

**Pontifícia Universidade Católica do Rio Grande do Sul**

**Faculdade de Farmácia**

**Mestrado Profissional em Biotecnologia Farmacêutica**

Gustavo Bellani Migott

**Determinação de energia livre de ligação por métodos *in silico* para ligantes da enzima InhA (EC 1.3.1.9) de *Mycobacterium tuberculosis***

Porto Alegre  
2016

Gustavo Bellani Migott

**Determinação de energia livre de ligação por métodos *in silico* para ligantes da enzima InhA (EC 1.3.19) de *Mycobacterium tuberculosis***

Dissertação apresentada como requisito parcial para obtenção do grau de mestre em Biotecnologia Farmacêutica pela Pontifícia Universidade Católica do Rio Grande do Sul.

Orientador: Prof. Dr. Osmar Norberto de Souza  
Coorientador: Dr. José Fernando Ruggiero Bachega

Porto Alegre  
2016

## **DEDICATÓRIA**

Dedico esta dissertação aos meus pais Ana Maria Bellani Migott e Luiz Carlos Migott, por todo amor, carinho e incentivo durante toda minha vida e na realização deste mestrado.

## **AGRADECIMENTOS**

Ao Prof. Dr. Osmar Norberto de Souza, por ter me aceitado em seu laboratório e por ter me apresentado ao mundo da bioinformática. Acreditado que nestes dois anos de convivência poderíamos realizar um excelente trabalho juntos.

Aos colegas de Mestrado pelo companheirismo e disponibilidade nas disciplinas e á equipe do LABIO/FarmInf, por toda a ajuda, amizade e força nestes dois anos de trabalhos conjuntos. Agradeço em particular ao Dr. José Fernando e ao Dr. Luís Fernando pelas discussões e ensinamentos sobre o presente estudo.

Por último agradecer em especial a minha mãe Ana Maria Bellani Migott por toda força, ajuda e incentivo para a realização deste mestrado, e ao meu pai Luiz Carlos Migott pelos ensinamentos.

## RESUMO

A tuberculose é uma doença infectocontagiosa responsável por cerca de 1,3 milhão de mortes anualmente a nível mundial. Apesar do aparecimento de cepas multirresistentes de *Mycobacterium tuberculosis* (Mtb), desde a década de 80 se observa uma lacuna no desenvolvimento de novos antimicrobianos. Com o advento da bioinformática e biofísica molecular computacional, tornou-se possível, a partir de um alvo molecular preestabelecido, testar inúmeros candidatos a fármacos, com destaque para a predição da energia livre de ligação. Nesta dissertação, foram selecionados 14 compostos com conhecida atividade frente a enzima 2-*trans*-enoil-ACP redutase (InhA, EC 1.3.1.9) de Mtb. Estas moléculas foram divididas em 3 grupos. Grupo 1: 5 compostos com valores esparsos de energia livre. Grupo 2: 9 compostos com valores similares de energia livre e derivadas do composto Genz-10850. Grupo 3: 14 compostos correspondendo à união dos Grupos 1 e 2. Amostragens por simulações de dinâmica molecular de 2 ns, em solvente explícito, permitiram estimar os valores da energia livre de ligação destes compostos com os métodos MM/GBSA, MM/PBSA, (QM)MM/GBSA, LigScore, DrugScore, AutoDock e SQM. O ranqueamento dos compostos foi baseado na correlação ( $R^2$ ) entre valores experimentais e estimados de energia livre. Os resultados apontaram valores de  $R^2$  similares entre os métodos testados. Técnicas mais robustas, como SQM e (QM)MM/GBSA, não obtiveram resultados mais acurados em comparação àquelas mais simples, como LigScore e DrugScore. No geral, foram obtidos valores moderados de correlação ( $R^2$  de 0,00-0,80) para o Grupo 1. Os Grupos 2 e 3 exibiram correlações fracas ( $R^2 < 0,40$ ). Apesar dos resultados satisfatórios para o Grupo 1, os métodos utilizados apresentaram limitações e não foram capazes de predizer e ranquear corretamente compostos com valores próximos de energia livre e estruturas moleculares similares, como os do Grupo 2.

Palavras chave: Enzima InhA, *Mycobacterium tuberculosis*, receptor flexível, dinâmica molecular, energia livre de ligação.

## ABSTRACT

Tuberculosis is an infectious disease responsible for about 1.3 million deaths annually worldwide. Despite the appearance of multi-resistant strains of *Mycobacterium tuberculosis* (Mtb), since the 80's there is a gap in the development of new antimicrobials. With the advent of bioinformatics and computational molecular biophysics, became possible to test, from a established molecular target, numerous molecules, especially associated with the prediction of binding free energy. In the current dissertation, were selected 14 compounds with recognized activity against the enzyme 2-*trans*-enol-ACP reductase (InhA, EC 1.3.1.9) of Mtb. These molecules were divided into three groups. Set 1: 5 compounds with distant values of binding free energy. Set 2: 9 compounds with close binding free energy values and similar molecular structures (derived from Genz 10850). Set 3: 14 ligands, corresponding to the sum of the set's 1 and 2. Sampling obtained from molecular dynamics and 2 ns of simulations, in explicit solvent, allowed to estimate the free energy of bind associated with the methods MM/GBSA, MM/PBSA, (QM)MM/GBSA, LigScore, DrugScore, AutoDock and SQM. The ranking of the compounds were based in the correlation ( $R^2$ ) between the predicted and experimental values. Results showed similar values of  $R^2$  in all tested methods. More accurate methods, such as SQM and (QM)/MM/GBSA, not obtained better correlations in comparison with simplified methods, as LigScore and DrugScore. In general, Set 1 obtained a moderate correlation ( $R^2$  of 0.00-0.80). Set's 2 and 3, showed weak correlation ( $R^2 < 0.40$ ). Despite the satisfactory results of Set 1, the tested methods presented limitations in the ranking of compounds with close values of experimental binding free energy and similar molecular structures, as Set 2.

Key words: InhA enzyme, *Mycobacterium tuberculosis*, flexible receptor, molecular dynamics, free energy of binding.

## Lista de Figuras

<b>Figura 1:</b> Reação de redução de um tioéster $\alpha,\beta$ -insaturado catalisada pela enzima InhA.....	17
<b>Figura 2:</b> Biossíntese de ácidos graxos em Mtb.....	18
<b>Figura 3:</b> Estrutura tridimensional da subunidade C da enzima InhA.....	19
<b>Figura 4:</b> Representação do tipo ribbons do homotetrâmero da InhA.....	20
<b>Figura 5:</b> Representação do tipo ribbons da cadeia principal da subunidade C destacando o sítio ativo da enzima.....	21
<b>Figura 6:</b> Sobreposição de cinquenta e uma estruturas de InhA de Mtb (2014).....	22
<b>Figura 7:</b> Conversão do NAD em INH-NAD catalisado pela enzima KatG.....	24
<b>Figura 8:</b> Estrutura molecular dos ligantes retirados do PDB.....	34
<b>Figura 9:</b> Composto Genz-10850, adaptado de Rotta et al., 2015.....	35
<b>Figura 10:</b> InhA de Mtb código PDB 1P44 envolto pela caixa de água TIP3P (10 Å).....	41
<b>Figura 11:</b> Ilustração da enzima InhA complexada ao ligante GEQ (PDB ID: 1BVR).....	47
<b>Figura 12:</b> Ilustração da fração do complexo considerado na região quântica, formada pelos resíduos que se encontram a 4 Å de distância do ligante.....	50
<b>Figura 13:</b> Comparação das simulações da enzima InhA código PDB: 1P44 pelo método de dinâmica molecular, comparando CPU e GPU.....	53
<b>Figura 14:</b> Diferença entre os campos de força testados para o Grupo 1 pelo método MM/GBSA.....	60
<b>Figura 15:</b> Diferença entre os campos de força testados para o Grupo 2 pelo método MM/GBSA.....	61
<b>Figura 16:</b> Diferença entre os campos de força testados para o Grupo 3 pelo método MM/GBSA.....	61
<b>Figura 17:</b> Diferença entre os campos de força testados para o Grupo 1 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2).....	64

<b>Figura 18:</b> Diferença entre os campos de força testados para o Grupo 2 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2).....	65
<b>Figura 19:</b> Diferença entre os campos de força testados para o Grupo 3 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2).....	66
<b>Figura 20:</b> Grupo 1, diferença entre as abordagens do ligante+receptor (Ligand/Complex) e múltiplas trajetórias (Multiple Trajectories).....	68
<b>Figura 21:</b> Grupo 2, diferença entre as abordagens do ligante+receptor (Ligand/Complex) e múltiplas trajetórias (Multiple Trajectories).....	68
<b>Figura 22:</b> Grupo 3, diferença entre as abordagens de trajetória do ligante+receptor (Ligand/Complex) e múltiplas trajetórias (Multiple Trajectories).....	69
<b>Figura 23:</b> Grupo 1, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.....	70
<b>Figura 24:</b> Grupo 2, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.....	70
<b>Figura 25:</b> Grupo 3, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.....	71
<b>Figura 26:</b> Grupo 1, Valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM.....	72
<b>Figura 27:</b> Grupo 2, valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM.....	73
<b>Figura 28:</b> Grupo 3, valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM.....	74
<b>Figura 29:</b> Grupo 1, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM.....	75
<b>Figura 30:</b> Grupo 2, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM.....	76
<b>Figura 31:</b> Grupo 3, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM.....	77
<b>Figura 32:</b> Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 1.....	80

<b>Figura 33:</b> Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 2.....	81
<b>Figura 34:</b> Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 3.....	82
<b>Figura 35:</b> Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 1.....	83
<b>Figura 36:</b> Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 2.....	84
<b>Figura 37:</b> Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 3.....	84
<b>Figura 38:</b> Resultados preditos pelo LigScore para o Grupo 1.....	85
<b>Figura 39:</b> Resultados preditos pelo LigScore para o Grupo 2.....	86
<b>Figura 40:</b> Resultados preditos pelo LigScore para o Grupo 3.....	86
<b>Figura 41:</b> Resultados preditos pelo DrugScore para o Grupo 1.....	87
<b>Figura 42:</b> Resultados preditos pelo DrugScore para o Grupo 2.....	87
<b>Figura 43:</b> Resultados preditos pelo DrugScore para o Grupo 3.....	88
<b>Figura 44:</b> Comparação entre métodos de SQM para o Grupo 1.....	88
<b>Figura 45:</b> Comparação entre métodos de SQM para o Grupo 2.....	89
<b>Figura 46:</b> Comparação entre métodos de SQM para o Grupo 2.....	89

## Lista de Tabelas

<b>Tabela 1:</b> Moléculas ligadas a InhA de Mtb, retiradas do Banco de dados de proteínas (PDB). Período de análise entre julho de 2014 á julho de 2015.....	34
<b>Tabela 2:</b> Modificações propostas para os ligantes derivados do Genz-10850.....	35
<b>Tabela 3:</b> Comparação da velocidade das simulações em sistemas de aproximadamente 36.500 átomos, comparando-se CPU e GPU.....	53
<b>Tabela 4:</b> Diferenças de $\Delta G$ calculadas pelo método MM/GBSA de acordo com tempo de simulação, da enzima InhA.....	58
<b>Tabela 5:</b> Diferenças de $\Delta G$ calculadas pelo método MM/PBSA para os ligantes 8PS e TCL. Resultados em kcal/mol.....	79
<b>Tabela 6:</b> Resultados dos métodos avaliados frente aos Grupos 1, 2 e 3 para o campo de força ff14SB.....	93

## Lista de Abreviaturas

**AA:** Aminoácidos

**ACP:** Acyl-carrier-protein – Proteína Carreadora de Acila

**TB:** Tuberculose

**OMS:** Organização Mundial de Saúde

**Mtb:** *Mycobacterium tuberculosis*

**LABIO:** Laboratório de Bioinformática, Modelagem e Simulação de Biosistemas

**FarmInf:** Laboratório de Farminformática

**FAS:** Fatty Acid Synthase - Sistema de Síntese de Ácidos Graxos

**SDR:** Short Chain Dehydrogenase - Família das desidrogenases de cadeia curta

**PDB:** Protein Data Bank - Banco de Dados de Proteínas

**ENR:** Enoil redutase

**InhA:** Enzima 2-*trans*-enoil-ACP (CoA) Redutase (CE 1.3.1.9)

**XDR-TB:** TB extensivamente resistente a drogas

**MDR-TB:** TB multi-resistente a drogas

**TDR-TB:** TB totalmente resistente a drogas

**PBSA:** Poisson–Boltzmann Surface Area

**GBSA:** Generalized Born Surface Area

**FF:** Force Field

**AMBER:** Assisted Model Building with Energy Refinement

**CHARMM:** Chemistry at HARvard Molecular Mechanics

**MM:** Mecânica Molecular

**DM:** Dinâmica Molecular

**GPU:** Graphical Processing Unity

**RMSD:** Root Mean-Square Deviation – Raiz Quadrada do Desvio Médio Quadrado

**QM:** Mecânica Quântica

## Lista de Símbolos

**Å:** Angstrom

**K:** Kelvin

**ns:** Nanossegundo

**ps:** Picossegundo

**fs:** Femtossegundo

**$\Delta G$ :** Energia livre de ligação

**$K_i$ :** Constante de inibição

**t:** tempo

**°C:** Graus Celsius

## Sumário

1. INTRODUÇÃO.....	15
1.1 TUBERCULOSE HUMANA.....	15
1.2 A ENZIMA InhA DE <i>Mycobacterium tuberculosis</i> .....	16
1.3 ÁCIDOS MICÓLICOS.....	17
1.4 A ESTRUTURA DA InhA.....	19
1.5 SÍTIO DE LIGAÇÃO AO SUBSTRATO DA InhA.....	20
1.6 DESENVOLVIMENTO DE CEPAS MULTIRRESISTENTES.....	23
1.7 DESENHO RACIONAL DE FÁRMACOS.....	25
1.7.1 Desenvolvimento de novos fármacos com abordagens <i>in silico</i> .....	25
1.7.2 Predição de energia livre de ligação por métodos <i>in silico</i> .....	26
2. OBJETIVOS GERAIS.....	28
2.1 OBJETIVOS ESPECÍFICOS.....	28
3. JUSTIFICATIVA.....	29
4.1 LOCAL DE EXECUÇÃO.....	31
4.2 O PACOTE DE SIMULAÇÃO AMBER.....	31
4.3 OBTENÇÃO E ESCOLHA DOS LIGANTES.....	32
4.3.1 Montagem dos modelos tridimensionais.....	35
4.3.2 Grupos de ligantes.....	36
4.4 AMOSTRAGEM POR DINÂMICA MOLECULAR.....	36
4.4.1 Dinâmica molecular em GPU.....	38
4.4.2 Protocolos de DM.....	38
4.4.2.1 Preparação dos arquivos de entrada.....	38
4.4.2.2 Campos de Força.....	40
4.4.2.3 Equilíbrio e minimização do sistema.....	40
4.4.2.4 Fase de Aquecimento.....	40
4.4.2.5 Fase de produção.....	41
4.5 MÉTODOS DE PREDIÇÃO DE ENERGIA LIVRE DE LIGAÇÃO <i>IN SILICO</i> .....	42
4.5.1 MM/GBSA e MM/PBSA.....	43
4.5.1.2 (QM)MM/GBSA.....	44
4.5.2 AutoDock 4.2.6.....	46
4.5.3 LigScore e DrugScore.....	46
4.6 QUÍMICA QUÂNTICA NA DETERMINAÇÃO DE ENERGIA LIVRE.....	47
5. RESULTADOS E DISCUSSÃO.....	50
5.2 ANÁLISE DA DINÂMICA MOLECULAR.....	51
5.2.1 Análise dos ligantes e suas interações com a InhA.....	52
5.3 RESULTADOS MM/GBSA.....	55
5.4.1 Tempo de simulação em MM/GBSA.....	55
5.4.2.1 Campos de Força com raios atômicos otimizados.....	61
5.4.3 Campo de força ff14SB com múltiplas trajetórias.....	66
5.4.4 Campo de força ff14SB com entropia baseada em Modos Normais.....	68

5.4.5 (QM)MM/GBSA.....	70
5.4.5.1 Ligante considerado na região de QM.....	70
5.5 RESULTADOS MM/PBSA.....	77
5.5.1 Tempo de simulação em MM/PBSA.....	77
5.6 RESULTADOS AUTODOCK.....	80
5.7 RESULTADOS LIGSCORE.....	82
5.8 RESULTADOS DRUGSCORE.....	85
5.9 RESULTADOS SQM.....	86
6. COMPARAÇÃO ENTRE METODOLOGIAS.....	89
7. CONSIDERAÇÕES FINAIS.....	92
9. ANEXOS.....	99

# 1. INTRODUÇÃO

## 1.1 TUBERCULOSE HUMANA

A tuberculose (TB) humana é umas das doenças infecciosas mais antigas descritas pela comunidade médico científica. Dados alarmantes da Organização Mundial de Saúde (OMS), caracterizam a tuberculose como sendo a segunda doença infectocontagiosa mais prevalente mundialmente, com cerca de um terço da população global colonizada. Dentre os indivíduos infectados, aproximadamente 8,6 milhões deles desenvolvem a TB ativa todos os anos, e em média 1,3 milhão de pessoas morrem anualmente por causa desta doença (OMS 2015).

Uma patologia que acomete pessoas principalmente na faixa etária entre 20 e 49 anos, a TB pode apresentar-se sobre diferentes manifestações clínicas, que são normalmente associadas ao órgão acometido. As manifestações da forma pulmonar contudo são as mais frequentes, visto que o *Mycobacterium tuberculosis* tem uma predileção pelo tecido pulmonar. A forma pulmonar tem como principal manifestação clínica a tosse seca ou produtiva com expectoração purulenta, com presença ou não de sangue (hemoptise). Formas extrapulmonares ocorrem principalmente em pacientes imunocomprometidos devido a casos associados ao HIV ou tratamentos quimioterápicos. (San Pedro et al., 2013; Costa et al., 2012).

Com relação ao tratamento efetivo da TB, é necessário uma associação de diagnóstico precoce com o uso medicamentoso adequado. O tratamento preconizado pela OMS é a base de potentes antibióticos (rifampicina, etambutol, isoniazida e pirazinamida) por um tempo médio de seis meses (OMS 2014). O uso indiscriminado de antibióticos, associado a uma lacuna importante na descoberta de novos antimicrobianos a partir da década de 80, contribuíram para o crescimento e o ressurgimento da tuberculose nos últimos tempos. Este fato se agrava principalmente nos países subdesenvolvidos e em desenvolvimento, como é o caso do Brasil

Por ser o maior responsável pelos casos de tuberculose, o Mtb sempre foi o alvo principal do estudo para o desenvolvimento de drogas anti-TB. É um bacilo patogênico da família *Mycobacteriaceae* descoberto em 1882 por Roberto Koch e comumente chamado de bacilo de Koch (BK), em homenagem ao seu descobridor. Isolada no ano de 1905, a cepa H37Rv teve seu genoma sequenciado somente em 1998. O seu genoma completo possui 4.411.529 pares de bases e aproximadamente 4.000 genes (Cole et al., 1998; Camus et al., 2002). Na natureza, depende de parasitismo para sobreviver, portanto ele não se encontra em sua forma livre. As micobactérias podem entrar em estado de dormência, onde são capazes de sobreviver sem se dividir. Esta

característica permite que elas continuem a colonizar o ser humano por longos períodos, dificultando a sua erradicação. Poucos microrganismos produzem uma variedade de moléculas lipofílicas como o Mtb. Este patógeno é capaz de produzir desde moléculas simples, como palmitato, até ácidos graxos de cadeias longas (ácidos micólicos). Possui em média 250 enzimas distintas envolvidas no metabolismo dos ácidos graxos, contra apenas 50 na *Escherichia coli*, por exemplo (Glickman et al., 2001).

No estudo de novos fármacos contra a TB provocada pelo *Mycobacterium tuberculosis*, destaca-se a enzima InhA (2-*trans*-enoil-ACP (CoA) Redutase) (E.C 1.3.1.9). A InhA é um conhecido alvo molecular, local de ação do antimicrobiano isoniazida. Por se caracterizar em um alvo de interesse já estabelecido e validado para o desenvolvimento de drogas anti-TB, a InhA ainda se mostra promissora neste tipo de estudos contra novos alvos moleculares (Agüero et al., 2008).

## 1.2 A ENZIMA InhA DE *Mycobacterium tuberculosis*

A enzima InhA, é uma das enzimas participantes da via de síntese de ácidos graxos, a FAS II (Fatty Acid Synthase). Em micro-organismos, incluindo o Mtb, a FAS divide-se em dois subtipos: FAS I e FAS II, sendo que a FAS II não se encontra em mamíferos (Quemard et al., 1995). Esta exclusividade faz com que as enzimas desta via sejam potenciais alvos para novos inibidores, pois não interferem com a homeostase do corpo humano.

Pertencente a família de enzimas SDR's ou desidrogenases/redutases de cadeia curta, a InhA é codificada pelo gene *inhA* e possui 269 aminoácidos em cada subunidade. Essa enzima utiliza substratos derivados 2-*trans*-enoil-ACP de ácidos graxos de cadeias curtas e longas, sendo esta uma reação NADH-dependente (utiliza NADH como coenzima). Nesta etapa um íon hidreto (H<sup>-</sup>) é transferido da coenzima para o substrato, o que resulta na redução da dupla ligação (entre C2 e C3) do ácido graxo (Figura 1). A InhA utiliza principalmente ácidos graxos de cadeia longa, entre 12-24 carbonos, como sendo precursores dos ácidos micólicos. A sua inativação, sozinha, é suficiente para induzir o acúmulo de ácidos graxos saturados e este processo leva a alterações e lise da parede celular do Mtb (Vilchêze et al., 2000).

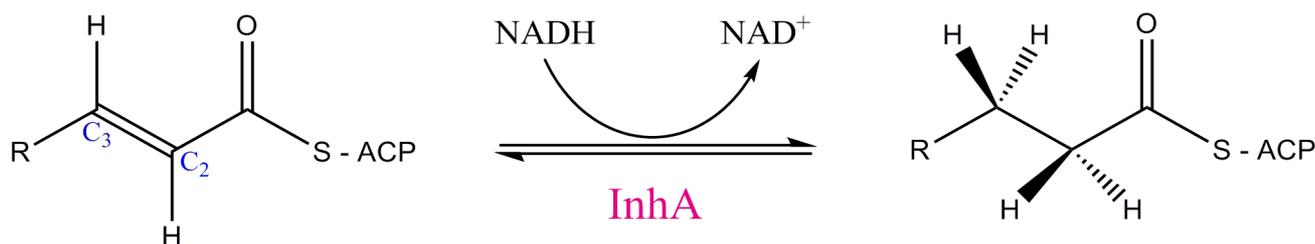


Figura 1: Reação de redução de um tioéster  $\alpha,\beta$ -insaturado catalisada pela enzima InhA (Mills, 2006).

### 1.3 ÁCIDOS MICÓLICOS

A formação da parede celular do *Mycobacterium tuberculosis* é composta por várias moléculas complexas. Predominantemente podemos citar três delas atreladas por ligações covalentes: o peptidoglicano, o arabinogalactano e os ácidos micólicos (sendo este o componente majoritário da parede celular). Os ácidos micólicos caracterizam-se por serem ácidos graxos  $\alpha$ -alquil  $\beta$ -hidroxilados com um alto peso molecular. São compostos de duas cadeias, uma menor com cerca de 20-24 átomos de carbono e uma cadeia maior com 50-60 átomos de carbono (Barry et al., 1998). No sistema FAS I todas as atividades enzimáticas são codificadas por um ou dois polipeptídeos e são responsáveis pela produção de ácidos graxos de cadeia pequena ( $C_{16}$ - $C_{18}$ ) e de cadeia longa ( $C_{24}$ - $C_{26}$ ). A via FAS II é composta por quatro enzimas monofuncionais distintas que são responsáveis pelo alongamento da cadeia de ácidos graxos no Mtb. Primeiro ocorre a condensação de uma unidade de malonil-AcpM à cadeia do ácido graxo, podendo esta reação ser catalisada por duas  $\beta$ -cetoacil-ACP sintases, a KasA e a KasB (Schaeffer et al., 2001), resultando na liberação do AcpM e  $CO_2$ . Desta ligação resulta um  $\beta$ -cetoéster, e este grupo ceto é reduzido pela enzima MabA ( $\beta$ -cetoacil-ACP redutase NADPH-dependente), a um  $\beta$ -hidroxiacil-ACP. Na terceira etapa ocorre uma eliminação do grupo  $\beta$ -OH, por uma enzima hidroxiacil-ACP desidratase, resultando na formação de um derivado 2-*trans*-enoil-AcpM. A etapa final desta reação consiste na redução da ligação dupla do enoil-AcpM pela enzima InhA (Rosado et al., 2012; Kim et al., 2013). Uma representação esquemática das vias FAS I e FAS II pode ser observada na Figura 2 da página 18.

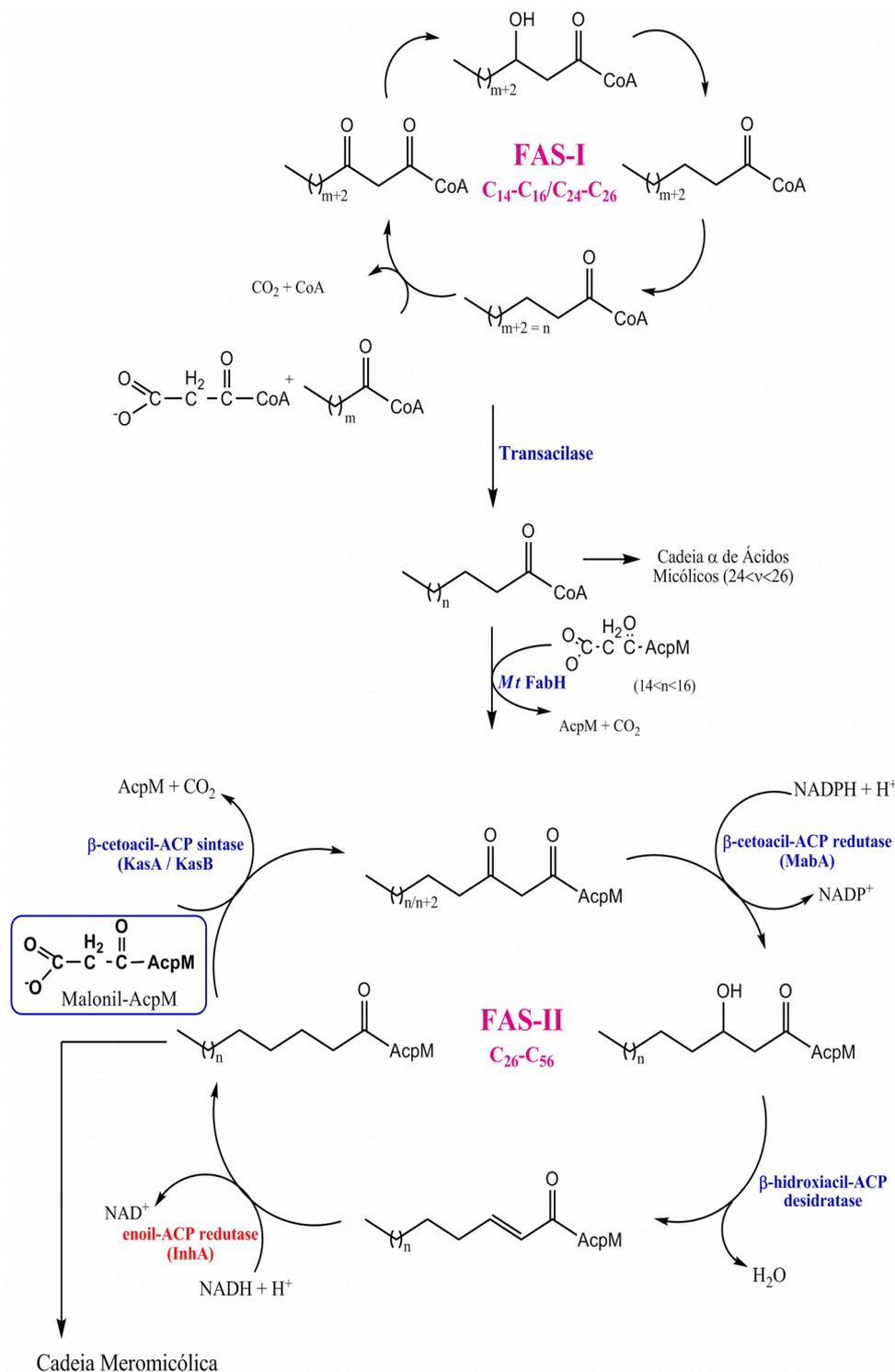


Figura 2: Biossíntese de ácidos graxos em *Mtb*. FAS I sintetiza ácidos graxos numa distribuição bimodal C<sub>16-18</sub> e C<sub>24-28</sub>, que são removidos do complexo enzimático por uma transacetilase na forma de acil-CoA de cadeia longa. A enzima FabH, com especificidade pelos produtos de cadeia menor do sistema FASI, catalisa a condensação desses precursores acil-CoA com o malonil-ACP. O produto é alongado pela FAS II, resultando em cadeias carbônicas longas, precursoras da cadeia meromícólica (C<sub>50-60</sub>) dos ácidos micólicos. Os produtos mais longos do sistema FASI (C<sub>24-26</sub>) são utilizados (provavelmente ligados à CoA) como substrato na formação da cadeia  $\alpha$  dos ácidos micólicos (Figura adaptada de Schroeder et al., 2005).

## 1.4 A ESTRUTURA DA InhA

Existem hoje no Banco de Dados de Proteínas (PDB) até o presente momento 68 estruturas de InhA de Mtb resolvidas por cristalografia de proteínas. Dentre estas, se encontram estruturas tanto em sua forma selvagem (do inglês *wild type*) como é o caso da estrutura com o código de entrada 1ENY (Dessen et al., 1995), como também em suas formas mutantes, como é o caso da 2AQQ (Oliveira et al., 2006). A InhA caracteriza-se por apresentar uma topologia onde cada subunidade se comporta como um domínio único, na topologia do tipo Rossmann *fold*, e em palavras esquemáticas, a sua estrutura se parece com uma “cadeira” composta por oito  $\alpha$  hélices e sete folhas  $\beta$  (Dessen et al., 1995) (Figura 3).

A coenzima se liga entre o espaço composto pelo “encosto” da cadeira e o “assento”, e o sítio de ligação localiza-se em uma cavidade em cima do “encosto” da cadeira (Dessen et al., 1995). A unidade biológica da InhA apresenta-se na forma de um homotetrâmero (Figura 4), fato este confirmado pela resolução das primeiras estruturas da enzima e por análises de cromatografia de exclusão por tamanho.

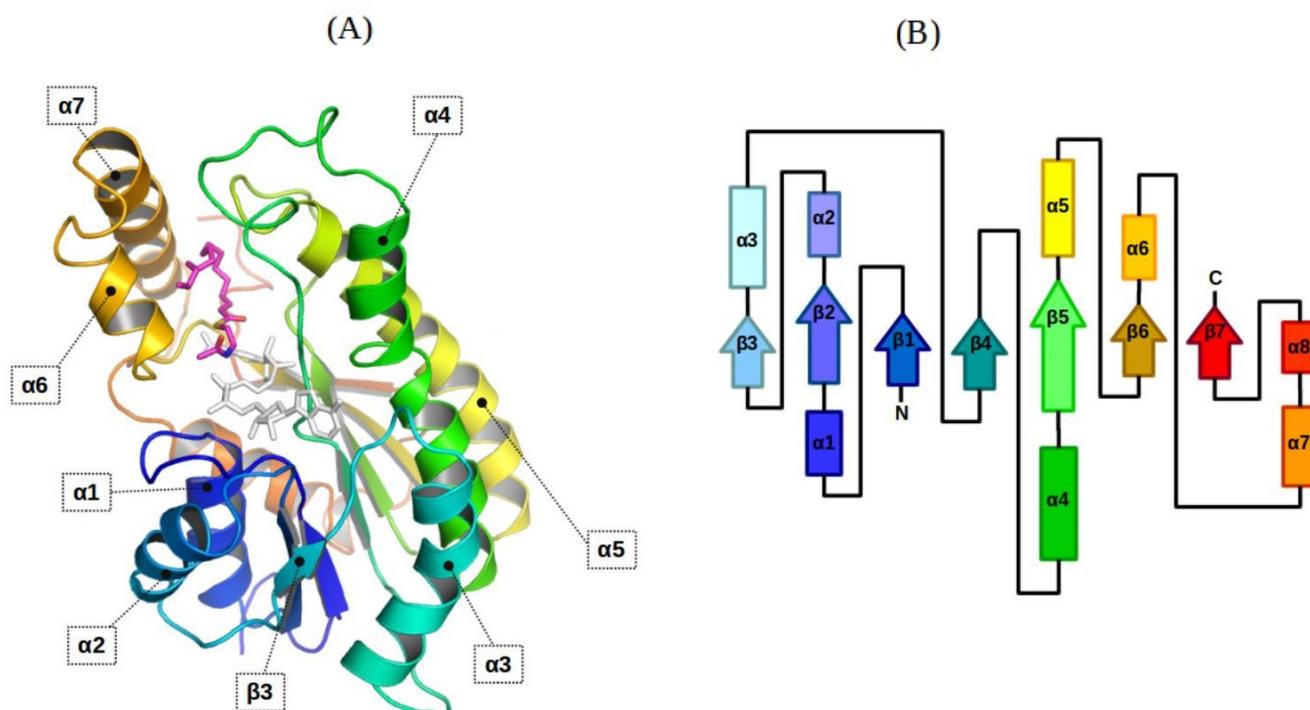


Figura 3: (A) Estrutura tridimensional da subunidade C da enzima InhA complexada como NADH (cinza) e o análogo C16 do substrato (lilás) código PDB: 1BVR (Rozwarski et al., 1999) . A figura mostra (B) a topologia que esta estrutura assume em uma escala de cores que ajuda a identificar a ordem em que ocorrem os elementos de estrutura secundária. Cada subunidade consiste em um domínio único com o núcleo central do tipo Rossmann fold onde se encontra o sítio de ligação de NADH.

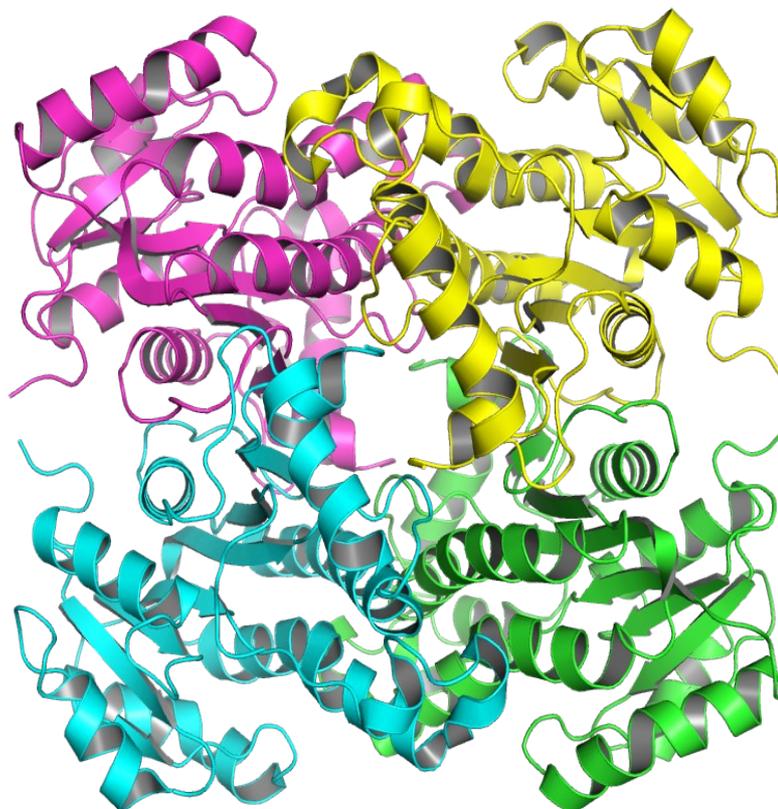


Figura 4: Representação do tipo ribbons do homotetrâmero da InhA, colorido por subunidades: A (verde); B (ciano); C (amarelo); D (lilás). Código de entrada PDB: 3FNE (Freundlich et al., 2009).

### 1.5 SÍTIO DE LIGAÇÃO AO SUBSTRATO DA InhA

Comparando a InhA com outras enzimas da classe das enoil redutases (ENRs), sua alça de ligação ao substrato possui um tamanho consideravelmente maior, o que está de acordo com a especificidade da ENR de Mtb de reduzir substratos maiores (Rozwarski et al., 1999). Este motivo estrutural é composto por duas  $\alpha$  hélices transversais ( $\alpha 6$  e  $\alpha 7$ ) conectadas por uma pequena alça sem estrutura secundária definida (Figura 5). Do lado oposto encontram-se duas grandes alças que compreendem a cavidade para a ligação do substrato no sítio ativo, conhecidas como alças A e B. Em sua forma biológica a enzima é um homotetrâmero, com sítios ativos separados por grandes distâncias (aproximadamente 40 Å) e sem evidências que ocorram efeitos de cooperatividade entre as partes. Por esta razão a unidade monomérica é utilizada como objeto de estudo para simulações de docagem dos ligantes e dinâmica molecular (Cohen et al., 2011).

