

---

UNIVERSIDADE FEDERAL DE SÃO  
CARLOS

CENTRO DE CIÊNCIAS EXATAS E DE TECNOLOGIA

DEPARTAMENTO DE FÍSICA

CURSO DE DOUTORADO EM FÍSICA BÁSICA

GRUPO FÍSICA TEÓRICA DE BIOMOLÉCULAS E MATERIAIS

Estudos in silico de proteínas alostéricas e sua interação com  
ligantes

Autor: Thales Souza Freire\*

Orientador: Profa. Dra. Ignez Caracelli

São Carlos-SP

mês / 2023

---

\* Bolsista CNPq

UNIVERSIDADE FEDERAL SÃO CARLOS  
CENTRO DE CIÊNCIAS EXATAS E DA TERRA  
DEPARTAMENTO DE FÍSICA  
DOUTORADO EM FÍSICA BÁSICA

Thales Souza Freire

Estudos in silico de proteínas alostéricas e sua interação com  
ligantes

São Carlos-SP

2023

THALES SOUZA FREIRE

Estudos in sílico de proteínas alostéricas e sua interação com ligantes

Dissertação apresentada ao Curso de  
Doutorado em Física da Universidade  
Federal de São Carlos, como requisito  
parcial para obtenção do Grau de Doutor.  
Área de Concentração: Física Básica.

Orientador: Profa. Dra. IGNEZ CARACELLI

São Carlos-SP

2023

Dedico este trabalho aos meu pais Valdineide e Hildebrando (in memorian), que com seus ensinamentos me possibilitaram chegar até aqui.

Às minha irmãs, Thálita e Thayse, que sempre estiveram ao meu lado me dando forças. Aos meu tios Edenilto e Ivone por toda ajuda que deram.

À minha namorada Caroline, por me acompanhar nessa jornada, sempre me apoiando e incentivando nos momentos mais difíceis.

# Agradecimentos

A todos aqueles que deram suas contribuições para a realização deste trabalho, deixo-lhes aqui o meu agradecimento sincero.

À minha família, pelo apoio incondicional que me deram, especialmente ao meu pai por sempre me incentivar e direcionar nos estudos.

À Profa. Dra. Ignez Caracelli, pela orientação, confiança e incentivo desde o início deste trabalho.

Ao Prof. Julio Zukerman e todos do grupo BioMat, pelas discussões e correções que melhoraram a qualidade da minha pesquisa.

Ao Prof. Ran Friedman pelas ótimas oportunidades de trabalho, orientação e disponibilização de recursos, sem os quais não seria possível realizar o volume de simulações para a conclusão desta tese.

À minha namorada Caroline, por me dar suporte e ajudar a melhorar meus pontos fracos.

Aos meus amigos e colegas de curso, que me proporcionaram novas ideias durante o decorrer deste trabalho.

Ao DF-UFSCar, pela utilização de sua infraestrutura para a realização deste Doutorado.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) juntamente com Programa de Pós Graduação em Física (PPGFIS), pela concessão da bolsa de Doutorado.

A Capes por tornar possível acesso aos periódicos, sem o que seria difícil estar em dia com a literatura.

# Lista de Abreviações

**Abl1** FMS-Like Tirosina Quinase 3

**BAR** Relação de Aceitação de Bennet

**CP** Componentes Principais

**DM** Dinâmica Molecular Clássica

**EDS** *Essential dynamics sampling*

**FEP** Perturbação de Energia Livre

**FLT3** FMS-Like Tirosina Quinase 3

**HSA** Albumina Sérica Humana

**ITD** *Internal tandem duplication*

**LMC** Leucemia Mieloide Crônica

**PDB** Protein Data Bank

**RMSD** *Root Mean Square Deviation*

**wT-metaD** Metadinâmica Bem Temperada

# Lista de Figuras

1	Modelo tridimensional do grupo heme . . . . .	1
2	Diferença de conformação entre os estados R e T da hemoglobina . . . . .	2
3	Velocidade da reação: Hill (Bioquímica) . . . . .	2
4	Fluxograma uma simulação de docking . . . . .	7
5	Algoritmo genético e o Programa GOLD . . . . .	9
6	Esquema do método de integração padrão do gromacs . . . . .	12
7	Esquema dos parâmetros do potencial intramolecular . . . . .	13
8	Perfil de energia livre em função de $s_0$ . . . . .	23
1.1	Mudança de estado da enzima Abl1 . . . . .	26
1.3	Esquema geral das simulações para cálculo de $\Delta G$ da mutação . . . . .	29
1.4	Plot gerado pelo programa PMX . . . . .	29
1.5	Interações intermoleculares para o dasatinib e a estrutura nativa da Abl1 . . . . .	31
2.1	Domínios e subdomínios da HSA . . . . .	35
2.2	Superposição das estruturas de HSA . . . . .	37
2.3	Estado N e B da HSA superpostos . . . . .	37
2.4	rmsd das trajetórias de EDS e caminho de referência para metadinâmica para HSA . . . . .	41
2.5	Superposição entre os estados básico e neutro da albumina, evidenciando os centros de massa dos subdomínios . . . . .	42
2.6	Distância entre os resíduos Tyr138 e Tyr161 para o caso estado a HSA . . . . .	43
2.7	DM da 1E7E após retirar ácidos graxos . . . . .	44
2.8	Metadinâmica da HSA com variáveis coletivas geométricas . . . . .	45
2.9	Metadinâmica da HSA com variáveis coletivas de caminho . . . . .	47

2.10	Estruturas da HSA comparando os mínimos de energia da Metadinâmica .	47
3.1	Estrutura do gilteritinib dentro do sítio da FLT3 . . . . .	51
3.2	rmsd das trajetórias de EDS e caminho de referência para metadinâmica para FLT3 . . . . .	56
3.3	Estruturas ao longo da trajetória de ativação . . . . .	57
3.4	Superfície de Energia Livre e Caminho de Mínima Energia da FLT3 . . . .	58
3.5	Estrutura do gilteritinib . . . . .	59



# Sumário

<b>Agradecimentos</b>	<b>iv</b>
<b>Lista de Abreviações</b>	<b>v</b>
<b>Lista de Figuras</b>	<b>vii</b>
<b>Resumo</b>	<b>xi</b>
<b>Abstract</b>	<b>xii</b>
<b>INTRODUÇÃO</b>	<b>1</b>
Alosterismo . . . . .	1
Objetivos . . . . .	3
<b>REFERENCIAL TEÓRICO</b>	<b>5</b>
Seleção, Visualização e Edição de Estruturas . . . . .	6
Docking . . . . .	6
Dinâmica Molecular . . . . .	10
Condições Iniciais . . . . .	11
Método de integração . . . . .	11
Campo de Força . . . . .	12
Minimização de energia . . . . .	14
Ensembles . . . . .	15
Análise de Componentes Principais . . . . .	16
Perturbação de Energia Livre (FEP) . . . . .	17
Metadinâmica . . . . .	21

<b>1</b>	<b>ABELSON TIROSINA QUINASE</b>	<b>26</b>
1.1	Introdução . . . . .	26
1.2	Métodos . . . . .	28
1.2.1	Software, campo de força e parâmetros para simulações de DM . . .	28
1.2.2	Perturbação de Energia Livre . . . . .	28
1.3	Discussão e Resultados . . . . .	29
1.4	Conclusão . . . . .	32
1.5	Informações adicionais . . . . .	33
<b>2</b>	<b>ALBUMINA SÉRICA HUMANA</b>	<b>35</b>
2.1	Introdução . . . . .	35
2.2	Métodos . . . . .	36
2.2.1	Estrutura . . . . .	36
2.2.2	Software, campo de força e parâmetros para simulações de DM . . .	38
2.2.3	Dinâmica Molecular . . . . .	38
2.2.4	<i>Essential dynamics sampling</i> (EDS) . . . . .	39
2.2.5	Metadinâmica . . . . .	40
2.3	Resultados e Discussão . . . . .	42
2.3.1	Dinâmica Molecular . . . . .	42
2.3.2	Metadinâmica . . . . .	44
2.4	Conclusão . . . . .	48
2.5	Informações Adicionais . . . . .	48
<b>3</b>	<b>FMS-LIKE TIROSINA QUINASE 3</b>	<b>50</b>
3.1	Introdução . . . . .	50
3.2	Métodos . . . . .	52
3.2.1	Software, campo de força e parâmetros para simulações de DM . . .	52

	x
3.2.2 Modelos . . . . .	52
3.2.3 Estrutura do estado ativo ligada ao gilteritinib . . . . .	53
3.2.4 Perturbação de energia livre . . . . .	53
3.2.5 Essential dynamics sampling . . . . .	54
3.2.6 Metadinâmica . . . . .	55
3.3 Discussão e Resultados . . . . .	59
3.4 Conclusão . . . . .	62
3.5 Informações Adicionais . . . . .	63
<b>CONSIDERAÇÕES FINAIS</b>	<b>64</b>
Referências . . . . .	71
<b>A Anexo 1</b>	<b>72</b>
<b>B Anexo 2</b>	<b>90</b>

# Resumo

Proteínas alostéricas são de extrema importância para a regulação de diversas funções no organismo, como o transporte de moléculas pelo sangue e membranas celulares, bem como a fosforilação de outras proteínas. Neste contexto, o presente trabalho se dividiu no estudo de três proteínas: FMS-Like Tirosina Quinase 3 (Abl1), Albumina Sérica Humana (HSA) e FMS-Like Tirosina Quinase 3 (FLT3). Cada proteína foi estudada separadamente, com utilização de métodos de simulação computacional, como Dinâmica Molecular Clássica e Docking Molecular. Para a Abl1 foi estudado o surgimento de resistência para o medicamento dasatinib devido a mutações. Aqui, três mutações foram simuladas com o método de Perturbação de Energia Livre, onde foi calculada a diferença na variação de energia livre entre ligar o inibidor na forma nativa e mutante da enzima, obtendo uma boa concordância com valores experimentais. Já para a albumina foi estudada a mudança de conformação mediada pela interação com ácidos graxos de cadeia média. Para isso, foi usada o método Metadinâmica bem temperada para facilitar a transição entre os estados e obter caminho de mínima energia livre entre um estado e outro, com resultados compatíveis com as observações estruturais. Já para FLT3, iniciou-se um estudo similar ao da Abl1, para o fármaco gilteritinib, porém a dinâmica desta enzima se mostrou mais complexa do que a da Abl1. Como os resultados teóricos não concordavam com os valores experimentais, optou-se por investigar o efeito das mutação no perfil de energia livre de ativação, com utilização da Metadinâmica com variáveis coletivas de caminho. Tal ferramenta possibilitou revelar o efeito das mutações na dinâmica de ativação, onde o efeito de resistência é em parte devido a uma redução na barreira de ativação da enzima. Nos três casos, apenas a dinâmica molecular pura não é capaz de obter as informações desejadas, sendo necessário lançar mão de métodos de melhoramento de amostragem, como a perturbação de energia livre e metadinâmica, os quais se mostraram eficientes para os casos tratados.

Palavras-chave: Alostéria, proteínas, enzimas, dinâmica molecular, docking, metadinâmica, perturbação de energia livre, FLT3, Abl1, HSA, mutações

# Abstract

Allosteric proteins are extremely important for the regulation of several functions in the organism, such as the transport of molecules through the blood and cell membranes, as well as the phosphorylation of other proteins. In this context, the present work was divided into the study of three proteins: Abelson Tyrosine Kinase (Abl1), Human Serum Albumin (HSA) and FMS-Like Tyrosine Kinase 3 (FLT3). Each protein was studied separately, using computational simulation methods, such as Classical Molecular Dynamics and Molecular Docking. For Abl1 was studied the emergence of resistance to the drug dasatinib upon mutations. Here, three mutations were simulated with the Free Energy Perturbation method, where the difference in free energy variation between binding the inhibitor to the native and mutant form of the enzyme was calculated, obtaining a good agreement with experimental values. As for albumin, the change in conformation mediated by the interaction with medium-chain fatty acids was studied. In this case, the well-tempered Metadynamics method was used to facilitate the transition between states and obtain the path of minimum free energy between one state and another, the results were compatible with the experimental observations. As for FLT3, a study similar to that of Abl1 was initiated, for the drug gilteritinib, however the dynamics of this enzyme proved to be more complex than that of Abl1. As the theoretical results did not agree with the experimental values, it was decided to investigate the effect of mutations on the activation free energy profile, using Metadynamics with collective path variables. Such a tool made it possible to reveal the effect of mutations on the activation dynamics, where the resistance effect is partly due to a reduction in the enzyme activation barrier. In the three cases, pure molecular dynamics alone was not able to give the desired information, making it necessary to resort to enhanced sampling methods, such as free energy perturbation and metadynamics, which showed to be efficient for the treated cases.

Keywords: Allostery, proteins, enzymes, molecular dynamics, docking, metadynamics, free energy perturbation, FLT3, Abl1, HSA, mutations

# INTRODUÇÃO

As proteínas estudadas nessa tese apresentam como característica comum o alosterismo, cada uma com suas particularidades. Por esse motivo, a próxima seção irá introduzir detalhes importantes sobre este efeito. Uma discussão mais detalhada pode ser encontrada na minha dissertação de mestrado ([1]), onde os modelos encontrados na literatura são apresentados e comparados para motivar, em seguida, o desenvolvimento de novas equações de  $IC_{50}$  e velocidade de reação para enzimas alostéricas com ferramentas da Mecânica Estatística.

## Alosterismo

A estrutura das proteínas podem se comportar de maneiras variadas dependendo das interações com solvente, pH, íons e ligantes. No mais simples dos casos, a posição média dos aminoácidos no tempo não muda significativamente, sendo possível aproximar o comportamento da estrutura para um corpo rígido [2,3]. Para essas proteínas, considera-se a existência de apenas uma conformação da estrutura tridimensional, tanto para a forma livre quanto para o complexo proteína-ligante.

Porém, alguns processos fisiológicos requerem enzimas com maior sensibilidade a variações na concentração de substrato, ou ainda de proteínas que tenham a sua afinidade pelo ligante aumentada em situações específicas. Um exemplo deste tipo de proteína é a hemoglobina. Ela é formada por 4 subunidades, tendo no centro de cada monômero um grupo heme (Figura 1), no qual o oxigênio se liga. Esta proteína existe em duas conformações distintas, tendo um estado livre (T) e outro ligado (R) à moléculas de oxigênio (Figura 2). A forma T tem baixa afinidade por moléculas de  $O_2$ , porém à medida que os monômeros são ocupados por oxigênio a afinidade dos sítios livres é aumentada até que atinge o máximo quando três sítios estão ligados à moléculas de oxigênio. Nesta situação, o quarto

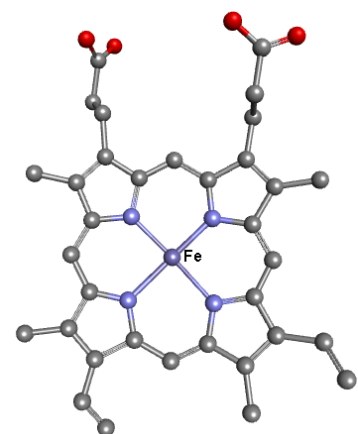


Figura 1: Modelo tridimensional do grupo heme, evidenciando o elemento Fe no centro da molécula (Fonte: <https://www.rcsb.org/structure/6FQF>)

sítio terá a maior probabilidade possível de ser ocupado por uma molécula de oxigênio. Neste ponto a proteína finaliza a sua transição para a forma R, como mostrado na Figura 2.

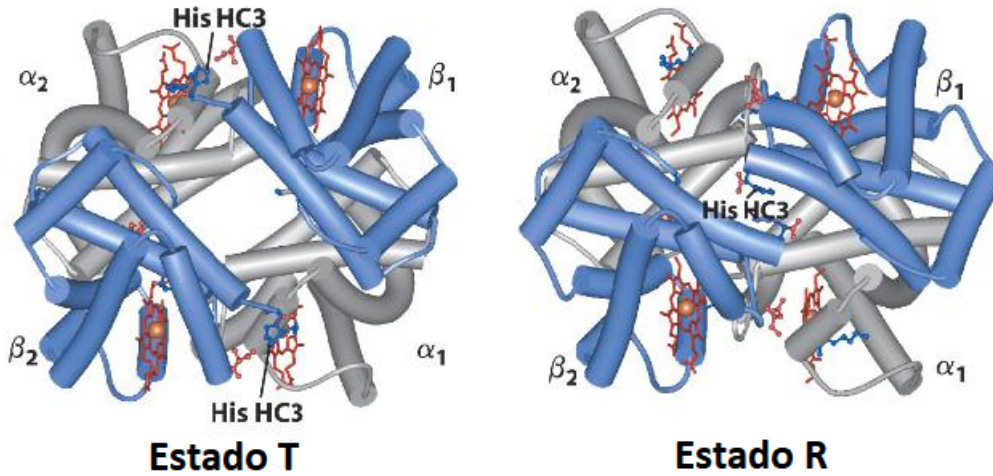


Figura 2: Esquema mostrando a diferença de conformação entre os estados R e T da hemoglobina. (Fonte: <https://fayllar.org/lecture-3-proteins.html>)

Quem primeiro tentou ajustar os dados experimentais para proteínas com esse comportamento foi Archibald Vivian Hill, em 1910 [4]. Em seus trabalhos foi proposta uma equação semelhante a:

$$v = V_{max} \frac{([S]K_{0,5}^{-1})^h}{1 + ([S]K_{0,5}^{-1})^h} \quad (1)$$

Uma equação semelhante a Eq. 1 foi formulada a partir de uma modificação do modelo de Michaelis-Menten para um ajuste dados experimentais da hemoglobina. A constante  $K_{0,5}$  faz um papel similar à constante de Michaelis-Menten e o parâmetro  $h$  é uma correção empírica para melhorar o ajuste da curva, sem significado microscópico. A Equação de Hill consegue explicar o efeito de aumento da probabilidade de ligação para incrementos na concentração de oxigênio, o que experimentalmente se evidencia, para o caso de

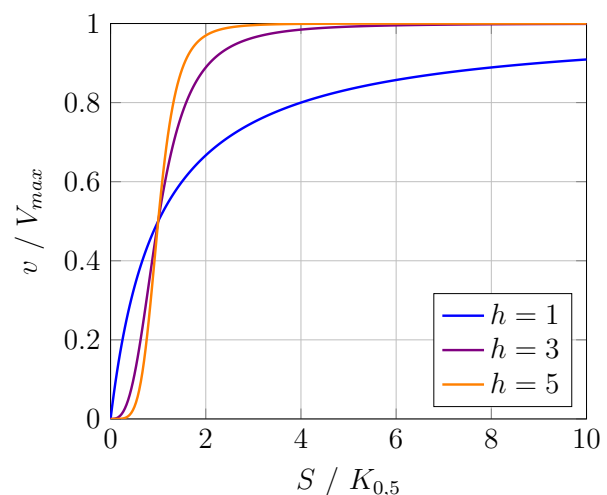


Figura 3: Velocidade relativa da reação segundo a equação de Hill. Para valores de  $h > 1$  existe ponto de inflexão para a curva, o que torna a catálise mais sensível a variações de concentração do substrato no intervalo de  $0,1V_{max}$  a  $0,9V_{max}$  [5]. (Figura elaborada pelo autor)

enzimas, pelo surgimento de um ponto de inflexão na curva  $v \times [S]$  (Figura 3). Como consequência, a resposta do sistema cresce consideravelmente para  $h > 1$ , por exemplo no caso das curvas azul e roxa da Figura 3, verifica-se que para atingir 90% da velocidade máxima é necessário aproximadamente 4 vezes mais substrato para  $h = 1$  (Modelo de Michaelis-Menten) do que  $h = 3$ . Fatos como esse aliados ao conhecimento das proteínas estudadas sugeriam uma interpretação para o fator  $h$ , pois o aparecimento de curvas sigmoidais aparecem comumente em proteínas com mais de 1 sítio para ligantes. Como nesses casos, a entrada de uma molécula nos sítios proteicos parecia cooperar para a ligação de uma próxima molécula cunhou-se o termo *cooperatividade* para descrever esse efeito de interação entre os sítios [5].

## Objetivos

Este trabalho tem como objetivo estudar o comportamento de proteínas alostéricas e a sua interação com ligantes. Mais especificamente, serão estudadas as variações de energia livre associadas às mutações no estado ativo das enzimas Abl1 e FLT3 e de mudanças de conformação da proteína HSA.

Primeiramente serão analisadas mutações que causam resistência na enzima Abl1 para o inibidor dasatinib. Utilizando-se de conceitos da Mecânica Estatística associada a simulações de Dinâmica Molecular Clássica, serão calculadas as diferenças de variação de energia livre para cada mutação. Estes valores de energias serão relacionados com a magnitude da resistência causada por cada mutação.

Em seguida, será realizado as transições alostéricas da proteína HSA e sua interação com ácidos-graxos e fármacos. Para isso, também serão utilizadas as ferramentas de Dinâmica Molecular aliadas a técnicas avançadas para cálculo da variação energia livre da transição alostérica. Estes resultados possibilitarão dizer qual a consequência da mudança de estado da HSA para a cinética destes ligantes no meio fisiológico.

Por fim, pretende-se realizar um estudo similar ao da Abl1 com a FLT3. O fármaco analisado será do gilteritinib, um inibidor também da forma ativa desta quinase. O intuito desta terceira parte é testar o modelo aplicado à Abl1 e verificar quão similar são os dois casos, verificando se apenas mudanças de interação causada pelas mutações são suficientes



para explicar o surgimento de resistências ao medicamento.

# REFERENCIAL TEÓRICO

Na primeira parte do trabalho foram estudadas três mutações da enzima Abl1 e suas consequências para a afinidade do estado ativo com o inibidor dasatinib: V299L, T315I e F317L<sup>1</sup>. Devido ao fato de o fármaco ligar-se apenas ao estado ativo da Abl1, as simulações foram realizadas apenas para essa conformação da enzima. Neste caso, como o inibidor e o substrato competem pelo mesmo sítio do estado ativo a cinética da inibição se aproxima de uma enzima de Michaelis Menten. Para cada uma das mutações citadas acima, foram calculadas as variações de energia livre de ligação do dasatinib com a conformação ativa da enzima Abl1 utilizando Dinâmica Molecular Clássica. Este resultado possibilitou verificar se apenas a interação da enzima com o inibidor é suficiente para explicar o aumento de resistência.

Já a segunda parte do trabalho será focada no estudo das transições alostéricas da albumina sérica humana e tentar entender o papel dos ácidos graxos na dinâmica desta proteína. Na primeira parte deste estudo foram feitas simulações para entender a movimentação da estrutura proteica ao se retirar os ácidos graxos. Em seguida, foi avaliada o perfil de energia livre da transição entre dois estados conformacionais, utilizando dois métodos diferentes de Metadinâmica. A comparação entre os três estágios deste estudo possibilitou entender como a proteína muda de conformação na ausência de ácidos graxos, tanto de forma global quanto para efeitos locais, como a mudança na posição relativa das Tyr138 e Tyr161. Além disso, foi possível entender o papel de estabilização de um dos estados alostéricos promovido pelos ácidos graxos.

Na terceira parte, foi estudado o efeito de duas mutações (D698N e F691L) na enzima FLT3 inibida pelo fármaco Gilteritinib. O início deste estudo é similar ao da Abl1, porém os valores de energia livre calculados não foram compatíveis com os obtidos experimentalmente. Por esse motivo, foi investigado o impacto das mutações na probabilidade de ativação da enzima. A Metadinâmica foi utilizada para realizar simulações de transição entre os inativo e ativo, obtendo também o perfil de energia livre de um caminho de mínima energia entre os dois estados. Os resultados destes cálculos possibilitaram entender a forma como as

---

<sup>1</sup>A representação de uma mutação é geralmente escrita da seguinte forma: (aminoácido nativo)(posição do aminoácido nativo)(aminoácido da mutação), onde geralmente se usa o código de uma letra para os aminoácidos (T: tirosina, I: isoleucina; V: valina, F: fenilalanina, L: leucina)

mutações promoveram o surgimento resistência ao gilteritinib.

## Seleção, Visualização e Edição de Estruturas

Para realizar as simulações foram selecionadas estruturas do Protein Data Bank (PDB) [6]. A visualização 3D, análise e parte da preparação das estruturas foram feitas com o software Discovery Studio versão 2021 [7]. Algumas proteínas encontradas no PDB estão com parte das suas cadeias laterais incompletas e para resolver este problema foi utilizada a ferramenta online *WHAT IF* [8]. A edição e parametrização de moléculas pequenas foi feita com uso da ferramenta online *CHARMM-GUI Input Generator* [9,10].

## Docking Molecular

Docking molecular é o procedimento de se encontrar a melhor pose entre duas moléculas, o que significa achar o melhor encaixe entre duas moléculas tridimensionais. É possível realizar esta busca por meio de programas gráficos, visualizando as moléculas e buscando a melhor situação que satisfaça algum critério já conhecido, como alguma interação intermolecular específica. Na situação mais comum, uma das moléculas é chamada de receptor (proteínas, enzimas, DNA, etc.) e a outra molécula de ligante (geralmente menor em relação ao receptor). O docking é um método de simulação computacional cujo objetivo é avaliar a formação de complexos receptor-ligante. Em muitos casos, o docking é a ferramenta principal do estudo para se entender mecanismos de reação, enquanto em outros, os cálculos utilizados como *screening virtual* [11,12].

De modo geral, uma simulação de docking segue os seguintes passos:

- planejamento do experimento in silico, com a escolha dos receptores e ligantes, seleção do sítio de ligação e raio da esfera de trabalho onde o algoritmo irá procurar a melhor pose;
- realização do experimento computacional para a formação de possíveis complexos;
- validação da simulação (redocking);

- avaliação dos resultados da simulação. (escores do docking, análise das interações receptor-ligantes em tela gráfica, correlação com dados experimentais).

Tal procedimento está resumido na Figura 4



Figura 4: Fluxograma resumindo os passos comuns em uma simulação de docking. A obtenção das estruturas para o experimento pode ser feita pelos bancos de dados (subseção anterior) ou por modelagem molecular. O algoritmo de docking varia de um software para outro, podendo utilizar mecânica molecular, método de Monte Carlo, algoritmo genético ou outro (Figura elaborada por Ignez Caracelli)

A seguir será descrito o programa utilizado neste trabalho para os estágios onde foi necessário utilizar cálculos de docking.

## O programa GOLD

Os principais métodos estocásticos são o Monte Carlo e o Algoritmo Genético. No primeiro é utilizado especificamente o método Metrópolis, onde o movimento aleatório do ligante no sistema é aceito ou rejeitado com base na probabilidade de Boltzmann. O segundo está baseado em algumas das características principais da evolução Darwiniana, utilizando ideias baseadas na linguagem da genética natural e evolução biológica. Constitui-se em uma poderosa ferramenta para a solução de problemas complexos. É um procedimento de busca não-determinístico que procura fornecer rapidamente boas, mas não necessariamente ótimas, soluções para problemas de otimização combinatorial que demandam recursos computacionais pesados demais para serem investigados utilizando algoritmos convencionais.

Os algoritmos genéticos têm sido utilizados em aplicações como o reconhecimento molecular e design de novas moléculas, visto que estes são, por natureza, problemas inerentemente combinatoriais [13, 14]. No programa GOLD, o algoritmo genético leva em conta o mapeamento entre átomos doadores em ligações de hidrogênio e pares de elétrons isolados no ligante e na proteína, além dos ângulos torcionais das moléculas do ligante e da proteína [15]. A Figura 5 mostra um fluxograma esquemático de um algoritmo genético relacionando-o com o Programa GOLD. Pode ser observada uma relação entre a linguagem da genética e a do algoritmo genético:

- gene representa cada parâmetro a ser explorado na otimização. No caso do GOLD, os parâmetros são translação, rotação e ângulos de torção dos ligantes;
- cromossomo, também chamado de indivíduo, é a cadeia de genes que representa uma solução possível para o problema. Para o GOLD, representa a posição, orientação e conformação do ligante;
- população é o conjunto de soluções no espaço de busca.

A seguir, apresenta-se a descrição das etapas do algoritmo genético e seu papel no Programa GOLD, conforme o esquema da Figura 5

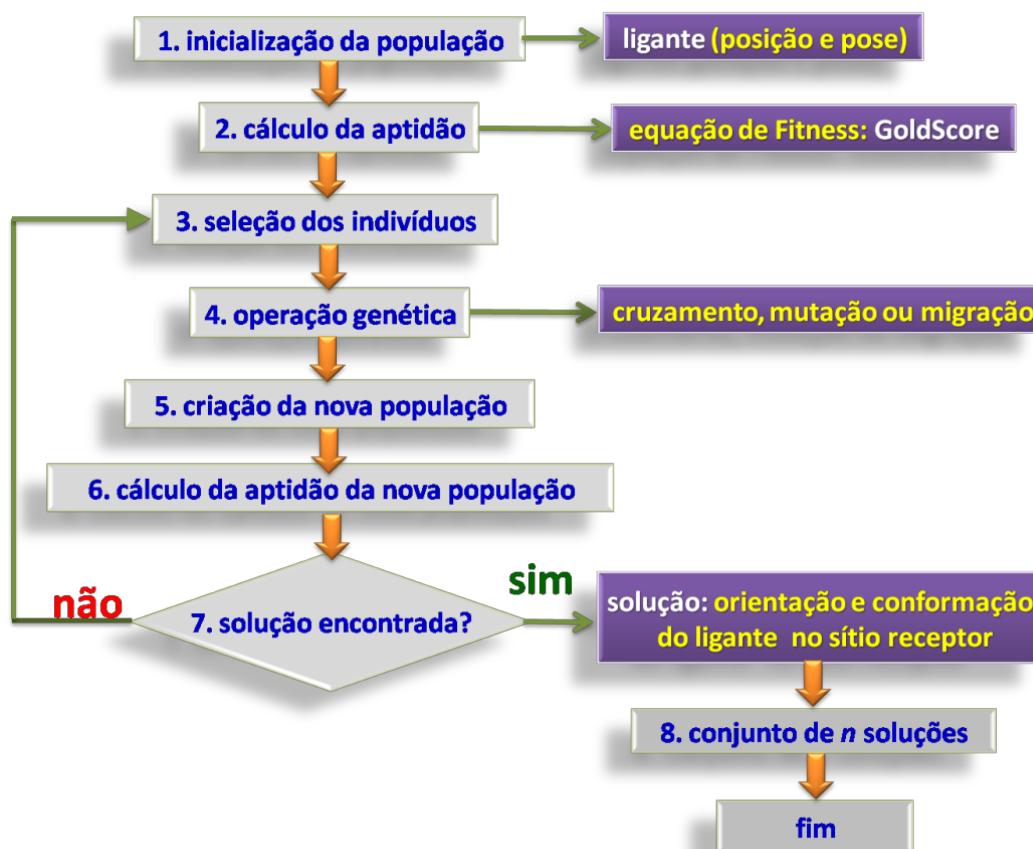


Figura 5: Algoritmo genético e o Programa GOLD (Figura elaborada por Ignez Caracelli)

A visualização gráfica é parte do processo de docking na seleção das moléculas, no estudo das interações, no estudo do padrão de orientações dos ligantes em um dado sítio, na comparação de moléculas e de sítios de ligação. É nesta etapa que o pesquisador toma suas decisões, ainda não possíveis de serem automatizadas pelos programas computacionais. São avaliadas as ligações de hidrogênio, as interações de van der Waals e as interações  $\pi$ . Quando se trata de interações  $\pi$  a situação piora, pois não é “detectável” pelos programas computacionais de docking nem está ainda “embutida” nos parâmetros dos programas de modelagem molecular ou de dinâmica molecular. A detecção destas interações só tem sido possível graças à análise criteriosa realizada em tela gráfica pelo pesquisador. A informação proporcionada pelo programa GOLD são as poses (orientação e conformação) do ligante em relação ao sítio receptor, uma pontuação que depende da função de ajuste escolhida. O programa trabalha com um método de ajuste do ligante ao sítio, considerando os aspectos conformacional e de energia do ligante e da macromolécula. No GOLD existem muitas funções de score disponíveis [15]. A função de ajuste mais utilizada é a GOLDScore que compreende quatro componentes:

$$GOLDScore\ Fitness = S(hb_{ext}) + 1,3750S(vdw_{ext}) + S(hb_{int}) + S(vdw_{int}) \quad (2)$$

$S(hb_{ext})$ : energia de ligação de hidrogênio do complexo proteína-ligante;

$S(vdw_{ext})$ : energia de Van der Waals entre proteína-ligante;

$S(vdw_{int})$ : energia de Van der Waals no ligante;

$S(hb_{int})$ : energia de ligação de hidrogênio intramolecular do ligante.

O escore  $vdw_{ext}$  é multiplicado por um fator de 1,375 quando o escore total é calculado. Isto é uma correção empírica para aumentar a probabilidade de proteína-ligante realizar contato hidrofóbico. Para converter o valor final de energia para um score, resultado total é multiplicado por -1. O campo de força (mecânica molecular) utilizado no programa GOLD é o Tripos 5.2 Force Field [16]. Para análise das interações inter e intramoleculares [17] e das características estruturais das enzimas, serão utilizados programas de visualização gráfica (DS Visualizer [7], Chimera [18]).

## Dinâmica Molecular (DM)

A Dinâmica Molecular Clássica (DM) tem como técnica central a resolução da segunda lei de Newton para sistemas com  $N$  átomos interagentes, como mostrado na Eq. 3

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = F_i, \quad (i = 1, \dots, N) \quad (3)$$

Onde a força é calculada pela derivada do potencial:

$$F_i = -\frac{\partial}{\partial \mathbf{r}_i} V(\mathbf{r}_1, \dots, \mathbf{r}_N) \quad (4)$$

Estas equações são resolvidos simultaneamente, dependendo do método de integração escolhido, com um passo de tempo pequeno (geralmente 2 fs). A forma do potencial irá depender do campo de força escolhido. Ao usar a segunda lei de Newton algumas simplificações devem ser feitas para adaptar comportamentos do sistema que são de origem quântica, como as ligações covalentes e interações que envolvam átomos de hidrogênio [19–21].

## Condições Iniciais

Em uma sequência padrão de passos para realizar uma corrida de Dinâmica Molecular, primeiro se define as condições iniciais do sistema: escolhe-se os vetores que irão delimitar forma e tamanho da caixa, condições periódicas de contorno e a posição da molécula principal em relação às paredes da caixa. Após isso, realiza-se a minimização de energia (explicada mais adiante) e, em seguida, o sistema precisa ser inicializado da forma correta, dependendo do método de integração. Neste estágio, existe apenas informação das coordenadas, porém alguns algoritmos de integração, como o integrador leap-frog (mais conhecido como método de integração MD), requer a informação sobre as velocidades que precedem o tempo inicial da simulação ( $t = t_0 - \frac{1}{2}\Delta t$ ). Além disso, caso o arquivo salvo no final de um dos estágios da dinâmica tenha extensão .pdb no lugar de .gro, apenas as coordenadas serão salvas para iniciar o próximo.

Em ambos os casos acima será necessário gerar as velocidades iniciais de cada átomo do sistema. O gromacs consegue fazer isso escolhendo randomicamente um valor de velocidade na distribuição de Maxwell-Boltzman (Eq. 5)

$$\rho(v_i) = \sqrt{\frac{m_i}{2\pi kT}} \exp\left(-\frac{m_i v_i^2}{2kT}\right) \quad (5)$$

A velocidade obtida é então multiplicada pelo desvio padrão da distribuição normal ( $kT/m_i$ ) e a energia resultante é corrigida, removendo-se o movimento do centro de massa e escalando as velocidades para corresponderem à temperatura T, segundo o teorema da equipartição de energia:

$$\left\langle \frac{1}{2} \sum_{i=1}^N m_i v_i^2 \right\rangle = \frac{1}{2} N_f kT \quad (6)$$

onde  $N_f = 3N - N_c - N_{com}$ , ou seja, o número de graus de liberdade considerando as restrições ( $N_c$ ) e a remoção do movimento do centro de massa ( $N_{com}$ ).

## Método de integração

A integração das equações de movimento na Dinâmica Molecular depende do método de integração utilizado. No caso do gromacs, o integrador padrão é chamado algoritmo



leap-frog [22]. Neste método de integração, posição e velocidade são atualizados relativos ao passo de tempo  $\Delta t$  de modo intercalado, como mostrado nas equações a seguir:

$$\mathbf{v}(t + \frac{1}{2}\Delta t) = \mathbf{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\mathbf{F}(t) \quad (7)$$

$$\mathbf{r}(t + \Delta t) = \mathbf{r}(t) + \Delta t\mathbf{v}(t + \frac{1}{2}\Delta t) \quad (8)$$

Desse modo, velocidade e posição são atualizadas em intervalos que saltam um sobre os outros (Figura 6), como sapos, de onde deriva o nome leap-frog.

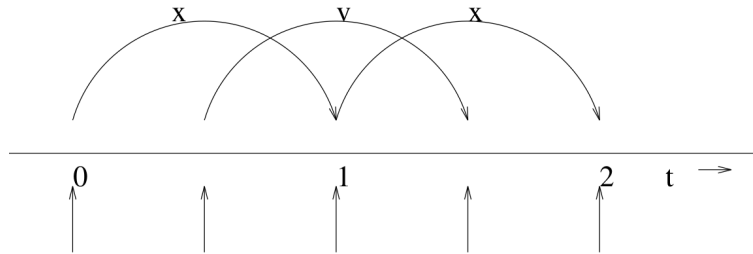


Figura 6: Esquema do método de integração padrão do gromacs, integrador Leap-frog (Figura adaptada do manual do gromacs de 2022)

As equações acima são modificadas de acordo com o acoplamento de temperatura e pressão escolhidos, assim o ensemble escolhido irá adaptar as equações do integrador para manter constantes apenas a temperatura (NVT) ou ambos (NPT). Mais informações podem ser consultadas no manual do gromacs (<https://manual.gromacs.org/current/index.html>).

