| **183L** | -5,13 | -6,86 | -4,33 | -4,65 | -2,02 | -2,34 | -4,28 | -4,59 |
| **184L** | -6,51 | -10,58 | -7,66 | -8,15 | -5,47 | -5,82 | -7,70 | -8,17 |
| **185L** | -4,89 | -7,84 | -4,87 | -4,63 | -1,65 | -1,45 | -4,82 | -4,60 |
| **186L** | -6,70 | -13,17 | -9,87 | -10,18 | -7,50 | -7,79 | -9,96 | -10,22 |
| **187L** | -4,67 | -5,65 | -2,20 | -0,23 | -0,21 | -0,32 | -2,13 | -2,25 |
| **188L** | -4,60 | -9,2 | -3,13 | -3,50 | -1,20 | -1,59 | -3,09 | -3,44 |
| **1NHB** | -5,76 | -7,11 | -0,82 | -0,82 | 1,34 | 1,33 | -0,81 | -0,78 |
| **Correlação** | | 0,73 | 0,71 | 0,73 | 0,72 | 0,71 | 0,71 | 0,71 |

Fonte: Elaborada pelo autor.

Na Tabela 13, $\Delta G_{Minh}$ representa a energia livre de ligação calculada por David Minh utilizando o truncamento de primeira ordem que eles mencionam em seu trabalho para computar a entropia.

Ao compararmos nossas energias livres de ligação com os obtidos por Minh e seus colaboradores vemos que a correlação com os dados experimentais que eles obtiveram foi bastante razoável, isto é, r = 0,73, um valor que também obtivemos para o mesmo conjunto de dados utilizado a função *SF3* com Poisson. Isso pode ser justificado pelo fato de ele utilizar as funções padrão do *AmberTools*, ou seja, utiliza um potencial do tipo AMBER clássico, como nós utilizamos para a função SF3. Como ele também utiliza um meio dielétrico constante, o nosso resultado que melhor se assemelha ao dele ser o que utiliza Poisson também era esperado.

Embora eles tenham apresentado bons resultados, seus dados derivam de 200 ns de simulações de DM, seguidos por cálculos de MM-GBSA, que, se realizados em GPUs, costumam durar mais de 24h utilizando o *AmberTools*, o que é consideravelmente custoso. Além disso, apesar da boa correlação, tanto eles como nós não acertamos os valores absolutos das energias livres de ligação, apenas o perfil característico.

A Tabela 14 compara nossos resultados com os obtidos por Enrico Purisima e colaboradores.[70] Em seu trabalho foram feitos testes de *docking* exaustivos como uma alternativa para MC e DM a fim de calcular grandezas termodinâmicas de um sistema proteína-ligante, como energia livre de ligação, entalpia e entropia. Para realizar esse cálculo eles incluíram termos de solvatação baseado em resolver a equação de Poisson, usando dois modelos eletrostáticos diferentes.

Tabela 14 - Comparação entre os valores de energia livre de ligação calculados por Purisima e colaboradores[70] com os valores calculados pelo nosso programa.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta G_{Pur1}$ (kcal/mol) | $\Delta G_{Pur1}$ (kcal/mol) | $\Delta G_{SF3\text{-}C}$ (kcal/mol) | $\Delta G_{SF3\text{-}P}$ (kcal/mol) | $\Delta G_{SF2\text{-}C}$ (kcal/mol) | $\Delta G_{SF2\text{-}P}$ (kcal/mol) | $\Delta G_{SF1\text{-}C}$ (kcal/mol) | $\Delta G_{SF1\text{-}P}$ (kcal/mol) |
|---|---|---|---|---|---|---|---|---|---|
| **182L** | -5,46 | -6,99 | -8,72 | -3,58 | -3,62 | -0,67 | -0,68 | -3,53 | -3,52 |
| **183L** | -5,13 | -5,14 | -6,47 | -4,33 | -4,65 | -2,02 | -2,34 | -4,28 | -4,59 |
| **184L** | -6,51 | -7,35 | -8,31 | -7,66 | -8,15 | -5,47 | -5,82 | -7,70 | -8,17 |
| **185L** | -4,89 | -3,47 | -6,89 | -4,87 | -4,63 | -1,65 | -1,45 | -4,82 | -4,60 |
| **186L** | -6,70 | -7,50 | -8,40 | -9,87 | -10,18 | -7,50 | -7,79 | -9,96 | -10,22 |
| **187L** | -4,67 | -3,63 | -4,63 | -2,20 | -0,23 | -0,21 | -0,32 | -2,13 | -2,25 |
| **188L** | -4,60 | -3,38 | -4,41 | -3,13 | -3,50 | -1,20 | -1,59 | -3,09 | -3,44 |
| **3DMX** | -5,35 | -5,32 | -6,45 | -1,13 | -1,12 | 0,95 | 0,97 | -1,28 | -1,27 |
| **Correlação** | | 0,91 | 0,81 | 0,82 | 0,83 | 0,92 | 0,91 | 0,83 | 0,83 |

Fonte: Elaborada pelo autor.

Na Tabela 14, $\Delta G_{Pur1}$ e $\Delta G_{Pur2}$ representa as energias livres de ligação obtidas a partir dos modelos 1 e 2 descrito por Purisima[70] em seu trabalho, para o conjunto de ligantes de lisozima de vírus T4.

Como pode ser visto, ambos os modelos usados por eles apresentaram boa correlação com os dados experimentais, sendo possível chegar a valores calculados bastante próximos dos valores experimentais. Porém, se compararmos com nossos resultados é possível ver que mesmo que nós não tenhamos obtido os valores absolutos de energia, apenas os perfis, as correlações foram bastante altas (r = 92, para o nosso melhor caso).

Uma comparação mais justa seria utilizando nosso modelo que leva em consideração a solvatação, o *SF2*. Ao olharmos pra ele vemos que independente do modelo eletrostático utilizado, nossa correlação foi igual ou superior à apresentada por Purisima, indicando a competitividade entre os métodos.

Apesar do bom desempenho do método que utiliza *docking* exaustivo, ele ainda apresenta limitações, como não lidar muito bem com ligantes muito flexíveis, um problema que tentamos resolver por meio da amostragem *full_flex.*

É válido dizer que todos esses autores trabalharam com um conjunto de ligantes para lisozima de T4 menor do que o nosso, e todos eles utilizaram ligantes cujos receptores possuíam apenas a mutação L99A, a qual gera a cavidade apolar no interior da enzima, enquanto que nosso conjunto teste possui, além dessa mutação, região polar no sítio de ligação (M102Q) ou sítio carregado (M102E).

Ao ampliarmos o conjunto de moléculas analisadas, adicionando dados de ciclodextrina, disponíveis no trabalho de David Mobley e seus colaboradores,[4] e utilizarmos nosso melhor conjunto de parâmetros para realizar as simulações, isto é, *full*_flex como modelo de amostragem, Poisson como modelo eletrostático e considerarmos a solvatação, chegamos ao que pode ser visto na Figura 33.

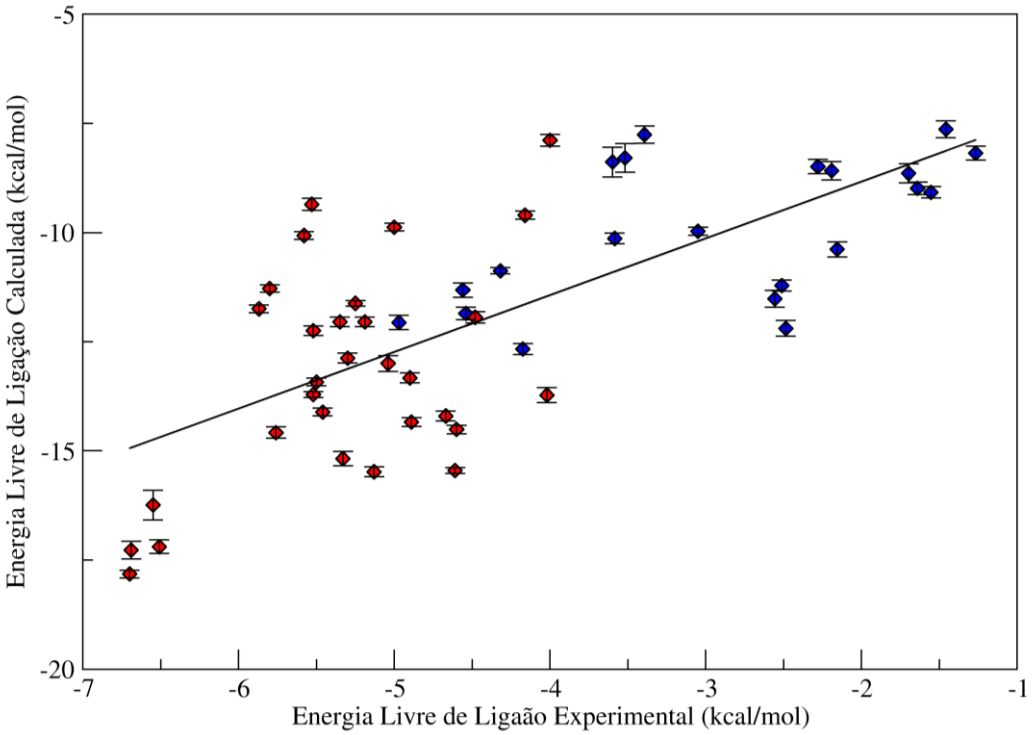



Figura 33 - Comparação entre os dados experimentais e os valores obtidos para as simulações de MC, à 250K, utilizando *full_flex* como modelo de amostragem, considerando a dessolvatação, com α = 0,2 kcal/(mol.e$^{-2}$) e β = -0,005 kcal/mol como parâmetros. Em azul estão representados os dados obtidos para ciclodextrina e em vermelho, os dados de lisozima de vírus T4.

