

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ÍCARO ARIEL SIMON

**MODELAGEM MOLECULAR NO ESTUDO DAS INTERAÇÕES RECEPTOR-
LIGANTE E NO DESENHO RACIONAL DE INIBIDORES DA BIOSÍNTESE DE
PETROBACTINA EM *Bacillus anthracis*: DEIDROSHIKIMATO DESIDRATASE
COMO ALVO DE NOVAS TERAPIAS ANTI-ANTRAZ**

DISSERTAÇÃO DE MESTRADO

Porto Alegre, setembro de 2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

ÍCARO ARIEL SIMON

**MODELAGEM MOLECULAR NO ESTUDO DAS INTERAÇÕES RECEPTOR-
LIGANTE E NO DESENHO RACIONAL DE INIBIDORES DA BIOSÍNTESE DE
PETROBACTINA EM *Bacillus anthracis*: DEIDROSHIKIMATO DESIDRATASE
COMO ALVO DE NOVAS TERAPIAS ANTI-ANTRAZ**

Dissertação apresentada como requisito
parcial para a obtenção do título de Mestre em
Química.

Prof. Dr. Paulo Augusto Netz
Orientador

Porto Alegre, setembro de 2017

“Então disse o Senhor a Moisés e a Arão: Tomai vossas mãos cheias de cinza do forno, e Moisés a espalhe para o céu diante dos olhos de Faraó; E tornar-se-á em pó miúdo sobre toda a terra do Egito, e se tornará em sarna, que arrebente em úlceras, nos homens e no gado, por toda a terra do Egito.”

Êxodo 9:8,9

AGRADECIMENTOS

À Deus, que a partir da luz e da energia criou o Universo, a matéria, a química, as enzimas, a vida! Ainda que me digam que Deus está morto ou que Ele não passa de um delírio, não consigo deixar de acreditar! Assim como Albert sugeriu, quanto mais penetro nos segredos da natureza, através da nossa limitada ciência, mais percebo que por trás de cada lei e de cada fenômeno existe algo “sutil, intangível e inexplicável” e eu sei que esse algo é Tu, meu Deus! Obrigado pela força, pela graça, a inspiração e por teu amor incondicional.

Ao meu orientador, Prof. Dr. Paulo Augusto Netz, um dos homens mais inteligentes e sábios que já conheci! Obrigado pela tua paciência – muita paciência, diga-se de passagem! Obrigado por ter acreditado em mim e nesse projeto, lá em 2014, e por ter me orientado ao longo dessa caminhada no mestrado nos últimos dois anos. Obrigado por tuas ideias, sugestões, revisões de textos e por todas as conversas enriquecedoras que tivemos!

Ao Prof. Dr. Andrew T. Koppisch e sua esposa Cindy C. Browder da Northern Arizona University, pessoas inspiradoras que me acolheram em seu laboratório no Summer Break de 2014, em uma das experiências mais enriquecedoras da minha vida! E foi lá onde esse projeto começou! Thank you guys! Without you, none of these pages would ever exist! Thanks Dr. K!

Aos meus pais, Vilmar e Marlene Simon que com suas conversas me alegraram, com seus conselhos me instruíram e com seu amor me fortaleceram ao longo dessa caminhada. Obrigado pelo suporte emocional e financeiro. Amo muito vocês!

À minha família, que não entende porque tenho de estudar tanto e que se pergunta sobre quando vou começar a ganhar dinheiro! Também me pergunto sobre isso, pessoal! Obrigado pelo apoio de todos, mas em especial do meu irmão Victor Simon, da ‘Koda’ Vera, da vó Valeska e da Sandra Simon, minha parceira da corrida à cerveja e terapeuta nas horas vagas! Te amo!

Aos grandes amigos da minha vida, Marlon dos Santos, Raul Batista, Eduarda Nascimento e Roana Oliveira. É muito bom poder compartilhar tantas memórias boas e inesquecíveis com vocês! Vocês são minha inspiração!

Aos incríveis colegas e amigos que fiz nessa caminhada, em especial à Vanessa Petry, que me ajudou muito desde o começo nessa vida de químico teórico, ao meu ‘Bruder’ de mestrado Matheus Jacobs, a parceria mais ‘top’, a quem desejo todo o sucesso do mundo no doutorado que se inicia, aos demais colegas do GQT e aos ‘roomies’ Guilherme Silva, Lucas Cunha e Patrícia Freitas.

Aos professores e amigos do GQT e do Instituto de Química da UFRGS, à minha Deutschlehrerin Luana Mayer, a quem desejo muita força e saúde, à ‘profe’ Cleciane Mutti e todos os grandes mestres da minha vida, sem os quais nada disso seria possível. Afinal, se tenho visto mais longe, é só por estar sobre o ombro de gigantes!

À Capes e CNPq pelo suporte financeiro e ao Grupo de Química Teórica e o Centro Nacional de Supercomputação (CESUP/UFRGS) pelo suporte operacional.

Devo estar esquecendo de alguém (espero que não), mas se estiver, por favor considere-se abraçado! Você também foi muito importante para a realização desse trabalho!

RESUMO

O antraz é uma doença infecciosa aguda grave, com uma taxa de mortalidade superior a 90% em sua forma respiratória, causada pelo *Bacillus anthracis*, uma bactéria altamente virulenta, que está desenvolvendo resistência e que tem potencial aplicação como arma biológica e agente de bioterrorismo. Nesse trabalho, a inibição de deidroshikimato desidratase do *B. anthracis* foi estudada por meio *docking*, dinâmica molecular e *ensemble docking*. Essa enzima é responsável por uma etapa chave na biossíntese de petrobactina, molécula através da qual o *B. anthracis* adquire ferro – micronutriente essencial para seu desenvolvimento e proliferação no hospedeiro. O *docking* de 25 compostos com ação inibitória conhecida na estrutura cristalográfica da enzima indicou interações importantes com os resíduos His144, His175, Phe211, Tyr217 (ligações de hidrogênio), Arg102 (ponte salina), His144 e Phe255 (interações π - π). Ligantes estruturalmente semelhantes ao cristalográfico (3,4-DHBA) foram docados adequadamente no sítio ativo, enquanto ligantes mais volumosos foram docados na entrada do sítio, resultando em baixa correlação entre as energias livres de ligação experimentais e os escores de *docking* ($R^2 = 0,1295$; R-Pearson = 0,360) e desvios de 23%, em média, frente ao experimental. Simulações de dinâmica molecular mostraram que essa proteína apresenta uma grande rigidez estrutural intrínseca, porém porções do seu sítio ativo, sobretudo da estrutura em forma de laço que o recobre, apresentaram flexibilidade significativa. A presença de ligantes induz a alterações conformacionais que proporcionam o alargamento do sítio e permitem a entrada de ligantes mais volumosos, indicando que o sítio cristalográfico era, de fato, muito restrito. A atividade inibitória aparenta estar relacionada com a formação de uma rede de ligações de hidrogênio entre os ligantes e resíduos do sítio ativo, sendo as principais entre grupos 3-OH do anel aromático dos ligantes e a His175; entre o grupo carboxílico e a Arg102 (ponte salina); entre o grupo 4-OH e a Phe211 e principalmente entre o grupo 5-OH e a His144, um resíduo importante no mecanismo enzimático. O *ensemble docking* em três estruturas extraídas das simulações de dinâmica molecular permitiu aprimorar a correlação entre os escores de *docking* e atividade inibitória experimental, com $R^2 = 0,363$ e R-Pearson = 0,602 considerando a totalidade dos ligantes ou com $R^2 = 0,8157$ e R-Pearson = 0,903 considerando-se os dez ligantes mais potentes (contra $R^2 = 0,5683$ e R-Pearson = 0,754 na estrutura cristalográfica), evidenciando a necessidade de se considerar a flexibilidade do receptor para o *docking* adequado. Esse modelo linear juntamente com essa compreensão mais profunda dos mecanismos relacionados com a inibição dessa enzima permitirão o desenho e a triagem *in silico* de novas moléculas com potência e seletividade aprimoradas e com potencial aplicação como uma nova terapia contra o *Bacillus anthracis*.

Palavras chaves: *docking*, dinâmica molecular, *ensemble docking*, inibição enzimática, petrobactina, *Bacillus anthracis*.

ABSTRACT

Anthrax is a serious acute infectious disease with a mortality rate higher than 90% in its inhalational form. This disease is caused by *Bacillus anthracis*, a highly virulent bacterium that is developing resistance and which has potential application as a biological weapon and bioterrorism agent. In this work, the inhibition of dehydroshikimate dehydratase from *B. anthracis* was studied through *docking*, molecular dynamics and *ensemble docking*. This enzyme is responsible for a key step in the biosynthetic pathway of petrobactin, a molecule released by *B. anthracis* to acquire iron, an essential micronutrient for its development and proliferation within the host. Molecular *dockings* of 25 compounds with known inhibitory activity against dehydroshikimate dehydratase in the crystallographic structure of this enzyme indicated important interactions with the residues His144, His175, Phe211, Tyr217 (hydrogen bonds), Arg102 (salt bridge), His144 and Phe255 (π - π interactions). Ligands structurally similar to the crystallographic (3,4-DHBA) were appropriately docked within the active site, while bulkier ligands were docked at the site's entrance, resulting in a low correlation between the experimental binding free energies and the *docking* scores ($R^2 = 0,1295$; R-Pearson = 0,360), as well as a deviation of 23%, on average, compared to the experimental data. Molecular dynamics simulations showed that this protein has a high structural rigidity, however portions of its active site, especially the loop-like structure that covers it, showed a significant mobility. The presence of ligands induced conformational changes that lead to the widening of the site and allowed bulkier ligands to enter it, what indicates the crystallographic site was, in fact, very restricted. The inhibitory activity appears to be related with the formation of a network of hydrogen bonds between ligands and active site residues, mainly between the 3-OH moiety in the aromatic ring of ligands and His175; between the carboxylic group and Arg102 (salt bridge); between the 4-OH moiety and Phe211 and specially between the 5-OH group and His144, a residue with an important role in the enzymatic mechanism. *Ensemble docking* with three structures extracted from molecular dynamics simulations allowed to improve the correlation between *docking* scores and experimental inhibitory activity, with $R^2 = 0,363$ and R-Pearson = 0,602, when considering all ligands, and $R^2 = 0,8157$ and R-Pearson = 0,903 when considering the ten ligands of higher activity (against the values of $R^2 = 0,5683$ and R-Pearson = 0,754 for their *docking* in the crystallographic structure). This point out the need to account for receptor's flexibility for an appropriate *docking*. This linear model coupled with this deeper understanding about the mechanisms related with enzymatic inhibition will allow the *in silico* drug design and screening of new molecules with improved potency and selectivity and with potential application as a new therapy against *Bacillus anthracis*.

Keywords: *docking, molecular dynamics, ensemble docking, enzymatic inhibition, petrobactin, Bacillus anthracis.*

Produção científica no período do mestrado (08/2015 – 09/2017)

SIMON, ÍCARO. A.; FOX, DAVID. T. S. ; KOPPISCH, ANDREW. T. ; BROWDER, CINDY. C.; NETZ, PAULO. A. *Docking, molecular dynamics and ensemble docking in the study of petrobactin biosynthesis inhibition in Bacillus anthracis*. XIX Simpósio Brasileiro de Química Teórica. Águas de Lindóia, **2017**. (Apresentação de Trabalho/Congresso).

SIMON, ÍCARO. A.; RODIN, ALEXIS. J. ; NETZ, PAULO. A. ; FOX, DAVID. T. S. ; KOPPISCH, ANDREW. T. ; BROWDER, CINDY. C. *Development of inhibitors of petrobactin biosynthesis: dehydroshikimate dehydratase as a target for anti-anthrax therapeutics*. 249th American Chemical Society National Meeting & Exposition. Denver, **2015**. (Apresentação de Trabalho/Congresso).

LISTA DE ACRÔNIMOS E ABREVIATURAS

3-DHS	ácido 3-dehidroshikímico ou 3-dihidroshikimato
3,4-DHBA	ácido 3,4-dihidroxibenzoico ou 3,4-dihidroxibenzoato
ADT	AutoDockTools
Ala	alanina
Arg	arginina
Asn	asparagina
AsbF	deidroshikimato desidratase (grupo gênico)
Asp	ácido aspártico
COD	coeficiente de determinação
DM	dinâmica molecular
Gln	glutamina
Glu	ácido glutâmico
Gly	glicina
IC ₅₀	concentração inibitória média
Ile	isoleucina
Leu	leucina
LH	ligação de hidrogênio
Lys	lisina
Met	metionina
PDB	Protein Data Bank
Phe	fenilalanina
Pro	prolina
RMSD	deslocamento quadrático médio
RMSE	erro quadrático médio
RMSF	flutuação quadrática média
Ser	serina
Thr	treonina
Trp	triptofano
Val	valina
VdW	Van der Waals
Vina	AutoDock Vina

SUMÁRIO

1	INTRODUÇÃO	15
2	OBJETIVOS	23
2.1	OBJETIVO GERAL.....	23
2.2	OBJETIVOS ESPECIFICOS.....	23
3	FUNDAMENTAÇÃO TEÓRICA	24
3.1	PROTEÍNAS: COMPOSIÇÃO E ESTRUTURA	24
3.1.1	Aminoácidos	25
3.1.2	Organização Estrutural Proteica	28
3.1.3	Aspectos estruturais dinâmicos e enovelamento das proteínas	31
3.2	ENZIMAS	34
3.2.1	Catálise enzimática	35
3.2.2	O Sítio Ativo	37
3.2.3	Especificidade Enzima-Substrato	37
3.2.4	Mecanismos de Catalise Enzimática	39
3.2.5	Inibição Enzimática	41
3.2.6	Afinidade Enzima-Ligante e Energia Livre	43
3.2.6.1	<i>Contribuição da restrição de movimento</i>	46
3.2.6.2	<i>Contribuição da restrição conformacional</i>	46
3.2.6.3	<i>Interações eletrostáticas e ligações de hidrogênio</i>	47
3.2.6.4	<i>Contribuição Hidrofóbica</i>	49
3.2.6.5	<i>Contribuição não-polares (Interações de Van der Waals)</i>	50
3.3	MODELAGEM MOLECULAR	51
3.3.1	Docking Molecular	51
3.3.1.1	<i>Algoritmos de busca</i>	53
3.3.1.2	<i>Função de Escore</i>	54
3.3.2	Dinâmica Molecular	55
3.3.3	Ensemble docking	Erro! Indicador não definido.
4	METODOLOGIA	60
4.1	PRERARAÇÃO DAS MOLÉCULAS	60
4.1.1	Preparação do Receptor	60
4.1.2	Preparação de Ligantes	61
4.2	DOCKING MOLECULAR	63
4.3	DINÂMICA MOLECULAR	64
4.4	ENSEMBLE DOCKING	Erro! Indicador não definido.
5	RESULTADOS E DISCUSSÃO	73
5.1	DOCKING MOLECULAR	73
5.2	DINÂMICA MOLECULAR	87