## Campo de Força

A forma analítica das equações de movimento dependem do campo de força escolhido. Este, por sua vez, terá um conjunto de parâmetros específicos para cada átomo com seus primeiros ligantes e interações do tipo não-ligadas. Assim, é possível dividir o potencial em duas porções, uma mediada por ligações covalentes (intra) e outra por interações intermoleculares (inter), como mostrado na Eq. 9

$$U = U_{intra} + U_{inter} \quad (9)$$

Para o gromacs existem diversos campos de força disponíveis. Cada um deles levam

em consideração aproximações diferentes na parametrização dos elementos químicos e suas interações. Para este trabalho foi utilizado o campo de força Charmm36 [23, 24], no qual cada parte do potencial é tratado da seguinte forma:

$$U_{intra} = \sum_{\text{ligados}} K_b(b - b_0)^2 + \sum_{\text{angulos}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedros}} K_\psi(1 + \cos(n\psi - \delta)) + \sum_{\text{improprios}} K_\phi(\phi - \phi_0)^2 + \sum_{\text{Urey-Bradley}} K_{UB}(r_{1,3} - r_{1,3,0})^2 \quad (10)$$

$$U_{inter} = \sum_{\text{n\~{o}a-ligados}} \frac{q_i q_j}{4\pi D r_{ij}} + \epsilon_{ij} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{min,ij}}{r_{ij}} \right)^6 \right] \quad (11)$$

Na Eq. 10 vê-se que os termos referentes às ligações, aos ângulos e os diedros impróprios (torções fora do plano) são tratados como osciladores harmônicos, com as suas respectivas constantes de mola ( $K_b$ ,  $K_\theta$  e  $K_\phi$ , com o termo 1/2 da lei de Hooke incorporado nas constantes), enquanto os diedros são descritos como termos sinusoidais. O ultimo termo desta equação (Urey-Bradley), também tratado como um oscilador harmônico, se refere à interação

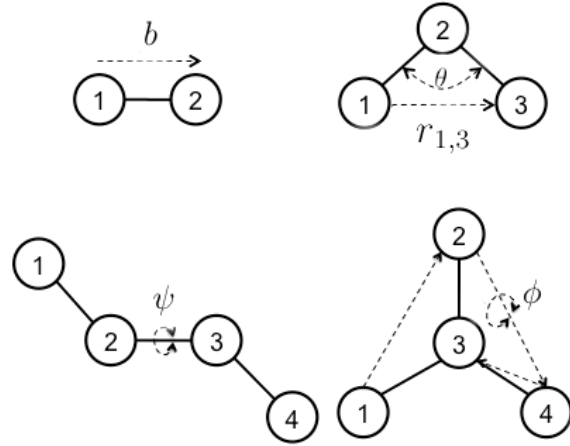


Figura 7: Esquema mostrando os parâmetros da parte do potencial intramolecular (Adaptado de [25]).

entre os átomos 1 e 3 dos três átomos que descrevem o ângulo  $\theta$  (veja um esquema de cada variável na Figura 7). Os parâmetros  $b$ ,  $\psi$ ,  $\phi$  se referem ao comprimento da ligação covalente, ângulo de torção dos diedros e diedros impróprios, respectivamente, com o subscrito 0 para indicar os valores das posições de equilíbrio.

Já a Eq. 11 trata das interações entre átomos não ligados, sendo o primeiro termo referente às interações Coulombianas, onde  $q$  é a carga parcial dos átomos interagentes. O segundo termo corresponde ao potencial de Lennard-Jones, com a profundidade do poço de potencial dada por  $\epsilon_{ij}$  e o raio mínimo de interação é dado por  $R_{min,ij}$ . Em ambos os termos  $r_{ij}$  é a distância entre o par de átomos da interação

## Minimização de energia

Um estágio fundamental da dinâmica consiste em levar o sistema para o mínimo local mais próximo da configuração inicial. Para este método, ao gerar o modelo 3D da proteína, a informação sobre os hidrogênios se perde e é necessários adicioná-los por meio de algum software (neste caso o Discovery Studio). Isto, somado ao passo de adicionar as moléculas de água e íons em posições aleatórias na construção da caixa de simulação, faz surgir forças de interação muito altas. Por esse motivo, uma curta minimização de energia é necessária para corrigir a posição desses novos componentes [26]. Além disso, o fato de o sistema estar em condições diferentes das encontradas na formação do cristal também faz dessa necessária para evitar artefatos como forças excessivas entre átomos, e consequentemente falhas na dinâmica.

O gromacs dispõe de três métodos: steepest descent, conjugate gradients e L-BFGS. Todos esses métodos não tem a intenção de encontrar o mínimo global do sistema, mas sim o mínimo local mais próximo da configuração inicial. Neste trabalho foi utilizado apenas o primeiro método dos três listados acima. A seguir é feita uma breve explicação de como ele funciona. Mais detalhes sobre os métodos de minimização de energia podem ser consultados no manual do gromacs.

### Steepest Descent

A ideia principal deste algoritmo é de atualizar o vetor posição dos átomos do sistema  $\mathbf{r}$  da seguinte forma:

$$\mathbf{r}_{k+1} = \mathbf{r}_k + \frac{\mathbf{F}_k}{\max(|\mathbf{F}_k|)} d_k \quad (12)$$

sendo  $d_k$  é o deslocamento máximo e  $\mathbf{F}$  força entre os átomos. Como  $F = -\nabla V$ , o segundo termo do lado direito da equação resultará em um deslocamento de módulo  $d_k$  na direção negativa do gradiente do potencial  $V$ , com um fator de normalização dado pela força máxima entre átomos do sistema. O algoritmo é iniciado com um valor inicial para o deslocamento ( $d_0$ ) e para cada interação ocorre o seguinte:

Se ( $V_{k+1} < V_k$ ), o deslocamento ocorreu no sentido da minimização; o passo é aceito e a posição é atualizada pra  $d_{k+1} = 1.2d_k$ .

Caso contrário, se ( $V_{k+1} \geq V_k$ ), a nova posição é rejeitada e o deslocamento é reduzido para  $d_k = 0.2d_k$ .

Assim fica evidente que este algoritmo não irá atravessar barreiras de potencial para procurar o mínimo global, muito menos um mínimo local menor, pois são aceitos apenas os deslocamentos que reduzem o potencial [27]. Mas apenas essas duas condições por si só gerariam interações infinitas, precisando assim de mais uma condição. Neste caso o usuário, via `mdp`<sup>2</sup> script, um valor de força mínima a ser alcançada ou um número máximo de passos. Alcançados um desses dois limites, o processo de minimização é finalizado.

## Ensembles

Ensemble estatístico, primeiramente definido por Josiah Willard Gibbs [28], é uma abstração matemática que consiste em considerar múltiplas réplicas de um mesmo sistema, cada uma representando uma possível configuração do espaço de fase do sistema real, também chamado microestado. Com o número de réplicas grande o suficiente, é possível fazer a aproximação de que a porcentagem de estados repetidos é igual a probabilidade de que esse estado ocorra no sistema real. No caso de um ensemble termodinâmico, além da definição acima é necessário que o sistema esteja em equilíbrio [29].

O conceito de ensemble é importante pois para uma dada condição do experimento é mais apropriado uma formulação teórica do que outra. Por exemplo, pode-se fazer um experimento isolando o sistema totalmente do meio externo, mantendo o seu conteúdo de energia (E), volume (V) e número de partículas (N) fixos, mas deixando temperatura (T) e pressão (P) variarem. Esta situação corresponde ao ensemble chamado de Microcanônico (ou NVE), o qual tem uma formulação teórica específica para se calcular a probabilidade de cada microestado e para implementação na dinâmica molecular.

Usando a mesma lógica, pode-se escolher manter a temperatura (NVT) ou ainda temperatura e pressão ao mesmo tempo (NPT). No caso de um experimento, para manter esses parâmetros utiliza-se os chamados reservatórios de temperatura e pressão, que podem consistir em sistemas mecânicos maiores em contato com sistema de interesse, em que as trocas de calor e variações de volume são desprezíveis. Já em uma simulação de dinâmica molecular, as equações de movimento são modificadas com base em algoritmos

---

<sup>2</sup>Formato de arquivo que o gromacs reconhece para leitura de parâmetros

de acoplamento de temperatura e pressão (termostatos e barostatos) [30–33].

É comum que em laboratórios os experimentos com proteínas estejam sujeitos às condições de temperatura e pressão do local, as quais são aproximadamente constantes. Devido a isso é comum que se escolha o ensemble NPT para as simulações de dinâmica molecular. Por esse motivo, neste trabalho foi utilizado este ensemble em todas as simulações.

## Análise de Componentes Principais

Sistemas com grande número de átomos, como proteínas em uma solução de água e íons, tem uma quantidade grande de coordenadas para se analisar, mais especificamente  $3N$  graus de liberdade cartesianos e  $3N$  das velocidades. Ao tentar descrever uma mudança estrutural, como uma transição alostérica, é possível ver que boa parte dessa movimentação não é relevante para o fenômeno a ser estudado. Por esse motivo se torna interessante filtrar as movimentação harmônicas em torno de um mínimo, restando apenas as vibrações que irão contribuir para a transição de conformação geométrica.

É com essa ideia que surge o método *Essential Dynamics Sampling* [34], baseado na ideia de Análise de Componentes Principais. Este método consiste em usar as flutuações de movimento dos átomos para construir um novo eixo de coordenadas, onde os vetores principais representem a maior movimentação do sistema. Para isso, inicialmente calcula-se a matriz de covariância:

$$C_{ij} = \left\langle M_{ii}^{1/2}(x_i - \langle x_i \rangle) M_{jj}^{1/2}(x_j - \langle x_j \rangle) \right\rangle \quad (13)$$

onde  $\mathbf{x} = (x_1, \dots, x_{3N})$  é um vetor coluna das coordenadas da trajetória de uma dinâmica molecular realizada previamente, implicando em uma matriz de covariância  $3N \times 3N$ ; o símbolo  $\langle \rangle$  denota média temporal; já  $M$  é uma matriz diagonal com as massas atômicas, que podem ser escolhidas como 1 para uma análise sem massas com pesos para as coordenadas atômicas. A matriz  $C$  pode ser diagonalizada utilizando a matriz  $R$  dos autovetores:

$$\mathbf{R}^T \mathbf{C} \mathbf{R} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_{3N}) \quad (14)$$

Cada coluna de  $\mathbf{R}$  é um autovetor da matriz de covariância, também chamados de modos principais ou essenciais. O autovetores  $\lambda_1$  a  $\lambda_{3N}$  são organizados numa sequência decrescente ( $\lambda_1 \geq \lambda_2 \geq \dots$ ). Com isso, é possível obter vetores ortogonais, chamados de componentes principais ( $p_i(t)$ ), sobre as quais é possível projetar a trajetória:

$$\mathbf{p}(t) = \mathbf{R}^T \mathbf{M}^{1/2} (\mathbf{x}(t) - \langle \mathbf{x} \rangle) \quad (15)$$

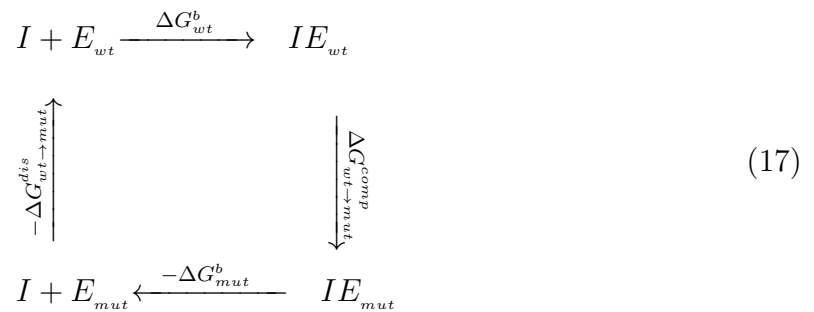
Como  $\lambda_i$  é o quadrado da média das flutuações da componente principal  $i$ , quanto maior o seu valor, maior será a movimentação do sistema naquela direção. Assim, a componente principal  $p_1$  representa a direção preferencial de movimento, seguida por  $p_2$ ,  $p_3$  e assim por diante. Por consequência, uma quantidade pequena dos primeiros componentes principais irão representar uma porcentagem grande da movimentação do sistema [34].

O gromacs possibilita realizar a projeção da trajetória sobre os autovetores e ainda filtrar para que apenas a movimentação sobre determinadas componentes principais seja representada. Assim, a trajetória projetada sobre a componente principal  $i$  fica:

$$\mathbf{x}(t) = \langle \mathbf{x} \rangle + M^{1/2} R_{*i} p_i(t) \quad (16)$$

## Perturbação de Energia Livre (FEP)

Considere a inibição de uma enzima (E) por um inibidor (I), onde tal enzima pode sofrer mutação um de seus aminoácidos. Para avaliar a diferença de variação de energia livre ( $\Delta\Delta G^b$ ) associada a esta mutação, é possível construir o seguinte ciclo termodinâmico (Eq. 17):



onde *wt* e *mut* indicam a forma nativa<sup>3</sup> e mutante, respectivamente; *comp* indica que a mutação ocorre com complexo enzima-inibidor formado e *dis* para a o dissociado.

Como este ciclo é fechado, a diferença de variação de energia livre pode ser calculada de duas formas:

$$\Delta\Delta G^b = \Delta G_{mut}^b - \Delta G_{wt}^b \quad (18)$$

$$\Delta\Delta G^b = \Delta G_{wt \rightarrow mut}^{comp} - \Delta G_{wt \rightarrow mut}^{dis} \quad (19)$$

Dependendo da equação escolhida (Eq. 18 ou 19) utiliza-se um método computacional diferente, porém ambos se baseiam na Perturbação de Energia Livre (FEP<sup>4</sup>). Para explicar mais detalhadamente este método, primeiro considere somente a situação em que o sistema vai do estado nativo para a mutação. Nesta situação é possível escrever a função de partição no ensemble das pressões (NPT) com base na hamiltoniana ( $\mathcal{H}_i$ ) de cada estado:

$$Q_i(N, P, T) = \frac{1}{V_0 N! h^{3N}} \int_0^\infty dV \int e^{-\beta(\mathcal{H}_i(\mathbf{r}, \mathbf{p}) + PV)} d^N \mathbf{p} d^N \mathbf{r} \quad (20)$$

sendo  $V_0$  um volume de referência,  $N$  é o número de átomos do sistema,  $h$  é a constante de Planck,  $\beta = 1/k_B T$ ,  $P$  é a pressão do sistema,  $(\mathbf{r}, \mathbf{p})$  são os vetores do espaço de fase e o subscrito  $i$  indica um dos quatro vértices do ciclo termodinâmico.

Neste ensemble, o potencial termodinâmico de Gibbs é dado por:

$$G_i(N, P, T) = -k_B T \ln(Q_i(N, P, T)) \quad (21)$$

É possível usar as Eqs. 20 e 21 para calcular a variação de energia livre de qualquer uma das transições mostradas no ciclo termodinâmico. Como neste trabalho optou-se pelo método da Eq. 19, segue que:

$$\Delta G_{wt \rightarrow mut} = -k_B T \ln\left(\frac{Q_{mut}}{Q_{wt}}\right) = -k_B T \ln\left(\frac{1}{Q_{wt}} \int_0^\infty dV \int e^{-\beta(\mathcal{H}_{mut}(\mathbf{r}, \mathbf{p}) + PV)} d^N \mathbf{p} d^N \mathbf{r}\right) \quad (22)$$

---

<sup>3</sup>a forma nativa é normalmente designada como *wild* em inglês)

<sup>4</sup>do inglês Free Energy Perturbation

Reescrevendo  $e^{-\beta(\mathcal{H}_{mut}(\mathbf{r},\mathbf{p})+PV)} = e^{-\beta(\mathcal{H}_{wt}(\mathbf{r},\mathbf{p})+PV)}e^{-\beta(\mathcal{H}_{mut}(\mathbf{r},\mathbf{p})-\mathcal{H}_{wt}(\mathbf{r},\mathbf{p}))}$  é possível notar que a quantidade dentro do logaritmo natural no membro da direita da Eq. 22 é, por definição, uma média de ensemble sobre o estado  $wt$ :

$$\Delta G_{wt \rightarrow mut} = -k_B T \ln \left\langle e^{-\beta(\mathcal{H}_{mut}(\mathbf{r},\mathbf{p})-\mathcal{H}_{wt}(\mathbf{r},\mathbf{p}))} \right\rangle_{wt} \quad (23)$$

Geralmente os estados do sistema no ciclo termodinâmico estão separados por uma diferença de energia grande o suficiente para que a exponencial da Eq. 23 seja muito pequena. Desta forma, as configurações do sistema terão um peso estatístico muito pequeno e a energia livre irá convergir lentamente. Para resolver este problema é possível segmentar a transformação de um estado para outro em estados intermediários dado por uma combinação linear dos dados extremos:

$$\mathcal{H}(\lambda) = (1 - \lambda)\mathcal{H}_{wt} + \lambda\mathcal{H}_{mut} \quad , \quad (24)$$

com  $\lambda(t = 0) = 0$  e  $\lambda(t = \tau) = 1$ , onde  $\lambda$  cresce linearmente para cada passo de tempo partindo de 0 até 1, e  $\tau$  é o tempo para uma transição completa.

Sendo estes estados intermediários apenas uma combinação entre as hamiltonianas dos extremos, eles não possuem significado físico. Estes são denominados **estados alquímicos** e servem apenas para melhorar a convergência no cálculo de energia livre, fazendo com que a diferença da hamiltoniana entre cada estado seja apenas uma pequena perturbação em relação aos estados vizinhos. Com isso, usando a Eq 24 a função de partição fica:

$$Q(N, P, T, \lambda) = \frac{1}{V_0 N! h^{3N}} \int_0^\infty dV \int e^{-\beta(\mathcal{H}(\mathbf{r},\mathbf{p},\lambda)+PV)} d^N \mathbf{p} d^N \mathbf{r} \quad (25)$$

Escrevendo a definição do potencial termodinâmico de Gibbs com a Eq. 25 e derivando em relação a  $\lambda$ , fica:



$$\begin{aligned}
\frac{\partial G}{\partial \lambda} &= -\frac{-k_B T}{Q} \frac{\partial Q}{\partial \lambda} = \\
&= -\frac{-k_B T}{Q} \frac{\partial}{\partial \lambda} \int_0^\infty dV \int e^{-\beta(\mathcal{H}(\mathbf{r}, \mathbf{p}, \lambda) + PV)} d^N \mathbf{p} d^N \mathbf{r} \\
&= \frac{1}{Q} \int_0^\infty dV \int \left( \frac{\partial \mathcal{H}}{\partial \lambda} \right) e^{-\beta(\mathcal{H}(\mathbf{r}, \mathbf{p}, \lambda) + PV)} d^N \mathbf{p} d^N \mathbf{r} \\
&= \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle
\end{aligned}$$

Com isso, obtém-se:

$$\Delta G_{wt \rightarrow mut} = \int_0^1 \frac{\partial G}{\partial \lambda} d\lambda = \int_0^1 \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_{\lambda=0} d\lambda \quad (26)$$

A Eq. 26 tem melhor convergência quanto menor for a diferença de energia entre os estados alquímicos. Para se aplicar esta equação em uma simulação é preciso que a transição ocorra de forma lenta o suficiente para que o  $\Delta \lambda$  seja pequeno e o sistema não se afaste muito do equilíbrio. Utilizando-se das mesmas estratégias é possível ainda mostrar que em uma transformação fora do equilíbrio o trabalho realizado no caminho da transição pode ser calculado como:

$$W = \int_0^1 \frac{\partial \mathcal{H}}{\partial \lambda} d\lambda \quad (27)$$

Esta expressão possibilita derivar expressões para cálculo de energia livre com transições que ocorram de forma mais rápida. Neste contexto, existem três principais equações, cuja deduções seguem princípios similares ao exemplificado acima e podem ser vistas em detalhes nas referências [35–37]. A ideia geral destes métodos consiste em construir um ensemble em equilíbrio para cada extremo da transformação e realizar múltiplas transições na direção direta e reversa. Cada transição terá um trabalho associado, dado pela Eq. 27. Dentre estes três métodos foi utilizado neste trabalho a Relação de Aceitação de Bennet (BAR)<sup>5</sup> por ter melhor acurácia e menos viés no cálculo de energia livre do que os outros métodos que utilizam médias de exponenciais [38], além de ser o único que possui um estimador de erro analítico.

---

<sup>5</sup>Bennet's Acceptance Ratio em inglês

A energia livre pode ser obtida com o método BAR resolvendo-se a seguinte equação:

$$\sum_{i=1}^{N_f} \frac{1}{1 + \frac{N_f}{N_r} e^{\beta(W_i - \Delta G)}} = \sum_{j=1}^{N_r} \frac{1}{1 + \frac{N_r}{N_f} e^{-\beta(W_j - \Delta G)}} \quad (28)$$

onde  $N_f$  e  $N_r$  são o número de transições na direção direta e reversa da transformação.

Já a convergência pode ser avaliada para o total de transições pelo estimador de erro do método BAR [38]:

$$\frac{1}{\beta^2 N_{tot}} \left[ \left\langle \frac{1}{2 + 2 \cosh [\beta(W_i - \Delta G)]} \right\rangle^{-1} - \left( \frac{N_{tot}}{N_f} + \frac{N_{tot}}{N_r} \right) \right] \quad (29)$$

## Metadinâmica

Em um sistema de  $N$  átomos é necessário resolver as equações de Hamilton, que formam um sistema de  $2N$  equações do tipo:

$$\begin{aligned} \dot{\mathbf{q}}_i &= \frac{\mathbf{p}_i}{m_i} \\ \dot{\mathbf{p}}_i &= -\frac{\partial \mathcal{H}}{\partial \mathbf{q}_i} \end{aligned} \quad (30)$$

onde  $\mathbf{q}_i$  e  $\mathbf{p}_i$  são vetores posição e momento do átomo  $i$  do sistema e  $\mathcal{H}$  a hamiltoniana, dada por:

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}) \quad (31)$$

A função energia potencial  $U(\mathbf{q})$  descreve a interação interatômica, cuja forma dependerá do campo de força utilizado na simulação.

No ensemble das pressões (NPT), utilizando a Eq. 20, é possível calcular a distribuição probabilidade de se encontrar o sistema em um dado estado  $(\mathbf{p}, \mathbf{q})$  usando a relação:

$$\rho(\mathbf{p}, \mathbf{q}, V) = \frac{e^{-\beta(\mathcal{H}(\mathbf{p}, \mathbf{q}) + PV)}}{Q(N, P, T)} \quad (32)$$

Caso a simulação corra por tempo suficientemente grande, é possível dizer que todos os estados pertencentes ao ensemble NPT serão acessados, com base na *hipótese ergódica* [39], ou seja, médias temporais equivalem à médias sobre um ensemble do sistema em questão. Porém, com os algoritmos e o poder computacional atuais, alguns eventos raros, como transição entre estados separados por barreiras energéticas relativamente grandes, precisarão de um tempo de simulação muito grande. Por esse motivo é preciso procurar outras maneiras de melhorar a amostragem em uma região específica do espaço de fase.

Além disso, no espaço 3D, considerando as Eqs.30, será preciso 6N equações para descrever o sistema, tornando inviável a utilização delas separadamente para caracterizar as mudanças de estado desse sistema. Por esse motivo recorre-se ao conceito de variáveis coletivas [40] (CV<sup>6</sup>). Pode-se chamar de CV uma função das coordenadas atômicas que seja capaz de descrever as alterações ocorridas no sistema para processo em questão. Ângulos torcionais de uma molécula e distância ente dois átomos ou entre centros de massa em regiões específicas de uma proteína são exemplos de CVs que podem ser utilizadas no estudo de transições entre mínimos de energia do sistema.

Para isso, sendo  $\mathbf{s}(q)$  um vetor das variáveis coletivas no ensemble NPT, é possível mostrar que a densidade de probabilidade do sistema pode ser escrita em função deste vetor utilizando a energia livre de Gibbs ( $G(\mathbf{s})$ ):

$$\rho(\mathbf{s}) \propto e^{-\beta G(\mathbf{s})} \quad (33)$$

onde se define  $G(\mathbf{s})$  da seguinte forma [41]:

$$G(s) = -k_B T \ln \int \delta(s(\mathbf{q}) - s) e^{-\beta U(\mathbf{q})} d\mathbf{q} + C \quad (34)$$

sendo C é uma constante arbitrária e  $\delta$  é a função delta de Dirac, a qual seleciona todas as configurações  $s$  e resulta em um perfil de energia livre em função das CVs.

Dessa forma, reduz-se a análise de 6N coordenadas à dimensão do vetor  $\mathbf{s}(q)$ . Isto torna mais simples a tarefa de identificar os estados físicos que se quer estudar. O perfil de energia livre gerado pela Eq. 34 permite estimar a taxa de transição entre estados por

---

<sup>6</sup>do inglês *Collective Variables*

meio da teoria da transição estado, segundo a qual essa taxa ( $\eta$ ) é uma função da barreira energética ( $\Delta G^\ddagger$ ) que separa o estado inicial do final:

$$\eta = \eta_0 e^{-\beta \Delta G^\ddagger} \quad (35)$$

onde  $\eta_0$  é um pré-fator que depende de constantes de tempo do sistema em questão.

Como exemplo, considere a situação de uma proteína alostérica com dois estados possíveis. Entre essas duas conformações existe um estado energeticamente instável, denominado estado de transição (Figura 8), onde  $\Delta G(s) = \Delta G^\ddagger$ . A Eq. 35 diz que quanto maior essa barreira menor será a taxa de transição e, conseqüentemente um maior tempo para que se possa observar uma transição. A maioria das

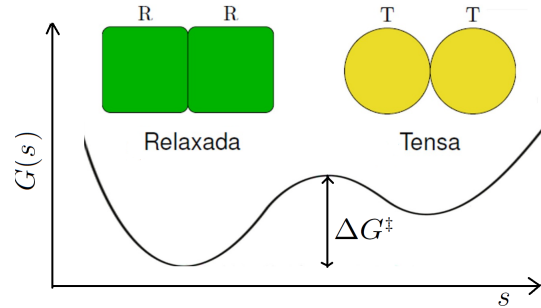


Figura 8: Perfil de energia livre em função de  $s_0$  de uma proteína alostérica formada por duas subunidades (Figura elaborada pelo autor).

proteínas alostéricas possuem  $\Delta G^\ddagger$  significativamente grande, de modo que seria necessário simulações da ordem microssegundos [42]. Com o poder computacional atual, seguindo os algoritmos de resolução das equações de movimento de todos os átomos do sistema, simulações dessa magnitude podem levar alguns meses, tornando inviável a utilização de dinâmica molecular pura. Por esse motivo é necessário recorrer a técnicas avançadas de amostragem do espaço de fase, onde as variáveis coletivas são utilizadas para acelerar a transição entre dois estados de forma que não se perca informações relevantes da física do sistema em estudo, introduzindo um viés que possa ser fatorado e recuperar a informação sem viés.

A seguir será descrita uma das técnicas de amostragem melhorada do espaço de fase com base em variáveis coletivas, a metadinâmica bem temperada.

## Metadinâmica Bem Temperada (wT-metaD)

Existem várias maneiras de se aplicar a metadinâmica. A ideia central é a de adicionar um potencial dependente das variáveis coletivas à hamiltoniana do sistema, de forma a facilitar sua passagem pela barreira de energia livre que separa dois estados. Como

o potencial é aditivo será possível recuperar a física sem viés do sistema da seguinte forma:

considere a nova hamiltoniana  $\tilde{\mathcal{H}}$  gerada pela adição do potencial  $\tilde{U}(s(\mathbf{q}))$ :

$$\mathcal{H} = \sum_{i=1}^N \frac{\mathbf{p}_i^2}{2m_i} + U(\mathbf{q}) \rightarrow \tilde{\mathcal{H}} = \mathcal{H} + \tilde{U}(s(\mathbf{q})) \quad (36)$$

o que implica, pela Eq. 34, em

$$\begin{aligned} \tilde{G}(s) &= -k_B T \ln \int \delta(s(\mathbf{q}) - s) e^{-\beta[U(\mathbf{q}) + \tilde{U}(s(\mathbf{q}))]} d\mathbf{q} + \tilde{C} \\ &= -k_B T \ln e^{-\beta[\tilde{U}(s(\mathbf{q}))]} \int \delta(s(\mathbf{q}) - s) e^{-\beta[U(\mathbf{q})]} d\mathbf{q} + \tilde{C} \\ &= G(s) + \tilde{U}(s) + \tilde{C} + C \end{aligned} \quad (37)$$

Assim, conhecendo a forma analítica de  $\tilde{U}(s(\mathbf{q}))$  é possível recuperar a energia livre original  $G(s)$  da obtida com a adição do potencial  $\tilde{G}(s)$ . Para a metadinâmica bem temperada (wT-metaD) o potencial de viés proposto é da seguinte forma [43]:

$$\tilde{U}(\vec{s}, t) = w \int_0^{t'} e^{-\tilde{U}(\vec{s}, t)/k_B \Delta T} \exp\left(-\sum_{i=1}^{N_{cv}} \frac{(s_i - s_i(t'))^2}{2\sigma_i^2}\right) dt' \quad (38)$$

onde a primeira exponencial tem a função de reduzir a altura inicial  $w$  da gaussiana e  $\Delta T$  é um parâmetro com dimensão de temperatura, de modo que as regiões já visitadas tenham valores de potencial depositados menores do que as regiões ainda não visitadas, fazendo o sistema se distanciar das configurações do espaço de fase inicial. Dessa maneira, o sistema irá explorar o espaço de fase, depositando gaussianas em regiões de poços de potencial até o seu preenchimento e, conseqüentemente, cruzando as barreiras energéticas. No longo limite de tempo, devido à sua forma analítica, é possível mostrar que o potencial de viés tenderá a [40]:

$$\tilde{U}(s) = -\frac{\Delta T}{T + \Delta T} (G(s) - C(t)) \quad (39)$$

Nessa situação, das Eqs. 33, 37 e 39 obtém-se:

$$\rho(\mathbf{s}) \propto e^{-\frac{G(\mathbf{s})}{k_B(T + \Delta T)}} \quad (40)$$

Dessa forma, o potencial de viés terá efeito similar ao de aumentar a temperatura

em  $\Delta T$  do ensemble para as variáveis coletivas, enquanto as demais variáveis cartesianas terão sua dinâmica determinada pela temperatura  $T$ .

# Capítulo 1

## ABELSON TIROSINA QUINASE <sup>1</sup>

### 1.1 Introdução

A classe de proteínas conhecida como *Quinase* é responsável por transferir grupos fosfato para substratos específicos. Essa transferência ocorre geralmente em duas etapas: primeiro uma molécula doadora de fosfato (o ATP<sup>2</sup>) se liga à enzima e transfere um grupo fosfato para um dos aminoácidos do sítio; em seguida o substrato se liga ao sítio ativo e recebe o grupo fosfato, finalizando o processo. Este mecanismo é denominado fosforilação e implica na formação de um estado ativo para a enzima (fosforilada) e um inativo (não fosforilada) e podem ser considerados estados alostéricos [5, 46].

A maioria das quinases sofrem mudança de conformação após serem fosforiladas, que é o caso da enzima Abelson Tirosina Quinase (Abl1) (foco deste trabalho). Como mostrado na Figura 1.1, a Abl1 possui um sítio para o ATP localizado entre dois loops, que ao se ligar à enzima um grupo fosfato é transferido para o loop-P. Isso causa uma mudança nas interações intermoleculares, resultando na abertura do loop-A. Este mecanismo torna o sítio do substrato (o peptídeo Abltide) acessível e a enzima pode finalizar o processo de fosforilação [45].

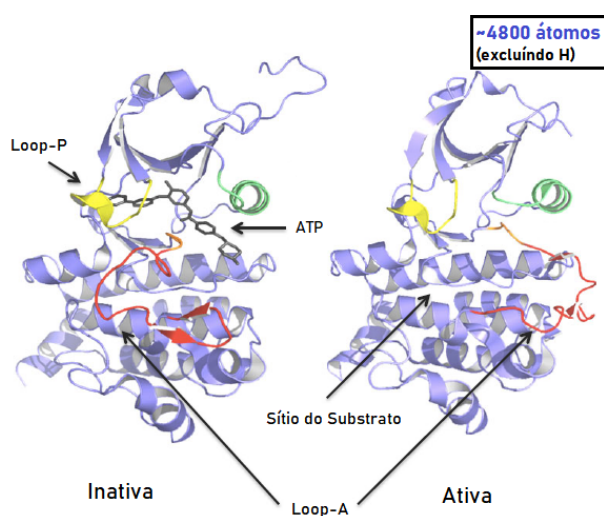


Figura 1.1: Esquema mostrando a mudança de estado da enzima Abl1. As regiões coloridas em verde, amarelo e vermelho sofrem modificações conformacionais após a transferência de um grupo fosfato pelo ATP, sendo a abertura do loop-A a mudança mais perceptível. (Fonte: Adaptada da referência [45])

<sup>1</sup> Trabalho publicado (Anexo 1): [44] Dávila-Rodríguez, M. J., Freire, T. S., Lindahl, E., Caracelli, I., Zukerman-Schpector, J., and Friedman, R., "Is breaking of a hydrogen bond enough to lead to drug resistance" *Chemical Communications*, vol. 56, no. 49, p. 6727-6730, 2020.

<sup>2</sup>do inglês Adenosine triphosphate

O aumento na atividade dessa enzima está relacionado com o surgimento de câncer, sendo a Leucemia Mieloide Crônica (LMC) o caso mais comum. Esta enzima pode ainda sofrer diferentes mutações que implicam em mudanças na estrutura e interações intermoleculares com ligantes [47]. Isto faz surgir múltiplos estados possíveis para a mesma enzima, onde inibidor e substrato tem diferentes afinidades para cada estado. Como consequência, a eficácia do tratamento é comprometida devido ao aumento na resistência para maioria dos medicamentos usados no combate ao câncer.

O Dasatinib é um inibidor seletivo da ABL1 utilizado no tratamento da LMC. Este é o único fármaco que se liga apenas ao estado ativo da enzima, tornando mais simples os cálculos computacionais. Existem algumas mutações que podem afetar o tratamento da LMC, porém apenas três delas levam a casos importantes de resistência ao dasatinib: V299L, T315I e F317L. Cada uma dessas mutações ocasionam resistências com diferentes graus. Isto se deve ao fato de que cada troca de um resíduo de aminoácido por outro resultará em novas interações intermoleculares espe-

cíficas e, conseqüentemente, em novos microestados para o complexo, ou seja, o fármaco irá buscar novos mínimos de energia. Neste cenário, é fácil de entender o mecanismo da resistência causada pela mutação T315I, pois a Thr<sup>315</sup> interage com dasatinib por meio de uma ligação de hidrogênio, a qual é perdida na forma mutante Ile<sup>315</sup> (ver Figura 1.2). Por outro lado, nas demais mutações existem apenas interações fracas tanto para o estado nativo quanto mutante.

Por esse motivo, primeiramente foi utilizado o método de Perturbação de Energia Livre para calcular a diferença entre a variação energia livre ( $\Delta\Delta G$ ) para a formação do complexo com a forma nativa e a mutante da Abl1. Esta etapa é importante para testar a precisão da Dinâmica Molecular, visto que existem muitas fontes de erros, como o próprio modelo da proteína utilizado juntamente com a escolha de parâmetros da própria DM e o

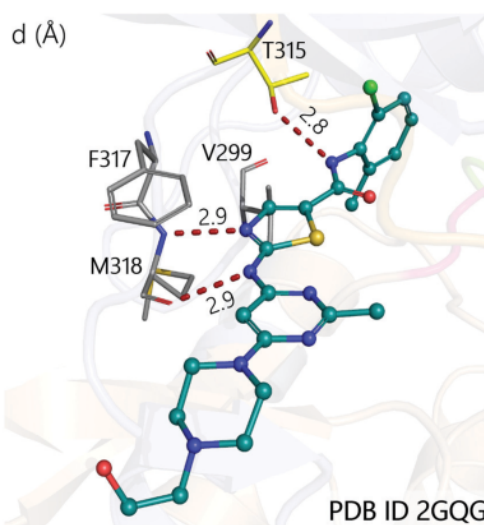


Figura 1.2: Modelo 3D mostrando as interações do dasatinib com os resíduos que iram sofrer mutação. As ligações de hidrogênio são representadas por linhas pontilhadas em vermelho com números sobre a linha informando a distância em Å entre o doador e receptor (Adaptado da referência [44]).



campo de força.

## 1.2 Métodos

### 1.2.1 Software, campo de força e parâmetros para simulações de DM

As simulações de DM foram realizadas usando gromacs [19, 21, 48] versão 2019.2. O campo de força CHARMM36 [49] foi usado para átomos de soluto, e o modelo TIP3P [50] foi empregado para moléculas de água. Uma distância de corte de 1,2 nm foi usada para calcular as interações de Van der Waals e Coulomb. As interações eletrostáticas de longo alcance foram modeladas usando o método PME. O algoritmo LINCS [51] foi usado para restringir ligações envolvendo átomos de hidrogênio nos solutos, e o algoritmo SETTLE [52] foi empregado para o mesmo objetivo com as moléculas de água. As simulações foram realizadas em temperatura constante de  $T = 300$  K, usando o termostato de reescalonamento de velocidade [33] ( $\tau_T = 0,1$  ps), e a pressão constante,  $P = 1$  bar. A pressão foi mantida constante pelo barostato Berendsen [30] para a etapa de equilíbrio e pelo Parrinello-Rahman barostat [31] para a etapa de produção (com  $\tau_P = 1$  ps).

### 1.2.2 Perturbação de Energia Livre

As simulações de DM da primeira parte deste trabalho foram feitas com o gromacs juntamente com o programa PMX [53]. A estrutura da Abl1 ligada ao dasatinib [54] foi obtida do PDB (PDB id: 2GQG). As moléculas de água cristalográficas foram retiradas para melhorar a convergência dos cálculos. Utilizando o gromacs juntamente com PMX foram construídos os estados *wt* e *mut* para cada uma das três mutações citadas anteriormente (T315I, V299L e F317L). O sistema foi simulado como uma caixa com condições periódicas de contorno, contendo a enzima (ou o complexo) em uma solução de água com íons  $K^+$  e  $Cl^-$  para simular o ambiente fisiológico e neutralizar as cargas do sistema, com uma concentração de 0.15 M. Dessa forma, para cada mutação, foram realizadas duas simulações de DM, uma apenas com a enzima e outra com o complexo enzima-inibidor.