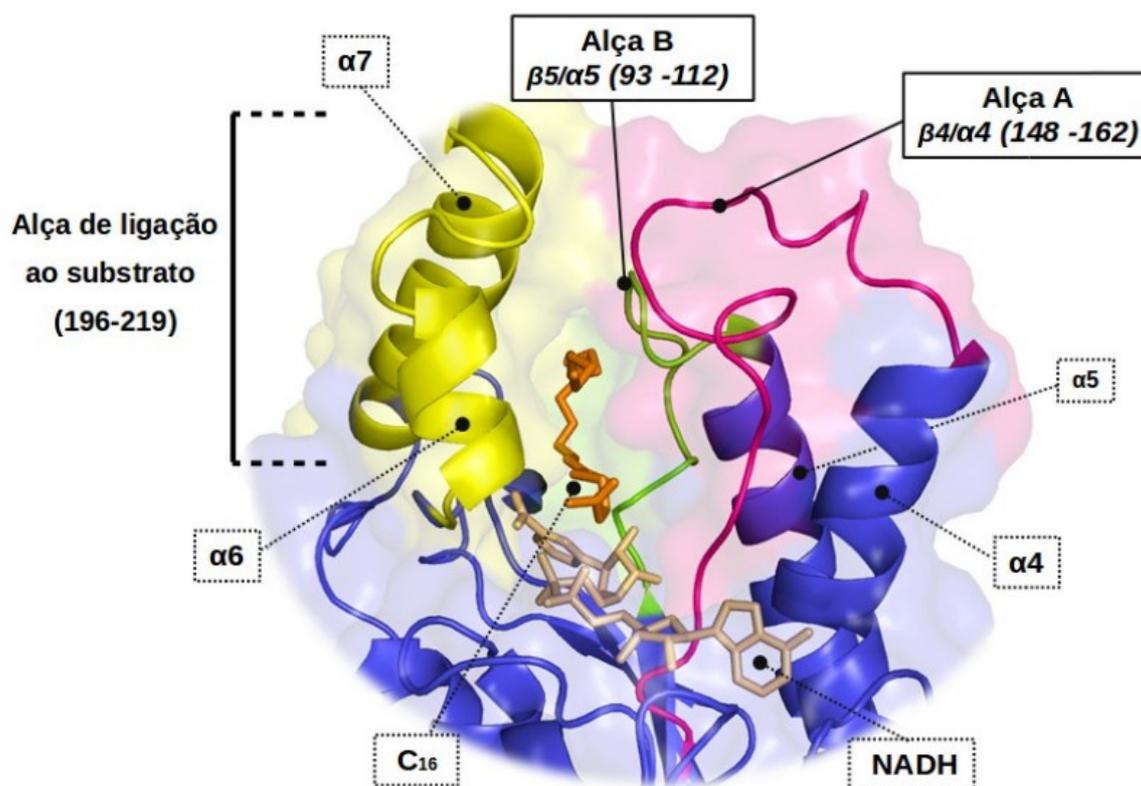


Figura 5: Representação do tipo ribbons da cadeia principal da subunidade C destacando o sítio ativo da enzima, código PDB :1BVR (Rozwarski et al., 1999). Alças A e B mostradas em roxo e verde, o NADH está na cor salmão, e o análogo do substrato na cor laranja. Os valores entre parênteses indicam o intervalo de resíduos que formam as estruturas. Adaptado de (adaptado de Pauli et al., 2011).

O sítio de ligação ao substrato de cada subunidade compreende um formato oval com dimensões de aproximadamente  $16 \text{ \AA} \times 13 \text{ \AA} \times 7 \text{ \AA}$ , compatível com os substratos volumosos que essa enzima utiliza. Considerando a entrada de ácidos graxos acima de 16 carbonos, por exemplo, Rozwarski e colaboradores propõem que as hélices  $\alpha 6$  e  $\alpha 7$  sofrem um movimento de deslocação, abrindo um espaço para a entrada destes substratos. Isso ocorre devido à mobilidade desta parte da proteína, que é confirmada pelas estruturas cristalográficas resolvidas, onde os fatores de temperatura nesta região são mais elevados que nas demais porções da enzima. Uma sobreposição realizada com base em 51 estruturas de InhA de Mtb resolvidas com diversos ligantes mostra as flutuações na cadeia principal, especialmente nas regiões das hélices  $\alpha 6$  e  $\alpha 7$  (**Figura 6**). Essa movimentação é favorecida devido ao fato da alça de ligação ao substrato estar na parte externa do tetrâmero, não entrando em contato com as outras subunidades.

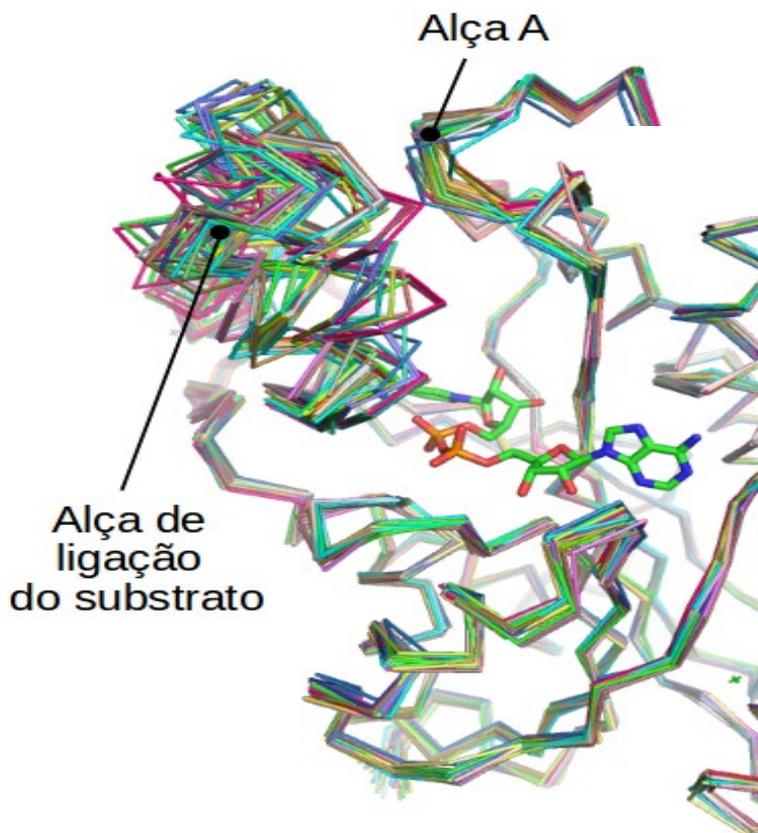


Figura 6: Sobreposição de cinquenta e uma estruturas de InhA de Mtb (2014), demonstrando flutuações principalmente no que diz respeito a alça de ligação do substrato.

A primeira estrutura da InhA resolvida e depositada no banco de dados de proteínas, código PDB: 1ENY (Dessen, et al., 1995) permitiu compreender melhor o complexo formado entre a enzima, e seus ligantes. A resolução da estrutura de InhA complexada com a NAD-acetil-isonicotínico (Rozwarski et al., 1998) (código PDB: 1ZID) contribuiu para o entendimento do mecanismo de ação da isoniazida em Mtb. Mecanismo este que envolve um ataque ao carbono da posição 4 do anel nicotinamida do NAD, resultando na formação de um aduto acilpiridina/NAD (INH-NAD). Esta estrutura provem base para o desenvolvimento de novos agentes anti-TB, inibindo a ação da InhA sem a necessidade da ativação pela KatG (caso da isoniazida) (Pasqualotto et al., 2003). A Figura 7 mostra a conversão do NAD em INH-NAD e como essas duas estruturas se ligam no sítio de ligação de NAD da InhA.

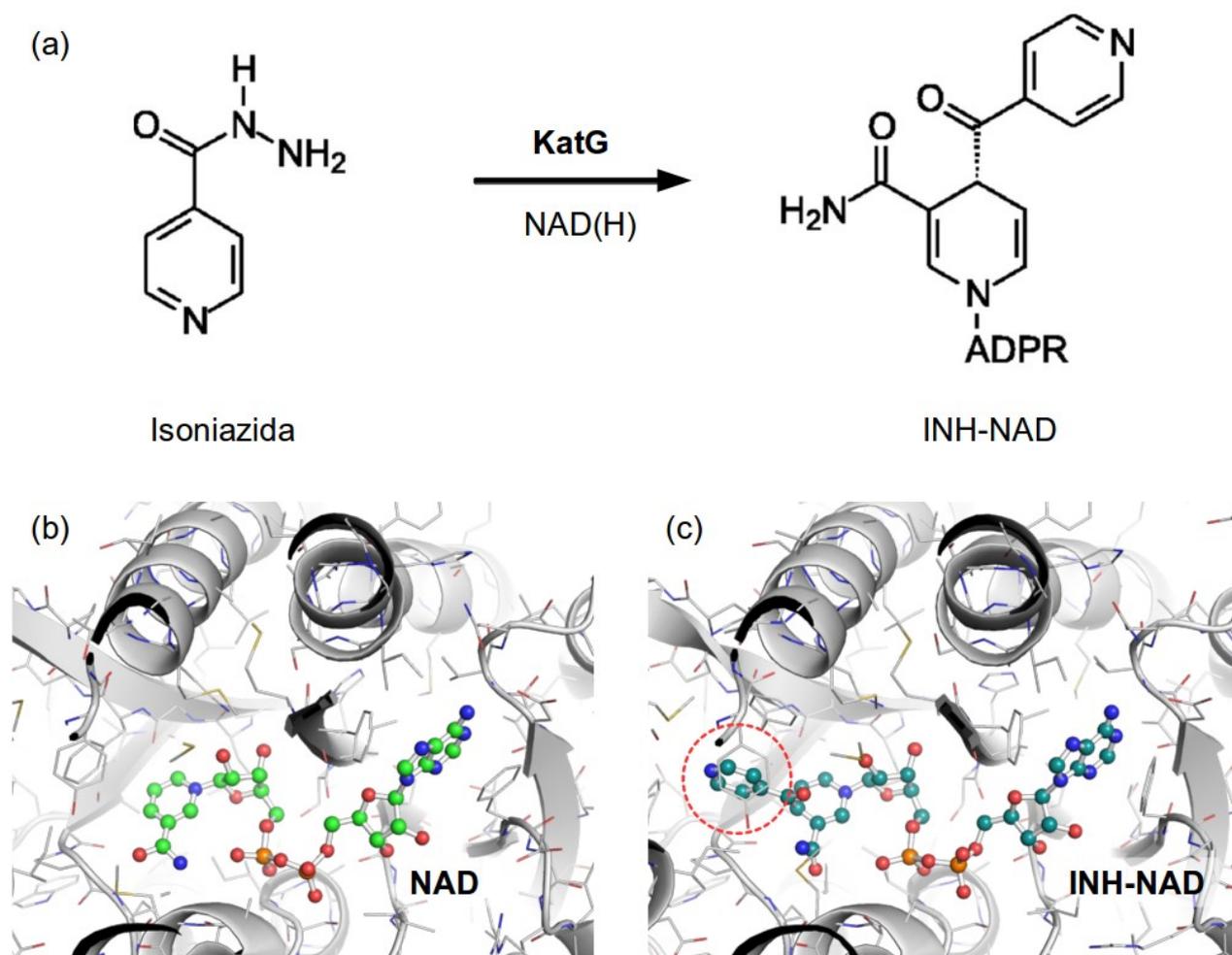


Figura 7: (a) conversão do NAD em INH-NAD catalisado pela enzima KatG, adaptado de Rawar et al., 2003; (b) estrutura da molécula de NAD ligada a InhA, código PDB: 1ENY (Dessen et al., 1995); (c) estrutura da molécula de INH-NAD ligada a InhA código PDB: 1ZID (Rozwarski et al., 1998). A circunferência em vermelho mostra a posição onde o grupo acilpiridina é introduzido na estrutura do NAD formando o aduto.

Com o intuito de se estudar os eventos moleculares e a sua relação com o aparecimento de resistência a fármacos na InhA, Schroeder e colaboradores utilizaram a técnica de dinâmica molecular aplicada a enzima InhA selvagem e seus mutantes INH-resistentes (I21V e I16T). Essas mutações que ocorrem na alça rica em glicina tornam esta região da estrutura mais flexível e diminuem a afinidade do NAD pelo seu sítio de ligação. Como a inibição pelo fármaco isoniazida depende da ligação covalente ao NAD para formar o aduto, a diminuição desta interação contribuiu para o aumento da resistência de *Mtb* frente a este antibiótico (Schroeder et al., 2005).

## 1.6 DESENVOLVIMENTO DE CEPAS MULTIRRESISTENTES

Um dos primeiros fármacos a serem utilizados como antibiótico para tratamento da TB, escolhido como primeira linha desde 1952 é a isoniazida. A isoniazida é um pró-fármaco ativado

oxidativamente pela enzima catalase-peroxidase codificada pelo gene *katG*. O mecanismo de ação desta droga consiste em uma ligação com o  $\text{NAD}^+$ , criando um inibidor competitivo que se liga a InhA com uma constante de inibição ( $K_i$ ) de 0,75 nM (Rawat et al., 2003). Além da isoniazida, outro grupo de fármacos anti-TB, utiliza a inibição da InhA como mecanismo de ação para o potencial antibiótico. Este grupo é conhecido como as tioamidas, representadas pela etionamida e protionamida. São pró-fármacos que utilizam a mono-oxigenase dependente de flavina (codificada pelo gene *ethA*) para formarem adutos com o NAD. Os motivos da indução de resistência destes antibióticos tem relação com os genes ativadores das enzimas *ethA* e *katG*, e não com a InhA diretamente.

Como inibidor direto da enzima InhA (sem a necessidade de genes ativadores) temos o triclosano (McMurry et al., 1998). Essa molécula possui uma biodisponibilidade baixa, e causa uma inibição relativamente fraca ( $K_i = 0,2 \mu\text{M}$ ), por isso não é utilizada para os fins de tratamento da tuberculose atualmente. Apesar disso, esta molécula tem sido utilizada como um composto líder no planejamento de fármacos com base na estrutura (McMurry et al., 1998) e foi testado neste presente estudo, frente a enzima InhA.

O tratamento padrão de “curta duração” DOTS (*Directly Observed Treatment Short Course*) para tuberculose consiste em uma associação de medicamentos. Nos dois primeiros meses se utiliza, concomitantemente, isoniazida, rifampicina, pirazinamida e estreptomicina (ou etambutol). Ao fim deste período, por pelo menos, quatro meses adicionais, é utilizada a combinação de isoniazida e rifampicina (Ducati et al., 2006). Devido a esta associação de medicamentos e o longo período necessário (seis meses) muitos pacientes abandonam o tratamento, principalmente após o desaparecimento dos sintomas. Este fato contribui para o aumento das cepas resistentes a estes fármacos, pois o tratamento incompleto, permite que algumas colônias de bactérias permaneçam viáveis no organismo.

Cepas de TB multirresistentes a fármacos (MDR-TB) ocorrem quando uma cepa de *Mtb* é resistente a isoniazida e rifampicina. Em 2007 havia-se estimado a existência de meio milhão de casos de MDR-TB. No ano de 2008, 55 países haviam reportado casos de cepas de TB extensivamente resistentes a fármacos (XDR-TB). Cepas XDR são resistentes aos antibióticos isoniazida, rifampicina e também as fluorquinolonas, e a pelo menos um dos três fármacos injetáveis de segunda linha, utilizados nos casos de MDR, como a capreomicina, a canamicina e a amicacina (Jain et al., 2008). Foi descrito em literatura casos de cepas apresentando resistência total (TDR-TB) *in vitro*, tanto aos fármacos de primeira linha, como também aos de segunda linha

(Velayati et al., 2009), o que corrobora as ações de promoção de saúde, controle de infecções hospitalares e desenvolvimento de novos fármacos contra doenças infectocontagiosas (Jain et al., 2008).

## 1.7 DESENHO RACIONAL DE FÁRMACOS

Estima-se que a indústria farmacêutica possui, em média, uma taxa de 14 anos desde a identificação até a aprovação efetiva de um fármaco. Os custos associados a identificação de novos candidatos a fármacos é muito elevado e normalmente a maior parte deste investimento advém do período de desenvolvimento (Kapetanovic et al., 2008).

Em contraste com os métodos de desenvolvimento de fármacos tradicionais, que são baseados em testes de tentativa e erro para ensaios de cultura de célula e modelo animal, o planejamento racional de fármacos começa com a hipótese de modelar moléculas a partir de um alvo específico. A escolha deste alvo depende majoritariamente de dois fatores. Primeiro o alvo precisa ser estabelecido e validado. O segundo está relacionado a capacidade do alvo de se ligar a pequenas moléculas (*drugable target*). Através de alvos já conhecidos como é o caso da InhA, propõe-se o planejamento de moléculas que se “encaixem” neste alvo, levando-se em consideração propriedades estruturais, físico-químicas e farmacológicas. Pelo fato de testes *in silico* serem mais baratos e rápidos quando comparados a testes *in vitro*, inúmeras moléculas podem ser testadas para um mesmo alvo, podendo passar da casa dos milhões de compostos analisados. Essa mesma abordagem para ensaios *in vitro* requer alto custo e tecnologias automatizadas de altíssima performance (*high throughput*) (Lechartier et al., 2014). Portanto, para o planejamento racional de fármacos o conhecimento da estrutura tridimensional das proteínas em questão é de grande importância (Leach et al., 2007).

### 1.7.1 Desenvolvimento de novos fármacos com abordagens *in silico*

Computadores são ferramentas essenciais na química medicinal moderna e são importantes para ambos; descoberta de moléculas bioativas/fármacos e desenvolvimento de fármacos. O acelerado avanço no desenvolvimento de computadores mais eficientes e softwares com maior acessibilidade significaram que, muitas operações as quais eram realizadas somente por grupos de pesquisa especializados (com infraestrutura diferenciada e pesquisadores altamente especializados em métodos computacionais), agora podem ser conduzidos em computadores pessoais. O descobrimento de novas drogas é ainda um desafio para a pesquisa farmacêutica. O planejamento de fármacos baseado em estrutura representa um papel significativamente importante nesta tarefa e

agora é parte fundamental da química medicinal. Foi determinado que para vários alvos a estrutura 3D pode ser utilizada para o planejamento de pequenas moléculas capazes de ligar-se fortemente à proteínas.

Um dos objetivos das abordagens *in silico* é a de desenvolver e aplicar métodos capazes de prever as interações entre proteína ligante. Uma maneira elegante de descrever/determinar o comportamento desses sistemas moleculares é utilizar métodos de química/física quântica para descrever os potenciais de interação entre os átomos do sistema. Essa é uma abordagem bastante realista, uma vez que permite descrever fenômenos de considerável complexidade envolvendo rearranjo eletrônico, como por exemplo, a quebra e formação de ligações químicas. Existem diversas abordagens em química quântica (diferentes níveis de teoria) e sua aplicabilidade depende do tipo de problema a ser descrito, porém, de maneira geral, os métodos quânticos podem ser aplicados somente aos sistemas com um número reduzido de átomos (atualmente estimando em poucas centenas). Quando os sistemas de interesse contém uma grande quantidade de átomos (e muitas vezes uma grande quantidade de átomos é necessária para descrever de forma realista o sistema em estudo) pode-se fazer uso de modelos mais simplificados, como por exemplo, a mecânica molecular (MM) que faz uso dos campos de força clássicos (*force fields* ou *FF*). Os *FF* são utilizados por possuírem uma ampla aplicabilidade no planejamento de fármacos baseado em estrutura, como por exemplo: modelagem comparativa, docagem molecular, dinâmica molecular, simulação de Monte Carlo entre outros. A dinâmica molecular, por exemplo, pode servir para o estudo de diferentes aspectos, como (i) a preparação da proteína antes da docagem, para otimizar a estrutura da proteína e considerar a sua flexibilidade, (ii) auxiliar no processo de refinamento dos complexos docados, (iii) incluir o efeito de solventes e considerar o encaixe induzido (*induced fit*), (iv) calcular a energia livre de ligação proteína-ligante, e (v) fornecer uma classificação de possíveis ligantes (Alonso et al., 2006). As simulações por DM, quando combinadas com informações obtidas por meios de técnicas experimentais, abrem caminho para o entendimento do comportamento termodinâmico e funcional de processos biológicos (Dessen et al., 1995).

### **1.7.2 Predição de energia livre de ligação por métodos *in silico***

Dentro do vasto campo da bioinformática e das várias técnicas disponíveis *in silico* para o desenho de fármacos, temos a determinação de energia livre. A medida de energia livre de ligação é a energia necessária para “separar” dois compostos em ligação. Simbolizada pelo  $\Delta G$ , consiste em uma métrica para avaliar a ligação entre duas moléculas, sendo elas macro ou pequenas moléculas (medidas em kcal/mol) (Kitamura et al., 2014). Os valores de  $\Delta G$  são inversamente proporcionais,

por este motivo quanto menor o seu valor (de preferência um valor negativo), mais afinidade um composto apresenta pelo seu par de ligação, e mais difícil de ocorrer a dissociação voluntária dos mesmos. Um sistema associado normalmente apresenta uma energia menor, do que a soma das duas partes, mantendo assim o complexo unido.

O principal objetivo da predição de energia livre de ligação *in silico* no que tange o desenvolvimento de fármacos e testes com novas moléculas, advêm dos potenciais ganhos em relação ao tempo e os recursos financeiros investidos. Um robusto método de determinação, em teoria, tem a capacidade de prever dentre um grupo de ligantes quais seriam os que fortemente se ligariam a um alvo específico e ranqueá-los conforme afinidade, para posteriores métodos *in vitro* (Ucisik et al., 2013). Este tipo de abordagem permite que não se pesquise compostos “as cegas”, e sim de uma forma racional. Associado ao *Drug Design*, a determinação de energia livre permite prever os valores de energia livre quando modificados radicais ou átomos em um determinado composto de interesse (composto “mãe”). Deste modo, é possível avaliar a contribuição positiva/negativa na interação dos grupos adicionados ou removidos.

Existem variados métodos de prever a energia livre de ligação *in silico* para avaliar a interação entre o complexo proteína-ligante. Atualmente dispomos de ferramentas que variam desde simples funções *scoring* via web (LigScore) até refinados e conseqüentemente custosos métodos computacionais (QM/MM-GBSA) (Genheden et al., 2015).

No presente trabalho foram utilizados sete diferentes abordagens de determinação de energia livre de ligação *in silico* sem calibração prévia. Foram avaliados compostos com conhecida atividade frente a enzima InhA de Mtb. Esta abordagem permitiu atribuir uma correlação entre os valores preditos e os valores experimentais dos ligantes, avaliando a acurácia dos métodos e parâmetros propostos neste estudo.

## 2. OBJETIVOS GERAIS

Predizer e comparar computacionalmente os valores de energia livre de ligação entre a enzima InhA (2-*trans*-enoil-ACP (CoA) redutase) (EC 1.3.1.9) de *Mycobacterium tuberculosis* e seus ligantes.

### 2.1 OBJETIVOS ESPECÍFICOS

- Auxiliar na compreensão do processo de interação proteína-ligante, com base nas simulações de dinâmica molecular para a proteína InhA de Mtb e seus complexos com 14 compostos.
- Estimar a energia livre de ligação desses complexos por meio dos seguintes parâmetros e métodos: (i) campos de força, (ii) empíricos, (iii) função *scoring* baseada em conhecimento (iv) métodos híbridos e (v) métodos de química quântica.
- Ranquear os compostos conforme os valores de energia livre de ligação *in silico* e compará-los com os dados experimentais pela técnica de regressão linear.
- Comparar os parâmetros e metodologias empregadas.

### 3. JUSTIFICATIVA

Podemos dividir em duas partes as motivações que norteiam este trabalho: os avanços científicos e tecnológicos da descoberta de novos fármacos no Brasil, e o aspecto social, referente as medidas de saúde pública e busca de novos tratamentos de doenças infectocontagiosas que acometem a população de uma forma geral.

A tuberculose foi considerada controlada na década de 1980, equivocadamente, especialmente nos países desenvolvidos como Estados Unidos, Japão e Europa ocidental. No ano de 1993, a Organização Mundial da Saúde (OMS) declarou um estado de urgência em relação à tuberculose, chamando atenção de governos, comunidade científica e sociedade civil para redobram seus esforços no controle desta doença. No ano 2000, todos os 189 estados-membros das Nações Unidas se comprometeram e assumiram responsabilidade por um programa de metas chamado “Desenvolvimento do Milênio”, onde se visou o combate a várias doenças infecciosas, entre elas o HIV/AIDS, a malária e a tuberculose (Barreira et al., 2007).

Entre os 22 países responsáveis por 80% do total de casos de TB no mundo, o Brasil ocupa o 19º lugar. Segundo dados do Sistema de Informação de Agravos de Notificação (SINAN) do Ministério da Saúde (MS), a partir do ano de 2009 o Brasil detectou 86% dos seus novos casos. Devido ao tamanho do país e ao potencial de disseminação, a TB situa-se entre as doenças de notificação compulsória no Brasil (Costa et al., 2012).

Podemos citar entre as principais causas desta ressurgência dos casos de tuberculose: (i) o advento e disseminação da AIDS, (ii) o aparecimento de cepas de *Mtb* multirresistentes a drogas, (iii) avanço de técnicas imunossupressoras principalmente em relação ao tratamento de neoplasias e (iv) problemas habitacionais, sanitários e de saúde em países subdesenvolvidos (Anderson et al., 2014).

Além dos problemas citados presenciamos uma lacuna na descoberta de novos fármacos para combater doenças infectocontagiosas que foram negligenciadas principalmente nas décadas de 80 e 90 (Barreira et al., 2007). Com o advento da bioinformática, os esforços estão concentrados na descoberta de novas moléculas e na pesquisa utilizando compostos já consolidados como fármacos, porém, para outras patologias de interesse (Barreira et al., 2015; Harling et al., 2014). Apesar do tempo e recursos necessários para o desenvolvimento de novas drogas, visamos promover uma melhoria no acesso a tratamentos efetivos em saúde, por meio de ferramentas que auxiliem na busca de novas moléculas do tipo fármaco contra a tuberculose.

As técnicas de bioinformática no desenvolvimento de fármacos tem a capacidade tanto de auxiliar de forma significativa no conhecimento das interações entre alvos e moléculas, e também através destes estudos reduzir custos, tempo e uso de animais nas pesquisas. Assim como as grandes empresas farmacêuticas possuem seus departamentos de pesquisa em bioinformática, o Brasil começa a avançar nos últimos anos nesta importante área de estudo por meio de parcerias público-privadas.

## 4. METODOLOGIA

A metodologia utilizada nesta dissertação foi descrita sem a ênfase em modelos e detalhamentos matemáticos, visto que já existe uma vasta bibliografia descrevendo estes métodos (McCammon et al., 1977; Karplus e Petsko, 1990; van Gunsteren e Mark, 1998; Karplus e McCammon, 2002; Adcock e McCammon, 2006). Desta forma, o objetivo foi descrever, para cada técnica em questão as abordagens e parâmetros utilizados neste trabalho para permitir a confiabilidade e reprodutibilidade dos resultados obtidos.

### 4.1 LOCAL DE EXECUÇÃO

As simulações e análises computacionais foram realizadas no LABIO (Laboratório de Bioinformática, Modelagem e Simulação de Biosistemas), e FarmInf (Laboratório de FarmInformática) integrantes do Instituto Nacional de Ciência e Tecnologia em Tuberculose (INCT-TB) da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), localizados na Av. Ipiranga, 6681, prédio 32, sala 602 e prédio 12, Sala 117, respectivamente.

Foi utilizada para este trabalho uma *workstation* composta de um Intel® Core™ i7-3770 CPU @ 3.40 GHz x 4, 2 TB HD e um Intel® Core™ i7-3770 CPU @ 3.40 GHz x 4, 2 TB HD.

### 4.2 O PACOTE DE SIMULAÇÃO AMBER

O AMBER (*Assisted Model Building with Energy Refinement*) foi desenvolvido primeiramente pela equipe de pesquisadores do grupo de Peter Andrew Kollman (Kollman et al., 1995) da Universidade da Califórnia, San Francisco. Quando nos referimos ao software AMBER, não estamos falando apenas de um programa único, mas sim de um conjunto de programas que trabalham de forma combinada e que permitem configurar, executar e analisar simulações por dinâmica molecular. Parte do pacote AMBER, denominado AmberTools, é distribuída de forma independente e gratuita (possuindo código aberto). O AmberTools inclui os programas *tLeap* e *antechamber*, que foram utilizados nesse trabalho para o preparo dos arquivos de entrada (*input files*) necessários para as simulações de dinâmica molecular. Os principais programas utilizados nesse trabalho dentro do pacote AMBER e AmberTools foram:

- *tLeap*: preparo dos arquivos de entrada para as simulações de DM (arquivos contendo topologias/parâmetros *-.top*, e coordenadas-*.crd*). Inclui um editor que permite construir

resíduos e manipular moléculas.

- antechamber: constrói a topologia para pequenas moléculas utilizando os tipos atômicos do GAFF e atribui as cargas parciais aos átomos. O programa gera arquivos que são interpretados no tLeap.
- sqm: calcula as cargas parciais utilizando o método AM1-BCC que são utilizadas pelo antechamber no processo de atribuição de cargas.
- sander: do inglês “Simulated annealing with NMR-derived energy restraints”, é um dos principais programas do pacote AMBER, para executar as simulações de DM.
- PMEMd: é uma evolução do programa sander, com performance otimizada para cálculos executados em GPUs (tecnologia que o grupo LABIO tem acesso).
- ptraj / cpptraj: analisa as simulações de DM. Esse programa permite calcular uma série de propriedades importantes, tais como: RMSD, análise de ligações de hidrogênio, comportamento de difusão e muitas outras.
- mmGB/PBSA: é o programa que executa os cálculos de energia livre de ligação proteína ligante com base nas simulações de DM geradas pelo sander/PMEMd.

#### 4.2.1 Programas para visualização e edição de estruturas

A visualização, manipulação de estruturas, geração de imagens e filmes foram realizadas através dos programas: SwissPdbViewer (Guex et al., 1997), PyMol (DeLano et al., 2002), VMD (Humphrey et al., 1996) e Chimera (Pettersen et al., 2004).

#### 4.3 OBTENÇÃO E ESCOLHA DOS LIGANTES

Foram selecionados 14 ligantes para os cálculos de predição de energia livre de ligação ( $\Delta G$ ) por meio de critérios pré estabelecidos: (i) pequenos compostos candidatos a fármacos, (ii) com valores de  $K_i$  calculados experimentalmente (iii) excluindo moléculas que apresentavam apenas valores de IC50, como é o caso do ligante GEQ (PDB:1P44), (iv) sendo os ligantes avaliados até junho de 2015 (PDB).

Destes 14 ligantes, cinco foram retirados do PDB, e tiveram suas estruturas cristalográficas resolvidas com estes compostos, conforme indicado na Tabela 1. A estrutura molecular de cada composto (PDB) esta representada na sequência de figuras 9-12.

Tabela 1: Moléculas ligadas a InhA de Mtb, retiradas do Banco de dados de proteínas (PDB). Período de análise entre julho de 2014 á julho de 2015.

Ligante	Código (PDB)	Referência	Resolução
8PS	2B37	Sullivan et al. 2006	2.6 Å
TCL	1P45	Kuo et al. 2003	2.6 Å
TCU	2X22	Luckner et al. 2010	2.1 Å
5PP	2B36	Sullivan et al. 2006	2.8 Å
3KX	4R9R	Manjunatha et al 2015	2.9 Å

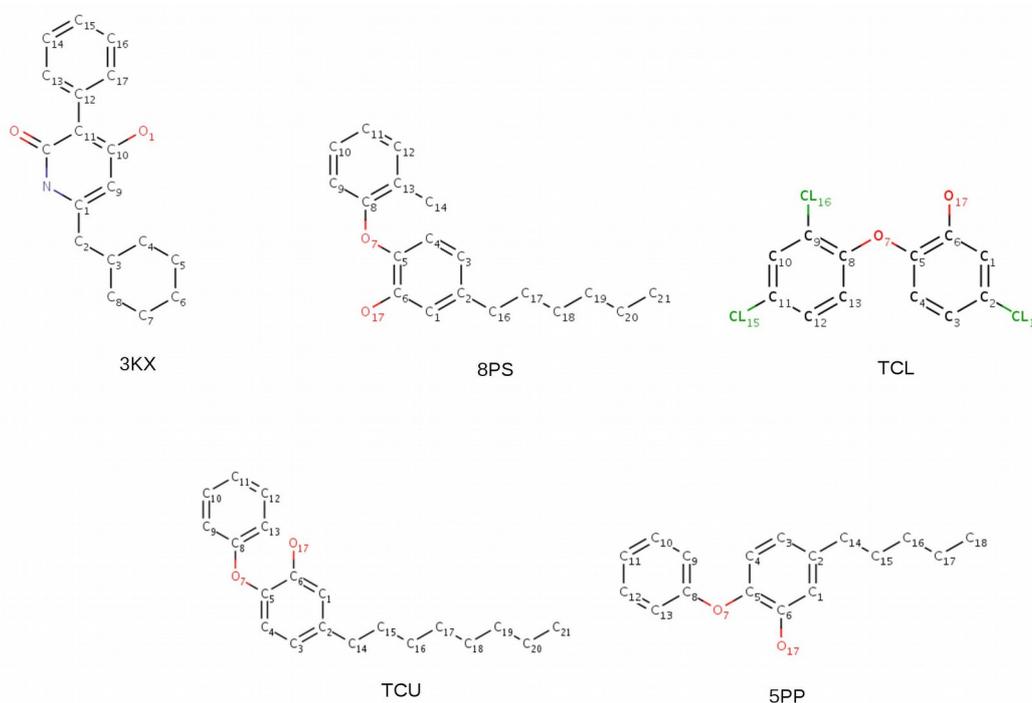


Figura 8: Estrutura molecular dos 5 ligantes retirados do PDB, utilizados neste estudo, frente a enzima InhA de Mtb. A referência bibliográfica e o código PDB de cada um dos ligantes selecionados está relacionado na Tabela 1. Da esquerda para a direita: 3KX, 8PS, TCL, TCU e 5PP. Nesta imagem é possível observar as diferenças moleculares entre as moléculas testadas.

Nove compostos (Tabela 2) compreendem moléculas que sofreram modificações derivadas de um composto original (composto “mãe”), o ligante Genz 10850 (Figura 9). Estas modificações foram propostas por um grupo de pesquisadores da Pontifícia Universidade Católica do Rio Grande do Sul (PUCRS), ligados ao Instituto Nacional de Ciência e Tecnologia em Tuberculose (Rotta et al., 2015).

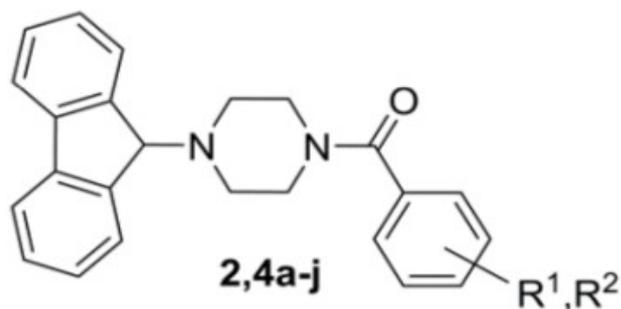


Figura 9: Composto Genz-10850, adaptado de Rotta et al., 2015.

Tabela 2: Modificações propostas para os ligantes derivados do Genz-10850. Adaptado de Rotta et al., 2015.

Ligante	R'	R''
2AX	H	H
4AX	4-O Me	H
4BX	4-CL	H
4CX	4-F	H
4DX	4-NO2	H
4FX	3-ME	H
4HX	3-F	H
4IX	2-F	H
4JX	2-F	5-F

Na Tabela 3 estão expostos os resultados de  $\Delta G$  dos 14 ligantes de InhA de Mtb escolhidos para este estudo, calculados com base nos valores experimentais de  $K_i$ . Para a conversão de  $K_i$  em  $\Delta G$  a fórmula utilizada foi:

$$\ln(K_i) = \Delta G_{\text{ligação}} / RT \quad (\text{eq.1})$$

Os valores resultantes desta conversão, foram os pontos de referência utilizados neste estudo, com o intuito de comparar os métodos computacionais aqui propostos. Partindo destes valores foi possível medir o índice de correlação entre os resultados preditos e os obtidos experimentalmente com o auxílio da técnica de regressão linear.

Tabela 3: Valores de energia livre de ligação calculados com base nos valores de  $K_i$ . Resultados em kcal/mol.

Ligante	$\Delta G_{\text{experimental}}$
2AX	-9.2843
4AX	-7.7832
4BX	-9.1924
4CX	-8.1313
4DX	-8.8052
4FX	-9.1274
4HX	-9.3564
4IX	-8.8333
4JX	-8.0314
5PP	-10.8081
TCL	-6.9012
TCU	-11.0598
8PS	-12.2147
3KX	-8.5369

#### 4.3.1 Montagem dos modelos tridimensionais

A montagem dos sistemas utilizados foi baseada nas estruturas de InhA retiradas do PDB, conforme Tabela 1 para cada composto selecionado (arquivo: proteína.pdb). No caso dos compostos derivados do Genz-10850, foi utilizada a estrutura de InhA cristalográfica retirada do código PDB:1P44 subunidade C. Apesar da InhA se comportar como um tetrâmero biologicamente, suas 4 subunidades se comportam independentemente, e por esta razão a unidade monomérica foi escolhida para realizar as simulações.

O desenho dos ligantes que não se encontravam no PDB, foram concebidos através do programa Avogadro (Hanwell et al., 2012), este sendo uma referência para o desenho de pequenas moléculas no meio científico da bioinformática. Dentro do próprio Avogadro os compostos foram parametrizados utilizando o campo de força MMFF94 (Zhu et al., 2014), também indicado para pequenas moléculas.