5.2.1	Enzima livre em solução.....	88
5.2.2	Enzima em complexo com substrato.....	95
5.2.3	Enzima em complexo com produto.....	104
5.2.4	Enzima em complexo com ácido gálico (ligante 4).....	111
5.2.5	Enzima em complexo com ligantes 27 e 37.....	120
5.2.6	Enzima em complexo com ligantes 13 e 46.....	128
5.3	<i>ENSEMBLE DOCKING</i>	Erro! Indicador não definido.
6	CONCLUSÕES E PERSPECTIVAS	153
7	REFERÊNCIAS BIBLIOGRÁFICAS	159
	APÊNDICES E ANEXOS	165

LISTA DE FIGURAS

Figura 1 - <i>Bacillus anthracis</i> em sua forma vegetativa, colorida com corante de Gram e após esporogênese, mostrando os esporos coloridos com o corante Schöeffer Fulton	17
Figura 2 - Ferida típica formada no estágio avançado de infecção cutânea por antraz.....	18
Figura 3 - Estruturas dos sideróforos bacilobactina e petrobactina	20
Figura 4 - Reção de desidratação do ácido 3-dehidroshikímico a ácido 3,4-dihidroxibenzoico catalisada pela enzima AsbF.	21
Figura 5 - Mecanismo proposto para a catálise enzimática da desidratação de ácido 3-dehidroshikímico a ácido 3,4-dihidroxibenzoico.	22
Figura 6 - Estrutura geral de um aminoácido.	25
Figura 7 - Os nomes, abreviaturas e estrutura química dos 20 aminoácidos proteínogênicos classificados pela natureza de suas cadeias laterais.	26
Figura 8 - Estados de protonação predominantes para as cadeias laterais dos sete aminoácidos com grupos laterais ionizáveis e seus respectivos pKas.....	28
Figura 9 - Os quatro níveis de organização estrutural das proteínas.....	29
Figura 10 - Estrutura de uma hélice- α mostrando o padrão de ligações de hidrogênio entre os aminoácidos que a formam e a orientação das cadeias laterais para o exterior da hélice	30
Figura 11 - Estruturas e padrões de interações entre folhas- β antiparalelas e paralelas e entre curvas- β , também chamadas de voltas- β , dos tipos I e II	30
Figura 12 - Termodinâmica do enovelamento proteico, mostrado como um funil de energia livre com diferentes perfis e mecanismos de dobramento.....	33
Figura 13 - Diagrama de coordenada de reação, comparando uma reação catalisada por enzima com uma reação não-catalisada.....	36
Figura 14 - Mecanismo geral da catálise enzimática ácida e básica	39
Figura 15 - Mecanismo da enzima acetoacetato decarboxilase ilustrando como se processa a catálise enzimática nucleofílica ou covalente.	40
Figura 16 - Mecanismo de atuação de termolisina, demonstrando a atuação de um cofator eletricamente carregado na catálise enzimática.....	41
Figura 17 - Esquema da interação enzima-substrato e da ação de inibidores competitivos clássicos, anti-competitivos e não-competitivos	42
Figura 18 - O equilíbrio químico que determina a afinidade proteína-ligante.....	43
Figura 19 - Ácido palmítico ligado à proteína de ligação de lipídios adipócitos e comparação entre a conformação preferencial em solução e a conformação sítio ativo.....	47
Figura 20 - Exemplos de diferentes tipos de ligações de hidrogênios presentes em sistemas biológicos.....	48
Figura 21 - O potencial eletrostático molecular do benzeno na presença de um cátion.	49
Figura 22 - Curva de energia potencial em função da distância interatômica para interações do tipo Van der Waals.....	50
Figura 23 - Estrutura do substrato enzimático e dos compostos com potencial inibitório mensurável experimentalmente utilizados no presente estudo computacional.	62
Figura 24 - Grid utilizado para a realização do <i>docking</i> específico no sítio ativo, com os resíduos que compõem o sítio em destaque.....	64
Figura 25 - Número de clusters formados em função do tempo de simulação para a réplica 1 da simulações enzima-substrato. Clusterização por RMSD dos carbonos- α	70
Figura 26 - Sobreposição do melhor modo de ligação do 3,4-dihidroxibenzoato obtido no <i>docking</i> ampliado com a estrutura cristalográfica depositada no PDB.	74
Figura 27 - Padrão de interações determinado pelo programa ADT para o 3,4-DHB conforme estrutura cristalográfica e o melhor modo de ligação do <i>docking</i>	74
Figura 28 - Mínimos conformacionais possíveis para o 3-deidroshikimato, com os grupos hidroxila em posição axial ou equatorial.....	75

Figura 29 - Modos de ligação do substrato na conformação axial obtidos por <i>docking</i> cego. O cofator magnésio em destaque aponta a posição aproximada do sítio ativo da AsbF.....	76
Figura 30 - Principais interações do substrato com o sítio ativo da enzima AsbF para a modo de ligação de menor energia segundo o <i>docking</i>	76
Figura 31 - Modo de ligação mais favorável do substrato ao sítio ativo, com destaque para o sítio de ligação do magnésio, os ácidos glutâmicos 142 e 253.	77
Figura 32 - Posicionamento da Tyr217 em proximidade com os hidrogênios do carbono-5 do produto na estrutura docada indicando sua provável atuação como base.	78
Figura 33 - Energias de interação do <i>docking</i> cego e do <i>docking</i> de sítio ativo em comparação com ΔG experimental a 300K e respectivas equações de regressão linear para <i>docking</i> cego e <i>docking</i> de sítio ativo.81	
Figura 34 - Padrão de interações do ligante 4 obtido pelo <i>docking</i> restrito ao sítio ativo.	82
Figura 35 - Padrão de ligações de hidrogênio apontado pelo <i>docking</i> para os ligantes 27 e 37 com as ligações estão pontilhadas em verde.	84
Figura 36 - Sobreposição dos modos de ligação mais favoráveis obtidos tanto no <i>docking</i> restrito quanto no <i>docking</i> geral para os ligantes 13 e 46.....	86
Figura 37 - Modo de interação principal do ligante 13 e do ligante 46 após <i>docking</i> sem restrição. A entrada e os resíduos principais do sítio ativo estão mostradas como superfícies.	87
Figura 38 - Deslocamento quadrático médio das posições dos carbonos-alfa (em nanômetros) em relação à estrutura inicial, em função do tempo de simulação, para as quatro simulações da enzima livre em solução.....	88
Figura 39 - RMSD acumulado dos carbonos-alfa, em comparação com a estrutura inicial, em função da fração do número de carbono- α considerado na análise.....	89
Figura 40 - RMSD das frações mais móveis e das mais rígidas em relação a estrutura inicial para as quatro simulações da enzima livre em solução.....	90
Figura 41 - Sobreposição de 120 frames extraídos a cada nanossegundo das quatro simulações da enzima livre em solução.	91
Figura 42 - Flutuação quadrática média (RMSF) em nanômetros, em relação a estrutura inicial para todos os resíduos das quatro simulações da enzima livre em solução.	93
Figura 43 - Sobreposição da estrutura inicial e da estrutura obtida após 120 ns para a simulação 2 com o cofator magnésio está destacado para indicar a localização do sítio ativo.....	94
Figura 44 - Flutuação Quadrática Média (RMSF) em relação à estrutura inicial para os resíduos que compõem o sítio ativo para cada simulação da enzima livre em solução.	94
Figura 45 - Evolução temporal do desvio quadrático médio (RMSD) das posições dos carbonos- α em relação à estrutura inicial para as três réplicas do sistema enzima-substrato.....	96
Figura 46 - RMSD dos carbonos- α em função da fração de átomos consideradas no ajuste da estrutura para as três réplicas da simulação enzima-substrato.	96
Figura 47 - Sobreposição de 100 frames extraídos a cada nanossegundo para as três reproduções da enzima em complexo com o substrato coloridos de acordo com menores e maiores valores de RMSD. Em destaque o cofator Mg.	97
Figura 48 - Flutuação quadrática média (RMSF) de todos os resíduos em relação à estrutura inicial para as réplicas do sistema enzima-substrato e média das triplicatas.	98
Figura 49 - RMSF médio dos resíduos do sítio ativo em relação à estrutura inicial das simulações enzima-substrato.....	99
Figura 50 – Evolução temporal do número de ligações de hidrogênio que ocorrem entre a enzima e o substrato para cada réplica realizada.	100
Figura 51 - Ocupação dos resíduos em ligações de hidrogênio, acima do limite de 10%, para cada réplica da simulação enzima-substrato.	100
Figura 52 – Padrão de ligações de hidrogênio apontado pelo programa ADT para o quadro de 69590 ps da réplica 2 da simulação enzima-substrato.	102

Figura 53 - Ocupação parcial em ligações de hidrogênio em função do tempo de simulação para o sistema enzima-substrato.....	103
Figura 54 - Padrão de ligações de hidrogênio entre enzima e substrato no tempo de 33330 ps da réplica 2.....	104
Figura 55 - Evolução temporal do RMSD dos carbonos- α em relação a estrutura inicial para as 3 reproduções da simulação enzima-produto e sua média.	105
Figura 56 - RMSD acumulado dos carbonos- α em função da fração de átomos consideradas no ajuste para as três réplicas da simulação enzima-produto.	105
Figura 57 - Sobreposição de 100 frames extraídos a cada nanossegundo para as três reproduções da simulação enzima-produto coloridos segundo RMSD dos carbonos- α	106
Figura 58 - Flutuação quadrática média de todos os resíduos para as três réplicas da simulação enzima-produto e para a média.	107
Figura 59 - RMSF médio dos resíduos do sítio ativo em relação à estrutura inicial das simulações da enzima em complexo com seu produto.	107
Figura 60 - Evolução temporal do número de ligações de hidrogênio entre enzima e produto para as três réplicas da simulação.	108
Figura 61 - Ocupação dos resíduos em ligações de hidrogênio para as três réplicas da simulação enzima-produto ao longo dos 100 ns das trajetórias.	109
Figura 62 - Evolução temporal da ocupação dos resíduos em ligações de hidrogênio para o complexo enzima-produto.	110
Figura 63 - Padrão de ligação de hidrogênio entre enzima e produto para o frame de 59600 ps da réplica 1, que representam o máximo de ligações simultâneas.....	110
Figura 64 - Desvio quadrático das posições atômicas dos carbonos- α para as três reproduções da simulação da enzima em complexo com o ácido gálico.	112
Figura 65 - RMSD dos carbonos- α das três réplicas da simulação enzima-ligante 4 em função da fração de átomos considerada no alinhamento.....	112
Figura 66 - Sobreposição das estruturas geradas por dinâmica em 100 frames de 1 ns para o sistema enzima-ligante 4 para as três réplicas.....	113
Figura 67 - Flutuação quadrática média de todos os resíduos para as três reproduções do sistema AsbF-ligante 4 e valor médio.	114
Figura 68 - RMSF médio dos resíduos do sítio ativo das três reproduções do complexo enzima-ligante 4.....	115
Figura 69 - Evolução do número de ligações de hidrogênio simultâneas formadas entre o ligante 4 e o receptor para as três reproduções desse complexo.	115
Figura 70 - Sobreposição do primeiro frame e do último para a terceira réplica do sistema enzima-ligante 4.....	116
Figura 71 - Ocupação dos resíduos em ligação de hidrogênio com o ácido gálico para cada réplica da simulação desse sistema.	117
Figura 72 - Evolução temporal da ocupação em ligações de hidrogênio ao longo da réplica 2 do complexo enzima-ligante 4	118
Figura 73 - Padrão ligações de hidrogênio que ocorrem entre o ligante 4 e o sítio ativo enzimático no frame de 97680 ps da réplica 1 e de 17020 ps da réplica 3 e que correspondem aos máximos de LHs simultâneas de suas trajetórias.....	119
Figura 74 - RMSD das posições dos carbonos- α ao longo das simulações da enzima em complexo com ligante 27 e em complexo com ligante 37 e suas respectivas médias.	120
Figura 75 - RMSD acumulado em função da fração de carbonos- α da estrutura considerada na análise para as duas réplicas da enzima em complexo com os ligantes 27 e 37.	121
Figura 76 - Sobreposição de 100 frames ao longo das trajetórias geradas para a enzima em complexos com o ligante 27 e com o ligante 37.	122
Figura 77 - Flutuação quadrática média dos resíduos ao longo das simulações da enzima em complexo com ligante 27 e com ligante 37 e respectivas médias.	123

Figura 78 - Evolução temporal do número de ligações de hidrogênio simultâneas para cada réplica da simulação da enzima em complexo com os ligante 27 e 37	124
Figura 79 - Ocupação dos resíduos em ligações de hidrogênio para as duas réplicas da simulação da enzima em complexo com o ligante 37.	124
Figura 80 - Evolução temporal da ocupação dos resíduos em ligações de hidrogênio para a réplica 2 do complexo enzima-ligante 37.	126
Figura 81 - Ligações de hidrogênio formadas entre o ligante 37 e os resíduos do sítio ativo no último frame da réplica 1 e da réplica 2 desse sistema.	126
Figura 82 - Ligações de hidrogênio entre o ligante 27 e o sítio ativo enzimático no último frame da réplica 1 da réplica 2 desse sistema.	127
Figura 83 - Temporal da ocupação dos resíduos em ligações de hidrogênio com o ligante 27 na réplica 1 desse sistema.	128
Figura 84 - Evolução temporal do RMSD dos carbonos- α em relação à estrutura inicial para as duas réplicas da simulação da enzima em complexo com o ligante 13 e em complexo com o ligante 46 e respectivas médias.	129
Figura 85 - RMSD acumulado em função da fração de carbonos- α consideradas na análise para cada réplica das simulações da enzima em complexo com os ligantes 13 e 46.	130
Figura 86 - Sobreposição de 100 frames, em cada nanossegundos de trajetória, para os sistemas enzima em complexo com ligante 13 e enzima em complexo com ligante 46	131
Figura 87 - Distância entre o cofator Magnésio e os ligantes 13 e 46 ao longo das trajetórias para cada réplica das simulações.....	132
Figura 88 - Sobreposição das estruturas inicial e após 100 ns de simulação do ligante 13 para a réplicas 1 e a réplica 2	132
Figura 89 - Sobreposição das estruturas do ligante 46 no tempo inicial e ao final de 100 ns de simulação para a primeira e a segunda réplica do sistema.	133
Figura 90 - Evolução temporal da ocupação dos resíduos em ligações de hidrogênio com o ligante 46, na primeira e segunda reprodução.	134
Figura 91 - Ligações de hidrogênio entre a enzima e o ligante 46 ao final de 100 ns de simulação, mostrando que o inibidor entra parcialmente no sítio ativo.	134
Figura 92 - Evolução temporal do número de ligações de hidrogênio entre o ligante 13 e ligante 46 e o receptor, para cada réplica realizada.....	135
Figura 93 - Complexos enzima-ligante 13 e enzima-ligante 46 nos tempos de 45440 ps e 45000 ps, respectivamente, momento em que ocorre a formação do maior número de ligações de hidrogênio simultâneas com a enzima.....	140
Figura 94 - Coeficientes de correlação de Pearson e inclinação da curva de correlação entre a média dos resultados de <i>docking</i> e as energias livres experimentais	146
Figura 95 - Relações entre os resultados de <i>docking</i> e os valores experimentais de energia livre de ligação e curva de correlação obtidas para o <i>Ensemble 2</i> em comparação com o obtido pelo <i>docking</i> na estrutura cristalográfica	148
Figura 96 - Sobreposição do modo de ligação de menor energia nas estruturas que compõem o <i>Ensemble 2</i> , extraídas das simulações da enzima com ligantes 23, 37 e 13.	149
Figura 97 - Modo de ligação de menor energia para o ligante 04 obtido por <i>docking</i> nas estruturas do <i>Ensemble 2</i> obtidas das simulações com ligante 27 e com o ligante 37	150
Figura 98 - Modo de ligação de menor energia para o produto segundo <i>docking</i> nas estruturas do <i>Ensemble 2</i> obtidas das simulações com ligante 13 e com o ligante 37	150
Figura 99 - Modo de ligação de menor energia para o ligante 13 e o ligante 46 segundo <i>docking</i> nas estruturas do <i>Ensemble 2</i> obtida a partir da simulação com o ligante 37.	151
Figura 100 - Relações entre os resultados de <i>docking</i> e os valores experimentais de energia livre de ligação e curva de correlação obtidas para o <i>Ensemble 2</i> em comparação com o obtido pelo <i>docking</i> na estrutura cristalográfica onsiderando-se os 10 ligantes mais ativos	152