Cada simulação seguiu a seguinte sequência: primeiramente o sistema foi submetido a uma minimização de energia para evitar erros durante a integração das equações de movimento; em seguida, uma equilibração curta (20 ps) foi realizada com intuito de reorganizar as moléculas de água e íons ao redor da proteína, mantendo os demais componentes do sistema fixos. Após estas etapas de preparação, o ensemble NPT<sup>3</sup> foi construído realizando-se uma equilibração de 5 ns para cada estado (*wt* e *mut*), sendo esta etapa replicada dez vezes para o cálculo de médias; por fim, o primeiro ns foi eliminado de cada uma trajetória temporal com o objetivo de melhorar a precisão da amostragem das configurações do sistema em equilíbrio; com os 4 ns restantes foram realizadas 100 transições na direção direta (*wt* para *mut*) e reversa (*mut* para *wt*), cada transição partindo de um ponto diferente da trajetória (veja a Figura 1.3)<sup>4</sup>.

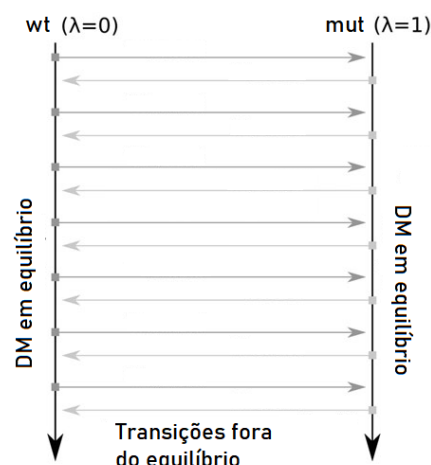


Figura 1.3: Esquema geral das simulações para cálculo de  $\Delta G$  da mutação. 100 “snapshots” foram tirados em espaçamentos de tempo iguais da trajetória de 4 ns (*wt* e *mut*). Entre cada par de “snapshots” realiza-se uma transição curta de 80 ps. (Figura adaptada de [55])

### 1.3 Discussão e Resultados

O intuito desta parte do trabalho foi analisar se as mutações reduzem a afinidade do inibidor pela enzima, sendo a diferença na variação de energia livre um parâmetro muito útil para isto. Como foi explicado anteriormente, foi calculado o  $\Delta G$  associado a cada mutação, obtendo-se um valor para a enzima livre e um para o complexo Abl1+dasatinib. A diferença entre esses dois valores resulta no  $\Delta\Delta G^b$  da Eq. 19.

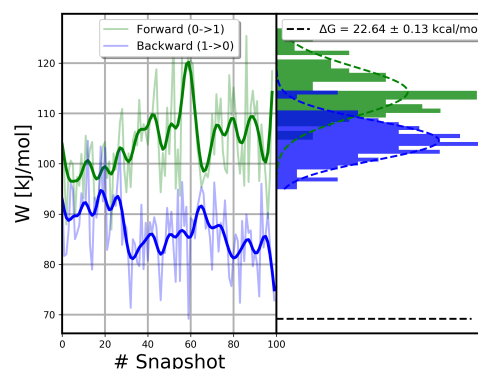


Figura 1.4: Plot gerado pelo programa PMX para uma das dez simulações da mutação T315I. Para cada par de “snapshot”, calcula-se o trabalho nas duas direções. Com isso, obtém-se duas distribuições normais de valores de trabalho, uma em cada direção. (Figura elaborada pelo autor)

<sup>3</sup>temperatura e pressão mantidos constante

<sup>4</sup>Veja a referência [44], Seção 1.3 do material de suporte para informações detalhadas sobre a simulação (comandos, pacotes e referências utilizados em cada software).

Para isso, os valores do trabalho realizado de cada transição (como exemplificado na Figura 1.3) foi calculado com a Eq. 27, utilizando funções nativas do gromacs e PMX [44]. Este programa ajusta uma curva normal para a distribuição de valores de trabalho, uma para o sentido direto e outra para o reverso da transição, como mostrado na Figura 1.4. A partir disto, avaliou-se a variação de energia livre para cada mutação utilizando a Eq. 28 e o erro associado com a Eq. 29 para as dez réplicas, obtendo-se o valor médio dessas simulações. Por fim, o valor médio de  $\Delta\Delta G$  para cada mutação foi obtido pela Eq. 19 e estão apresentados na Tabela 1.1, juntamente com as medidas experimentais.

Como todas as mutações apresentam  $\Delta\Delta G^b = \Delta G_{mut}^b - \Delta G_{wt}^b > 0$ , isto indica que o complexo Abl1-dasatinib é mais estável para a enzima nativa, ou seja, as mutações provocam resistência para formação do complexo. Dentre elas, a mutação T315I é a que mais afeta o tratamento de LMC com o dasatinib. Como é possível ver pela Tabela 1.1, os cálculos de FEP tiveram uma boa concordância com as medidas experimentais<sup>5</sup>.

Tabela 1.1: Diferença de variação de energia livre da ligação do dasatinib com a Abl1. Todos os valores estão em kcal mol<sup>-1</sup>. Os valores teóricos foram calculados por FEP (calc) utilizando o método BAR. Os valores experimentais (exp) foram calculados com a aproximação  $\Delta\Delta G_{exp}^b = RT \ln(IC_{50}^{mut}/IC_{50}^{wt})$ , onde os valores de  $IC_{50}$  são médias dos encontrados nas referências [56–59].

	T315I	V299L	F317L
$\Delta\Delta G_{calc}^b$ (kcal/mol)	$3.7 \pm 0.8$	$0.9 \pm 0.4$	$0.6 \pm 0.3$
$\Delta\Delta G_{exp}^b$ (kcal/mol)	$3.0 \pm 0.3$	$1.5 \pm 0.1$	$1.0 \pm 0.1$

Outras simulações de dinâmica molecular foram realizadas separadamente para analisar as interações intermoleculares (ver referência [44]). Ao examinar o número de ligações de hidrogênio (HBs) observado nas simulações para a proteína nativa, observou-se revela que o número médio de HB's entre o fármaco e proteínas é  $2,3 \pm 0,1$ , ou seja, mais próximo de dois do que três. Este número é um pouco menor para o mutante T315I,  $1,9 \pm 1$ . Isso indica que uma vez Thr315 sofre mutação para isoleucina, há alguma compensação para o perda de potenciais ligações de hidrogênio. Aparentemente, a ligação hidrogênio entre o nitrogênio da amina do fármaco e o *back-bone* do oxigênio da Met318 está ocupado em apenas 50% dos quadros obtido a partir das simulações da proteína

<sup>5</sup>Não foram reportados os erros das medidas experimentais para as três mutações estudadas aqui, de forma que os valores de erro da Tabela 1.1 são apenas desvios da média

nativa comparada com 78% nas simulações de T315I. Por outro lado, apenas em 33% das configurações da proteína nativa as ligações de hidrogênio entre o fármaco e Thr315 e Met318 estavam presentes simultaneamente. No geral, a diminuição do número de HBs entre a proteína e o fármaco, de 2,3 para 1,9 não foi suficiente para tornar o mutante T315I tão resistente quanto ele (mais de 70 vezes, se convertido o valor de  $\Delta\Delta G^b$  para aumento de  $IC_{50}$ ). Em vez disso, o número de contatos proteína-fármaco, que contribuem para a entalpia de ligação através de interações do tipo van der Waals (vdW) também é diminuída, levando a um aumento de 10,7 kcal mol<sup>-1</sup> na energia de interação proteína-fármaco. A energia de interação direta com o resíduo 315 é aumentada em 4,1 kcal mol<sup>-1</sup>.

Ainda nessas simulações foi avaliado a contribuição da entropia para cada variante da proteína, tanto a entropia absoluta quanto as contribuições vindas da dinâmica da proteína da pose do fármaco no sítio. Com isso foi possível inferir que o aumento de energia potencial é, em parte, compensado pelo aumento da entropia absoluta do fármaco no sítio, que tem um perfil de energia conformacional mais ampla (ver material suplementar da referência [44]), e por uma entropia conformacional mais alta para a proteína.

Tendo estabelecido que a resistência do dasatinibe para a mutação T315I da Abl1 é parcialmente devida à perda de uma ligação de hidrogênio (que contribuiu com cerca de 40% para a perda de entalpia de ligação) e em parte devido à perda de interações vdW com a proteína, foram analisados os mutantes V299L e F317L. À primeira vista, o padrão HB proteína-fármaco desses mutantes foi semelhante ao da proteína. O número de contatos proteína-fármaco também foi semelhante, um pouco menor para V299L ( $66 \pm 6$ ) do que para FLT3 ( $68 \pm 4$ ). Para V299L, onde o valor  $\Delta\Delta G^b$  calculado foi 0,9 kcal mol<sup>-1</sup>, aparentemente ocorre uma mudança geral na energia de ligação (que aumenta em 2,1 kcal mol<sup>-1</sup> com a mutação)

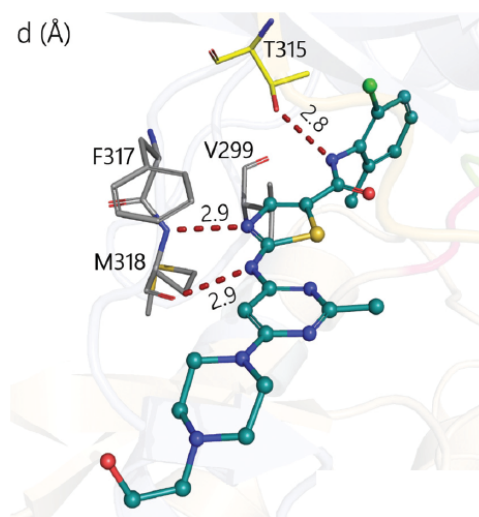


Figura 1.5: Interações intermoleculares para o dasatinibe e a estrutura nativa da Abl1. Ela forma ligações Figura adaptada de M. J. Dávila-Rodríguez et al, "Is breaking of a hydrogen bond enough to lead to drug resistance?," *Chemical Communications*, vol. 56, no. 49, 2020.

que determina a resistência aos medicamentos e é o resultado de múltiplas contribuições. Não é a mudança de valina para leucina sozinha que impulsionou esse aumento da energia de ligação. Em vez disso, a mutação levou a mudanças na dinâmica da proteína e fármaco que foi suficiente para conduzir a resistência. Uma das principais contribuições para esse aumento para o mutante V299L foi que as interações entre Thr315 e o fármaco se tornaram menos favorável (cerca de  $0,8 \text{ kcal mol}^{-1}$ ).

Para F317L as mudanças são ainda mais sutis, já que os cálculos levaram a uma energia de ligação semelhante para as duas variantes da proteína. A entropia absoluta do fármaco no sítio também foi semelhante. Como nenhuma contribuição entálpica estava foi observada, esta mutação deve ter alterado a entropia do complexo. Analisando as distribuições de ligações de hidrogênio e contatos curtos do complexo proteína-fármaco, observou-se uma distribuição mais estrita para o mutante em ambos os casos, o que sugere que a variação de energia livre calculada de  $0,6 \text{ kcal mol}^{-1}$  é devida a uma ligeira diminuição na entropia conformacional da proteína. Os cálculos com simulações independentes das daquelas usadas para o FEP indicaram que este é realmente o caso, pois a diferença na entropia observada foi de  $2 \text{ cal mol}^{-1} \text{ K}^{-1}$ , o que poderia diminuir a energia livre total em  $0,6 \text{ kcal mol}^{-1}$  a  $300 \text{ K}$ . Como isso não foi o suficiente para explicar a mudança geral no  $\Delta\Delta G^b$  após a mutação, foi também considerada a contribuição das moléculas de solvente. Análise do número de ligações de hidrogênio para proteína-água mostra uma diminuição apenas para o mutante F317L, sugerindo que as interações proteína-solvente contribuem ainda mais para a energia livre diferença entre o mutante ligado ao fármaco e a variante nativa da Abl1.

## 1.4 Conclusão

Usamos cálculos extensivos para determinar por que as mutações em Abl1 levam à resistência ao dasatinib. Simulações DM reproduziram o  $\Delta\Delta G_b$  após mutação em de  $0,7 \text{ kcal mol}^{-1}$  ou menos dos valores experimentais, e reproduziu corretamente a ordem de severidade das mutações que causam resistência aos medicamentos. As simulações atribuíram resistência ao dasatinib para o mutante T315I da Abl1 apenas parcialmente à perda do HB. A mutação da Thr315, um resíduo *gatekeeper* [60,61], alterou a dinâmica da proteína, baixou a energia de interação entre o fármaco e múltiplos resíduos e aumentou

a entropia conformacional do fármaco ligado. As mutações V299L e F317L têm efeitos entálpicos e entrópicos sutis, com o primeiro levando a resistência devido a um aumento de  $\Delta H_b$  e o último aparentemente devido a uma diminuição da entropia conformacional na estrutura da proteína inibida e a uma mudança na solvatação do complexo. Nosso estudo destaca a sutileza das interações proteína-fármaco, onde pequenas mudanças na mutação podem afetar a dinâmica de muitos resíduos. Além disso, mostra que simulações de DM fornecem um meio preciso para calcular os valores de  $\Delta\Delta G_b$  quando o fármaco se liga ao estado ativo de uma proteína (a menos que haja grandes variações na constante de ligação ( $K_M$ ) [62]). Sendo este o caso, espera-se que cálculos semelhantes, se feitos corretamente, irão subestimar o  $\Delta\Delta G_b$  da mutação para fármacos que se ligam ao estado inativo de uma enzima se a mutação levar a mudanças no estado ativo.

Por fim, mostra-se que a quebra de uma ligação de hidrogênio foi em grande parte compensada por outras interações e, portanto, não foi a única (ou principal) força motriz da resistência aos medicamentos. As mutações afetam o sistema não só de forma local, nos resíduos de aminoácidos da proteína que interagem com o fármaco, como também na dinâmica da conformação da proteína e na interface proteína-solvente. Quando se trata de design de fármacos destinadas a evitar resistência, as simulações podem ajudar a identificar tais efeitos e entender melhor o mecanismo que cada mutação exerce para gerar resistência a um determinado medicamento.

## 1.5 Informações adicionais

Este capítulo faz parte de um trabalho desenvolvido durante um estágio de três meses na Universidade de Linnaeus, em Kalmar, na Suécia (de 8 de setembro a 4 de dezembro de 2019). O trabalho completo foi publicado em 2020 (ver Anexo 1).

Mais detalhes:

- Cooperação Brasil-Suécia

Coordenador Brasil: Ignez Caracelli (Grupo BioMat)

Coordenador Suécia: Ran Friedman (Grupo CCBG)

- Projeto número 2014-4406: *Computer-aided development of potent inhibitors for wild type and mutant kinases*. Subvencionado pelo *The Swedish Research Council*.
- Estágio na Suécia → Duração de 3 meses

*Department of Chemistry and Biomedical Sciences of Linnaeus University*

Acesso ao Lunarc (<http://www.lunarc.lu.se/>)

# Capítulo 2

## ALBUMINA SÉRICA HUMANA <sup>1</sup>

### 2.1 Introdução

A Albumina sérica humana (HSA<sup>2</sup>) é a proteína mais abundante no plasma sanguíneo [63] e tem a função fisiológica de manter a viscosidade do sangue. Devido a sua estrutura e conformação com vários sítios, ela transporta várias moléculas exógenas, como medicamentos, e endógenas, como hormônios da tireoide, grupo heme e ácidos graxos. Esse comportamento faz da albumina uma proteína de extrema importância no estudo da farmacocinética de diversos medicamentos, pois a sua afinidade por um fármaco pode afetar o modo com que ele é liberado na região do alvo biológico [64].

Diferente da maioria dos casos de alosterismo, a HSA é um monômero e pode ser dividida em três domínios contendo dois subdomínios cada (Figura 2.1). Ela pode assumir duas conformações bem definidas: um estado conhecido como Neutro (N), ocorre em pH próximo de 7.0 livre de ácidos graxos; e o estado chamado de Básico (B), ocorre em pH maior que 8.0 e em pH neutro com a proteína ligada a ácidos graxos [65]. Assim, a estrutura pode mudar de estado pela alteração do pH ou por intermédio de uma dada concentração de ácidos graxos de cadeia longa no plasma. Esta molécula pode, então, ser vista como um ligante alostérico [66] no sentido de que o equilíbrio químico entre as

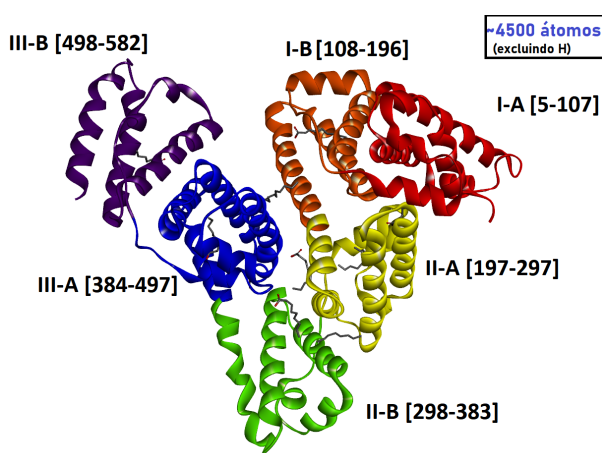


Figura 2.1: Modelo mostrando cada domínio (algarismos romanos) e subdomínios (letras A e B) da HSA. Entre colchetes tem-se a numeração dos resíduos de aminoácidos do início e do final de cada subdomínio. (PDB id: 1e7e). (Figura elaborada pelo autor)

<sup>1</sup>Será submetido

<sup>2</sup>do inglês Human Serum Albumin



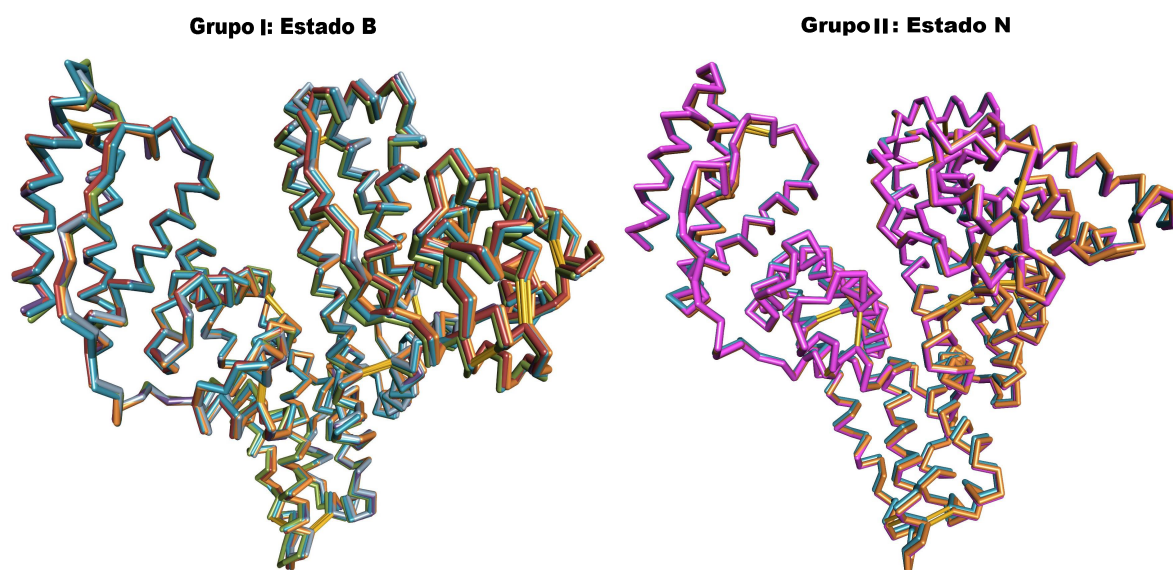
espécies Neutra e Básica da HSA pode ser deslocado no sentido da segunda forma pela presença de ácidos graxos de cadeia longa.

Se por um lado esta proteína é importante no transporte de muitos fármacos, como para aumentar a solubilidade de compostos no plasma sanguíneo, por outro ela pode afetar o tratamento de certas enfermidades para as quais os fármacos tenham afinidade excessivamente alta para esta proteína. Nestes casos, uma dose muito alta do medicamento é necessária para se obter a concentração plasmática terapêutica, visto que parte ficará ligada à albumina [67]. Devido a essas características, é importante entender o mecanismo de transição de um estado alostérico para o outro. Não está claro na literatura qual o nível de importância dos ácidos graxos para a estabilidade do estado B e qual a probabilidade de a transição ocorrer independente destes ligantes alostéricos. Com isso, este trabalho tem o intuito de estudar os mecanismos envolvidos na transição do estado N para o B da HSA por meio da Dinâmica Molecular, aliada a métodos de melhoramento de amostragem derivados da Mecânica Estatística.

## 2.2 Métodos

### 2.2.1 Estrutura

Em pesquisas no Protein Data Bank (PDB) [68] é possível encontrar estruturas da HSA que possuem entre quatro e onze ácidos graxos ligados nos sítios de cada subdomínio da proteína, bem como estruturas livres de ácido graxos ou apenas com fármacos. Com essa variedade, foi possível comparar alterações geométricas causadas por diferentes combinações e quantidade de ligantes. Dentre as estruturas encontradas no PDB, a série 1E7\* (PDB ID) foi escolhida para análise [69, 70]: 1E7H, 1E7E, 1E7F, 1E7G, 1E7I, 1E7A, 1E7B, 1E78 e 1E7C. Dentre elas, uma estrutura possui ácidos graxos e outros ligantes (1E7C), duas estruturas com ligantes e livre de ácidos graxos (1E7A e 1E7B), cinco estruturas com apenas ácidos graxos (1E7H, 1E7E, 1E7F, 1E7G e 1E7I) e uma livre de qualquer ligante (1E78). Ao sobrepor estas estruturas, utilizando o programa BIOVIA Discovery Studio [7] (Figura 2.2), dois clusters foram formados, um com as estruturas ligadas a ácidos graxos (Grupo 1, Figura 2.2a) e outro com as estruturas livres de ácidos graxos (Grupo 2, Figura 2.2b).



(a) Grupo 1: estruturas contendo ácidos graxos

(b) Grupo 2: estruturas sem ácidos graxos

Figura 2.2: Grupos 1 e 2 superpostos separadamente. (As figuras não foram feitas com a mesma perspectiva, logo servem apenas para mostrar a similaridade dentro de cada grupo). (Figuras elaboradas pelo autor)

Dentre as estruturas escolhidas, o fator comum para estabilização de cada forma alostérica foi a presença ou ausência de ácidos graxos ligados aos sítios da HSA, o que foi notado também em outros cristais disponíveis no PDB. Ao superpor as 1E7E e 1E78 (Figura 2.3), tomando o subdomínio II-B como referencial, nota-se uma movimentação relativa dos domínios I e III. Além dessas modificações globais, outras alterações locais também ocorrem, como o encurtamento da distância entre as Tyr138 e Tyr161 na ausência de ácidos graxos no sítio do grupo heme (sítio IB).

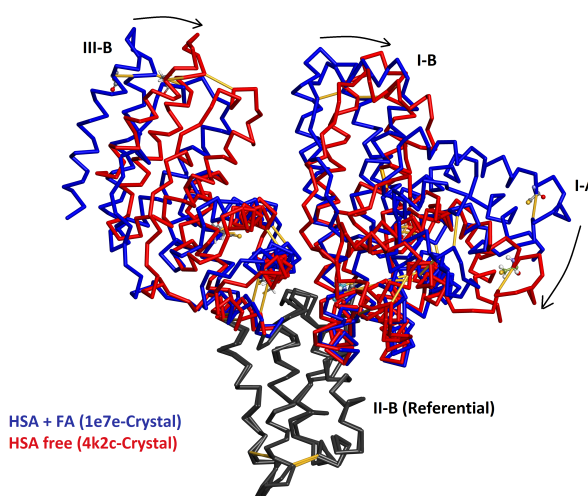


Figura 2.3: Modelo mostrando a superposição dos dois estados alostéricos da HSA. A estrutura de código pdb 1e7e tem seus sítios ocupados por ácidos graxos (estado B) enquanto a 1E78 está livre de qualquer ligante. (Fonte: Elaborada pelo autor com base nas estruturas encontradas no PDB). (Figura elaborada pelo autor)

Assim, a ligação de ácidos graxos de cadeia longa parece estabilizar uma das conformações da HSA, gerando um efeito similar ao aumento do pH. De um ponto de vista da Física Estatística, é esperado que para a HSA livre ácidos graxos o microestado da forma neutra seja mais provável de ocorrer do que o da forma básica, descrevendo um perfil de energia livre com pelo menos dois mínimos e uma barreira de energia entre eles, com o mínimo

global ocorrendo para a forma neutra.

Dessas observações, surgem dúvidas sobre como a presença de ácidos graxos afeta a conformação da HSA, visto que a transição pode ser induzida pela ligação dos ácidos ou a proteína pode existir nos dois estados e os ácidos graxos apresentam maior probabilidade de se ligarem aos sítios da albumina no estado B e estabilizar esta conformação. Do ponto de vista computacional levanta-se a questão: seria possível verificar estas possibilidades adicionando ácidos graxos aos sítios da HSA no estado N, ou ainda retirar ácidos de uma estrutura no estado B e ver a transição em um determinado tempo?

Dada a dependência temporal dessas simulações, foi escolhido o método da Dinâmica Molecular associada a métodos de melhoramento de amostragem com base nos conhecimentos da Mecânica Estatística, que serão detalhados na próxima seção.

### 2.2.2 Software, campo de força e parâmetros para simulações de DM

As simulações de DM foram realizadas usando gromacs [19, 21, 48] versão 2019.4. Simulações de metadinâmica foram realizadas usando a versão do gromacs modificado pela ferramenta PLUMED [71, 72] versão 2.5.4. O campo de força CHARMM36 [49] foi usado para átomos de soluto, e o modelo TIP3P [50] foi empregado para moléculas de água. Uma distância de corte de 1,2 nm foi usada para calcular as interações de van der Waals e Coulomb. As interações eletrostáticas de longo alcance foram modeladas usando o método PME. O algoritmo LINCS [51] foi usado para restringir ligações envolvendo átomos de hidrogênio nos solutos, e o algoritmo SETTLE [52] foi empregado para o mesmo objetivo com as moléculas de água. As simulações foram realizadas em temperatura constante de  $T = 300$  K, usando o termostato de reescalonamento de velocidade [33] ( $\tau_T = 0,1$  ps), e a pressão constante,  $P = 1$  bar. A pressão foi mantida constante pelo barostato Berendsen [30] para a etapa de equilíbrio e pelo Parrinello-Rahman barostat [31] para a etapa de produção (com  $\tau_P = 1$  ps).

### 2.2.3 Dinâmica Molecular

Como uma exploração inicial da estrutura da albumina, optou-se por verificar o que acontece ao se retirar os ácidos graxos de uma estrutura no estado B e realizar simulações

de Dinâmica Molecular de 30 ns. Simultaneamente, outra simulação foi realizada com a albumina no estado N para ver se o estado B seria visitado nesta mesma escala de tempo.

Para a primeira parte deste experimento computacional escolheu-se o PDB ID 1E7E como representante do estado B. Já na segunda parte, o PDB ID 1E78 foi utilizado como modelo do estado N. No geral, a DM seguiu os seguintes passos: o sistema foi construído como um dodecaedro, colocando a enzima no centro, a uma distância mínima de 1,2 nm de cada borda da caixa. A proteína foi solvatada, foram adicionando íons  $K^+$  e  $Cl^-$  para neutralizar as cargas e atingir a concentração de 0,15 mM. Uma minimização de energia foi realizada usando o algoritmo *steepest descent* seguido por uma simulação DM curta de 20 ps, onde restrições de posição foram impostas a todos os átomos pesados do soluto para equilibrar as moléculas de água e íons ao redor da proteína. Em seguida, o sistema foi equilibrado por 30 ns depois de remover as restrições. Esta etapa foi repetida vinte vezes para cada estado.

Como o cristal foi tirado do equilíbrio ao se remover os ácidos graxos, optou-se por pular a tradicional sequência de simulação da DM, onde a etapa de produção viria após uma equilibrção mais curta (de 5 a 10 ns) pois notou-se que as mudanças de conformação começavam a ocorrer nos primeiros 5 ns de simulação. Assim a etapa de equilibrção sem restrição de posições foi estendida e replicada para melhorar a amostragem.

Devido à limitação do passo de 2 fs, necessário para resolver as vibrações dos átomos de hidrogênio [73], alguns eventos não acontecem na escala de tempo usual da Dinâmica Molecular pura, na faixa de nano para micro segundos [74]. Assim, para observar tais eventos raros, como a transição alostérica de algumas proteínas, é necessário empregar técnicas de melhoramento de amostragem, o que será tratado na próxima seção.

#### **2.2.4 *Essential dynamics sampling (EDS)***

A preparação de estruturas para o EDS seguiu a mesma abordagem da DM da seção anterior, com cada estado passando pelas etapas de minimização de energia, solvatados e equilibrados da mesma maneira, optando aqui por apenas 10 ns e equilibrção sem restrição de posições. Em seguida, 50 ns de corrida na etapa de produção foram realizadas para cada estado. Porém, como a estrutura representante do estado B teve seus ligantes removidos, a equilibrção e produção não representavam mais a conformação usual, o que

será melhor discutido na seção 2.3.1. Por esse motivo, apenas a trajetória da simulação DM do estado N foi usada para análise de cluster com base nos átomos de  $C\alpha$  da proteína, usando o método Gromos [75] com **gmx cluster**. Foi utilizado um corte de 0,2 nm no valor de gromacs dos  $C\alpha$ 's. Posteriormente, uma janela de 10 ns da trajetória de 50 ns, na qual a proteína permaneceu no mesmo cluster, foi usada para realizar uma análise de covariância, usando o comando **gmx covar**. Um quadro extraído deste intervalo foi usado como origem e como quadro final foi utilizado as coordenadas finais da etapa de minimização de energia do estado B para construir o arquivo de entrada para a simulação de EDS, usando o comando **gmx make edi**. O algoritmo EDS [76] foi usado como implementado no gromacs para executar simulações de 1 ns. As coordenadas e energias foram salvas a cada 1 ps.

## 2.2.5 Metadinâmica

O método da Metadinâmica foi utilizado para simular a transição da HSA do estado neutro para o básico, primeiramente utilizando como variáveis coletivas a distância entre os centros de massa dos subdomínios IA e IIIB e a distância entre os centros geométricos dos anéis das Tyr138 e Tyr161. Posteriormente, novas variáveis coletivas foram empregadas para verificar qual melhor representava o sistema durante a transição. As que se mostraram mais efetiva foram variáveis especiais, chamadas de Variáveis Coletivas de Caminho (ou Path Collective Variables, no inglês) [77–79],  $s(R)$  e  $z(R)$ . Elas descrevem o progresso da simulação ao longo de um caminho pre determinado, sendo  $s(R)$  a variável relativa à posição de cada quadro do caminho-referência e  $z(R)$  a distância em relação a um determinado quadro desse caminho. Assim, a primeira irá indicar em que ponto da transição a simulação se encontra, e a segunda dirá quão perto do caminho de mínima energia o quadro de referência se encontra. Tais variáveis coletivas são definidas como:

$$s(R) = \frac{\sum_{i=1}^{N-1} i e^{-\lambda \|X - X_i\|^2}}{\sum_{i=1}^{N-1} e^{-\lambda \|X - X_i\|^2}} \quad (2.1)$$

$$z(R) = -\frac{1}{\lambda} \ln \sum_{i=1}^{N-1} e^{-\lambda \|X - X_i\|^2} \quad (2.2)$$

onde o número de quadros do caminho de referência é dado por  $N$  e o termo  $\|X - X_i\|^2$  é o deslocamento quadrático médio (MSD) entre uma configuração geométrica instantânea  $X$  e a de uma quadro do caminho de referência  $X_i$ . O termo  $\lambda$  é proporcional ao MSD entre quadros consecutivos do caminho e pode ser estimado pela seguinte equação [77–79]:

$$\lambda = \frac{2.3(N - 1)}{\sum_{i=1}^{N-1} \|X_{i+1} - X_i\|^2} \quad (2.3)$$

Para construir esse caminho de referência para a transição da estrutura 1E78 para a 1E7E foram utilizadas ferramentas nativas do gromacs, como o *Essential Dynamics Sampling*. Tal trajetória foi montada selecionando apenas os quadros com *Root Mean Square Deviation* (RMSD) em um intervalo de 0,045 a 0,060 nm em relação aos seus primeiros vizinhos. Uma segunda condição foi imposta para garantir que cada quadro seguinte se aproxime do quadro final e se afaste do inicial, gerando, assim, uma curva decrescente monotônica de  $\text{rmsd} \times \text{número de quadros}$ . Este procedimento gerou uma trajetória composta por 45 quadros. O  $\lambda$  calculado para esse caminho foi de  $805 \text{ nm}^2$ . Em seguida, uma corrida de 1 ns de wT-metaD sem potencial viés foi realizada para estimar o desvios padrão das variáveis  $s(R)$  e  $z(R)$ . O quadro da estrutura de final foi adicionado como ultimo da trajetória 2.4. Os desvios calculados foram usados como estimativa das larguras das gaussianas na direção  $s(R)$  e  $z(R)$ , obtendo os valores  $1.0 \text{ nm}^2$  e  $0.005 \text{ nm}^2$ , respectivamente.

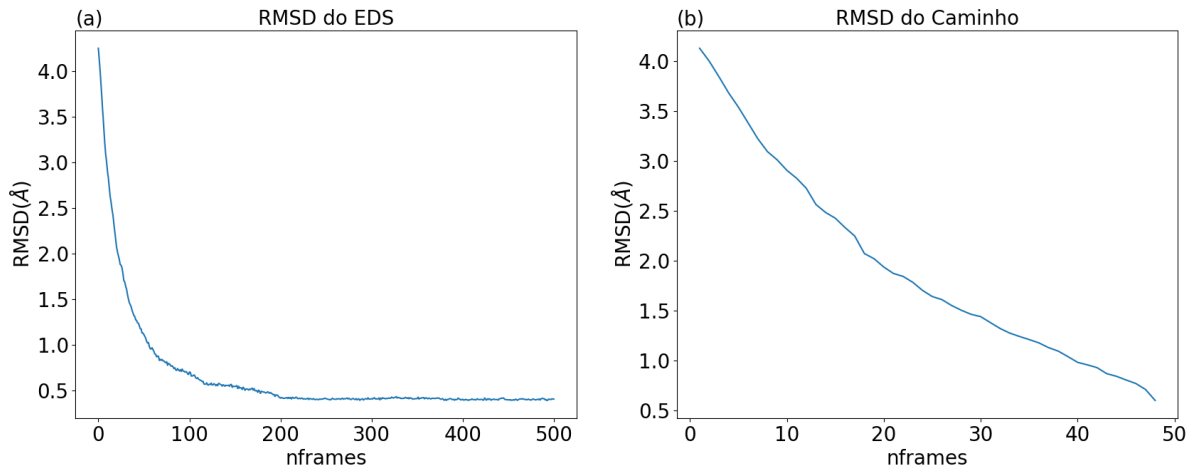


Figura 2.4: rmsd relativo ao ultimo quadro da trajetória de (a) EDS pura e (b) quadros selecionados com rmsd entre quadros aproximadamente constante e monotonicamente decrescente. (Figura elaborada pelo autor)

As simulações wT-metaD foram executadas com os parâmetros calculados acima. Gaussianas foram depositadas a cada 500 passos com altura inicial ajustada em 1,0 kJ/mol; a altura foi reduzida com um *bias factor* de 10. Para limitar a exploração do espaço de fase e evitar que a simulação gaste tempo em estados com energias muito altas na direção  $z(R)$ , uma força restritiva de  $10^4 \text{ kJ mol}^{-1} \text{ \AA}^{-2}$  foi aplicada quando  $z(R)$  excedeu  $0,4 \text{ \AA}^2$ . Esta etapa foi repetida dez vezes. Perto de 150 ns algumas das corridas convergiram para uma superfície de energia livre (FES) razoável, com mínimos próximos dos estados inicial e final.

## 2.3 Resultados e Discussão

### 2.3.1 Dinâmica Molecular

Um total de vinte simulações, de 30 ns cada, foram realizadas partindo da estrutura 1E7E livre de ácidos graxos. Ao final de cada uma delas, foi medida a distância entre o centro de massa dos subdomínios IA e IIIB ( $d_{IA-IIIB}$ ), os ângulos formados pelos centros de massa dos subdomínios IB-III A-III B ( $\theta_{III B}$ ) e IIB-II A-IA ( $\theta_{IA}$ ) (Figura 2.5) e a distância entre o centrômero dos anéis das Tyr138 e Tyr161 ( $d_{Y138-Y161}$ ) (Figura 2.6). Os valores de referência para cada um dos parâmetros estão representados em linhas pontilhadas na Figura 2.7.

Dentre as vinte simulações foi possível identificar que: oito atingiram o valor de referência do estado N para o parâmetro  $d_{IA-IIIB}$ ; cinco terminaram perto do  $\theta_{III B}$  da 1E78; nenhuma das réplicas mostraram mudanças significativas que indicassem transição de estado alostérico para  $\theta_{IA}$ ; em dezenove das vinte simulações a distância entre as tirosinas  $d_{Y138-Y161}$  foi de 8 Å para 5 Å, enquanto em uma delas esse valor permaneceu próximo de 8 Å.

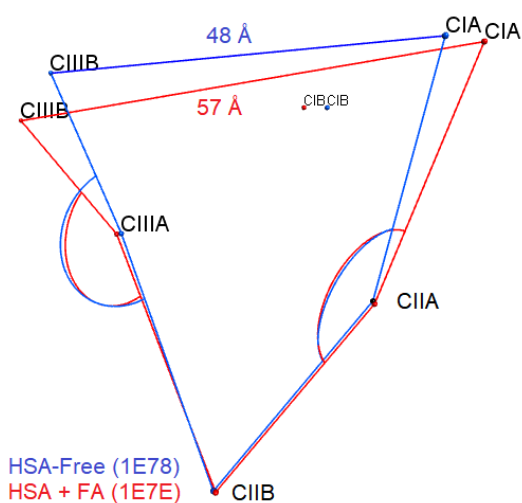


Figura 2.5: Superposição entre os estados básico e neutro da albumina, evidenciando os centros de massa dos subdomínios. A distância entre os centros de massa de interesse são mostradas nas suas respectivas cores (vermelho para o estado neutro e azul para o neutro). (Figura elaborada pelo autor)

Dos resultados apresentados acima é possível perceber que ao se remover os ácidos graxos dos sítios da HSA, a estrutura sofre uma rápida reconfiguração em intervalos relativamente curtos (entre 5 e 10 ns). Para o caso da distância entre as Tyr138 e Tyr161 a transição é ainda mais rápida, ocorrendo dentro do primeiro nano segundo. Por outro lado, neste tempo de 30 ns de dinâmica molecular, não foi possível observar a transição completa do estado B para o estado N. Isso chama a atenção ao fato do subdomínio IA apresentar uma movimentação mais restrita se comparado ao subdomínio IIIB. Como é possível ver na Figura 2.3, para se atingir a conformação neutra, o domínio I deve sofrer uma rotação sobre o II, ao passo que o subdomínio IIIB tem a liberdade de translação em direção à fenda formada entre os domínios I e III sem translação significativa do subdomínio IIIA.