#### 4.3.2 Grupos de ligantes

Dos 14 ligantes previamente escolhidos para este estudo, os mesmos foram divididos em três grupos para melhor avaliar e analisar os resultados obtidos:

- Grupo 1 (TCL, 3KX, 5PP, TCU, 8PS): formam este grupo os 5 ligantes retirados do PDB, apresentando uma diferença de energia livre experimental que varia do -6.9 kcal/mol até -12,2 kcal/mol. A diferença entre o composto de pior resultado e o melhor colocado no ranqueamento experimental neste Grupo é de -5.3 Kcal/mol (valores esparsos).
- Grupo 2 (4AX, 4AX, 4BX, 4CX, 4DX, 4FX, 4HX, 4IX, 4JX): este grupo é formado pelos 9 compostos derivados do Genz-10850, devido a suas proximidades estruturais e valores experimentais de energia livre (-7.78 kcal/mol à -9.31 kcal/mol). A diferença entre o pior e o melhor colocado neste Grupo é de -1.53 kcal/mol (valores similares).
- Grupo 3: formado pela soma dos Grupo 1 e Grupo 2. A diferença entre o pior composto e o melhor ranqueado experimentalmente deste Grupo é de -5.3 kcal/mol. Este grupo apresenta a maior complexidade de predição neste estudo, pois os 14 ligantes estão muito próximos entre si em relação aos valores experimentais.

O objetivo desta divisão em grupos foi avaliar a acurácia dos diferentes métodos testados frente a conjuntos de compostos com complexidades de predição variadas. Conforme proposto pelos autores deste estudo, foi considerado uma diferença menor que 1,0 kcal/mol entre os compostos do grupo como de alta complexidade. Comparando-se os grupos testados, o Grupo 1 apresenta uma média complexidade, e os Grupo 2 e 3 alta complexidade.

#### 4.4 AMOSTRAGEM POR DINÂMICA MOLECULAR

A dinâmica molecular (DM) tornou-se uma ferramenta importante e amplamente utilizada por químicos, (bio)físicos e biólogos para modelar detalhes microscópicos do comportamento dinâmico de sistemas diferentes, incluindo gases, líquidos, sólidos, superfícies e aglomerados (Tuckerman et al., 2000). A DM é uma evolução da mecânica molecular e nela as equações clássicas de movimento são resolvidas numericamente. Conhecendo os potenciais que agem sobre os átomos do sistema, é fácil determinar a força, pelo gradiente do potencial:

$$F = -\nabla v \quad (\text{eq.2})$$

Resultando na equação de movimento de Newton:

$$F = m \frac{d^2 x}{dt^2} \quad (\text{eq.3})$$

Para o estudo da dinâmica de um sistema, ou para uma simulação de dinâmica molecular, a equação de movimento deve ser resolvida para cada partícula. A metodologia da DM é fundamentada nos princípios da Mecânica Clássica e por meio de sua simulação se obtém uma visão microscópica do comportamento dinâmico de átomos individuais que constituem um sistema, como uma proteína ou um complexo proteína-ligante. A DM funciona como um microscópio computacional (Phillips et al., 2009). Em 1977, McCammon e colaboradores realizaram a primeira simulação de DM envolvendo proteínas. Esta simulação foi realizada no vácuo e o tempo de simulação foi de  $8,8 \times 10^{-12}$  s (McCammon et al., 1977). Desde então, a técnica de DM foi se aprimorando, deixando os sistemas construídos *in silico* mais realísticos. Este progresso se deve tanto a avanços na área da química, com o melhoramento dos parâmetros dos campos de força, quanto da computação, com o desenvolvimento de máquinas mais robustas, que permitem a realização de simulações mais longas, chegando a  $10^{-9}$  s e  $10^{-6}$  s. O ambiente (solvente) em volta da proteína é incluído de forma explícita juntamente a cofatores, íons, substratos ou ligantes. Por exemplo, para uma proteína dimérica, como a glutathione S-transferase (GST), com 434 resíduos de aminoácidos e 7.346 átomos, incluindo os hidrogênios, foram necessárias 17.098 moléculas de água para a solvatação do sistema de forma realística (Norberto de Souza et al., 1997). A célula principal da simulação continha 59.642 átomos (Norberto de Souza et al., 1999). Essas dimensões correspondem a uma concentração de proteína equivalente a 0.012M, o que é próximo da condição de solução diluída (Tinoco et al., 1998).

Como resultado, uma simulação por DM produz uma sequência de fotos (*snapshots*) da proteína em função do tempo. Esta sequência de fotos constitui o que é chamado de *ensemble* na Mecânica Estatística. Quando o *ensemble* está em equilíbrio, o valor médio de parâmetros termodinâmicos como a pressão, temperatura, volume e calor específico podem ser calculados, assim como a estrutura média da proteína. A DM é empregada em várias áreas, desde o refinamento de estruturas cristalográficas, predição de estruturas proteicas e avaliação da interação ligante receptor.

#### 4.4.1 Dinâmica molecular em GPU

Para o desenvolvimento deste trabalho foram necessárias 150 simulações por DM. Com o intuito de acelerar estas simulações e permitir a realização no prazo estipulado de 2 anos, foram utilizadas GPU's (graphics processing unit ou unidades de processamento gráfico) como parte do estudo. Uma GPU, é uma espécie de microprocessador especializado na parte de gráficos em computadores pessoais, ou videogames. Normalmente é utilizada em placas de vídeo e é

responsável pela execução de cálculos e rotinas que resultam nas imagens exibidas no monitor de vídeo do computador (Pierce et al., 2012). O uso desta tecnologia permitiu por consequência uma extensiva busca para o refinamento dos resultados obtidos.

O pacote AMBER14, sofreu algumas modificações referentes as suas versões anteriores para aprimorar o uso das GPU's em simulações. Entre estas melhorias se pode destacar o suporte do GBSA (utilizado neste trabalho), o aumento de performance de 30% em relação a simulações em única GPU, e o suporte para as últimas gerações de GPU's e CUDA 5.0, 5.5. e 6.0 (Case et al., 2014)

Para realizar a etapa de produção das simulações por meio de dinâmica molecular, o programa Pmemd (pacote AMBER), foi o escolhido devido a sua capacidade tecnológica de realizar processos em paralelo (ganho de performance) e principalmente de utilizar processadores gráficos.

#### 4.4.2 Protocolos de DM

A DM consiste em parte primordial para o desenvolvimento desta pesquisa, pois os *snapshots* utilizados pelos métodos de determinação de energia livre de ligação, são derivados desta etapa. É necessário um cuidado em todas as fases da DM, pois erros acumulados, podem influenciar de maneira significativa nos resultados preditos. Com base nas recomendações do próprio pacote de simulações AMBER14 e seguindo protocolos já estabelecidos e validados pelo grupo de pesquisas LABIO (Norberto de Souza et al., 1999; Schroeder et al., 2005; Schroeder et al., 2010), a DM foi dividida em quatro etapas complementares entre si:

- Preparação dos arquivos de entrada;
- Equilíbrio e minimização;
- Aquecimento do sistema;
- Produção.

##### 4.4.2.1 Preparação dos arquivos de entrada

O preparo dos arquivos de entrada ou *inputs* foi realizada através do programa antechamber (integrante do pacote AmberTools). Nele é possível preparar tanto os ligantes como o receptor (InhA). A parametrização dos ligantes ocorreu por meio do campo de força GAFF (referência para pequenas moléculas), utilizando cargas AM1-BCC, padrão do pacote AMBER (Case et al., 2005). Existem duas formas básicas de solvatação com água quando estudamos biosistemas, uma com solvente implícito (método Generalized-born) e outra com solvente explícito (na presença de água). Para realizar simulações em solvente explícito, foi escolhido o método de *grid* (caixa de água)

entorno do complexo (proteína-ligante). O modelo de água escolhido foi o TIP3P (transferable intermolecular potential 3P) (Mark et al., 2001), com um tamanho de *grid* de 10Å, considerado ótimo para o estudo da interação proteína-ligantes (Figura 10) (Case et al., 2005; Cohen et al., 2011; Genheden et al., 2015).

Em relação às cargas totais do sistema, *grids* que possuem menor ou o maior potencial eletrostático, podem ser estabilizados utilizando contra íons como  $\text{Na}^+$  ou  $\text{Cl}^-$ . O intuito deste trabalho foi o de realizar as simulações, com complexos de carga zero. Para este fim, a carga total do NADH foi considerada -2, pois os dois grupos fosfatos estão desprotonados. A carga da enzima InhA foi considerada -4 devido à diferença entre aminoácidos de caráter ácido e básico. Somando os valores do receptor e do NADH a carga formal do complexo estudado (InhA+NADH) foi de -6. Para neutralizar, foram adicionados seis íons sódio ( $\text{Na}^+$ ), resultando na carga formal esperada igual a zero.

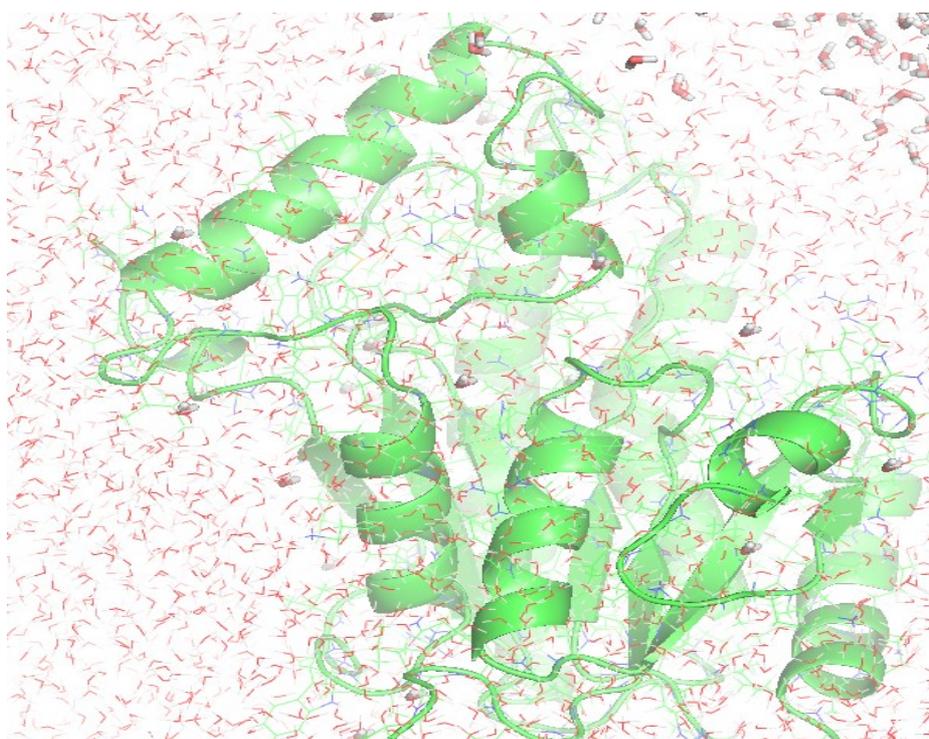


Figura 10: InhA de Mtb código PDB 1P44 envolto pela caixa de água TIP3P (10 Å). Figura gerada pelo programa Pymol (Delano et al., 2005).

Estabelecidos estes parâmetros gerais, foi obtido um sistema total de aproximadamente 36.500 átomos, dependendo do tamanho do ligante testado, pois os ligantes não possuem o mesmo número de átomos. Este sistema é formado pela enzima InhA, o cofator NADH, o ligante escolhido, seis íons sódio, águas cristalográficas (em média de 20-26 dependendo da estrutura cristalográfica

escolhida) e o restante composto de moléculas de água TIP3P.

#### 4.4.2.2 Campos de Força

Considerando os campos de força como um dos pilares para simulações por DM, inúmeros FF's tem sido desenvolvidos para se adequar aos padrões quânticos/experimentais. No pacote AMBER14 vários FF's são disponibilizados, dentre eles ff12SB e ff14SB.

Não existe um consenso até o momento de qual FF apresenta uma melhor performance no estudo da interação proteína-ligante (Xu, et al., 2013). Estes resultados são ainda mais escassos no que se refere ao estudo de ligantes frente a enzima InhA de Mtb. Por esta razão foram testados 8 FF diferentes (ff99, ff99SB, ff99SBLDN, ff99Bnmr, ff98, ff03, ff12SB e ff14SB) para o GBSA e 4 campos de força para o PBSA (ff99, ff99SB, ff03 e ff14SB). No restante das metodologias testadas (LigScore, DrugScore, (QM)MM/GBSA, SQM e AutoDock), utilizou-se o campo de ff14SB. Esta escolha se deve ao fato do ff14SB ser o campo de força mais recente liberado junto ao pacote AMBER14 e por apresentar resultados promissores (Case et al., 2014).

#### 4.4.2.3 Equilíbrio e minimização do sistema

A fase de equilíbrio da caixa de água tem o objetivo de “acomodar” o solvente junto a proteína, evitando a formação de “bolhas” no sistema, permitindo a difusão da água pelo complexo. Com relação aos métodos laboratoriais, poderíamos comparar esta etapa a homogenização da amostra. Para esta etapa foi estabelecido um protocolo de 100 ps de equilíbrio com a proteína fixa.

Depois de equilibrado o solvente, foi realizada uma etapa de minimização de energia do complexo (proteína-ligante). Inicialmente foi estabelecido uma restrição, do inglês *restraint*, de 25 kcal/mol, com redução gradual em etapas de 5 kcal/mol até o complexo estar desimpedido (25 kcal/mol → 20 kcal/mol → 15 kcal/mol → 10 kcal/mol → 5 kcal/mol → 0 kcal/mol).

O *input* utilizado pelo nosso grupo de pesquisa e referência para todas as fases de minimização encontram-se no Anexo 1 deste trabalho.

#### 4.4.2.4 Fase de Aquecimento

O aquecimento gradual do sistema não é sempre necessário. Porém é importante pois permite que se alcance a temperatura desejada em etapas, resultando em uma maior estabilidade do sistema, evitando conformações espúrias (artefatos). Conforme mencionado na etapa de equilíbrio, a rampa de aquecimento também previne a formação de “bolhas” no sistema.

Os testes laboratoriais de determinação de  $K_i$  em enzimas são normalmente realizados a temperatura ambiente (25 °C). Como o intuito deste trabalho é o de comparação entre os dados

experimentais e os dados computacionais preditos, a referência de temperatura adotada foi a experimental de 25 °C ou 298.16 K (o programa AMBER trabalha com a escala Kelvin).

A fase de aquecimento do sistema foi dividida em seis pequenas etapas de 100 ps, com aquecimento gradual até o período de estabilização 298.16 K (temperatura desejada). Depois de alcançada a temperatura desejada, foram realizados mais 400 ps de estabilização, resultando em um total de 1 ns de aquecimento. Aquecimento: (0-50 K → 50-100 K → 100-150 K → 150-200 K → 200-250 K → 250-298.16 K). O *input* utilizado para a fase de aquecimento neste trabalho encontra-se no Anexo 2, ao final deste trabalho.

#### 4.4.2.5 Fase de produção

A fase de produção é a etapa na qual foi realizada a coleta dos dados (*snapshots*) das simulações. Para uma melhor compreensão dos resultados da etapa de produção, os *outputs* (arquivos de saída) foram divididos a cada 2 ns de simulação com cada passo de 5 ps, resultando em 400 *snapshots* de simulação por arquivo de saída. Esta divisão não é primordial, mas permite uma melhor organização dos arquivos de saída e do posterior estudo de qualidade da DM e análise dos dados, como RMSD. A temperatura foi mantida em 298.16 K durante todo processo e tanto a proteína quanto a caixa de água (*grid*) não sofreram nenhum processo de *restraint* (impedimento). O *input* completo da fase de produção encontra-se demonstrado no Anexo 3 ao final deste trabalho, como referência dos parâmetros utilizados na DM em todo processo.

Dependendo do método, em especial o GBSA e PBSA, o tempo de simulação pode afetar os resultados preditos de energia livre de ligação (Xu et al., 2013). Esta informação demonstra que se torna crucial os testes com diferentes tempos de simulação. Para testar o efeito do tempo de simulação, foram realizados estudos prévios variando desde 1 ns até 50 ns o tamanho das simulações. Foi dada ênfase aos valores entre 1 ns e 5 ns, pois conforme Xu e colaboradores, tempos de simulação em torno de 2-4 ns normalmente resultam nos melhores valores preditos de energia livre para os métodos de GBSA e PBSA (Xu et al., 2013).

## 4.5 MÉTODOS DE PREDIÇÃO DE ENERGIA LIVRE DE LIGAÇÃO *IN SILICO*

Existem cinco classes gerais de função *scoring* em química computacional, e estas se diferenciam pelo tipo de abordagem que utilizam para prever a energia livre de ligação no estudo proteína-ligantes (Oda et al., 2006). São elas:

1. Modelos baseados em campos de força (Force Field-based methods): As afinidades são estimadas pela somatória das forças intermoleculares de van der Waals e interações

eletrostáticas entre todos os átomos do complexo usando cálculos de mecânica clássica. Normalmente as interações ocorrem na presença de água, levando em conta a energia de dessolvatação da proteína e do ligante. Entre os métodos de campo de força mais utilizados podemos citar o MM/GBSA e MM/PBSA.

2. Modelos baseados em métodos empíricos (empirical scoring function): Modelo baseado nos vários tipos de interação entre as moléculas. A contagem é baseada no número de átomos do ligante e do receptor em contato, calculando a superfície acessível ao solvente (SASA) do complexo, comparada a proteína e o ligante separados. Dentre os programas conhecidos baseados em métodos empíricos podemos citar o AutoDock e o AutoDock Vina.
3. Função *score* baseada em conhecimento (knowledge-based methods): Baseados em observações entre contatos intermoleculares em larga escala usando bancos de dados 3D, como o PDB. Este método assume que determinados grupos funcionais e interações entre átomos previamente conhecidos serão favoráveis ou desfavoráveis para a afinidade entre ligante e proteína. Existem vários programas que utilizam esta abordagem, podendo citar o LigScore e o DrugScore. Como exemplo destas interações temos:
  - Contato hidrofóbico – hidrofóbico (favorável)
  - Contato hidrofóbico – hidrofílico (desfavorável)
  - Número de ligações de hidrogênio (favorável)
  - Número de ângulos rotáveis imobilizados no complexo (desfavorável)
4. Métodos quânticos: A química quântica descreve o comportamento fundamental da matéria na escala molecular, situando-se na barreira entre a química e a física. Normalmente as teorias de QM são utilizadas para descrever o processo de um composto ativo, por exemplo, reações envolvendo formação/quebra de ligações. Comparada as outras metodologias a QM no estudo dos cálculos de energia livre de ligação é mais recente, porém promissora. Os métodos atuais permitem apenas que algumas centenas de moléculas sejam descritas na parte quântica, devido a limitações dos métodos (Mandal et al., 2009).
5. Métodos híbridos: Combinam os métodos de química clássica com química quântica, como o (QM)MM/GBSA. Neste tipo de abordagem normalmente o receptor é tratado por química clássica e o ligante tratado por química quântica, devido ao número de átomos do ligante ser menor que do receptor (Reddy et al., 2007).

O ponto chave deste estudo, foi realizar testes amplos utilizando as diversas abordagens *in silico* descritas. Para contemplar o objetivo, as cinco classes de predição de energia livre descritas foram utilizadas. Os escolhidos dentro de cada metodologia foram: MM/GBSA e MM/PBSA, AutoDock, LigScore, DrugScore, (QM)MM/GBSA e SQM.

#### 4.5.1 MM/GBSA e MM/PBSA

Os métodos de MM/PBSA (*Molecular Mechanics/Poisson–Boltzmann Surface Area*) e MM/GBSA (*Molecular Mechanics/Generalized Born Surface Area*) são extensamente utilizados para a predição da energia livre de ligação e considerados mais eficientes que métodos rigorosos como o FEP (*Free Energy Perturbation*) e o TI (*Thermodynamic Integration*). A energia livre de ligação para um sistema de proteína-ligante calculada por PBSA/GBSA utiliza as seguintes equações (Xu et al., 2013):

$$\Delta E_{MM} = \Delta E_{interna} + \Delta E_{eletrostática} + \Delta E_{vdw} \quad (\text{eq.4})$$

$$\Delta G_{ligação} = G_{complexo} - (G_{ligante} + G_{receptor}) \quad (\text{eq.5})$$

$$\Delta G_{ligação} = \Delta H - T \Delta S \approx \Delta E_{MM} + \Delta G_{solv} - T \Delta S \quad (\text{eq.6})$$

$$\Delta G_{solv} = \Delta G_{(PB/GB)} + \Delta G_{SA} \quad (\text{eq.7})$$

O  $\Delta G_{ligação}$  representa a energia livre de ligação do receptor e do ligante, o  $\Delta E_{MM}$  é a soma das interações eletrostáticas calculadas em fase gasosa,  $\Delta G_{solv}$  é a energia livre de dessolvatação, o  $T \Delta S$  compreende a entropia conformacional do sistema, o  $\Delta E_{interna}$ , é a soma dos ângulos, ligações e a energia dos diedros,  $\Delta E_{eletrostática}$  compreende a energia eletrostática,  $E_{vdw}$  como sendo o termo da energia de Van der Waals, e  $\Delta G_{solv}$  é a soma dos componentes eletrostáticos ( $\Delta G_{(PB/GB)}$ ) e não eletrostáticos  $\Delta G_{SA}$ . A dessolvatação é calculada por modelos de solvatação implícitos, como o *Poisson–Boltzmann* (PB) e o *Generalized Born* (GB), não sendo fácil distinguir ou apontar qual dos métodos seria mais eficiente, pois estes apresentam performances variadas dependendo do tipo de sistema a ser utilizado (Xu et al., 2013).

Estes dois métodos utilizam como abordagem os campos de força, porém não existe um padrão recomendado para os cálculos de energia livre (Yang et al., 2011; Xu et al., 2013). O MM/PBSA é o

mais custoso computacionalmente dos dois métodos, portanto foram testados quatro campos de força, o ff03, ff99 o ff99SB e o ff14SB. Para o MM/GBSA foram testados os campos de força ff99, ff99Bldn, ff99Bnmr, ff99SB, ff12SB, ff14SB, ff98 e ff03.

No MM/GBSA além dos variados campos de forças testados foram realizadas três abordagens distintas na tentativa de refinar os resultados, utilizando:

1. Raios otimizados: Modelo proposto pelo grupo de Onufriev e colaboradores. Este modelo foi desenvolvido em especial para o estudo de proteínas com pequenas moléculas (ligantes). Foram testados os modelos de raios atômicos *Bondi radii* (GB<sup>OBC</sup>) e a versão otimizada dos raios de Onufriev para estudo de proteínas+ligantes *Bondi radii (mbondi2)* (GB<sup>OBC2</sup>)(Onufriev et al., 2004).
2. Campo de força ff14SB levando em consideração a entropia por modos normais: Os cálculos de GBSA/PBSA consideram a contribuição entropica com base em modelos matemáticos implícitos. Para fins de comparação, no campo de força ff14SB foram realizados testes considerando os valores de entropia por modos normais na determinação de energia livre.
3. Campo de força ff14SB com múltiplas trajetórias: As coordenadas de entrada para os cálculos por GBSA/PBSA são provenientes da fase de produção da DM. Pode-se utilizar apenas as trajetórias do complexo, conhecido como *single trajectory* ou utilizar trajetórias separadas do ligante, da proteína e do complexo, conhecido como métodos de *Multiple Trajectories*. O método de *single trajectory* é o mais utilizado, tem um menor custo computacional e apresenta os melhores resultados (Xu et al., 2013). Neste estudo o método *Multiple Trajectories* e o Ligand/Complex (trajetórias do ligante+ trajetórias do complexo) foram testados para o campo de força ff14SB.

#### 4.5.1.2 (QM)MM/GBSA

O cálculo de energia livre de ligação para ligantes flexíveis requer a avaliação da reprodutibilidade dos parâmetros moleculares dos ligantes. Esta restrição pode ser superada usando uma mistura de mecânica quântica/mecânica molecular (QM/MM), onde o ligante é tratado quanticamente (Gräter et al., 2005). Utilizando QM/MM-GBSA temos a combinação dos métodos de química quântica com os métodos de *Generalized-Born*. Normalmente nos estudos envolvendo QM/MM, as proteínas (receptores) são tratadas por mecânica clássica e os ligantes por mecânica quântica (esta caracterização é apenas uma convenção, lembrando que resíduos da proteína também podem fazer parte da mecânica quântica). Este formato permite as seguintes vantagens: (i) descreve

a maior parte do sistema de forma clássica, permitindo um ganho na velocidade dos cálculos, (ii) não requer um conjunto específico de parâmetros para cada novo ligante, ou seja, a parametrização do ligante e (iii) a inclusão de efeitos quânticos como polarização das ligações (Gräter et al., 2005).

A qualidade dos métodos semi-empíricos depende das aproximações teóricas que lhe deram origem e da amplitude e qualidade da informação usada em sua parametrização. (Leal et al., 2010). Para realizar os cálculos, os mesmos *snapshots* provenientes da DM e o pacote de simulação AMBER foram utilizados. Os modelos semi-empíricos testados foram o AM1, PM3, PM6, e RM1. Além do ligante tratado na região quântica dos cálculos (Gräter et al., 2005), neste trabalho outra abordagem adicional foi testada: tratando o ligante e o NADH (considerado integrante do receptor) na região quântica do cálculo (Figura 11).

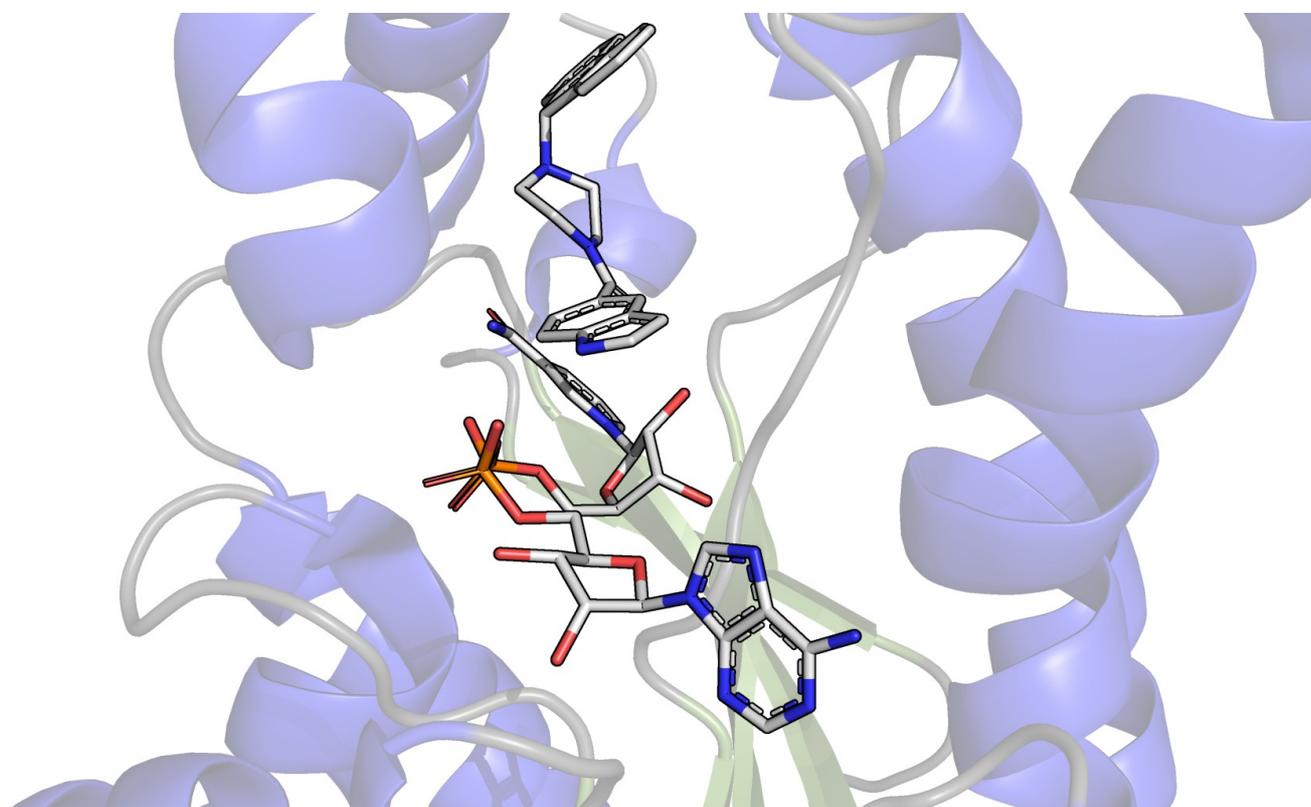


Figura: 11: Ilustração da enzima InhA complexada ao ligante GEQ (PDB ID: 1BVR). Nesta figura esta representado (na forma de *sticks*) o ligante e o NADH, tratados por QM neste estudo.

#### 4.5.2 AutoDock 4.2.6

Desde que foi lançado em 1990, o AutoDock tem provado ser uma ferramenta efetiva, rápida e acurada para predizer as melhores orientações de pequenas moléculas em seus receptores. É um software de modelagem molecular, que serve tanto para o *docking* quanto para o cálculo de energia

livre de ligação entre proteínas e ligantes. (Rezácová et al., 2008; Morris et al., 2009)

O programa AutoDock utiliza campos de força híbridos (*semi empirical free energy force field*) para calcular a energia livre de ligação entre pequenas moléculas e seus alvos macromoleculares. O termos empíricos, aliados a um baixo custo computacional, tornam a predição de energia livre normalmente menos acurada pelo AutoDock, do que utilizando métodos baseados em campos de força (MM/GBSA e MM/PBSA). Apesar desta desvantagem frente a outros métodos, o AutoDock apresenta resultados satisfatórios considerando o ranqueamento de pequenas moléculas no estudo de complexos com proteínas e ligantes (Morris et al., 2009).

Com base nas coordenadas provenientes da DM, o AutoDock determina um valor de energia livre de ligação para cada *snapshot*, ao contrario das técnicas envolvendo mecânica clássica e métodos híbridos. Neste estudo o valor de referência para cada determinação foi o valor mais populado (em forma de *cluster*) entre os *snapshots* conforme exemplificado pelos histogramas dos Anexo 4 e 5. Para a parametrização dos ligantes foram utilizadas dois tipos de cargas, a AM1-BCC (mesma do pacote AMBER) e a carga Gasteiger (padrão do AutoDock) para fins de comparação.

Os seguintes parâmetros foram utilizados no AutoDock: (i) *grid* centralizado nos ligantes com um espaçamento de malha de 0.375 Å com 70 pontos para coordenada X, 80 para Y e 80 para Z, (ii) as trajetórias foram alinhadas utilizando como centro de *grid* -2,027 para X, 0,442 para Y e -1,021 para Z (com o intuito de usar o mesmo centro de *grid* para os ligantes). O tempo de simulação escolhido para os testes com o AutoDock foi de 2 ns.

#### 4.5.3 LigScore e DrugScore

As abordagens *knowledge-based* possuem acurados sistemas de *scoring* com a finalidade de ranquear pequenas moléculas baseados na interação proteína-ligante. Além da boa acurácia estas abordagens se tornam interessante pelo custo computacional relativamente baixo em comparação com outras ferramentas, como o GBSA/PBSA.

O programa LigScore está disponível como parte do IMP *toolkit* e também como uma ferramenta web (disponível online e de forma gratuita)(Krammer et al., 2005). Para esta abordagem, foram utilizadas os *snapshots* provenientes da DM resultando em um total de 2 ns analisados (400 *snapshots*). Cada *snapshot* resulta em um valor de energia livre de ligação independente e portanto o método escolhido para determinar o resultado final para cada ligante foi valor mais populado em forma de *cluster*. (semelhante ao utilizado no programa AutoDock) (Anexo 4).

O DrugScore consiste em um método baseado na interação entre pequenas moléculas orgânicas

já descritas em bancos de dados (Velec et al., 2005). As informações estruturais são retiradas de estudos de complexos proteína-ligante obtidos por cristalografia usando ReLiBase e convertendo os dados em preferências entre pares de átomos e área acessível ao solvente (SAS). Um estudo conduzido por Velec e colaboradores utilizando uma amostragem de 100 compostos, apresentou resultados satisfatórios, com um coeficiente de correlação de 0,624 quando comparadas as tendências dos compostos preditos com os experimentais (Velec et al., 2005).

Os *snapshots* utilizados para os cálculos via DrugScore foram obtidos da DM. O protocolo estabelecido foi o uso de 400 *snapshots* resultando em um total de 2 ns analisados. Assim como no caso do LigScore e AutoDock, foram utilizados histogramas na forma de *cluster* para se obter o resultado mais populado de cada ligante.

O DrugScore e o LigScore não estimam os valores de energia livre de ligação em kcal/mol. Este dois métodos utilizam funções *scoring* inerentes a cada programa. Por este motivo não se pode comparar os valores preditos com os valores experimentais, apenas a correlação dos compostos analisada por regressão linear.

#### 4.6 QUÍMICA QUÂNTICA NA DETERMINAÇÃO DE ENERGIA LIVRE

O campo da química quântica apresentou significativos avanços nas últimas décadas devido principalmente ao fato das melhorias e avanços em computadores, *hardwares* e *softwares*. Além disso, a implementação de métodos de QM (*quantum mechanics*) mais robustos, eficientes e precisos permitiram a extensão de seu uso para problemas “do mundo real” (como desenvolvimento de fármacos). Nos dias atuais a QM é amplamente utilizada nas indústrias mais modernas (principalmente química e farmacêutica) como uma ferramenta para racionalizar e prever o curso de reações químicas, promover o desenvolvimento e caracterização de materiais, assim como acessar as propriedades de estados excitados e fundamentais de moléculas (Deglmann et al., 2015).

Para realizar os cálculos por SQM (*Semi-empiricum quantum mechanichs*), o software MOPAC foi utilizado. O MOPAC (Molecular Orbital PACkage) é um popular programa de computador usado em química computacional, distribuído de forma livre. Foi desenhado com o intuito de implementar algoritmos semi-empíricos de química quântica. A última versão disponível do programa é o MOPAC2012, utilizada neste trabalho, com a implementação do método PM7. Os métodos semi-empíricos PM3, PM6-DHX, PM6-DH+ e PM7 foram utilizados para obter os cálculos de orbitais moleculares para os cálculos de química quântica. Foram tratados na região quântica o ligante e uma área de 4 Å ao seu entorno (Figura 16). Ao contrário dos métodos híbridos (que consideram a região quântica e a região clássica), a química quântica utilizada nesta

abordagem leva em consideração apenas as interações da região quântica (Stewart et al., 2013). Na Figura 16, está exposto o ligante GEQ (precursor dos ligantes 2AX, 4A-JX) e a área considera por QM neste estudo. Esta região compreende o ligante, mais a coenzima NADH e onze resíduos de aminoácidos que se encontram a 4 Å de distância do centro do ligante.

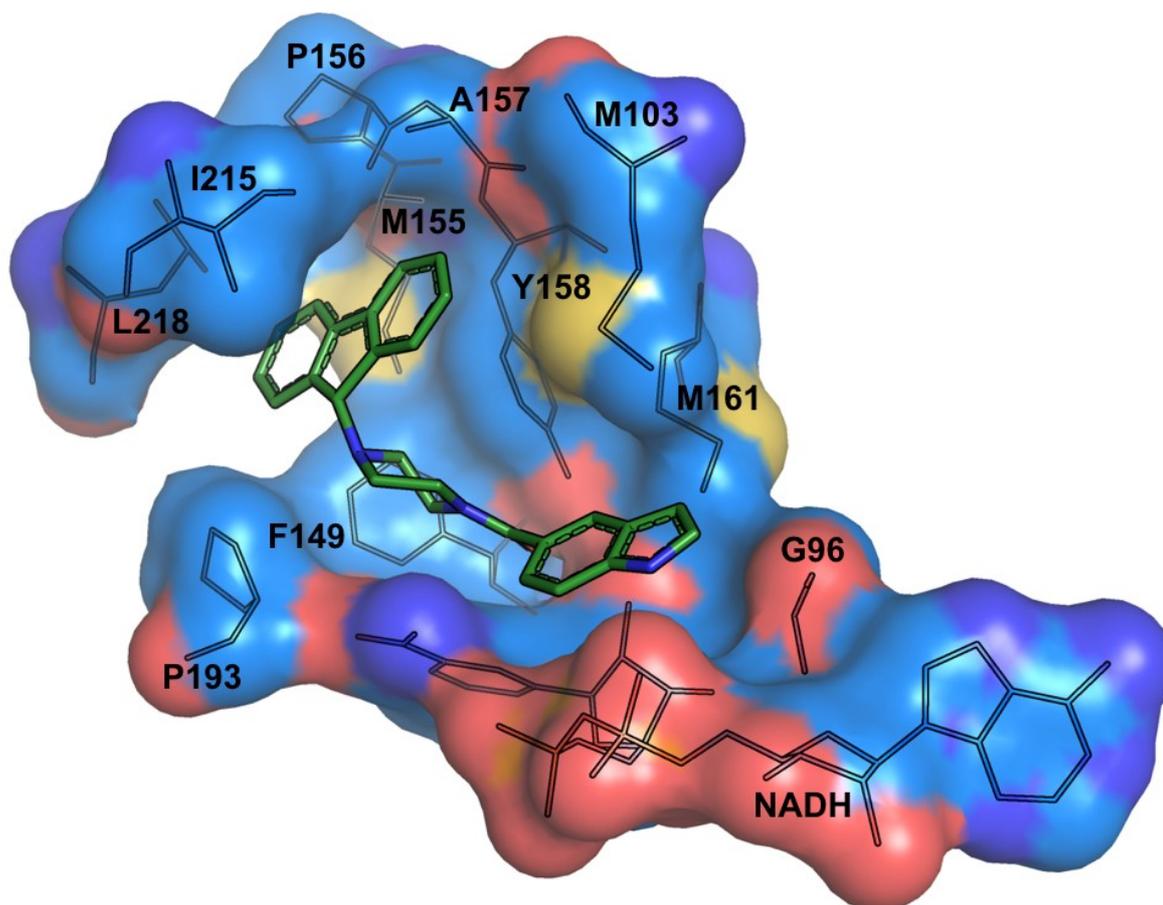


Figura 12: Ilustração da fração do complexo considerado na região quântica, formada pelos resíduos que se encontram a 4 Å de distância do ligante. Imagem: InhA (PDB ID: 1BVR com o ligante GEQ).