LISTA DE TABELAS

Tabela I - Número de reproduções e estrutura do ligante para cada sistema simulado por dinâmica molecular.....	66
Tabela II - Número de ligações de hidrogênio e localização dos frames em que ocorrem o maior número de ligações de hidrogênio simultâneas para cada réplica de cada ligante.....	69
Tabela III - Número total de cluster formados e número de cluster principais em cada metodologia de clusterização aplicada para cada simulação.....	71
Tabela IV - Valores de IC ₅₀ determinados experimentalmente por Koppisch e colaboradores, em ordem ascendente, e energia livre de Gibbs de ligação a 303 K.	73
Tabela V - Afinidades de ligação estimadas pelo <i>docking</i> com Vina utilizando-se grid envolvendo toda a enzima ou restrito ao sítio ativo e comparação com valores de ΔG experimentais a 300 K.	80
Tabela VI - Principais interações entre os ligantes e a enzima obtidos pelo programa BINANA após <i>docking</i> restrito ao sítio ativo.....	83
Tabela VII - Coeficientes de correlação de Pearson, coeficiente de determinação, inclinação da reta de correlação e erro quadrático médio entre as energias livres experimentais e os resultados de <i>docking</i> com cada estrutura final gerada.	137
Tabela VIII - Coeficientes de correlação de Pearson, coeficiente de determinação, inclinação da reta de correlação e erro quadrático médio entre as energias livres experimentais e os resultados de <i>docking</i> obtidos com a estruturas que maximizam o número de ligações de hidrogênio simultâneas enzima-ligante.	139
Tabela IX - Parâmetros de qualidade dos modelos de correlação linear entre as energias livres experimentais e os resultados de <i>docking</i> com a estruturas principais obtidas pela clusterização por RMSD dos carbonos- α	141
Tabela X - Parâmetros de qualidade dos modelos de correlação linear entre as energias livres experimentais e os resultados de <i>docking</i> com a estruturas principais obtidas pela clusterização por RMSD do sítio ativo.	142
Tabela XI - Valores médios dos parâmetros de qualidade e da inclinação das curvas de correlação considerando-se todas as estruturas geradas de cada sistema, em ordem crescente de correlação de Pearson.....	143
Tabela XII - Parâmetros de qualidade e da inclinação das retas de correlação em ordem crescente de R-Pearson para as estruturas com desempenho de <i>docking</i> superior à estrutura cristalográfica geradas pelas três estratégias de clusterização empregadas.	144
Tabela XIII - Escores de <i>docking</i> e desvio percentual em relação aos valores experimentais dos dois <i>ensembles</i> finais gerados a partir dos melhores resultados das estruturas individuais. Resultados em ordem crescente de ΔG de ligação experimental.	147

1 INTRODUÇÃO

Um dos problemas de saúde pública mais graves a ser enfrentado ao longo das próximas décadas será o aumento da resistência de bactérias patogênicas aos antibióticos atualmente disponíveis, tanto em nível regional e nacional quanto à nível global.¹

Todos os anos, somente nos Estados Unidos, em torno de 2 milhões de pessoas adquirem algum tipo de infecção resistente a um ou mais antibióticos, de acordo com dados do Centro de Controle e Prevenção de Doenças do país (CDC). Desses 2 milhões, cerca de 23 mil acabam morrendo anualmente como consequência direta dessas infecções resistentes a antibióticos.²

Segundo o relatório mais recente da Organização Mundial de Saúde (OMS) sobre o tema, publicado em 2016, altas taxas de resistência têm sido observadas em bactérias comuns como *Escherichia coli*, *Klebsiella pneumoniae* e *Staphylococcus aureus* responsáveis por problemas de saúde triviais até então, como pneumonias e infecções urinárias.³

No Brasil ainda não há dados oficiais de abrangência nacional sobre o número de vítimas por bactérias resistentes, porém se estima que número de óbitos anuais causados por essas infecções seja próximo, se não superior ao observado nos Estados Unidos.⁴

Ao longo dos últimos anos, inúmeros casos de infecções por bactérias super-resistentes foram registrados em todo o Brasil. Dados da Agência Nacional de Vigilância Sanitária, publicados em 2015, mostram que 77,4% das infecções da corrente sanguínea por *Acinetobacter spp.* foram causadas por cepas resistentes.⁵

Em 2015, duas UTIs neonatal foram fechadas na cidade de Porto Alegre devido à contaminação por *Klebsiella pneumoniae* resistente a carbapenemas (ou KPC)⁶. Em março ano seguinte, seis pessoas morrem e mais de 40 foram contaminadas em Americana, interior do Estado de São Paulo, também por infecção por KPC.⁷ Cepas dessa mesma bactéria foram encontrada por pesquisadores da Universidade Federal do Rio de Janeiro, nas águas da baía de Guanabara e das principais praias da orla carioca, em julho de 2016, às vésperas da realização dos Jogos Olímpicos na cidade.⁸

Ainda mais alarmante foi a descoberta, publicada no final de novembro de 2015, de que bactérias da família *Enterobacteriaceae*, que inclui os gêneros *Escherichia* e *Salmonella*, desenvolveram um mecanismo de transmissão da resistência à Colistina – considerado o último recurso na luta contra agentes infecciosos bacterianos – por meio da transferência do gene de resistência MCR-1 através de plasmídeos, pequenos fragmentos circulares de DNA facilmente trocados entre bactérias.⁹

Essa resistência foi observada inicialmente na China, e foi relacionada ao uso de Colistina para acelerar o ganho de peso em animais de produção, porém, menos de quatro meses após a descoberta, em março de 2016, cepas de bactéria contendo o gene MRC-1 já haviam sido reportadas na África e Europa.¹⁰ Em maio do mesmo ano, bactérias com a resistência também foram reportadas nos Estados Unidos¹¹ e no mês seguinte, cepas de *E. coli* contendo o gene MRC-1 foram encontradas no nordeste do Brasil.¹²

Em 2014, a Organização Mundial de Saúde já havia alertado que a distribuição generalizada ao redor do mundo de cepas bacterianas resistentes e o rápido desenvolvimento de tolerância aos medicamentos ainda ativos se constituiria "uma ameaça às realizações da medicina

moderna, com vias de tornar o século XXI em uma Era pós-antibiótica".³ Ninguém poderia prever, entretanto, que essa Era estaria tão próxima.

Em janeiro de 2017 foi reportado o primeiro caso de uma infecção bacteriana resistente a *todos* os antibióticos disponíveis. O caso foi registrado em Nevada, nos Estados Unidos, e envolveu uma mulher com cerca de 70 anos de idade admitida na emergência hospitalar com síndrome da resposta inflamatória sistêmica causada por *Klebsiella pneumoniae*, em 18 de agosto de 2016. Após a falha de todas as estratégias terapêuticas, a paciente faleceu de choque séptico duas semanas após dar entrada no hospital.¹³

Não bastasse o rápido desenvolvimento de resistência e até mesmo *pan*-resistência a antibióticos e a rápida propagação dessas cepas ao redor do mundo, bactérias e outros agentes infecciosos considerados como controlados ou até mesmo erradicados estão ressurgindo como consequência do aquecimento do planeta. Na medida em que as temperaturas globais se elevam, induzem ao rápido derretimento das áreas de *permafrost* – uma camada de solo permanentemente congelado nas áreas polares e subpolares que recobre cerca de 25% das áreas emersas do hemisfério norte – expondo restos humanos e de animais contaminados com os mais diferentes patógenos, como os causadores da varíola, gripe espanhola e antraz.¹⁴

Em agosto de 2016, temperaturas anormalmente altas foram registradas na região da península de Yamal, no norte da Rússia, causando o derretimento de camadas superficiais do Permafrost e expondo restos de animais que morreram vítimas de *antraz* em 1941. A exposição desses corpos liberou no ambiente esporos da bactéria causadora da doença, fazendo com que um menino de 12 anos de idade morresse e quase 100 pessoas fossem hospitalizadas, além de provocar a morte de mais de 2300 renas.^{15,16}

O antraz, também conhecido como carbúnculo hemático, é uma doença infecciosa aguda e altamente letal, com manifestações clínicas semelhantes às da febre hemorrágica.¹⁷ É considerada como uma zoonose, na medida em que afeta principalmente animais herbívoros domésticos e selvagens, porém todos os mamíferos e inclusive os seres humanos são sucessíveis a doença. As infecções em humanos – predominantemente cutâneas – resultam do contato com animais infectados ou com produtos contaminados de origem animal, como couros e lã.¹⁸

Conhecida desde a Antiguidade, acredita-se que a doença tenha sido registrada pela primeira vez na Bíblia, no capítulo 9 do livro de Êxodo, que descreve a quinta e sexta pragas do Egito como uma “pestilência gravíssima” que afetou inicialmente o gado e que após Moisés “espalhar um pó miúdo por todo o Egito” produziu em sarna e úlceras nos homens e no gado.^{19,20}

As evidências atualmente disponíveis, entretanto, apontam a África subsaariana com o local mais provável de surgimento do Antraz, de onde se espalhou para o norte da África, Eurásia e América do Norte seguindo os fluxos migratórios humanos.²¹ Atualmente, a doença é endêmica em praticamente todo o mundo, inclusive nos Sul do Brasil, e seu agente etiológico está presente no solo de todos os continentes, exceto a Antártida.^{22,23}

Ao longo do século XIX e no começo do século XX, o antraz foi uma das principais causas de mortalidade na pecuária em todo mundo, chegando a matar 60 mil cabeças de gado na África do Sul, somente em 1923.¹⁹ Nas décadas seguintes, o desenvolvimento de vacinas e a adoção de melhores práticas sanitárias levou a uma queda abrupta no número de casos de antraz no gado doméstico na maioria dos países, com a consequente redução do número de casos humanos.²¹

Apesar disso, o maior surto já registrado de antraz em humanos ocorreu entre 1979 e 1980 no Zimbábue, quando mais de 6.000 pessoas foram infectadas, a maior parte com a forma cutânea da doença e em torno de 100 pessoas morreram.²⁴ Em 1987, o maior surto já registrado da doença na América do Sul ocorreu no Paraguai, quando 25 pessoas adquiriram a doença após o abate de uma única vaca infectada, porém sem vítimas fatais.²⁵

Atualmente, o antraz é tido uma doença ocupacional em humanos, limitada a categorias específicas de trabalhadores, como agricultores, pastores, veterinários, trabalhadores de matadouros e pessoas que tem contato direto com produtos de origem animal não processados.²⁵ Todo os anos são registrados entre 20 mil e 100 mil casos de infecções por antraz em todo o mundo, o que a torna uma doença relativamente rara.²⁶

O antraz é causado por um micro-organismo da família Bacillaceae, o *Bacillus anthracis*, uma bactéria aeróbica gram-positiva, imóvel, encapsulada e que tem o formato de bastonete, medindo entre 1-1,5 μm de largura por 5-6 μm de comprimento (Figura 1A).²⁷ Em condições adversas como calor ou frio extremo, baixa umidade ou carência nutricional esse micro-organismo forma esporos de aspecto ovalado (Figura 1B), que são liberados para o ambiente após a lise celular da bactéria.¹⁷

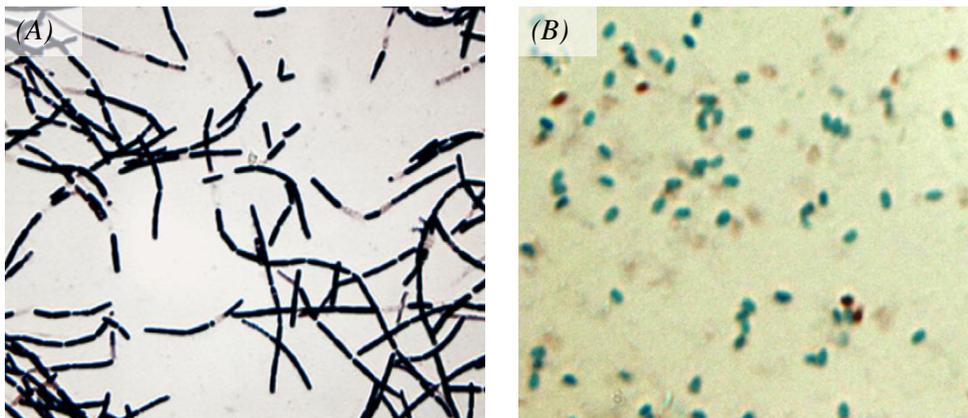


Figura 1 - *Bacillus anthracis* em sua forma vegetativa, colorida com corante de Gram (A) e após esporogênese, mostrando os esporos coloridos com o corante Schöeffer Fulton. Adaptado de Fasanella et al. (2010).¹⁷