O comportamento observado nessas simulações traz informações importantes sobre a física dessa transição alostérica. O fato de ocorrer mudanças significativas de conformação em um curto intervalo de dinâmica indica que as barreiras energéticas são relativamente baixas, pelo menos no sentido  $B \rightarrow N$ . Neste contexto, os parâmetros medidos revelam pelo menos três constates de tempo diferentes: a aproximação das tirosinas ocorre em uma escala de tempo até dez vezes menor do que a movimentação do subdomínio IIIB, sendo ainda menor se comparada ao subdomínio IA, para o qual não foi possível de observar transição  $B \rightarrow N$  no tempo de dinâmica utilizado. Assim, espera-se valores crescentes de barreira de energia livre, seguindo a relação  $\Delta G^\ddagger(d_{Y138-Y161}) < \Delta G^\ddagger(\theta_{IIIB}) < \Delta G^\ddagger(\theta_{IA})$ .

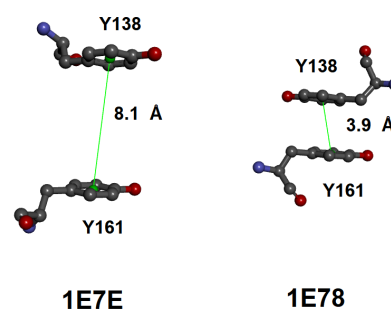


Figura 2.6: Distância entre os resíduos Tyr138 e Tyr161 para o cada estado a HSA. (Figura elaborada pelo autor)



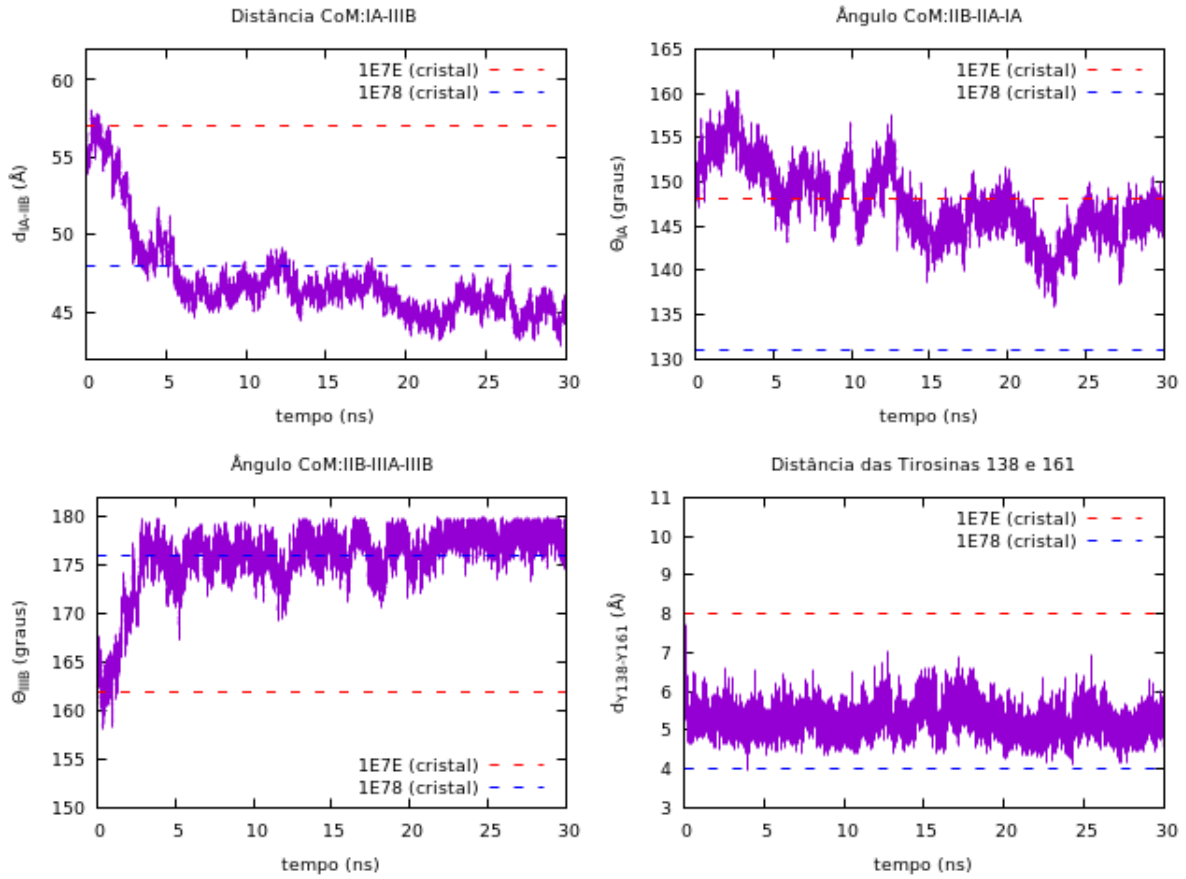


Figura 2.7: Figura mostrando a variação dos parâmetros no tempo da simulação que mais se aproximou da transição completa do estado B para o N. É possível notar que  $d_{IA-IIB}$ ,  $\theta_{IIIB}$  e  $d_{Y138-Y161}$  transitaram rapidamente para os valores esperados do estado N, porém  $\theta_{IA}$  se manteve próximo do valor inicial. (Figura elaborada pelo autor)

Essas simulações serviram como uma exploração inicial do sistema para gerar uma intuição sobre a física da transição alostérica da HSA. Com essa noção a respeito das barreiras de energia livre desse sistema foi possível otimizar a próxima etapa, na qual se utilizou a técnica de Metadinâmica.

### 2.3.2 Metadinâmica

Ao realizar uma série de simulações sem viés foi possível identificar variáveis coletivas de rápida e lenta difusão pelo espaço de fase. Isso é importante pois ao ignorar variáveis coletivas lentas é possível que o sistema nunca convirja numa simulação de metadinâmica [40]. A recomendação é combinar pelo menos uma variável lenta com uma rápida na Eq. 38 para se obter melhor desempenho. Com base nessas considerações, foram escolhidas as variáveis coletivas  $d_{IA-IIB}$  e  $d_{Y138-Y161}$  e realizada uma simulação de metadinâmica com

30 ns de duração, partindo do estado B (conformação cristalográfica da 1E7E, retirando-se os ácidos graxos). Os resultados são mostrados na Figura 2.8.

Nota-se que a distância entre os centros de massa IA e IIIB oscilou próximo do esperado, porém não percorreu novamente a transição no sentido contrário. Já a distância entre as tirosinas cruzou a barreira de energia três vezes, sendo duas no sentido direto da transição. Isso se deve ao fato de esta segunda variável envolver barreiras energéticas mais baixas e de valores muito próximos nos dois sentidos. Assim, à medida que as gaussianas do potencial de viés preenchem o primeiro mínimo, a barreira energética é ultrapassada e o mesmo acontece no segundo mínimo, até que o sistema possa difundir pelo espaço de fase em um potencial aproximadamente plano.

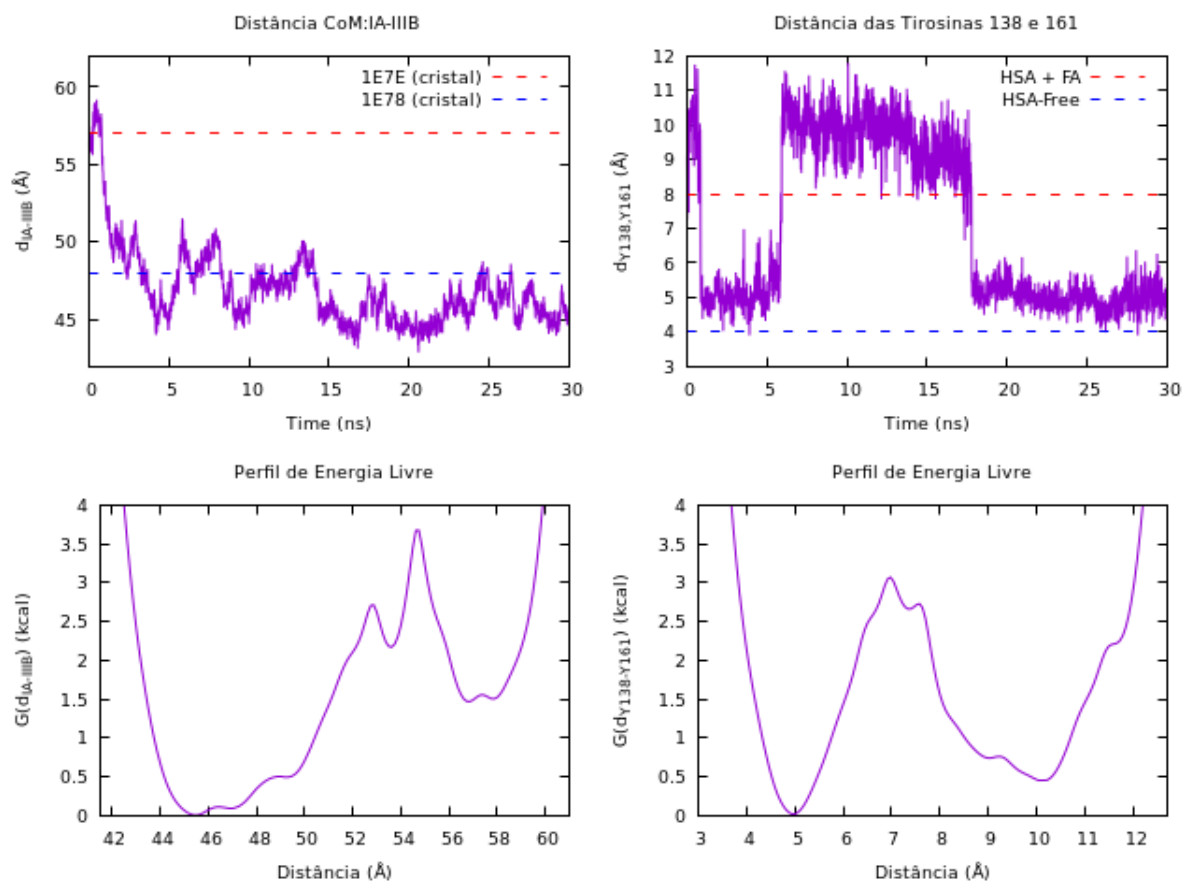


Figura 2.8: Figura mostrando a variação dos parâmetros no tempo da simulação de metadinâmica completa do estado B para o N. É possível notar que  $d_{IA-IIIB}$ , e  $d_{Y138-Y161}$  transitaram para os valores esperados do estado N, tendo essa segunda uma variação entre dois estados um pouco mais afastados do observado no cristal. (Figura elaborada pelo autor)

Assim como se esperava, a barreira de energia livre na direção  $B \rightarrow N$  é relativamente mais baixa se comparada com a transição  $N \rightarrow B$ , produzindo um perfil de energia livre na forma de um duplo poço assimétrico. Vale lembrar que isso ocorre pois os ácidos graxos

foram retirados antes da simulação, comportamento que é compatível com os dados do PDB, já que as estruturas livre de ácidos graxos encontram-se na conformação do estado N. Porém, este resultado aponta que a HSA pode transitar naturalmente entre os dois estados alostéricos e possivelmente os ácidos graxos irão selecionar a estrutura de melhor afinidade, neste caso o estado B [64, 65, 67, 80]. Ao se ligar a proteína, os ácidos graxos modificariam o perfil de energia livre, produzindo um poço duplo assimétrico de potencial com mínimo global em torno da conformação do estado B.

Outras combinações de variáveis coletivas geométricas foram testadas, como os ângulos formados pela combinação de três centros de massa de cada subdomínio (ou e diedros ao se combinar quatro centros de massa), mas não foi possível obter um bom resultado. Devido a essa dificuldade, optou-se pelo uso das variáveis coletivas de caminho para simular a transição no sentido  $N \rightarrow B$ . A vantagem dessas novas variáveis sobre as geométricas está na ideia de não ser necessário um conhecimento prévio do sistema. Apenas tendo os estados final e inicial é possível montar uma sequência de configurações para as quais o rmsd decresça monotonicamente ao se afastar da configuração inicial e tenda a zero próximo do estado final. Caso esse caminho referencial da transição esteja próximo o suficiente do caminho de mínima energia os valores esperados da variável  $z(R)$  devem ser pequenos. Assim, as variáveis  $s(R)$  e  $z(R)$  formarão um tubo de busca em torno do caminho de referência, preenchendo os mínimos e atravessando as barreiras de mais baixa energia.

A Figura 2.9 mostra o Perfil de Energia Livre (FEP em inglês) à esquerda onde a variável  $s(R)$  vai de zero (quadro do estado N) a 1 (quadro referente ao estado B). A variável  $z(R)$ , medida em  $\text{Å}^2$ , se afasta pouco do zero, mostrando que o caminho inicial não estava distante do Caminho de Mínima Energia Livre (MFEP em inglês). É possível observar a existência de mínimos próximos dos estados finais e iniciais, porém o primeiro mínimo abaixo de 1 kcal/mol acontece próximo de  $s(R) = 0.1$ , ao passo que o último mínimo local ocorre próximo de  $s(R) = 0.9$ . Isso indica que para os parâmetros utilizados na dinâmica as conformações cristalográficas estão ligeiramente afastados do equilíbrio.

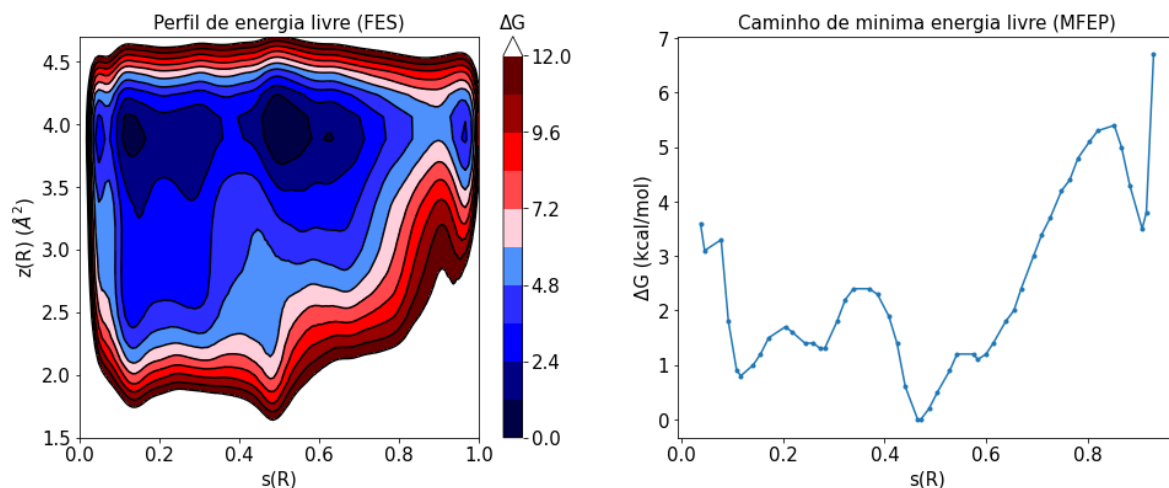


Figura 2.9: Perfil de energia livre (FES) (esquerda) e Caminho de mínima energia livre (direita). Em ambos os gráficos é possível notar a forma assimétrica da energia livre na transição alostérica, indicando maior probabilidade de encontrar a estrutura mais próxima do estado N. (Figura elaborada pelo autor)

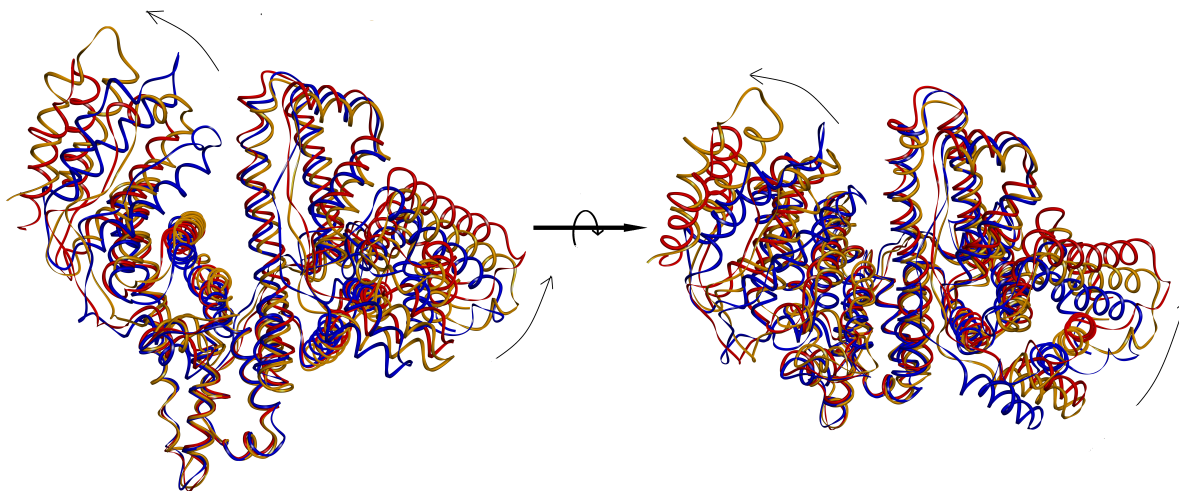


Figura 2.10: Comparação entre os três principais mínimos de energia, primeiro mínimo em azul, segundo em castanho e terceiro em vermelho. Pode-se notar que o subdomínio IIIB chega mais próximo da sua conformação final logo na transição do primeiro pro segundo mínimo, enquanto o subdomínio IIIA conclui encontra-se ainda em uma conformação intermediária. (Figura elaborada pelo autor)

Já no gráfico da direita na Figura 2.9, o MFEP mostra barreiras relativamente baixas até  $s(R) = 0.6$ . Depois deste ponto surge uma barreira maior, porém de menor valor no sentido oposto. Dividindo a trajetória em 50 quadros, ao superpor as estruturas dos quadros 6 (azul), 24 (castanho) e 48 (vermelho) (Figura 2.10) referente a estes mínimos, é possível notar que do primeiro mínimo para o segundo existe uma prevalência da movimentação do subdomínio IIIB, enquanto o subdomínio IIIA conclui a sua transição apenas no último mínimo. Este comportamento está de acordo com o observado nas simulações de dinâmica molecular pura.

## 2.4 Conclusão

Neste trabalho foi possível estudar o comportamento da albumina sérica humana por meio das técnicas de dinâmica molecular pura e associada ao método de melhoramento de amostragem da metadinâmica. No primeiro caso, ao obter uma fração considerável de estruturas que transitaram do estado básico para o neutro, mesmo que de forma incompleta, em um tempo de dinâmica relativamente curto traz a informações sobre as barreiras energética envolvidas na transição no sentido da  $B \rightarrow N$ . Com esse resultado foi possível escolher as variáveis coletivas e outros parâmetros para a realização da metadinâmica. Esta ferramenta possibilitou calcular a variação de energia livre durante a transição com uso de variáveis coletivas geométricas e relativas a um caminho referencial da transição e ainda obter uma trajetória com possíveis estados intermediários.

Em partes, o que foi observado com dinâmica molecular pura foi confirmado pela metadinâmica. A facilidade com que a albumina passa do estado B para o N se deve a baixa barreira energética neste sentido da transição. Já no sentido oposto é observado barreiras maiores, especialmente no caso em que se utiliza variáveis coletivas de caminho. Em todas as situações foi possível observar que o subdomínio IIIB se movimenta em uma escala de tempo menor do que o subdomínio IA. Assim, a transição alostérica da albumina sérica humana ocorre em pelo menos duas etapas, com o subdomínio IIIB chegando à sua conformação final antes do subdomínio IA.

## 2.5 Informações Adicionais

Este capítulo faz parte do seguinte projeto:

- Projeto 2019 - 2022:

Chamada MCTIC/CNPq N<sup>o</sup> 28/2018 - Universal/Faixa B -

Estudos *in silico* do modo de Ligação e seletividade de ligantes complexados com albumina sérica HSA.

- Situação: Em andamento; Natureza: Pesquisa.

Alunos envolvidos: Graduação: (1) / Doutorado: (3)

Integrantes: Iñez Caracelli - Coordenador / Zukerman-Schpector, Julio - Integrante / Thales Souza Freire - Integrante.

- Aquisição da estação virtual: CNPq-457255/2014-5

Compreensão da Seletividade e Mecanismo Molecular da Inibição de COX-2

Parte das simulações foram realizadas também no Lunarc, em parceria com Departamento de Química e Ciências Biomédicas da Universidade de Linnaeus.

# Capítulo 3

## FMS-LIKE TIROSINA QUINASE 3 <sup>1</sup>

### 3.1 Introdução

As proteínas quinases fosforilam proteínas e, assim, mediam transdução de sinal. A hiperatividade de tais quinases pode levar a diferentes condições patológicas, tornando-os excelentes alvos para o desenvolvimento de medicamentos [82], como mencionado no Capítulo 1. Especificamente, os receptores de tirosina quinases (RTKs) são alvos promissores para tratamento de câncer [83]. As proteínas quinases adotam principalmente um dos dois estados, ativo e inativo, com características estruturais distintas [84]. A maioria dos inibidores de quinase (KIs) tem como alvo um desses estados [85]. Os inibidores do tipo 1 inibem os estados ativos e também pode se ligar ao inativo, embora com menor afinidade, enquanto os inibidores do tipo 2 ligam-se ao estado inativo. A resistência aos KIs é um desafio significativo, já que é mediado por mutações no alvo molecular do fármaco e por outros mecanismos [86, 87]. A compreensão dos mecanismos precisos que medeiam a resistência é importante para o desenvolvimento de melhores terapias [88]. Além disso, elucidar o papel de aminoácidos específicos e interações intermoleculares é fundamental no caso de mutações no alvo do fármaco [47, 89], pois pode permitir o desenvolvimento de medicamentos melhores. Simulações computacionais em detalhe atômico são de extrema importância para este objetivo, uma vez que elas muitas vezes podem lançar luz sobre detalhes que não podem ser facilmente obtidos diretamente de experimentos [90, 91]. Simulações usando técnicas avançadas de amostragem, como a perturbação de energia livre (FEP) [92], especialmente *Essential Dynamics Sampling* [76] (EDS) e metadinâmica [71, 72] são particularmente úteis para tais estudos.

Gilteritinib é um inibidor eficaz e seletivo da FLT3 RTK e é usado na terapia de

---

<sup>1</sup>Trabalho Publicado (Anexo 2): [81] T. S. Freire, I. Caracelli, J. Zukerman-Schpector, and R. Friedman, “Resistance to a tyrosine kinase inhibitor mediated by changes to the conformation space of the kinase,” *Physical Chemistry Chemical Physics*, vol. 25, no. 8, p. 6175–6183, 2023.

leucemia mieloide aguda (LMA) desde 2018. É um inibidor do tipo I, e, portanto, mutações que ativam a proteína (D835F/V/Y e Y842C/H) não o afetam. Infelizmente, a maioria, se não todos pacientes ainda dependem da terapia com gilteritinib [93]. Muitos desses pacientes expressam uma forma mutante, FLT3-ITD/F691L (*Internal tandem duplication* (ITD) é uma mutação que não modifica o sítio ativo [42,94,95]). Recentemente, foi relatado que outra mutação, D698N leva à resistência marcada *in vitro* [96]; esta mutação, no entanto, não foi relatada em pacientes. Por que essas mutações levam à resistência não está claro. Ao contrário de outros inibidores do tipo I (por exemplo, dasatinib [44]) não há indicação de que os resíduos mutantes interagem diretamente com fármaco (Figura 3.1). Os resíduos que interagem com o fármaco incluem Glu<sup>692</sup> e Cys<sup>694</sup> que são ligados por hidrogênio ao gilteritinibe através de seus esqueletos de carbono e a Leu<sup>616</sup> que se liga ao fármaco através de interações vdW (dentro de  $< 0,4$  nm dos átomos de proteína), com outros resíduos mais longe.

O objetivo deste estudo foi entender a razão por trás da resistência causada por essas mutações. Como não há estrutura experimental da FLT3 no estado ativo ligado ao gilteritinibe, nós primeiro modelamos o fármaco ligado ao estado ativo (Fig. 2). FEP foi usado posteriormente para calcular a mudança de energia de Gibbs mutando para F691L e D698N. Para obter uma imagem mais completa na ativação da variante nativa e resistente, a ativação processo foi posteriormente estudado por uma combinação de EDS e metadinâmica bem-temperada (wT-metaD [43]). Os resultados dessas simulações foram ainda usadas para racionalizar por que as mutações levam resistência a o fármacos contra gilteritinibe.

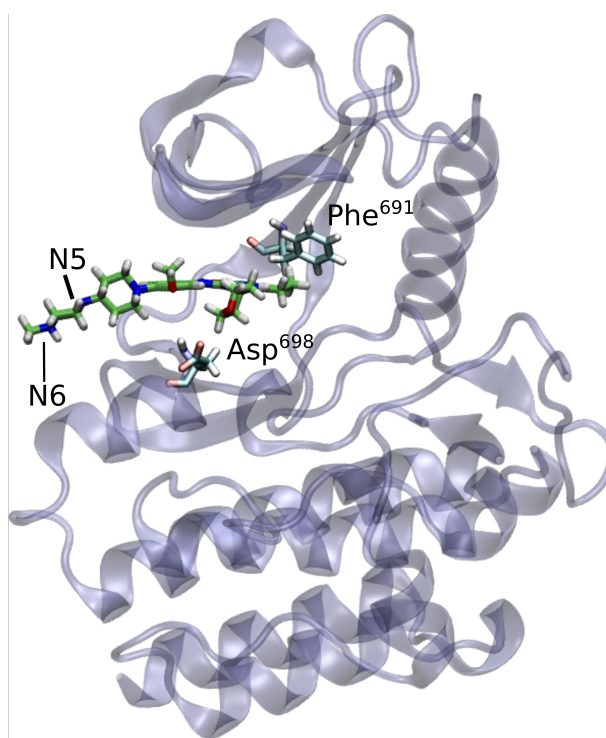


Figura 3.1: Estrutura do FLT3 no estado ativo com gilteritinibe. Resíduos Phe<sup>691</sup> e Asp<sup>698</sup>, que estão mutados em variantes resistentes são indicados. Os dois nitrogênios da porção piperazina, que são denominados aqui N5 e N6, também são indicados. N5 está mais próximo do centro de massa da proteína e menos exposto ao solvente, portanto, é menos provável que seja protonado. (Fonte: Elaborada pelo autor)



## 3.2 Métodos

### 3.2.1 Software, campo de força e parâmetros para simulações de DM

As simulações de DM foram realizadas usando gromacs [19, 21, 48] versão 2019.4. Simulações de metadinâmica foram realizados usando a versão do gromacs que foi corrigida com o PLUMED [71, 72] versão 2.5.4. O campo de força CHARMM36 [49] foi usado para átomos de soluto, e o modelo TIP3P [50] foi empregado para moléculas de água. Uma distância de corte de 1,2 nm foi usada para calcular as interações de van der Waals e Coulomb. As interações eletrostáticas de longo alcance foram modeladas usando o método PME. O algoritmo LINCS [51] foi usado para restringir ligações envolvendo átomos de hidrogênio nos solutos, e o algoritmo SETTLE [52] foi empregado para o mesmo objetivo com as moléculas de água. As simulações foram realizadas em temperatura constante de  $T = 300$  K, usando o termostato de reescalonamento de velocidade [33] ( $\tau_T = 0,1$  ps), e a pressão constante,  $P = 1$  bar. A pressão foi mantida constante pelo barostato Berendsen [30] para a etapa de equilíbrio e pelo Parrinello-Rahman barostat [31] para a etapa de produção (com  $\tau_P = 1$  ps).

### 3.2.2 Modelos

O estado inativo foi modelado com base na estrutura cristalina PDB ID 6JQR [97]. Para o estado ativo, foi usado um modelo produzido por homologia, relatado em [89]. Ambas as estruturas foram completadas e corrigidas pela ferramenta online WHAT IF, usando a sequência do estado ativo (que era mais curto) como referência, terminando com o mesmo número de resíduos de aminoácidos e átomos. As enzimas mutantes foram obtidas a partir das nativas usando o conformero gerado pela ferramenta de edição do programa UCSF Chimera [98] (v1.13.1) para as rodadas de metadinâmica e pela ferramenta online pmx [53, 99] para as rodadas de perturbação de energia livre .

### 3.2.3 Estrutura do estado ativo ligada ao gilteritinib

O modelo ativo e a estrutura inativa de raios-X (ligada ao gilteritinib) foram sobrepostas por sequência de aminoácidos pelo programa Discovery Studio, v.2021 (BIOVIA, Dassault Systèmes). Em seguida, o sítio ativo de ambas as estruturas foi selecionado e a função de superposição foi aplicada novamente para melhorar a sobreposição de resíduos nesta região. Depois desta etapa, a estrutura da enzima inativa foi deletada, mantendo-se apenas o modelo ativo e o ligante. Este arquivo foi salvo como uma extensão mol2 para ser usado em uma corrida de Docking Molecular. A modelagem adicional foi realizada por docking usando o software GOLD, v.2020.1 (CCDC, Cambridge, Reino Unido <sup>2</sup>). Os átomos de hidrogênio foram adicionados à proteína e o ligante. O sítio de ligação para a corrida de docking foi posteriormente definido como a região ocupada pelo ligante mais 0,8 nm de distância de cada átomo, produzindo uma cavidade em forma de elipsoide. A enzima foi mantida rígida, permitindo flexibilidade apenas para o ligante, com uso das opções padrão do programa para esta etapa. A seguir, a função de *fitness* GOLD score foi usada para executar o algoritmo genético. As saídas do docking mostrou um cluster de estruturas próximo à pose inicial do ligante. A melhor pose deste cluster foi escolhida para ser usada na DM.

### 3.2.4 Perturbação de energia livre

A perturbação de energia livre foi usada para calcular os valores de  $\Delta\Delta G$  para gilteritinib após a mutação, ou seja,  $\Delta\Delta G_{S \rightarrow R}$  [86], onde *S* é a variante sensível e *R* é a variante resistente. O pacote pmx [53, 99] foi usado para preparação das topologias híbridas (nativa e mutante no mesmo arquivo). As mutações (F691L e D698N) foram geradas com o módulo **pmx mutate**. Topologias híbridas, contendo informações sobre o estado nativo e o mutante da enzima, foram construídas a partir dos arquivos pdb das formas nativas, com uso do módulo **pmx gentop**. O sistema foi construído como caixa dodecaedro, colocando a enzima no centro, a uma distância mínima de 1,2 nm de cada borda da caixa. O complexo proteína-fármaco foi solvatado, e foram adicionados íons  $K^+$  e  $Cl^-$  para neutralizar as cargas e atingir a concentração de 0,15 mM. Para cada mutação, uma única minimização de energia foi realizada usando o algoritmo *steepest*

---

<sup>2</sup><https://www.ccdc.cam.ac.uk/>

*descent* seguido por uma simulação DM curta de 20 ps, onde restrições de posição foram impostas a todos os átomos pesados do soluto para equilibrar as moléculas de água e íons ao redor da proteína. Em seguida, o sistema foi equilibrado por 10 ns depois de remover as restrições. A etapa de equilíbrio foi repetida dez vezes para cada estado da enzima para garantir uma melhor precisão por meio de amostragem. Cem quadros (salvos em arquivos no formato gro) foram extraídos das trajetórias de equilíbrio em intervalos de tempo regulares. De cada um desses quadros, foi realizada uma única simulação curta de DM de 100 ps, partindo do wt (transição direta) e do mutante (transição reversa). Isso resultou em 100 transições em cada direção. O barostato Parrinello-Rahman foi usado para essas simulações, e as velocidades foram geradas em cada corrida de transição. O comando **gmx mdrun -dhdl** foi usado para calcular a mudança no Hamiltoniano para cada etapa da simulação, onde a integral ao longo de todo o caminho é igual ao trabalho feito naquela corrida específica. Finalmente, a energia livre de Gibbs média para cada mutação foi obtida com a equação da taxa de aceitação Bennett, usando o módulo **pmx analyse** sobre todos os valores, fazendo a média das dez simulações.

### 3.2.5 Essential dynamics sampling

A preparação de (apo) estruturas para o EDS seguiu a mesma abordagem que para o FEP, ou seja, os estados nativo e mutantes foram minimizados em energia, solvatados e equilibrados da mesma maneira. Depois disso, 50 ns de corrida na etapa de produção foram realizadas para cada estado (ativo e inativo). Usando as trajetórias de cada simulação DM, análises de cluster foram realizadas com base nos átomos de  $C\alpha$  da proteína usando o método Gromos [75] com **gmx cluster**. Foi utilizado um corte de 0,15 nm no valor de rmsd dos  $C\alpha$ 's, o que produziu resultados semelhantes à escolha de um corte de 0,20 nm para átomos pesados. Todos os três estados inativos mostraram um número menor de clusters em comparação com cada respectivo estado ativo. Posteriormente, uma janela de 10 ns da trajetória de 50 ns, na qual a proteína permaneceu no mesmo cluster, foi usada para realizar uma análise de covariância, usando o comando **gmx covar**. Um quadro extraído de cada um desses intervalos foi usado como origem (trajetória do estado inativo) e um como quadro final (trajetória de estado ativo) para construir o arquivo de entrada para a simulação de EDS usando o comando **gmx make edi**. O algoritmo EDS [76] foi usado como implementado no gromacs para executar simulações de 1 ns para das enzimas

nativas e mutantes. As coordenadas e energias foram salvas a cada 1 ps.

### 3.2.6 Metadinâmica

Simulações de metadinâmica bem temperada (wT-metaD) foram realizadas como em [100]. Resumidamente, o rmsd dos átomos  $C\alpha$  da trajetória EDS em relação à estrutura alvo foi calculado primeiro. Usando o perfil de rmsd, os quadros foram selecionados para atender aos critérios de ter um rmsd com decréscimo monotônico e valores aproximadamente constantes entre dois quadros consecutivos, como pode ser visto na Fig. 3.2. Finalmente, uma corrida de wT-metaD de 1 ns sem potencial de viés foi realizado para estimar o desvios padrão das variáveis  $s(R)$  e  $z(R)$ .  $s(R)$  descreve o progresso da reação ao longo de uma caminho de referência predefinido (sequência de quadros selecionados), e  $z(R)$  a distância para este caminho. O quadro da estrutura de destino foi adicionado ao caminho no seu final.

As simulações EDS dos mutantes convergiram para um rmsd de menos de 0,1 nm usando as Componentes Principais (CP) dos estados mutantes, enquanto que para o nativo foi necessário usar os CP do estado ativo [101]. Usando o EDS como uma entrada (trajetórias das Fig. 3.2 (b,d,f)), as simulações wT-metaD convergiram de forma bastante monotônica. Usando o caminho completo foi calculado o rmsd entre os quadros e o  $\lambda$ , que é proporcional a distância média entre quadros vizinhos, foi estimado usando a equação:

$$\lambda = \frac{2.3(N - 1)}{\sum_{i=1}^{N-1} \|X_{i+1} - X_i\|^2} \quad (3.1)$$

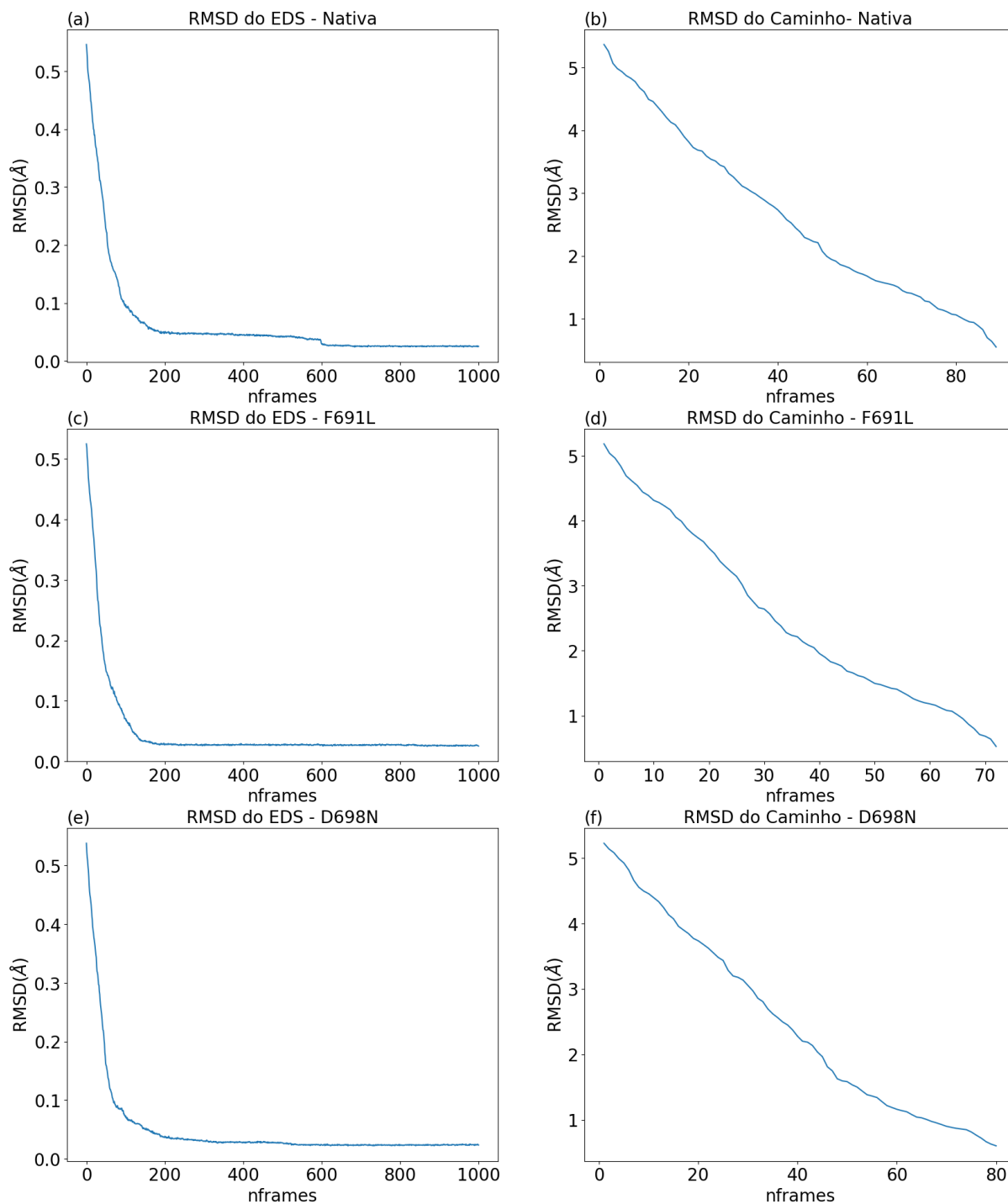


Figura 3.2: rmsd relativo ao ultimo quadro da trajetória de (a,c,e) EDS pura e (b,d,f) quadros selecionados com rmsd entre quadros aproximadamente constante e monotonicamente decrescente. (Figura elaborada pelo autor)

Todos os caminhos foram construídos selecionando apenas os quadros com rmsd em relação aos seus vizinhos em um intervalo de 0,045 a 0,060 nm. Este procedimento resultou em arquivos de caminho com 90, 73 e 81 quadros para forma nativa, F691L e D698N, respectivamente. Os valores de  $\lambda$  calculados para esses caminhos foram 885,5, 863,3 e

847,0 nm<sup>2</sup> para o estado nativo, F691L e D698N, respectivamente. A partir das corridas de 1 ns, o rmsd para s(R) e z(R) foi calculado para ser usado como uma estimativa das larguras das gaussianas. Os valores obtidos foram 0,45 nm<sup>2</sup> (F691L), 0,60 nm<sup>2</sup> (D698N) e 0,50 nm<sup>2</sup> (nativa), em s(R). Para z(R), os rmsd foram 0,002, 0,003 e 0,005 nm<sup>2</sup> para o F691L, D698N e a nativa, respectivamente.