Fonte: Elaborada pelo autor.

Comparando os dados de energia livre calculados com os dados experimentais, para todo o conjunto, isto é, analisando 51 ligantes, obtemos uma correlação de Pearson r = 0,72, ou, por *bootstrapping*, 0,75 ± 0,01. Ainda aqui, não acertamos os valores absolutos das energias, porém, a tendência de correlação com os dados experimentais mostrou-se bastante satisfatória.

## 5.2 Amostragem

As configurações amostradas para as moléculas nas seções sobre amostragem (4.2.1.1 e 4.2.2.1) apresentaram resultados condizentes entre si, isto é, independente do tipo de amostragem utilizado ou do modelo eletrostático escolhido, os perfis de distribuições dos desvios com relação à configuração cristalográfica

variando com a energia apresentou perfis semelhantes. Utilizar o modelo de amostragem *full_flex* resultou numa distribuição de pontos ao longo dos gráficos de energia livre versus RMSD menos uniforme, de modo que aglomerados de pontos surgissem de maneira mais distinta nos perfis apresentados. Esses perfis se deram dessa maneira pelo fato de termos menos pontos nos gráficos, isto é, as simulações feitas para esse modelo foram mais curtas. Entretanto, utilizar *full_flex* como modelo corresponde a uma amostragem mais correta, em termos de algoritmo de Metropolis, corrigindo a necessidade do termo jacobiano ao usar um sistemas de coordenadas diferente do cartesiano (ângulos de Euler, no caso).

Os ângulos de diedro amostrados para os ligantes que possuíam ângulos de torção foram bastante semelhantes entre as simulações de Monte Carlo, independentemente do conjunto de parâmetros utilizados para gerá-los, isto é, o modelo eletrostático e o tipo de amostragem, pouco influenciavam nos ângulos de diedro. Além disso, as simulações de MC também apresentaram distribuições mais uniformes quando comparadas com a dinâmica molecular, apesar de manter o perfil dos diedros amostrados apresentadas por ela. Essa distribuição mais homogênea era esperada, uma vez que as simulações de MC se baseiam em processos estocásticos enquanto que as simulações de DM são processos determinísticos.

Embora o *full_flex* seja um modelo mais rigoroso para amostrar as configurações das simulações de MC, ele ainda precisa passar por melhorias. Isso porque os cálculos levam mais tempo para convergir e descorrelacionar que o modelo *sample_torsion*, estimando energias ligeiramente piores do que esse último. Numa comparação justa, a correlação com os dados experimentais entre os 27 ligantes na configuração SF3 Poisson *sample_torsion* foi 0,58, enquanto o mesmo conjunto de parâmetros, com *full_flex*, resultou numa correlação de Pearson de 0,51. Além disso, assim como *sample_torsion*, o modelo mais rigoroso também superestima os valores das energias calculadas.

De modo geral, quando comparamos ambos os métodos podemos percebemos que utilizar *sample_torsion*, dentro das devidas proporções, retorna resultados bastante significativos (r = 57, para 27 ligantes, considerando a solvatação), com boas correlações com dados experimentais e num tempo bastante razoável (cerca de 4 horas). Em contrapartida, não explorar todo o espaço de

conformações para os ângulos de torção das moléculas que formam diedros implica em aceitar que o algoritmo não realiza a amostragem da forma mais rigorosa possível, perdendo estados de energia e conformações que deveriam ser amostrados.

Se comparamos os resultados obtidos a partir do nosso método com os resultados obtidos por dos cálculos que partiram de simulações de dinâmica molecular é possível ver que nenhum dos métodos utilizados para obter as energias por DM apresentou resultados melhores do que os advindos das simulações de MC, utilizando Poisson como modelo eletrostático, *sample_torsion* como modelo de amostragem e levando em conta a solvatação, ao se retirar os 3 ligantes problemáticos (3HU8, 3HUA e 3HUK). Além disso, utilizar o método apresentado aqui ainda apresenta a vantagem do custo computacional e, consequentemente, o tempo gasto nas simulações, ser menor, principalmente porque não é necessário utilizar GPUs para obter resultados rápidos.

Etapas futuras de desenvolvimento do método devem permitir que o receptor seja flexível durante as simulações para tornar o sistema mais real, além de aprimorar a amostragem do tipo *full_flex*. Além disso, considerar diferentes tipos de campo de força que não tenhamos testado pode ser também uma alternativa. Minh afirma que erros provenientes dos campos de força aumentam porque os campos de força de mecânica molecular apenas se aproximam da mecânica quântica, contabilizando efeitos locais do ambiente, como polarização e rearranjo de ligações, tais como protonação e tautomerização de maneira inadequada.[36]

## 5.3 Efeito do Modelo Eletrostático

Na seção **4.2.1.4** apresentamos os dados gerados para as energias livres de ligação, utilizando as diferentes formas de calcular as energias, isto é, utilizar ou não a solvatação bem como a utilização de um potencial de van der Waals amortecido. Nela foi possível ver que os valores calculados utilizando Poisson/DelPhi como modelo eletrostático implicaram em melhores correlações, independente da forma como a energia foi calculada. Esse resultado era esperado, uma vez que o modelo

eletrostático empregado é mais robusto e representa melhor o comportamento do meio. A imagem a seguir representa o comportamento padrão que a utilização desse modelo resultou.

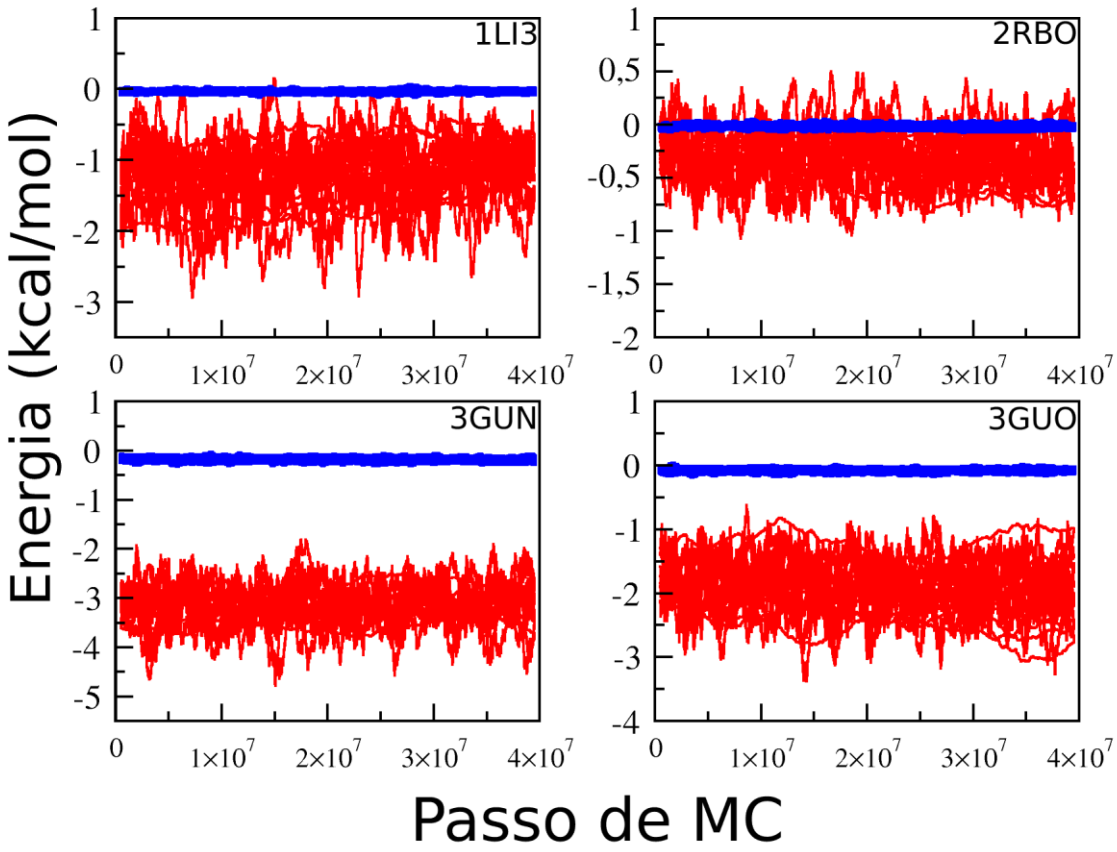



Figura 34 - Comparação entre as médias das contribuições das energias eletrostáticas utilizando os diferentes modelos. Azul, representa a média para a componente eletrostática utilizando Poisson como modelo para os quatro ligantes indicados e vermelho, representa a média para a componente eletrostática utilizando Coulomb como modelo. Aqui, a temperatura foi de 250K, a solvatação não foi considerada e não foi utilizado um potencial amortecido.