Esse esporos podem se manter viáveis por décadas, tendo uma meia vida de cerca de 100 anos, e são resistentes à dessecação, solventes, extremos de temperatura, pressão e pH bem como à radiação ultravioleta e até à radiação ionizante.^{28,29} Além disso, esses esporos são facilmente dispersáveis no ambiente, podendo adentrar em mamíferos por vias respiratórias, cutânea ou gastrointestinal, quando então germinam e retornam à forma vegetativa e de alta virulência, que rapidamente estabelece infecção e eventualmente mata o hospedeiro.³⁰

Cerca de 95% dos casos de antraz reportados em humanos são da forma cutânea da doença, que ocorre quando os esporos do *B. anthracis* penetram nas camadas mais profundas da pele através de cortes ou abrasão. Os esporos germinam e produzem um edema local, que entre 3 e 5 dias forma uma ferida pruriginosa que evolui para necrose e ressecamento do tecido, deixando uma escara negra característica (Figura 2).²⁷ O nome antraz, inclusive, é derivado da palavra grega *anthrax*, que significa carvão, em referência a aparência dessa ferida.³¹ Sem o tratamento adequado e imediato a mortalidade é de cerca de 20%, na medida em que bactéria pode ingressar na corrente sanguínea e induzir ao choque séptico.²⁷



Figura 2 - Ferida típica formada no estágio avançado de infecção cutânea por antraz. Adaptado de Sweeney et al. (2011).²⁷

A antraz gastrointestinal é causado pela principalmente ingestão de carne malcozida contendo esporos da bactéria e apresenta sintomas iniciais pouco específicos, como náusea, vômitos e febre, os quais rapidamente progridem para ulceração e edema severo do trato gastrointestinal eventualmente levando à sepse e morte. Essa forma de antraz tem uma taxa de mortalidade superior a 50%, a qual está relacionada com a demora no tratamento adequado devido à falta de especificidade dos sintomas.³²

É justamente a falta de sintomas iniciais específicos que torna a forma respiratória da doença a mais letal de todas, com uma taxa de mortalidade superior a 90%. A infecção ocorre quando os esporos do *B. anthracis*, com menos de 5 µm são inalados e atingem o trato respiratório inferior, de onde são transportados pelos macrófagos até os nodos linfáticos, onde esses esporos germinam e começam a produzir toxinas.²⁷

Após um período de incubação, que dura de um a seis dias, os pacientes desenvolvem sintomas parecidos com a gripe, como tosse, febre baixa, fadiga e dores peitorais e abdominais, que duram desde algumas horas até alguns dias, e então melhoram brevemente.³² O segundo estágio da doença, geralmente fulminante, é caracterizado por um aumento da febre, edema no tórax e pescoço, dificuldade cardíaca e pulmonar que culmina com o choque séptico e a morte.^{27,32}

A alta resistência dos esporos dessa bactéria aliada à sua facilidade de obtenção, produção e dispersão no ambiente e em conjunto com a letalidade extrema que a infecção detém em sua forma respiratória colocam o *B. anthracis* na lista dos agentes biológicos mais perigosos do mundo, com potencial aplicação em bioterrorismo e como arma de destruição em massa.³³

Em 2001, logo após os ataques terroristas ao World Trade Center e ao Pentágono, envelopes contendo esporos de *B. anthracis* foram enviados a empresas de comunicação e funcionários do governo dos Estados Unidos, entre outubro e novembro daquele ano, infectando 22 pessoas, das quais 11 desenvolveram antraz respiratório e 11 desenvolveram a forma cutânea da infecção. Todos os pacientes que contraíram antraz respiratório necessitaram de hospitalização e terapia intensiva e dentre esses, cinco foram a óbito.^{19,34}

A Organização Mundial de Saúde estima, com base em cálculos científicos, que a liberação de 50 kg de *B. anthracis* em pó ao longo de duas horas em uma cidade com 500.000 habitantes resultaria na perda de 95.000 vidas humanas, com a incapacitação adicional de outras 125.000 pessoas, e consequente ruptura da infraestrutura civil e o esgotamento dos recursos médicos ao longo de poucos dias.^{18,35}

Atualmente, existem três formas de se enfrentar a doença: a vacinação, para prevenir a infecção bacteriana; o uso de antibióticos, para barrar a proliferação do micro-organismo no hospedeiro, impedindo que a infecção se alastre; e os tratamentos antitoxina, que inibem os efeitos nocivos das substâncias liberadas pela bactéria no organismo do hospedeiro.¹

Apesar dos programas de vacinação terem controlado o antraz em animais de criação e de existirem vacinas para humanos contra o *B. anthracis* desde 1954, aspectos práticos e econômicos tornam a vacinação em massa contra esse agente biológico impraticável. Além disso, apenas a vacinação não seria efetiva para prevenção num cenário pós-exposição em função da demora na resposta imune do organismo.³⁶

A existência de toxinas do *B. anthracis* foi postulada por Robert Koch no século XIX e comprovada na década de 50 do século XX, quando se descobriu que a alta virulência da bactéria estava relacionado com a presença de dois plasmídeos, pXO1 de 181 kb e pXO2 de 93.5 kb.³⁷

O plasmídeo de virulência pXO1 contém o gene de três proteínas diferentes, chamadas de antígeno protetor, fator letal e fator de edema, que se combinam para formar duas toxinas diferentes, enquanto que o plasmídeo pXO2 codifica uma cápsula antifagocitose composta por poli-D-glutamato, um polipeptídeo de alto peso molecular que inibe a fagocitose da forma vegetativa do *B. anthracis* pelas células do sistema imunológico.^{38,39}

O fator letal é uma protease de zinco, um tipo de enzima que contém zinco e que quebra outras proteínas, enquanto que o fator de edema pertence à classe das adenilato ciclases. A combinação do fator letal com antígeno protetor – que, por sua vez, é uma proteína que se liga a receptores celulares – forma a toxina letal, ao passo que a combinação do antígeno com o fator de edema dá origem à toxina de edema.^{38,40}

Diferente de outras infecções bacterianas, como o botulismo, tétano e difteria, a alta letalidade e virulência do *B. anthracis* aparenta não estar relacionada à ação direta de toxinas, inclusive alguns estudos sugerem que a toxina letal, por exemplo, requer quantidades da ordem de miligramas para causar morte, tendo, portanto, uma potência significativamente baixa.⁴¹

Acredita-se que o papel dessas toxinas seja apenas danificar as células de defesa, levando ao mal funcionamento do sistema imune e permitindo assim que a bactéria se estabeleça e se multiplique dentro do organismo hospedeiro. Nos estágios finais de infecção, entretanto, essas toxinas já estariam presentes em quantidades grandes na corrente sanguínea, contribuindo diretamente com a morte do organismo hospedeiro.⁴²

Por outro lado, estudos em ratos com cepas de *B. anthracis* que não apresentavam o plasmídeo pXO1 e que eram, portanto, incapazes de sintetizar as toxinas continuaram apresentando elevada virulência⁴³.

Isso mostra que, das três estratégias existentes de enfrentamento ao antraz, apenas o uso de antibióticos permanece como um pilar central no tratamento e como resposta adequada a uma eventual crise sanitária. Por mais que terapias antitoxina possam ser consideradas como coadjuvantes importantes, sobretudo na mitigação dos efeitos tóxicos que são consequência da infecção, de nada servem se a proliferação do micro-organismo não for inibida, em primeiro lugar.⁴²

Ainda que a maioria das cepas do *B. anthracis* permaneçam responsivas a antibióticos como ciprofloxacina, doxiciclina e a penicilina G, Athamna e colaboradores (2004) mostraram que é possível – e relativamente fácil – induzir o bacilo à resistência a estes antibióticos.⁴⁴ Além

disso, a emergência de bactérias *multi* e até mesmo *pan*-resistentes e que facilmente trocam essa resistência umas com as outras torna imperativo e crucial a identificação de novos alvos terapêuticos e o desenho de novas drogas contra esse micro-organismo tão perigoso e mortal.

Um dos alvos possíveis para o desenvolvimento desses novos agentes antimicrobianos são os mecanismos bacterianos de aquisição de nutrientes essenciais. Um desses nutrientes é o ferro, que serve como cofator para uma grande variedade de enzimas que são responsáveis por reações cruciais como a síntese de RNA, transferência de elétrons e resistência a espécies reativas de oxigênio, entre outras.⁴⁵

O íon ferro (III) é muito insolúvel em condições biológicas, o que faz com que a concentração desse íon livre em um ser humano, por exemplo, varie entre 10^{-15} e 10^{-24} mol.L⁻¹, enquanto que bactérias patogênicas necessitam de aproximadamente 1 μ mol.L⁻¹ para obter crescimento ótimo e proliferação significativa dentro do organismo hospedeiro, o que levou esse microrganismo a desenvolver uma variedade de sistemas para extrair ferro dos hospedeiros.⁴⁶

No *Bacillus anthracis* essa aquisição é realizada por meio da liberação de dois sideróforos catecolatos – moléculas quelantes com alta afinidade por ferro – que são a petrobactina, que contém grupos 3,4-dihidroxibenzoila, e a bacilobactina, formada por grupos 2,4-dihidroxibenzoila (Figura 3).^{47,48}

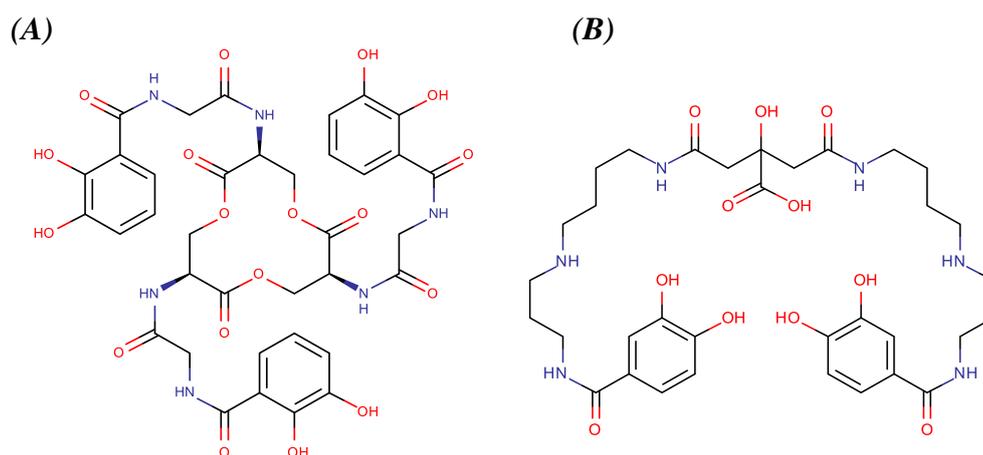


Figura 3 - Estruturas dos sideróforos bacillobactina (a) e petrobactina (b).

Entretanto, apenas a petrobactina se mostrou importante no estabelecimento pleno de virulência em mamíferos, sendo o primeiro sideróforo liberado pela bactéria em condições de depleção desse nutriente.^{49,50}

Dentre as razões que justificam essa preferência estão a alta afinidade deste composto por ferro, cerca de 100 vezes superior à afinidade da transferrina, que é a principal proteína transportadora de ferro no plasma sanguíneo, bem como a facilidade com que esse composto escapa de ser sequestrado pela siderocalina, uma proteína do sistema imunológico especializada em capturar sideróforos com grupos catecolatos, a qual apresenta alta afinidade pela bacillobactina porém com afinidade negligenciável pela petrobactina.^{51,52}

A petrobactina foi originalmente isolada da bactéria marinha *Marinobacter hydrocarbonoclasticus*, responsável pela degradação de hidrocarbonetos presente no petróleo, de onde o composto herdou seu nome.⁵³ O grupo de genes (operon) *asb*ABCDEF do *B. anthracis* apresenta homologia com o operon responsável pela biossíntese de petrobactina em

Marinobacter, o que permitiu, por meio de investigações experimentais, a elucidação do mecanismo de síntese biológica desse sideróforo em *B. anthracis*.⁵⁰

Esse mecanismo envolve a participação de três unidades biossintéticas básicas - citrato, espermidina e ácido 3,4-dihidroxibenzoico (3,4-DHBA).⁵⁴ No entanto, enquanto citrato e espermidina são metabólitos prontamente disponíveis à célula bacteriana, 3,4-DHBA é um precursor pouco usual especialmente quando se considera que grande maioria dos sideróforos com grupos catecolatos utiliza o isômero estrutural 2,3-DHB.⁵⁵ Essa diferença pode estar relacionada com a dificuldade de captura da petrobactina pelo sistema imunológico, sendo então crucial para a elevada virulência desse micro-organismo, quem em última análise é também responsável por sua alta letalidade.⁵⁰

Através de mutações genéticas nos diferentes segmentos do operon *asb*ABCDEF com o propósito de inativar as enzimas por eles codificadas, Lee e colaboradores (2007) observaram que todos os mutantes eram capazes de produzir 3,4-DHBA, exceto quando a mutação ocorria no gene *AsbF*.⁵⁶

Posteriormente, Fox e colaboradores (2008) demonstraram através da superexpressão desse gene em um hospedeiro *Escherichia coli* que a enzima *asbF* atua na desidratação do 3-dihidroschikimato (3-DHS), um metabólito da rota do ácido shikímico – relacionada com a biossíntese dos aminoácidos fenilalanina, tirosina e triptofano em bactérias – convertendo esse composto em 3,4-DHBA, o que a caracteriza como uma *deidroshikimato desidratase* (DSD).⁵⁷ A reação catalisada por essa enzima está esquematizada na Figura 4.

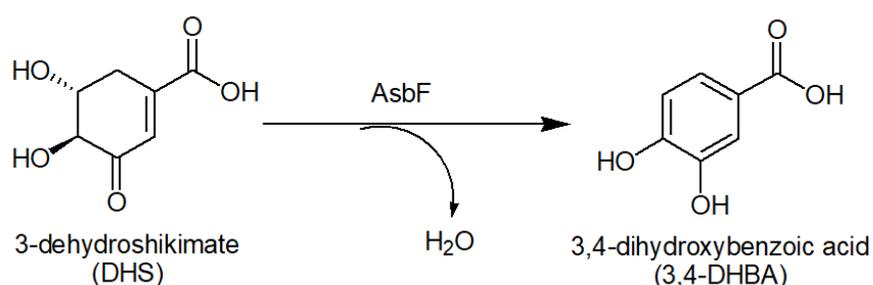


Figura 4 - Desidratação do ácido 3-dehidroschikímico a ácido 3,4-dihidroxibenzoico catalisada pela enzima AsbF.