As simulações wT-metaD foram executadas com os parâmetros calculados acima. Gaussianas foram depositadas a cada 500 passos com altura inicial ajustada em 1,0 kJ/mol; a altura foi reduzida com um *biás factor* de 25. Para limitar a exploração do espaço de fase e evitar que a simulação gaste tempo em estados com energias muito altas na direção z(R), uma força restritiva de 10<sup>4</sup> kJ mol<sup>-1</sup> Å<sup>-2</sup> foi aplicada quando z(R) excedeu 0,8 Å<sup>2</sup> para a nativa, ou 0,7 para os mutantes. Esta configuração foi repetido dez vezes para cada estado FLT3. Perto de 100 ns algumas das corridas convergiram para um valor razoável de estimativa para superfície de energia livre (FES). Estas simulações continuaram até 200 ns. As barreiras mais baixas observadas para o FES convergiu em diferentes tempos de simulação: 145 ns, 155 ns e 180 ns para nativa, D698N e F691L, respectivamente (Fig. 3.4).

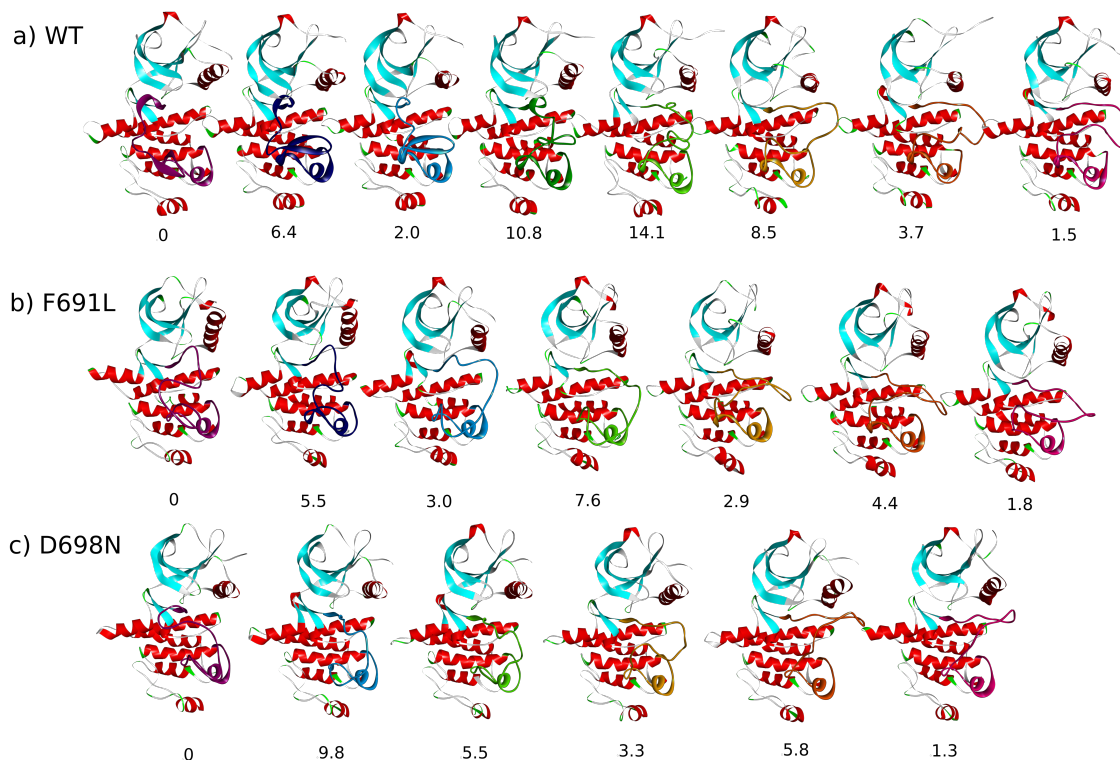


Figura 3.3: Estruturas ao longo da trajetória de ativação para wt (a), F691L (b) e D698N (c). Os valores abaixo de cada estrutura correspondem a diferença de energia de Gibbs em relação ao estado inativo.(Figura elaborada pelo autor)

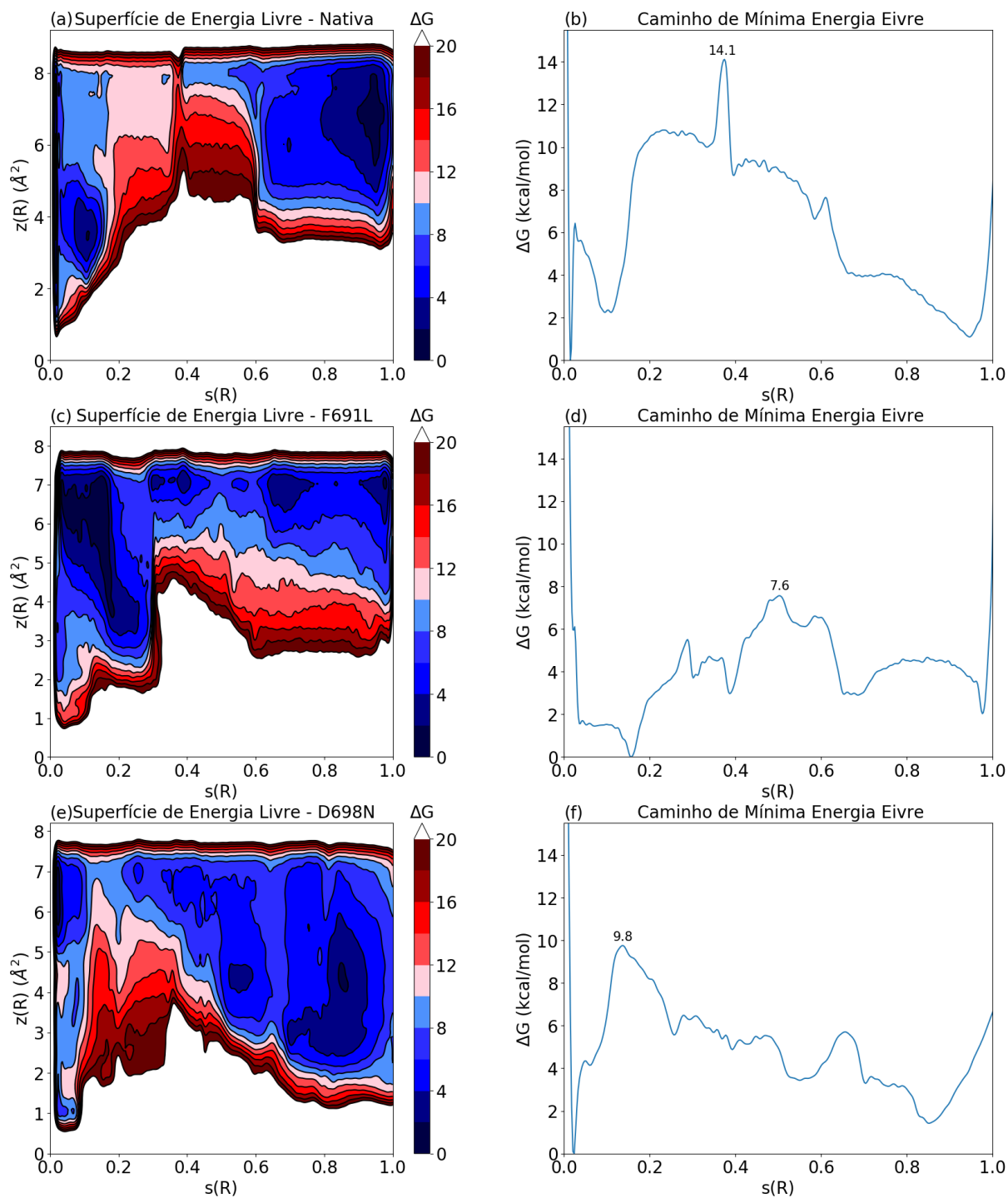


Figura 3.4: Superfície de Energia Livre (a,c,e) e Caminho de Mínima Energia (b,d,f) mostrando a barreira de ativação para a transição do inativo para o ativo no três casos.(Figura elaborada pelo autor)

### 3.3 Discussão e Resultados

O FEP é considerado um método preciso para calcular mudanças energia livre de Gibbs sob mutações no alvo do fármaco [44, 95] e foi usado aqui para este objetivo. Valores experimentais de  $\Delta\Delta G_b$  foram estimados a partir de medições de aumento de  $IC_{50}$  [95] em células Ba/F3, as quais foram feitas para dependem de FLT3/ITD para seu crescimento. Uma vez que as diferenças de pKa entre os nitrogênios proximais e distais (N5 e N6 na Figura 3.5) são comumente muito pequenas, os valores de  $\Delta\Delta G_b$  foram calculados com N5 ou N6 protonado, mantendo um total de +1 na carga eletrônica total do medicamento. Os resultados (Tabela 3.1) sugerem que a ligação é mais favorável com N6 protonado. Como esse nitrogênio é mais acessível ao solvente uma vez que o fármaco está ligado, também é mais provável que seu pKa seja maior. Em F691L, os cálculos mostram que o mutante liga-se ao fármaco com menor afinidade do que a proteína nativa. Os resultados das simulações subestimaram a variação na diferença de energia livre, mas a diferença foi inferior a  $0,6 \text{ kcal mol}^{-1}$ , ou seja, menor que a energia térmica. Para D698N, onde a carga total da proteína também é modificada,  $\Delta\Delta G_b$  é aproximadamente zero. Isso sugere que (1) os cálculos não são capazes de capturar algumas mudanças sutis que interferem na ligação ou (2) existe outro fator que pode contribuir para a diferença obtida experimentalmente. Também é possível que as medições no aumento de  $IC_{50}$  em células Ba/F3 não podem ser traduzidas com tanta precisão para valores de  $\Delta\Delta G_b$ , pois esses valores não são medidos direto na proteína e sim em células. Por exemplo, se FLT3/ITD/D698N é mais eficiente na catálise, as células que transportam este mutante tem mais chances de sobreviver, mesmo que a diferença de afinidades proteína-fármaco não seja tão grande quanto o valor  $\Delta\Delta G_b$  estimado do experimento.

Para entender melhor a diferença entre a nativa e os dois mutantes, analisamos as ligações de hidrogênio e os contatos envolvendo a proteína e o fármaco. Surpreendentemente, os resultados (Tabela 3.2) mostram um aumento no número de ligações de hidrogênio e

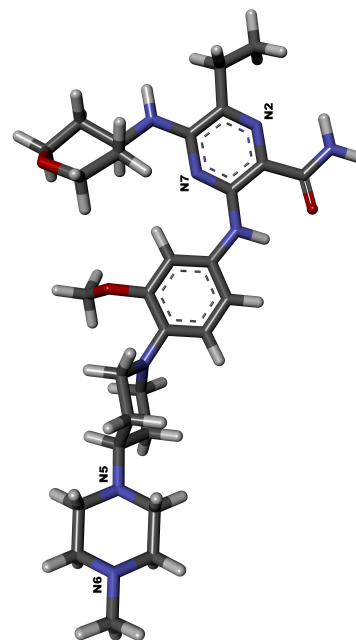


Figura 3.5: A estrutura do gilteritinib. A porção piperazina é básica, portanto, nitrogênios N5 e N6 podem ser protonados. (Fonte: Elaborada pelo autor com base nas estruturas encontradas no PDB)



Tabela 3.1: Diferenças de energia livre de Gibbs (mutantes versus nativa) para a ligação de gilteritinib com FLT3. Todos os valores estão em kcal mol<sup>-1</sup>. A diferença de energia de Gibbs foram calculadas por FEP (Calc.). Estimativas experimentais (Exp.) são dados como referência.

Protonação	F691L	D698N
N5	0.41 ± 0.31	-0.64 ± 0.44
N6	0.75 ± 0.32	-0.19 ± 0.35
Exp*	1.30 ± 0.12	1.43 ± 0.25

\*Os valores experimentais foram estimados a partir de medições de aumento de IC<sub>50</sub> em células Ba/F3 que dependem da nativa ou mutante FLT3/ITD para crescimento [96]

Tabela 3.2: O número de ligações de hidrogênio e contatos proteína-fármaco (excluindo átomos de hidrogênio, d < 0,4 nm) conforme observado em simulações MD. Percentual de ocupação para ligações de hidrogênio específicas com ocupação >10% dos quadros da trajetória de simulação também são fornecidos na tabela.

	wt	F691L	D698N
# ligações de hidrogênio (HB)	1.4 ± 0.1	2.1 ± 0.8	2.2 ± 0.6
# contatos	42 ± 16	51 ± 13	51 ± 11
Percentual de ocupação: (HB específicas)			
Gil:N3–E692:O	25	87	94
Gil:N1–C694:O	0	18	7
Gil:N3–C694:O	29	18	0
C694:N–Gil:N3	31	14	9
C694:N–Gil:O2	27	86	97

contatos após mutação. Isso, no entanto, é compensado por um menor Gibbs energia da mutante livre após a mutação, o que leva as diferenças de energia de Gibbs mostradas na Tabela 3.1. Os mesmos resíduos estão envolvidos em ligações de hidrogênio em todos os casos, ou seja, Glu<sup>692</sup> e Cys<sup>694</sup>. Estes também são os resíduos que se ligam ao fármaco através de pontes de hidrogênio na estrutura cristalina do estado inativo da FLT3 ligada ao gilteritinibe [97]. Parece haver uma flexibilidade um tanto maior na Cys<sup>694</sup> na wt, onde

o resíduo pode ligar tanto para nitrogênio N3 quanto para oxigênio O2 do gilteritinibe, enquanto em nos mutantes a interação é principalmente com o oxigênio.

Outro aspecto que pode contribuir para a resistência é o desenvolvimento de mutantes mais eficientes na catálise. Devido à farmacocinética e farmacodinâmica, espera-se que uma pequena fração da proteína alvo seja ativa mesmo em pacientes tratados; se esta fração é mutante e mais eficiente, isso pode levar à resistência ao fármaco, como mostrado para a Abl1 quinase [47]. A preferência pelo estado ativo nos mutantes [102] provavelmente desempenha um papel na resistência de alguns mutantes. Assim, nos propusemos a examinar o perfil de energia de Gibbs ao longo da ativação da proteína para o wt e mutantes, usando nossa abordagem recentemente desenvolvida [103], que combina amostragem dinâmica essencial (EDS) e metadinâmica bem temperada (wT-metaD [43]).

Os perfis de energia livre e os caminhos de energia de ativação ao longo das coordenadas de transição  $s(R)$  [41] são mostradas na Fig. 3.4. Em todas as variantes, o estado inativo é favorecido por cerca de  $2 \text{ kcal mol}^{-1}$  (espera-se que a autofosforilação modifique a diferença entre o estados [84]). Dois mínimos claros (correspondentes aos estados inativo e ativo) são observados para todas as variantes, com algumas diferenças na dinâmica e presença de mínimos rasos adicionais. A barreira, calculada em função de  $s(R)$  difere entre as variantes. É maior para a nativa ( $14,1 \text{ kcal mol}^{-1}$ ) e menor para os mutantes ( $7,6 \text{ kcal mol}^{-1}$  e  $9,8 \text{ kcal mol}^{-1}$  para F691L e D698N, respectivamente), o que sugere que os mutantes poderiam mais facilmente tornar-se ativas. A barreira para a nativa foi previamente estimada como  $17,8 \text{ kcal mol}^{-1}$  com base em estruturas retiradas diretamente do EDS [42]. Aparentemente, o refinamento do caminho por wT-metaD fornece uma estimativa mais baixa que provavelmente é mais precisa.

A ativação da proteína envolve mudanças em várias características, principalmente o loop de ativação e a hélice  $\alpha C$  [42], como mostrado na Figura 3.3. A transição entre estruturas inativas e ativas começa com o desenrolar do loop de ativação, que é visto no topo dos dois lobos da quinase. No intermediário de alta energia, o loop é desenrolado, mas não adota sua conformação final. Na proteína wt e mutante F691L, hélice  $\alpha C$  (a única hélice no lóbulo N, que está na parte superior da figura) já adota uma conformação próxima à do estado ativo. No mutante D698N, a hélice ainda está próxima de sua localização inicial no ponto intermediário mais alto energético. Caso contrário, as transições são bastante semelhantes entre o peso e as proteínas mutantes.

### 3.4 Conclusão

Nossos cálculos mostram que o aumento na energia de ligação sobre mutação da quinase FLT3 é semelhante ao valor esperado baseado em experimentos de inibição em células Ba/F3 para um mutante (F691L) mas não o outro (D698N). O primeiro mutante é indicado como uma fonte de resistência em um ambiente clínico, enquanto o segundo até agora não foi identificado em pacientes, apesar de levar a um maior grau de resistência em um modelo celular. A descoberta de que os mutantes têm barreiras de energia livre de Gibbs para ativação mais baixas traduz também para outros KIs, onde a resistência pode depender de uma combinação de (1) modificação das interações proteína-fármaco e (2) mudanças na dinâmica entre os estados ativo e inativo. Nossos resultados sugerem que ambos os mutantes são mais prováveis do que a nativa a serem ativados, o que pode explicar em parte o perfil de resistência, como visto no modelo celular. Em um ambiente clínico, a farmacodinâmica (absorção e distribuição do fármaco, ligação a outras proteínas), bem como farmacocinética (metabolismo e excreção) provavelmente levaria a que alguma fração da proteína total não se ligasse. Isso, por sua vez, sugere que uma variante com menor barreira entre os estados tem um benefício inerente, mesmo que se ligue ao inibidor em um modo semelhante. Quão significativa pode ser tal variação para inibidores de quinase do tipo 1 é difícil de estimar. Também resta ver se a D698N será observada em pacientes. A LMA é uma doença relativamente rara e o número de pacientes que foram tratados por gilteritinib é bastante pequeno, tornando difícil concluir se esta mutação leva a resistência em pacientes. Além disso, com gilteritinib visto como um fármaco de último recurso, a maioria dos pacientes não passa por mais testes para mutações na FLT3 após a falha do teste clínico, pois não há outros inibidores que podem ajudar [88].

É difícil julgar a precisão dos resultados, pois as estimativas experimentais das barreiras de ativação estão longe de ser triviais. Nossos resultados anteriores com Abl1 [100] estavam de acordo com outras estimativas baseadas em diferentes teorias, que dão alguma credibilidade à abordagem aqui utilizada. Além disso, evidências como observação de estados intermediários que se assemelham a estruturas proteicas encontradas pelo nosso método [42, 101] indicam que as trajetórias ao longo da ativação caminho são razoáveis.

Finalmente, pode ser interessante considerar como utilizar nossos resultados para o design de fármacos. As lições de outras proteínas quinases têm mostrado que é possível de-

desenvolver fármacos que inibam também mutantes específicos (inibidores de pan-resistência). Também pode ser possível desenvolver fármacos que se ligam a intermediários ao longo da via de ativação [42]. Tais fármacos inibem a proteína e também são esperadas a impedir sua ativação.

### **3.5 Informações Adicionais**

O trabalho deste capítulo foi desenvolvido em colaboração com Department of Chemistry and Biomedical Sciences, Linnaeus University at Kalmar, na Suécia, coordenado pelo prof. Dr. Ran Friedman (Grupo CCBG). As simulações foram realizadas com uso do cluster Lunarc (<http://www.lunarc.lu.se/>).

# CONSIDERAÇÕES FINAIS

Proteínas alostéricas realizam funções extremamente importantes nos mais diversos organismos, promovendo uma regulação diferenciada na concentração de ligantes disponíveis em solução e a variação da velocidade de reações. Entender tais mecanismos moleculares dá a possibilidade de corrigir eventuais disfunções fisiológicas, sendo o desenvolvimento de medicamentos um dos principais frutos de estudos como este.

Nos três casos tratados aqui, foi possível notar a eficiência de métodos baseados em física para o entendimento molecular de mecanismos biológicos. Ferramentas como o Docking e Dinâmica Molecular Clássica permitem entender e complementar informações obtidas experimentalmente, além de gerar nos resultados que podem ser testados do por meio de experimentos. Porém, por mais eficientes que sejam estes métodos computacionais, a modelagem ainda é um trabalho complexo, pois o efeito estudado pode não se resumir apenas a um conjunto parâmetro. Os casos da Abl1 e FLT3 exemplificam esse problema, visto que enzimas tão similares possam exibir dinâmicas distintas para seus inibidores frente a determinadas mutações.

Por mais que tenham sido foco de estudos separados, as proteínas estudadas aqui se conectam pela farmacodinâmica. Assim, entender o mecanismo conjunto dessas moléculas pode resultar na formulação de inibidores mais efetivos, reduzindo efeitos colaterais e aumentando o poder de inibição.

Estudando a HSA verificamos diferenças estruturais entre a HSA com e sem ácidos graxos. Estas mudanças foram estudadas por dinâmica molecular, e foi possível verificar alguns estados de transição. Iniciamos com uma estrutura do PDB com ácidos graxos, e após retirarmos os ácidos graxos, conseguimos atingir a estrutura da HSA sem ácidos graxos. Apesar de ser um monômero, ela apresenta estados alostéricos que estão relacionados com a sua função de transporte, porém não é claro na literatura como a sua interação com seus ligantes afeta a transição entre um estado alostérico e outro, ou ainda como as mudanças de estados podem afetar a afinidade dos ligantes.

Os resultados dos estudos de dinâmica molecular utilizando metadinâmica possibilitaram calcular a variação de energia livre durante a transição com uso de variáveis

coletivas geométricas e relativas a um caminho referencial da transição e ainda obter uma trajetória com possíveis estados intermediários.

Percebe-se que desta forma, será possível estudar a associação de fármacos ou outros compostos exógenos ou entender a associação de inacessíveis ou criar novos sítios de ligação. O fato de entender como a presença de ácidos graxos pode mudar a conformação de HSA, pode ser importante para entender a associação de fármacos exógenos ou compostos endógenos e as constantes de associação ou dissociação dos compostos com HSA. Assim será possível melhorar os fármacos e sua associação com HSA e sua liberação.

# Referências Bibliográficas

- [1] Freire, T. S., “Inibição enzimática analisada com uma abordagem da mecânica estatística,” Master’s thesis, Universidade Federal de São Carlos, 2019. [Online]. Available: <https://repositorio.ufscar.br/handle/ufscar/11408>
- [2] Michaelis, L. and Menten, M. L., “Die kinetik der invertinwirkung,” *Biochem Z*, vol. 49, pp. 333–369, 1913.
- [3] Johnson, K. A. and Goody, R. S., “The original Michaelis constant: Translation of the 1913 Michaelis-Menten paper,” *Biochemistry*, vol. 50, pp. 8264–8269, 2011.
- [4] Hill, A. V., “The possible effects of the aggregation of the molecules of hemoglobin on its dissociation curves,” *Journal of Physiology*, vol. 40, pp. 4–7, 1910.
- [5] Cornish-Bowden, A., *Fundamentals of Enzyme Kinetics*, 4th ed. Wiley-Blackwell, 2012.
- [6] Berman, H. M., “The protein data bank,” *Nucleic Acids Research*, vol. 28, no. 1, p. 235–242, 2000.
- [7] BIOVIA, D. S., “Biovia discovery studio. version 2021.San Diego: Dassault Systèmes, 2022.” [Online]. Available: <https://www.3ds.com/products-services/biovia/products/molecular-modeling-simulation/biovia-discovery-studio/>
- [8] Vriend, G., “What if: A molecular modeling and drug design program,” *Journal of Molecular Graphics*, vol. 8, no. 1, p. 52–56, 1990.
- [9] Jo, S., Kim, T., Iyer, V. G., and Im, W., “Charmm-gui: A web-based graphical user interface for charmm,” *Journal of Computational Chemistry*, vol. 29, no. 11, pp. 1859–1865, 2008. [Online]. Available: <https://onlinelibrary.wiley.com/doi/abs/10.1002/jcc.20945>
- [10] Lee, J., Cheng, X., Swails, J. M., Yeom, M. S., Eastman, P. K., Lemkul, J. A., Wei, S., Buckner, J., Jeong, J. C., Qi, Y., Jo, S., Pande, V. S., Case, D. A., Brooks, C. L., MacKerell, A. D., Klauda, J. B., and Im, W., “Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field,” *Journal of Chemical Theory and Computation*, vol. 12, no. 1, pp. 405–413, 2016, pMID: 26631602. [Online]. Available: <https://doi.org/10.1021/acs.jctc.5b00935>
- [11] Cunha, R. L., Zukerman-Schpector, J., Caracelli, I., and Comasseto, J. V., “Revisiting the addition reaction of tecl4 to alkynes: The crystal structure and docking studies of 1-chloro-2-trichlorotelluro-3-phenyl-propen-2-ol,” *Journal of Organometallic Chemistry*, vol. 691, no. 23, p. 4807–4815, 2006.
- [12] Vega-Tejido, M., Caracelli, I., and Zukerman-Schpector, J., “Conformational analyses and docking studies of a series of 5-nitrofuran- and 5-nitrothiophen-semicarbazone derivatives in three possible binding sites of trypanothione and glutathione reductases,” *Journal of Molecular Graphics and Modelling*, vol. 24, no. 5, p. 349–355, 2006.
- [13] Jones, G., Willett, P., and Glen, R. C., “Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation,” *Journal of Molecular Biology*, vol. 245, no. 1, p. 43–53, 1995.
- [14] Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R., “Development and validation of a genetic algorithm for flexible docking 1 (edited by f. e. cohen),” *Journal of Molecular Biology*, vol. 267, no. 3, p. 727–748, 1997.
- [15] Sapundzhi, F., Prodanova, K., and Lazarova, M., “Survey of the scoring functions for protein-ligand docking,” *PROCEEDINGS OF THE 45TH INTERNATIONAL CONFERENCE ON APPLICATION OF MATHEMATICS IN ENGINEERING AND ECONOMICS (AMEE’19)*.
- [16] Clark, M., Cramer, R. D., and Van Opdenbosch, N., “Validation of the general purpose tripos 5.2 force field,” *Journal of Computational Chemistry*, vol. 10, no. 8, p. 982–1012, 1989.
- [17] Ermondi, G. and Caron, G., “Recognition forces in ligand–protein complexes: Blending information from different sources,” *Biochemical Pharmacology*, vol. 72, no. 12, p. 1633–1645, 2006.

- [18] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E., "Ucsf chimera—a visualization system for exploratory research and analysis," *Journal of Computational Chemistry*, vol. 25, no. 13, p. 1605–1612, 2004.
- [19] Berendsen, H. J. C., Spoel, D. V. D., and Drunen, R. V., "Gromacs: A message-passing parallel molecular dynamics implementation," *Computer Physics Communications*, vol. 91, no. 1-3, p. 43–56, 1995.
- [20] James, M. A., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E., "Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers," *SoftwareX*, vol. 1-2, p. 19–25, 2015.
- [21] Spoel, D. V. D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., and Berendsen, H. J. C., "Gromacs: Fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, p. 1701–1718, 2005.
- [22] Van Gunsteren, W. F. and Berendsen, H. J., "A leap-frog algorithm for stochastic dynamics," *Molecular Simulation*, vol. 1, no. 3, p. 173–185, 1988.
- [23] Mackerell, A. D., "Empirical force fields for biological macromolecules: Overview and issues," *Journal of Computational Chemistry*, vol. 25, no. 13, p. 1584–1604, 2004.
- [24] Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., and Mackerell Jr., A. D., "Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields," *Journal of Computational Chemistry*, 2009.
- [25] Swails, J., "Free energy simulations of complex biological systems at constant ph," Ph.D. dissertation, 07 2013.
- [26] Lindahl, E. R., "Molecular dynamics simulations," *Methods in Molecular Biology*, p. 3–23, 2008.
- [27] Meza, J. C., "Steepest descent," *WIREs Computational Statistics*, vol. 2, no. 6, p. 719–722, 2010.
- [28] Gibbs, J. W., "Elementary principles in statistical mechanics developed with especial reference to the rational foundation of thermodynamics / by j. willard gibbs." 1902.
- [29] Tuckerman, M. E., *Statistical mechanics: Theory and molecular simulation*. Oxford University Press, 2018.
- [30] Berendsen, H. J. C., Postma, J. P. M., Gunsteren, W. F. V., Dinola, A., and Haak, J. R., "Molecular dynamics with coupling to an external bath," *The Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [31] Parrinello, M. and Rahman, A., "Polymorphic transitions in single crystals: A new molecular dynamics method," *Journal of Applied Physics*, vol. 52, no. 12, pp. 7182–7190, 1981.
- [32] Andersen, H. C., "Molecular dynamics simulations at constant pressure and/or temperature," *The Journal of Chemical Physics*, vol. 72, no. 4, p. 2384–2393, 1980.
- [33] Bussi, G., Donadio, D., and Parrinello, M., "Canonical sampling through velocity rescaling," *The Journal of Chemical Physics*, vol. 126, no. 1, p. 014101, Jul 2007.
- [34] Amadei, A., Linssen, A. B., and Berendsen, H. J., "Essential dynamics of proteins," *Proteins: Structure, Function, and Genetics*, vol. 17, no. 4, p. 412–425, 1993.
- [35] Crooks, G. E., "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences," *Physical Review E*, vol. 60, no. 3, pp. 2721–2726, Jan 1999.
- [36] Jarzynski, C., "Nonequilibrium equality for free energy differences," *Physical Review Letters*, vol. 78, no. 14, pp. 2690–2693, Jul 1997.
- [37] Bennett, C. H., "Efficient estimation of free energy differences from monte carlo data," *Journal of Computational Physics*, vol. 22, no. 2, pp. 245–268, 1976.
- [38] Shirts, M. R., Bair, E., Hooker, G., and Pande, V. S., "Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods," *Physical Review Letters*, vol. 91, no. 14, Feb 2003.
- [39] Salinas, S. R. A., *Introdução a Física Estatística*, 2nd ed. Edusp, 2005.



- [40] Bussi, G. and Branduardi, D., *Free-Energy Calculations with Metadynamics: Theory and Practice*, 05 2015, pp. 1–49.
- [41] Barducci, A., Bonomi, M., and Parrinello, M., “Metadynamics,” *WIREs Computational Molecular Science*, vol. 1, no. 5, p. 826–843, 2011.
- [42] Todde, G. and Friedman, R., “Conformational modifications induced by internal tandem duplications on the *flt3* kinase and juxtamembrane domains,” *Physical Chemistry Chemical Physics*, vol. 21, no. 34, p. 18467–18476, 2019.
- [43] Barducci, A., Bussi, G., and Parrinello, M., “Well-tempered metadynamics: A smoothly converging and tunable free-energy method,” *Phys. Rev. Lett.*, vol. 100, p. 020603, Jan 2008. [Online]. Available: <https://link.aps.org/doi/10.1103/PhysRevLett.100.020603>
- [44] Dávila-Rodríguez, M. J., Freire, T. S., Lindahl, E., Caracelli, I., Zukerman-Schpector, J., and Friedman, R., “Is breaking of a hydrogen bond enough to lead to drug resistance?” *Chemical Communications*, vol. 56, no. 49, p. 6727–6730, 2020.
- [45] Soverini, S., Mancini, M., Bavaro, L., Cavo, M., and Martinelli, G., “Chronic myeloid leukemia: the paradigm of targeting oncogenic tyrosine kinase signaling and counteracting resistance for successful cancer therapy,” *Molecular Cancer*, vol. 17, no. 1, 2018.
- [46] Voet, D., Voet, J. G., and Pratt, C. W., *Fundamentals of Biochemistry*, 4th ed. John Wiley and Sons, Inc, 2013.
- [47] Georgoulia, P. S., Todde, G., Bjelic, S., and Friedman, R., “The catalytic activity of *abl1* single and compound mutations: Implications for the mechanism of drug resistance mutations in chronic myeloid leukaemia,” *Biochimica et Biophysica Acta (BBA) - General Subjects*, vol. 1863, no. 4, p. 732–741, 2019.
- [48] Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., and Lindahl, E., “Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers,” *SoftwareX*, vol. 1-2, pp. 19–25, 2015.
- [49] Best, R. B., Zhu, X., Shim, J., Lopes, P. E., Mittal, J., Feig, M., and MacKerell, A. D., “Optimization of the additive charmm all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_2$  dihedral angles,” *Journal of Chemical Theory and Computation*, vol. 8, no. 9, p. 3257–3273, 2012.
- [50] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W., and Klein, M. L., “Comparison of simple potential functions for simulating liquid water,” *The Journal of Chemical Physics*, vol. 79, no. 2, p. 926–935, 1983.
- [51] Hess, B., Bekker, H., Berendsen, H. J., and Fraaije, J. G., “Lincs: A linear constraint solver for molecular simulations,” *Journal of Computational Chemistry*, vol. 18, no. 12, p. 1463–1472, 1997.
- [52] Miyamoto, S. and Kollman, P. A., “Settle: An analytical version of the shake and rattle algorithm for rigid water models,” *Journal of Computational Chemistry*, vol. 13, no. 8, p. 952–962, 1992.
- [53] Gapsys, V., Michielssens, S., Seeliger, D., and Groot, B. L. D., “pmx: Automated protein structure and topology generation for alchemical perturbations,” *Journal of Computational Chemistry*, vol. 36, no. 5, p. 348–354, 2014.
- [54] Tokarski, J. S., Newitt, J. A., Chang, C. Y. J., Cheng, J. D., Wittekind, M., Kiefer, S. E., Kish, K., Lee, F. Y. F., Borzilleri, R., Lombardo, L. J., and et al., “The structure of dasatinib (bms-354825) bound to activated *abl* kinase domain elucidates its inhibitory activity against imatinib-resistant *abl* mutants,” *Cancer Research*, vol. 66, no. 11, p. 5790–5797, 2006.
- [55] Aldeghi, M., de Groot, B. L., and Gapsys, V., “Accurate calculation of free energy changes upon amino acid mutation,” *Methods in Molecular Biology*, p. 19–47, 2018.
- [56] Bradeen, H. A., Eide, C. A., Ohare, T., Johnson, K. J., Willis, S. G., Lee, F. Y., Druker, B. J., and Deininger, M. W., “Comparison of imatinib mesylate, dasatinib (bms-354825), and nilotinib (amn107) in an n-ethyl-n-nitrosourea (enu)based mutagenesis screen: high efficacy of drug combinations,” *Blood*, vol. 108, no. 7, pp. 2332–2338, Jan 2006.