Fonte: Elaborada pelo autor.

A Figura 34 apresenta as médias da componente eletrostática para os ligantes 1LI3 e 2RBO, cujos receptores possuem as mutações L99A e M102Q, e para os ligantes 3GUN e 3GUO, cujos receptores possuem as mutações L99A e M102E. Ao utilizar Coulomb, os valores médios da componente eletrostática das energias são mais negativos, indicando que o modelo superestima tal componente. Entretanto, se olharmos para as médias obtidas utilizando Poisson, vemos que não

ocorre essa superestimava, resultando numa contribuição eletrostática menor, como era de se esperar para tal modelo.[14]

Como pode ser visto em todos os resultados apresentados na seção anterior, ao calcularmos as energias livres de ligação pelos dois modelos eletrostáticos, Poisson sobressaiu como o melhor modelo a ser utilizado, apresentando as melhores correlações com os dados experimentais. Esses resultados ficam ainda melhores quando adicionamos os termos de solvatação. Isso se deve ao fato de que tanto Poisson como considerar a solvatação são modelos mais robustos e que se aproximam mais de modelos reais para realização dos cálculos das energias.

Outro reflexo da média da contribuição eletrostática utilizando o modelo PB ser menor pode ser observado nos gráficos de energia variando com o RMSD. Neles, ao utilizarmos Coulomb permitimos que os ligantes acessassem estados de energia mais favoráveis, isto é, mais negativos quando comparado com aqueles apresentados por PB.

## 5.4 Entalpia e Entropia de Ligação

Como foi visto nas Tabela 8 e Tabela **9**, o valor da contribuição entálpica para a energia de ligação foi bastante considerável e maior do que a contribuição entrópica, mesmo para os ligantes que deveriam ter uma contribuição dessa última mais elevada – ou seja, aqueles que são estritamente apolares no sítio estritamente apolar.

Morton e seus colaboradores[39] mostraram em seu trabalho que a contribuição entálpica é maior para o caso dos mutantes que possuem apenas L99A como mutação, devido às interações entre o ligante e o receptor não se basearem na transferência de solvente e sim numa cavidade enterrada e apolar que acomoda um ligante também apolar, o que compensaria a perda entrópica.

Além disso, experimentalmente, os ligantes polares cujos receptores possuíam a mutação M102E, apesar de serem dominados pela entalpia, deveriam também apresentar uma contribuição entrópica considerável, uma vez que ela

advém da necessidade de estabilizar a rede de ligações de hidrogênio.[40] Tal contribuição não foi vista nos resultados das simulações.

Assim como as energias livres de ligação médias calculadas são maiores que as determinadas experimentalmente,[40-41] os valores de entalpia e entropia obtidos pelos métodos de MC também não foram os mesmos. Isso pode ser justificado pela rigidez do receptor e, consequentemente, do sítio de ligação contribui bastante para discrepância entre os valores obtidos de entropia e entalpia, uma vez que o receptor não pode se “mover” e não garante que todas as regiões de mínimo possam ser amostradas.

Os ligantes cujos receptores possuíam as mutações L99A e M102Q também apresentaram resultados de entalpia mais elevados (valores não apresentados aqui, uma vez que não há dados experimentais para comparação), como era de se esperar dado que eles apresentam regiões polares que podem interagir com a parte polar adicionada à cavidade pela nova mutação.

# 6 CONCLUSÃO

A abordagem utilizada para a determinação da energia livre de ligação, a qual combinou energias médias no *ensemble* de Monte Carlo e estimativa da entropia pela aproximação de primeira ordem mostrou-se eficiente em termos de tempo computacional e para comparação com energias livres de ligação determinadas experimentalmente. As propriedades termodinâmicas que foram avaliadas diretamente possibilitam a determinação de perfis termodinâmicos para os ligantes, os quais apresentaram uma boa correlação quando comparados a outros métodos. Sendo o algoritmo capaz de encontrar mínimos de energia e mapear o sítio de ligação a fim de elucidar a maneira como as interações entre ligante e receptor estão acontecendo.

Apesar do modelo descrito aqui como *full_flex* ser mais rigoroso e representar melhor o que se deveria esperar do comportamento das moléculas durante a simulação de Monte Carlo e, consequentemente, fazer uma amostragem mais adequada dos estados disponíveis para os ligantes, o modelo *sample_torsion*, que foi amplamente testado, foi capaz de apresentar resultados que podem ser utilizados como forma preliminar de análise de dados de energia livre de ligação por métodos computacionais.

A utilização de métodos mais robustos, como a dinâmica molecular e seus métodos de calcular as energias livre de ligação entre os ligantes e seus respectivos receptores é bastante custosa computacionalmente falando, apesar de resultar em dados com maior correlação com os experimentais para quantidades maiores de complexos.

O programa, no geral, apresenta uma boa estimativa das energias livres de ligação usando os dois métodos de amostragem, com resultados ligeiramente melhores para as que utilizaram o modelo eletrostático Poisson. O custo computacional para esse tipo de simulação é bastante reduzido quando comparado com as DM, vide Tabela 15, de modo que ele se mostra uma alternativa bastante satisfatória para análises preliminares.

Tabela 15 - Tempo gasto para obter as coordenadas por meio da dinâmica molecular e Monte Carlo e para o cálculo das variações de energia livre a partir das coordenadas geradas.

| Método | Simulação | Tempo Gasto | Processamento |
|---|---|---|---|
| DM | Coordenadas | ~ 15h | 1x GTX 980 |
| | | | |
| DM | Energia – LIE | < 5 min | 1x Xeon E5640 |
| DM | Energia – Vácuo | ~ 3h | 1x GTX 980 |
| DM | Energia – MM-GBSA | ~ 7h | 1x GTX 980 |
| DM | Energia – MM-PBSA | ~ 12h | 1x GTX 980 |
| | | | |
| MC | *sample_torsion* | ~ 4h | 1x Xeon E5640 |
| MC | *full_flex* | ~12h | 1x Xeon E5640 |

Fonte: Elaborada pelo autor.

Utilizar um modelo eletrostático mais robusto, como o Poisson, resultou, no geral, uma melhora significativa na correlação entre os valores das energias livres de ligação experimentais e os calculados. Apesar de ainda errarmos os valores absolutos para essas energias, os perfis observados são bastante interessantes, o que confirma a possibilidade de utilizar o módulo de Monte Carlo do programa LiBELa para análises preliminares dos perfis energéticos de moléculas de interesse.

Apesar da aproximação de primeira ordem para o cálculo de entropia se mostrar uma alternativa interessante para obter essa propriedade, os resultados aqui obtidos não foram tão satisfatórios quanto poderiam ser, de modo que futuras melhorias no programa deverão ser feitas a fim de aprimorar esse quesito. Além disso, permitir que o receptor seja flexível também auxiliaria a gerar resultados mais acurados.

# REFERÊNCIAS


1 NELSON, C. **Lehninger principles of biochemistry**. New York: W.H. Freeman, 2013.

2 BRANDSDAL, B. O. et al. Free energy calculations and ligand binding. **Advances in Protein Chemistry**, v. 66, p. 123–158, 2003. doi: 10.1016/S0065-3233(03)66004-3.

3 LEACH, A. R. **Molecular modelling**: principles and applications. New York: Prentice Hall, 2001.

4 MOBLEY, D. L.; GILSON, M. K. Predicting binding free energies: frontiers and benchmarks. **Annual Review of Biophysics**, v. 46, n. 1, p. 531–558, 2017. doi: 10.1146/annurev-biophys-070816-033654.

5 MEIROVITCH, H. Methods for calculating the absolute entropy and free energy of biological systems based on ideas from polymer physics. **Journal of Molecular Recognition**, v. 23, n. 2, p. 153–172, 2010.

6 SCHLICK, T. **Molecular modeling and simulation**: an interdisciplinary guide. 2nd ed. New York: Springer, 2010. (Interdisciplinary applied mathematics, v. 21.

7 GONZÁLEZ, M. A. **Force fields and molecular dynamics simulations**. 2011. Disponível em: <https://www.neutron-sciences.org/articles/sfn/pdf/2011/01/sfn201112009.pdf>. Acesso em: 23 jan. 2018.