Fox e colaboradores também demonstraram que essa enzima tem uma exigência absoluta por metais bivalentes, sobretudo magnésio, manganês e cobalto, tendo a eficiência reduzida quando zinco, cálcio ou níquel são utilizados. Além disso, o pH de funcionamento ótimo foi determinado como sendo entre 8,3 e 8,8, com uma atividade residual inferior a 15% em pH fisiológico.⁵⁷ Nesse mesmo ano Pflieger e colaboradores publicaram a estrutura cristalográfica da enzima.⁵⁸

A análise da estrutura cristalográfica obtida utilizando-se manganês como cofator metálico mostrou que a enzima apresenta estrutura terciária em forma de barril, composta por oito hélices- α e oito folhas- β paralelas que se alternam, com a porção N-terminal se localizando na parte interna do barril e a porção C-terminal na superfície.⁵⁸

O produto da reação (3,4-DHBA) foi encontrado coordenado ao sítio ativo, cercado por resíduos aromáticos (Phe104, Phe211 e Phe255), tirosina-217 e histidina-144 e em interação com as tirosinas 70 e 217, arginina-102, histidina-175, entre outros resíduos, além de interagir com o íon metálico.

Através dessas observações estruturais e de mutagênese sítio-específica com o propósito de se descobrir os resíduos chaves na catálise enzimática, Pflieger e colaboradores propuseram um mecanismo do tipo E1cB (eliminação unimolecular via base conjugada), no qual a His144 atua como base abstraindo o hidrogênio do carbono alifático C4 do 3-DHS (Figura 5). A proximidade desse resíduo do ácido glutâmico-142 poderia elevar sua basicidade pela formação de ligação de hidrogênio desse resíduo com o segundo nitrogênio do anel imidazol dessa histidina, favorecendo sua habilidade na abstração de prótons.⁵⁸

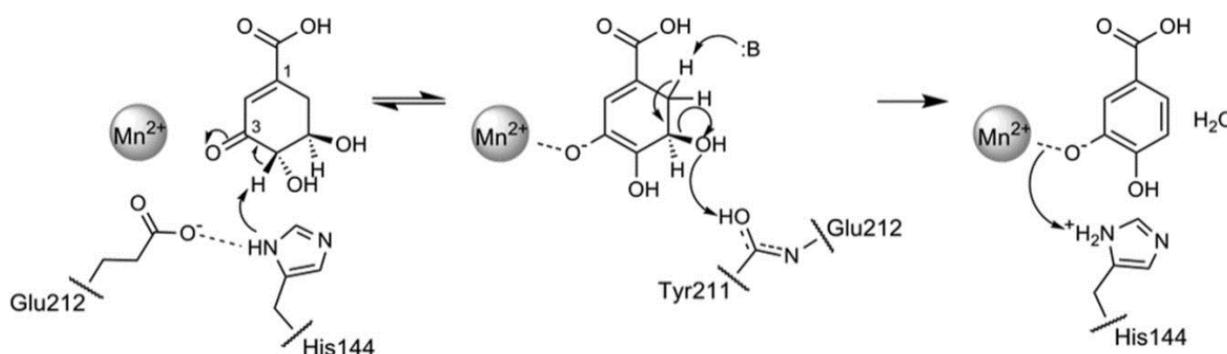


Figura 5 - Mecanismo proposto para a catálise enzimática da desidratação de ácido 3-dehidroshikímico a ácido 3,4-dihidroxibenzoico. Adaptado de Pflieger e colaboradores (2008).⁵⁸

A remoção desse hidrogênio do carbono C4 do 3-DHS levaria à formação de um intermediário enolato, não ligado covalentemente, cuja carga poderia ser estabilizada pelo cátion metálico bivalente.⁵⁸

A etapa seguinte, por sua vez, seria a eliminação da hidroxila em posição C5 por meio de catálise ácido-base geral com consequente formação do produto após aromatização. Pflieger e colaboradores sugerem que essa remoção é catalisada pela interação com a carbonila da cadeia principal da Phe-211, que é parte de uma pequena hélice- α que compõem o laço helicoidal que liga a folha- β 7e a hélice- α 7 e que recobre o sítio ativo.⁵⁸

Considerando-se o papel primordial da petrobactina no estabelecimento de virulência e, conseqüentemente, na alta taxa de mortalidade da infecção por *Bacillus anthracis* bem como a importância que o ácido 3,4-dihidroxibenzoico desempenha dentre as unidades biossintéticas desse composto, a inibição da enzima deidroshikimato desidratase se constitui de um alvo interessante para o desenvolvimento de novas terapias anti-antraz. Esse alvo é particularmente estratégico na medida em que não há outro mecanismo conhecido de geração de 3,4-DHB nesse micro-organismo tampouco existem enzimas homólogas a esta em seres humanos.

Dessa forma, esse trabalho se propõem à aplicação de métodos de modelagem molecular na compreensão das características estruturais da enzima deidroshikimato desidratase bem como no estudo das interações entre esse alvo macromolecular e potenciais ligantes, permitindo um entendimento mais aprofundado dos mecanismos relacionados à inibição enzimática e auxiliando no futuro desenho racional de novos compostos com maior potência e seletividade e com potencial aplicação como novas drogas para o tratamento do antraz.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Estudar a interação entre a enzima deidroshikimato desidratase do *Bacillus anthracis* e ligantes com possível ação inibitória por meio de modelagem molecular através das metodologias computacionais de *docking*, *ensemble docking* e simulação por dinâmica molecular, visando a compreensão das características estruturais e dos processos relacionados à inibição enzimática e o planejamento racional de moléculas com maior potência, eficácia e seletividade e com potencial aplicação como novos fármacos contra o antraz.

2.2 OBJETIVOS ESPECÍFICOS

Avaliar o modo e a intensidade de interação entre potenciais inibidores e a enzima por meio do *docking* molecular e da correlação desses dados com os resultados experimentais disponíveis a fim de se compreender os padrões estruturais e eletrônicos fundamentais responsáveis pela atividade biológica desses compostos.

Analisar o comportamento estrutural da enzima e de suas interações com os ligantes em função do tempo por meio da simulação por dinâmica molecular em busca de alterações importantes que impactem o modo, a intensidade ou a seletividade dessas interações.

Aperfeiçoar o *docking* através da obtenção de um conjunto de estruturas distintas da enzima, extraídas das simulações por dinâmica molecular, capaz de amostrar a mobilidade natural do receptor em condições fisiológicas e assim permitir a construção de um modelo teórico que possibilite predições quantitativas confiáveis para a triagem virtual de novas moléculas (*ensemble docking*).

Planejar racionalmente, desenhar e testar virtualmente por *docking* novas moléculas visando o aumento da potência e eficácia (diminuição do IC₅₀) e da seletividade em busca de compostos com qualidades terapêuticas aprimoradas e potencial aplicação como medicamento no combate a antraz e outras infecções por bactérias do gênero *Bacillus*.

Futuramente, através da parceria com o Dr. Andrew Koppisch e a Dra. Cindy Browder da Northern Arizona University (Flagstaff, Estados Unidos) as moléculas virtualmente mais promissoras serão sintetizadas e testadas *in vitro* e *in vivo*.

3 FUNDAMENTAÇÃO TEÓRICA

3.1 PROTEÍNAS: COMPOSIÇÃO E ESTRUTURA

As proteínas são certamente as macromoléculas mais abundantes e com a maior diversidade de funções nos seres vivos, atuando em praticamente todos os processos vitais, incluindo o controle e regulação metabólica (hormônios peptídicos), a sinalização intercelular (receptores) e o transporte de íons, água e outros nutrientes através das membranas celulares (poros e canais). Além disso, também atuam como componentes estruturais, formando a matriz de ossos e tecidos, bem como na contração muscular, na replicação, transcrição e tradução dos genes e especialmente na catálise de reações bioquímicas através das *enzimas*, entre outras inúmeras funções.⁵⁹

Apesar da multiplicidade de funções, todas as proteínas têm a mesma composição química básica, sendo constituídas de sequências lineares de α -aminoácidos unidas por ligações amido, também chamadas de *ligações peptídicas*, que são formadas através da reação de condensação entre o grupo $-NH_2$ de um aminoácido com o grupo $-COOH$ do outro, com a consequente liberação de uma molécula de água. A cadeia polimérica assim obtida recebe o nome de *polipeptídeo* e os monômeros que a formam são chamados de *resíduos*.⁶⁰

As proteínas consistem em uma ou mais cadeias polipeptídicas e são formadas por no mínimo 40 resíduos cada. Elas se diferem uma das outras apenas na sequência de seus aminoácidos, a qual, por sua vez, é determinada por uma sequência de nucleotídeos em uma porção particular de DNA chamada de *gene*, que decodifica cada proteína específica em um determinado organismo.⁶⁰

Inicialmente, essa sequência de DNA é transcrita para uma sequência de RNA mensageiro – uma molécula quimicamente mais lábil que o DNA e mais facilmente construída e desmontada. Essa sequência de RNA é então traduzida em proteína em conjuntos de três bases nucleotídicas (*códon*), com cada trio decodificando um único aminoácido em um código universal, seguido pelos mecanismos de biossíntese de proteínas em todos os organismos.⁽²³⁾

Existem apenas 22 aminoácidos codificados no genoma, chamados de aminoácidos proteínogênicos, dos quais apenas 20 possuem pelo menos um *códon* específico no código genético (aminoácidos padrão). Os outros dois, pirrolisina e selenocisteína, podem ser incorporados em mecanismos especiais de translação.⁵⁹

Apesar disso, essas 20 possibilidades diferentes para cada resíduo da cadeia polipeptídica produzem 20^n possibilidades de sequências diferentes para uma proteína de n resíduos. Assim, para uma proteína pequena com cerca de 100 resíduos haveria cerca de $20^{100} \approx 1,27 \times 10^{130}$ cadeias polipeptídicas diferentes com esse comprimento, o que representam mais opções do que o número total de estrelas no Universo, que é estimado em 9×10^{78} .⁵⁹

3.1.1 Aminoácidos

Os aminoácidos proteínogênicos compartilham a mesma estrutura comum, representada na Figura 6, que é um átomo de carbono central (carbono- α ou CA) ao qual estão covalentemente ligados um grupo carboxílico, um grupo amino e um átomo de hidrogênio. O quarto grupo, representado por R, é chamado de cadeia lateral e é o que diferencia os 20 aminoácidos um dos outros.⁵⁹

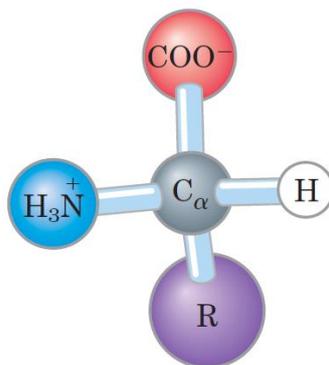


Figura 6 - Estrutura geral de um aminoácido. Adaptado de Nelson et al. (2014).⁶¹

Apesar dos aminoácidos existirem como enantiômeros L e D, na medida em que o carbono- α é um centro assimétrico, todas as proteínas são compostas exclusivamente por L-aminoácidos, o que as torna moléculas intrinsecamente quirais. As estruturas, nomes e abreviaturas dos 20 aminoácidos proteínogênicos estão mostradas na Figura 7, onde os mesmos estão classificados em função da natureza de suas cadeias laterais.⁵⁹

Os aminoácidos alifáticos apresentam cadeias laterais com grupamentos alquila e incluem a glicina, alanina, valina, leucina e isoleucina. O mais simples e o único aminoácido sem centro assimétrico é a glicina, cujo grupo lateral é simplesmente um átomo de hidrogênio, e por isso não se enquadra muito bem em nenhuma categoria. Essa cadeia lateral pequena faz com que a glicina promova o menor impedimento estérico dentre todos os aminoácidos, sendo por isso encontrada em curvas e alças da cadeia polipeptídica.⁶²

A alanina e os aminoácidos de cadeia ramificada, entretanto, apresentam cadeias apolares maiores e mais volumosas, tendo um alto índice hidrofóbico. Isso faz com que, nas proteínas, as cadeias laterais desses aminoácidos tendam a se agrupar para formar núcleos hidrofóbicos.⁶²

A prolina é o único aminoácido alifático cíclico, contendo um anel relativamente rígido de cinco membros entre seu carbono- α e o grupo α -amino, ambos parte do esqueleto peptídico. A restrição da rotação em torno da ligação N- C α limita a participação da prolina nas conformações da cadeia peptídica que ela interrompa a regularidade de elementos da estrutura secundária.⁶²

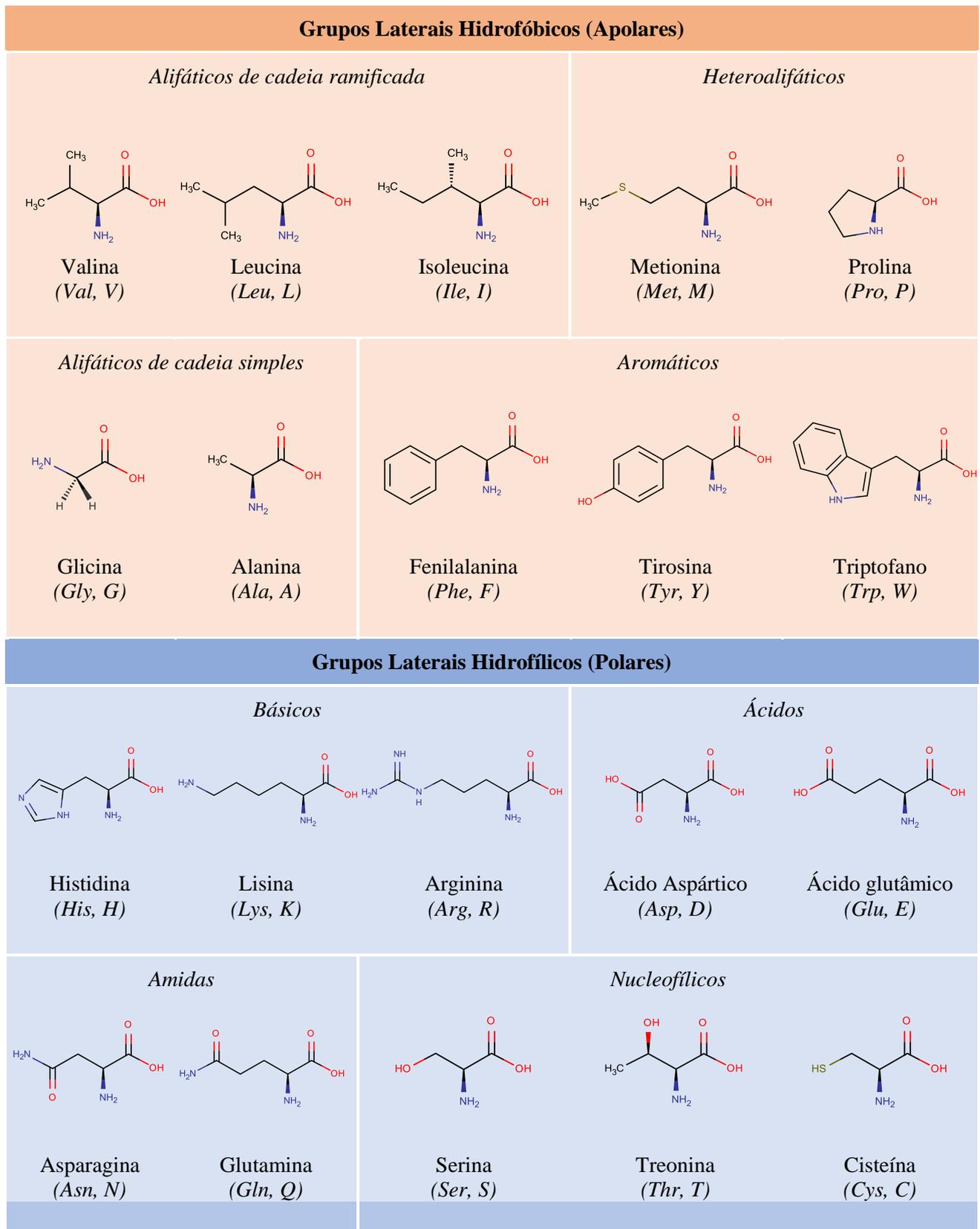


Figura 7 - Os nomes, abreviaturas e estrutura química dos 20 aminoácidos proteinogênicos classificados pela natureza de suas cadeias laterais.