- [57] Zabriskie, M. S., Eide, C. A., Tantravahi, S. K., Vellore, N. A., Estrada, J., Nicolini, F. E., Khoury, H. J., Larson, R. A., Konopleva, M., Cortes, J. E., and et al., “Bcr-abl1 compound mutations combining key kinase domain positions confer clinical resistance to ponatinib in ph chromosome-positive leukemia,” *Cancer Cell*, vol. 26, no. 3, pp. 428–442, 2014.
- [58] Redaelli, S., Mologni, L., Rostagno, R., Piazza, R., Magistroni, V., Ceccon, M., Viltadi, M., Flynn, D., and Gambacorti-Passerini, C., “Three novel patient-derived bcr/abl mutants show different sensitivity to second and third generation tyrosine kinase inhibitors,” *American Journal of Hematology*, vol. 87, no. 11, Sep 2012.
- [59] Gruber, F. X., Ernst, T., Porkka, K., Engh, R. A., Mikkola, I., Maier, J., Lange, T., and Hochhaus, A., “Dynamics of the emergence of dasatinib and nilotinib resistance in imatinib-resistant cml patients,” *Leukemia*, vol. 26, no. 1, pp. 172–177, May 2011.
- [60] Lee, C.-C., Chuang, Y.-C., Liu, Y.-L., and Yang, C.-N., “A molecular dynamics simulation study for variant drug responses due to fms-like tyrosine kinase 3 g697r mutation,” *RSC Advances*, vol. 7, no. 47, p. 29871–29881, 2017.
- [61] Mishra, P. and Günther, S., “New insights into the structural dynamics of the kinase jnk3,” *Scientific Reports*, vol. 8, no. 1, 2018.
- [62] Yun, C.-H., Mengwasser, K. E., Toms, A. V., Woo, M. S., Greulich, H., Wong, K.-K., Meyerson, M., and Eck, M. J., “The t790m mutation in egfr kinase causes drug resistance by increasing the affinity for atp,” *Proceedings of the National Academy of Sciences*, vol. 105, no. 6, p. 2070–2075, 2008.
- [63] Anderson, N. L. and Anderson, N. G., “The human plasma proteome: History, character, and diagnostic prospects,” *Molecular & Cellular Proteomics*, vol. 1, no. 11, pp. 845–867, 2002. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1535947620342468>
- [64] Fanali, G., Masi, A. D., Trezza, V., Marino, M., Fasano, M., and Ascenzi, P., “Human serum albumin: From bench to bedside,” *Molecular Aspects of Medicine*, vol. 33, no. 3, p. 209–290, 2012.
- [65] Ascenzi, P. and Fasano, M., “Allostery in a monomeric protein: The case of human serum albumin,” *Biophysical Chemistry*, vol. 148, no. 1–3, p. 16–22, 2010.
- [66] Chuang, V. T. and Otagiri, M., *Pharmaceutical Research*, vol. 19, no. 10, p. 1458–1464, 2002.
- [67] Ghuman, J., Zunszain, P. A., Petitpas, I., Bhattacharya, A. A., Otagiri, M., and Curry, S., “Structural basis of the drug-binding specificity of human serum albumin,” *Journal of Molecular Biology*, vol. 353, no. 1, p. 38–52, 2005.
- [68] Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T., Weissig, H., Shindyalov, I., and Bourne, P., “The protein data bank,” pp. 235–242. [Online]. Available: [rcsb.org](http://rcsb.org)
- [69] Bhattacharya, A. A., Curry, S., and Franks, N. P., “Binding of the general anesthetics propofol and halothane to human serum albumin,” *Journal of Biological Chemistry*, vol. 275, no. 49, p. 38731–38738, 2000.
- [70] Bhattacharya, A. A., Grüne, T., and Curry, S., “Crystallographic analysis reveals common modes of binding of medium and long-chain fatty acids to human serum albumin 1 edited by r. huber,” *Journal of Molecular Biology*, vol. 303, no. 5, p. 721–732, 2000.
- [71] Bonomi, M., Branduardi, D., Bussi, G., Camilloni, C., Provasi, D., Raiteri, P., Donadio, D., Marinelli, F., Pietrucci, F., and Broglia, R. A. e. a., “Plumed: A portable plugin for free-energy calculations with molecular dynamics,” *Computer Physics Communications*, vol. 180, no. 10, pp. 1961–1972, 2009.
- [72] Tribello, G. A., Bonomi, M., Branduardi, D., Camilloni, C., and Bussi, G., “Plumed 2: New feathers for an old bird,” *Computer Physics Communications*, vol. 185, no. 2, p. 604–613, 2014.
- [73] Choe, J.-I. and Kim, B., “Determination of proper time step for molecular dynamics simulation,” *Bulletin of the Korean Chemical Society*, vol. 21, 04 2000.
- [74] Hartmann, C., Banisch, R., Sarich, M., Badowski, T., and Schütte, C., “Characterization of rare events in molecular dynamics,” *Entropy*, vol. 16, no. 1, pp. 350–376, 2013.

- [75] Daura, X., Gademann, K., Jaun, B., Seebach, D., van Gunsteren, W., and Mark, A., "Peptide folding: When simulation meets experiment," *Angewandte Chemie International Edition*, vol. 38, no. 1-2, p. 236–240, 1999.
- [76] Amadei, A., Linssen, A. B., de Groot, B. L., van Aalten, D. M., and Berendsen, H. J., "An efficient method for sampling the essential subspace of proteins," *Journal of Biomolecular Structure and Dynamics*, vol. 13, no. 4, p. 615–625, 1996.
- [77] Berteotti, A., Cavalli, A., Branduardi, D., Gervasio, F. L., Recanatini, M., and Parrinello, M., "Protein conformational transitions: The closure mechanism of a kinase explored by atomistic simulations," *Journal of the American Chemical Society*, vol. 131, no. 1, pp. 244–250, 2008.
- [78] Formoso, E., Limongelli, V., and Parrinello, M., "Energetics and structural characterization of the large-scale functional motion of adenylate kinase," *Scientific Reports*, vol. 5, no. 1, 2015.
- [79] Branduardi, D., Gervasio, F. L., and Parrinello, M., "From a to b in free energy space," *The Journal of Chemical Physics*, vol. 126, no. 5, p. 054103, 2007.
- [80] Monod, J., Wyman, J., and Changeux, J.-P., "On the nature of allosteric transitions: a plausible model." *Journal of Molecular Biology*, vol. 12, p. 88–118, 1965.
- [81] Freire, T. S., Caracelli, I., Zukerman-Schpector, J., and Friedman, R., "Resistance to a tyrosine kinase inhibitor mediated by changes to the conformation space of the kinase," *Physical Chemistry Chemical Physics*, vol. 25, no. 8, p. 6175–6183, 2023.
- [82] Attwood, M. M., Fabbro, D., Sokolov, A. V., Knapp, S., and Schiöth, H. B., "Trends in kinase drug discovery: Targets, indications and inhibitor design," *Nature Reviews Drug Discovery*, vol. 20, no. 11, p. 839–861, 2021.
- [83] Cohen, P., Cross, D., and Jänne, P. A., "Kinase drug discovery 20 years after imatinib: Progress and future directions," *Nature Reviews Drug Discovery*, vol. 20, no. 7, p. 551–569, 2021.
- [84] Friedman, R. and Bjelic, S., "Simulations studies of protein kinases that are molecular targets in cancer," *Israel Journal of Chemistry*, vol. 60, no. 7, p. 667–680, 2020.
- [85] Ayala-Aguilera, C. C., Valero, T., Lorente-Macías, A., Baillache, D. J., Croke, S., and Unciti-Broceta, A., "Small molecule kinase inhibitor drugs (1995–2021): Medical indication, pharmacology, and synthesis," *Journal of Medicinal Chemistry*, vol. 65, no. 2, p. 1047–1131, 2021.
- [86] Friedman, R., "Computational studies of protein–drug binding affinity changes upon mutations in the drug target," *WIREs Computational Molecular Science*, vol. 12, no. 1, 2021.
- [87] Lindström, H. J., de Wijn, A. S., and Friedman, R., "Stochastic modelling of tyrosine kinase inhibitor rotation therapy in chronic myeloid leukaemia," *BMC Cancer*, vol. 19, no. 1, 2019.
- [88] Yang, J., Lindström, H. J., and Friedman, R., "Combating drug resistance in acute myeloid leukaemia by drug rotations: The effects of quizartinib and pexidartinib," *Cancer Cell International*, vol. 21, no. 1, 2021.
- [89] Georgoulia, P. S., Bjelic, S., and Friedman, R., "Deciphering the molecular mechanism of ft3 resistance mutations," *The FEBS Journal*, vol. 287, no. 15, p. 3200–3220, 2020.
- [90] Friedman, R., Boye, K., and Flatmark, K., "Molecular modelling and simulations in cancer research," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1836, no. 1, p. 1–14, 2013.
- [91] Moretto, L., Ušaj, M., Matusovsky, O., Rassier, D. E., Friedman, R., and Månsson, A., "Multistep orthophosphate release tunes actomyosin energy transduction," *Nature Communications*, vol. 13, no. 1, 2022.
- [92] Williams-Noonan, B. J., Yuriev, E., and Chalmers, D. K., "Free energy methods in drug design: Prospects of "alchemical perturbation" in medicinal chemistry," *Journal of Medicinal Chemistry*, vol. 61, no. 3, p. 638–649, 2018.
- [93] Buelow, D. R., Bhatnagar, B., Jeon, J. Y., Zavorka Thomas, M., Eisenmann, E., Stromatt, J., Orwick, S., Pabla, N., Blachly, J. S., Baker, S., and et al., "High-dimensional analysis identifies mechanisms of gilteritinib resistance in ft3-mutated aml," *Blood*, vol. 138, no. Supplement 1, p. 207–207, 2021.

- [94] Todde, G. and Friedman, R., “Pattern and dynamics of flt3 duplications,” *Journal of Chemical Information and Modeling*, vol. 60, no. 8, p. 4005–4020, 2020.
- [95] Friedman, R., “The molecular mechanisms behind activation of flt3 in acute myeloid leukemia and resistance to therapy by selective inhibitors,” *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer*, vol. 1877, no. 1, p. 188666, 2022.
- [96] Tarver, T. C., Hill, J. E., Rahmat, L., Perl, A. E., Bahceci, E., Mori, K., and Smith, C. C., “Gilteritinib is a clinically active flt3 inhibitor with broad activity against flt3 kinase domain mutations,” *Blood Advances*, vol. 4, no. 3, p. 514–524, 2020.
- [97] Kawase, T., Nakazawa, T., Eguchi, T., Tsuzuki, H., Ueno, Y., Amano, Y., Suzuki, T., Mori, M., and Yoshida, T., “Effect of fms-like tyrosine kinase 3 (flt3) ligand (fl) on antitumor activity of gilteritinib, a flt3 inhibitor, in mice xenografted with fl-overexpressing cells,” *Oncotarget*, vol. 10, no. 58, p. 6111–6123, 2019.
- [98] Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E., “Ucsf chimera visualization system for exploratory research and analysis,” *Journal of Computational Chemistry*, vol. 25, no. 13, p. 1605–1612, 2004.
- [99] Seeliger, D. and de Groot, B. L., “Protein thermostability calculations using alchemical free energy simulations,” *Biophysical Journal*, vol. 98, no. 10, p. 2309–2316, 2010.
- [100] Oruganti, B. and Friedman, R., “Activation of abl1 kinase explored using well-tempered metadynamics simulations on an essential dynamics sampled path,” *Journal of Chemical Theory and Computation*, vol. 17, no. 11, p. 7260–7270, 2021.
- [101] Friedman, R. and Caffisch, A., “Pepsinogen-like activation intermediate of plasmeprin ii revealed by molecular dynamics analysis,” *Proteins: Structure, Function, and Bioinformatics*, vol. 73, no. 4, p. 814–827, 2008.
- [102] Friedman, R., “The molecular mechanism behind resistance of the kinase flt3 to the inhibitor quizartinib,” *Proteins: Structure, Function, and Bioinformatics*, vol. 85, no. 11, p. 2143–2152, 2017.
- [103] Oruganti, B., Lindahl, E., Yang, J., Amiri, W., Rahimullah, R., and Friedman, R., “Allosteric enhancement of the bcr-abl1 kinase inhibition activity of nilotinib by cobinding of asciminib,” *Journal of Biological Chemistry*, vol. 298, no. 8, p. 102238, 2022.

# Apêndice A

## Anexo 1



## Is breaking of a hydrogen bond enough to lead to drug resistance?<sup>†</sup>

Cite this: DOI: 10.1039/d0cc02164d

Received 24th March 2020,  
Accepted 6th May 2020

DOI: 10.1039/d0cc02164d

rsc.li/chemcomm

 María José Dávila-Rodríguez,<sup>a</sup> Thales Souza Freire,<sup>b</sup> Erik Lindahl,<sup>c</sup>  
Igne Caracelli,<sup>b</sup> Julio Zukerman-Schpector<sup>a</sup> and Ran Friedman<sup>\*c</sup>

Drug resistance is a serious problem in cancer, viral, bacterial, fungal and parasitic diseases. Examination of crystal structures of protein–drug complexes is often not enough to explain why a certain mutation leads to drug resistance. As an example, the crystal structure of the kinase inhibitor dasatinib bound to the Abl1 kinase shows a hydrogen bond between the drug and residue Thr<sup>315</sup> and very few contacts between the drug and residues Val<sup>299</sup> and Phe<sup>317</sup>, yet mutations in those residues lead to drug resistance. In the first case, it is tempting to suggest that the loss of a hydrogen bond leads to drug resistance, whereas in the other two cases it is not known why mutations lead to drug resistance in the first place. We carried out extensive molecular dynamics (MD) simulations and free energy calculations to explain drug resistance to dasatinib from a molecular point of view and show that resistance is due to a multitude of subtle effects. Importantly, our calculations could reproduce the experimental values for the binding energies upon mutations in all three cases and shed light on their origin.

Targeted cancer therapy, where specific proteins are targeted by drugs instead of, or in addition to chemotherapy, has gained much success in the last two decades. In particular, the availability of small molecules that inhibit the Abl1 kinase has tremendously decreased the five-year mortality from chronic myeloid leukaemia (CML). Unfortunately, resistance to targeted therapy is limiting the success of drugs in CML and many other cancers.<sup>1</sup> Mutations in the gene encoding the drug target protein are perhaps the most common factor leading to drug resistance. In few cases, structural studies can reveal why a certain mutation lead to drug resistance, *e.g.*, through direct interference with binding of the drug. In most cases, however, the underlying reason for resistance owing to mutations is not clear.

Kinases adopt two main states, active and inactive, with distinct conformational differences. Most Abl1 inhibitors bind the inactive state, and evidence shows that some mutations may work by shifting the equilibrium between the states.<sup>2,3</sup> It is possible to use advanced simulation methods such as free energy perturbation (FEP) to calculate the free energy difference between two structures (drug-sensitive and drug-resistance) upon mutation, but this requires that any changes to the free energy of binding  $\Delta\Delta G^b$  would not involve changes to the conformational equilibrium. In practice, this is often not the case, as changes upon resistance mutations often involve the active state.<sup>4,5</sup>

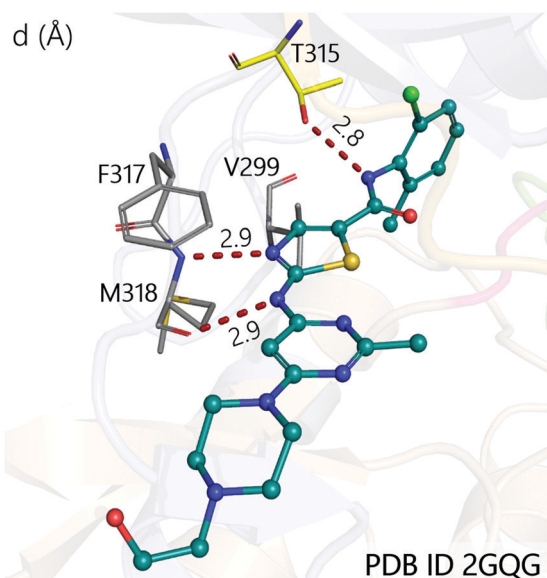
Dasatinib is an Abl1 inhibitor that is used in CML. This is the only clinically-available inhibitor that binds the active state of the protein, which makes the protein–drug complex a great model for studies of resistance mutations by computational methods. There are three mutations that lead to dasatinib resistance: V299L, T315I and F317L. Val<sup>299</sup> and Phe<sup>317</sup> form part of the binding pocket (Fig. 1) but only interact with the inhibitor *via* dispersion (*i.e.*, weak attractive and non-directional) interactions<sup>6</sup> (ESI,<sup>†</sup> Fig. S1a and b, non-covalent interactions<sup>7,8</sup>). It is therefore hard to see how a mutation to a rather similar hydrophobic residue would drive a change in  $\Delta\Delta G^b$  that would be enough to yield drug resistance. Thr<sup>315</sup> on the other hand is hydrogen bonded to the drug's amide nitrogen through its side-chain hydroxyl and through weaker interactions between the CH<sub>3</sub> group of the threonine and the drug's 2-chloro-6-methylphenyl ring (Fig. S1c, ESI<sup>†</sup>). The hydrogen bond is clearly lost upon mutation from threonine to isoleucine (Fig. S1d, ESI<sup>†</sup>).

The T315I mutation leads to an increase of  $3.0 \pm 0.3$  kcal mol<sup>-1</sup> in  $\Delta\Delta G^b$ .<sup>‡</sup> To examine if this is due to the loss of the hydrogen bond, we first calculated  $\Delta G^b$  by using a minimal, four residues model of the binding site and full quantum mechanics (QM) and then estimated the contribution of the hydrogen bond (ESI,<sup>†</sup> Fig. S2 and S3). The QM calculation yielded  $\Delta G^b = -13.3$  kcal mol<sup>-1</sup>, in good agreement with the experimental value of  $-14.4 \pm 0.1$  kcal mol<sup>-1</sup>.<sup>§</sup> The contribution of the hydrogen bond to

<sup>a</sup> Federal University of São Carlos, Department of Chemistry, São Carlos-SP, Brazil<sup>b</sup> Federal University of São Carlos, Department of Physics, São Carlos-SP, Brazil<sup>c</sup> Linnæus University, Department of Chemistry and Biomedical Sciences,

391 82 Kalmar, Sweden. E-mail: ran.friedman@lnu.se

<sup>†</sup> Electronic supplementary information (ESI) available: Methods, supplemental tables and supplemental figures. See DOI: 10.1039/d0cc02164d



**Fig. 1** Dasatinib binding and resistance mutations. Mutations in residues Val<sup>299</sup>, Thr<sup>315</sup> and Phe<sup>317</sup> lead to resistance against dasatinib. Met<sup>318</sup> interacts with dasatinib via two hydrogen bonds to its backbone atoms. Hydrogen bonds are represented by dashed lines in red with numbers on top of the lines representing donor–acceptor distances in Å.

Thr<sup>315</sup> model was  $-5.4 \text{ kcal mol}^{-1}$ . The use of larger models for the QM region was also considered. Whereas such larger models may in principle capture additional protein–drug interactions, they may on the other hand underestimate the strain that is imposed on each of the complex's components (protein and ligand), which can result in a binding energy that is artificially too favourable even if the QM calculation itself is accurate. Such artefacts increase with system size. In addition, using density functional theory (DFT) is associated with the so-called self interaction error (SIE), which also becomes larger with system size.<sup>14</sup> Indeed, the calculations resulted in  $\Delta G^b$  values of  $-15.8$  and  $-23.6 \text{ kcal mol}^{-1}$  for models 2 (intermediate, five residues) and 3 (large, nine residues), respectively (ESI,† Table S1). Additional details are given in the ESI.†

Although the notion that the loss of a hydrogen bond leads to resistance upon mutation seems appealing, hydrogen bonds are dynamic in the sense that they break and form in different parts of the protein and can involve parts of the drug that are solvent exposed in the crystal.<sup>15,16</sup> Moreover, the loss of a hydrogen bond cannot explain why V299L and F317L become resistant. For this reason, we studied the protein–drug interactions using multiple MD simulations, with the aim of following on protein–drug interactions.

The accuracy of MD simulations depends on many aspects including the simulation set-up and how accurate is the given forcefield (that is used to calculate the potential energy) for the problem at hand. For this reason, before simulating the bound proteins we calculated the binding free energies for the binding of dasatinib to the kinase domain (KD) of Abl1. Encouragingly, the calculations agreed to within  $0.7 \text{ kcal mol}^{-1}$  or less with the experimental  $\Delta\Delta G^b$  (Table 1).

**Table 1** Free energy differences of binding of dasatinib to Abl1 (mutants versus wt). All values are in  $\text{kcal mol}^{-1}$ . The free energy differences were calculated by FEP (calc.). Experimental values<sup>9–12</sup> (exp.) are given as reference. Values in parentheses are standard deviations given at the last significant digit, e.g.  $1.5(1) = 1.5 \pm 0.1$

V299L		T315I		F317L	
Calc.	Exp.	Calc.	Exp.	Calc.	Exp.
0.9(4)	1.5(1)	3.7(8)	3.0(3)	0.6(3)	1.0(1)

Examination of the number of hydrogen bonds (HBs) observed in the simulations for the wt protein reveals that the average number of protein–drug HBs is  $2.3 \pm 0.1$ , i.e., closer to two than to three. The number is smaller for the T315I mutant, but only by a fraction (Table 2, the distributions of HB numbers are shown in the ESI,† Fig. S4). This indicates that once Thr<sup>315</sup> is mutated to isoleucine, there is some compensation for the loss of the potential hydrogen bond. Apparently, the hydrogen bond between the amine nitrogen of the drug and the backbone oxygen of Met<sup>318</sup> is occupied in only 50% of the snapshots obtained from the simulations of the wt protein compared with 78% in the simulations of T315I (Table 2 and ESI,† Fig. S5). Of note, only in 33% of the configurations of the wt protein the hydrogen bonds between the drug and both Thr<sup>315</sup> and Met<sup>318</sup> were present simultaneously. Overall, the decrease in the number of HBs between the protein and the drug, from 2.3 to 1.9 was not enough to make the T315I mutant as resistance as it is (more than 70-fold). Rather, the number of protein–drug contacts, which contribute to the enthalpy of protein–drug binding through van der Waals (vdW) interactions is also decreased, leading to an increase of  $10.7 \text{ kcal mol}^{-1}$  in the protein–drug interaction energy,  $\Delta E^b$  (Table 2). The direct interaction energy with residue 315 is increased by  $4.1 \text{ kcal mol}^{-1}$  (ESI,† Table S2). The increase in  $\Delta E^b$ ,

**Table 2** Analysis of the MD simulations of wt and mutant Abl1. Numbers are averages calculated over 39 (wt, T315I, F317L) or 40 (V299L) separate 50 ns long simulations for each system, with standard deviation (last significant digit) in parentheses. #HB – number of hydrogen bonds. # contacts – number of heavy atom contacts between the protein and the drug ( $d < 4 \text{ Å}$ ).  $\Delta E^b$  – protein–drug interaction energy ( $\text{kcal mol}^{-1}$ ).  $S_m$  – absolute entropy for the bound drug.  $S_{\text{conf}}$  – configurational entropy of the protein's heavy atoms. Entropy values ( $\text{cal mol}^{-1} \text{ K}^{-1}$ ) were calculated from the simulations based on the quasiharmonic approximation<sup>17</sup>

	wt	V299L	T315I	F317L
#HB, protein–drug	2.3(1)	2.3(2)	1.9(1)	2.5(1)
HB occupancy (%):				
T315:OH	73(10)	66(24)	0(0)	74(16)
M318:N	99(1)	97(2)	97(1)	97(1)
M318:O	50(16)	60(16)	78(7)	70(11)
T315:OH & M318:O <sup>a</sup>	33(11)	50(24)	0(0)	62(20)
#HB, protein–water	662(27)	663(26)	663(27)	658(26)
# contacts	69(3)	66(6)	59(6)	68(4)
$\Delta E^b$	$-72.4(7)$	$-70.3(8)$	$-61.7(6)$	$-72.8(6)$
$S_m$	74.9(1.9)	75.9(2.1)	78.2(2.3)	74.4(2.4)
$S_{\text{conf}}$	1866(13)	1874(12)	1872(16)	1864(12)

<sup>a</sup> Hydrogen bonds T315:OH–dasatinib and M318:O–dasatinib present at the same time.

that is very close to the enthalpic contribution $\ddagger$  is somewhat offset by the increase in the absolute entropy of the bound drug (Table 2), that has a wider conformational energy landscape (ESI, $\dagger$  Fig. S6), and by a higher conformational entropy for the protein.

Having established that dasatinib resistance upon T315I Abl1 mutation is partly due to the loss of a hydrogen bond (which contributed about 40% to the loss of binding enthalpy) and partly due to the loss of vdW interactions with the protein residues, we set to examine the V299L and F317L mutants. At first glance, the protein–drug HB pattern of these mutants was similar to that in the wt protein. The number of protein–drug contacts was also similar, perhaps slightly lower for V299L (Table 2). For V299L, where the calculated  $\Delta\Delta G^b$  value was 0.9 kcal mol<sup>-1</sup> it is apparently an overall change in the binding energy ( $\Delta\Delta E^b$ , which increases by 2.1 kcal mol<sup>-1</sup> upon mutation) that dictates drug resistance and is the result of multiple contributions. It is not the change from valine to leucine alone that drove this increase in  $\Delta E^b$  (ESI, $\dagger$  Table S2). Rather, the mutation led to changes in the dynamics of the protein and the drug that was enough to drive resistance. One of the major contributions to the increase in  $\Delta\Delta E^b$  for the V299L mutant was that the interactions between Thr<sup>315</sup> and the drug became less favourable (by 0.8 kcal mol<sup>-1</sup>, Table S2).

For F317L the changes are even more subtle, as even  $\Delta E^b$  was similar for the wt and mutant protein. The absolute entropy of the bound drug was also similar. Since no enthalpic contribution was in effect, the mutation must have changed the entropy of the complex. Examination of the distributions of protein–drug hydrogen bonds and short contacts (ESI, $\dagger$  Fig. S4 and S7) reveals a narrower distribution for the mutant in both cases which suggests that the calculated free energy change of 0.6 kcal mol<sup>-1</sup> is due to a slight decrease in the conformational entropy of the protein. Our calculations indicated that this is indeed the case as the difference in entropy is  $-2$  cal mol<sup>-1</sup> K<sup>-1</sup> (Table 2), which would decrease the overall free energy by 0.6 kcal mol<sup>-1</sup> at 300 K. As this was not enough to explain the overall change in  $\Delta\Delta G^b$  upon mutation, we have also considered contribution from the solvent molecules. Examination of the number of protein–water hydrogen bonds shows a decrease only for the F317L mutant (Table 2), suggesting that protein–solvent interactions further contribute to the free energy difference between the drug-bound mutant KD and wt Abl1.

It may seem surprising that a change in the binding free energy of about 1 kcal mol<sup>-1</sup> as seen for V299L and F317L is enough to confer drug resistance. However, it should be remarked that pharmacokinetics and pharmacodynamics may also play a role. It has been shown that maintaining exposure of CD34+ cells to dasatinib at a concentration  $>IC_{50}$  could improve prognosis.<sup>18</sup> As  $IC_{50}$  is directly proportional to  $K_I$ , an increase of several-fold (0.6 kcal mol<sup>-1</sup> amounts to about three fold increase in  $K_I$ ) is enough to drive resistance to dasatinib.

In conclusion, we used extensive calculations to determine why mutations in Abl1 lead to dasatinib resistance. QM calculations with a rather small model could reproduce the experimental binding free energy and suggest that the loss of a hydrogen bond between the drug and residue Thr<sup>315</sup> amounted to 5.4 kcal mol<sup>-1</sup> which is more than the binding free energy

difference upon mutation. MD simulations reproduced  $\Delta\Delta G^b$  differences upon mutation to within 0.7 kcal mol<sup>-1</sup> or less of the experimental values, and correctly reproduced the order of severity of the mutations in causing drug resistance. The simulations attributed resistance to dasatinib in the Abl1 T315I mutants only partially to the loss of the HB. The mutation of Thr<sup>315</sup>, a key gatekeeper residue<sup>19,20</sup> changed the dynamics of the protein, lowered the interaction energy between the drug and multiple residues and increased the conformational entropy of the bound drug. The V299L and F317L mutations have subtle enthalpic and entropic effects, with the former leading to resistance due to an increase of  $\Delta H^b$  and the latter apparently due to a decrease of the conformational entropy of the inhibited protein and to a change in solvation of the complex. Our study highlights the subtlety of protein–drug interactions, where small changes upon mutation can affect the dynamics of many residues. In addition, it shows that MD simulations provide an accurate means to calculate  $\Delta\Delta G^b$  values when the drug binds the active state of a protein (unless there are large changes in  $K_m$ <sup>21</sup>). As this is the case, it is expected that similar calculations, if done properly, will underestimate  $\Delta\Delta G^b$  upon mutation for drugs that bind to the inactive state of an enzyme if the mutation leads to changes in the active state.

Referring to the question posed in the title, it is shown that the breaking of a hydrogen bond was to a large extent compensated by other interactions and hence was not the only (or main) driving force for drug resistance. In this context, it should also be noted that the Thr<sup>315</sup>:OH–dasatinib HB was less persistent than the HB where Met<sup>318</sup>:N was the donor. It is hence likely that the latter would not be compensated in the same manner. When it comes to drug design aimed at avoiding resistance, simulations can help in identifying such persistent HBs to the protein backbone, which could limit the benefit from mutations, as seen *e.g.*, in CDK4/6 inhibitors.<sup>22,23</sup>

This study was supported by funding from the Swedish Research Council (grant number VR 2014-4406), and the Swedish Cancer Society (Cancerfonden, grant number CAN 2018/362). In addition, the Brazilian agency for Coordination for the Improvement of Higher Education Personnel is acknowledged for a scholarship to MJDR (CAPES Finance Code 001), and the National Council for Scientific and Technological Development (CNPq) is acknowledged for grants 312210/2019-1, 433957/2018-2, and 406273/2015-4 to IC, for a fellowship 303207/2017-5 to JJS, and for a scholarship 141923/2019-9 to TSF. The calculations were performed on resources provided by the Swedish National Infrastructure for Computing (project grant SNIC 2019/3-29) at Lunarc (Lund University) and at PDC (KTH, Stockholm).

## Conflicts of interest

There are no conflicts to declare.

## Notes and references

$\ddagger$  Calculated as  $\Delta\Delta G_{wt \rightarrow mut}^b \approx RT \ln(IC_{50}^{mut}/IC_{50}^{wt})$  using values averaged from ref. 9–12.

$\S$  Calculated from the  $K_d$  value reported in ref. 13.



¶  $\Delta E^b \approx \Delta H^b$  since the change in the volume is negligible.

|| We note, however, that the standard deviation in the calculation of the entropy is higher than the difference in  $S_{\text{conf}}$  between the wt and mutants.

- 1 R. Friedman, *Oncotarget*, 2016, 7, 11746–11755.
- 2 N. Lesgidou, E. Eliopoulos, G. N. Goulielmos and M. Vlassi, *Bioinformatics*, 2018, 34, i781–i786.
- 3 P. S. Georgoulia, G. Todde, S. Bjelic and R. Friedman, *Biochim. Biophys. Acta, Gen. Subj.*, 2019, 1863, 732–741.
- 4 R. Friedman, *Proteins*, 2017, 85, 2143–2152.
- 5 P. S. Georgoulia, S. Bjelic and R. Friedman, *FEBS J.*, 2020, DOI: 10.1111/febs.15209.
- 6 M. A. Strauss and H. A. Wegner, *Eur. J. Org. Chem.*, 2019, 295–302.
- 7 E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen and W. Yang, *J. Am. Chem. Soc.*, 2010, 132, 6498–6506.
- 8 J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan and W. Yang, *J. Chem. Theory Comput.*, 2011, 7, 625–632.
- 9 H. A. Bradeen, C. A. Eide, T. O'Hare, K. J. Johnson, S. G. Willis, F. Y. Lee, B. J. Druker and M. W. Deininger, *Blood*, 2006, 108, 2332–2338.
- 10 M. S. Zabriskie, C. A. Eide, S. K. Tantravahi, N. A. Vellore, J. Estrada, F. E. Nicolini, H. J. Khoury, R. A. Larson, M. Konopleva, J. E. Cortes, H. Kantarjian, E. J. Jabbour, S. M. Kornblau, J. H. Lipton, D. Rea, L. Stenke, G. Barbany, T. Lange, J. C. Hernandez-Boluda, G. J. Ossenkoppele, R. D. Press, C. Chuah, S. L. Goldberg, M. Wetzler, F. X. Mahon, G. Etienne, M. Baccarani, S. Soverini, G. Rosti, P. Rousset, R. Friedman, M. Deininger, K. R. Reynolds, W. L. Heaton, A. M. Eiring, A. D. Pomietter, J. S. Khorashad, T. W. Kelley, R. Baron, B. J. Druker, M. W. Deininger and T. O'Hare, *Cancer Cell*, 2014, 26, 428–442.
- 11 S. Redaelli, L. Mologni, R. Rostagno, R. Piazza, V. Magistroni, M. Ceccon, M. Viltadi, D. Flynn and C. Gambacorti-Passerini, *Am. J. Hematol.*, 2012, 87, E125–E128.
- 12 F. X. Gruber, T. Ernst, K. Porkka, R. A. Engh, I. Mikkola, J. Maier, T. Lange and A. Hochhaus, *Leukemia*, 2011, 26, 172–177.
- 13 L. M. Wodicka, P. Ciceri, M. I. Davis, J. P. Hunt, M. Floyd, S. Salerno, X. H. Hua, J. M. Ford, R. C. Armstrong, P. P. Zarrinkar and D. K. Treiber, *Chem. Biol.*, 2010, 17, 1241–1249.
- 14 A. Fouda and U. Ryde, *J. Chem. Theory Comput.*, 2016, 12, 5667–5679.
- 15 C. A. Menéndez, S. R. Accordino, D. C. Gerbino and G. A. Appignanesi, *PLoS One*, 2016, 11, e0165767.
- 16 S. Maganhi Hernandez, P. Jensen, I. Caracelli, J. Zukerman Schpector, S. Fröhling and R. Friedman, *Protein Sci.*, 2017, 26, 870–879.
- 17 I. Andricioaei and M. Karplus, *J. Chem. Phys.*, 2001, 115, 6289–6292.
- 18 Y. Ishida, K. Murai, K. Yamaguchi, T. Miyagishima, M. Shindo, K. Ogawa, T. Nagashima, S. Sato, R. Watanabe, S. Yamamoto, T. Hirose, S. Saitou, M. Yonezumi, T. Kondo, Y. Kato, N. Mochizuki, K. Ohno, S. Kishino, K. Kubo, T. Oyake and S. Ito, *Eur. J. Clin. Pharmacol.*, 2016, 72, 185–193.
- 19 C.-C. Lee, Y.-C. Chuang, Y.-L. Liu and C.-N. Yang, *RSC Adv.*, 2017, 7, 29871–29881.
- 20 P. Mishra and S. Günther, *Sci. Rep.*, 2018, 8, 9435.
- 21 C. H. Yun, K. E. Mengwasser, A. V. Toms, M. S. Woo, H. Greulich, K. K. Wong, M. Meyerson and M. J. Eck, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, 105, 2070–2075.
- 22 M. Poratti and G. Marzaro, *Eur. J. Med. Chem.*, 2019, 172, 143–153.
- 23 R. Friedman, *Isr. J. Chem.*, 2020, DOI: 10.1002/ijch.202000015.

# ESI for: Is breaking of a hydrogen bond enough to lead to drug resistance?

María José Dávila-Rodríguez<sup>1</sup>, Thales Souza Freire<sup>2</sup>, Erik Lindahl<sup>3</sup>, Ignez Caracelli<sup>2</sup>, Julio Zukerman-Schpector<sup>1</sup> and Ran Friedman<sup>3\*</sup>

## About

This is the Electronic Supplementary Information

<sup>1</sup>Federal University of São Carlos, Department of Chemistry, São Carlos-SP, Brazil

<sup>2</sup>Federal University of São Carlos, Department of Physics, São Carlos-SP, Brazil

<sup>3</sup>Linnæus University, Department of Chemistry and Biomedical Sciences, Kalmar, Sweden

\*Corresponding author

## Contents

<b>1</b>	<b>Methods</b>	<b>1</b>
1.1	Quantum chemistry . . . . .	1
	Models • Binding free energy • Hydrogen bond energy	
1.2	Molecular dynamics simulations of Abl1 with dasatinib . . . . .	2
	Model preparation • Simulation protocol • CPU usage • Dasatinib parametrisation • Analysis of the simulations	
1.3	Free energy perturbation . . . . .	2
	Theory • Software and simulation details • Preparation Stage • Energy minimisation and equilibration • Transitions • CPU usage	
1.4	Non-covalent interaction plots . . . . .	4
<b>2</b>	<b>Supplemental Tables</b>	<b>5</b>
<b>3</b>	<b>Supplemental Figures</b>	<b>6</b>
	<b>References</b>	<b>12</b>

## 1. Methods

### 1.1 Quantum chemistry

All QM calculations were performed with NWChem, version 6.8.1 [1].

#### 1.1.1 Models

The free energy of binding of dasatinib to Abl1 was estimated based on the models presented in Figure S2. The initial coordinates were taken from the crystal structure (PDB id: 2GQG [2]). To make fully-QM calculations feasible, only the part of the protein which is closest to dasatinib is included. Ideally, such a calculation should capture the most important interactions, although it may underestimate the binding free energy since many other interactions are missing. The success of such a model depend on its ability to capture the most important interactions. It is clear that many weak protein-drug interactions are not included, which to some degree is compensated in that the conformation of the complex is strained

since the binding residues cannot freely move. Since the X-ray structure was solved for the complex, the stability of the complex is expected to be higher than that of the free protein and drug together. In principle, it is possible to remove this strain by optimising the protein and drug separately. However, especially for the protein this is likely to introduced artefacts since it is not possible to consider all of the protein residues in a fully QM system studied at a high level of accuracy due to computational cost. In addition, in density functional theory (DFT) the interaction of an electron with the entire electron density is included in the coulomb term, leading to the so-called self interaction error (SIE). The exchange-correlation functional should remove this error. Most functionals do not remove all of the SIE [3], which can become severe for large systems [4]. Together, these two effect can paradoxically make a smaller model more accurate than a larger one, as long as such a model indeed captures the essential protein-ligand interactions.

To examine the effect of enlarging the model, the free energy of binding is calculated for three models (Figure S2). The contribution of the hydrogen bond was considered only with the smallest one.

#### 1.1.2 Binding free energy

The structure was optimised in gas phase with the def2-SV(P) basis set [5] and M06 functional [6]. The free energy of binding  $\Delta G^b$  was then approximated as:

$$\Delta G^b = G_{\text{complex}} - G_{\text{dasatinib}} - G_{\text{protein}} \quad (1)$$

Each of the free energies in Equation 1 was calculated with the def2-TZVP basis set. The solvent (water) was approximated by the SMD solvation model [7]. The M06 functional (hybrid meta-GGA, 27% HF exchange) was used. We report the results in Table S1.

#### 1.1.3 Hydrogen bond energy

The contribution of the hydrogen bond between fragment A and dasatinib to  $\Delta G^b$  was estimated in the following way.

The distance between fragment A and dasatinib is varied in steps of 0.1 Å. For each distance the interaction energy was calculated with def2-TZVP/M06 and the SMD solvent model. The hydrogen bond energy was then be estimated by

$$\Delta G_{\text{HB}} = \Delta G_{(d_1)}^b - \Delta G_{(d_2)}^b \quad (2)$$

with  $d_1 = 2 \text{ \AA}$  and  $d_2 = 4.5 \text{ \AA}$ .