8 VLACHAKIS, D. et al. Current State-of-the-Art Molecular Dynamics Methods and Applications. **Advances in protein chemistry and structural biology**. v. 94, 269–313, 2014. doi: 10.1016/B978-0-12-800168-4.00007-X.

9 KOLLMAN, P. A. et al. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. **Accounts of Chemical Research**, v. 33, n. 12, p. 889–897, 2000.

10 BROOKS, B. R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. **Journal of Computational Chemistry**, v. 4, n. 2, p. 187–217, 1983.

11 HERMANS, J. et al. A consistent empirical potential for water-protein interactions. **Biopolymers**, v. 23, n. 8, p. 1513–1518, 1984.

12 WEINER, S. J. et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. **Journal of the American Chemical Society**, v. 106, n. 3, p. 765–784, 1984.

13 ROCCHIA, W.; NESHICH, G. Electrostatic potential calculation for biomolecules - creating a database of pre-calculated values reported on a per residue basis for all PDB protein structures. **Genetics and Molecular Research**, v. 6, n. 4, p. 923–936, 2007.

14 LI, C. et al. Progress in developing Poisson-Boltzmann equation solvers. **Molecular Based Mathematical Biology**, v. 1, 2013. doi: 10.2478/mlbmb-2013-0002.

15 ALLEN, M. **Introduction to molecular dynamics simulations**. 2004. Disponível em: <https://udel.edu/~arthij/MD.pdf>. Acesso em 23 jan. 2018.

16 VERLET, L. Computer “experiments” on classical fluids. I. Thermodynamical properties of Lennard-Jones molecules. **Physical Review**, v. 159, n. 1, p. 98–103, 51967.

17 SWOPE, W. C. et al. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. **Journal of Chemical Physics**, v. 76, n. 1, p. 637–649, 1982.

18 VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C. A Leap-frog algorithm for stochastic dynamics. **Molecular Simulation**, v. 1, n. 3, p. 173–185, 1988.

19 BEEMAN, D. Some multistep methods for use in molecular dynamics calculations. **Journal of Computational Physics**, v. 20, n. 2, p. 130–139, 1976.

20 NEWMAN, M.; BARKEMA, G. **Monte Carlo methods in statistical physics**. Oxford : Clarendon Press, 1999.

21 METROPOLIS, N. et al. Equation of state calculations by fast computing machines. **Journal Chemical Physics**, v. 21, n. 6, p. 1087–1092, 1953.

22 DOS SANTOS MUNIZ, H.; NASCIMENTO, A. S. Ligand-and receptor-based docking with LiBELa. **Journal of Computer-Aided Molecular Design**, v. 29, n. 8, p. 713–723, 2015.

23 ÅQVIST, J.; MEDINA, C.; SAMUELSSON, J. E. A new method for predicting binding affinity in computer-aided drug design. **Protein Engineering, Design and Selection**, v. 7, n.3, p. 385-391, 1994.

24 SUN, H. et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. entropy effects on the performance of end-point binding free energy calculation approaches. **Physical Chemistry Chemical Physics**, v. 20, n. 21, p. 14450–14460, 2018.

25 WANG, C. et al. Recent developments and applications of the MMPBSA method. **Frontiers in Molecular Biosciences**, v. 4, , p. 87, 2018. doi: 10.3389/fmolb.2017.00087.

26 STILL, W. C. et al. Semianalytical treatment of solvation for molecular mechanics and dynamics. **Journal of the American Chemical Society**, v. 112, n. 16, p. 6127–6129, 1990.

27 PURISIMA, E.; SULEA, T. Solvation models: theory and validation. **Current Pharmaceutical Design**, v. 20, n. 20, p. 3266–3280, 2014.

28 LIU, H.-Y.; KUNTZ, I. D.; ZOU, X. Pairwise GB/SA scoring function for structure-based drug design. **Journal of Physical Chemistry B**, v. 108, n. 17, p. 5453–5462, 2004.

29 ZWANZIG, R. W. High-temperature equation of state by a perturbation method. I. nonpolar gases. **Journal of Chemical Physics**, v. 22, n. 8, p. 1420–1426, 1954.

30 KOLLMAN, P. Free energy calculations: applications to chemical and biochemical phenomena. **Chemical Reviews**, v. 93, n. 7, p. 2395–2417, 1993.

31 BRANDSDAL, B. O. et al. Free energy calculations and ligand binding. **Advances in Protein Chemistry**, v. 66, p. 123–158, 2003. doi: 10.1016/S0065-3233(03)66004-3.

32 STEINBRECHER, T. B. et al. Accurate binding free energy predictions in fragment optimization. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2411–2420, 2015.

33 EDHOLM, O.; BERENDSEN, H. J. C. Entropy estimation from simulations of non-diffusive systems. **Molecular Physics**, v. 51, n. 4, p. 1011–1028, 1984.

34 PANDEY, K. C.; KAXIRAS, E. Entropy calculation beyond the harmonic approximation: application to diffusion by concerted exchange in Si. **Physical Review Letters**, v. 66, n. 7, p. 915–918, 1991.

35 CHANG, C.-E.; CHEN, W.; GILSON, M. K. Evaluating the accuracy of the quasiharmonic approximation. **Journal of Chemical Theory and Computation**, v. 1, n. 5, p. 1017–1028, 2005.

36 MENZER, W. M. et al. Simple entropy terms for end-point binding free energy calculations. **Journal of Chemical Theory and Computation**, v. 14, n. 11, p. 6035–6049, 2018.

37 KILLIAN, B. J.; YUNDENFREUND KRAVITZ, J.; GILSON, M. K. Extraction of configurational entropy from molecular simulations via an expansion approximation. **Journal of Chemical Physics**, v. 127, n. 2, p. 024107, 2007.

38 WEI, B. Q. et al. A model binding site for testing scoring functions in molecular docking. **Journal of Molecular Biology**, v. 322, n. 2, p. 339–355, 2002.

39 MORTON, A.; BAASE, W. A.; MATTHEWS, B. W. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. **Biochemistry**, v. 34, n. 27, p. 8564–8575, 1995.

40 LIU, L.; BAASE, W. A.; MATTHEWS, B. W. Halogenated Benzenes Bound within a Non-polar Cavity in T4 Lysozyme Provide Examples of I···S and I···Se Halogen-bonding. **Journal of Molecular Biology**, v. 385, n. 2, p. 595–605, 2009.

41 BOYCE, S. E. et al. Predicting ligand binding affinity with alchemical free energy methods in a polar model binding site. **Journal of Molecular Biology**, v. 394, n. 4, p. 747–763, dez. 2009.

42 ERIKSSON, A. E. et al. A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. **Nature**, v. 355, n. 6358, p. 371–373, 1992.

43 WORD, J. M. et al. Asparagine and glutamine: using hydrogen atom contacts in the choice of side-chain amide orientation. **Journal of Molecular Biology**, v. 285, n. 4, p. 1735–1747, 1999.

44 CASE, D. A. **AMBER 16.** University of California, San Francisco. 2016. Disponível em: <http://ambermd.org/doc12/Amber16.pdf>. Acesso em 23 jan. 2018.

45 WANG, J. et al. Development and testing of a general amber force field. **Journal of Computational Chemistry**, v. 25, n. 9, p. 1157-1174, 2004.

46 JAKALIAN, A. et al. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: I. method. **Journal of Computational Chemistry**, v. 21, n. 2, p. 132–146, 2000.

47 MAIER, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. **Journal of Chemical Theory and Computation**, v. 11, n. 8, p. 3696–3713, 2015.

48 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-alkanes. **Journal of Computational Physics**, v. 23, n. 3, p. 327–341, 1977.

49 ROE, D. R.; CHEATHAM, T. E. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. **Journal of Chemical Theory and Computation**, v. 9, n. 7, p. 3084–3095, 2013.

50 PETTERSEN, E. F. et al. UCSF Chimera - a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, out. 2004.

51 MUNIZ, H. DOS S. **Métodos híbridos em docagem molecular**: implementação, validação e aplicação. 2018. 113 p. Tese (Doutorado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2018.

52 CUNHA, J. V. S. **Aplicação de Monte Carlo para a geração de ensembles e análise termodinâmica da interação biomolecular**. 2016. 112 p. Dissertação (Mestrado em Ciências) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2016..

53 LUTY, B. A. et al. A molecular mechanics/grid method for evaluation of ligand-receptor interactions. **Journal of Computational Chemistry**, v. 16, n. 4, p. 454–464, 1995.

54 MENG, E.; SHOICHET, B.; KUNTZ, I. Automated docking with grid-based energy evaluation. **Journal of Computational Chemistry**, v. 13, n. 4, p. 505–524, 1992.

55 O'BOYLE, N. M. et al. Open babel: an open chemical toolbox. **Journal of Cheminformatics**, v. 3, n. 1, p. 33, 2011. doi: 10.1186/1758-2946-3-33.