#### 4.7 MÉTODO DE COMPARAÇÃO

A comparação entre as metodologias descritas e os valores experimentais, foi realizada pela ferramenta estatística de regressão linear. A regressão linear foi escolhida, por ser o padrão de comparação entre dados experimentais e *in silico* envolvendo predição de energia livre de ligação para uma conhecida série de compostos. (Greenidge et al, 2013). Os cálculos de regressão linear e os gráficos de correlação foram obtidos por meio do programa QtiPlot versão para linux (Maestre-Reyna et al., 2013). Foram mantidas as mesmas escalas nos gráficos para o eixo X (valores experimentais), e variadas as escalas no eixo y (valores preditos), devido a grande variância nos valores preditos, permitindo uma melhor interpretação visual dos resultados.

A regressão linear consiste em uma equação para se estimar uma variável  $y$  (condicional e esperada), com base nos dados de uma outra variável  $x$  (variável base). Os dados experimentais foram considerados no eixo  $x$  (como o valor de referência) e os dados preditos foram alocados no eixo  $y$ . O  $R^2$  é uma medida do sucesso de determinada estimativa (Fausto et al., 2008). A interpretação dos resultados foi baseada nos seguintes parâmetros: valores de  $R^2$  menores que 0,10 foram considerados não correlacionados, entre 0,10 e 0,30 fracamente relacionados, entre 0,30 e 0,60 moderadamente relacionados, acima de 0,60 altamente relacionados e valores iguais ou superiores à 0,90 foram considerados ótimos (Fausto et al., 2008).

## 5. RESULTADOS E DISCUSSÃO

Neste capítulo versaremos sobre os resultados obtidos de acordo com as metodologias exemplificadas no capítulo 4. Os resultados foram divididos em subitens para uma melhor compreensão das abordagens utilizadas e devido ao grande número de gráficos expostos. A forma de gráficos foi escolhida para permitir avaliar tanto os resultados (valores), como a inclinação da reta em cada caso. Os resultados finais foram comparados com os dados experimentais, com intuito de se obter o perfil de correlação, por meio da regressão linear.

### 5.1 TESTES EM GPU

Devido ao grande número de simulações (150) necessárias para o desenvolvimento deste trabalho, foram realizados testes preliminares em GPU para avaliar o seu funcionamento. Os testes foram realizados primeiramente com o objetivo de avaliar o ganho computacional descrito na literatura para outros sistemas. Assim como, a reprodutibilidade dos resultados para o complexo de InhA e seus ligantes (Greenidge et al., 2013). O programa Pmemd (integrante do pacote AMBER) foi utilizado em paralelo (multiprocessadores) com o auxílio de duas unidades de processadores gráficos.

Com base no tempo final da fase de produção de cada simulação, foi possível verificar um aumento de 22,4x na velocidade das simulações, quando comparadas ao uso em CPU (8 núcleos em paralelo), conforme demonstrado na Tabela 3.

Na fase dos testes de velocidade das simulações, três estruturas de InhA complexadas com o ligante GEQ (Genz 10850) (PDB ID: 1P44) foram escolhidas, totalizando aproximadamente 36.500 átomos em cada sistema (Tabela 3). Para fins de análise da qualidade das simulações, foram comparados o volume, a temperatura, a densidade e a energia potencial total do sistema. A partir dos gráficos de comparação com sobreposição de valores foi possível concluir que não ocorreram diferenças nos quesitos avaliados nas simulações entre GPU e CPU (Figura 13).

Tabela 3: Comparação da velocidade das simulações em sistemas de aproximadamente 36.500 átomos, comparando-se CPU (Intel® Core™ i7-3770 CPU @ 3.40 GHz x 4, 16 GB RAM, 2 TB HD) e GPU (Intel® Core™ i7-3970 x CPU @ 3.50 GHz x 12,64 Bit, 2 TB HD, GTX 680 nvidia geforce).

Ligante	CPU	GPU
1P44	2.5 ns/dia	56 ns/dia
1ZID	2.5 ns/dia	56 ns/dia
1ENY	2.5 ns/dia	56 ns/dia

Devido ao ganho computacional expressivo, e a manutenção da qualidade/reprodutibilidade das simulações, as fases de produção de dinâmica molecular utilizadas no presente trabalho foram realizadas em GPU. Segundo Case e colaboradores, desenvolvedores do pacote AMBER, o uso de processadores gráficos foi amplamente testado nos últimos anos, sendo recomendado e validado tanto para a pesquisa acadêmica quanto industrial. (Case et al., 2005; Case et al., 2012; Case et al., 2014).

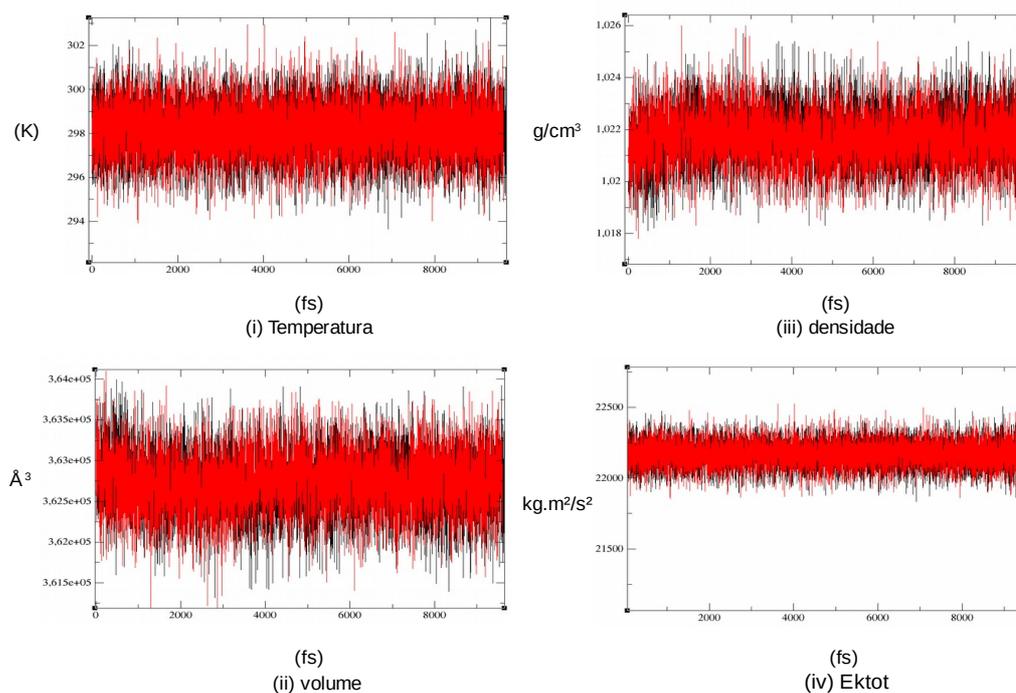


Figura 13: Comparação das simulações da enzima InhA código PDB: 1P44 pelo método de dinâmica molecular (4ns), comparando CPU (Intel® Core™ i7-3770 CPU @ 3.40 GHz x 4, 2 TB HD) em vermelho e GPU (Intel® Core™ i7-3970 x CPU @ 3.50 GHz x 12,64 Bit, 2 TB HD, GTX 680 e 780 nvidia geforce) em preto, com relação a (i) temperatura (ii) volume (iii) densidade (iv) e energia total do sistema.

## 5.2 ANÁLISE DA DINÂMICA MOLECULAR

Componentes das simulações por DM foram analisados com o intuito de monitorar as mesmas e garantir uma boa performance e reprodutibilidade dos resultados. Ao término de cada simulação foram analisados o RMSD, a temperatura e o comportamento visual do ligante e da proteína através do programa VMD. Comparando-se com metodologias laboratoriais e de produção, esta etapa é equivalente ao controle de qualidade, garantindo o bom funcionamento dos experimentos.

A realização desta etapa é importante pois todas as metodologias descritas neste trabalho, para estimar a energia livre de ligação, foram baseadas nos dados provenientes da DM. Portanto um erro nesta etapa pode comprometer os valores preditos. O objetivo de avaliar a temperatura foi verificar

a estabilidade do sistema, que deve se manter em torno dos 298.16 K. O RMSD permite avaliar possíveis mudanças estruturais no complexo, decorrentes, por exemplo, de saídas do ligante do sítio de ligação. Em todas simulações, além do RMSD, a qualidade e manutenção visual das hélices alfa e fitas beta foi avaliada. Nesta etapa foram verificadas as 150 simulações realizadas ao longo do estudo, conforme os três parâmetros descritos e não foram encontradas nenhuma irregularidade nas mesmas.

### 5.2.1 Análise dos ligantes e suas interações com a InhA

A análise das interações entre proteína e ligante, tornam-se passo primordial para o entendimento e a compreensão do modo de ligação e posterior caracterização da energia livre de ligação por meio de métodos *in silico*. Para contabilizar estas ligações, o programa LIGPLOT, que gera representações esquemáticas em 2 D das ligações entre proteína e ligante foi utilizado (Wallace et al., 1995). Os 14 ligantes utilizados neste trabalho, passaram pela análise de ligação via LIGPLOT. Nas figuras esquemáticas 14 e 15, estão representados seis ligantes (5 PDB + 1 composto derivado do Genz 10850), demonstrando os tipos de interação que ocorrem predominantemente entre os ligantes descritos e a proteína InhA. É preciso ressaltar, que as figuras 14 e 15 são representativas do *frame* inicial da dinâmica molecular, podendo outras ligações serem formadas/quebradas ao longo da DM.

Uma análise interessante, recai sobre a tirosina 158. Conforme os primeiros estudos referentes a enzima InhA e a resolução de sua estrutura, (Dessen et al., 1995, Rozwarski et al., 1998, Rozwarski et al., 1999) este resíduo de aminoácido se mostra importante para a formação de uma ligação coesa entre proteína e ligante. A perda da ligação de hidrogênio entre a tirosina 158 e seus ligantes, pode acarretar a diminuição das forças de contato e desenvolvimento de resistência ao ligante, como é o caso da Isoniazida (Dessen et al., 1995, Rozwarski et al., 1998, Rozwarski et al., 1999). Entre os compostos estudados, aqueles que apresentaram ligação de hidrogênio com a tirosina 158 (5PP, TCU,8PS), apresentaram *in vivo*, os melhores valores de energia livre de ligação. Quando se leva em consideração as interações entre o NADH e os ligantes selecionados, não foi possível estabelecer correlação entre que demonstre a ligação de hidrogênio com o NADH e a diminuição da energia livre de ligação *in vivo*. O número de contatos hidrofóbicos por si só, também não se mostrou uma boa métrica para correlacionar os valores de energia livre de ligação. Compostos como o TCL (-6,91 kcal/mol) apresentam o mesmo número de contatos no *frame* inicial que o TCU (-11,05 kcal/mol).

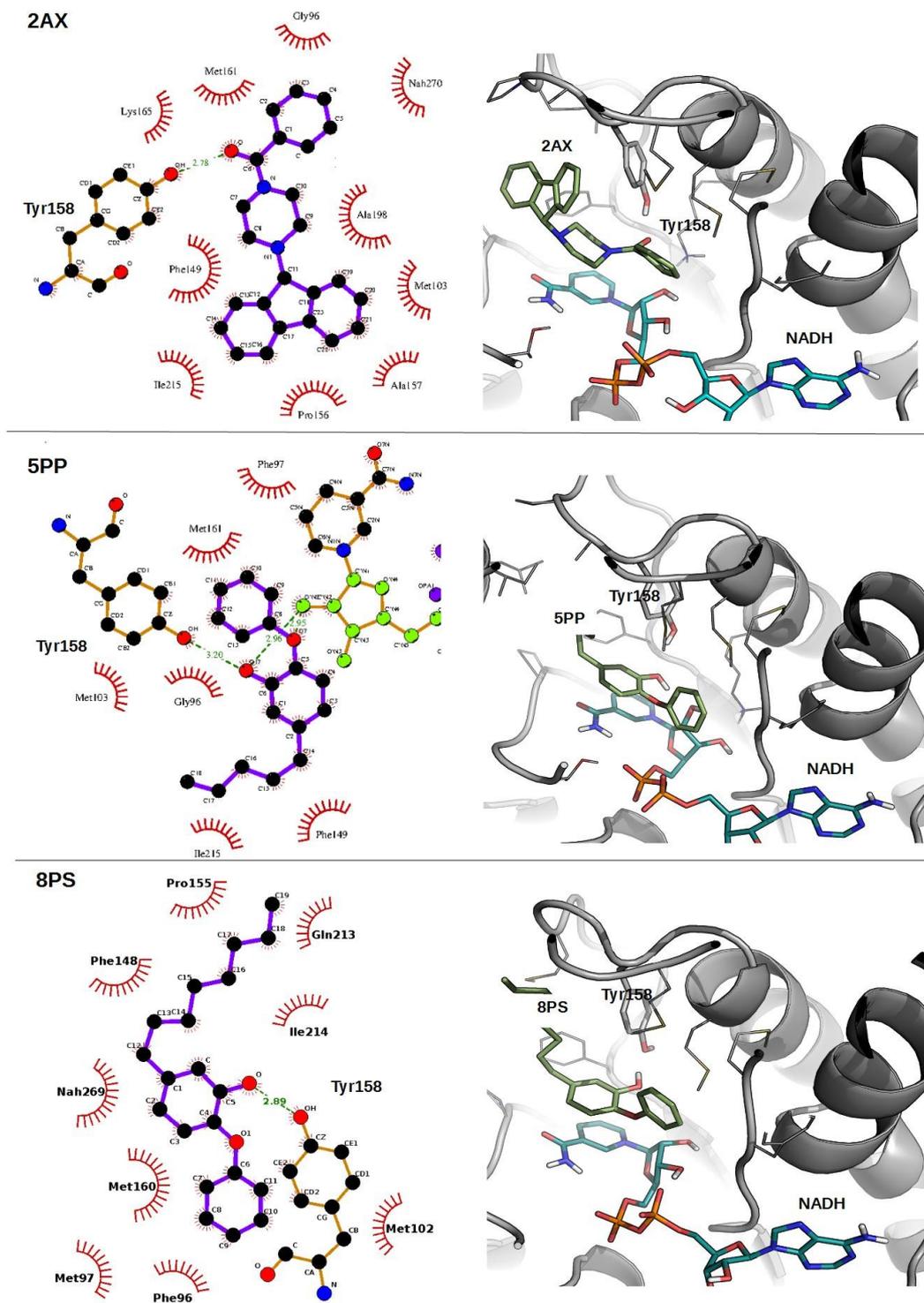
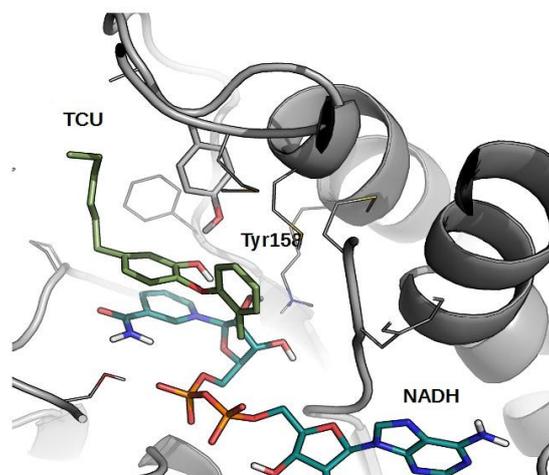
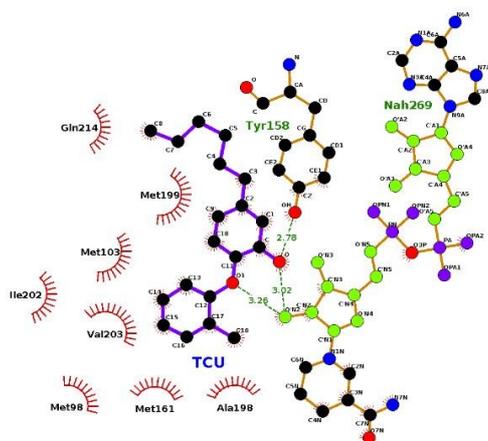
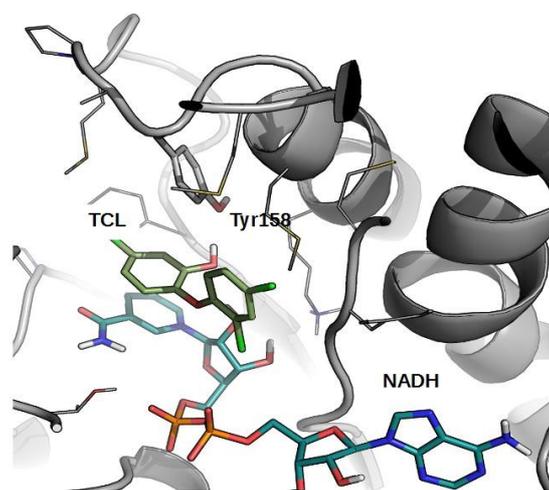
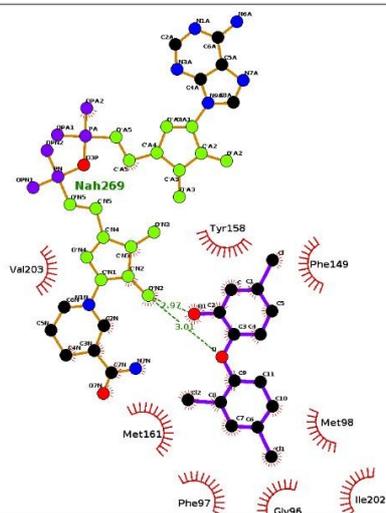


Figura 14: Imagem representativa dos tipos interações que ocorrem entre os ligantes selecionados e a proteína InhA de Mtb. As linhas tracejadas representam as ligações de hidrogênio e os arcos vermelhos os contatos hidrofóbicos. Em verde está representada a molécula NADH, considerada integrante da proteína.

TCU



TCL



3KX

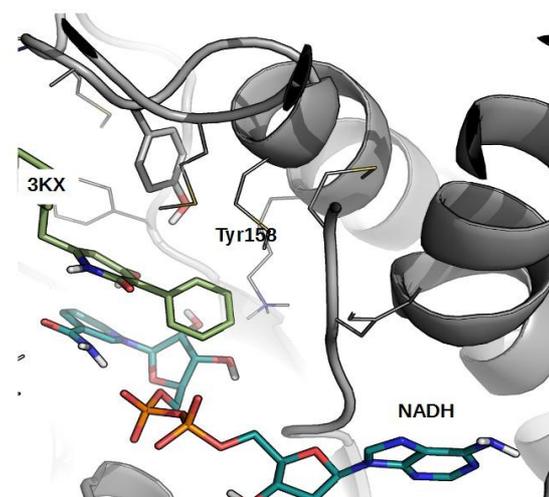
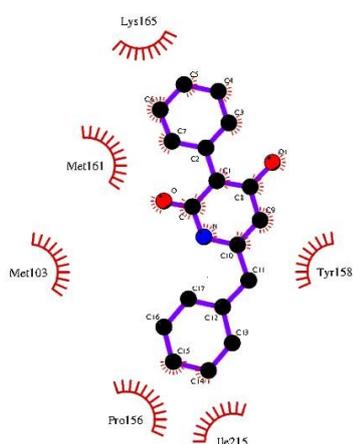


Figura 15: Imagem representativa dos tipos interações que ocorrem entre os ligantes selecionados e a proteína InhA de Mtb. As linhas tracejadas representam as ligações de hidrogênio e os arcos vermelhos os contatos hidrofóbicos. Em verde está representada a molécula NADH, considerada integrante da proteína.

### 5.3 RESULTADOS MM/GBSA

Após os testes em GPU e avaliação da fase de produção de cada ligante, foi possível passar para a fase de predição de energia livre. A primeira ferramenta escolhida para o desenvolvimento deste trabalho foram os cálculos utilizando o método MM/GBSA, por ser o mais frequente e amplamente utilizado.

Segundo Su e colaboradores, os resultados obtidos por MM/GBSA são sensíveis as variações no tempo de simulação (Su et al., 2015). Com base nessa informação, os primeiros testes realizados foram variando o tamanho de simulação associado ao MM/GBSA. O seguinte *workflow* foi desenvolvido para o GBSA: (i) variação do tempo de simulação, (ii) variação dos campos de força, (iii) variação dos campos de força com raios atômicos otimizados para GBSA, (iv) MM/GBSA contabilizando os valores de entropia para o campo de força ff14SB e (v) MM/GBSA utilizando múltiplas trajetórias para o campo de força ff14SB. Este *workflow* foi desenvolvido nesta sequência, pois cada etapa é complementar a anterior (Su et al., 2015).

O primeiro fato inerente ao GBSA observado, foram os valores de energia livre obtidos. Os valores absolutos foram consideravelmente maiores que os valores experimentais medidos em kcal/mol. Os dados experimentais estão situados entre -6.9 kcal/mol e -12.2 kcal/mol, e os valores preditos por GBSA encontraram-se entre -30 e -65 kcal/mol dependendo do campo de força testado. Esta diferença entre os valores preditos já era esperada conforme trabalhos de referência realizados com o GBSA (Genheden et al., 2015a; Greenidge et al., 2013). Os valores obtidos reforçam os dados descritos na literatura (Hou et al., 2011), onde os métodos de GBSA não são acurados para prever os valores experimentais aproximados. Porém, estes métodos descrevem valores relativos, apresentando a tendência entre um grupo de compostos.

#### 5.4.1 Tempo de simulação em MM/GBSA

O tamanho de simulação pode influenciar os resultados de MM/GBSA dependendo do sistema utilizado e o tempo de convergência da simulação. Nos estudos entre a interação proteína e ligante por meio de DM, simulações entre 2-6 ns são amplamente utilizadas (Xu et al., 2013; Hou et al., 2011; Kumari et al., 2014), e por este motivo foi dada maior importância para esta faixa de tempo.

Nesta fase de testes preliminares, dois ligantes foram escolhidos, o 8PS e o TCL. Foram utilizados os tempos de simulação de 1ns, 2ns, 4ns, 6ns, 10ns e 50ns com o campo de força ff14SB (Tabela 4).

Os valores médios de energia livre para os compostos 8PS e TCL foram de  $-38,3256 \pm 0,8626$

kcal/mol e  $-35,5308 \pm 0,5282$  kcal/mol, respectivamente. A partir destes resultados não foram encontradas diferenças significativas nos valores de  $\Delta G$  em relação aos tempos de simulação analisados frente a este complexo (InhA + ligantes). Conforme Hou e colaboradores, quando não existe diferença significativa variando-se o tempo de simulação nos valores preditos, preconiza-se o uso entre 2 ns e 6 ns. (Hou e al., 2011). Com base nestes dados, o tempo de simulação escolhido para o restante do trabalho foi de 2 ns, devido ao ganho computacional e a estabilidade verificada do complexo InhA+ligante no presente trabalho.

Tabela 4: Diferenças de  $\Delta G$  calculadas pelo método MM/GBSA de acordo com tempo de simulação, da enzima InhA. PDB ID: 1P44 associado aos ligantes 8PS e TCL. **Resultados em kcal/mol.**

Tempo	8PS	TCL
1 ns	-37.9212	-36.4019
2 ns	-36.9234	-36.0411
4 ns	-38.9845	-35.4422
5 ns	-38.1011	-35.5495
10 ns	-39.7451	-34.9992
50 ns	-38.9284	-34.9111

#### 5.4.2 Campos de Força

Não existe um consenso em relação a qual campo de força apresenta os melhores resultados na predição de  $\Delta G$  pelo método de GBSA. Somando-se a este fato, a maioria dos estudos com GBSA não foram realizadas utilizando a enzima InhA como base. Com o intuito de garantir uma ampla varredura nos campos de força, foram testados os seguintes: ff98, ff03, ff99, ff99SB, ff99Bldn, ff99Bnrm, ff12SB e ff14SB. O número após cada campo de força indica o ano em que este foi lançado, lembrando que não necessariamente um campo de força mais moderno apresenta melhores valores para a predição de  $\Delta G$ .

Os resultados obtidos (Bondi radii ( $GB^{OBC}$ )) para o Grupo 1 foram muito variados (Figura 16), apresentando valores de  $R^2$  entre 0,70 (ff99) e 0,00 (ff12SB). Exemplificando neste caso o que foi mencionado anteriormente sobre campos de força mais modernos e sua acurácia. Apesar da alta correlação obtida pelo campo ff99, a maioria dos campos de força apresentaram uma fraca correlação com valores de  $R^2$  entre 0,20-0,30. O Grupo 2, que possui alta complexidade de predição (Figura 17), apresentou resultados com uma fraca correlação. Porém, neste caso, o melhor dos

campos de força testados foi o ff12SB, que no Grupo 1 apresentou o pior resultado. O Grupo 3 não apresentou correlação significativa pela técnica de regressão linear para nenhum dos campos de força testados. Este, apresenta a maior complexidade, pois é formado por 14 ligantes com valores experimentais de energia livre muito próximos entre si, dificultando o ranqueamento dos mesmos (Figura 18).

A partir dos resultados obtidos com GBSA podemos observar que neste caso não houve uma predileção por algum campo de força. O ff99 apesar de ter apresentado um excelente resultado no Grupo 1, não apresentou resultados satisfatórios para o Grupo 2 e Grupo 3, não podendo ser apontado como o melhor dos campos testados nesta etapa. Os melhores resultados para o Grupo 1 em detrimento dos outros, demonstram a limitação desta técnica em predizer e ranquear valores muito próximos de energia livre de ligação (Grupo 2 e Grupo 3). Analisando a média de todos campos de força, para o Grupo 1, obtivemos um  $R^2$  de 0,26 de correlação (correlação fraca).

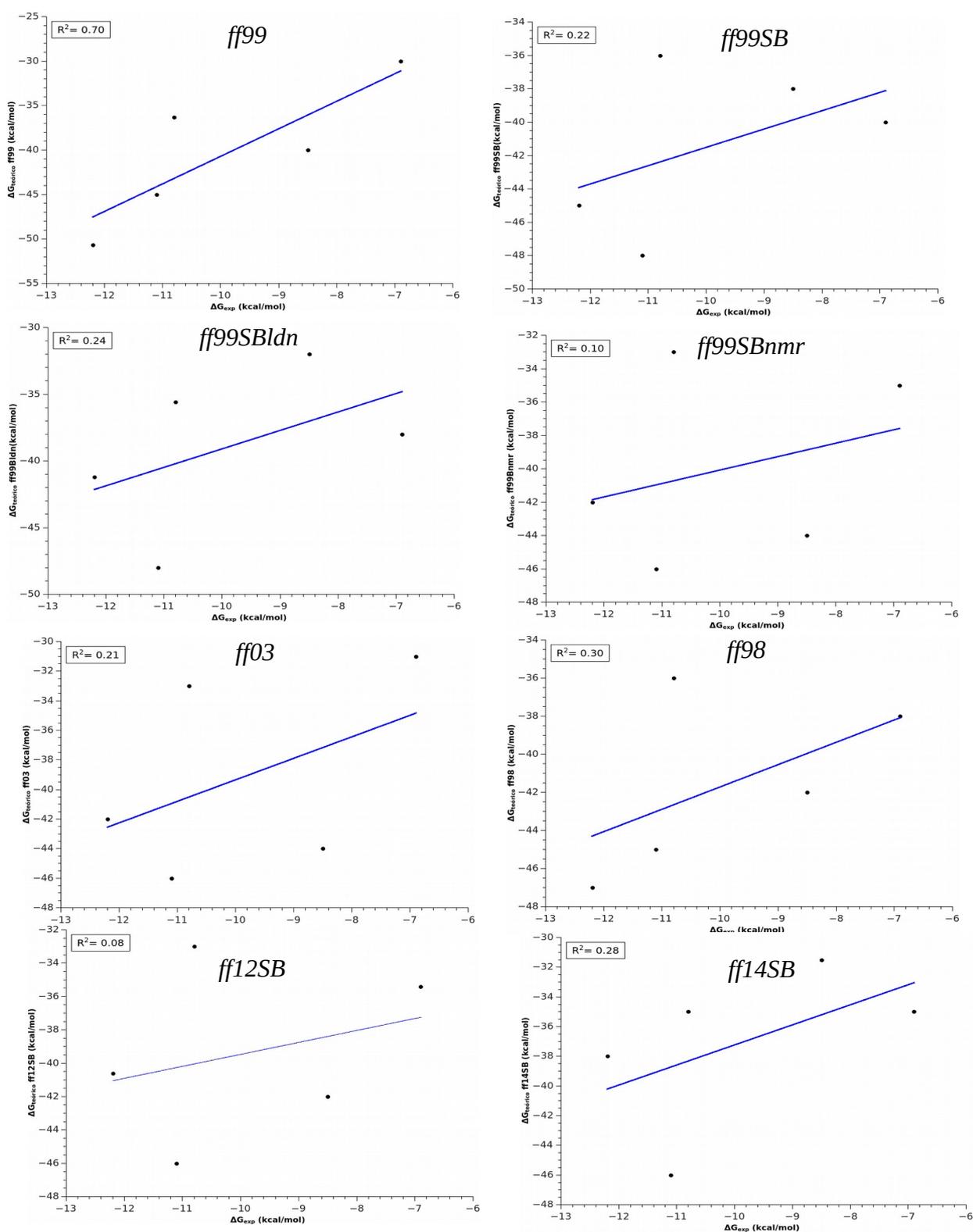


Figura 16: Diferença entre os campos de força testados para o Grupo 1 pelo método MM/GBSA. Da esquerda para a direita: ff99, ff99SB, ff99SBldn, ff99SBnmr, ff03, ff98, ff12SB, ff14SB.

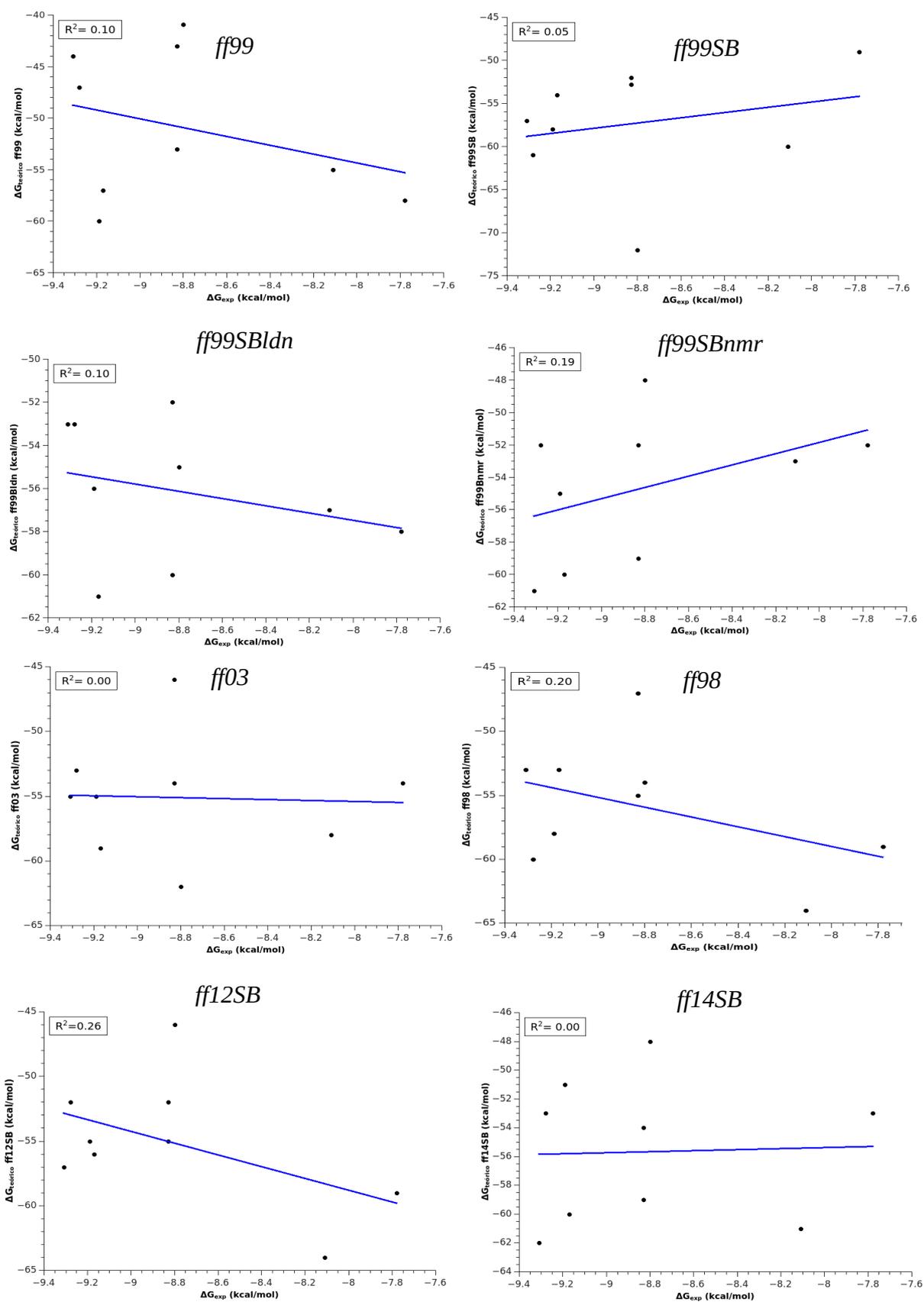


Figura 17: Diferença entre os campos de força testados para o Grupo 2 pelo método MM/GBSA. Da esquerda para a direita: ff99, ff99SB, ff99Bldn, ff99Bnmr, ff03, ff98, ff12SB, ff14SB.

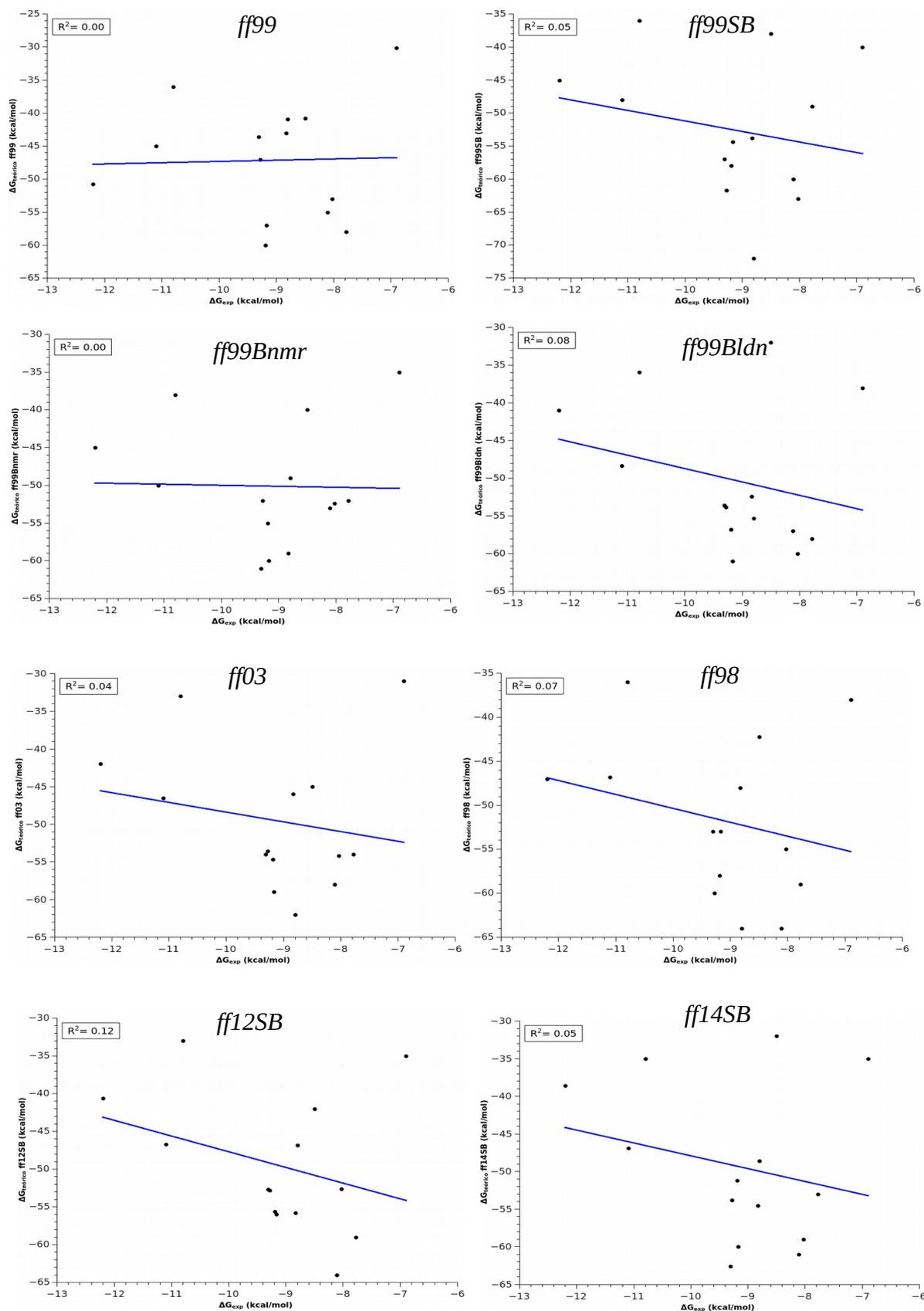


Figura 18: Diferença entre os campos de força testados para o Grupo 3 pelo método MM/GBSA. Da esquerda para a direita: ff99, ff99SB, ff99Bldn, ff99Bnmr, ff03, ff98, ff12SB, ff14SB.