## 1.2 Molecular dynamics simulations of Abl1 with dasatinib

MD simulations were carried out using GROMACS [8, 9, 10], version 2018.2. Bonds involving hydrogen atoms were constrained by the LINCS algorithm [11], except for the water molecules for which the SETTLE algorithm [12] was used. The CHARMM36 [13] forcefield was used for the protein. Mutations were introduced with PyMOL, version 1.8 (Schrödinger, LLC).

### 1.2.1 Model preparation

The structure of the kinase domain of Abl1 bound to dasatinib [2] was downloaded from the protein data bank (PDB id: 2GQG, resolution 2.4 Å). Residue pTyr<sup>393</sup> (phosphorylated tyrosine) was modelled as tyrosine. All crystallographic waters were removed, to avoid potential bias in the simulations of the mutant proteins or due to the removal of the phosphate group, and as none of the waters is involved in the protein-drug interactions. The complexes of Abl1(T315I), Abl1(V299L), and Abl1(F317L) with dasatinib were modelled by mutating the corresponding residues in PyMol.

### 1.2.2 Simulation protocol

The Abl1-dasatinib complex was put in a cubic box extending to at least 1.2 nm from each edge of the protein and solvated by approximately 25000 pre-equilibrated TIP3P water molecules [14]. Potassium and chlorine ions were used to neutralise the system and model a concentration of 0.15 mol L<sup>-1</sup> to mimic experimental ionic strength conditions. Each system was energy minimised until the maximal force on individual atoms was smaller than 100 KJ mol<sup>-1</sup> nm<sup>-1</sup>. Thereafter, a short (20 ps) run was performed with positional restraints on all protein and drug heavy atoms. 40 replicas for each system (Abl1 wt kinase domain, V299L, T315I, F317L) were generated and equilibrated for 5 ns without constraints which led to 40 independent structures in each case. This was followed by 50 ns long production NpT. The temperature (300 K) was kept constant using the velocity rescaling thermostat [15] and the pressure (1 bar) controlled with Berendsen's barostat [16] during position restraint and equilibration simulations, and the Parrinello-Rahman algorithm [17] during production runs. The simulation timestep was 2 fs with van der Waals forces truncated at 1 nm with a plain cutoff and long-range electrostatics treated with the PME method [18, 19]. For wt, V299L and T315I one simulation crashed due to hardware failure and the analysis was based on 39 simulations.

### 1.2.3 CPU usage

The simulations were carried on the Swedish National Infrastructure for Computations (SNIC) PDC system. For each system, production runs were carried on a system employing an Intel Xeon E5-2698 v3 2.3GHz 16-core - 32-thread processors (ensemble parallelisation, 40 processors in total) which took approximately 49 hours for 50 ns.

### 1.2.4 Dasatinib parametrization

Forcefield (CGenFF [20]) parameters for dasatinib were generated using CHARMM-GUI [21, 22]. The parameters were adjusted to ensure planarity of the tiazol and pyrimidine rings with the amine linker. Dasatinib is a base, and it was modelled as a cation with the charge on its distal piperazine nitrogen. Trial simulations were carried out with dasatinib in water to make sure that the compound is structurally stable and does not show any unexpected features. In these simulations, dasatinib's radius of gyration was almost constant ( $R_g=0.59\pm0.01 \text{ nm}$ ) and it was fully solvated forming on average six hydrogen bonds with the water.

### 1.2.5 Analysis of the simulations

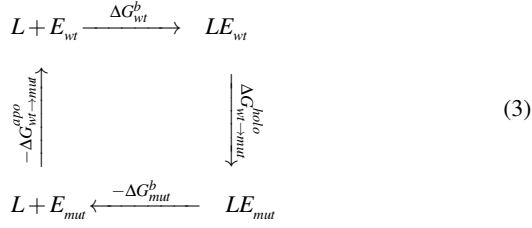
Analysis programs available in GROMACS were used with the default parameters unless otherwise indicated. The number of short contacts between the kinase domain and dasatinib (non-hydrogen atoms) was calculated within a cutoff of 0.4 nm. The free energy landscapes were built using *gmx distance* and *gmx sham* (20 bins per dimension). Interaction energies between dasatinib and the kinase domain correspond to the sum of the average short-range Coulombic interaction energies and the average short-range Lennard-Jones energies obtained with *gmx energy*. The absolute entropy for the bound dasatinib (non-hydrogen atoms) was estimated with the quasi-harmonic approximation [23] after a covariance analysis over all the simulated time of each trajectory using *gmx covar* and *gmx anaeig*. Determination of the configurational entropy of the kinase domain (non-hydrogen atoms residues 240 - 495) was performed as follows. First, cluster analysis was performed employing the algorithm developed by Daura and co-workers [24] with a cutoff of 0.15 nm. The covariance analysis for each trajectory was performed over a patch of five consecutive nanoseconds where the kinase domain was found to be in the same cluster. The entropy was determined based on this covariance analysis according to the quasiharmonic approximation using *gmx anaeig*.

## 1.3 Free energy perturbation

### 1.3.1 Theory

Consider the binding of a ligand (L) to an enzyme (E), which can undergo amino acid mutations. To evaluate the binding free energy difference upon this mutation ( $\Delta\Delta G^b$ ) we can

construct a thermodynamic cycle as shown at Eq. 3:



where the *wt* subscript stands for wild type and *mut* for mutated.

As this is a closed cycle it is possible to calculate the desired quantity in two different manners:

$$\Delta \Delta G^b = \Delta G_{mut}^b - \Delta G_{wt}^b \quad (4)$$

$$\Delta \Delta G^b = \Delta G_{wt \rightarrow mut}^{holo} - \Delta G_{wt \rightarrow mut}^{apo} \quad (5)$$

Each approach will need a different simulation setup. For this study the second method (Eq. 5) was used along with Non-equilibrium Statistical Mechanics (NESM) techniques, which required the equilibrium sampling for each state of the enzyme and an ensemble of short transitions between these states in the forward and reverse directions. For each transition the work done by the morphing of one amino acid into another was calculated as:

$$W = \int_0^1 \frac{\partial \mathcal{H}}{\partial \lambda} d\lambda \quad (6)$$

where the Hamiltonian depends on  $\lambda$  as

$$\mathcal{H}(\lambda) = (1 - \lambda)\mathcal{H}_{wt} + \lambda\mathcal{H}_{mut}, \quad (7)$$

with  $\lambda(t=0) = 0$  and  $\lambda(t=\tau) = 1$ .  $\lambda$  was increased linearly from 0 to 1 at each time step to perform the complete transition from one equilibrium state to another.

With this work ensemble in hand it is possible to use several methods derived from NESM [25, 26, 27] to calculate  $\Delta G_{wt \rightarrow mut}^{apo}$  and  $\Delta G_{wt \rightarrow mut}^{holo}$ . For accuracy reasons the Bennet's Acceptance Ratio (BAR) was chosen, as it has an analytical error estimation and a less biased way to calculate the free energy (Eq. 8) than other methods with exponential averaging [28]. Thus, the free energy upon mutation was calculated by solving the following equation:

$$\sum_{i=1}^{N_f} \frac{1}{1 + \frac{N_f}{N_r} e^{\beta(W_i - \Delta G)}} = \sum_{j=1}^{N_r} \frac{1}{1 + \frac{N_r}{N_f} e^{-\beta(W_j - \Delta G)}} \quad (8)$$

where  $N_f$  and  $N_r$  are the number of transitions in the forward and reverse direction and  $\beta = 1/k_B T$

The convergence of the free energy calculated for the full set of trajectories is determined by the error expression [28]:

$$\frac{1}{\beta^2 N_{tot}} \left[ \left\langle \frac{1}{2 + 2 \cosh[\beta(W_i - \Delta G)]} \right\rangle^{-1} - \left( \frac{N_{tot}}{N_f} + \frac{N_{tot}}{N_r} \right) \right] \quad (9)$$

where the  $\langle \rangle$  denotes the average over all work values.

### 1.3.2 Software and simulation details

MD simulations were performed with GROMACS [8, 9, 10], version 2019.2. The CHARMM36 forcefield [13] was used, together with the mutation library derived with the *pmx* program [29]. The TIP3P model [14] was used for water. Bond lengths were constrained by use of the LINCS algorithm [11], except for water molecules, where SETTLE [12] was used instead.

### 1.3.3 Preparation Stage

The structure of Abl1 bound to dasatinib [2] was downloaded from the protein data bank (PDB id: 2GQG). Crystallographic waters were removed. The GROMACS module *pdb2gmx* was used to build the topology for the kinase domain of the protein. Three mutations (V299L, T315I and F317L) were generated with the *pmx mutate* module of the *pmx* software [29]. Hybrid topologies, containing information about the wt and the mutated state of the enzyme, were built from these three pdb files, with the *pmx gentop* module. The system was simulated as dodecahedron box, with the enzyme at the centre at 1.2 nm from each side of the box. Then, the system was solvated and  $K^+$  and  $Cl^-$  ions were used to neutralise the charges and reach the concentration of 0.15 mM.

### 1.3.4 Energy minimisation and equilibration

For each mutation a single energy minimisation was performed using a steepest descent algorithm followed by a short MD simulation of 20 ps, where positional restraint were imposed on all protein atoms in order to equilibrate the water molecules around the protein. Next, the system was equilibrated for 5 ns after removing the restraints. The equilibration was repeated ten times for each state of the enzyme to ensure better accuracy through sampling. During the simulations, the Berendsen barostat [16] and velocity rescaling thermostat [15] were used to keep the pressure and temperature constant at 1 bar (with  $\tau_p = 1$  ps) and 300 K (with  $\tau_r = 0.1$  ps), respectively

### 1.3.5 Transitions

One hundred snapshots were extracted from each equilibration trajectory. From each of this snapshots, a single short MD simulation of 80 ps was spawned, starting from the wt (forward transition) and from the mutant (reverse transition). This resulted in simulations of 100 transitions in each direction. The Parrinello-Rahman barostat [17] was used for these

simulations, and velocities were generated at each transition run. The Gromacs function *gmx mdrun -dhdl* was used to calculate the change in the Hamiltonian for each simulation step, where the integral along the whole path is the work done at that specific run (Eq 6). Finally, the average free energy for each mutation was obtained with Bennett Acceptance Ratio [27]) using the *pmx analyse* module over all values and averaging over the ten simulations.

### 1.3.6 CPU usage

The simulations were carried on the SNIC Lunarc system. Each run (equilibration of each state plus 100 forward and 100 reverse transitions) required 45 hours on 10 nodes, each equipped with 2 Intel Xeon E5-2650 v3 (2.3 GHz, 10-core).

## 1.4 Non-covalent interaction plots

Non-covalent interaction plot (NCIplot) analysis was used in the present study in order to confirm the nature of some of the intermolecular contacts. This method is based on the electron density and its derivatives allowing the visualisation of density gradient isosurfaces. The isosurfaces correspond to the values of  $\text{sign}(\lambda_2)\rho$ , where  $\rho$  is the electron density and  $\lambda_2$  is the second eigenvalue of the Hessian matrix of  $\rho$  [30]. This analysis provides a convenient visualisation of the strength of any existing non-covalent interactions through a colour scheme [31] on the isosurface. Here, the viridis colour palette was used, where the gradient colour is changed gradually from purple to blue, green and finally yellow. Thus, purple tones indicate strong attractive interactions (the value of  $\text{sign}(\lambda_2)\rho$  at low-density gradients is negative). Green surfaces indicate weak interactions (the value of  $\text{sign}(\lambda_2)\rho$  at low-density gradients is close to zero), whereas yellow coloured surfaces are indicate of strong repulsive interactions (the value of  $\text{sign}(\lambda_2)\rho$  at low-density gradients is positive).

## 2. Supplemental Tables

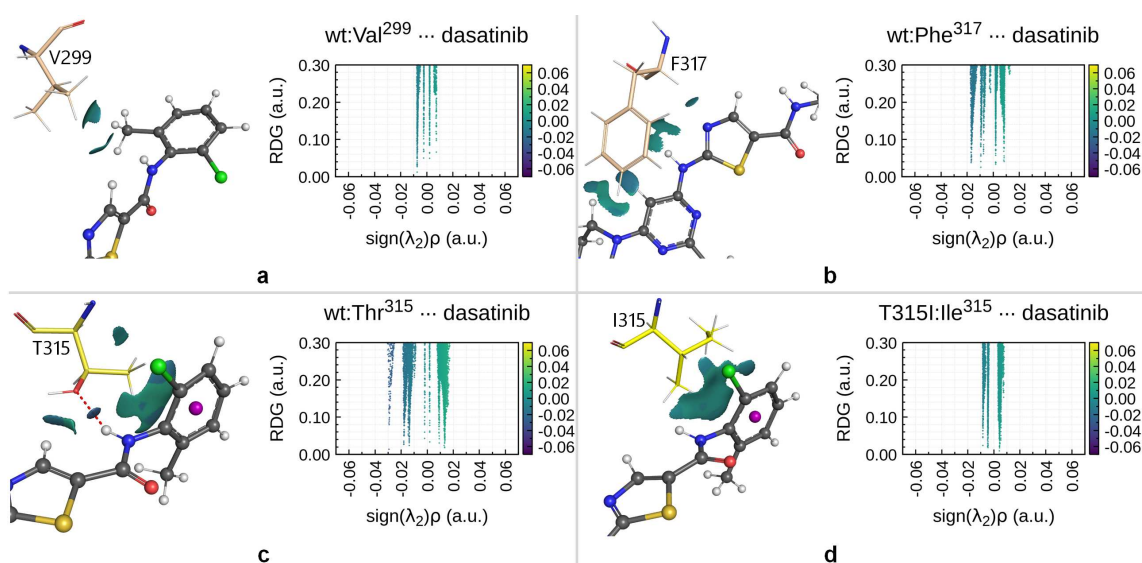
**Table S1.** The free energy of binding ( $\text{kcal mol}^{-1}$ ) between dasatinib and Abl1 calculated with full QM using three different models (Figure S2) with the def2-TVZP basis set. The rightmost column is the signed error (positive values indicate overbinding by the calculation).

	$\Delta G^b$	error
model 1, 4 residues	-13.3	-1.2
model 2, 5 residues	-15.8	+1.4
model 3, 8 residues	-23.6	+9.2

**Table S2.** Interaction energies between dasatinib and specific Abl1 residues ( $\text{kcal mol}^{-1}$ ). Numbers are averages calculated over 39 (wt, T315I, F317L) or 40 (V299L) separate 50 ns long simulations for each system, with standard deviation (last significant digit) in parentheses.

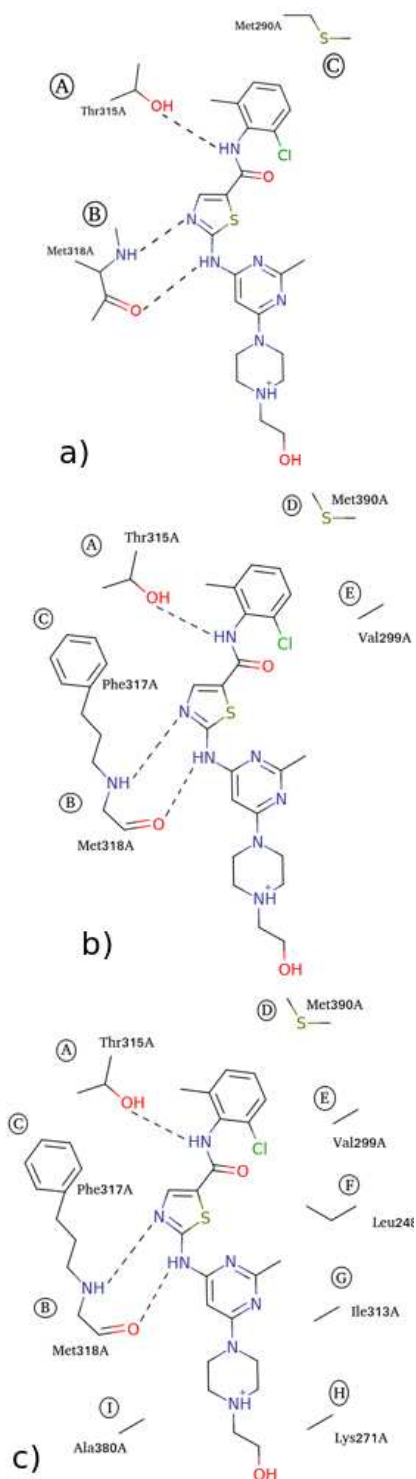
Res.	wt	V299L	T315I	F317L
299	V -1.3(02)	L -1.5(02)	V -0.6(04)	V -1.3(02)
315	T -6.7(07)	T -5.9(2)	I -2.6(1)	T -6.7(06)
317	F -5.5(04)	F -5.5(06)	F -5.9(04)	L -3.4(02)
318	M -11.7(04)	M -11.6(04)	M -11.7(03)	M -12.0(04)
All	-72.4(7)	-70.3(8)	-61.7(6)	-72.8(7)
$\Delta\Delta E^b$		2.1	10.7	-0.3

### 3. Supplemental Figures

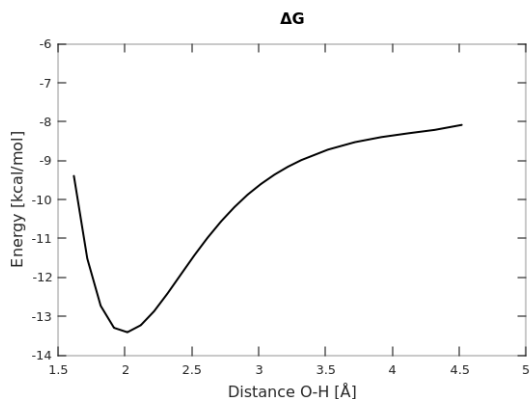


**Figure S1.** NCI analysis. Isosurfaces representing non-covalent interactions are presented and coloured according to  $\text{sign}(\lambda_2)\rho$  on the left side of which panel, whereas plots of reduced density gradient, RDG, versus  $\text{sign}(\lambda_2)\rho(r)$  are shown on the right. NCI plots are shown for dasatinib and (a) Val<sup>299</sup> (b) Phe<sup>317</sup> (c) Thr<sup>315</sup>, (d) Ile<sup>315</sup> in the T315I mutant. NCI plots are calculated based on an equilibrated structure from an MD simulation in each case. Protein figures were created with PyMOL.

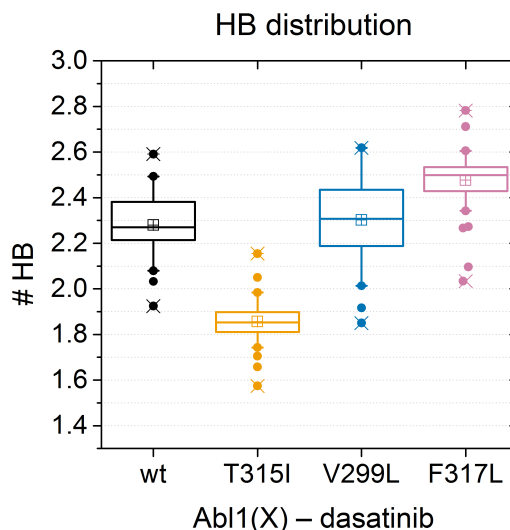




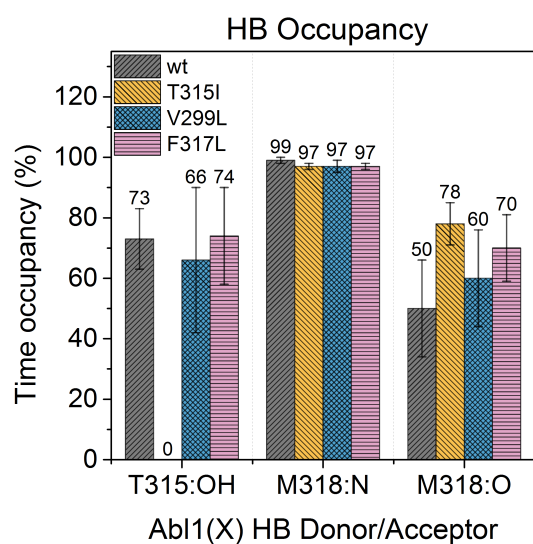
**Figure S2.** Models used to calculate the free energy of interaction between dasatinib and Abl1. Dotted lines indicate hydrogen bonds. a) The original (minimal) model, model 1. b) Model 2. c) Model 3.



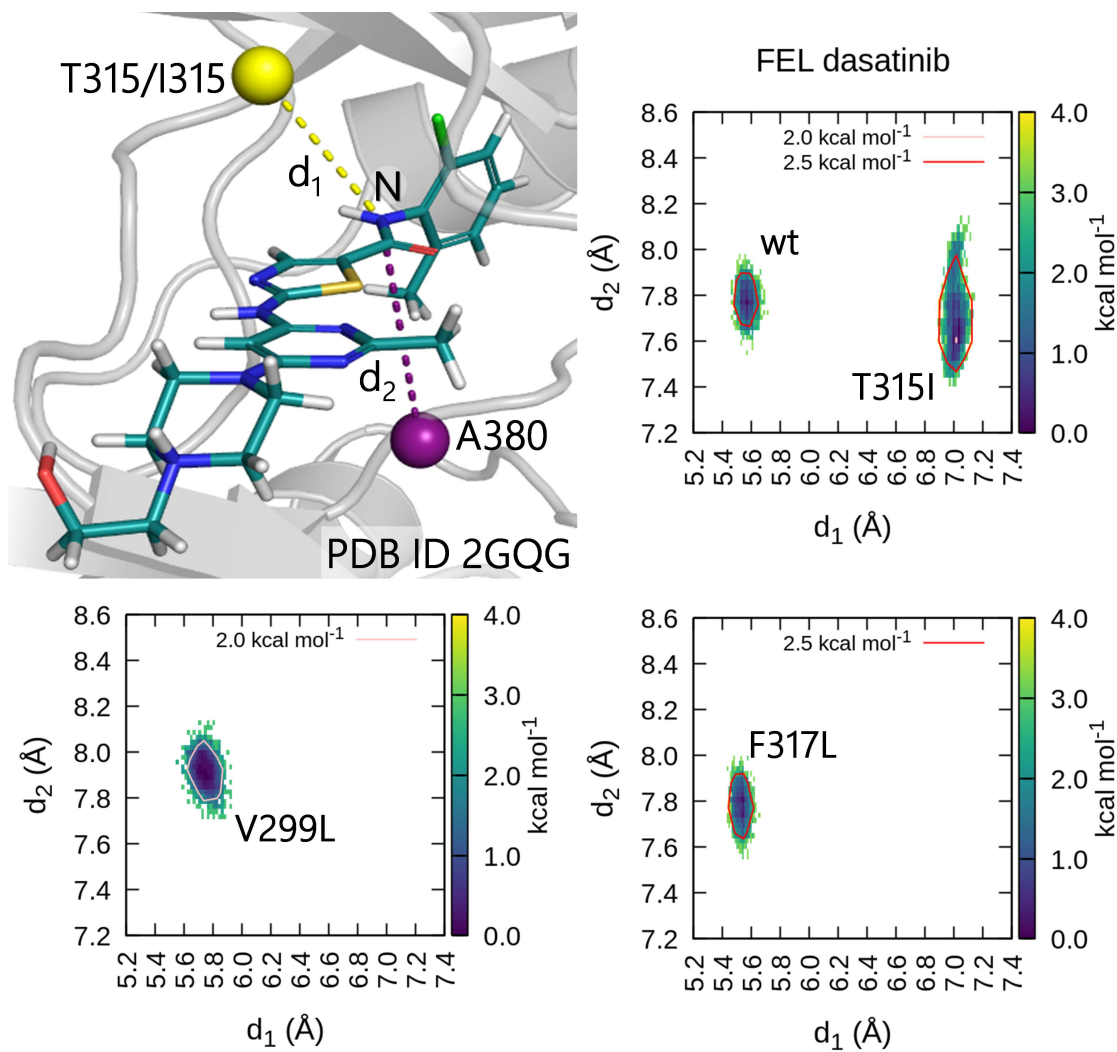
**Figure S3.** Interaction free energy for dasatinib's binding to Abl1, as a function of the T315:OH-NH:dasatinib hydrogen bond distance.



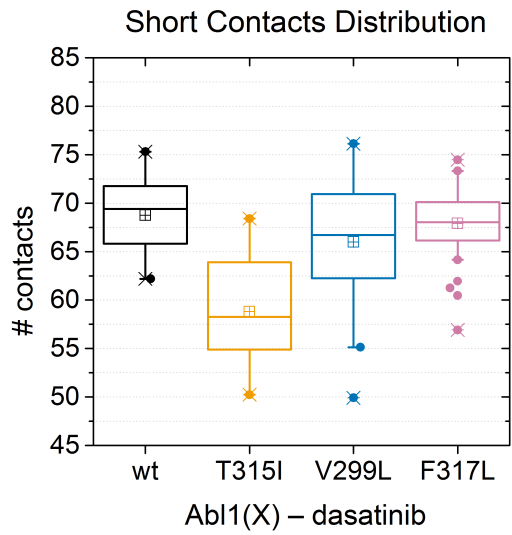
**Figure S4.** Distribution of the hydrogen bonds between Abl1 and dasatinib for the wt and mutants (box plot).



**Figure S5.** Hydrogen bond occupancies during the simulations.



**Figure S6.** Two dimensional free energy landscapes calculated as a function of the two distances  $d_1$  and  $d_2$  monitored for the four different simulations.  $d_1$  is the distance to the centre of mass (COM) of the backbone of the gatekeeper residue.  $d_2$  is the distance to the COM of the backbone of residue Ala<sup>380</sup> which precedes the DFG motif and activation loop.



**Figure S7.** Distribution of the number of contacts ( $d < 4\text{\AA}$ ) between Abl1 and dasatinib for the wt and mutants (box plot).

## References

- [1] M. Valiev, E.J. Bylaska, N. Govind, K. Kowalski, T.P. Straatsma, H.J.J. Van Dam, D. Wang, J. Nieplocha, E. Apra, T.L. Windus, and W.A. de Jong. Nwchem: A comprehensive and scalable open-source solution for large scale molecular simulations. *Comput. Phys. Commun.*, 181(9):1477–1489, 2010.
- [2] J S Tokarski, J A Newitt, C Y Chang, J D Cheng, M Wittekind, S E Kiefer, K Kish, F Y Lee, R Borzillerri, L J Lombardo, D Xie, Y Zhang, and H E Klei. The structure of Dasatinib (BMS-354825) bound to activated ABL kinase domain elucidates its inhibitory activity against imatinib-resistant ABL mutants. *Cancer Res*, 66:5790–5797, 2006.
- [3] J. L. Bao, L. Gagliardi, and D. G. Truhlar. Self-Interaction Error in Density Functional Theory: An Appraisal. *J Phys Chem Lett*, 9(9):2353–2358, May 2018.
- [4] Adam Fouda and Ulf Ryde. Does the dft self-interaction error affect energies calculated in proteins with large qm systems? *J. Chem. Theor. Comput.*, 12:5667–5679, 2016.
- [5] F Weigend and R Ahlrichs. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy. *Phys. Chem. Chem. Phys.*, 7:3297–3305, 2005.
- [6] Y Zhao and D G Truhlar. The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: Two new functionals and systematic testing of four m06 functionals and twelve other functionals. *Theor. Chem. Acc.*, 120:215–214, 2008.
- [7] A. V. Marenich, C. J. Cramer, and D. G. Truhlar. Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. *J Phys Chem B*, 113(18):6378–6396, May 2009.
- [8] H. J. C. Berendsen, D. van der Spoel, and R. Vandrunen. Gromacs - a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.*, 91:43–56, 1995.
- [9] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. Gromacs: Fast, flexible, and free. *J. Comput. Chem.*, 26:1701–1718, 2005.
- [10] Mark James Abraham, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindahl. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2:19–25, 2015.
- [11] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. LINCS: A linear constraint solver for molecular simulations. *J. Comp. Chem.*, 18:1463–72, 1997.
- [12] S. Miyamoto and P. A. Kollman. SETTLE: An analytical version of the SHAKE and RATTLE algorithms for rigid water models. *J. Comp. Chem.*, 13:952–962, 1992.
- [13] R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig, and A. D. Mackerell. Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone  $\phi$ ,  $\psi$  and side-chain  $\chi_1$  and  $\chi_1$  dihedral angles. *J Chem Theory Comput*, 8(9):3257–3273, Sep 2012.
- [14] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 79:926–935, 1983.
- [15] G Bussi, D Donadio, and M Parrinello. Canonical sampling through velocity rescaling. *J Chem Phys*, 126:014101, 2007.
- [16] H. J. C. Berendsen, J. P. M. Postma, A. DiNola, and J. R. Haak. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.*, 81:3684–90, 1984.
- [17] M. Parrinello and A. Rahman. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.*, 52:7182–7190, 1981.
- [18] T. Darden, D. York, and L. Pedersen. Particle mesh ewald: An N-log(N) method for ewald sums in large systems. *J. Chem. Phys.*, 98:10089–92, 1993.
- [19] Ulrich Essmann, Lalith Perera, Max L. Berkowitz, Tom Darden, Hsing Lee, and Lee G. Pedersen. A smooth particle mesh ewald method. *J. Chem. Phys.*, 103(19):8577–8593, 1995.
- [20] K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov, and A. D. Mackerell. Charmm general force field: A force field for drug-like molecules compatible with the charmm all-atom additive biological force fields. *J. Comput. Chem.*, 31(4):671–690, 2010.
- [21] S. Jo, M. Vargyas, J. Vasko-Szedlar, B. Roux, and W. Im. PBEQ-Solver for online visualization of electrostatic potential of biomolecules. *Nucleic Acids Res.*, 36(Web Server issue):W270–275, Jul 2008.
- [22] S. Kim, J. Lee, S. Jo, C. L. Brooks, H. S. Lee, and W. Im. CHARMM-GUI ligand reader and modeler for CHARMM force field generation of small molecules. *J Comput Chem*, 38(21):1879–1886, 06 2017.
- [23] I. Andricioaei and M. Karplus. On the calculation of entropy from covariance matrices of the atomic fluctuations. *J. Chem. Phys.*, 115:6289–6292, 2001.
- [24] Xavier Daura, Karl Gademann, Bernhard Jaun, Dieter Seebach, Wilfred F. van Gunsteren, and Alan E. Mark. Peptide folding: When simulation meets experiment. *Angew. Chem. Int. Ed.*, 38:236–240, 1999.

- [25] Gavin E. Crooks. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E*, 60(3):2721–2726, Jan 1999.
- [26] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690–2693, Jul 1997.
- [27] Charles H Bennett. Efficient estimation of free energy differences from monte carlo data. *J. Computat. Phys.*, 22(2):245–268, 1976.
- [28] Michael R. Shirts, Eric Bair, Giles Hooker, and Vijay S. Pande. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.*, 91(14), Feb 2003.
- [29] V. Gapsys, S. Michielssens, D. Seeliger, and B. L. de Groot. pmx: Automated protein structure and topology generation for alchemical perturbations. *J Comput Chem*, 36(5):348–354, Feb 2015.
- [30] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang. Revealing noncovalent interactions. *J. Am. Chem. Soc.*, 132(18):6498–6506, May 2010.
- [31] J. Contreras-García, E. R. Johnson, S. Keinan, R. Chaudret, J. P. Piquemal, D. N. Beratan, and W. Yang. NCI-PLOT: a program for plotting non-covalent interaction regions. *J Chem Theory Comput*, 7(3):625–632, Mar 2011.

# **Apêndice B**

## **Anexo 2**

Cite this: DOI: 10.1039/xxxxxxxxxx

# Resistance to a tyrosine kinase inhibitor mediated by changes to the conformation space of the kinase<sup>†</sup>

Thales Souza Freire <sup>a</sup>, Ignez Caracelli <sup>a</sup>, Julio Zukerman-Schpector <sup>b</sup> and Ran Friedman<sup>\*c</sup>

Received Date

Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Gilteritinib is a highly selective and effective inhibitor of the FLT3/ITD mutated protein, and is used successfully in treating acute myeloid leukaemia (AML). Unfortunately, tumour cells gradually develop resistance to gilteritinib due to mutations in the molecular drug target. The atomistic details behind this observed resistance are not clear, since the protein structure of the complex is only available in the inactive state, while the drug binds better to the active state. To overcome this limitation, we used a computer-aided approach where we docked gilteritinib to the active site of FLT3/ITD and calculated the Gibbs free energy difference between the binding energies of the parental and mutant enzymes. These calculations agreed with experimental estimations for one mutation (F691L) but not the other (D698N). To further understand how these mutations operate, we used metadynamics simulations to study the conformational landscape of the activation process. Both mutants show a lower activation energy barrier which suggests that they are more likely to adopt an active state until inhibited, making the mutant enzymes more active. This suggests that a higher efficiency of tyrosine kinases contributes to resistance not only against type 2 but also against type 1 kinase inhibitors.

## 1 Introduction

Protein kinases phosphorylate proteins and thereby mediate signal transduction. Overactivity of such kinases can lead to different pathological conditions, making them excellent molecular targets for drug development<sup>1</sup>. Specifically, receptor tyrosine kinases (RTKs) are promising targets for cancer treatment<sup>2</sup>. Protein kinases primarily adopt one of two states, active and inactive, with distinct structural features<sup>3</sup>. Most kinase inhibitors (KIs) target one of these states<sup>4</sup>. Type 1 inhibitors inhibit the active states and may also bind the inactive state albeit with lower affinity, whereas type 2 inhibitors bind the inactive state exclusively. Resistance to KIs is a significant medical challenge, and is mediated by mutations in the molecular drug target and by other mechanisms<sup>5,6</sup>. Understanding the precise mechanisms that mediate resistance is important for development of better therapies<sup>7</sup>. Furthermore, elucidating the role of specific amino acids and atomistic interactions is eminent in the case of mutations of the drug target<sup>8,9</sup>, as it can enable the development of better drugs. Computer simulations in atomic detail are instrumental to this aim, since such simulations can often shed light on details that cannot be easily obtained directly from experiments<sup>10,11</sup>. Simula-

tions using advanced sampling techniques such as free energy perturbation<sup>12</sup> (FEP), essential dynamics sampling<sup>13</sup> (EDS), and metadynamics<sup>14,15</sup> are particularly useful for such studies. Using FEP, it is possible to simulate a transition between a wild type (wt) and a mutant protein by artificially generating trajectories that smoothly transition between the two. Such simulations are referred to as alchemical, since they involve structures that are in-between the wt and mutant protein which do not represent the physical reality. Intermediate structures generated along FEP trajectories are however useful since together with the end-states they yield a computationally tractable free energy profile. EDS and metadynamics can be used to study transitions between states in proteins that cannot be sampled using equilibrium MD simulations, such as protein activation<sup>16</sup>. In EDS, the two states are studied first with MD simulations, and then a targeted simulation is started from an initial state, from where it continues smoothly if the RMSD to the target state does not increase. If the RMSD does increase, sampling is performed on a subspace of the principal components that is taken from the simulation of the target. In metadynamics, sampling is enhanced along a chosen set of collective variables (CVs) that need to be selected in such a way that the transition can be completed. Neither of these methods are guaranteed to succeed in simulating a transition that requires overcoming a free energy barrier. In EDS, this would depend on the overlap between the principal components of the initial and target state (if the overlap is not sufficient, the simulation will not converge to a small enough RMSD between the simulated protein and the target). In metadynamics, it can be challenging to select

<sup>a</sup> Federal University of São Carlos, Department of Physics, São Carlos-SP, Brazil

<sup>b</sup> Federal University of São Carlos, Department of Chemistry, São Carlos-SP, Brazil

<sup>c</sup> Linnæus University, Department of Chemistry and Biomedical Sciences, 391 82 Kalmar, Sweden. E-mail: ran.friedman@lnu.se

<sup>†</sup> Electronic Supplementary Information (ESI) available. See DOI: 10.1039/cXCP00000x/



the collective variables. Recently, we developed a method that combines EDS and metadynamics<sup>17</sup>, such that an initial EDS simulation generates a path that is further refined by metadynamics simulations, thus alleviating the difficulty with choosing the CVs (since in our experience EDS simulations often succeed in generating a smooth trajectory connecting the initial and target states). On top of providing a smooth trajectory, an additional advantage with the combination of EDS and metadynamics is that it is possible to generate free energy profiles of the transition and estimate the reaction barrier, which is not possible with EDS alone (where the potential energy is biased, but the bias is difficult to estimate).

Gilteritinib (Fig. 1) is an effective and selective inhibitor of the FLT3 RTK and is used in AML therapy since 2018. FLT3 is commonly mutated in AML by duplications in the protein, although these duplications do not modify the active site<sup>18–20</sup>. Such mutations, collectively termed FLT3/ITD (ITD stands for internal tandem duplication) make FLT3 overactive and (in most cases) also mean that the tumour becomes sensitive to FLT3 inhibitors. Clinically approved specific FLT3 inhibitors include gilteritinib (type 1: i.e., binding to the active state kinase), the highly specific inhibitor quizartinib (type 2: i.e., binding to the inactive state, currently used only in Japan), crenolanib (type 1, not typically used in AML) and sorafenib (type 2, not typically used in AML). All of these inhibitors may become inactive due to additional mutations in FLT3/ITD (resistance mutations<sup>21</sup>). Mutations in the kinase domain (D835F/V/Y and Y842C/H) activate the protein and lead to resistance against type 2 inhibitors. Patients that carry such mutations still respond to gilteritinib. Unfortunately, most if not all patients still relapse on gilteritinib therapy<sup>22</sup>. Many of these patients express a mutated form, FLT3-ITD/F691L. In an effort to identify additional mutations that might lead to drug resistance but which hitherto were not observed in patients (of note, gilteritinib is relatively new in the market and not all patients are screened for resistance mutations upon treatment failure), a cellular model was employed where Ba/F3 cells were made to depend on FLT3/ITD for growth and thereafter they were grown in the presence of gilteritinib<sup>23</sup>. Since those cells depend on the kinase for survival, growing them with gilteritinib in the medium leads to a strong evolutionary pressure to develop resistance. Several mutations were identified that could provide such resistance, with G697S and D698N leading to the highest increase in the drug concentration needed to limit the cellular growth by 50% (IC<sub>50</sub>). However, when counting the number of clones that lead to drug resistance only F691L and D698N were observed in excess when Ba/F3 cells were grown in the presence of 40 nM gilteritinib. Thus, these two mutations seem to be the most interesting from a preclinical point of view (although D698N was hitherto not observed in patients). Why these mutations lead to resistance is not clear. Unlike in other type 1 inhibitors (e.g., dasatinib<sup>24</sup>), the mutated residues here do not interact with the drug directly (Fig. 2). The residues that interact with the drug include Glu<sup>692</sup> and Cys<sup>694</sup> that are hydrogen bonded to gilteritinib through their backbone and Leu<sup>616</sup> which binds the drug through vdW interactions (within <0.4 nm of the protein atoms), with other residues further away.