56 GALASSI, M. et al. **GNU scientific library reference manual.** 2009. Disponível em: <https://www.gnu.org/software/gsl/doc/latex/gsl-ref.pdf>. Acesso em: 23 jan. 2018.

57 BROWN, S. P.; MUCHMORE, S. W.; HAJDUK, P. J. Healthy skepticism: assessing realistic model performance. **Drug Discovery Today**, v. 14, n. 7–8, p. 420–427, 2009.

58 SHARP, K. A.; HONIG, B. Electrostatic macromolecules : theory and applications. **Annual Review of Biophysics and Biophysical Chemistry**, v. 19, p. 301–332, 1990. doi: 10.1146/annurev.bb.19.060190.001505.

59 BINDER, K.; HEERMANN, D. W. **Monte Carlo simulation in statistical physics:** an introduction. New York, Springer, 1985.

60 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, 1996.

61 MORGON, N. H.; COUTINHO, K. **Métodos em química teórica e modelagem molecular**. São Paulo, Livraria da Física, 2007.

62 FERRARI, A. M. et al. Soft docking and multiple receptor conformations in virtual screening. **Journal of Medicinal Chemistry**, v. 47, n. 21, p. 5076–5084, 2004.

63 STOUTEN, P. F. W. et al. An effective solvation term based on atomic occupancies for use in protein simulations. **Molecular Simulation**, v. 10, n. 2–6, p. 97–120, 1993.

64 VERKHIVKER, G. M. et al. Towards understanding the mechanisms of molecular recognition by computer simulations of ligand-protein interactions. **Journal of Molecular Recognition**, v. 12, n. 6, p. 371–389, 1999.

65 MUNIZ, H. S.; NASCIMENTO, A. S. Towards a critical evaluation of an empirical and volume-based solvation function for ligand docking. **PLoS ONE**, v. 12, n. 3, p. 1–19, 2017.

66 LIU, L. et al. Use of stabilizing mutations to engineer a charged group within a ligand-binding hydrophobic cavity in T4 lysozyme. **Biochemistry**, v. 48, n. 37, p. 8842–8851, 2009.

67 PURISIMA, E. O.; HOGUES, H. Protein–ligand binding free energies from exhaustive docking. **Journal of Physical Chemistry B**, v. 116, n. 23, p. 6872–6879, 2012.

68 PODTELEZHNIKOV, A. A.; WILD, D. L. Inferring knowledge based potentials using contrastive divergence. In: HAMELRYCK, T.; MARDIA K.; FERKINGHOFF-BORG J. (Ed) **Bayesian methods in structural bioinformatics.** statistics for biology and health. Berlin: Springer, 2012 p. 135–155. doi: 10.1007/978-3-642-27225-7_5.

69 ZUCKERMAN, D. M. **Statistical physics of biomolecules**. Boca Raton:CRC Press, 2010.

70 HERSCHBACH, D. R.; JOHNSTON, H. S.; RAPP, D. Molecular partition functions in terms of local properties. **Journal of Chemical Physics**, v. 31, n. 6, p. 213–217, 1959.

# APÊNDICE A – Deslocamentos na Caixa de Simulação

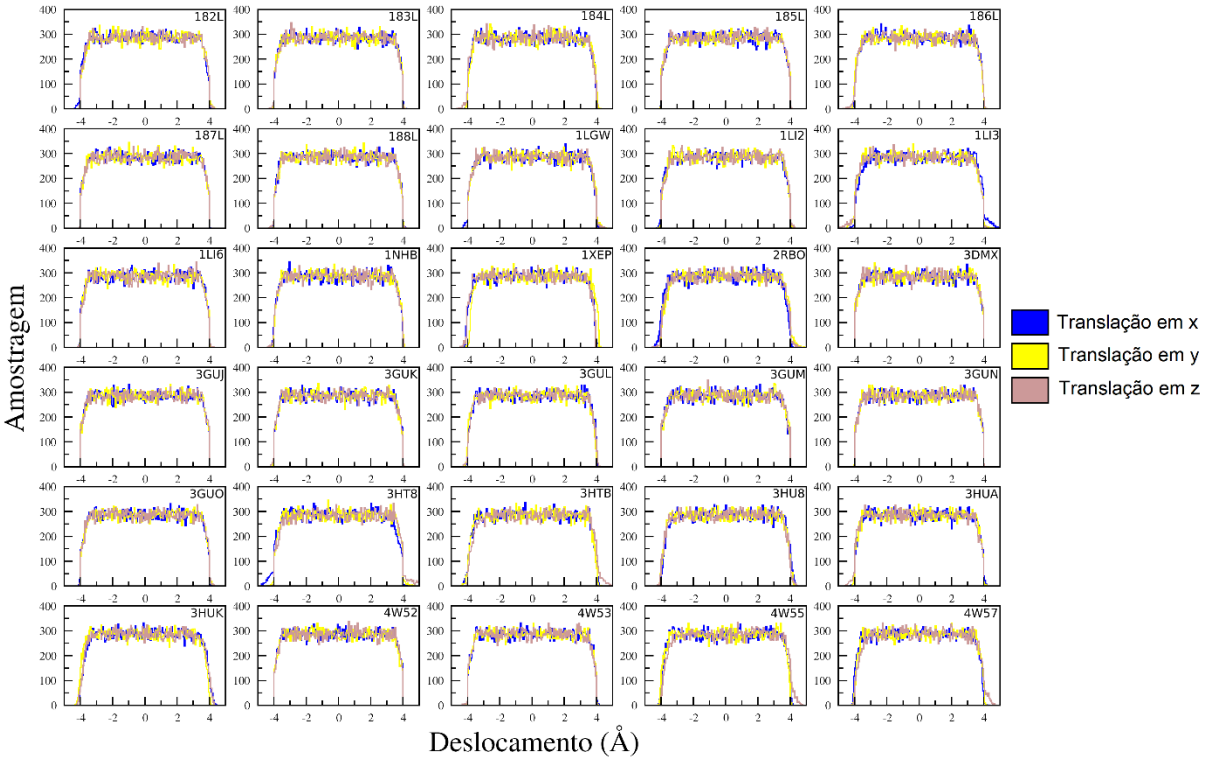



Figura 35 - Deslocamento dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa os deslocamentos na coordenada x do espaço, amarelo representa os deslocamentos na coordenada y e marrom, em z.

Fonte: Elaborada pelo autor.

A Figura 35 mostra o comportamento dos diferentes ligantes dentro da caixa cúbica delimitada durante as simulações de MC, usando o modelo eletrostático de Coulomb. Nela é possível ver que todos os ligantes se deslocaram em todo o espaço nas três coordenadas disponíveis, isto é, foi possível amostrar configurações que passaram por toda extensão da caixa de simulação. Além disso, para todas as coordenadas, o comportamento dos ligantes foi bastante semelhante.

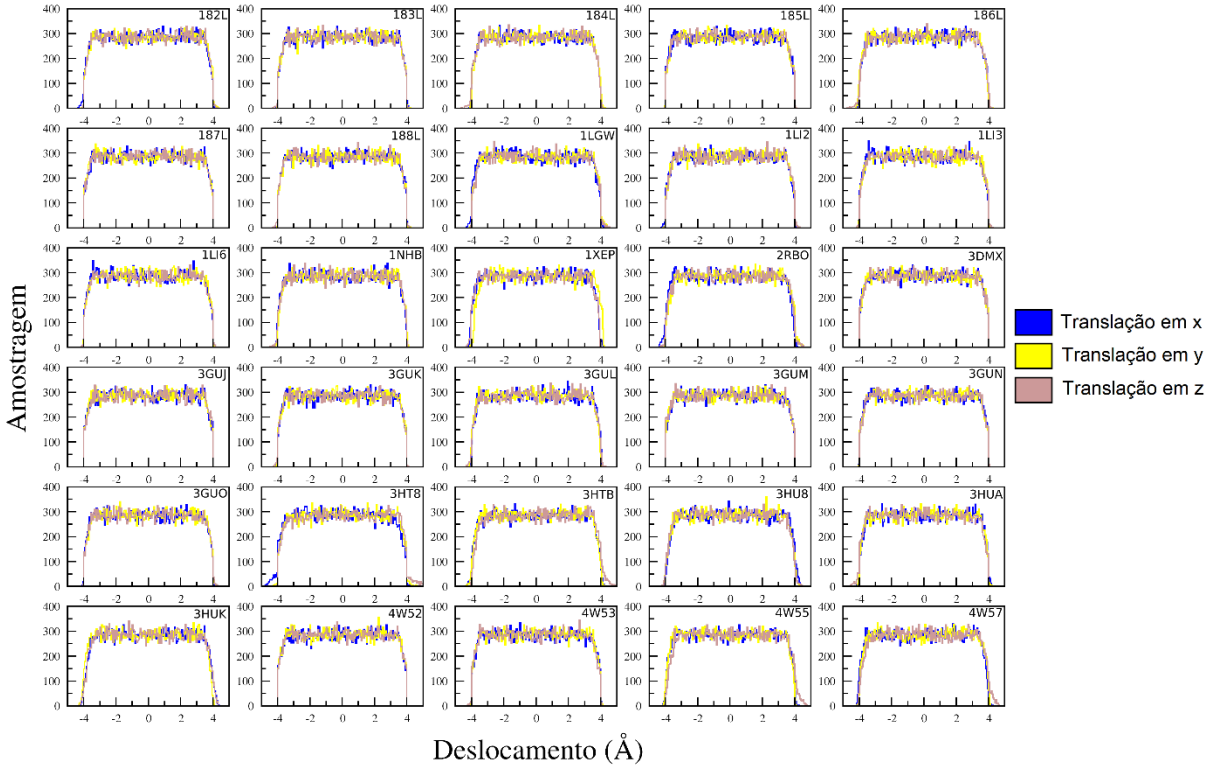



Figura 36 - Deslocamento dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Poisson como modelo eletrostático. Na figura, azul representa os deslocamentos na coordenada x do espaço, amarelo representa os deslocamentos na coordenada y e marrom, em z.