#### 5.4.2.1 Campos de Força com raios atômicos otimizados

O MM/GBSA é mais sensível que o MM/PBSA para variações de raios atômicos atribuídos, e estudos preliminares demonstraram o desenvolvimento desses raios otimizados de forma especial para o método de GB (Su et al., 2015; Onufriev et al., 2004). Os raios otimizados foram criados pelo grupo de Onufriev e colaboradores com o intuito de realizar estudos entre proteína e ligante e melhorar os resultados preditos de energia livre de ligação nesses casos específicos. Devido a este fato, foi realizado o mesmo estudo dos campos de força previamente descrito, porém, desta vez, adotando os raios otimizados de Onufriev (PBRadii mbondi2 (GB<sup>OB<sup>BC</sup>2</sup>)). Nesta etapa foram utilizadas as mesmas simulações por DM descritas no item (5.4.2), mudando apenas os valores dos raios na preparação das topologias.

Apesar desta mudança nos raios ser sugerida pelos próprios desenvolvedores do pacote AMBER, os resultados foram semelhantes aos obtidos sem os raios otimizados. Para o Grupo 1 (Figura 19), foi possível observar uma pequena melhora em alguns campos de força em comparação ao método anterior. O ff99, que previamente já havia predito um resultado satisfatório, obteve um  $R^2$  de 0,80, considerado alto para a técnica de regressão linear. O ff99Bnmr também obteve melhoras significativas apresentando um  $R^2$  de 0,57. Os resultados para o Grupo 2 se mantiveram no mesmo patamar, representando uma fraca correlação entre os valores preditos e os experimentais. De acordo com as análises para o Grupo 3, não houve correlação na maioria dos casos, exceto para o ff99Bnmr que apresentou uma correlação moderada ( $R^2$  de 0,40). Comparado ao teste anterior, não foi possível identificar um campo de força superior ao demais.

Analisando a média dos valores para o Grupo 1, que obteve os melhores resultados entre os grupos testados, temos um  $R^2$  de 0,30 (correlação fraca). Os resultados foram semelhantes aos obtidos sem a atribuição dos raios atômicos otimizados ( $R^2$  de 0,26), não ocorrendo uma melhora na média geral dos campos de força testados. Se levarmos em consideração apenas os melhores valores encontrados para o Grupo 1 utilizando o ff99 com os raios otimizados ( $R^2$  de 0,80), e com raios padrões ( $R^2$  de 0,70), estes valores se mostraram maiores que a média universal do GBSA que é aproximadamente um  $R^2$  de 0,60 (Xu et al., 2013 ;Swanson et al., 2005; Kumari et al., 2014).

A utilização de diferentes campos de força foi importante para demonstrar a influência que os mesmos apresentam nos valores de predição de energia livre. Visto que, os resultados para um mesmo ligante variam significativamente entre os diferentes campos de força testados. Desta forma, a escolha do campo de força não é uma opção trivial nos estudos de GBSA.

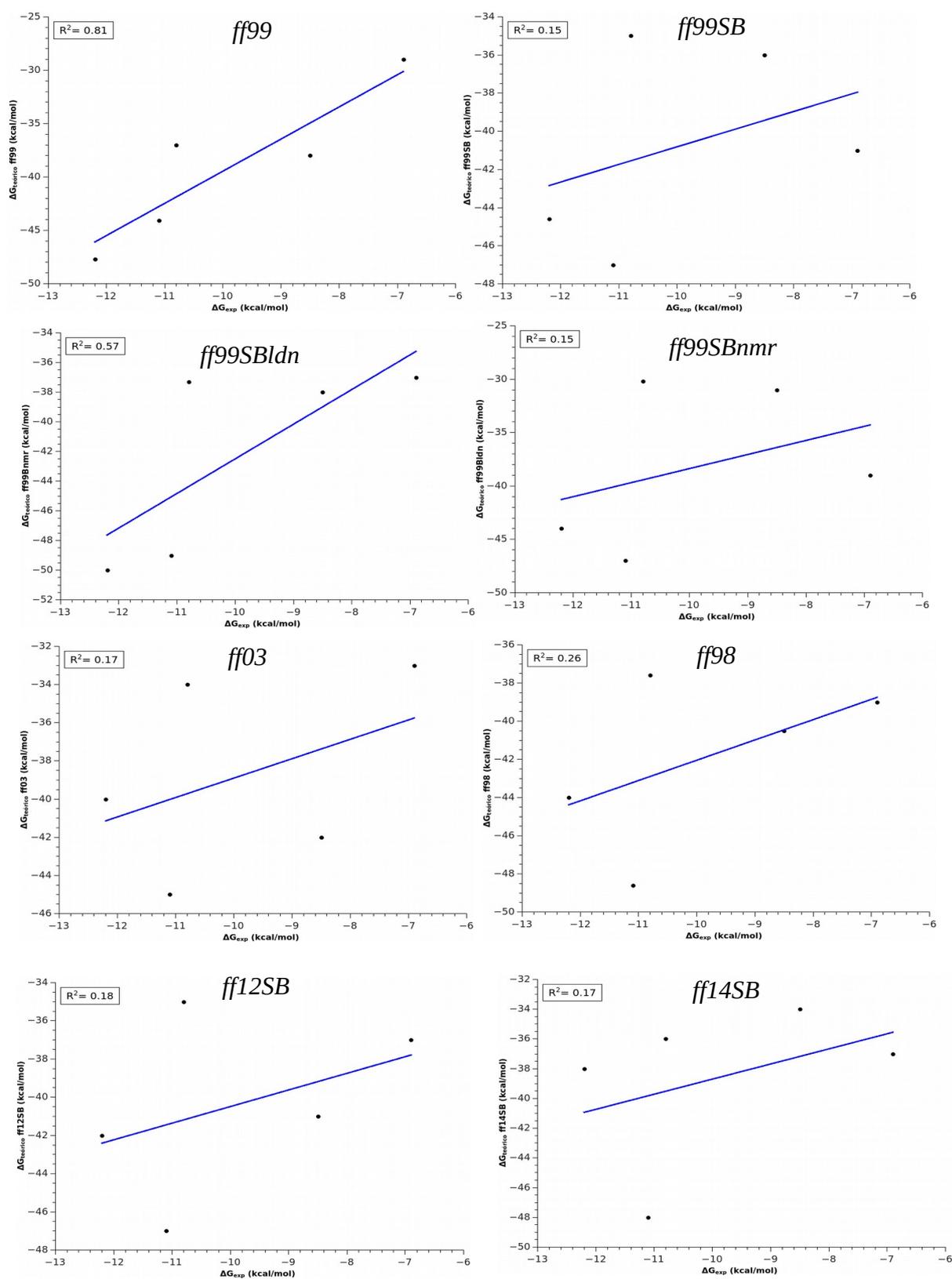


Figura 19: Diferença entre os campos de força testados para o Grupo 1 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2). Da esquerda para a direita: ff99, ff99SB, ff99SBldn, ff99SBnmr, ff03, ff98, ff12SB, ff14SB.

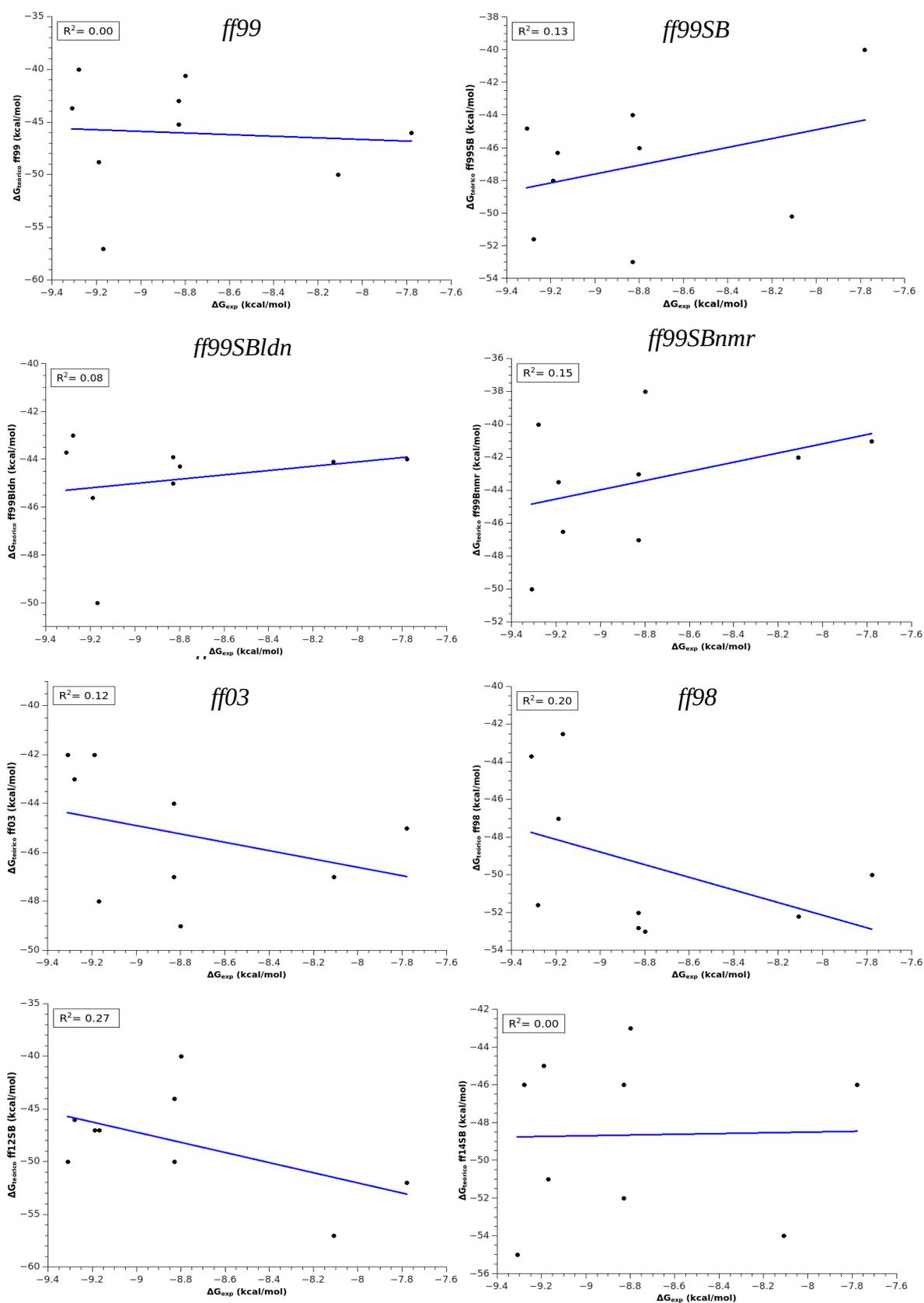


Figura 20: Diferença entre os campos de força testados para o Grupo 2 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2). Da esquerda para a direita: ff99, ff99SB, ff99SBldn, ff99SBnmr, ff03, ff98, ff12SB, ff14SB.

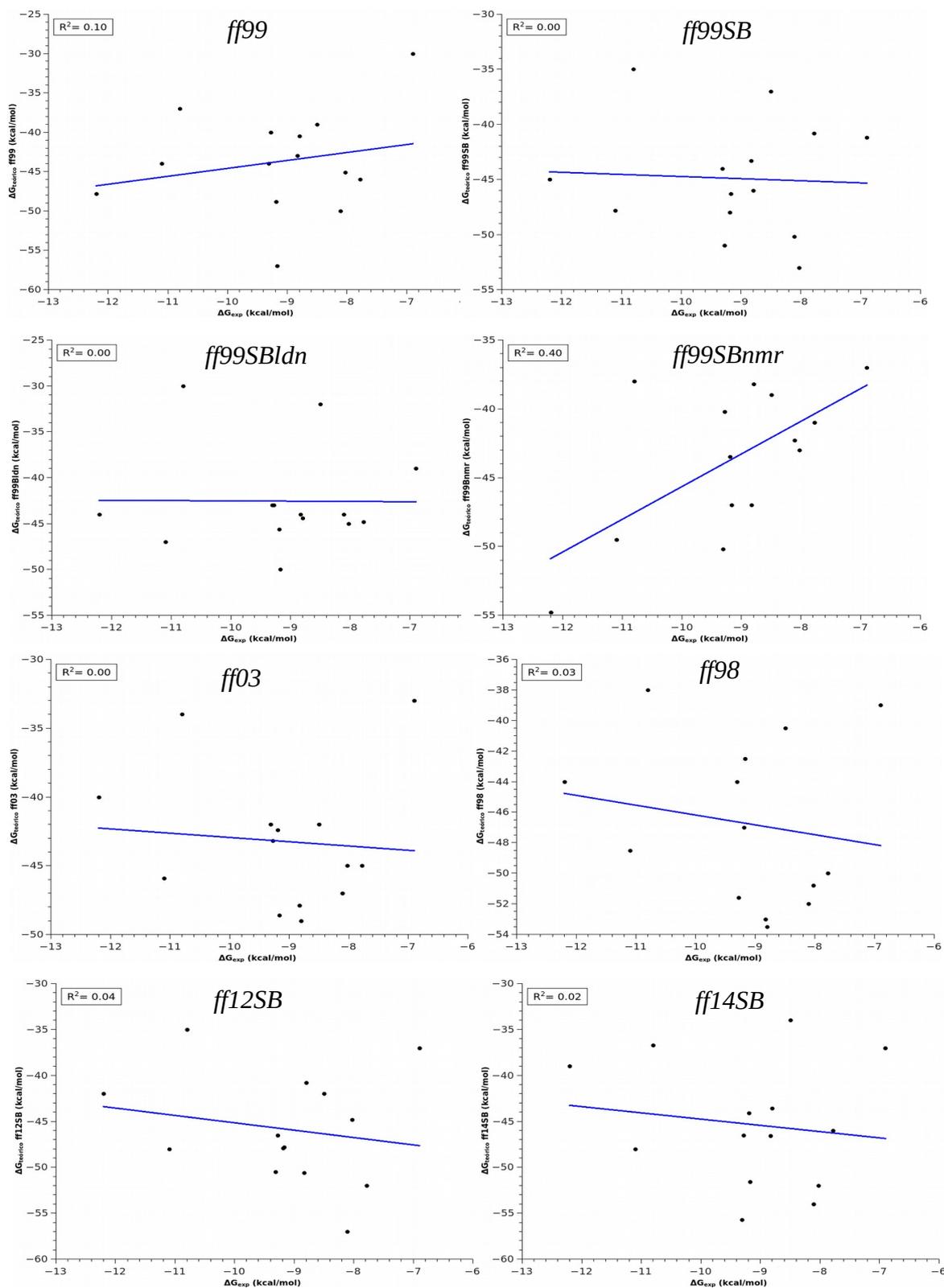


Figura 21; Diferença entre os campos de força testados para o Grupo 3 pelo método MM/GBSA utilizando raios otimizados de Onufriev, Bondi radii (mbondi2). Da esquerda para a direita: ff99, ff99SB, ff99Bldn, ff99Bnmr, ff03, ff98, ff12SB, ff14SB.

### 5.4.3 Campo de força ff14SB com múltiplas trajetórias

A partir dos resultados obtidos com os variados campos de força e os raios otimizados de Onufriev, foi realizado um estudo complementar testando a utilização de múltiplas trajetórias no método de GBSA para o campo de força ff14SB. As diferenças utilizando múltiplas trajetórias independentes costumam não interferir na determinação de energia livre nos métodos GBSA. Conforme o tempo de simulação vai aumentando, o método de múltiplas trajetórias tem sua acurácia diminuída em comparação ao método de uma única trajetória (método tradicional) (Su et al., 2015). Para esta fase de testes o tempo de 2 ns também foi o escolhido como padrão.

O campo de força ff14SB foi utilizado, pois se trata do último campo de força disponibilizado pelo pacote AMBER. Nesta etapa, foram testadas três abordagens para a determinação do  $\Delta G$ , o método de *Single Trajectory*, *Ligand/Complex trajectories* e *Multiple Trajectories*. Foi observado primeiramente que com o aumento do número de trajetórias de uma para duas, e posteriormente para três, ocorreu um aumento significativo do tempo de cálculo na ordem de 2x à 3x.

Apesar do aumento considerável do tempo de simulação, os métodos *Ligand/Complex* e *Multiple Trajectories* conseguiram melhorar a relação dos valores preditos para o Grupo 1. Os valores de  $R^2$  passaram de 0,17 (correlação fraca) para 0,40 (correlação moderada), conforme Figura 22. Apesar das consideráveis melhoras no Grupo 1, o Grupo 2 e Grupo 3 não obtiveram melhora na correlação (Figura 23 e Figura 24).

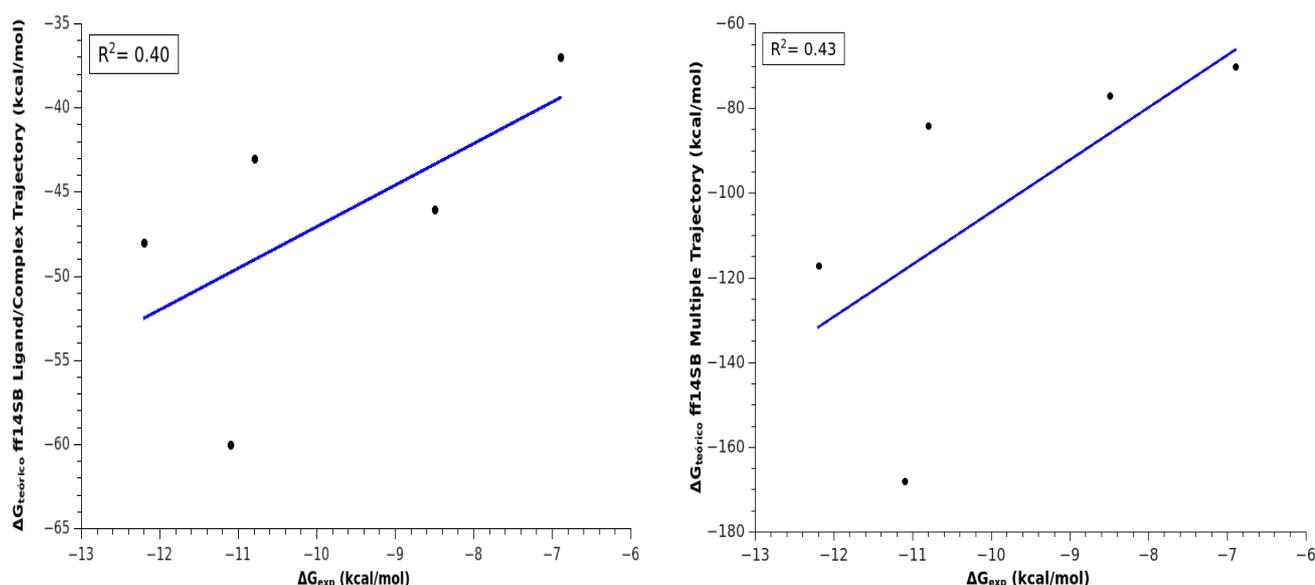


Figura 22: Grupo 1, diferença entre as abordagens do ligante+receptor (*Ligand/Complex*) e múltiplas trajetórias (*Multiple Trajectories*), este ultimo composto por trajetórias do ligante+receptor+complexo.

As metodologias de GBSA com múltiplas trajetórias também se mostraram ineficientes para prever os valores absolutos dos compostos. Os valores experimentais apresentados no Grupo 2 são próximos, fato que não ocorreu no método de múltiplas trajetórias, pois estes valores foram variados entre si.

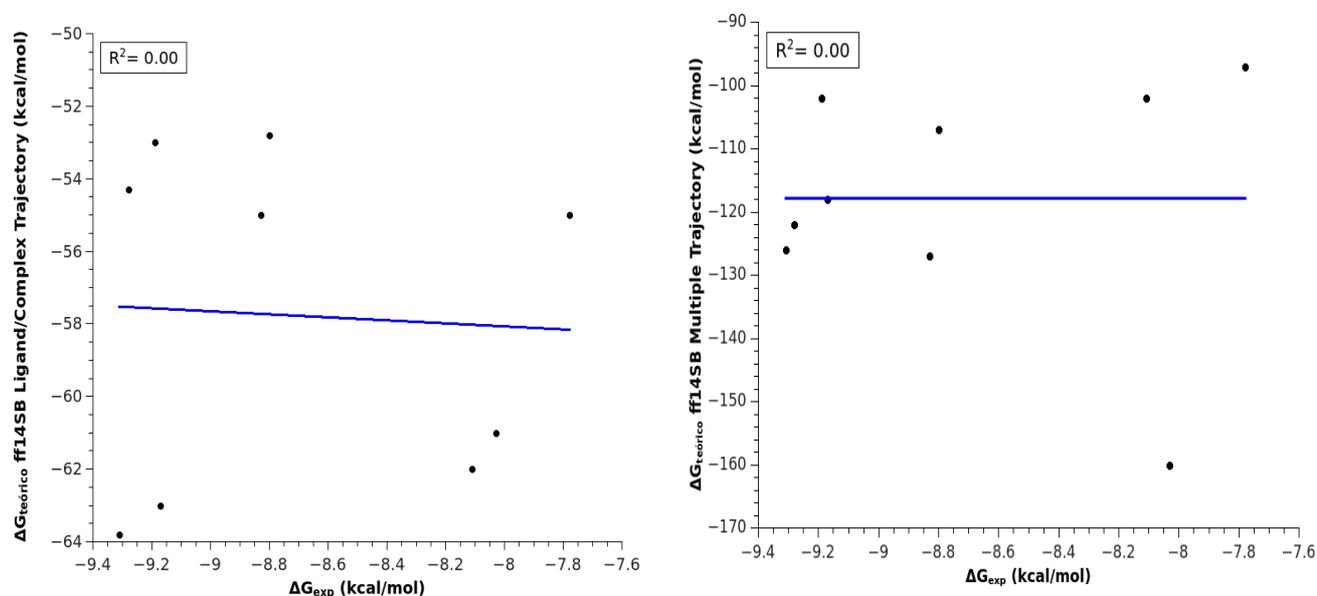


Figura 23: Grupo 2, diferença entre as abordagens do ligante+receptor (Ligand/Complex) e múltiplas trajetórias (Multiple Trajectories), este ultimo composto por trajetórias do ligante+receptor+complexo.

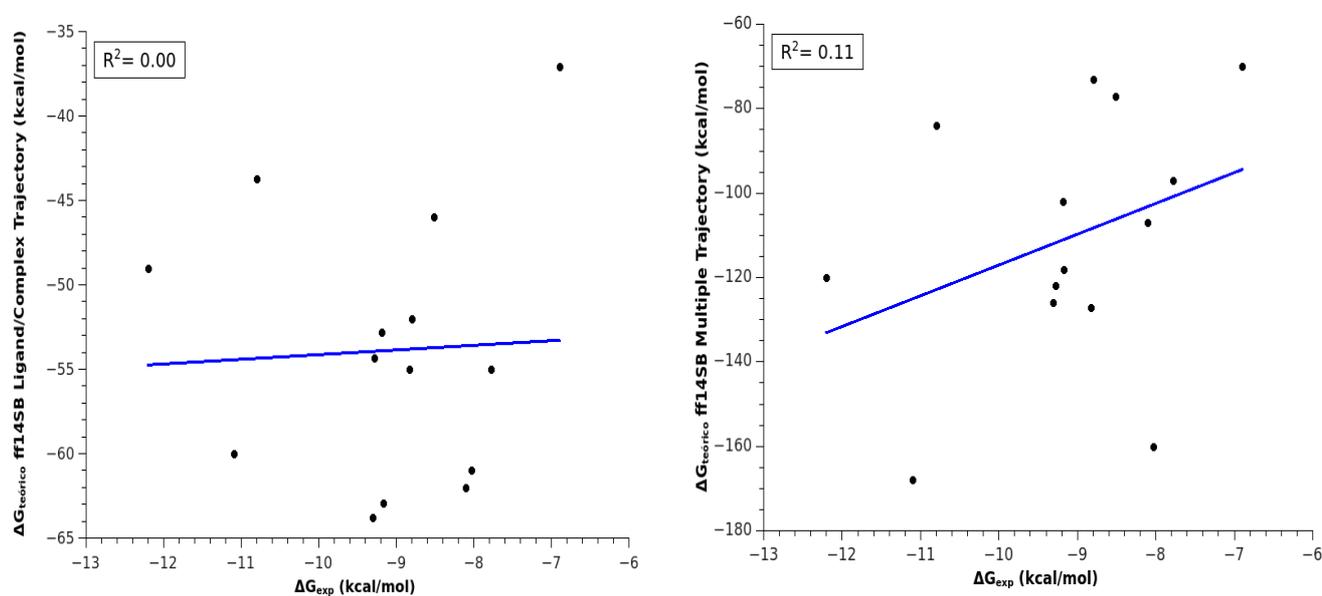


Figura 24: Grupo 3, diferença entre as abordagens de trajetória do ligante+receptor (Ligand/Complex) e múltiplas trajetórias (Multiple Trajectories), este ultimo composto por trajetórias do ligante+receptor+complexo.

Comparando os resultados em kcal/mol, estes se mostraram mais distantes que os cálculos pelo

método de *Single Trajectory*. O composto TCL, por exemplo, no ff14SB *Multiple Trajectories* apresentou um valor predito de energia livre de -170 kcal/mol, mostrando a distância em relação aos valores experimentais (-11.1 kcal/mol para o TCL) em uma ordem de aproximadamente 17x. Além desta distância nos valores absolutos, as distâncias entre eles (em relação aos compostos do grupo) também se mostrou maior que no método *Single Trajectory*.

#### 5.4.4 Campo de força ff14SB com entropia baseada em Modos Normais

Realizados os testes para o ff14SB com múltiplas trajetórias, a próxima etapa foi avaliar a contribuição da entropia nos cálculos de  $\Delta G$  *in silico*. Devido ao ganho computacional no MM/GBSA, o padrão é realizar cálculos desconsiderando os valores de entropia (Su et al., 2015). Além do aumento do custo computacional, os valores preditos levando em consideração a entropia não demonstram ganho efetivo em eficiência nos resultados (Kumari et al., 2014). Por este motivo um teste reduzido foi realizado levando em consideração a entropia associada aos métodos de *Single Trajectory* e *Multiple Trajectories* com o campo de força ff14SB. Os cálculos de entropia foram realizados no programa *MMPBSA.py* utilizando a análise dos modos normais (NMA).

Os resultados observados apresentaram valores distintos dos obtidos com a abordagem descrita anteriormente. No caso da entropia associada a múltiplas trajetórias, os valores variaram de -25 kcal/mol até -120 kcal/mol, aumentando os valores absolutos preditos, e assim, a distância em comparação aos valores experimentais. Os resultados para o Grupo 1 (Figura 25), apresentaram valores próximos aos obtidos com as abordagens anteriores. Com relação ao ff14SB isolado, testado no item 5.4.2 (Figura 22) os valores foram ligeiramente superiores, porém não apresentaram melhoras significativas.

Para o Grupo 2, não houve correlação entre os valores medidos pela regressão linear conforme apresentado na Figura 26. O Grupo 3 apresentou uma pequena melhora para a metodologia *Single Trajectory/Entropy* (Figura 27). Conforme demonstrado em estudos prévios com outros complexos proteína-ligante, os cálculos levando em consideração a entropia também não se mostraram efetivos frente a enzima InhA de Mtb e seu complexo com ligantes (Genheden & Ryde, 2015b). Ao contrário do senso comum, nem sempre quando adicionamos parâmetros e tornamos os cálculos mais demorados significa que os resultados vão ser refinados.

Os testes utilizando o NMA, vem ao encontro do que os desenvolvedores do pacote AMBER recomendam. Segundo Case e colaboradores, as atribuições entrópicas pelo método de NMA não estão bem atribuídas, principalmente para sistemas envolvendo um grande número de átomos e

ligações, como é o caso do sistema em estudo (Case et al., 2005). O número de variáveis aumenta quando é utilizado a abordagem de múltiplas trajetórias associadas ao NMA, demonstrando um decréscimo na qualidade das predições de energia livre no sistema estudado.

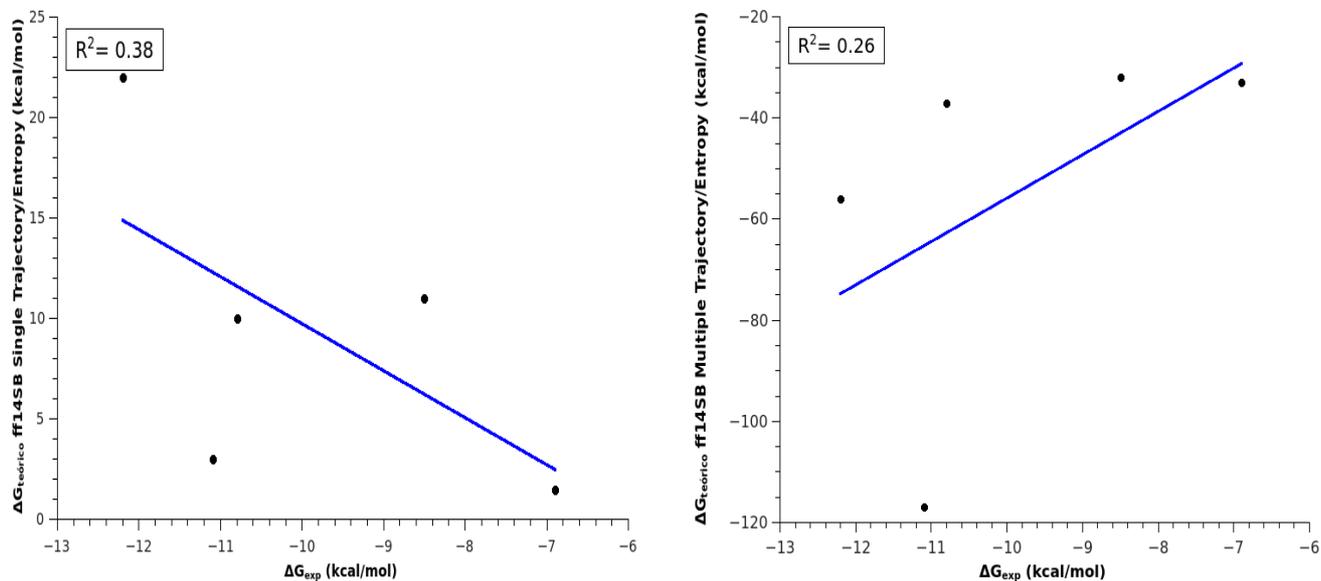


Figura 25: Grupo 1, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.

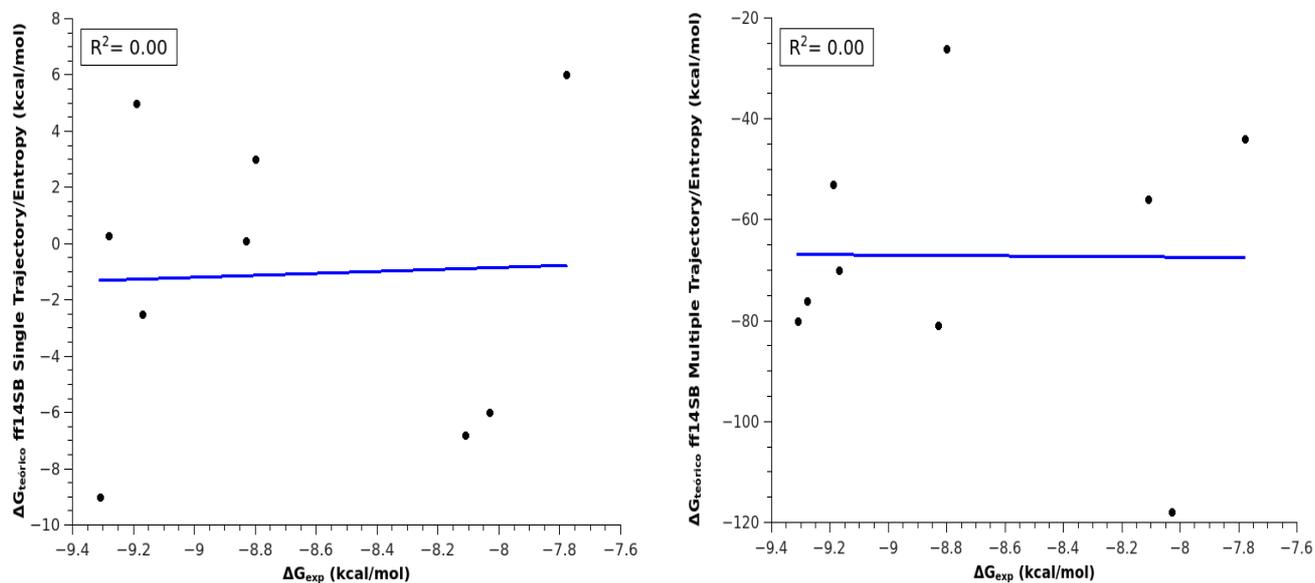


Figura 26: Grupo 2, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.

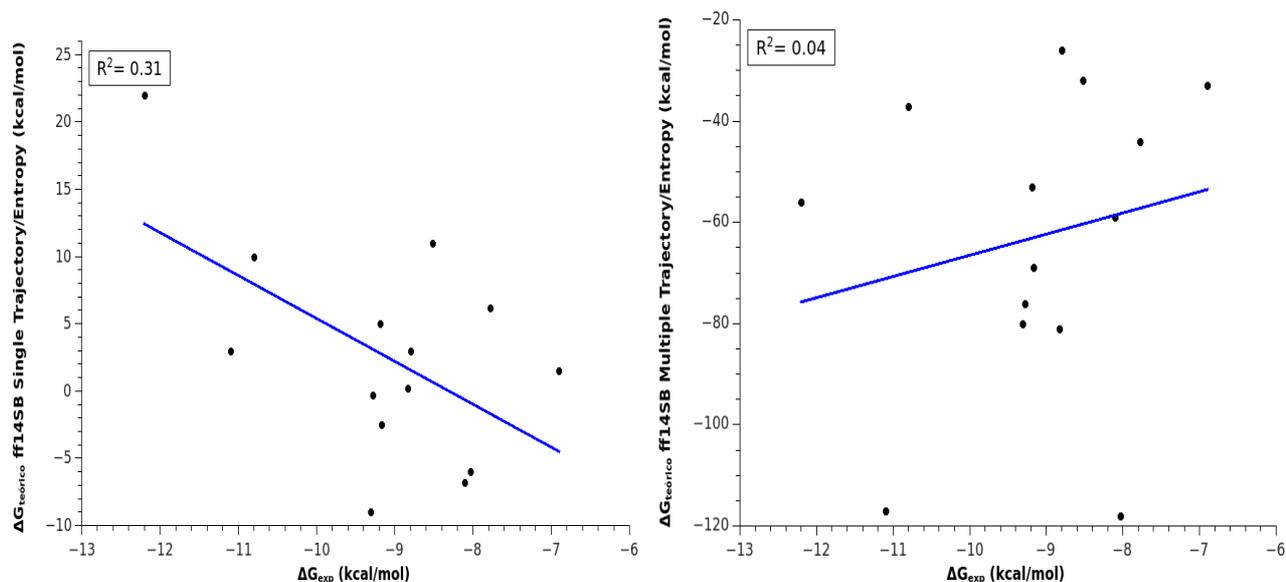


Figura 27: Grupo 3, diferença entre os cálculos levando em consideração a entropia com as abordagens de Single Trajectory e Multiple Trajectories.

#### 5.4.5 (QM)MM/GBSA

A última etapa dos testes envolvendo GBSA, compreendeu a predição de energia livre por métodos híbridos, associando o GBSA (método clássico) com QM (química quântica). Esta técnica permite selecionar a região a ser tratada com QM, podendo-se assim variar os resultados obtidos dependendo da região escolhida. O interesse deste trabalho foi avaliar a interação ligante-proteína, portanto a primeira abordagem consistiu em tratar o ligante na região quântica (abordagem usual). A segunda etapa, considerou o ligante+NADH (cofator) na região quântica.

##### 5.4.5.1 Ligante considerado na região de QM

Os valores preditos por (QM)MM/GBSA podem variar principalmente em relação aos métodos semi-empíricos utilizados e a região quântica selecionada (Su et al., 2015). O uso de potenciais híbridos nesta etapa, onde apenas o ligante é incluído na região QM, permite avaliar quais os efeitos das cargas parciais da região MM sobre a função de onda, portanto, levando em consideração os efeitos de rearranjo eletrônico (polarização). Foram avaliados quatro métodos: AM1, PM3, PM6 e RM1.