A metionina, apesar de conter um átomo de enxofre, também tem características apolares, graças ao tioéter de sua cadeia lateral, que se assemelha ao *n*-butil em muitas de suas características, inclusive na hidrofobicidade.⁶²

Os aminoácidos aromáticos fenilalanina, tirosina e triptofano contém anéis de com ligações duplas conjugadas, que podem ser anéis benzênicos (Phe, Tyr) ou grupos indol (Trp). Na fenilalanina o anel benzênico não contém substituintes, permitindo o compartilhamento igualitário da densidade eletrônica entre todos os carbonos do anel, o que resulta em uma estrutura altamente hidrofóbica e permite interações do tipo empilhamento- π com outros grupos aromáticos.⁶²

A presença do grupo fenol na cadeia lateral da tirosina a torna mais polar e hidrofílica, na medida em que seu o grupo hidroxila pode participar de ligações de hidrogênio. O mesmo ocorre com nitrogênio aromático do triptofano, motivo pelo qual algumas classificações os colocam numa categoria à parte dentro dos resíduos polares.⁶²

Realmente polares, entretanto, são os resíduos com cadeias laterais amido (asparagina e glutamina) e com grupos hidroxila (serina e treonina), que permitem a estes aminoácidos formarem ligações de hidrogênio com a água, com a cadeia peptídica, uns com os outros ou com compostos no sítio ativo das proteínas. Essa grande hidrofílicidade faz com que tais compostos possam ser comumente encontrados na superfície das proteínas globulares.⁶²

À parte está a cisteína, cuja cadeia lateral contém um grupo tiol, com pK_a para dissociação do seu hidrogênio em torno de 8,4, portanto normalmente sem carga e não-dissociada em pH fisiológico. Entretanto, a cisteína pode formar uma ligação covalente de dissulfeto com outra cisteína pela oxidação espontânea dos seus grupos sulfidril, formando o aminoácido cistina. Em proteínas, essas ligações S-S por manter unidas diferentes cadeias polipeptídicas ou diferentes porções da mesma cadeia.⁶²

Outros aminoácidos polares são os ácidos aspártico e glutâmico, cujas cadeias laterais contém laterais grupos carboxílicos hidrolisáveis, normalmente desprotonados em pH fisiológico, formando aspartato e glutamato, respectivamente, bem como os aminoácidos básicos lisina, arginina e histidina, que contém nitrogênio que podem ser protonados e ficar positivamente carregados em pH neutro ou inferior.⁶²

A histidina contém em sua cadeia lateral um anel imidazol, enquanto que a lisina possui uma amina primária em seu carbono- ϵ e a arginina possui um grupo guanidina. A carga positiva desses aminoácidos permite que eles formem ligações iônicas (interações eletrostáticas fortes) com grupos negativamente carregados, como as cadeias laterais de glutamatos e aspartato, grupos fosfatos de coenzimas e com porções como ácidos carboxílicos de ligantes em sítios ativos.⁶²

No pH fisiológico (7,4), os aminoácidos se encontram na forma de íons dipolares ou zwitteriônicos, com o grupo amino protonado como cátion amônio e o grupo carboxílico desprotonado, formando um ânion carboxilato. Em proteínas, entretanto, como os grupos carboxílicos e amino de dois aminoácidos adjacentes são condensados para formar as ligações peptídicas, apenas o grupo amino do primeiro aminoácido da cadeia – denominado N-terminal – e o grupo carboxílico do último resíduo (C-terminal) apresentam hidrogênios dissociáveis.⁵⁹

Todavia, sete aminoácidos possuem grupos laterais ionizáveis em suas cadeias laterais (Figura 8), os quais podem passar de neutros para negativamente carregados ou então de positivamente carregados para neutros dependendo das condições de pH.⁵⁹

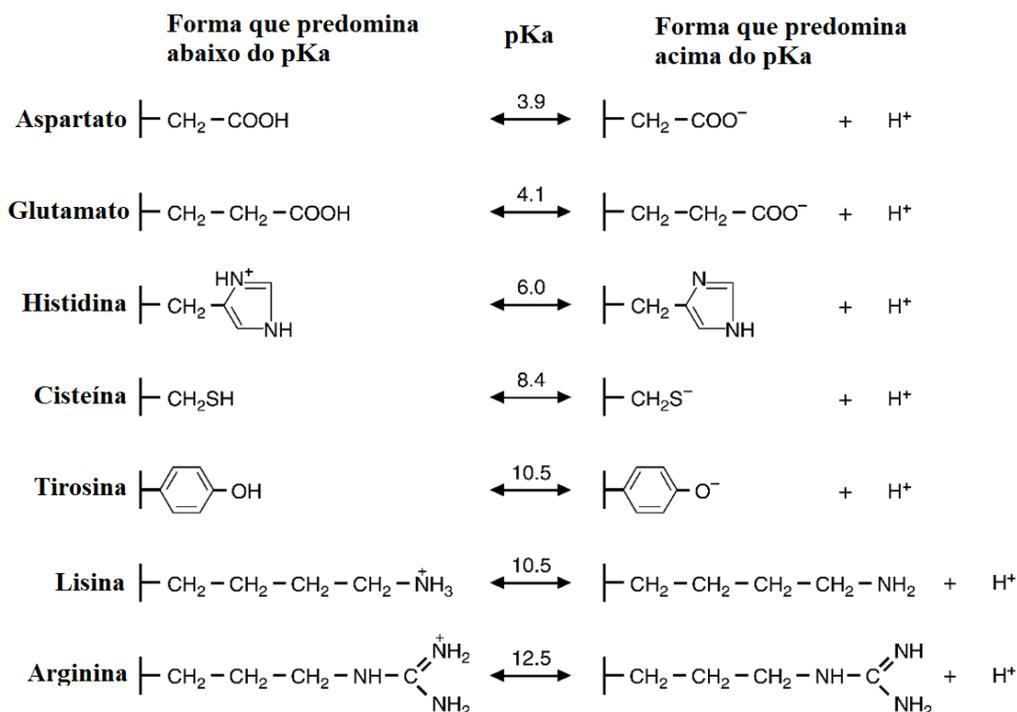


Figura 8 - Estados de protonação predominantes para as cadeias laterais dos sete aminoácidos com grupos laterais ionizáveis e seus respectivos pKas. Adaptado de Smith et al (2005).⁶²

Os ácidos aspártico e glutâmico perdem o hidrogênio do ácido carboxílico de suas cadeias laterais em pH 4 ou superiores, encontrando-se normalmente com carga negativa em condições fisiológicas. A cisteína e a tirosina perdem os prótons de suas cadeias laterais em pH 8.4 e 10.5, respectivamente, permanecendo, portanto, em sua forma neutra nos sistemas biológicos. Arginina e lisina, por sua vez, tem valores de pKa superiores a 10, o que faz com que suas formas protonadas predominem em pH fisiológico, conferindo às cadeias laterais desses aminoácidos cargas positivas.

O grupo imidazol da histidina, por outro lado, tem pKa em torno de 6, relativamente próximo das condições fisiológicas. Assim, ainda que grande parte dos resíduos histidina apresente a cadeia lateral neutra, uma porção não desprezível poderá ser encontrada positivamente carregada em meio fisiológico.

No interior das proteínas, entretanto, em função dos efeitos hidrofóbicos, da formação de ligações de hidrogênio e da proximidade entre grupos carregados, as cadeias laterais podem apresentar pKas muito diferentes daqueles observados para os aminoácidos livres em solução, fenômeno que tem papel muito importante na atuação das proteínas como catalisadores de reações orgânicas em meio biológico.

3.1.2 Organização Estrutural Proteica

A estrutura das proteínas pode ser descrita em termos de quatro níveis de organização (Figura 9). O primeiro deles, denominado de *estrutura primária* é a sequência linear dos aminoácidos na cadeia polipeptídica, a qual é determinada pela sequência de genes no DNA que expressa essa proteína.⁵⁹

Em geral, as proteínas contêm entre 100 e 1000 resíduos de aminoácido sendo o encadeamento desses resíduos o que determina a forma e, em última análise, a função de cada proteína.⁵⁹

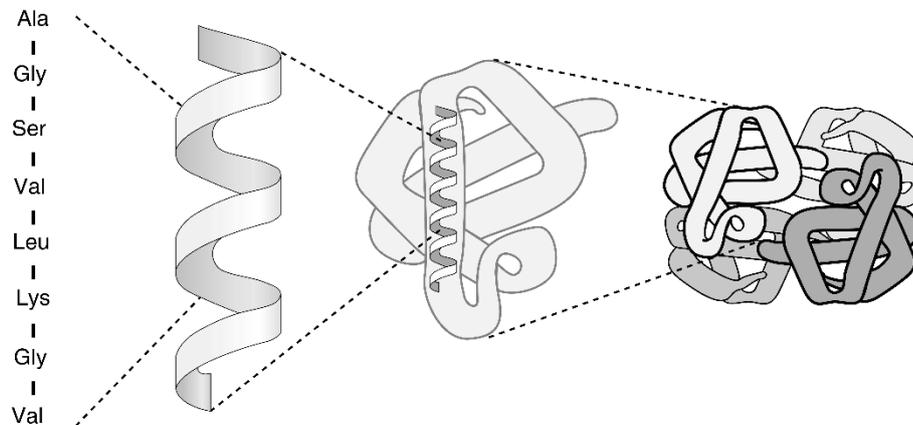


Figura 9 - Os quatro níveis de organização estrutural das proteínas. Adaptado de Smith et al. (2005).⁶²

Quando essa sequência linear de aminoácidos é formada a cadeia polipeptídica se dobra espontaneamente formando a estrutura tridimensional única de cada proteína, chamada de *conformação nativa*. Essa estrutura é flexível, porém, estável e contém sítios para a ligação para outras moléculas, ditando a função que a proteína irá desempenhar no organismo.⁶²

A organização dessa estrutura tridimensional se inicia em grupos pequenos, entre 10 e 20 aminoácidos, que formam estruturas locais estáveis e ordenadas entre si, estabilizadas por meio de ligações de hidrogênio entre os aminoácidos. Esse nível de organização local é chamado de *estrutura secundária* e suas três formas mais comumente observadas são hélices- α , folhas- β e curvas- β .⁶³

As hélices- α são formadas por uma única cadeia peptídica, onde as ligações de hidrogênio ocorrem entre o oxigênio carboxílico de um aminoácido e o N-H de outro, quatro resíduos à frente na cadeia, como mostrado na Figura 10.⁶³

Cada aminoácido forma duas ligações de hidrogênio, o que torna essa unidade estrutural muito estável. Além disso, o formato de hélice faz com que os grupos laterais dos aminoácidos apontem para fora, permitindo outras interações que influenciam a estrutura total da proteína.⁶⁴

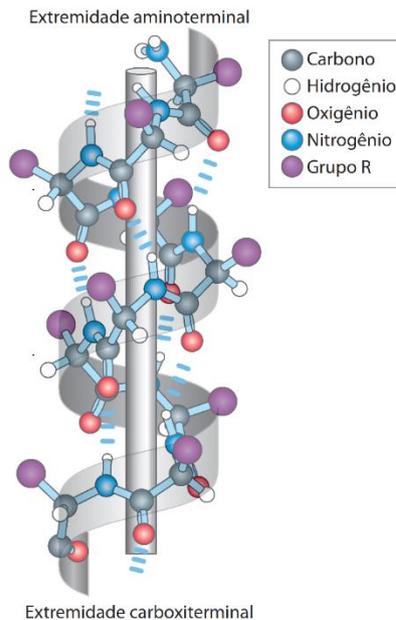


Figura 10- Estrutura de uma hélice- α mostrando o padrão de ligações de hidrogênio entre os aminoácidos que a formam e a orientação das cadeias laterais para o exterior da hélice. Adaptado de Nelson et al (2014).⁶¹

As folhas- β são formadas por duas ou mais cadeia peptídicas unidas por uma série de ligações de hidrogênio entre as cadeias, que podem ser paralelas quando seguem a mesma orientação amino-carbonila ou antiparalelas quando seguem direções opostas (Figura 11, lado esquerdo). Folhas- β antiparalelas acabam sendo mais estáveis devido a um melhor posicionamento das ligações de hidrogênio.⁶⁴

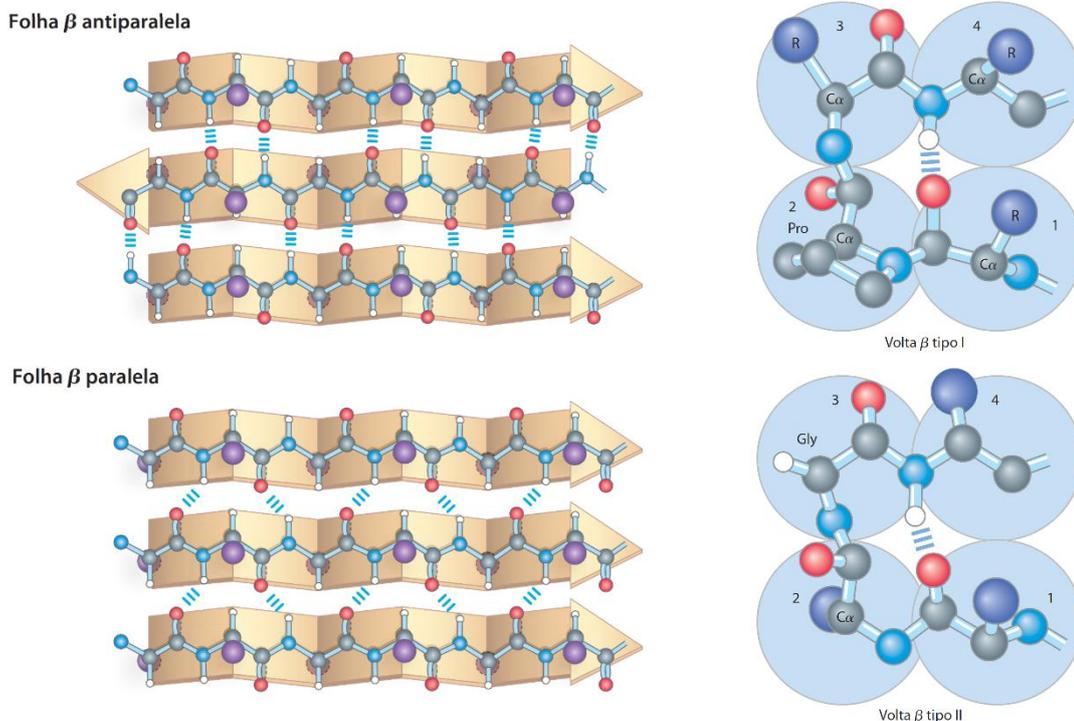


Figura 11- Estruturas e padrões de interações entre folhas- β antiparalelas e paralelas (esquerda) e entre curvas- β , também chamadas de voltas- β , dos tipos I e II (direita). Adaptado de Nelson et al (2014).⁶¹