The aim of this study was to fathom the reason behind resis-

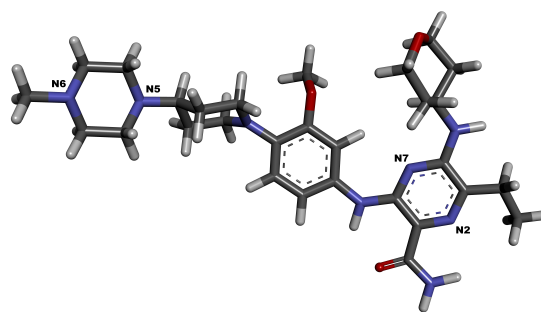


Fig. 1 The structure of gilteritinib. The piperazine moiety is basic, thus nitrogens N5 and N6 can be protonated.

tance caused by these mutations. As there is no experimental structure with FLT3 in the active state bound to gilteritinib, we first modelled the docked drug in the active state (Fig. 2). FEP was used thereafter to calculate the Gibbs energy change upon mutating for F691L and D698N. To get a more complete picture into the activation of the wt and resistant variant, the activation process was thereafter studied by a combination of EDS and well-tempered metadynamics (wT-metaD<sup>25</sup>). The results of these simulations were further used to rationalise why the mutations lead to drug resistance against gilteritinib.

## 2 Computational Methods

### 2.1 Quantum mechanical calculations of the protonated state of gilteritinib

Gilteritinib is a basic drug owing to its piperazine moiety. Typical first pK<sub>a</sub> values for piperazines are 9 or higher<sup>26</sup>. To examine whether gilteritinib has a preference for the protonation of any particular nitrogen, the structures of gilteritinib protonated at either N5 or N6 were energy minimised and their Gibbs energy difference was evaluated. This was performed using quantum mechanical (QM) calculations in implicit solvent, which for small molecules provide a favourable combination of accuracy and speed. Geometry optimisations were performed with the computer program ORCA, version 5.0.3<sup>27</sup>. The solvent (water) was modelled with the conductor-like polarisable continuum model (CPCM<sup>28</sup>). Density functional theory (DFT) was employed with the general-purpose composite method r<sup>2</sup>SCAN-3c<sup>29</sup>. The end point Gibbs energy values were used to estimate the relative stability of the two forms of gilteritinib.

### 2.2 Software, forcefield and parameters for MD simulations

MD simulations were carried out using GROMACS<sup>30–32</sup>, version 2019.4. Metadynamics simulations were carried out using the version that is patched with PLUMED<sup>14,15</sup>, version 2.5.4. The CHARMM36 force field<sup>33</sup> was used for solute atoms, and the TIP3P model<sup>34</sup> was employed for water molecules. A cutoff distance of 1.2 nm was used to compute van der Waals and Coulomb interactions. Long-range electrostatic interactions were modelled using the PME method<sup>35</sup>. The LINCS algorithm<sup>36</sup> was used to constrain bonds involving hydrogen atoms in the solutes, and the SETTLE algorithm<sup>37</sup> was employed to the same aim with wa-

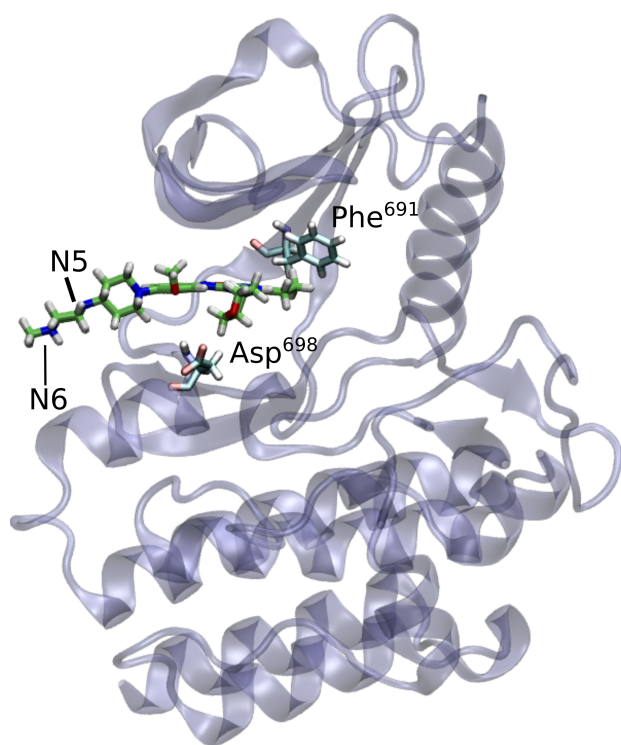


Fig. 2 Structure of FLT3 in the active state with gilteritinib. Residues Phe<sup>691</sup> and Asp<sup>698</sup>, which are mutated in resistant variants are indicated. The two nitrogens of the piperazine moiety, which are termed here N5 and N6 are also indicated. N5 is closer to the centre of mass of the protein and less solvent exposed thus it is less likely to be protonated.

ter molecules. The simulations were performed at constant temperature of  $T = 300$  K, using the velocity rescaling thermostat<sup>38</sup> ( $\tau_T = 0.1$  ps), and at constant pressure,  $P = 1$  bar. The pressure was kept constant by the Berendsen barostat<sup>39</sup> for the equilibration and by the Parrinello-Rahman barostat<sup>40</sup> for the production runs (with  $\tau_p = 1$  ps).

### 2.3 Models

The inactive state was modelled based on the crystal structure 6JQR<sup>41</sup>. For the active state, the homology modelled reported in<sup>9</sup> was used. Both structures were completed and corrected by the WHAT IF online tool<sup>42</sup>, using the sequence of the active state (which was shorter) as reference, ending up with the same number of amino acid residues and atoms. The mutated enzymes were obtained from the wild type using the conformer edit tool from the UCSF Chimera<sup>43</sup> program (v1.13.1).

### 2.4 Structure of the active state bound to gilteritinib

The active model and the X-ray inactive structure (bound to gilteritinib) were superimposed by sequence in Discovery Studio, v.2021 (BIOVIA, Dassault Systèmes). Then, the active site of both structures was selected and the superimpose function was applied again in order to improve the residue overlap in this region. After this step, the inactive enzyme structure was deleted, keeping only the active model and ligand. This was saved as a mol2 file,

to be used for a molecular docking run. Further modelling was performed by docking using the GOLD software, v.2020.1 (CCDC, Cambridge, UK). Hydrogen atoms were added first to the protein and the ligand. The binding site for the docking run was thereafter set as the region occupied by the ligand plus 0.8 nm from each atom, yielding an ellipsoid-like shape. The enzyme was kept rigid, allowing flexibility only for the ligand, with use of the default options of the program for this step. Next, the Gold Fitness scoring function was used to run the genetic docking algorithm. The output showed a cluster of structures close to the initial pose of the ligand. The best ranked pose from this cluster was chosen to be used for molecular dynamics (MD) runs.

### 2.5 Free energy perturbation

FEP was used to calculate  $\Delta\Delta G_b$  values for gilteritinib upon mutation, i.e.,  $\Delta\Delta G_{b(S \rightarrow R)}$ <sup>6</sup>, where  $S$  is the drug-sensitive variant and  $R$  is the resistant one. FEP simulations sample an alchemical trajectory which connects the initial (in this case, wt) and final (mutant) structures in order to calculate the free energy difference between the initial and final models. The pmx package<sup>44,45</sup> was used for preparation of the topologies. Mutations (F691L and D698N) were generated with the pmx mutate module. Hybrid topologies, containing information about the wt and the mutated state of the enzyme, were built from these pdb files, with the gentop module. The system was built as a dodecahedron box, placing the enzyme at the center, at a minimal distance of 1.2 nm from each edge of the box. The protein-drug complex was solvated, adding  $K^+$  and  $Cl^-$  ions to neutralise the charges and reach the concentration of 0.15 mM. For each mutation, a single energy minimisation was performed using a steepest descent algorithm followed by a short MD simulation of 20 ps, where positional restraints were imposed on all solute heavy atoms in order to equilibrate the water and ions around the protein. Next, the system was equilibrated for 10 ns after removing the restraints. The equilibration was repeated ten times for each state of the enzyme to ensure better accuracy through sampling. One hundred snapshots were extracted from the equilibration trajectories. From each of this snapshots, a single short MD simulation of 100 ps was spawned, starting from the wt (forward transition) and from the mutant (reverse transition). This resulted in simulations of 100 transitions in each direction. The Parrinello-Rahman barostat was used for these simulations, and velocities were generated at each transition run. The Gromacs command `gmx mdrun -dhd1` was used to calculate the change in the Hamiltonian for each simulation step, where the integral along the whole path corresponds to the work done at that specific run. Finally, the average Gibbs energy for each mutation was obtained with Bennett Acceptance Ratio using the pmx analyze module over all values and averaging over the ten simulations.

### 2.6 Essential dynamics sampling

EDS was used to generate the initial trajectory in which the transition from an inactive to an active state was sampled. The free energy barrier between the states is such that direct sampling, by means of unbiased MD simulations is not possible within rea-

sonable time, but EDS was shown to be feasible to use with FLT3 and yielded a reasonable trajectory<sup>46</sup>. Preparation of (apo) structures for the EDS followed the same approach as for FEP, i.e., the wt and modelled mutants were energy minimised, solvated and equilibrated in the same way. Thereafter, 50 ns long production runs were performed for each state (active and inactive). Using the trajectories of each MD simulation, cluster analysis were performed based on the C $\alpha$  atoms of the protein using Gromos method<sup>47</sup> with `gmx cluster`. A cut-off of 0.15 nm was used, which yielded results similar to choosing a cut-off of 0.2 nm for the heavy atoms. All three inactive states showed a smaller number of clusters compared to each respective active state. Subsequently, a 10 ns long trajectory, in which the protein remained in the same cluster, was used to perform a covariance analysis, using `gmx covar`. One frame dumped from each of these intervals was used as the origin (inactive state trajectory) and one as the target (active state trajectory) to build the EDS simulation input using `gmx make_edt`. The EDS algorithm<sup>13</sup> was used as implemented in Gromacs to run simulations of 1 ns for the wt and mutated ensembles. Coordinates and energies were saved every 1 ps.

## 2.7 Metadynamics

As mentioned above it is not possible to view the active-to-inactive transition in FLT3 by using MD simulations without enhancing the sampling along the transition. Simulating such processes would otherwise take a formidable amount of time. A trajectory can be prepared by EDS, but EDS in itself does not yield estimations of the Gibbs energies along the profile and hence cannot be used to estimate if there is a free energy difference between the wt and mutant when it comes to the activation of FLT3. We therefore used metadynamics simulations which provide atomistic and thermodynamic details on the process, taking advantage of the recently developed approach that combines EDS and metadynamics<sup>48</sup>. Well-tempered metadynamics (wT-metaD) simulations were performed as described in ref<sup>48</sup>. Briefly, root mean square deviations (RMSD) of C $\alpha$  atoms of the EDS trajectory with respect to the target structure were calculated first. Using the RMSD profile, frames were selected to fit the criteria of having a monotonic RMSD decrease and approximately constant RMSD between each two consecutive frames. Finally, a 1 ns wT-metaD without bias was performed to estimate the standard deviations of s(R) and z(R). s(R) describes the progress of the reaction along a predefined reference path (here, the EDS), and z(R) the distance to this path. The target structure frame was added to the path at its end. The set-up that was used for wT-metaD simulation was similar to the one used before<sup>48</sup> for a similar protein system.

The EDS simulations of the mutants could converge to an RMSD of less than 0.1 nm already using the PCA of the inactive state, whereas for the wt it was required to use that of the active state<sup>16</sup>. Using the EDS as an input, wT-metaD simulations converged rather monotonously. Using the full path file, the RMSD between the frames were calculated and  $\lambda$ , which is proportional to the average distance between neighbouring frames was estimated using the equation:

$$\lambda = \frac{2.3(N-1)}{\sum_{i=1}^{N-1} \|X_{i+1} - X_i\|^2} \quad (1)$$

All paths were built selecting only the frames with RMSD relative to their neighbours in a interval of 0.045 to 0.060 nm. Such intervals ensure that consecutive frames are not too far from each other, whereas lowering the interval further would considerably slow down the optimisation. This procedure resulted in path files with 90, 73 and 81 frames for the wt, F691L, and D698N variants, respectively. The  $\lambda$  values calculated for these paths were 885.5, 863.3 and 847.0 nm<sup>-2</sup> for the wt, F691L and D698N state, respectively. From the 1 ns runs, the RMSD for s(R) and z(R) were calculated to be used as an estimate of the Gaussian widths. The obtained values were 0.45 nm<sup>2</sup> (F691L), 0.60 nm<sup>2</sup> (D698N) and 0.50 nm<sup>2</sup> (wt), in s(R). For z(R), the RMSD were 0.005, 0.003 and 0.002 nm<sup>2</sup> for the wt, F691L, and D698N variants, respectively.

WT-metaD simulations were run with the parameters calculated above. Gaussians were deposited every 500 steps with the initial height set to 1.0 kJ/mol; the height was reduced with a bias factor of 25. Choosing the initial height was done in a way that balances convergence and accuracy. To limit the exploration of phase space with very high energies on the z(R) direction, a restraining force of 10<sup>4</sup> kJ mol<sup>-1</sup> Å<sup>-2</sup> was applied when z(R) exceeded 0.8 Å<sup>2</sup> to the wt, or 0.7 Å<sup>2</sup> to the mutants. This setup was repeated ten times for each FLT3 state. Close to 100 ns some of the runs converged to a reasonable free energy surface (FES) estimate. These simulation were continued until 200 ns. The lowest barriers were observed for FES converged at different simulation times, namely 145, 180, and 155 ns for wt, F691L and D698N, respectively. This shows that the choice of the bias was indeed reasonable with respect to sampling.

## 2.8 Structural convergence

See the supplementary information, Fig. S1 for RMSD values along the EDS and wT-metaD trajectories.

## 2.9 Representation of the protein structure

The figure showing the protein structure and the drug was generated with VMD<sup>49</sup>.

# 3 Results and Discussion

## 3.1 Calculations of binding Gibbs energy changes upon mutations account for resistance due to the gatekeeper mutation F691L

FEP is considered an accurate method to account for Gibbs energy changes upon mutations in the drug target<sup>6,24</sup> and was used here to this aim. Experimental values for  $\Delta\Delta G_b$  were estimated from fold-IC<sub>50</sub> measurements<sup>6</sup> in Ba/F3 cells that were made to depend on FLT3/ITD for growth<sup>23</sup>.

When studying inhibitors that include a piperazine ring it is important to examine different possibilities for the protonation, namely whether the proximal (N5) or distal (N6) nitrogen (Fig.1) is protonated. Here, it was assumed that the pK<sub>a</sub> differences be-

tween the proximal and distal nitrogens were very small, which was confirmed by DFT calculations showing that the difference in the Gibbs energy between the two forms was essentially zero ( $<0.1$  kcal mol<sup>-1</sup>). Gibbs binding energy differences between the native and mutant variants,  $\Delta\Delta G_{b(S\rightarrow R)}$  values were thus calculated with either N5 or N6 protonated, maintaining an overall +1 charge on the drug. We note, however, that the N6 nitrogen is more likely to be protonated since this nitrogen is more accessible to the solvent once the drug is bound (Fig. 2). We therefore refer in the text only to energy differences calculated with protonated N6; these had better agreement with the experiment.

In F691L, the calculations show that the mutant binds the drug with lower affinity than the native protein. The results of the simulations underestimated the binding energy difference, but the difference between the calculated and experimentally estimated values was less than 0.6 kcal mol<sup>-1</sup>, i.e., smaller than the thermal energy. For D698N, where the overall charge of the protein is also modified,  $\Delta\Delta G_b$  is approximately zero. There could be different factors contributing to this effect. First, the calculations might not be accurate enough, given that the elimination of charge from residue Asp<sup>698</sup> makes it more difficult to use FEP in this case. Another option is that the calculations are indeed correct, but that resistance need to be attributed to a different factor than the interactions between the *active* protein and the drug. In other words, the dynamics of the protein is perturbed to a greater degree than could be sampled by the FEP calculations. It is also possible that the fold-IC<sub>50</sub> measurements in Ba/F3 cells cannot so accurately translate to  $\Delta\Delta G_b$  values, since these values are measured in cells not with the free protein. For example, if FLT3/ITD/D698N is more efficient in catalysis, the cells carrying this mutant are more likely to survive even if the difference in protein-drug affinities is not as large as the  $\Delta\Delta G_b$  value estimated from the experiment.

Table 1 Gibbs energy differences (mutants versus wt) for the binding of gilteritinib to FLT3. All values are in kcal mol<sup>-1</sup>. The Gibbs energy differences were calculated by FEP. Experimental estimations (Exp.) are given as reference. Values in parentheses are standard deviations given at the last significant digit, e.g., 0.41(31) = 0.41 ± 0.31. N5 - protonated on N5. N6 - protonated on N6.

F691L			D698N		
N5	N6	Exp*	N5	N6	Exp*
0.41(31)	0.75(32)	1.30(12)	-0.64(44)	-0.19(35)	1.43(25)

\* Experimental values were estimated from fold-IC<sub>50</sub> measurements in Ba/F3 cells that depend on wt or mutant FLT3/ITD for growth<sup>23</sup>

### 3.2 Protein-drug interactions in the wt and mutants

In order to better understand the difference between the wt and two mutants, we have analysed the hydrogen bonds and contacts involving the protein and the drug. Surprisingly, the results (Table 2) show an increase in the number of hydrogen bonds and contacts upon mutation. This, however, is offset by a lower Gibbs energy of the mutant apo-protein upon mutation, which leads to the Gibbs energy differences shown in Table 1. The same residues are involved in hydrogen bonds in all cases, namely Glu<sup>692</sup> and Cys<sup>694</sup>. These are also the residues that are shown to bind the

drug through hydrogen bonds in the crystal structure of inactive FLT3 bound to gilteritinib<sup>41</sup>. There appears to be somewhat higher flexibility in Cys<sup>694</sup> in the wt, where the residue can bind either to nitrogen N3 or to oxygen O2 of Gilteritinib, whereas in the mutants the interaction is mostly with the oxygen.

Table 2 The number of hydrogen bonds and protein-drug contacts (non hydrogen atoms only,  $d < 0.4$  nm) as observed in MD simulations. Percent occupancy for specific hydrogen bonds with occupancy >10% of the simulation snapshots is also given in the table.

	wt	F691L	D698N
# hydrogen bonds	1.4(1.0)	2.1(8)	2.2(6)
# contacts	42(16)	51(13)	51(11)
Percent occupancy, specific hydrogen bonds:			
Gil:N3-E692:O	25	87	94
Gil:N1-C694:O	0	18	7
Gil:N3-C694:O	29	18	0
C694:N-Gil:N3	31	14	9
C694:N-Gil:O2	27	86	97

### 3.3 The activation barrier is larger for FLT3/ITD than for the resistant mutants

As mentioned above, another aspect that might contribute to resistance is the development of mutants that are better in catalysis. Preference for the active state in the mutants<sup>50</sup> is thus likely to play a role in the resistance of some mutants. In this context, it should be mentioned that although type 1 inhibitors bind the active state better, such inhibitors (including gilteritinib) can often bind the inactive state. For these reasons, we set to examine the Gibbs energy profile along the activation of the protein for the wt and mutants, using our recently developed approach<sup>48</sup>, which combines EDS<sup>13</sup> and well-tempered metadynamics (wT-metaD<sup>25</sup>).

The free energy profiles and activation energy paths along the transition coordinate  $s(R)$ <sup>51</sup> are shown in Fig. 3. In all variants, the inactive state is favoured by about 2 kcal mol<sup>-1</sup> (autophosphorylation is expected to modify the difference between the states<sup>3</sup>). Two clear minima (corresponding to the inactive and active states) are seen for all variants, with some differences in the dynamics and presence of additional shallow minima. The barrier, calculated as a function of  $s(R)$  differs between the variants. It is highest for the wt (14.1 kcal mol<sup>-1</sup>) and lower for the mutants (7.6 kcal mol<sup>-1</sup> and 9.8 kcal mol<sup>-1</sup> for F691L and D698N, respectively), which suggests that the mutants could more easily become activated. The barrier for the wt was previously estimated as 17.8 kcal mol<sup>-1</sup> based on structures taken directly from the EDS<sup>46</sup>. Apparently, refinement of the path by wT-metaD provides a lower estimation that is likely more accurate.

Activation of the protein involves changes to several structural features, mainly the activation loop and helix  $\alpha C$ <sup>46</sup>, as shown in Figure 4. The transition between inactive and active structures starts with unwinding of the activation loop, that is seen on top of the two lobes of the kinase. In the high-energy intermediate, the loop is unwound but does not adopt its final conformation. In the wt protein and F691L mutant, helix  $\alpha C$  (the only helix in the N-lobe, which is on the top of the figure) already adopts a conformation that is close to that of the active state. In the D698N

mutant, the helix is still close to its initial location in the highest energy intermediate. Otherwise, the transitions are rather similar between the wt and mutant proteins.

## 4 Conclusions

Our calculations show that the increase in binding energy upon mutation of the FLT3 kinase is similar to the value expected based on inhibition experiments in Ba/F3 cells for one mutant (F691L) but not the other (D698N). The first mutant is indicated as a source of resistance in a clinical setting, whereas the second has hitherto not been identified in patients, despite leading to a higher degree of resistance in a cellular model. The finding that the mutants have lower Gibbs activation energy barriers might translate also to other KIs, where resistance might depend on a combination of (1) modification of protein-drug interactions and (2) changes to the dynamics between active and inactive states. Our results suggest that both mutants are more likely than the wt to be activated, which may in part explain the resistance profile as seen in a cellular model. In a clinical setting, pharmacodynamics effects (absorption and distribution of the drug, binding to other proteins) as well as pharmacokinetics (metabolism and excretion) can affect the patient. This, in turn, suggests that a variant with a lower barrier between the states has an inherent benefit even if it binds the inhibitor in a similar fashion. How significant can such a variation be for type 1 kinase inhibitors is difficult to estimate. It also remains to be seen if the D698N is observed in patients. AML is a relatively rare disease and the number of patients that were treated by gilteritinib is rather small, making it difficult to conclude whether this mutation drives resistance in patients. Moreover, with gilteritinib seen as a drug of last resort, most patients do not undergo further tests for FLT3 mutations upon trial failure as there are no other inhibitors that might help<sup>7</sup>.

It is difficult to judge the accuracy of the results since experimental estimation of the activation barriers are far from trivial. Our previous results with Abl1<sup>17</sup> were in agreement with other estimates based on different theories, which give some credibility to the approach used here. Furthermore, evidence such as observation of intermediate states that resemble related protein structures<sup>16,46</sup> indicate that the trajectories along the activation path are reasonable. The FEP simulations were limited to the active state and sampling was enhanced along the alchemical path that connects the wt and mutant. Thus, any changes to the inactive or intermediate states are not sampled. This might affect the accuracy for binding of type 1 inhibitors since such inhibitors do not bind exclusively to the active state, even if they bind to it better than to the inactive one. This might be important in FLT3-binding where the inactive state is slightly more stable.

Finally, it might be interesting to consider how to utilise our finding for drug design. Lessons from other protein kinases have shown that it is possible to develop drugs that inhibit also specific mutants (pan-resistance inhibitors). It might also be possible to develop drugs that bind to intermediates along the activation pathway<sup>46</sup>. Such drugs inhibit the protein and are also expected to prevent its activation.

## Author Contributions

Thales Souza Freire: conceptualisation, methodology, software, validation, formal analysis, investigation, writing - review and editing, visualisation

Ignex Caracelli: resources, funding acquisition, supervision, project administration

Julio Zukerman-Schpector: resources, funding acquisition, supervision, project administration, writing - review and editing

Ran Friedman: conceptualisation, methodology, software, validation, formal analysis, investigation, resources, writing - original draft, writing - review and editing, visualisation, supervision, funding acquisition, project administration

## Conflicts of interest

There are no conflicts to declare

## Acknowledgements

The authors thank Dr. Baswanth Oruganti, SRM University-AP, Guntur, India, for his help in setting up the metadynamics simulations. This study was supported by funding from the Swedish Cancer Society (Cancerfonden, grant number 21 1423 Pj). In addition, and the National Council for Scientific and Technological Development (CNPq) is acknowledged for grants 312210/2019-1, 433957/2018-2, and 406273/2015-4 to IC, for a fellowship 303207/2017-5 to JZS, and for a scholarship 141923/2019-9 to TSF. The calculations were performed on resources provided by Lund University at Lunarc (project IDs lu2021-2-6 and lu2021-2-109), and PDC (KTH, Stockholm). PDC is part of the Swedish National Infrastructure for Computing (SNIC) partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

## Notes and references

- 1 M. M. Attwood, D. Fabbro, A. V. Sokolov, S. Knapp and H. B. Schiöth, *Nat. Rev. Drug Discovery*, 2021, **20**, 839–861.
- 2 P. Cohen, D. Cross and P. A. Jänne, *Nat. Rev. Drug Discovery*, 2021, **20**, 551–569.
- 3 R. Friedman, *Isr. J. Chem.*, 2020, **60**, 667–680.
- 4 C. C. Ayala-Aguilera, T. Valero, Á. Lorente-Macías, D. J. Bailache, S. Croke and A. Unciti-Broceta, *J. Med. Chem.*, 2022, **65**, 1047–1131.
- 5 H. J. G. Lindström, A. S. de Wijn and R. Friedman, *BMC Cancer*, 2019, **19**, 508.
- 6 R. Friedman, *WIREs Computational Molecular Science*, 2022, **12**, e1563.
- 7 J. Yang, H. Lindström and R. Friedman, *Cancer Cell Int*, 2021, **21**, 198.
- 8 P. S. Georgoulia, G. Todde, S. Bjelic and R. Friedman, *Biochim. Biophys. Acta Gen. Subj.*, 2019, **1863**, 732–741.
- 9 P. S. Georgoulia, S. Bjelic and R. Friedman, *FEBS J.*, 2020, **287**, 3200–3220.
- 10 R. Friedman, K. Boye and K. Flatmark, *Biochim Biophys Acta*, 2013, **1836**, 1–14.

- 11 L. Moretto, M. Ušaj, O. Matusovsky, D. E. Rassier, R. Friedman and A. Månsson, *Nat. Commun.*, 2022, **13**, 4575.
- 12 B. J. Williams-Noonan, E. Yuriev and D. K. Chalmers, *J. Med. Chem.*, 2018, **61**, 638–649.
- 13 A. Amadei, A. B. Linssen, B. L. de Groot, D. M. F. van Aalten and H. J. C. Berendsen, *J. Biomol Struct Dyn*, 1996, **13**, 615–625.
- 14 M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. A. Broglia and M. Parrinello, *Comput. Phys. Commun.*, 2009, **180**, 1961–1972.
- 15 G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
- 16 R. Friedman and A. Caflisch, *Proteins*, 2008, **73**, 814–827.
- 17 B. Oruganti and R. Friedman, *J. Chem. Theory Comput.*, 2021, **17**, 7260–7270.
- 18 G. Todde and R. Friedman, *Phys Chem Chem Phys*, 2019, **21**, 18467–18476.
- 19 G. Todde and R. Friedman, *J Chem Inf Model*, 2020, **60**, 4005–4020.
- 20 R. Friedman, *Biochim. Biophys. Acta Rev. Cancer*, 2022, **1877**, 188666.
- 21 R. Friedman, *Oncotarget*, 2016, **7**, 11746–11755.
- 22 D. R. Buelow, B. Bhatnagar, J. Y. Jeon, M. Zavorka Thomas, E. Eisenmann, J. Stromatt, S. Orwick, N. Pabla, J. S. Blachly, S. Baker and B. W. Blaser, *Blood*, 2021, **138**, 207–207.
- 23 T. C. Tarver, J. E. Hill, L. Rahmat, A. E. Perl, E. Bahceci, K. Mori and C. C. Smith, *Blood Adv*, 2020, **4**, 514–524.
- 24 M. J. Dávila-Rodríguez, T. S. Freire, E. Lindahl, I. Caracelli, J. Zukerman-Schpector and R. Friedman, *Chem Commun (Camb)*, 2020, **56**, 6727–6730.
- 25 A. Barducci, G. Bussi and M. Parrinello, *Phys. Rev. Lett.*, 2008, **100**, 020603.
- 26 F. Khalili, A. Henni and A. L. L. East, *J. Chem. Eng. Data*, 2009, **54**, 2914–2917.
- 27 F. Neese, *WIREs Comput. Mol. Sci.*, 2022, **12**, e1606.
- 28 V. Barone and M. Cossi, *J. Phys. Chem. A*, 1998, **102**, 1995–2001.
- 29 S. Grimme, A. Hansen, S. Ehlert and J.-M. Mewes, *J. Chem. Phys.*, 2021, **154**, 064103.
- 30 H. J. C. Berendsen, D. van der Spoel and R. Vandrunen, *Comput. Phys. Commun.*, 1995, **91**, 43–56.
- 31 D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, *J. Comput. Chem.*, 2005, **26**, 1701–1718.
- 32 M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, *SoftwareX*, 2015, **1–2**, 19–25.
- 33 R. B. Best, X. Zhu, J. Shim, P. E. Lopes, J. Mittal, M. Feig and A. D. Mackerell, *J Chem Theory Comput*, 2012, **8**, 3257–3273.
- 34 W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926–935.
- 35 T. Darden, D. York and L. Pedersen, *J. Chem. Phys.*, 1993, **98**, 10089–92.
- 36 B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comp. Chem.*, 1997, **18**, 1463–72.
- 37 S. Miyamoto and P. A. Kollman, *J. Comp. Chem.*, 1992, **13**, 952–962.
- 38 G. Bussi, D. Donadio and M. Parrinello, *J Chem Phys*, 2007, **126**, 014101.
- 39 H. J. C. Berendsen, J. P. M. Postma, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–90.
- 40 M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.
- 41 T. Kawase, T. Nakazawa, T. Eguchi, H. Tsuzuki, Y. Ueno, Y. Amano, T. Suzuki, M. Mori and T. Yoshida, *Oncotarget*, 2019, **10**, 6111–6123.
- 42 R. Rodriguez, G. Chinea, N. Lopez, T. Pons and G. Vriend, *Bioinformatics*, 1998, **14**, 523–528.
- 43 E. F. Pettersen, T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin, *J Comput Chem*, 2004, **25**, 1605–1612.
- 44 D. Seeliger and B. L. de Groot, *Biophys J*, 2010, **98**, 2309–2316.
- 45 V. Gapsys, S. Michielssens, D. Seeliger and B. L. de Groot, *J Comput Chem*, 2015, **36**, 348–354.
- 46 G. Todde and R. Friedman, *J Phys Chem B*, 2019, **123**, 5385–5394.
- 47 X. Daura, K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark, *Angew. Chem. Int. Ed.*, 1999, **38**, 236–240.
- 48 B. Oruganti, E. Lindahl, J. Yang, W. Amiri, R. Rahimullah and R. Friedman, *J. Biol. Chem.*, 2022, **298**, 102238.
- 49 W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
- 50 R. Friedman, *Proteins*, 2017, **85**, 2143–2152.
- 51 A. Barducci, M. Bonomi and M. Parrinello, *WIREs Comput Mol Sci*, 2011, **1**, 826–843.

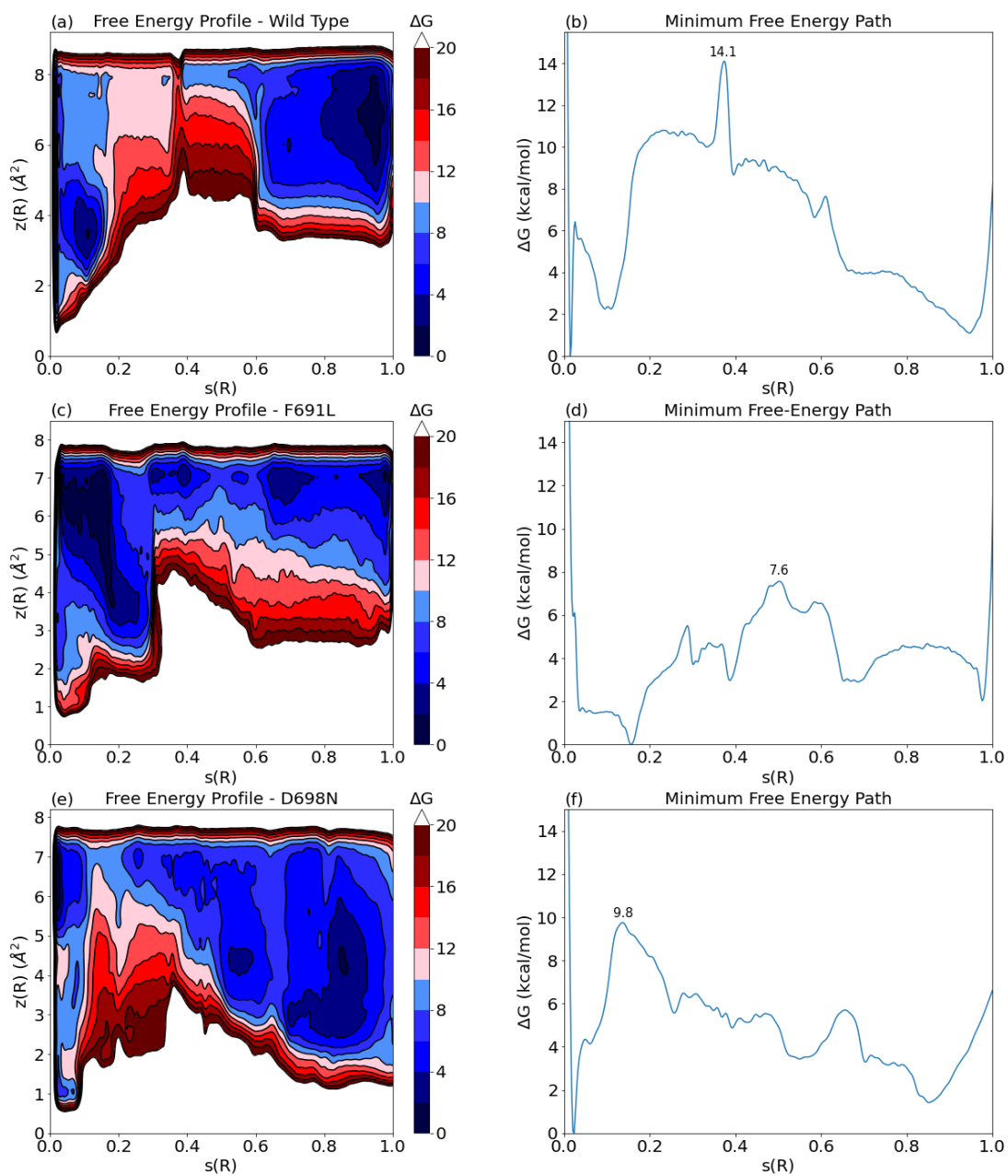


Fig. 3 Free energy profiles (a, c and e) and minimum Gibbs energy path (b, d and f) for activation of wt-FLT3 (a and b), F691L (c and d) and D698N (e and f). The energy barrier is indicated on each path.

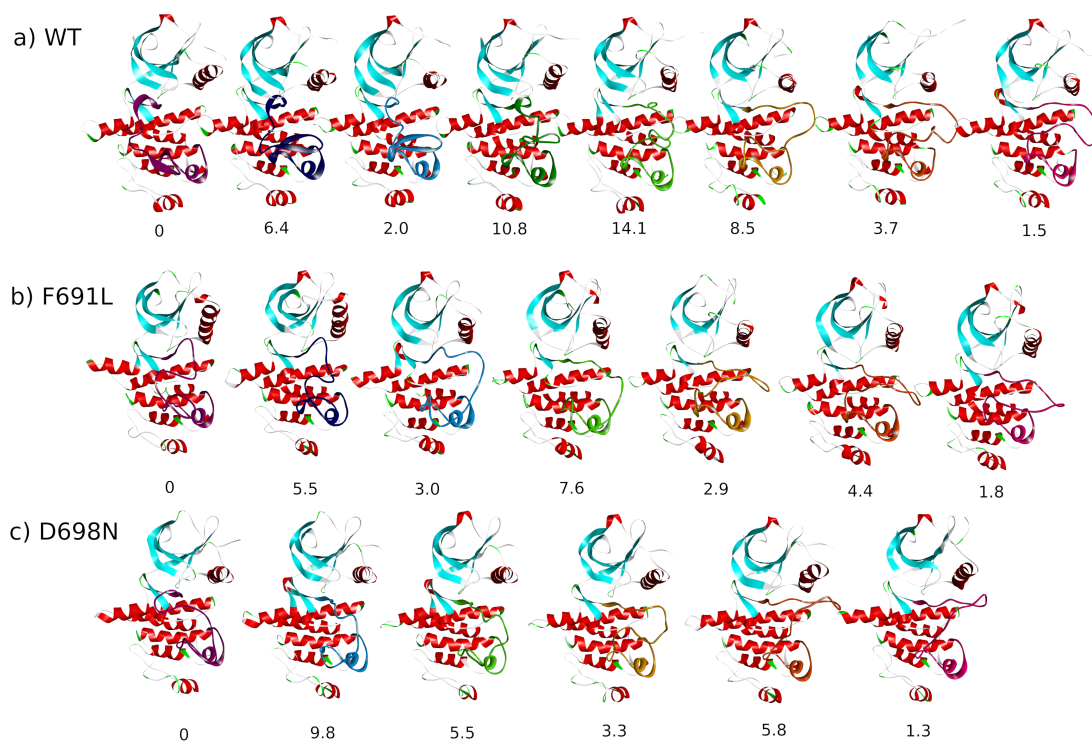


Fig. 4 Structures along the activation trajectory for wt (a), F691L (b) and D698N (c). The values below each structure correspond to the Gibbs energy difference with respect to the inactive state (see Figure 3 for the energetics).