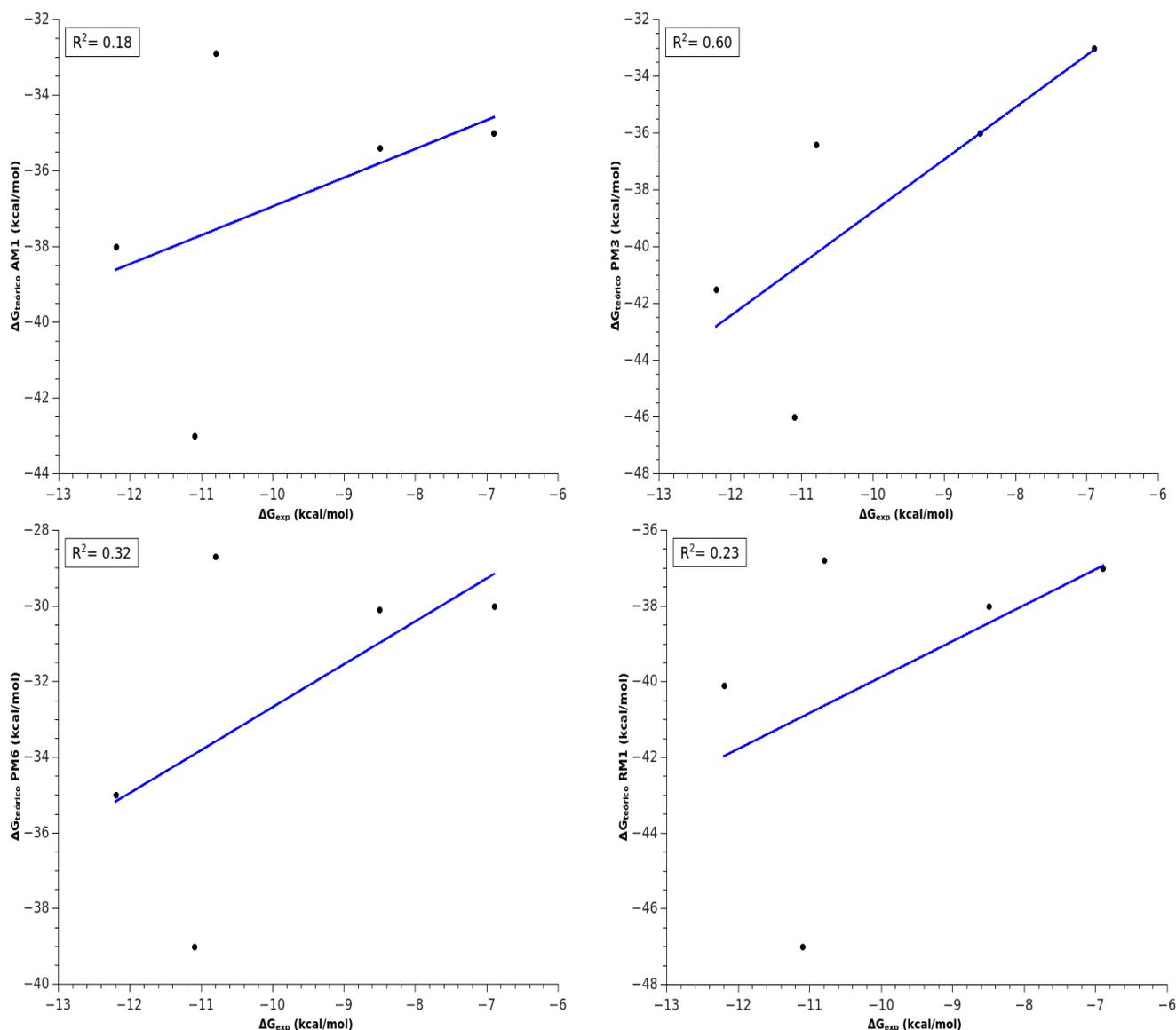


Figura 28: Grupo 1, Valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM. Correlação linear dos métodos (da esquerda para a direita) AM1, PM3, PM6 e RM1.

O Grupo 1 apresentou uma fraca correlação, com uma média geral dos  $R^2$  de 0,32, semelhante aos encontrados por GBSA sem o auxílio dos métodos quânticos. Com exceção para o uso do método PM3, que obteve um  $R^2$  de 0,60 (Figura 28). Assim como os resultados obtidos pelos métodos tradicionais de GBSA (utilizando-se de potenciais puramente clássicos), não houve correlação entre os valores obtidos para o Grupo 2 e 3 com o uso de potenciais híbridos (Figura 29 e Figura 30). Mesmo para o método PM3, que obteve um bom resultado para o Grupo 1, não houve a mesma performance quando comparado aos Grupo 2 e 3 deste estudo.

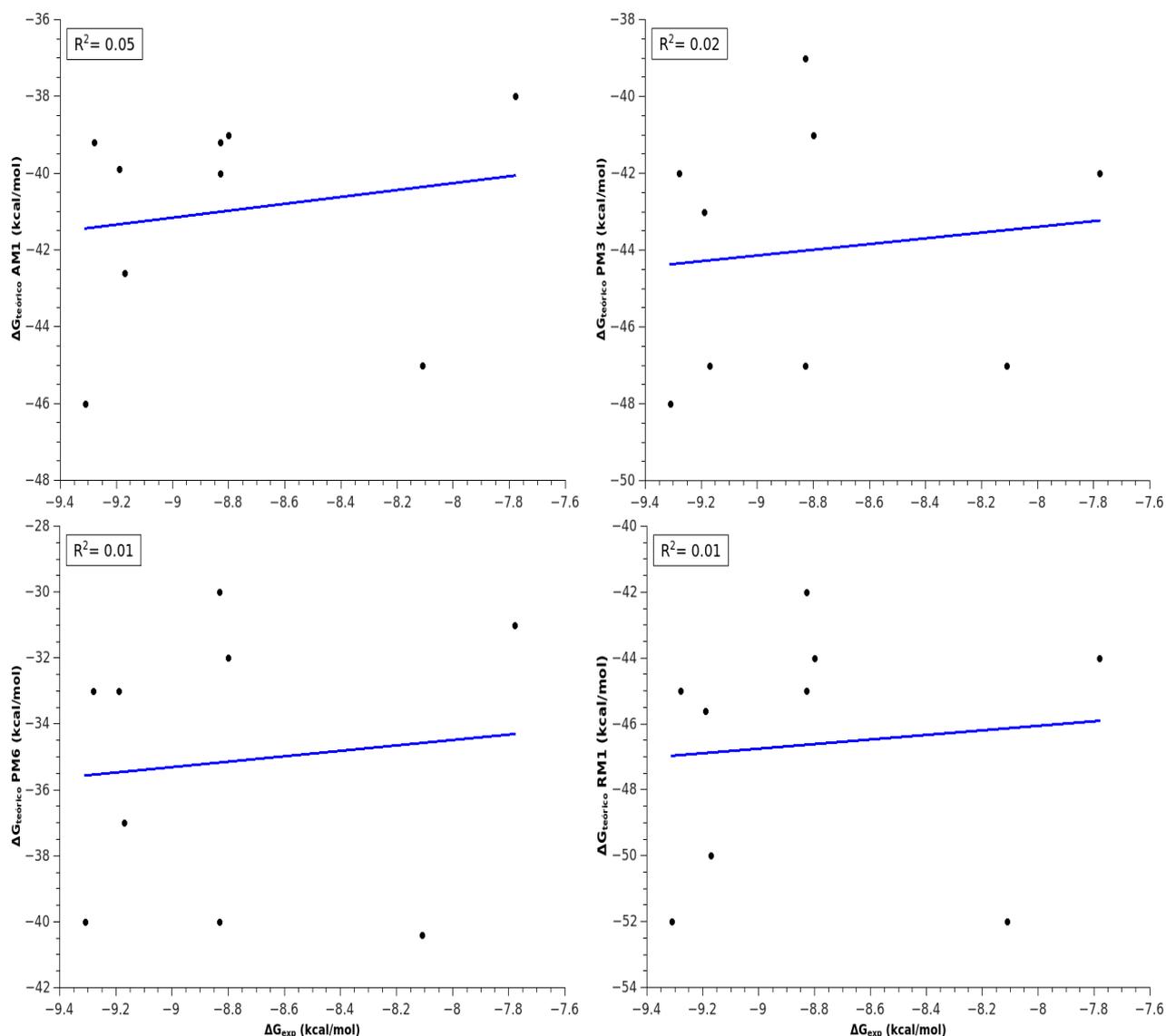


Figura 29: Grupo 2, valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM. Correlação linear dos métodos (da esquerda para a direita) AM1, PM3, PM6 e RM1.

Estes resultados mostram que a influência das cargas clássicas sobre a função de onda nos modelos semi-empíricos utilizados (efeito de polarização), não é significativa para diferenciar ligantes com estruturas e valores de energia livre de ligação, similares (Grupo 2).

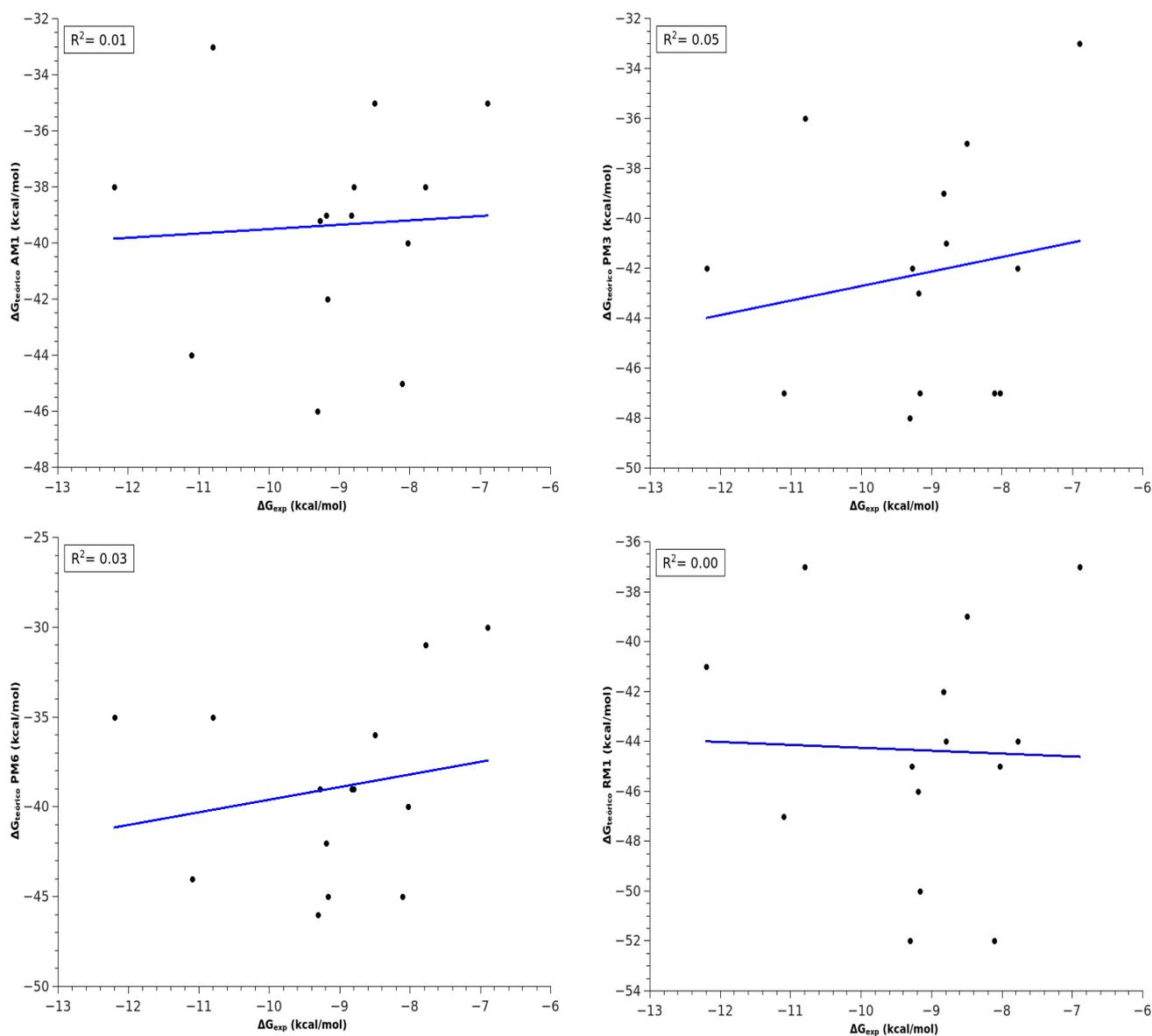


Figura 30: Grupo 3, valores preditos utilizando métodos híbridos e tratando o ligante na parte de QM. Correlação linear dos métodos ( da esquerda para a direita): AM1, PM3, PM6 e RM1.

#### 5.4.5.2 Ligante e NADH na região QM

Os mesmos quatro métodos semi-empíricos foram utilizados nesta segunda etapa, porém, a região tratada por química quântica envolveu o ligante e o cofator (NADH) da enzima InhA. Esta modificação tem como objetivo estimar a influência dos efeitos quânticos de interação entre o grupo de átomos vizinhos e não ligados no cálculo da energia livre.

Os resultados do Grupo 1 foram semelhantes aos obtidos pelo método anterior, tratando apenas o ligante na região quântica (fraca correlação  $R^2$  entre 0,20–0,30). Nesta abordagem o método PM3

não obteve resultados satisfatórios para o Grupo 1, resultando num  $R^2$  de 0,32. Porém para o Grupo 2 este obteve um valor melhor se comparado a etapa anterior (apenas ligante na região quântica) com um de  $R^2$  de 0,40. Com exceção do PM3, todos os outros resultados para o Grupo 2 não apresentaram uma boa correlação. O Grupo 3 não apresentou correlação em nenhum método.

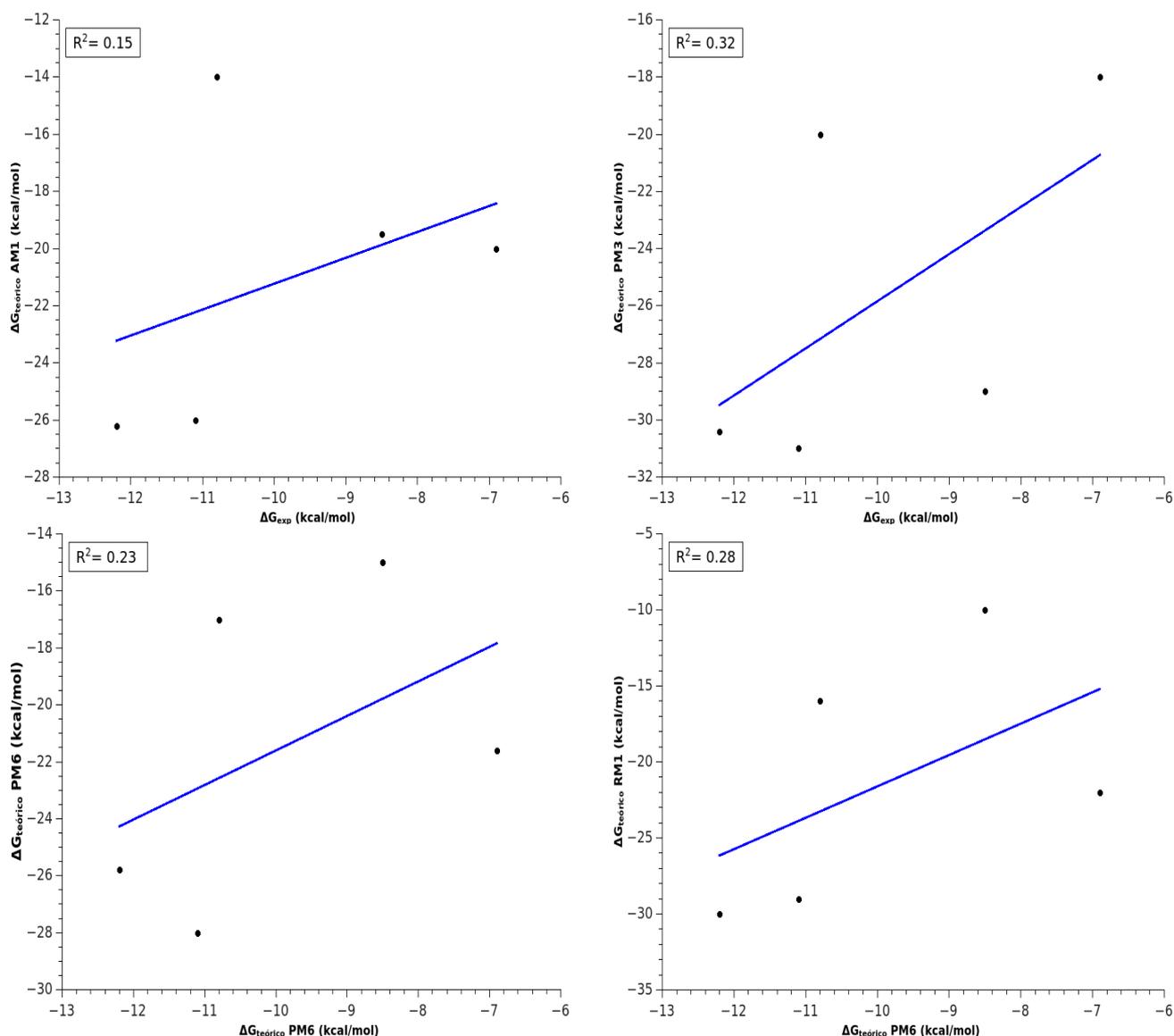


Figura 31: Grupo 1, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM. Correlação linear dos métodos (da esquerda para a direita) AM1, PM3, PM6 e RM1.

Apesar de ser uma ferramenta considerada mais acurada para determinar a energia livre de ligação *in silico*, os valores obtidos por QM associado ao GBSA neste estudo, não apresentaram vantagem significativa em relação aos métodos puramente clássicos.

Os resultados obtidos pelo método PM3, se mostraram ligeiramente superiores as outras três abordagens semi-empíricas. Apesar da pequena vantagem, não ocorreu uma melhora consistente nos dados, pois na primeira fase o Grupo 1 e o Grupo 2 apresentaram um  $R^2$  de 0,60 e 0,02, respectivamente. Já na segunda fase, o Grupo 1 e o Grupo 2 apresentaram um  $R^2$  de 0,32 e 0,40, respectivamente.

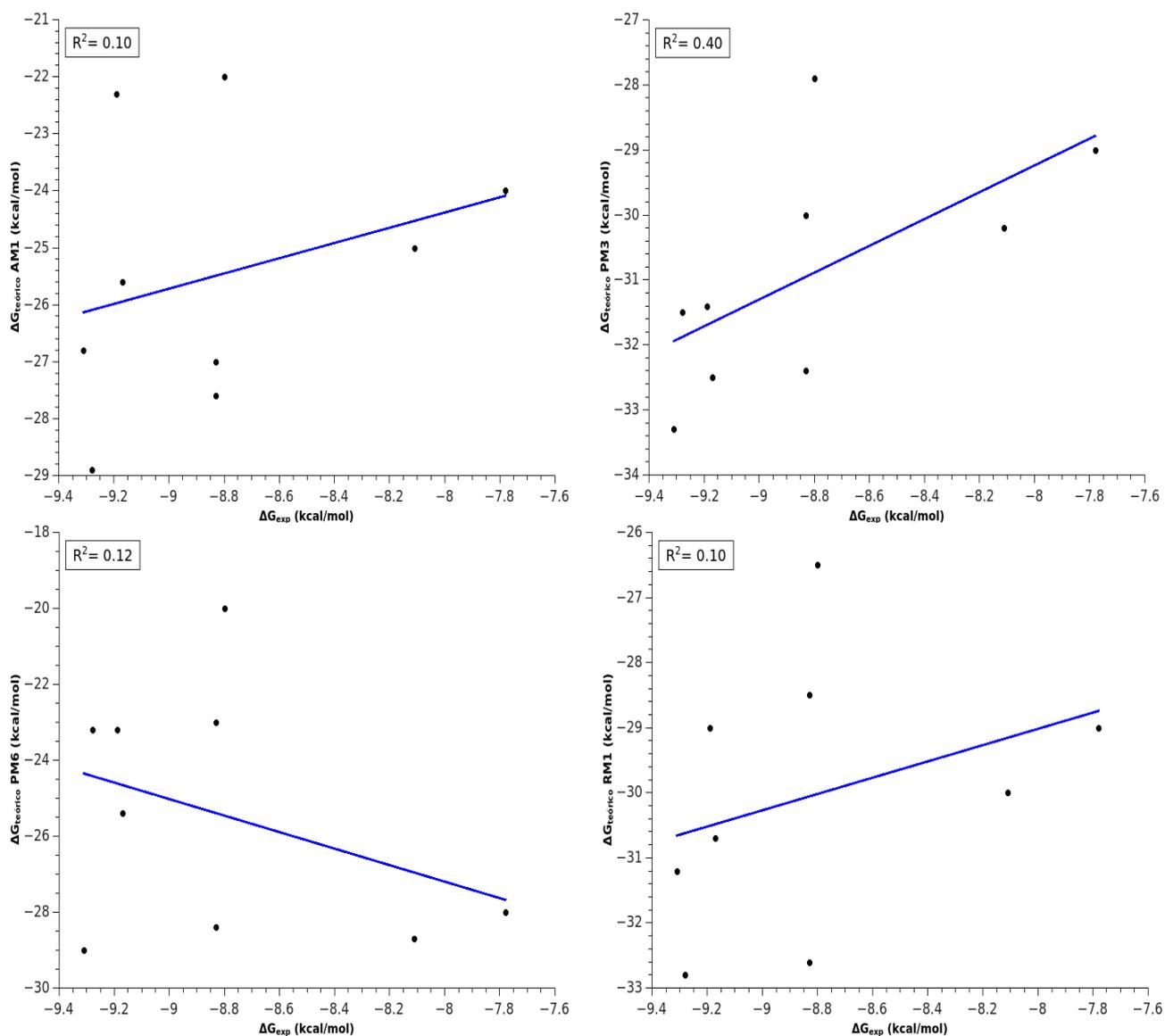


Figura 32: Grupo 2, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM. Correlação linear dos métodos (da esquerda para a direita) AM1, PM3, PM6 e RM1.

Em relação a média do Grupo 1, esta abordagem obteve um  $R^2$  de 0.24, inferior aos testes realizados tratando apenas o ligante na parte quântica. Portanto, não existe uma correlação direta entre a área quântica, e a melhora nos resultados de determinação de energia livre de ligação.

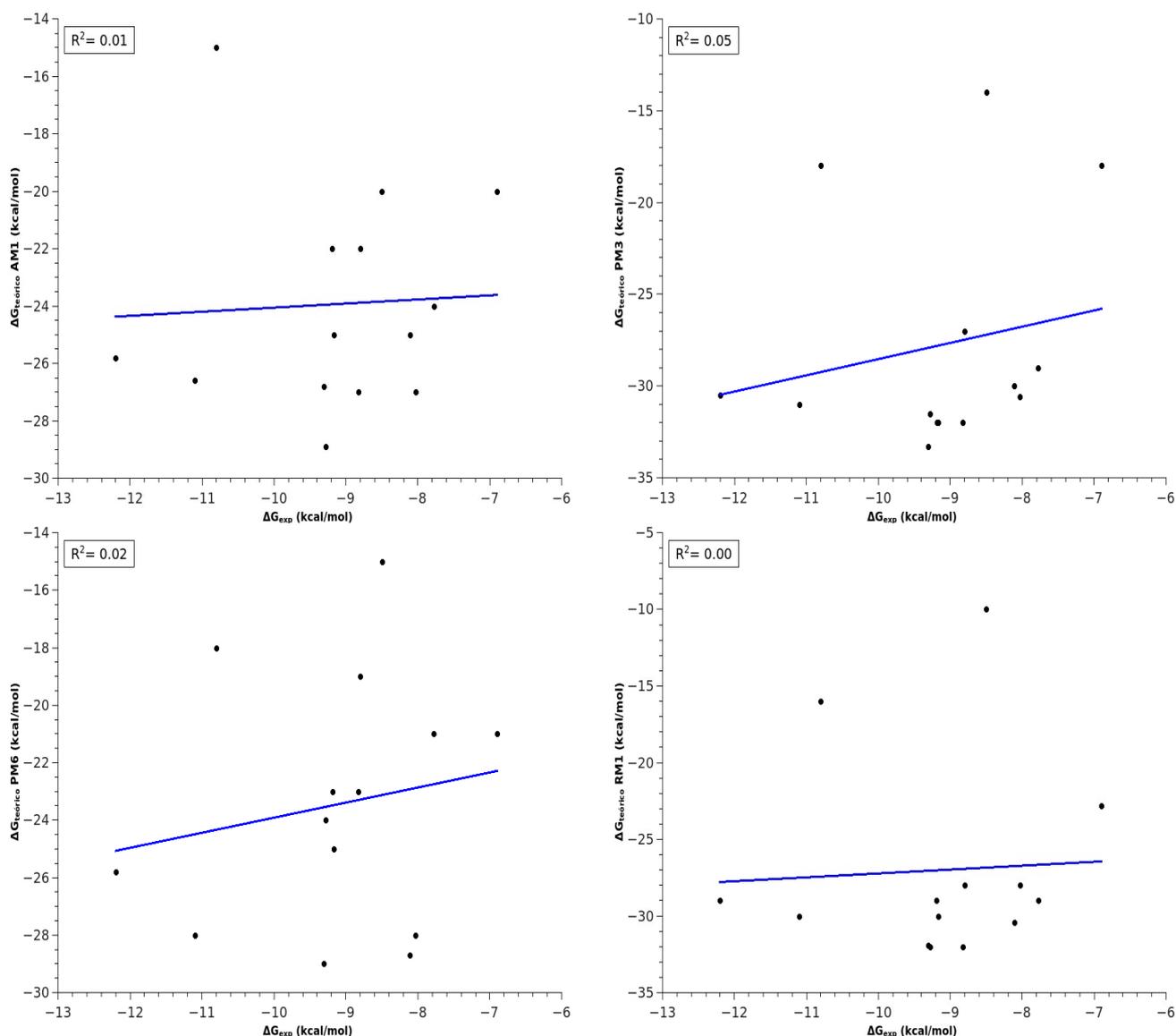


Figura 33: Grupo 3, valores preditos utilizando métodos híbridos e tratando o ligante e o NADH na parte de QM. Correlação linear dos métodos (da esquerda para a direita) AM1, PM3, PM6 e RM1.

### 5.4.5.3 Considerações parciais das abordagens utilizando o método GBSA

De acordo com Genheden e colaboradores, o sucesso e popularidade das técnicas envolvendo mecânica molecular combinada com *Generalized Born* são superestimados. Alguns parâmetros continuam questionáveis quanto a sua acurácia, estando entre eles, os valores de entropia e a contribuição dos valores de energia livre de moléculas de água no sítio de ligação. Aliado a estes parâmetros, a performance varia significativamente dependendo do sistema testado (Genheden et al., 2015b). Conforme observado no presente estudo, e reforçando as ideias propostas por Genheden e colaboradores, as tentativas de refinar os valores utilizando mecânica quântica e modelos

otimizados de solvatação, nos campos de força, não contribuíram significativamente com os resultados (Genheden et al., 2015). Mesmo com o auxílio de métodos quânticos na região do ligante, não foi possível descrever com acurácia os valores absolutos para a energia livre. Entretanto, um dos fatores que podem ter influenciado nos resultados é a proximidade dos valores de  $\Delta G$  experimentais do conjunto de compostos utilizados nas simulações. Outro fato inerente, se deve a flexibilidade da alça de ligação da enzima InhA. Esta alta flexibilidade pode influenciar de maneira negativa o desempenho dos cálculos de predição de energia livre.

## 5.5 RESULTADOS MM/PBSA

Os cálculos de MM/PBSA foram realizados pelo programa *MMPBSA.py* integrante do pacote AmberTools14. O *workflow* estabelecido para o MM/PBSA consistiu em duas partes: (i) variação nos tempos de simulação e (ii) variação nos campos de força, similar aos testes realizados com MM/GBSA. Neste estudo foi possível verificar que o método MM/PBSA foi em média 6x mais custoso computacionalmente que o método MM/GBSA e os resultados absolutos se aproximaram mais dos valores experimentais.

### 5.5.1 Tempo de simulação em MM/PBSA

Utilizando a mesma metodologia aplicada ao método MM/GBSA, foi testada a relação do tamanho de simulação com os valores preditos por PBSA. Os valores obtidos foram calculados por MM/PBSA utilizando o campo de força ff99, conforme a Tabela 5. O composto TCL obteve uma média de energia livre para os tempos de simulação testados de  $-0,7369 \pm 0,0573$  kcal/mol. O composto 8PS obteve uma média de  $-4,6636 \pm 0,2336$  kcal/mol. A partir desta análise, observou-se que o tempo de simulação não interferiu nos valores de energia livre para os compostos 8PS e TCL. Conforme mencionado por Hou e colaboradores, o tempo de simulação não interfere significativamente nos resultados preditos de energia livre de ligação pelo método PBSA (Hou et al., 2011).

Figura 5: Diferenças de  $\Delta G$  calculadas pelo método MM/PBSA para os ligantes 8PS e TCL. **Resultados em kcal/mol.**

Tempo	8PS	TCL
1 ns	-4.2525	-0.6321
2 ns	-4.6854	-0.7654
4 ns	-4.7845	-0.7031
5 ns	-4.9854	-0.8456
10 ns	-4.7712	-0.8034
50 ns	-4.5001	-0.7321

A tendência dos resultados obtidos variando o tempo de simulação em MM/PBSA foram semelhantes aos encontrados no MM/GBSA. O tempo de 2 ns foi o escolhido para o restante da metodologia, concordando com o tempo estipulado anteriormente para o GBSA e o tempo preconizado pela literatura na área (Xu et al., 2013; Hou et al., 2011; Kumari et al., 2014).

### 5.5.2 Campos de Força

Devido ao maior tempo computacional do PBSA comparado ao GBSA, foram testados quatro campos de força do pacote AMBER, o ff03, ff99, ff99SB e o ff14SB. Para o Grupo 1, os resultados de PBSA foram semelhantes aos resultados de GBSA. Os campos de força ff99, ff99SB, ff03 e ff14SB obtiveram valores de  $R^2$  de 0,59, 0,38, 0,00, e 0,51 respectivamente, como demonstrado na Figura 34.

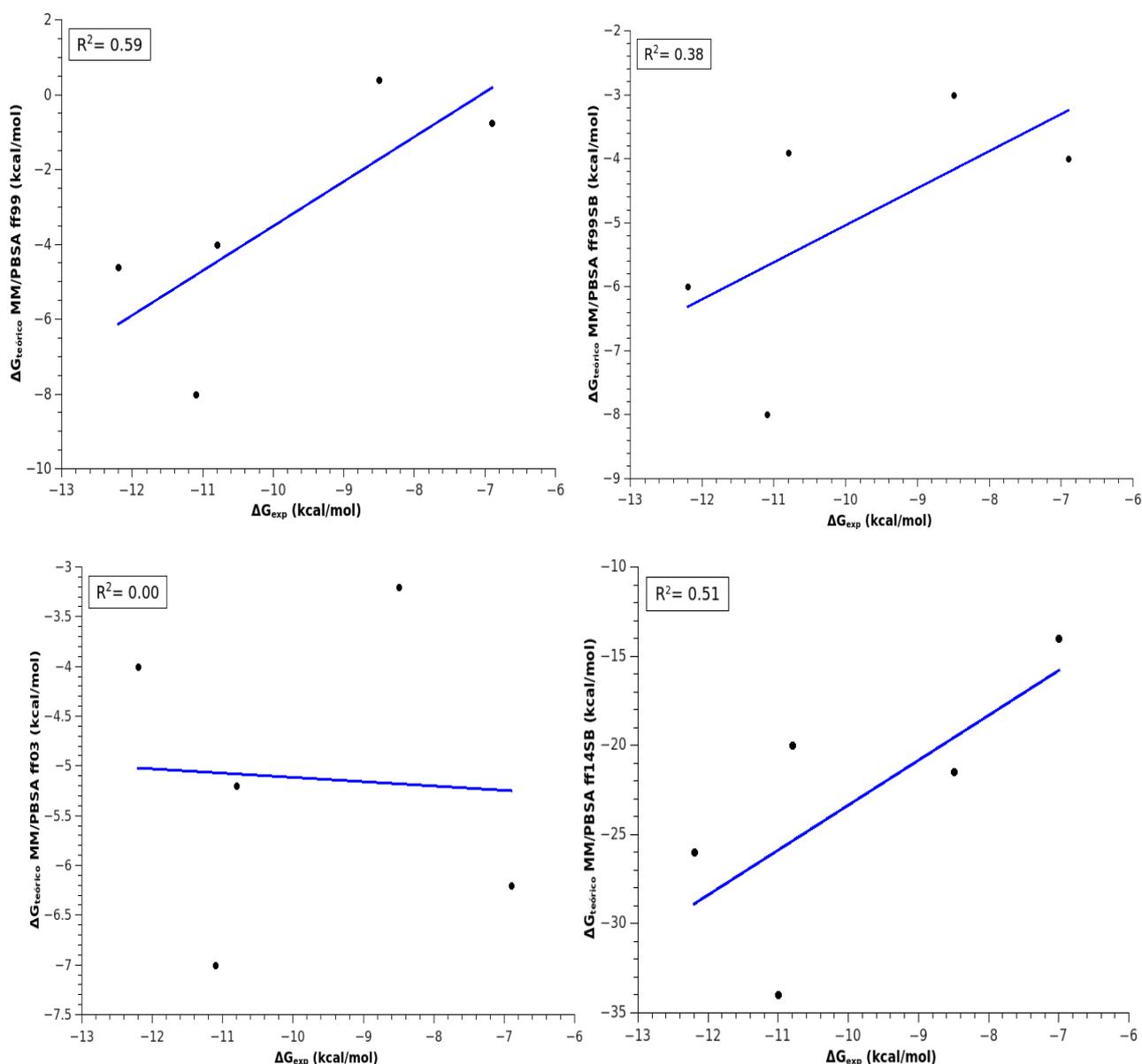


Figura 34: Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 1.

No Grupo 2 o ff99SB e o ff14SB apresentaram uma fraca correlação com  $R^2$  de 0,21 e 0,24 respectivamente (Figura 35). O Grupo 3 não apresentou correlação conforme Figura 36. Estes resultados demonstraram que apesar do incremento no tempo computacional, os valores de PBSA não se mostraram superiores quando comparados ao GBSA. Também foi possível observar que para o Grupo 2 e principalmente 3, o PBSA foi incapaz de prever a tendência correta dos compostos, devido em grande parte as suas semelhanças nos valores experimentais.

Um fato interessante que ocorreu no PBSA é que alguns valores preditos se mostraram positivos conforme indicado pelo ff99 e ff99SB nos Grupo 1, 2 e 3. Este evento não era esperado pois todos compostos testados apresentam valores negativos de energia livre de ligação.

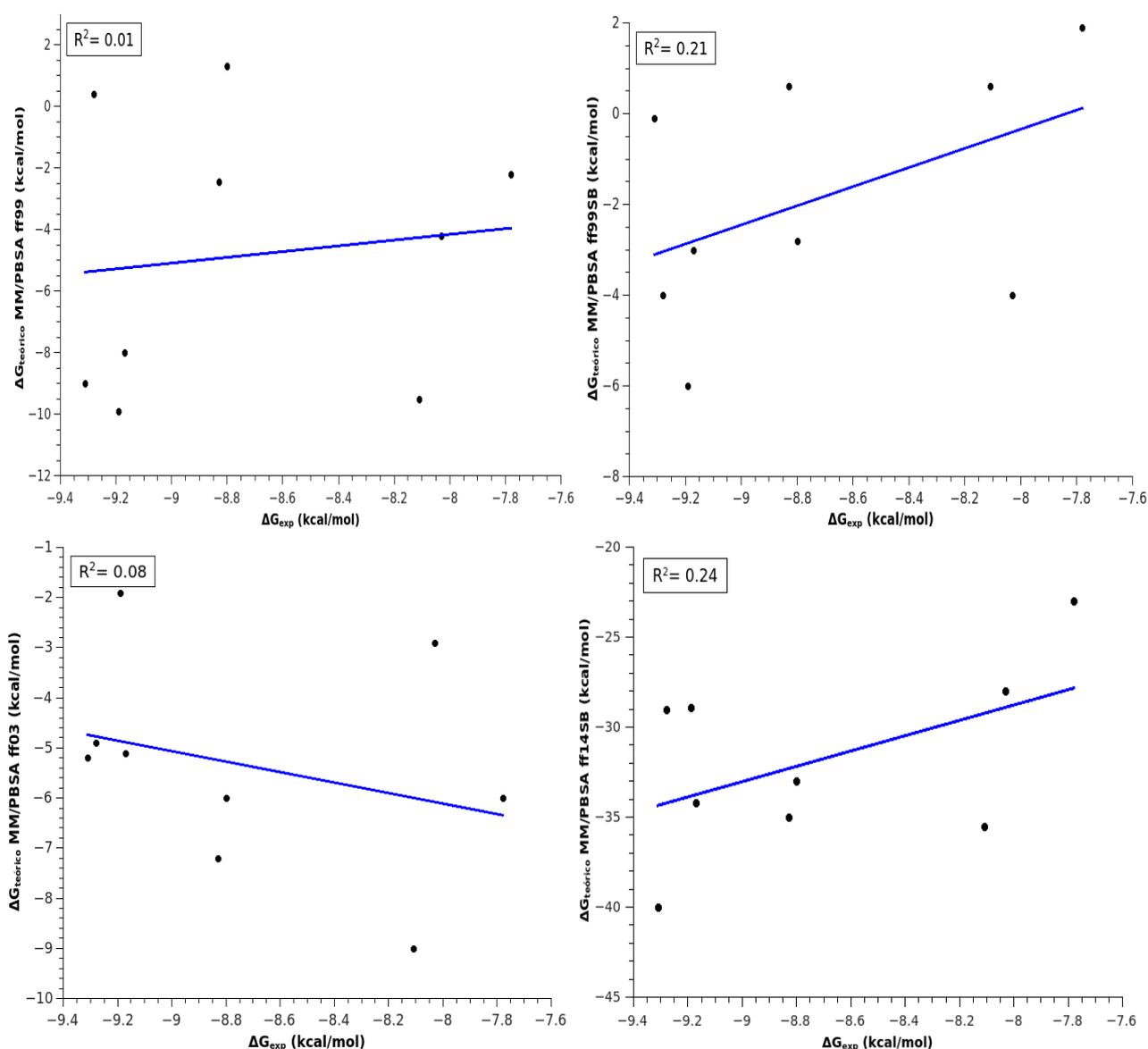


Figura 35: Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 2.