Fonte: Elaborada pelo autor.

Na Figura 36 estão representadas as amostragens para os deslocamentos nas três coordenadas para todos os ligantes simulados, utilizando como modelo eletrostático Poisson. Assim como no modelo anterior, aqui também é possível ver que os ligantes se deslocaram em todo o espaço permitido, amostrando configurações me toda a caixa cúbica de simulações. Novamente, o comportamento de todos os ligantes para todas as coordenadas foi bastante semelhante, isto é, não apresentaram sinais claros de favorecimento de uma delas em detrimento de outra. A utilização de um modelo eletrostático distinto não influenciou em gerar configurações que ocupassem todo o espaço disponível durante as etapas de translações.

Além das translações, as rotações dos ligantes durante as simulações também foram avaliadas, como pode ser visto a seguir.

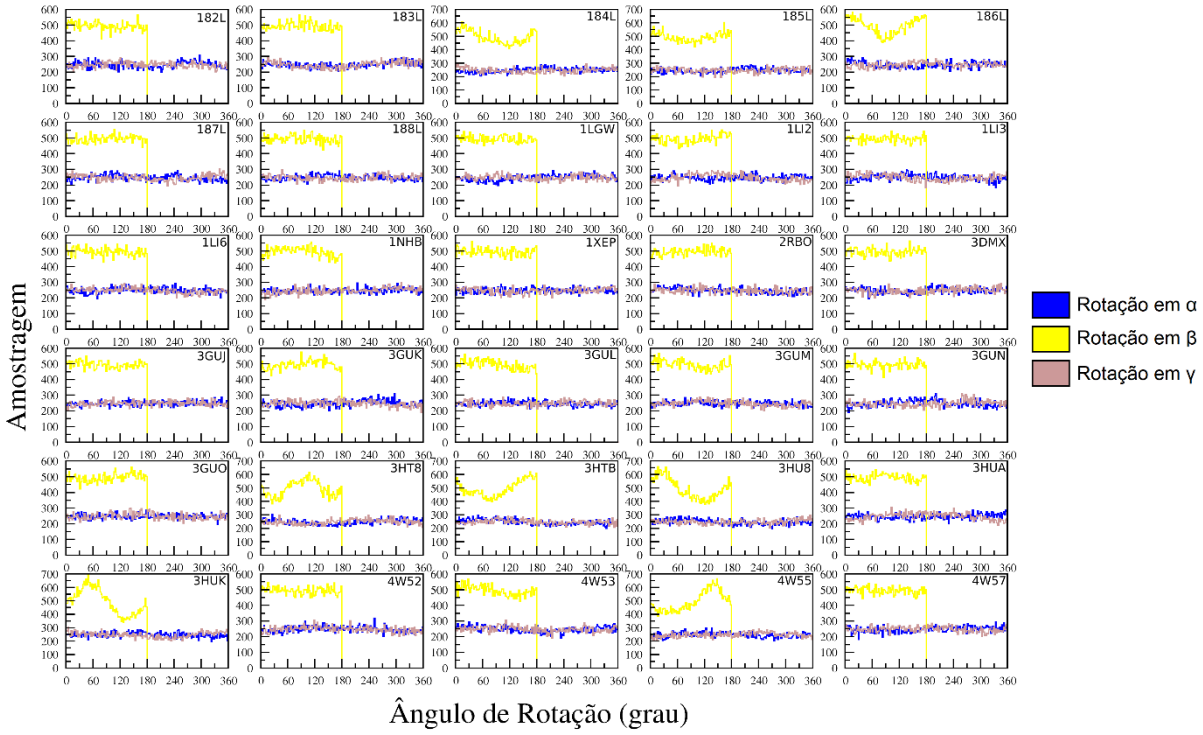



Figura 37 - Rotação dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa rotações em α no espaço rotacional, amarelo representa rotações em β e marrom, as rotações em γ.

Fonte: Elaborada pelo autor.

Na Figura 37 é possível ver o comportamento das rotações dos ligantes durante as simulações de Monte Carlo. Nela podemos notar que as rotações em α e γ no espaço de Euler, limitadas entre 0º e 360º, os ligantes amostraram configurações de maneira semelhante, isto é, não houve preferência por nenhuma delas. Para a rotação em β, definida entre 0º e 180º, o comportamento geral dos ligantes foi bastante semelhante, excetuando-se aqueles que apresentam ligações torcionáveis, que amostraram menos configurações em determinadas faixas de ângulos e mais em outras. Apesar disso, todo o espaço de rotações foi amostrado

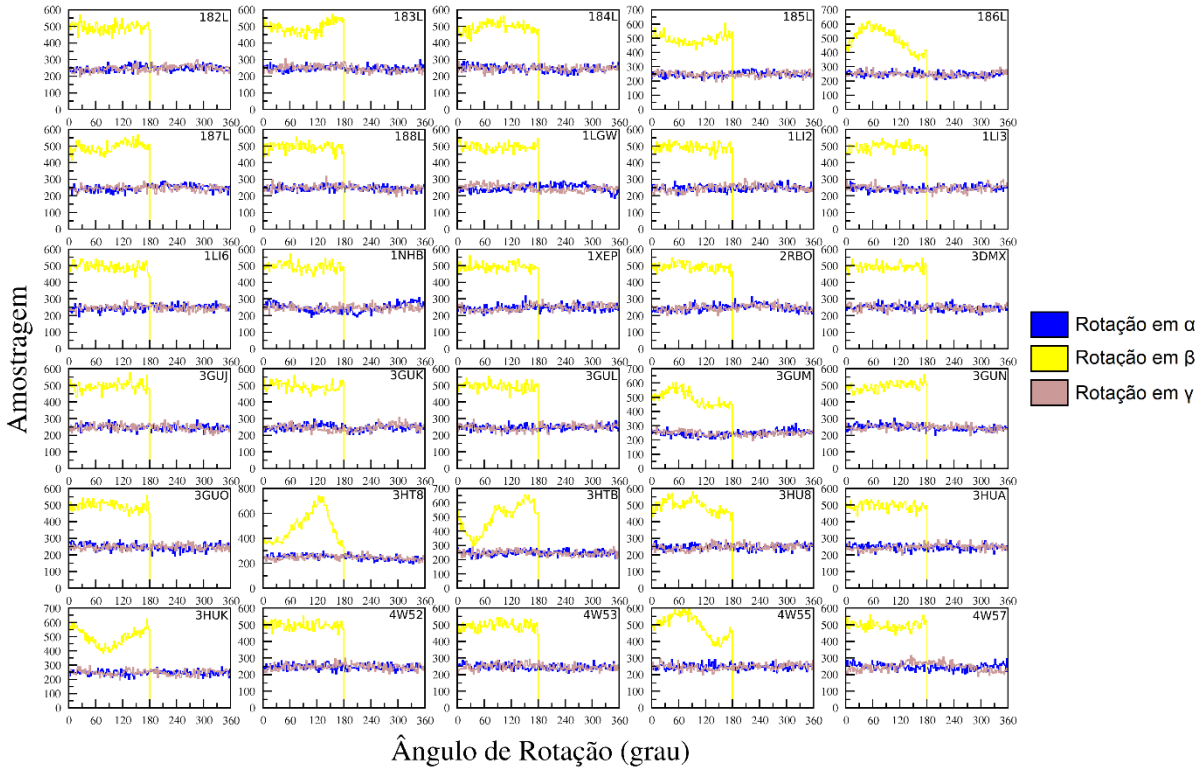



Figura 38 - Rotação dentro da caixa de simulação para as trajetórias dos ligantes durante MC, utilizando 250K como temperatura para o sistema, *sample_torsion* como modelo de amostragem e Coulomb como modelo eletrostático. Na figura, azul representa rotações em α no espaço rotacional, amarelo representa rotações em β e marrom, as rotações em γ.