Curvas- β também chamadas de *voltas reversas*, por sua vez, são estruturas formadas geralmente na porção final de uma folha- β e que levam a uma volta de 180° na direção da cadeia polipeptídica (Figura 11, lado direito). Essas estruturas ocorrem frequentemente na superfície das proteínas e envolvem quatro resíduos, sendo geralmente estabilizadas pela formação de uma ligação de hidrogênio entre o primeiro e o último resíduo. Há dois tipos de curvas- β , que se diferenciam por uma rotação de 180° na ligação peptídica que une os resíduos 2 e 3.⁶⁴

O empacotamento conjunto de diversos elementos de estrutura secundária forma uma conformação tridimensional global estável denominada de *estrutura terciária*. Em proteínas globulares como a AsbF, esse empacotamento está relacionado com o deslocamento dos grupos laterais hidrofóbicos dos aminoácidos para parte interna da proteína e dos grupos laterais hidrofílicos para a superfície da mesma. Dessa forma, há a formação de uma estrutura esférica que maximiza a interação da superfície proteica com a água, permitindo sua solubilidade em meio fisiológico.⁶⁴

A estrutura terciária de proteínas grandes é geralmente descrita em termos de regiões fisicamente independentes chamadas de *domínios estruturais*. Esses domínios compreendem sequências contínuas de aminoácidos da cadeia polipeptídicas que formam grupos de estrutura secundária independentes do resto da proteína, sendo conectados às demais porções proteicas por meio de estruturas simples como laços e alças.⁵⁹

O empacotamento da estrutura secundária para formar essa estrutura tridimensional possibilita a criação de sítios de ligação específicos para apenas uma molécula ou grupos de moléculas com propriedades estruturais similares, definindo a função biológica da proteína.⁶²

Algumas proteínas ainda podem combinar duas ou mais subunidades terciárias para formar uma macroestrutura chamada de *estrutura quaternária*. Esse nível mais elevado de organização pode ser formado por conjuntos da mesma subunidade terciária, que formam homodímeros, trímeros, tetrâmeros ou oligômeros, ou por unidades não idênticas (protômero) associadas de forma geométrica e estequiometricamente específica para formar uma única proteína funcional.⁶²

3.1.3 Aspectos estruturais dinâmicos e enovelamento das proteínas

O arranjo espacial dos átomos em uma proteína ou de parte da proteína é chamado de *conformação*. As conformações possíveis de uma proteína ou de qualquer segmento proteico incluem todos os estados estruturais que ela pode assumir sem a quebra de suas ligações covalentes. Isso inclui, por exemplo, todos os arranjos possíveis que uma proteína pode adotar causados pela rotação de suas ligações simples.⁶¹

Apesar da infinidade de conformações teoricamente possíveis, na medida em que as proteínas possuem normalmente entre centenas e milhares de ligações simples rotáveis, apenas uma ou algumas poucas predominam em condições biológicas, as chamadas *conformações nativas*. Essas conformações são normalmente as mais estáveis termodinamicamente, isto é, possuem as menores energias livres de Gibbs (G) dentre todas as conformações possíveis nas condições fisiológicas.⁶⁰

A diferença de energia livre de Gibbs (ΔG) que separa o estado nativo da forma desnaturada (totalmente desenovelada) da proteína, entretanto, é da ordem de apenas 20 a 65 kJ/mol, tornando a conformação nativa por pouco mais estável que a forma não dobrada.⁵⁹

Isso se deve ao fato da cadeia polipeptídica poder assumir, teoricamente, inúmeras conformações, o que confere ao estado não dobrado de uma proteína um alto grau de entropia conformacional. Essa entropia, junto com a formação de ligações de hidrogênio entre os diversos grupos da cadeia polipeptídica com a água, estabilizam o estado desnaturado.⁵⁹

Contrabalanceado esses efeitos e estabilizando as conformações nativas estão a formação de ligações dissulfeto (covalentes) e interações não-covalentes como as ligações de hidrogênio, interações hidrofóbicas e iônicas entre as cadeias no estado enovelado.⁵⁹

O fato da conformação nativa ser apenas um pouco mais estável que a forma desnaturada evidencia a existência de um balanço tênue entre a rigidez – que é essencial para a criação e manutenção dos sítios de ligação e para evitar que a estrutura se desfaça pela ação entrópica – e a flexibilidade e mobilidade estrutural, que permite à proteína se adaptar para o encaixe de outras proteínas e de pequenas moléculas, como substratos e cofatores e que é imprescindível para a execução apropriada de suas funções.⁶¹

Em condições adequadas, as proteínas desnaturadas podem se redobrar de volta à conformação nativa, readquirindo sua função original, o que mostra que o padrão de enovelamento de uma proteína depende essencialmente de sua estrutura primária.⁶²

Por outro lado, supondo-se que uma proteína seja formada por 100 resíduos de aminoácido, cada um podendo assumir, em média, 10 conformações distintas, números relativamente baixo quando comparados com a realidade, existiriam pelo menos 10^{100} conformações diferentes para a cadeia polipeptídica.⁵⁹

Se o dobramento da proteína ocorresse seguindo um processo aleatório no qual ela testasse todas as conformações possíveis em torno de cada uma das ligações de seu esqueleto até encontrar sua forma nativa, ainda que cada conformação fosse testada no menor tempo possível (10^{-13} segundos, que é o tempo necessário de uma única vibração molecular), levaria aproximadamente 10^{77} anos para testar todas as conformações possíveis, o que é mais do que toda a história do Universo.⁵⁹

Esse problema, apontado pela primeira vez por Cyrus Levinthal, em 1968, mostra que o enovelamento de proteínas não é um processo completamente aleatório, de tentativa e erro, especialmente quando se considera que muitos polipeptídeos adquirem estruturas semelhantes à nativa em frações de milissegundos depois do início do processo de dobramento.⁵⁹

Ao contrário, à medida que as proteínas se dobram, o espaço conformacional que precisa ser explorado pela estrutura vai ficando cada vez mais restrito. Esse espaço conformacional pode ser interpretado através de uma superfície de energia livre em formato de funil, com o ΔG representado como a profundidade e com a estrutura nativa (N) sendo o fundo desse funil, o ponto de menor energia-livre (Figura 12).⁶¹

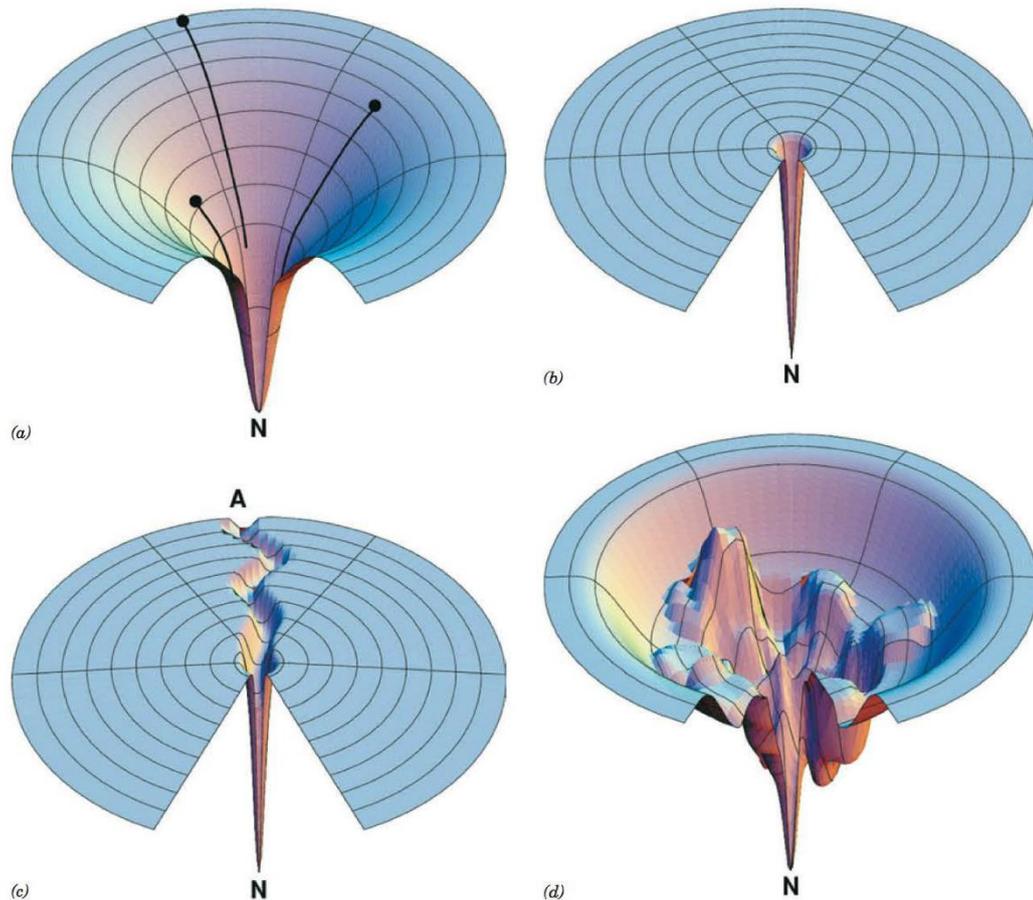


Figura 12- Termodinâmica do enovelamento proteico, mostrado como um funil de energia livre com diferentes perfis e mecanismos de dobramento (a-d). Adaptado de Voet et al (2013).⁵⁹

Os estados não dobrados apresentam um alto grau de entropia conformacional e, por consequência, uma energia livre relativamente alta. No processo de enovelamento, à medida em que proteína se aproxima de seu estado nativo, a entropia e a energia livre diminuem juntamente com o espaço conformacional que necessita ser amostrado até o estado nativo ser atingido.⁵⁹

Em virtude dessa superfície energética afunilada, não há uma única rota a ser seguida para que um polipeptídeo alcance a sua conformação nativa. Na medida em que cada polipeptídeo não dobrado adota uma conformação diferente em solução, o que energeticamente se traduz como uma localização diferente dentro do funil de dobramento (Figura 12A), cada um pode apresentar um conjunto de ajustes conformacionais distintos que reduzem sua energia livre e sua entropia até que o estado nativo comum seja atingido, isto é, até chegar no fundo do funil.⁶¹

Se um polipeptídeo realmente se enovelasse até seu estado nativo por meio da procura conformacional aleatória, a superfície de energia se assemelharia a um disco achatado com um único e pequeno orifício, tal qual a superfície de um campo de golfe (Figura 12B). Uma procura conformacional aleatória nessa superfície seria tão demorada (e ineficiente) quanto tentar acertar uma bola de golfe em seu buraco com os olhos vendados.⁶¹

Os funis podem ter uma variedade de formas dependendo da complexidade das rotas de enovelamento, da existência de conformações intermediárias semiestáveis ao longo dessas rotas, e do potencial desses intermediários formarem agregados de proteínas erroneamente dobradas.⁶³

Isso faz com que a superfície do funil de enovelamento da maioria das proteínas não seja necessariamente suave, como representado na Figura 12A, na qual um funil amplo e suave representa uma proteína com muitas rotas de enovelamento, porém nenhum intermediário significativamente estável.⁶³

Uma proteína com apenas uma ou com poucas rotas de enovelamento produtivas, por outro lado, seria mostrada como um funil com uma depressão estreita levando à forma nativa e sem intermediários significativos (Figura 12C).⁶³

Entretanto, o mais comum é que haja pequenas depressões ao longo das paredes do funil, que representam conformações semiestáveis que o polipeptídeo pode adotar ao longo do dobramento, como mostrado na Figura 12D.⁶³

Nessa situação, a proteína apresentaria intermediários de enovelamento com estabilidade substancial em praticamente qualquer via que leva ao estado nativo, ou seja, haveria um domínio particular sempre se dobra rapidamente, mas as outras partes da proteína se dobram mais lentamente e em ordem aleatória, o que é representado por um funil com uma depressão principal ao redor da depressão que leva à forma nativa.⁶³

3.2 ENZIMAS

O processamento contínuo de energia e informação dentro das células – a definição biológica de vida - consiste em milhares de reações químicas individuais, que para serem fisiologicamente úteis necessitam ocorrer em velocidades que se enquadrem às necessidades celulares e serem específicas, de tal forma que um determinado reagente sempre produza um produto em particular, minimizando-se a formação de subprodutos nocivos ou inutilizáveis. Por essa razão, os organismos vivos desenvolveram um grupo especial de moléculas capazes de aperfeiçoar a eficiência e seletividade dessas reações e impor ordem a este “caos” bioquímico intracelular, as *enzimas*.⁶³

De uma forma geral, enzimas podem ser definidas como uma classe de proteínas que catalisam reações bioquímicas. Essas macromoléculas são notáveis não apenas por sua alta complexidade estrutural tridimensional, resultado de um dobramento espontâneo de uma cadeia polipeptídica linear, mas principalmente por suas propriedades catalíticas impressionantes e muito superiores aos melhores catalisadores já desenvolvidos pelo Homem, muitas vezes demonstrando alta enantiosseletividade e enantioespecificidade com taxas de reações que se aproximam da “perfeição catalítica”.⁶⁴

As taxas de reação enzimáticas são normalmente de 10^6 a 10^{12} vezes maiores do que as taxas observadas para reações não catalisadas e várias ordens de magnitude mais rápidas que reações catalisadas quimicamente, mesmo ocorrendo em condições reacionais muito mais brandas que estas, como temperaturas inferiores a 100°C, pH neutro e pressão atmosférica.⁶⁴