Na abordagem por PBSA não foram modificados os raios atômicos, visto que estes não apresentam influência nos resultados de PB (Su et al., 2015 ; Wang, et al., 2001). A média geral dos  $R^2$  para o Grupo 1 foi 0,37, similar a encontrada para o MM/GBSA e o (QM)MM/GBSA. Os dois métodos baseados em campos de força utilizados no presente trabalho demonstraram performances semelhantes conforme indicado pela média do Grupo 1.

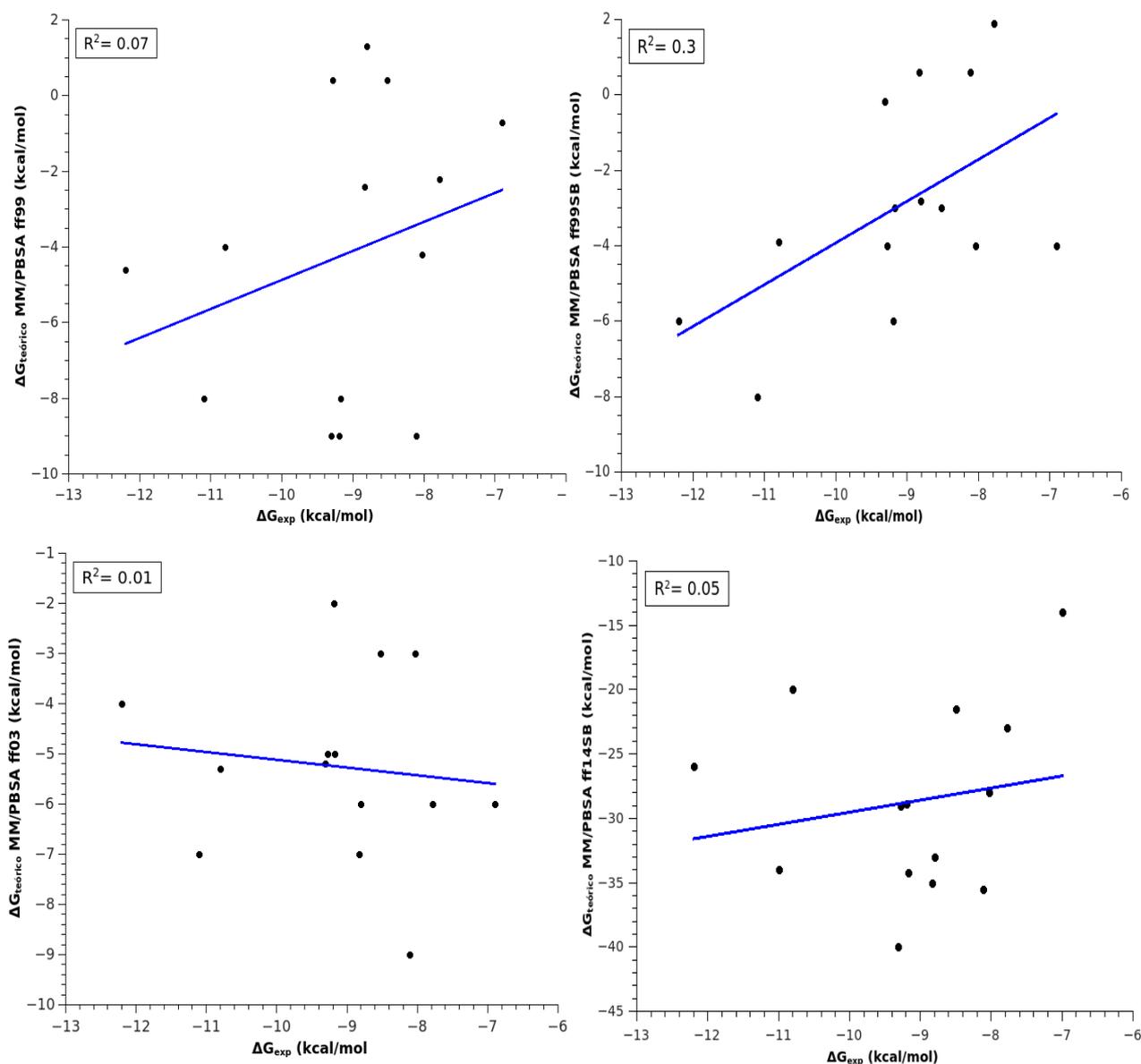


Figura 36: Resultados preditos pelo método MM/PBSA utilizando os campos de força ff99, ff99SB e ff03 e ff14SB para as moléculas que compõem o Grupo 3.

## 5.6 RESULTADOS AUTODOCK

A função *scoring* do programa AutoDock quando comparado ao GBSA e especialmente ao

PBSA é mais veloz computacionalmente. Ao contrário dessas outras duas metodologias, o AutoDock apresenta um valor para cada *snapshot* da simulação separado. O valor escolhido como referência neste estudo, foi baseado nos valores dos 400 *snapshots*, com base na técnica de histogramas, levando em consideração o resultado mais populado (*cluster*), apresentado para cada composto (Anexo 4). Foram testados dois tipos de cargas para a parametrização dos átomos nesta fase, Gasteiger e a AM1-BCC (utilizada pelo programa AMBER), para fins de comparação. Os resultados avaliados entre cada *snapshot* permitiram observar que, em alguns casos, ocorreu uma diferença de até 20%. Este dado corrobora com a não utilização da média aritmética como base dos resultados finais.

Os valores obtidos pelo AutoDock (-7 kcal/mol até -13 kcal/mol) em números absolutos (kcal/mol) ficaram próximos dos obtidos nos valores experimentais. Com relação aos resultados obtidos com o AutoDock, não houve diferenças significativas nos cálculos de energia livre quando comparados os tipos de cargas utilizados (Gasteiger para AM1-BCC). Os valores apresentados com demonstraram que este pode ser utilizado para determinar tanto a tendência como a predição dos valores absolutos.

Os valores de  $R^2$  para os Grupo 1 (Figura 37) e 3 (Figura 39) apresentaram uma fraca correlação (0,27 e 0,25, respectivamente) com os valores experimentais, utilizando ambos tipos de carga.

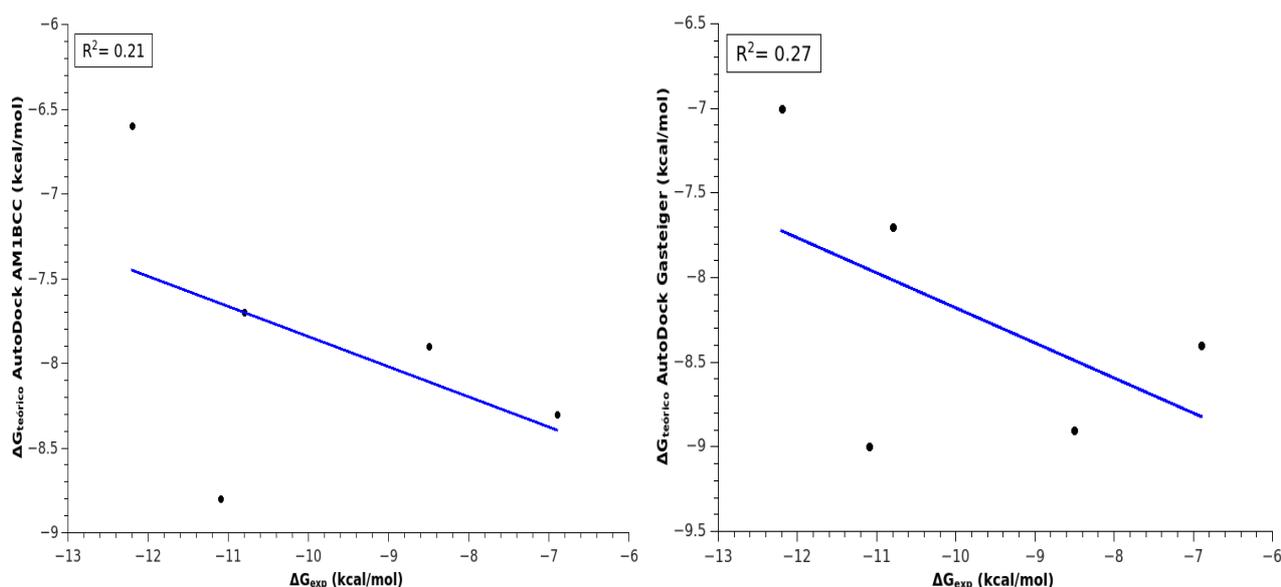


Figura 37: Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 1.

Conforme Figura 38 o Grupo 2 não apresentou correlação. O programa AutoDock também não foi capaz de prever a tendência dos compostos com valores e estrutura similares, demonstrando uma possível limitação desta técnica. A média para os  $R^2$  do Grupo 1 foi de 0,25, similar a encontrada com o MM/GB(PB)SA e o QM/MM-GBSA.

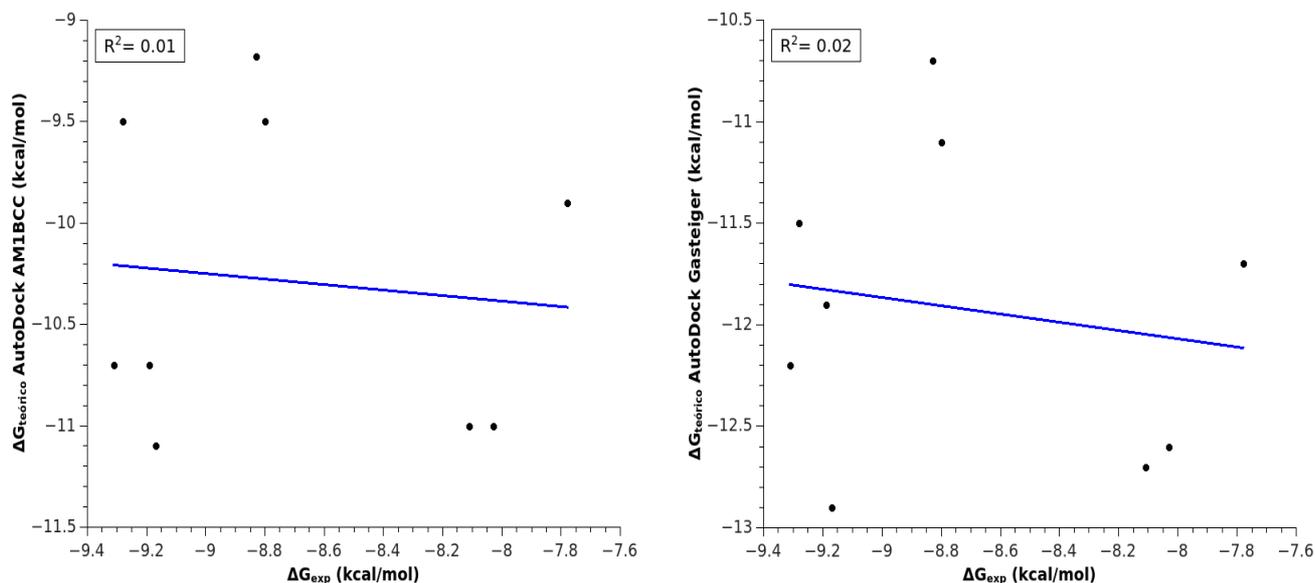


Figura 38: Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 2.

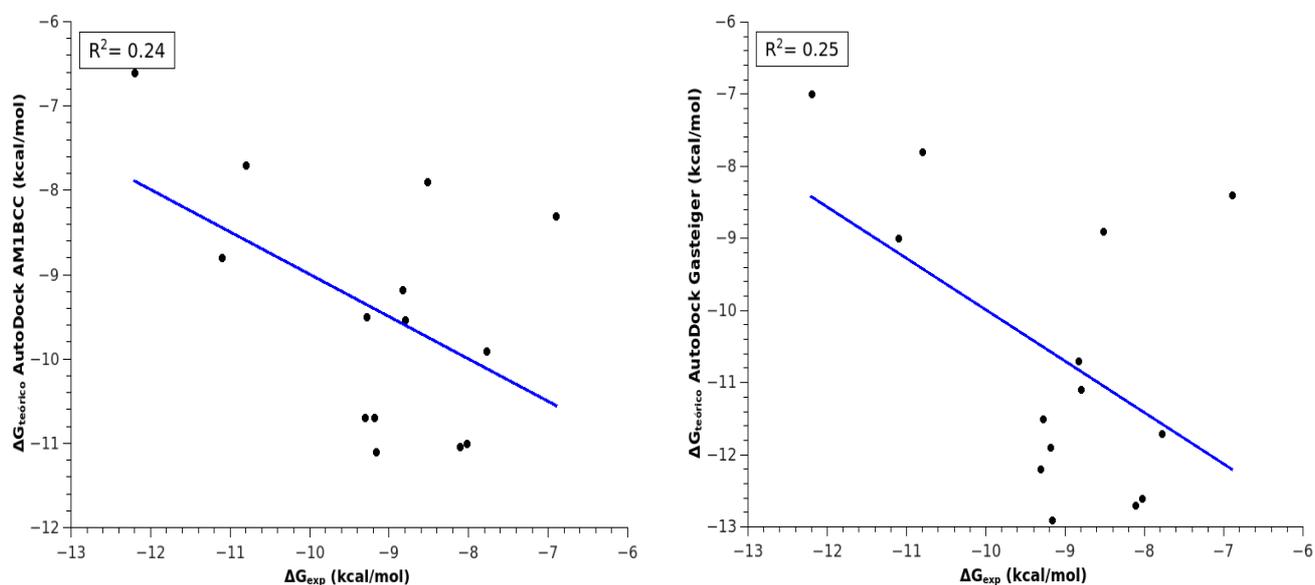


Figura 39 : Resultados preditos pelo AutoDock utilizando cargas AM1BCC e Gasteiger para as moléculas que compõem o Grupo 3.

## 5.7 RESULTADOS LIGSCORE

Os valores obtidos pelo LigScore não são estimados em kcal/mol, e sim em uma função

*scoring* desenvolvida para o próprio programa. Por este motivo só é possível comparar a tendência entre os grupos testados e não os valores absolutos dos mesmos. Comparado com as outras metodologias apresentadas, o LigScore tem uma grande vantagem em relação ao tempo computacional. Os cálculos são realizados em poucos minutos, podendo assim um grande número de moléculas serem testadas simultaneamente (a limitação consiste na amostragem por DM, que se torna um passo mais demorado e refinado). Assim como o AutoDock, o programa LigScore calcula um resultado de energia livre para cada *snapshot*. O tempo de simulação de 2 ns (400 snapshots) foi adotado como padrão, com base nos estudos realizados para o GB/PBSA

O Grupo 1 obteve moderada correlação com um  $R^2$  de 0,44 (Figura 40), similar a métodos mais robustos como o QM/MM-GBSA. Para os testes realizados com o Grupo 2 e Grupo 3 foi observado conforme Figura 41 e Figura 42 que não houve correlação entre os dados utilizando a regressão linear como método comparativo. Os resultados para o Grupo 2, demonstram que o LigScore não foi coerente nas pequenas variações apresentadas nas moléculas, resultando em uma tendência distinta da apresentada experimentalmente.

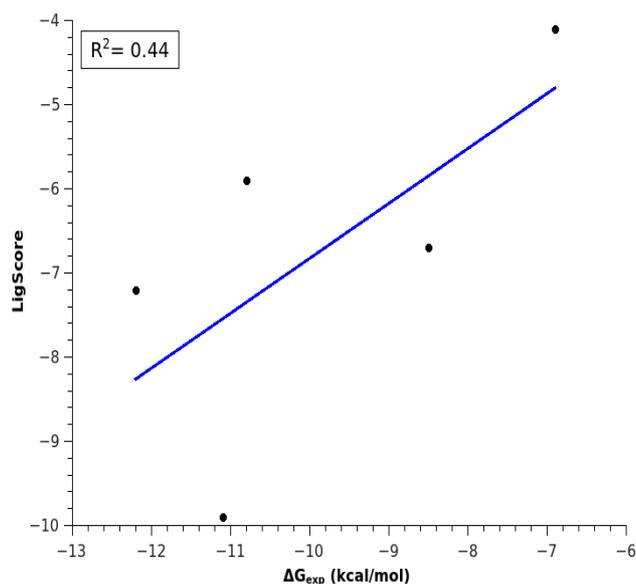


Figura 40: Resultados preditos pelo LigScore para as moléculas que compõem o Grupo 1.

Apesar das limitações, o programa LigScore pode ser indicado para uma triagem inicial de compostos não similares estruturalmente, onde se demanda um resultado rápido e um maior número de compostos analisados é necessário.

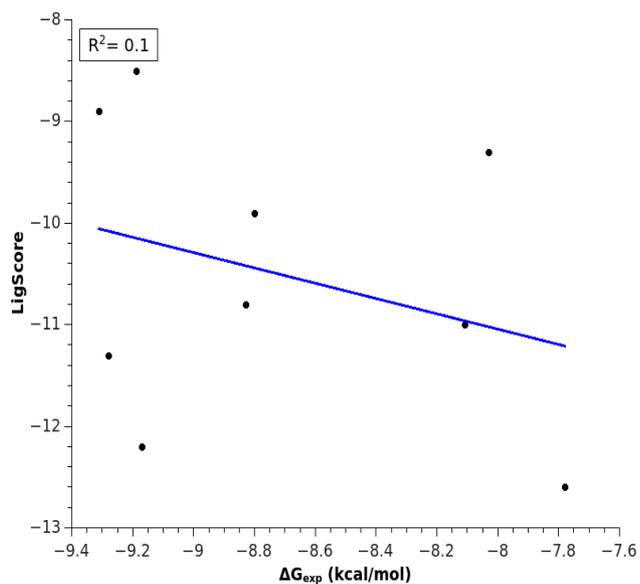


Figura 41: Resultados preditos pelo LigScore para as moléculas que compõem o Grupo 2.

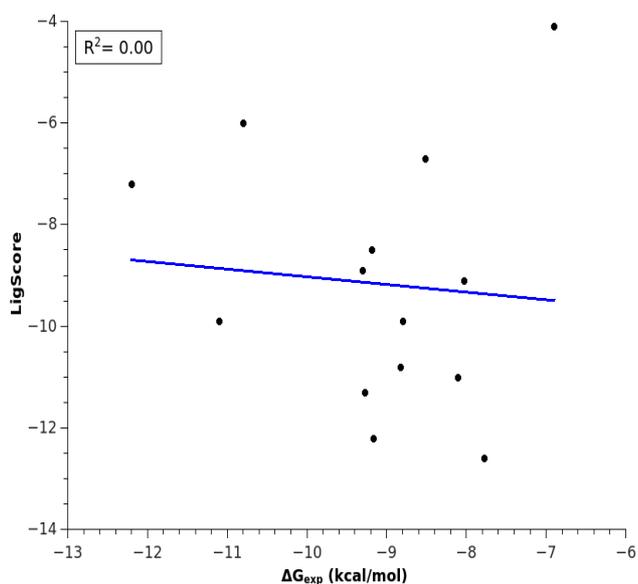


Figura 42: Resultados preditos pelo LigScore para as moléculas que compõem o Grupo 3.

## 5.8 RESULTADOS DRUGSCORE

Assim como o LigScore, o DrugScore também utiliza funções de *scoring* baseadas em conhecimento (Knowledge-based). Quando comparadas as duas metodologias, o DrugScore obteve resultados ligeiramente superiores. O gasto computacional entre os dois métodos foi similar, não ocorrendo vantagem neste quesito.

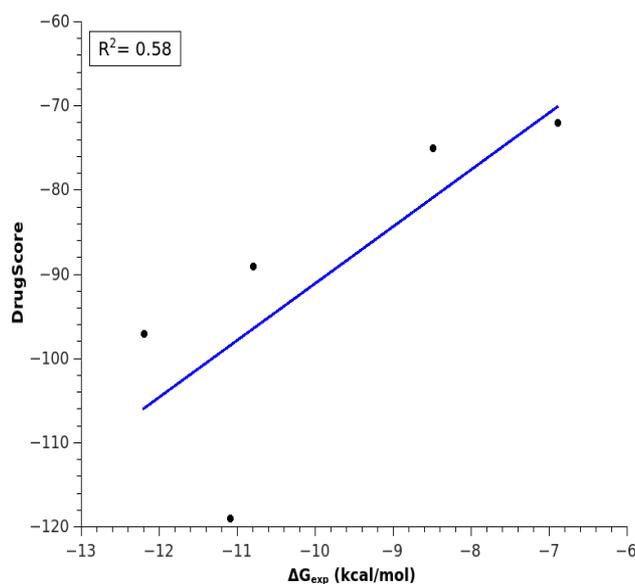


Figura 43 : Resultados preditos pelo DrugScore para as moléculas que compõem o Grupo 1.

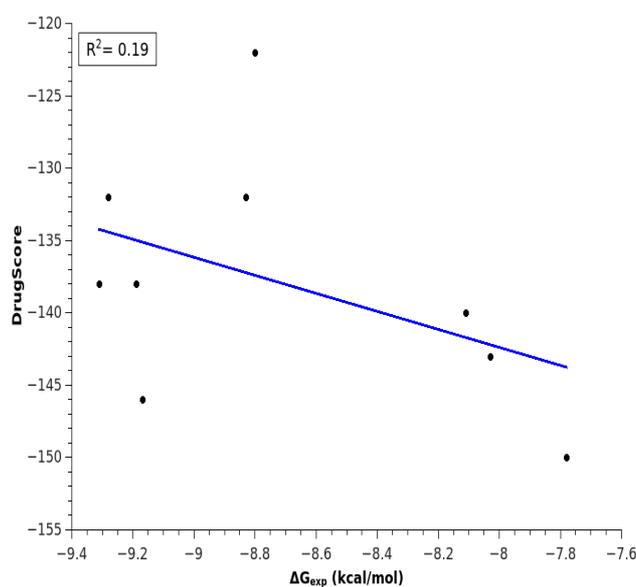


Figura 44 :Resultados preditos pelo DrugScore para as moléculas que compõem o Grupo 2.

O Grupo 1 obteve uma correlação de moderada à alta com um  $R^2$  de 0,58 (Figura 43), um pouco superior a encontrada pelo LigScore. Este valor de correlação encontra-se dentro da média encontrada em métodos mais sofisticados como o GBSA (média global dos  $R^2$  de 0,60). Com relação ao valor obtido para o Grupo 1, o DrugScore foi superior a maioria dos campos de força testados tanto no GBSA como no PBSA e aos métodos híbridos de QM/MM-GBSA.

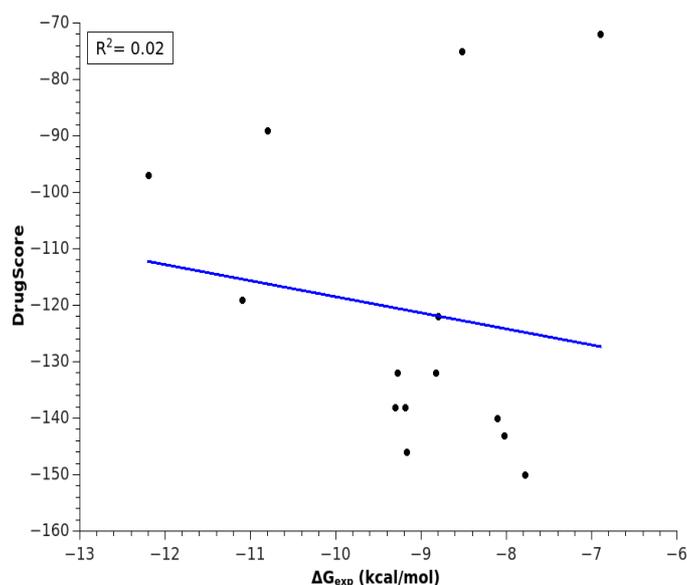


Figura 45 : Resultados preditos pelo DrugScore para as moléculas que compõem o Grupo 3

O Grupo 2 obteve uma correlação fraca com um  $R^2$  de 0,19 (Figura 44) e o Grupo 3 (Figura 45) não obteve correlação com os dados experimentais. Assim como os resultados obtidos pelo LigScore, e as técnicas de campos de força, o DrugScore não foi preciso para ranquear compostos muito próximos entre si. Porém, se mostrou útil para ranquear compostos não similares estruturalmente e com um custo computacional baixo.

## 5.9 RESULTADOS SQM

Comparada as outras metodologias utilizadas neste trabalho, o SQM demandou mais tempo computacional, na ordem de 16 horas por ligante para calcular os valores de energia livre. Ao contrário dos métodos híbridos combinados com GBSA utilizados no (QM)MM/GBSA, esta técnica leva em consideração os cálculos apenas da região quântica, desconsiderando a parte da proteína que não estiver nesta área específica. Foram testados quatro métodos de QM: PM3, PM6-DHX, PM6-DH+ e PM7. A abordagem escolhida para esta técnica foi considerar o ligante e o espaço até 4 Å de distância do mesmo na região quântica (átomos que estão à 4 Å)

Para o Grupo 1 (Figura 46), no método PM3 foi observado um  $R^2$  de 0,60, idêntica a ocorrida em (QM)MM/GBSA para o PM3. Para os outros métodos do Grupo 1 não houve correlação. Este fato demonstra a grande variância entre os valores obtidos quando comparados os métodos semi-empíricos (similar ao que ocorre com os campos de força em MM-GB/PBSA).

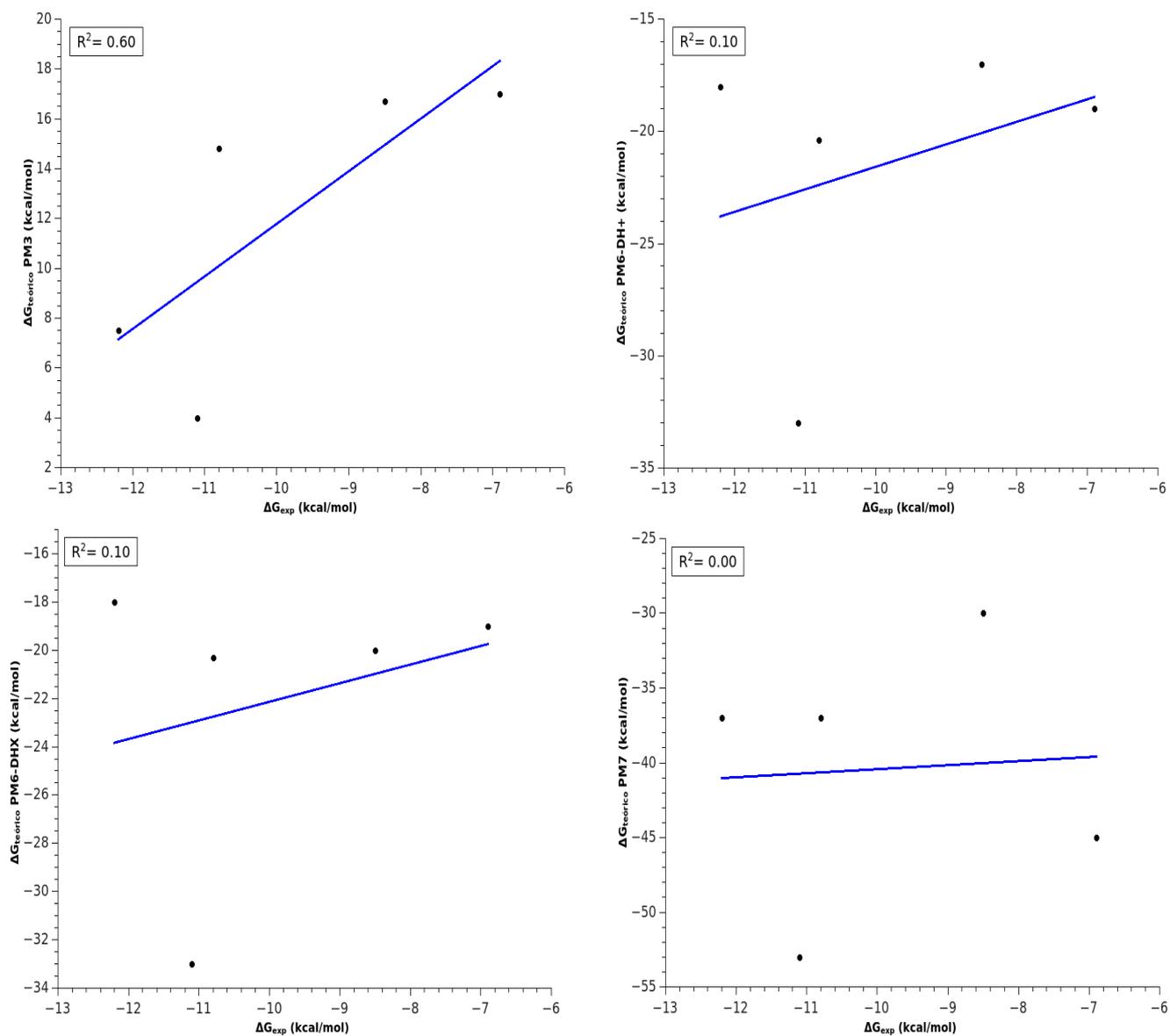


Figura 46: Comparação entre métodos de SQM para o Grupo 1. Da esquerda para a direita: PM3, PM6-DH+, PM6-DHX e PM7.

Os métodos mais acurados de SQM não foram capazes de distinguir corretamente as pequenas diferenças dos valores de energia livre para o Grupo 2 (Figura 47) e para o Grupo 3 (Figura 48). Os resultados de SQM neste trabalho não foram superiores a metodologias menos refinadas como LigScore e AutoDock, por exemplo.

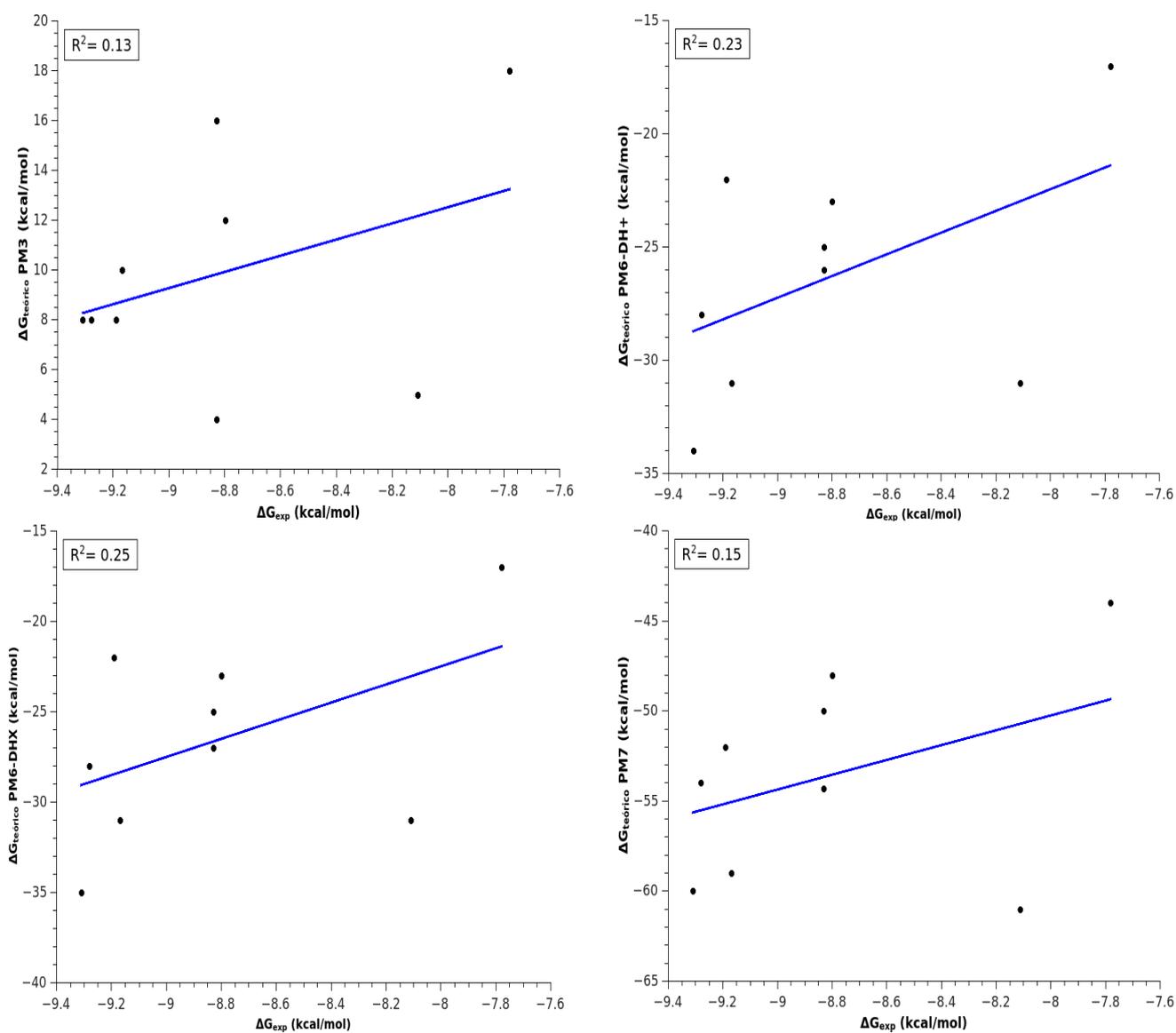


Figura 47: Comparação entre métodos de SQM para o Grupo 2. Da esquerda para a direita: PM3, PM6-DH+, PM6-DHX e PM7.

Entre os quatro métodos quânticos testados, o PM3 obteve uma fraca correlação no Grupo 3. Assim como no QM/MM-GBSA, o PM3 obteve vantagem em relação aos outros métodos, principalmente no Grupo 1. Com estes resultados nas duas abordagens, é possível supor que o PM3 se mostre superior aos outros métodos quânticos estudados levando em consideração o complexo InhA e seus ligantes.

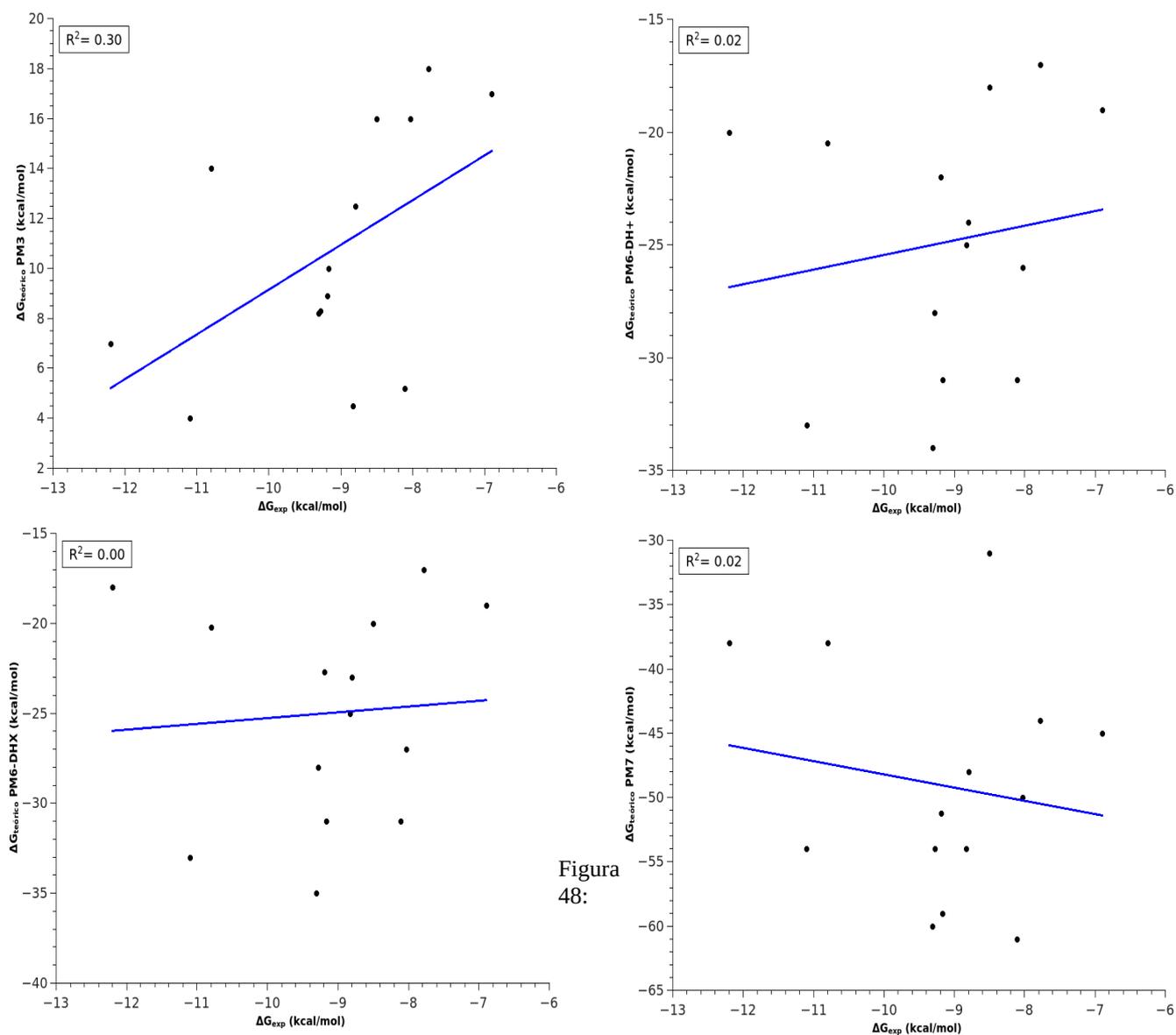


Figura 48:

Figura 46: Comparação entre métodos de SQM para o Grupo 3. Da esquerda para a direita: PM3, PM6-DH+, PM6-DHX e PM7.