Fonte: Elaborada pelo autor.

A Figura 38 apresenta as rotações dos ligantes amostradas durante as simulações de MC. Assim como no caso anterior, é possível ver um comportamento bastante semelhante para todos os ligantes nos ângulos α e γ. Entretanto, o ângulo β apresenta as diferenças, principalmente entre os ligantes que possuem ligações torcionáveis, novamente, amostrando menos configurações em determinadas faixas de ângulos e mais em outras. Para esse modelo eletrostático, os perfis amostrados em β para os ligantes com ligações rotacionáveis mostraram-se diferentes dos perfis amostrados usando Coulomb, isto é, as faixas de ângulos mais amostradas neste modelo foram as menos amostradas no outro.

# APÊNDICE B – Entropias e Entalpias

Tabela 16 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Coulomb como modelo eletrostático, sem solvatação, usando 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta H_{exp}$ (kcal/mol) | $-T\Delta S_{exp}$ (kcal/mol) | $\Delta G_{calc}$ (kcal/mol) | $\Delta H_{calc}$ (kcal/mol) | $-T\Delta S_{calc}$ (kcal/mol) |
|---|---|---|---|---|---|---|
| **182L** | -5,46 | -8,04 | 2,58 | -3,53 | -5,38 | 1,85 |
| **183L** | -5,13 | -8,31 | 3,18 | -4,28 | -6,19 | 1,91 |
| **184L** | -6,51 | -7,09 | 0,56 | -7,70 | -9,57 | 1,87 |
| **185L** | -4,89 | -11,23 | 6,34 | -4,82 | -6,80 | 1,98 |
| **186L** | -6,70 | -8,06 | 1,36 | -9,96 | -11,92 | 1,96 |
| **187L** | -4,67 | -6,97 | 2,30 | -2,13 | -3,73 | 1,60 |
| **188L** | -4,90 | -8,45 | 3,85 | -3,09 | -4,64 | 1,55 |
| **1NHB** | -5,76 | -6,76 | 1,00 | -0,81 | -2,50 | 1,70 |
| **3DMX** | -5,35 | -6,32 | 1,13 | -1,09 | -2,78 | 1,69 |
| **4W52** | -5,19 | -6,32 | 1,13 | -0,76 | -2,25 | 1,49 |
| **4W53** | -5,52 | -6,53 | 1,10 | -0,95 | -2,64 | 1,69 |
| **4W55** | -6,55 | -9,97 | 3,42 | -6,95 | -8,89 | 1,94 |
| **4W57** | -6,69 | -8,06 | 1,36 | -9,04 | -11,03 | 1,99 |
| **3GUJ** | -5,80 | -6,00 | 0,19 | -2,25 | -4,10 | 1,85 |
| **3GUK** | -5,30 | -5,30 | 0,03 | -2,85 | -4,85 | 2,00 |
| **3GUL** | -5,50 | -5,00 | -0,50 | -0,74 | -2,69 | 1,95 |
| **3GUM** | -4,90 | -3,70 | -1,16 | -1,46 | -3,32 | 1,86 |
| **3GUN** | -4,00 | -8,10 | 4,10 | -5,61 | -7,59 | 1,97 |
| **3GUO** | -5,00 | -9,20 | 4,20 | -3,26 | -5,20 | 1,94 |
| **Correlação de Pearson** | | | | **0,54** | **0,53** | **0,09** |
| **Correlação por média de Bootstrap** | | | | **0,54** | **0,50** | **0,13** |

Fonte: Elaborada pelo autor.

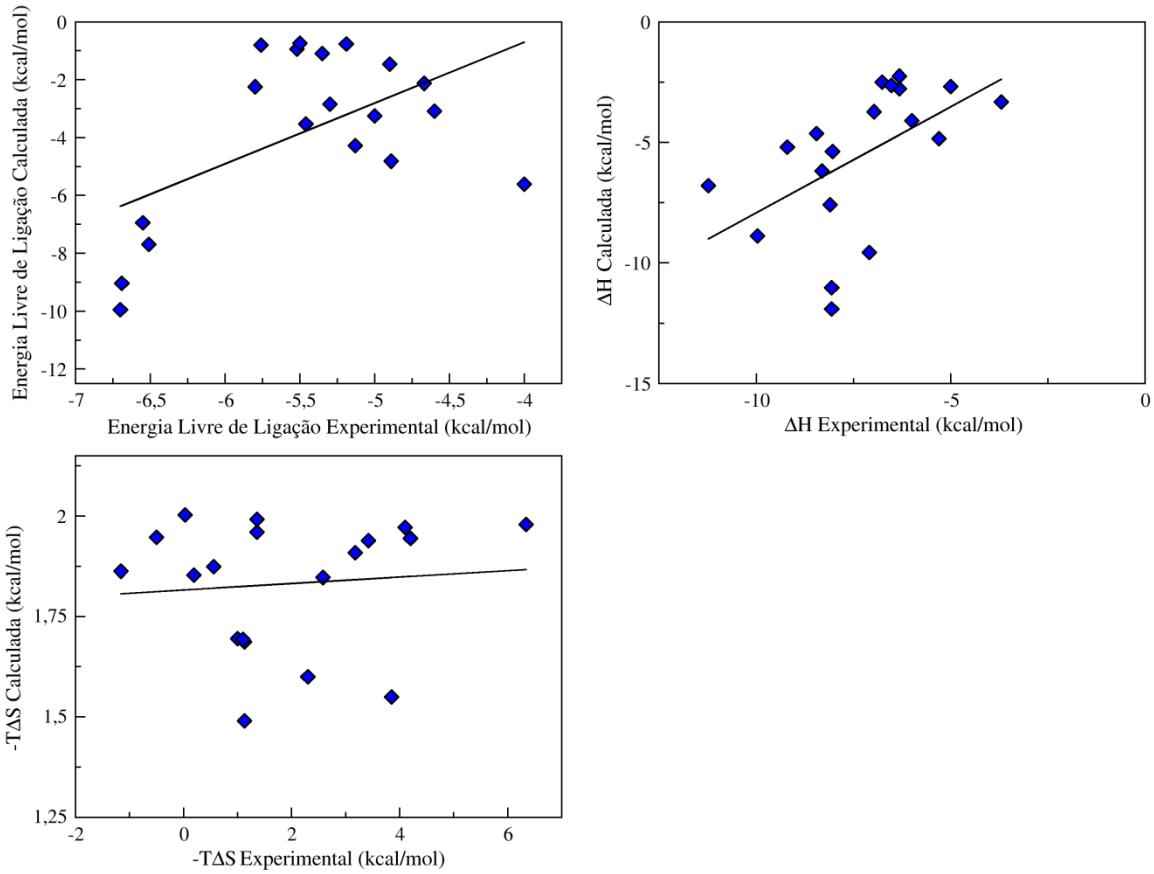



Figura 39 - Comparação entre os valores experimentais e calculados das energias livres de ligação, entalpia e entropia para 250K, usando Coulomb como modelo eletrostático, sem considerar solvatação, usando 0,5 Å como constante de amortecimento. Nos gráficos, ΔH é a entalpia e -TΔS é a entropia.

Fonte: Elaborada pelo autor.

Ao compararmos as energias livres de ligação experimentais e calculadas vemos que para esse conjunto de ligantes (Tabela 16 e Figura 39), temos que o valor da correlação entre tais grandezas é razoável, sendo r = 0,54. Os valores das entalpias calculadas comparados com os valores experimentais também apresentaram correlação razoável, isto é, r = 0,53. Entretanto, para as entropias o valor calculado foi bastante ruim, r = 0,09. É possível ver que a contribuição entálpica na composição da energia livre de ligação é muito maior do que a contribuição entrópica. Além disso, é valido dizer que para todas essas entidades, o programa também errou os valores absolutos, mas, para a energia livre de ligação e para a entalpia, a tendência do perfil termodinâmico está presente.

Tabela 17 - Análise da contribuição entrópica para a energia livre de ligação. Comparação entre os valores experimentais de energia, entalpia e entropia com os valores estimados pelo *LiBELa*, para a temperatura de 250K, Poisson como modelo eletrostático, sem solvatação, usando 0,5 Å como constante de amortecimento. Na tabela, $\Delta G_{exp}$, $\Delta H_{exp}$ e $-T\Delta S_{exp}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias experimentais. $\Delta G_{calc}$, $\Delta H_{calc}$ e $-T\Delta S_{calc}$ representam, respectivamente, as energias livres de ligação, entalpias e entropias calculadas.

| PDB ID | $\Delta G_{exp}$ (kcal/mol) | $\Delta H_{exp}$ (kcal/mol) | $-T\Delta S_{exp}$ (kcal/mol) | $\Delta G_{calc}$ (kcal/mol) | $\Delta H_{calc}$ (kcal/mol) | $-T\Delta S_{calc}$ (kcal/mol) |
|---|---|---|---|---|---|---|
| **182L** | -5,46 | -8,04 | 2,58 | -3,52 | -5,34 | 1,82 |
| **183L** | -5,13 | -8,31 | 3,18 | -4,59 | -6,49 | 1,90 |
| **184L** | -6,51 | -7,09 | 0,56 | -8,17 | -10,04 | 1,87 |
| **185L** | -4,89 | -11,23 | 6,34 | -4,60 | -6,57 | 1,97 |
| **186L** | -6,70 | -8,06 | 1,36 | -10,22 | -12,16 | 1,94 |
| **187L** | -4,67 | -6,97 | 2,30 | -2,25 | -3,84 | 1,59 |
| **188L** | -4,90 | -8,45 | 3,85 | -3,44 | -4,98 | 1,54 |
| **1NHB** | -5,76 | -6,76 | 1,00 | -0,78 | -2,46 | 1,68 |
| **3DMX** | -5,35 | -6,32 | 1,13 | -1,09 | -2,76 | 1,68 |
| **4W52** | -5,19 | -6,32 | 1,13 | -0,81 | -2,45 | 1,65 |
| **4W53** | -5,52 | -6,53 | 1,10 | -0,84 | -2,71 | 1,87 |
| **4W55** | -6,55 | -9,97 | 3,42 | -6,65 | -8,55 | 1,90 |
| **4W57** | -6,69 | -8,06 | 1,36 | -9,70 | -11,70 | 1,99 |
| **3GUJ** | -5,80 | -6,00 | 0,19 | -1,72 | -3,56 | 1,84 |
| **3GUK** | -5,30 | -5,30 | 0,03 | -1,70 | -3,69 | 2,00 |
| **3GUL** | -5,50 | -5,00 | -0,50 | 0,32 | -1,60 | 1,92 |
| **3GUM** | -4,90 | -3,70 | -1,16 | -0,34 | -2,19 | 1,84 |
| **3GUN** | -4,00 | -8,10 | 4,10 | -2,41 | -4,36 | 1,95 |
| **3GUO** | -5,00 | -9,20 | 4,20 | -1,21 | -3,14 | 1,93 |
| **Correlação de Pearson** | | | | **0,66** | **0,51** | **0,04** |
| **Correlação por média de Bootstrap** | | | | **0,66** | **0,45** | **0,08** |

Fonte: Elaborada pelo autor.

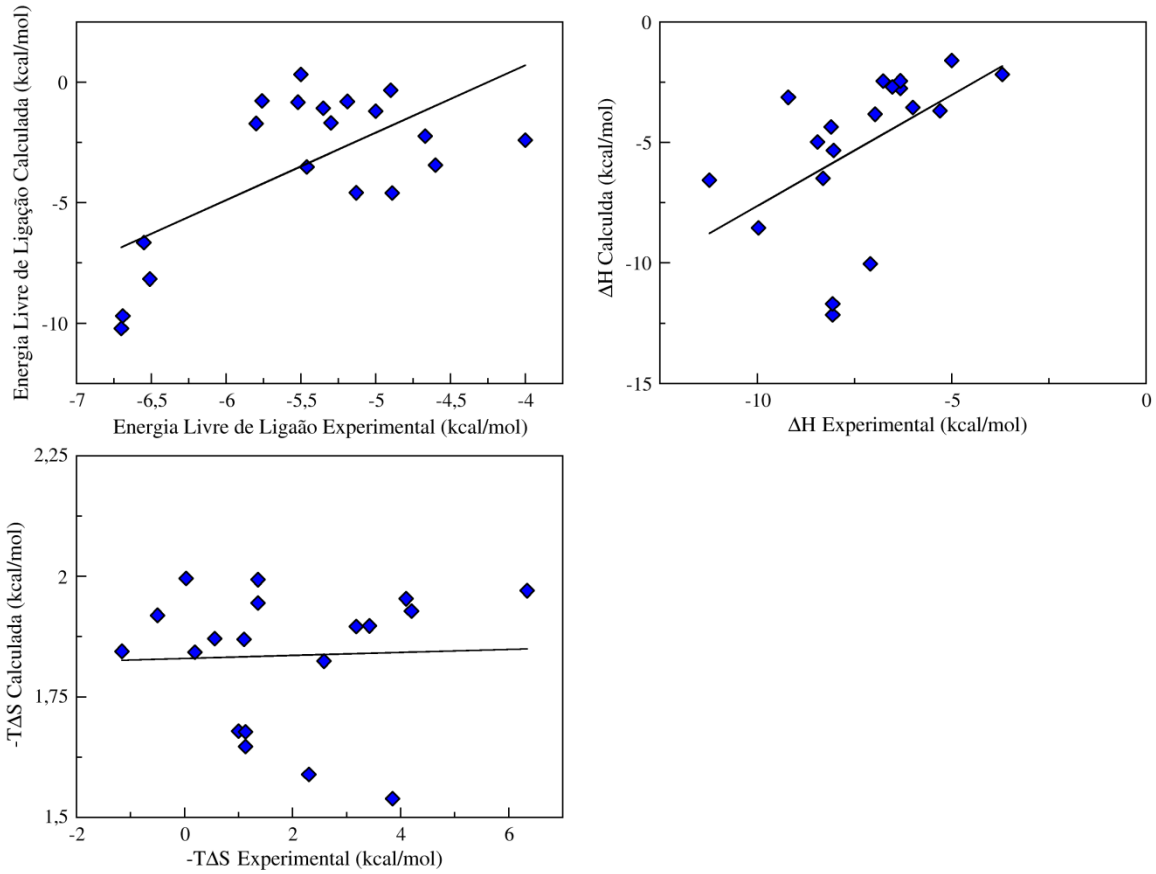



Figura 40 - Comparação entre os valores experimentais e calculados das energias livres de ligação, entalpia e entropia para 250K, usando Poisson como modelo eletrostático, sem considerar solvatação, usando 0,5 Å como constante de amortecimento. Nos gráficos, ΔH é a entalpia e -TΔS é a entropia.

Fonte: Elaborada pelo autor.

Comparando as energias livres de ligação experimentais com as calculadas pelo *LiBELa* (Figura 40 e Tabela 17), podemos ver que para esse conjunto de ligantes a correlação entre os valores foi de 0,66, enquanto que os valores calculados de entalpia tiveram uma correlação de 0,51 com os dados experimentais. Já para a entropia, as correlações foram consideravelmente baixas, chegando a -0,08. Aqui também fica clara a contribuição entálpica mais elevada na composição das energias livres de ligação e, novamente, os valores estimados não são iguais aos valores absolutos.

# APÊNDICE C – Input MCLiBELa

```
# mode

mode                    eq
dock_parallel           no
parallel_jobs           1

#
# input files
#

rec_mol2            ../../../../mol2/182L_DOCKPREP.mol2.gz
lig_mol2            ../../../../mol2/182L_LIG_DOCKPREP.mol2.gz
reflig_mol2         ../../../../mol2/182L_LIG_DOCKPREP.mol2.gz
multifile           multimol.dat

#
# force field parameters
#

deltaij                 0.0
deltaij_es              0.0
mol2_aa                 no
search_box              8.0 8.0 8.0
scoring_function        3
use_grids               yes
grid_spacing            0.4
grid_box                30.0    30.0    30.0
write_grids             McGrid
solvation_alpha         0.2
dielectric_model        r
diel                    1.0
ligand_energy_model     GAFF
atomic_model_ff         GAFF

#
# DelPhi Parameters
#

delphi_grid             ../../grids/phimap2.phi
use_delphi              yes
delphi_gsize            75

#
# Optimization
#

timeout                     120
minimization_tolerance      1.0e-10
minimization_delta          1.0e-5
dock_min_tol                1.0e-10
minimization_timeout        120
sort_by_energy              yes
overlay_optimizer           ln_auglag
energy_optimizer            direct
ignore_h                    no
deal                        no
```

```
elec_scale                      1.0
vdw_scale                       1.0

#
# MC Options
#
nsteps                          50000000
temperature                     250.0
cushion                         0.5
max_atom_displacement           0.04
rotation_step                   2.0
torsion_step                    2.0
sample_torsions                 no
mc_full_flex                    yes
compute_rotation_entropy        no
equilibration_steps             1000000
ligand_simulation               yes
mc_stride                       10000
seed                            26465
#entropy_rotation_bins          720
#entropy_translation_bins       20

#
# SA Options
#
sa_start_temp           100.0
sa_steps                10000
sa_mu_t                 1.1

#
# MCR Options
#
mcr_size                20
mcr_coefficients        1.090507733 1.090507733 1.090507733 1.189207115 1.189207115
1.189207115 1.189207115 1.189207115 1.189207115 1.189207115 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0 2.0
2.0


#
# output
#

output_prefix           182L_SF3_1
write_mol2               yes

#
# flexible ligands
#

generate_conformers             no
number_of_conformers            10
conformers_to_rank              1
conformer_generator             GA
conformer_min_steps             1000
```