## 6. COMPARAÇÃO ENTRE METODOLOGIAS

No que tange as metodologias testadas neste estudo, foi possível observar que o Grupo 1 apresentou os melhores resultados gerais. Este fato pode ser associado com a maior distância dos valores experimentais de  $\Delta G$  entre os compostos. Esta característica possibilitou um melhor ranqueamento dos ligantes. Os melhores resultados encontrados foram com o campo de força ff99 para o MM/GBSA tanto com os raios atômicos padrão do AMBER ( $R^2$  de 0,70), quanto para os raios otimizados ( $R^2$  de 0,80). Nas técnicas quânticas (QM/MM-GBSA e SQM), os melhores resultados foram provenientes do método PM3 ( $R^2$  de 0,60). Quando analisadas técnicas

consideradas menos refinadas, como o LigScore e o DrugScore, foi possível observar valores moderados de correlação para o Grupo 1 ( $R^2$  entre 0,40 e 0,60). A partir destes estudos, para o Grupo 1 foi possível verificar que todas as metodologias utilizadas obtiveram correlações moderadas, dependendo do parâmetro utilizado. É possível supor que compostos com valores experimentais mais distantes entre si, resultem em correlações melhores de predição de energia livre, pois os métodos testados conseguem diferenciá-los melhor.

Nenhuma das técnicas utilizadas neste trabalho apresentou uma correlação satisfatória (na faixa de  $R^2$  de 0,60) para os Sets 2 e 3. Este fato se deve principalmente aos valores de  $\Delta G$  experimentais serem semelhantes (variação máxima de -1.53 kcal/mol para o Grupo 2). Estas pequenas flutuações não puderam ser detectadas pelos métodos abordados no presente trabalho. Este resultado demonstra que é preciso cautela quando se associa técnicas de *Drug Design* e determinação de energia livre sem calibração prévia. Nestes casos, as predições de energia *in silico* podem acarretar erros de determinação, quando são variados radicais em uma mesma molécula.

A complexidade do presente trabalho utilizando a enzima InhA como base foi um fator impactante nos valores obtidos. Os estudos envolvendo a determinação de energia livre apresentam resultados variados, principalmente no que diz respeito a escolha da proteína alvo. A flexibilidade, especialmente na região da alça de ligação, da enzima InhA, pode ser apontada como um fator decisivo para os resultados obtidos. Segundo estudos prévios, receptores menos flexíveis apresentam menor variação e conseqüentemente menor interferência nas determinações de energia livre (Wang et al., 2001).

Outro fato importante observado, foi o fato de que pequenos desvios na tendência da reta, de apenas um composto (*outlier*), por exemplo, acarretam um decréscimo significativo na correlação pelo método de regressão linear. Na Tabela 6 estão listados os valores de correlação obtidos neste trabalho, para o campo de força ff14SB, para fins de comparação.

Sobre os compostos utilizados, foram selecionados apenas moléculas que apresentaram bons resultados *in vivo* frente a enzima InhA de Mtb. Neste quesito, as metodologias utilizadas obtiveram sucesso, pois os valores obtidos com exceção do PBSA (ff99 e ff99SB), foram sempre valores negativos de energia livre.

Se o objetivo for analisar apenas o valor isolado de um ligante por métodos *in silico* (valor desconhecido), este estudo demonstrou que o programa AutoDock e o método de PBSA obtiveram os melhores resultados aproximados, quando comparados aos valores experimentais.

Tabela 6: Resultados dos métodos avaliados frente aos Grupos 1, 2 e 3 para o campo de força ff14SB. Os valores apresentados são referentes as correlações obtidas pelo método de regressão linear ( $R^2$ ).

Método	Parâmetro	Grupo 1 $R^2$	Grupo 2 $R^2$	Grupo 3 $R^2$
GBSA	$GB^{OBC}$	0,28	0,00	0,05
	$GB^{OBC2}$	0,17	0,00	0,00
	<i>Complexo Trajetória+ Entropia</i>	0,38	0,00	0,31
	<i>Múltiplas trajetórias</i>	0,43	0,00	0,11
	<i>Múltiplas trajetórias + Entropia</i>	0,26	0,00	0,04
PBSA	<i>Ligante/Complexo trajetória</i>	0,40	0,00	0,00
	----	0,51	0,25	0,05
QM/MM	$AM1_{Ligante}$	0,18	0,05	0,01
	$AM1_{Ligante+ NADH}$	0,15	0,10	0,01
	$PM3_{Ligante}$	0,60	0,02	0,05
	$PM3_{Ligante+ NADH}$	0,32	0,40	0,05
	$PM6_{Ligante}$	0,33	0,01	0,03
	$PM6_{Ligante+ NADH}$	0,23	0,12	0,02
	$RM1_{Ligante}$	0,23	0,01	0,00
	$RM1_{Ligante+ NADH}$	0,28	0,10	0,00
SQM	$PM3$	0,60	0,13	0,30
	$PM6-DH+$	0,10	0,23	0,02
	$PM6+ DHX$	0,10	0,25	0,00
	$PM7$	0,00	0,15	0,02
AutoDock	<i>Gasteiger charges</i>	0,27	0,01	0,24
	<i>AM1-BCC charges</i>	0,21	0,02	0,25
LigScore	----	0,44	0,10	0,00
DrugScore	----	0,58	0,20	0,02

## 7. CONSIDERAÇÕES FINAIS

O presente trabalho forneceu detalhes sobre os variados e amplamente utilizados métodos *in silico* de predição de energia livre de ligação. A finalidade foi testar estas metodologias frente a enzima InhA de *Mycobacterium tuberculosis* e seus ligantes. Este estudo, foi um dos pioneiros neste tipo de abordagem referente a enzima InhA. Os resultados mostraram uma igualdade entre as metodologias testadas, não sendo possível entrar em um consenso sobre o melhor método a ser utilizado para a predição de energia livre. Métodos mais robustos e computacionalmente extensivos como o (QM)MM/GBSA e SQM não demonstraram vantagem significativa perante ferramentas mais simples como o LigScore e o DrugScore para o Grupo 1. Os parâmetros testados neste estudo são passos iniciais para o desenvolvimento de uma ferramenta que permita a união de várias metodologias para predizer computacionalmente a energia livre de ligação de pequenas moléculas em larga escala. O intuito desta futura ferramenta é de auxiliar na pesquisa de novas moléculas candidatas a fármaco, em especial, contra a tuberculose, poupando desta maneira tempo de pesquisa, recursos financeiros e maximizando os resultados encontrados.

Com isso podemos concluir que apesar das várias abordagens testadas, necessitam ainda estudos complementares para calibrar uma função frente a InhA de Mtb. Para esta finalidade é necessário abordar uma faixa maior de compostos, incluindo compostos com  $\Delta G$  positivo (que não possuem atividade frente a enzima). Além destes compostos, outras metodologias podem ser testadas como LIE (*linear interaction energy*) e FEP (*free energy perturbation*).

## 8. REFERÊNCIAS

- Agüero, F., Al-Lazikani, B., Aslett, M., Berriman, M., Buckner, F. S., Campbell, R. K., Verlinde, C. L. M. J. Genomic-scale prioritization of drug targets: the TDR Targets database. *Nature Reviews. Drug Discovery*. **2008**, 7(11), 900–7. <http://doi.org/10.1038/nrd2684>
- Alonso, H., Bliznyuk, A. A., Gready, J. E.. Combining docking and molecular dynamic simulations in drug design. *Medicinal Research Reviews*. **2011**, 26 531–68. <http://doi.org/10.1002/med.20067>
- Anderson, S. T., Kaforou, M., Brent, A. J., Wright, V. J., Banwell, C. M., Chagaluka, G., Eley, B. Diagnosis of childhood tuberculosis and host RNA expression in Africa. *The New England Journal of Medicine*. **2014**, 370(18), 1712–23. <http://doi.org/10.1056/NEJMoa1303657>
- Barreira, D., Grangeiro, A. Avaliação das estratégias de controle da tuberculose no Brasil Evaluation of tuberculosis control strategies in Brazil, *Revista Brasileira de Epidemiologia*. **2007**, 05, 23-26
- Barry, C. E., Lee, R. E., Mdluli, K., Sampson, A. E., Schroeder, B. G., Slayden, R. A., & Yuan, Y. Mycolic acids: Structure, biosynthesis and physiological functions. *Progress in Lipid Research*. **2008**, 37(2-3), 143–179. [http://doi.org/10.1016/S0163-7827\(98\)00008-3](http://doi.org/10.1016/S0163-7827(98)00008-3)
- Camus, J., Pryor, M. J., Me, C., & Cole, S. T. Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology*. **2002**, Oct;148(Pt 10):2967-7.
- Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Woods, R. J. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*. **2005**, 26(16), 1668–1688. <http://doi.org/10.1002/jcc.20290>
- Cohen, E. M. L., Machado, K. S., Cohen, M., Norberto de Souza, O. Effect of the explicit flexibility of the InhA enzyme from *Mycobacterium tuberculosis* in molecular docking simulations. *BMC Genomics*, **2011**, 12 Suppl 4, S7. <http://doi.org/10.1186/1471-2164-12-S4-S7>
- Cole, S. T., Brosch, R., Parkhill, J., Garnier, T., Churcher, C., Harris, D., Barrell, B. G. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. **1998**, 393(6685), 537–544. <http://doi.org/10.1038/31159>
- Costa, M., Tavares V R, Silva, A C., Campos, Bueno, S.K. Tuberculose: Uma Revisão De Literatura *Anais Brasileiros de Epidemiologia*. **2014**, V1.
- Deglmann, P., Schäfer, A., & Lennartz, C. Application of quantum calculations in the chemical industry-An overview. *International Journal of Quantum Chemistry*. **2015**, 115(3), 107–136. <http://doi.org/10.1002/qua.24811>
- Dessen, a, Quémard, a, Blanchard, J. S., Jacobs, W. R., & Sacchettini, J. C. Crystal structure and function of the isoniazid target of *Mycobacterium tuberculosis*. *Science (New York, N.Y.)*. **1995**, 267(5204), 1638–1641. <http://doi.org/10.1126/science.7886450>

- Ducati, R. G., Ruffino-Netto, A., Basso, L. A., & Santos, D. S. The resumption of consumption - A review on tuberculosis. *Memorias Do Instituto Oswaldo Cruz*. **2006**, 101(7), 697–714. <http://doi.org/10.1590/S0074-02762006000700001>
- Fausto, M. A., Carneiro, M., Antunes, C. M. D. F., Pinto, J. A., & Colosimo, E. A. O modelo de regressão linear misto para dados longitudinais: uma aplicação na análise de dados antropométricos desbalanceados. *Cadernos de Saúde Pública*. **2008**, 24(3), 513–524. <http://doi.org/10.1590/S0102-311X2008000300005>
- Genheden, S., Ryde, U. The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities. *Expert Opinion on Drug Discovery*. **2015**, 0-441 (October), 1–13. <http://doi.org/10.1517/17460441.2015.1032936>
- Glickman, M. S., Jacobs, W. R., York, N., Ave, M. P. *Mycobacterium tuberculosis*. Dawn of a Discipline Albert Einstein College of Medicine. **2001**, 104, 477–485
- Gräter, F., Schwarzl, S. M., Dejaegere, A., Fischer, S., & Smith, J. C. Protein/ligand binding free energies calculated with quantum mechanics/molecular mechanics. *Journal of Physical Chemistry B*. **2005**, 109(20), 10474–10483. <http://doi.org/10.1021/jp044185y>
- Greenidge, P. A., Kramer, C., Mozziconacci, J. C., & Wolf, R. M. MM/GBSA binding energy prediction on the PDBbind data set: Successes, failures, and directions for further improvement. *Journal of Chemical Information and Modeling*. 2013, 53(1), 201–209. <http://doi.org/10.1021/ci300425v>
- Hanwell, M. D., Curtis, D. E., Lonie, D. C., Vandermeersch, T., Zurek, E., Hutchison, G. R. Avogadro: An advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*. **2012**, 4(8), 1–17. <http://doi.org/10.1186/1758-2946-4-17>
- Hou, T., Wang, J., Li, Y., & Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *Journal of Chemical Information and Modeling*. **2011**, 51(1), 69–82. <http://doi.org/10.1021/ci100275a>
- Jain, A., & Mondal, R. Extensively drug-resistant tuberculosis: Current challenges and threats. *FEMS Immunology and Medical Microbiology*. **2008**, 53(2), 145–150. <http://doi.org/10.1111/j.1574-695X.2008.00400>
- Kapetanovic, I. M. Computer-aided drug discovery and development (CADD): in silico-chemico-biological approach. *Chemico Biological Interactions*. **2008** 171(2), 165–76. <http://doi.org/10.1016/j.cbi.2006.12.006>
- Kim, H., Seo, M., Park, Y. K., Yoo, J.-I., Lee, Y. S., Chung, G. T., & Ryoo, S. Evaluation of MGIT 960 System for the Second-Line Drugs Susceptibility Testing of *Mycobacterium tuberculosis*. *Tuberculosis Research and Treatment*. **2013**, 108401. <http://doi.org/10.1155/2013/108401>
- Kitamura, K., Tamura, Y., Ueki, T., Ogata, K., Noda, S., Himeno, R., & Chuman, H. Binding free-

energy calculation is a powerful tool for drug optimization: Calculation and measurement of binding free energy for 7-azaindole derivatives to glycogen synthase kinase-3 $\beta$ . *Journal of Chemical Inf. and Modeling*. **2014**, 54(6), 1653–1660. <http://doi.org/10.1021/ci400719v>

Krammer, A., Kirchhoff, P. D., Jiang, X., Venkatachalam, C. M., & Waldman, M. LigScore: A novel scoring function for predicting binding affinities. *Journal of Molecular Graphics and Modelling*. **2005**, 23(5), 395–407. <http://doi.org/10.1016/j.jmgm.2004.11.007>

Kumari, R., Kumar, R., & Lynn, A. G\_mmpbsa-A GROMACS Tool for High-Throughput MM-PBSA Calculations. *Journal of Chemical Information and Modeling*. **2015** <http://doi.org/10.1021/ci500020m>

Leal, R. C., Neto, J. M. M., Lima, F. D. C. A., & Feitosa, C. M. A química quântica na compreensão de teorias de química orgânica. *Quimica Nova*. **2010**, 33(5), 1211–1215. <http://doi.org/10.1590/S0100-40422010000500037>

Lechartier, B., Rybniker, J., Zumla, A., Cole, S. T. Tuberculosis drug discovery in the post-post-genomic era. *Molecular Medicin*. **2010** 6(2), 158–68. <http://doi.org/10.1002/emmm.201201772>

Maestre-Reyna, M., Wu, W. J., & Wang, A. H. J. Structural insights into RbmA, a biofilm scaffolding protein of *Vibrium cholerae*. *Plos One* **2013**, 8(12), 1–10. <http://doi.org/10.1371/journal.pone.0082458>

Mandal, S., Moudgil, M., & Mandal, S. K. Rational drug design. *European Journal of Pharmacology*. **2009**, 625(1-3), 90–100. <http://doi.org/10.1016/j.ejphar.2009.06.065>

Mark, P., & Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC / E Water Models at 298 K. *The Journal of Physical Chemistry A*. **2010**, 105(43), 9954–9960. <http://doi.org/10.1021/jp003020w>

McMurry, L. M., Oethinger, M., & Levy, S. B. Triclosan targets lipid synthesis. *Nature*. **1998**, 394(6693), 531–532. <http://doi.org/10.1038/28970>

Morris, G., & Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Modeling*. **2009**, 30(16), 2785–2791. <http://doi.org/10.1002/jcc.21256>.AutoDock4

Noberto de Souza, O. Ornstein, R. L. (1997). Effect of periodic box size on aqueous molecular dynamics simulation of a DNA dodecamer with particle-mesh Ewald method. *Biophysical Journal*. **1997**, 72 (6), 2395-2397

Norberto de Souza, O., Ornstein, R. L. Molecular dynamics simulations of a protein-protein dimer: particle-mesh Ewald electrostatic model yields far superior results to standard cutoff model. *Journal of Biomolecular Structure & Dynamics*. **1999**, 16(6), 1205–18. <http://doi.org/10.1080/07391102.1999.10508328>

Oda, A., Tsuchida, K., Takakura, T., Yamaotsu, N., & Hirono, S. Comparison of consensus scoring

- strategies for evaluating computational models of protein-ligand complexes. *Journal of Chemical Information and Modeling*. **2006**, 46(1), 380–391. <http://doi.org/10.1021/ci050283k>
- Oliveira, J. S., Pereira, J. H., Canduri, F., Rodrigues, N. C., de Souza, O. N., de Azevedo, W. F., Santos, D. S. Crystallographic and Pre-steady-state Kinetics Studies on Binding of NADH to Wild-type and Isoniazid-resistant Enoyl-ACP(CoA) Reductase Enzymes from *Mycobacterium tuberculosis*. *Journal of Molecular Biology*. **2006**, 359(3), 646–666. <http://doi.org/10.1016/j.jmb.2006.03.055>
- Onufriev, A., Bashford, D., Case, D. A. Exploring Protein Native States and Large-Scale Conformational Changes with a Modified Generalized Born Model. *Proteins: Structure, Function and Genetics*. **2006**, 55(2), 383–394. <http://doi.org/10.1002/prot.20033>
- Onufriev, A., Case, D. a., & Bashford, D. Effective Born radii in the generalized Born approximation: The importance of being perfect. *Journal of Computational Chemistry*. **2002**, 23(14), 1297–1304. <http://doi.org/10.1002/jcc.10126>
- Pasqualotto, A. C., Rosa, D. D., Fontoura Pereira, M. D. C., Targa-Ferreira, R. L., & Santos, B. R. Retrospective study of 668 cultures for mycobacteria in a reference hospital for AIDS in southern Brazil. *The Brazilian Journal of Infectious Diseases*. **2003**, 7(2), 126–128. <http://doi.org/S1413-86702003000200005> [pii]
- Phillips, J. C., & Stone, J. E. Probing Biomolecular Machines with Graphics Processors. *Queue*. **2009**, 7(9), 30. <http://doi.org/10.1145/1626135.1629155>
- Pierce, L. C. T., Salomon-Ferrer, R., Augusto, C., de Oliveira, F., McCammon, J. a, & Walker, R. C. Routine access to millisecond timescale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation*. **2012**, <http://doi.org/10.1021/ct300284c>
- Quemard, a, Sacchettini, J. C., Dessen, a, Vilcheze, C., Bittman, R., Jacobs, W. R., & Blanchard, J. S. Enzymatic Characterization of the Target for Isoniazid in *Mycobacterium-Tuberculosis*. *Biochemistry*. **1995**, 34(26), 8235–8241.
- Rawat, R., Whitty, A., & Tonge, P. J. The isoniazid-NAD adduct is a slow, tight-binding inhibitor of InhA, the *Mycobacterium tuberculosis* enoyl reductase: adduct affinity and drug resistance. *Proceedings of the National Academy of Sciences of the United States of America*. **1995**, 100(24), 13881–13886. <http://doi.org/10.1073/pnas.2235848100>
- Reddy, M. R., & Erion, M. D. Relative Binding Affinities of Fructose-1,6-Bisphosphatase Inhibitors Calculated Using a Quantum Mechanics-Based Free Energy Perturbation Method. *Journal of the American Chemical Society*. **2007**, 129(30), 9296–9297. <http://doi.org/10.1021/ja072905j>
- Rezácová, P., Borek, D., Moy, S. F., Joachimiak, A., & Otwinowski, Z. Crystal structure and putative function of small Toprim domain-containing protein from *Bacillus stearothermophilus*. *Proteins*. **2008**, 70(2), 311–319. <http://doi.org/10.1002/prot>
- Rosado, L. A, Caceres, R. A, de Azevedo, W. F., Basso, L. A, Santos, D. S.; Norberto de Souza, O.

Role of Serine140 in the mode of action of Mycobacterium tuberculosis beta-ketoacyl-ACP Reductase (MabA). *BMC Research Notes*. **2012**, 5(1), 526. <http://doi.org/10.1186/1756-0500-5-526>

- Rotta, M., Pissinate, K., Villela, A. D., Back, D. F., Timmers, L. F. S. M., Bachega, J. F. R., Machado, P. Piperazine derivatives: Synthesis, inhibition of the Mycobacterium tuberculosis enoyl-acyl carrier protein reductase and SAR studies. *European Journal of Medicinal Chemistry*. **2015**, 90, 436–447. <http://doi.org/10.1016/j.ejmech.2014.11.034>
- Rozwarski, D. A., Vilchèze, C., Sugantino, M., Bittman, R., & Sacchettini, J. C. Crystal structure of the Mycobacterium tuberculosis enoyl-ACP reductase, InhA, in complex with NAD<sup>+</sup> and a C16 fatty acyl substrate. *Journal of Biological Chemistry*. **1999**, 274(22), 15582–15589. <http://doi.org/10.1074/jbc.274.22.15582>
- Rozwarski, D.A., Grant, G.A., Barton, D.H.R, Jacobs Jr., W.R., Sacchetini, J.C. Modification of the NADH of the isoniazid Target (InhA) from Mycobacterium Tuberculosis. *Science*. **1998**, 98 – 102
- San Pedro, A., & Oliveira, R. M. De. Tuberculosis and socioeconomic indicators: systematic review of the literature. *Revista Panamericana de Salud Pública = Pan American Journal of Public Health*. **2009**, 33(4), 294–301. <http://www.ncbi.nlm.nih.gov/pubmed/23698179>
- Schaeffer, M. L., Agnihotri, G., Volker, C., Kallender, H., Brennan, P. J., & Lonsdale, J. T. Purification and Biochemical Characterization of the Mycobacterium tuberculosis B-Ketoacyl-acyl Carrier Protein Synthases KasA and KasB. *Journal of Biological Chemistry*. **2011**, 276(50), 47029–47037. <http://doi.org/10.1074/jbc.M108903200>
- Schroeder, E. K., Basso, L. A., Santos, D. S., Norberto de Souza, O. Molecular dynamics simulation studies of the wild-type, I21V, and I16T mutants of isoniazid-resistant Mycobacterium tuberculosis enoyl reductase (InhA) in complex with NADH: toward the understanding of NADH-InhA different affinities. *Biophysical Journal*. **2005**, 89(2), 876–84. <http://doi.org/10.1529/biophysj.104.053512>
- Schroeder, E. K., Norberto de Souza, O., Santos, D. S., Blanchard, J. S., & Basso, L. A. Drugs that Inhibit Mycolic Acid Biosynthesis. *Current Pharmaceutical Biotechnology*. **2002**, 3 (3), 197-225
- Stewart, J. J. P. Optimization of parameters for semiempirical methods VI: More modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*. **2003**, 19(1), 1–32. <http://doi.org/10.1007/s00894-012-1667-x>
- Su, P.-C., Tsai, C.-C., Mehboob, S., Hevener, K. E., & Johnson, M. E. Comparison of radii sets, entropy, QM methods, and sampling on MM-PBSA, MM-GBSA, and QM/MM-GBSA ligand binding energies of *F. tularensis* enoyl-ACP reductase (FabI). *Journal of Computational Chemistry*. **2015**, <http://doi.org/10.1002/jcc.24011>

- Swanson, J. M. J., Adcock, S. a., & McCammon, J. A. Optimized radii for Poisson-Boltzmann calculations with the AMBER force field. *Journal of Chemical Theory and Computation*. **2005**, 1(3), 484–493. <http://doi.org/10.1021/ct049834o>
- Tuckerman, M. E., & Martyna, G. J. Understanding Modern Molecular Dynamics: Techniques and Applications. *The Journal of Physical Chemistry B*. **2000**, 104(2), 159–178. <http://doi.org/10.1021/jp992433y>
- Velec, H. F. G., Gohlke, H., & Klebe, G. DrugScoreCSD-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*. 2005, 48(20), 6296–6303. <http://doi.org/10.1021/jm050436v>
- Vilchèze, C., Morbidoni, H. R., Weisbrod, T. R., Iwamoto, H., Kuo, M., Sacchettini, J. C., & Jr, W. R. J. Inactivation of the inhA Encoded Fatty Acid Synthase II (FASII) Enoyl-Acyl Carrier Protein Reductase. *Journal of Bacteriology*. **2000**, 182(14), 4059–4067. <http://doi.org/10.1128/JB.182.14.4059-4067.2000>.
- Wang, J., Morin, P., Wang, W., & Kollman, P. A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *Journal of the American Chemical Society*. **2001**, 123(22), 5221–5230. <http://doi.org/10.1021/ja003834q>
- Xu, L., Sun, H., Li, Y., Wang, J., & Hou, T. Assessing the performance of MM/PBSA and MM/GBSA methods. 3. The impact of force fields and ligand charge models. *The Journal of Physical Chemistry. B*. **2013**, 117(28), 8408–21. <http://doi.org/10.1021/jp404160y>
- Yang, T., Wu, J. C., Yan, C., Wang, Y., Luo, R., Gonzales, M. B., Ren, P. Virtual screening using molecular simulations. *Proteins: Structure, Function and Bioinformatics*. **2001**, 79(6), 1940–1951. <http://doi.org/10.1002/prot.23018>
- Zhu, H. Implementation and Application of the mmff94 Force Field. *Journal of Lincoln Nebraska University*. **2014**, 16, 140-5
- World Health Organization (WHO), Report On Tuberculosis, **2015**.

## 9. ANEXOS

```
#Equilibrando a Densidade da Caixa D'água por 100ps
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  ntxo=1, ntp=500,
  ntwr=500, iwrap=0, ntwx=500, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0, idecomp=0,
  ibelly=1, ntr=0, bellymask=':271-11021', restraint_wt=100.0, restraintmask=':1-270',
  nstlim=50000, nscm=500, t=0.0, dt=0.002, nrespa=1,
  isgld=0, tsgavg=0.2, tempsg=1.0, sgft=0.0, isgsta=1, isgend=1,
  ntt=1, temp0=298.16, tempi=10.0, ig=71277, tautp=1.0, gamma_ln=0, vrand=0, vlimit=20.0,
  ntp=1, pres0=1.0, comp=44.6, taup=2.0,
  ntc=2, tol=0.00001, jfastw=0,
  ivcap=0,
  ntf=2, ntb=2, dielc=1.0, cut=9.0, nsnb=25, ipol=0, ifqnt=0,
  igb=0, ievb=0,
/
&ewald
  verbose=0, ew_type=0, nbflag=1, use_pme=1, vdwmeth=1, eedmeth=1, netfrc=0,
/
```

Anexo 1: Input utilizado neste trabalho para a fase de minimização nas simulações por dinâmica molecular.

```
#Fase de aquecimento do sistema até 298.16K
&cntrl
  imin=0, ntx=1, irect=0, ntrx=1,
  ntxo=1, ntp=2500,
  ntwr=2500, iwrap=0, ntwx=2500, ntwv=0, ntwe=0,
  ioutfm=0, ntwprt=0, idecomp=0,
  ibelly=0, ntr=0, bellymask=':1-1', restraint_wt=1.0, restraintmask=':1-270',
  nstlim=50000, nscm=2500, t=0.0, dt=0.002, nrespa=1,
  isgld=0, tsgavg=0.2, tempsg=1.0, sgft=0.0, isgsta=1, isgend=1,
  ntt=1, temp0=50.0, tempi=10.0, ig=71277, tautp=1.0, gamma_ln=0, vrand=0, vlimit=20.0,
  ntp=1, pres0=1.0, comp=44.6, taup=2.0,
  ntc=2, tol=0.00001, jfastw=0,
  ivcap=0,
  ntf=2, ntb=2, dielc=1.0, cut=9.0, nsnb=25, ipol=0, ifqnt=0,
  igb=0, ievb=0,
/
&ewald
  verbose=0, ew_type=0, nbflag=1, use_pme=1, vdwmeth=1, eedmeth=1, netfrc=0,
/
```

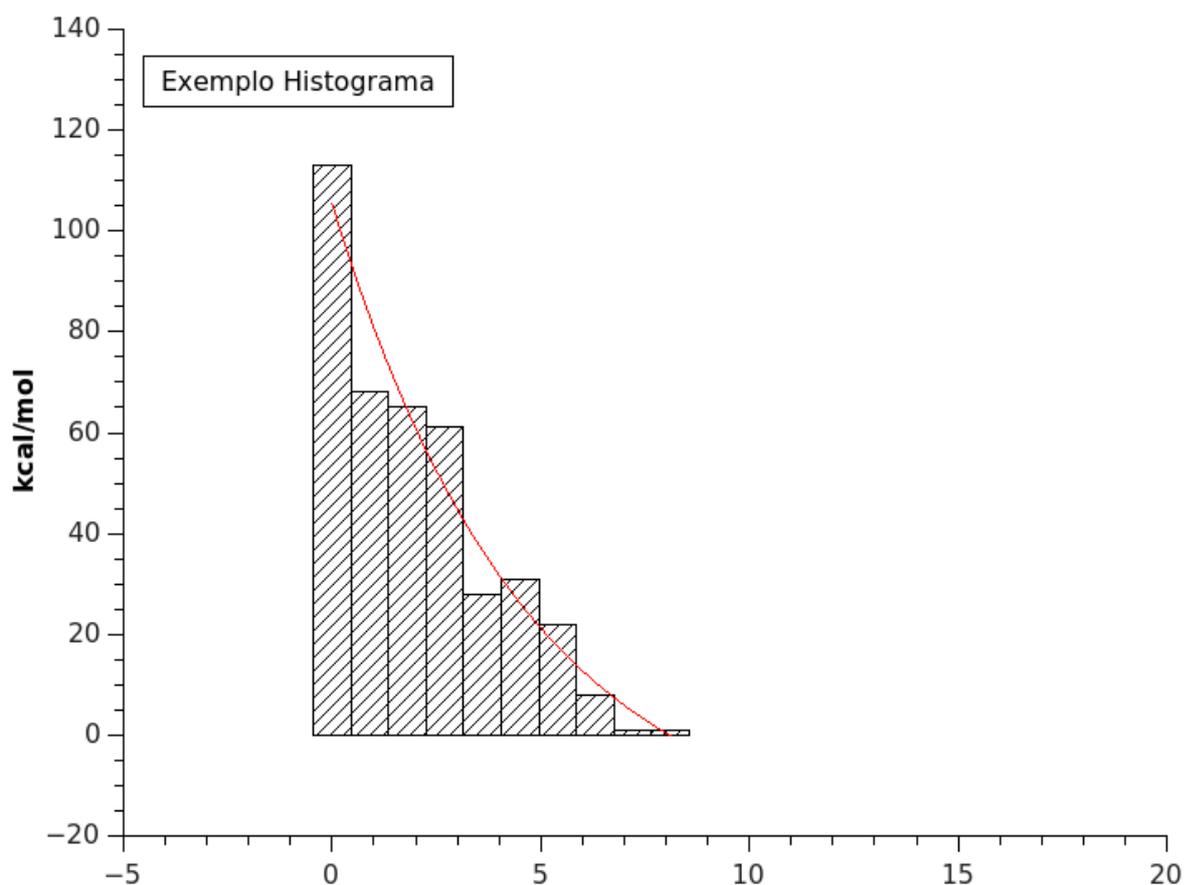
Anexo 2: Input utilizado neste trabalho para a fase de aquecimento gradual do sistema na dinâmica molecular. Para cada etapa os valores referentes a temperatura necessitam ser trocados (temp0 e tempi) até atingir a temperatura desejada.

```

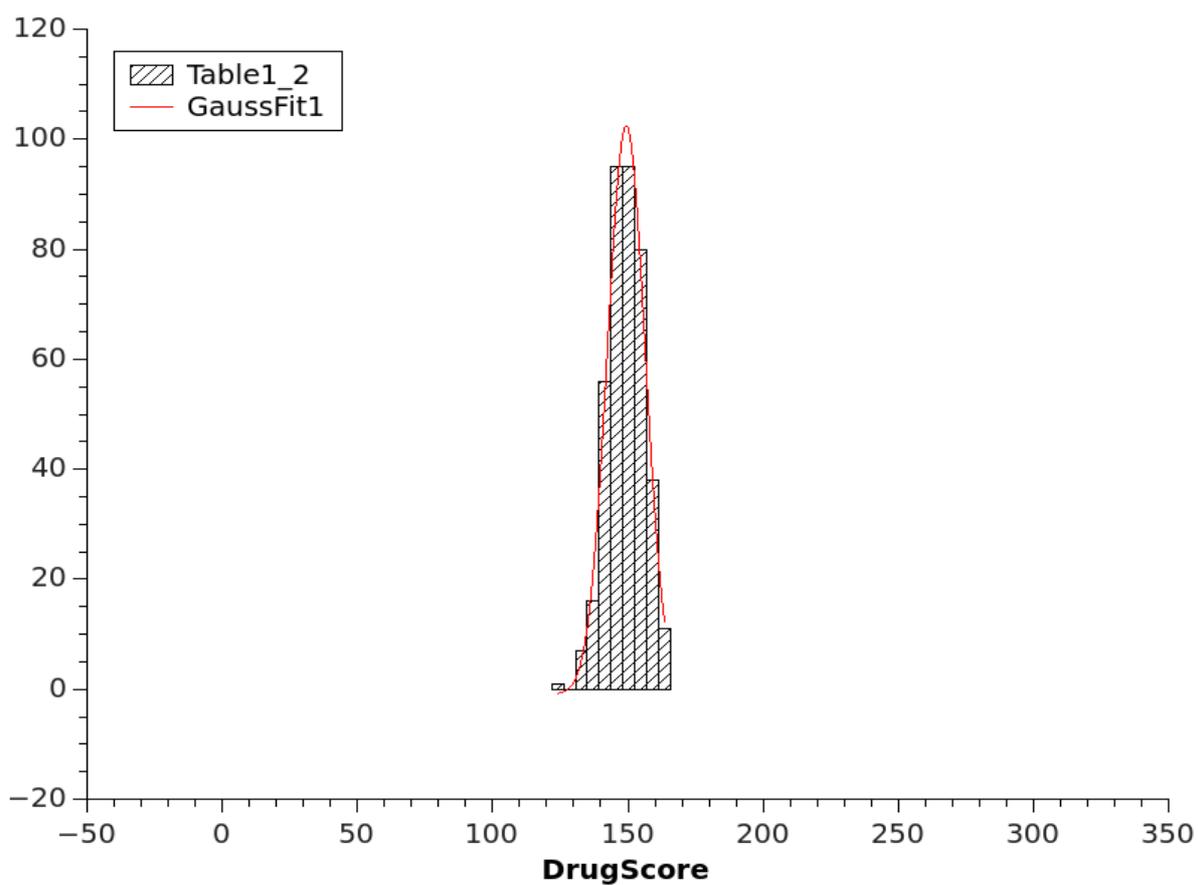
### DM 2ns##
&cntrl
  imin=0, ntx=5, irect=1, ntrx=1,
  ntxo=1, npr=2500,
  ntwr=2500, iwrap=0, ntwx=2500, ntwv=0, ntw=0,
  ioutfm=0, ntwprt=0, idecomp=0,
  ibelly=0, ntr=0, bellymask=':1-1', restraint_wt=1.0, restraintmask=':1-270',
  nstlim=1000000, nscm=2500, t=0.0, dt=0.002, nrespa=1,
  isgld=0, tsgavg=0.2, tempsg=1.0, sgft=0.0, isgsta=1, isgend=1,
  ntt=1, temp0=298.16, tempi=298.16, ig=71277, tautp=1.0, gamma_ln=0, vrand=0, vlimit=-1.0,
  ntp=1, pres0=1.0, comp=44.6, taup=2.0,
  ntc=2, tol=0.00001, jfastw=0,
  ivcap=0,
  ntf=2, ntb=2, dielc=1.0, cut=9.0, nsnb=25, ipol=0, ifqnt=0,
  igb=0, ievb=0,
/
&ewald
  verbose=0, ew_type=0, nbflag=1, use_pme=1, vdwmeth=1, eedmeth=1, netfrc=0,
/

```

Anexo 3: Input da fase de produção da DM. Para todas metodologias empregas, foi mantida os mesmos parametros para a fase de produção.



Anexo 4: Modelo de histograma, que permite calcular o resultado mais populado de uma determinada população.



Anexo 5: Método de gaussiana utilizada para calcular o valor mais populado de um conjunto de valores obtidos. Esta técnica foi empregada para calcular o resultado final para cada ligante nas metodologias do LigScore, DrugScore e AutoDock.