Além disso, enzimas apresentam grande capacidade de controle, na medida em que respondem a variações de concentração de outros compostos que não sejam reagentes e produtos, e raramente geram produtos secundários, sendo, portanto específicas tanto na reação que catalisam quanto nos reagentes que necessitam (substratos).⁶³

3.2.1 Catálise enzimática

Nas condições biológicas, as reações não catalisadas tendem a ser lentas pois a maioria das moléculas biológicas é muito estável nas condições internas das células com pH neutro, temperaturas amenas e ambiente aquoso. Além disso, muitos processos químicos corriqueiros, como a formação transitória de intermediários instáveis carregados ou a colisão de duas ou mais moléculas exatamente na orientação exata necessária para que as reações ocorram, são desfavoráveis ou improváveis no ambiente celular.⁶²

As enzimas contornam esses problemas ao proporcionarem um ambiente específico adequado para que uma dada reação possa ocorrer mais rapidamente. A propriedade característica das reações catalisadas por enzimas é que a reação ocorre confinada em um bolsão da enzima denominado sítio ativo. A molécula que liga no sítio ativo e sobre a qual a enzima age é denominada *substrato*.⁶⁴

Uma reação enzimática simples pode ser escrita em termos de três etapas básicas que são i) a ligação do substrato S à enzima E, formando o complexo enzima-substrato (ES); ii) conversão do substrato à produto, formando o complexo transitório enzima-produto (EP); e iii) liberação do produto P, com o retorno da enzima ao seu estado original (conformação nativa) para iniciar um novo ciclo catalítico.⁶²

Praticamente todas as reações químicas possuem uma barreira de energia que separa os reagentes e os produtos. Essa barreira, chamada de *energia livre de ativação*, é a diferença de energia livre que existe entre reagentes e um intermediário de alta energia que ocorre durante a formação do produto, chamado de *estado de transição*. Sem essas barreiras energéticas, as macromoléculas poderiam reverter espontaneamente para formas moleculares mais simples, e as estruturas complexas e altamente ordenadas bem como os processos metabólicos das células não poderiam existir.⁶¹

A Figura 13 apresentam um diagrama de energia em função da coordenada de reação que mostra as mudanças na energia livre que ocorrem durante a conversão de uma molécula do substrato S para o produto P à medida que a reação avança, passando pelo estado de transição (intermediário de alta energia). Devido à alta energia livre de ativação, as taxas de reações químicas não catalisadas são frequentemente lentas pois, para que as moléculas reajam, elas devem conter energia suficiente para superar a barreira de energia do estado de transição.⁶¹

Na ausência de uma enzima, apenas uma pequena porção de moléculas possui energia suficiente para atingir (e superar) a barreira energética do estado de transição entre reagente e produto. Em geral, quanto menor a energia livre de ativação, mais moléculas possuem energia suficiente para passar pelo estado de transição e, portanto, maior a velocidade da reação.⁶¹

Durante o curso da evolução, as enzimas desenvolveram-se para diminuir seletivamente as barreiras energéticas das reações necessárias para a sobrevivência celular, fornecendo um caminho de reação alternativo com uma menor energia livre de ativação. A enzima não altera as energias livres dos reagentes ou produtos e, portanto, não altera o equilíbrio da reação, mas sim acelera a taxa com a qual o equilíbrio é alcançado.⁶¹

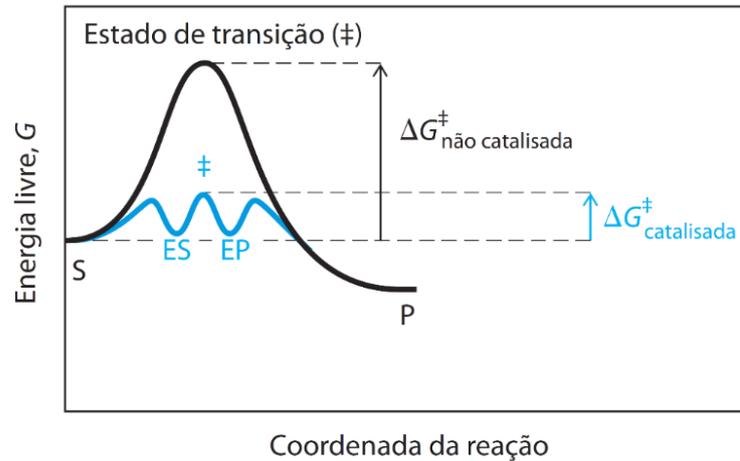


Figura 13- Diagrama de coordenada de reação, comparando uma reação catalisada por enzima com uma reação não-catalisada. Adaptado de Nelson et al (2014).⁶¹

A formação desse caminho de reação alternativo pelas enzimas está relacionada com mecanismos como o efeito de proximidade e orientação, aumento da tensão conformacional do substrato e à estabilização do estado de transição.⁶⁴

Muitas reações enzimáticas ocorrem entre substratos e grupos funcionais da enzima (grupos laterais de aminoácidos, íons metálicos e coenzimas), os quais geralmente se encontram numa porção do interior da enzima denominada *sítio ativo* da enzima, que possui resíduos que se ligam o substrato, englobando-o e, por vezes, até o sequestrando completamente da solução.⁶²

Esses grupos funcionais catalíticos mantêm-se próximos do substrato e em uma geometria ideal para que a reação ocorra, o que reduz a entropia dos reagentes (o substrato e os grupos catalíticos da enzima) na medida em que os coloca próximos e em uma orientação adequada, de forma análoga ao que ocorre no aumento da concentração dos reagentes.⁶⁴

Além disso, esses grupos catalíticos podem formar ligações covalentes transitórias com um substrato e ativá-lo para a reação, ou um grupo pode ser transitoriamente transferido do substrato para a enzima. Essas interações covalentes entre enzimas e substratos diminuem a energia de ativação na medida em que promovem o aumento da energia livre do substrato, atenuando a altura relativa da barreira energética.⁶²

Outro aspecto importante é que os sítios ativos das enzimas são complementares não aos substratos por si mesmos, mas aos estados de transição pelos quais os substratos passam ao serem convertidos em produtos durante a reação enzimática.⁵⁹

Na medida em que o substrato se liga no sítio ativo, as mesmas forças que estabilizam a estrutura das proteínas, incluindo ligações de hidrogênio, interações hidrofóbicas e iônicas, promovem a ligação entre a enzima e o substrato para formação do complexo ES. A formação de cada interação não-covalentes no complexo ES é acompanhada pela liberação de uma pequena quantidade de energia livre, denominada energia de ligação (ΔG_{lig}).⁵⁹

A liberação dessa energia de ligação estabiliza a estrutura desse complexo ao mesmo tempo em que promove alterações conformacionais, não apenas no sítio ativo, mas também nas alterações na estrutura global da enzima que força o substrato a adotar uma conformação mais próxima à do estado de transição, o que faz com que a diferença energética entre o estado ligado ES e o estado de transição seja reduzida, acelerando a reação.⁶¹

3.2.2 O Sítio Ativo

Apesar de quase todas as enzimas serem proteínas, nem todas as proteínas são enzimas. A diferença reside na atividade catalítica que as enzimas possuem e que está relacionado com uma parte da estrutura terciária denominada *sítio ativo*, que corresponde a 10 e 20 % do seu volume total, onde a “química enzimática” acontece.⁶⁴

O sítio ativo é geralmente uma fenda ou cavidade hidrofóbica que contém um arranjo de grupos laterais de aminoácidos capazes de se ligar ao substrato e realizar a reação enzimática. Os resíduos de aminoácidos que se ligam ao substrato podem vir de partes muito diferentes da sequência de aminoácidos linear da enzima.⁶²

Em alguns casos, o sítio ativo também contém um ou mais cofatores que podem ser íons metálicos, como Mn^{2+} e Mg^{2+} , complexos metálicos ou compostos orgânicos não proteicos cujo papel é auxiliar a catálise enzimática.⁵⁹

Este sítio não é um receptáculo passivo para ligar o substrato, mas uma máquina molecular complexa com uma diversidade de mecanismos químicos que auxiliam a conversão do substrato ao produto.⁶⁰

Uma série de fatores responsáveis pela eficiência catalítica das enzimas estão relacionados diretamente com aspectos estruturais do sítio ativo. Um dos principais é a estabilização de estado de transição, na medida em que esse sítio geralmente atua como um molde molecular flexível que liga o substrato ao mesmo tempo em que o induz a alterações conformacionais e eletrônicas forçando-o a adotar a estrutura do estado de transição, ao mesmo tempo em que estabiliza esse estado intermediário. Como consequência desse processo, há o aumento do número de moléculas que são capazes de atingir o estado de transição e serem convertidas em produto, aumentando a velocidade da reação.⁶⁴

Além disso, o sítio ativo também pode fornecer grupos catalíticos que aumentam a probabilidade de o estado de transição se formar. Em algumas enzimas, estes grupos podem participar da catálise geral ácido-base, em que os resíduos de aminoácidos fornecem ou aceitam prótons. Em outras enzimas, a catálise pode envolver a formação transitória de um complexo ES covalentemente ligado, como é o caso das quimotripsina, uma enzima que digere proteínas no intestino que contém em seu sítio ativo uma serina que se liga de forma covalente com o substrato em uma das etapas do mecanismo enzimático.⁶⁴

3.2.3 Especificidade Enzima-Substrato

A *especificidade* enzimática, isto é, a capacidade da enzima de reagir com apenas um substrato, resulta da disposição tridimensional de resíduos de aminoácidos específicos na enzima que formam sítios de ligação para os substratos e que os ativam durante o curso da reação.⁶²

A alta seletividade ao substrato – por vezes até enantiosseletividade – está relacionada com a natureza quiral do sítio ativo, que é capaz de executar uma série de interações não covalentes específicas entre enzima e substrato.⁶⁴

Ao longo do tempo, diversos modelos foram propostos para racionalizar a forma como essa interação específica enzima-substrato ocorre, dentre os quais dois se destacam. O primeiro deles, chamado de *modelo chave-fechadura* foi proposto por Emil Fischer em 1894.⁶²

Segundo esse modelo, as enzimas seriam estruturalmente complementares aos seus substratos de modo a se encaixarem como uma chave em uma fechadura. Essa interação específica (e, portanto, exclusiva) entre as duas moléculas biológicas seria mediada por superfícies moleculares com formas complementares.⁵⁹

O sítio de ligação do substrato conteria resíduos de aminoácidos dispostos em uma superfície tridimensional complementar que "reconhece" o substrato e o liga através de múltiplas interações hidrofóbicas, interações eletrostáticas ou ligações de hidrogênio. A ligação de compostos com uma estrutura que difere do substrato até mesmo em um pequeno grau seria evitada por impedimento estérico e repulsão de carga.⁵⁹

Por outro lado, uma enzima totalmente complementar ao seu substrato seria uma enzima muito pobre, na medida em que estabilizaria o substrato em vez de desestabilizá-lo. No diagrama de coordenadas da reação, esse tipo de complexo ES corresponderia a um poço de energia livre da qual o substrato teria dificuldade de escapar.⁶¹

A noção moderna da catálise enzimática, primeiramente proposta por Michael Polanyi (1921) e Haldane (1930), foi elaborada por Linus Pauling em 1946 e por William P. Jencks na década de 1970 e afirma que para poder catalisar reações, as enzimas devem ser complementares ao estado de transição da reação. Isso significa que interações ótimas entre substratos e enzimas só ocorrem no estado de transição.⁶¹

Assim, no complexo ES inicial haveria a formação de interações fracas, com a complementaridade total entre o substrato e a enzima ocorrendo apenas quando o substrato se encontrasse no estado de transição. A energia livre de ligação liberada durante a formação dessas interações compensaria parcialmente a energia necessária para que o substrato atingisse o estado de transição, no topo da curva de energia, tornando a barreira energética a ser vencida é muito menor do que para a reação não catalisada.⁶¹

Dessa forma, enzima também sofreria mudanças de conformação quando o substrato se ligasse a ela, induzidas por essas múltiplas interações fracas com o substrato que, em última análise, levariam o substrato a atingir estado de transição. Esse mecanismo de interação enzima-substrato é chamado de *encaixe-induzido* e foi postulado por Daniel Koshland em 1958.⁵⁹

Esses movimentos podem afetar apenas uma pequena parte da enzima nas proximidades do sítio ativo ou podem envolver mudanças no posicionamento de domínios inteiros da enzima. Em geral, uma rede de movimentos acoplados ocorre por toda a enzima, que finalmente leva às mudanças necessárias no sítio ativo. O encaixe-induzido serve para levar grupos funcionais específicos da enzima para uma posição apropriada para catalisar a reação. As mudanças conformacionais também permitem a formação de ligações fracas adicionais no estado de transição. Em ambos os casos, a nova conformação da enzima apresenta propriedades catalíticas aumentadas.⁶¹

Mais recentemente, um mecanismo alternativo foi proposto, a chamada "*seleção conformacional e mudança de população*". Esta proposição se baseia em conceitos derivados da paisagem de energia livre e sugere que as proteínas existem em solução como conjuntos amplos de diferentes conformações. Entre os estados conformacionais presentes nesse conjunto (ou

ensemble de conformações) deve haver alguns no qual o sítio de ligação corresponde à forma (e à química) do ligante.⁶⁵

Entretanto, a energia conformacional desses estados pode ser alta e eles podem estar escassamente povoados em solução. A aproximação do ligante, com a consequente formação do complexo proteína-ligante e liberação da energia livre de ligação teria como consequência a estabilização ou dessas conformações ativas, induzindo a uma mudança na população dessas estruturas em relação aos demais confôrmeros do *ensemble*. Assim, a atuação do ligante seria em selecionar e estabilizar uma conformação em detrimento às demais.⁶⁶

Atualmente, acredita-se que a seleção conformacional seja o mecanismo prevalecente, com encaixe-induzido dominando os casos onde a concentração do ligante é extremamente alta. O mais provável, entretanto, é que haja um efeito sinérgico entre esses dois mecanismos, com a seleção de conformações pelo ligante, por um lado, e a indução de conformações que maximizem as interações e estabilizem esse confôrmero, por outro.⁶⁶

3.2.4 Mecanismos de Catalise Enzimática

Uma vez tendo complexado o substrato, a enzima então realiza a catálise da sua reação específica por meio de grupos catalíticos do sítio ativo e finalmente libera seu produto de volta à solução. Três tipos principais de mecanismos podem ser definidos para a compreensão da ação enzimática: catálise ácido-base, catálise covalente ou nucleofílica e catálise eletrostática.⁵⁹

A *catálise ácido-básica* está envolvida em todos os processos enzimáticos que envolvem transferência de prótons, sendo poucas as enzimas que não possuem grupos ácidos ou básicos cataliticamente ativos. A catálise ácida envolve a ativação do substrato por meio da sua protonação por um resíduo cataliticamente ativo, enquanto que na catálise básica o substrato é desprotonado ou então água é desprotonada antes de atacar o substrato (Figura 14).⁶⁴

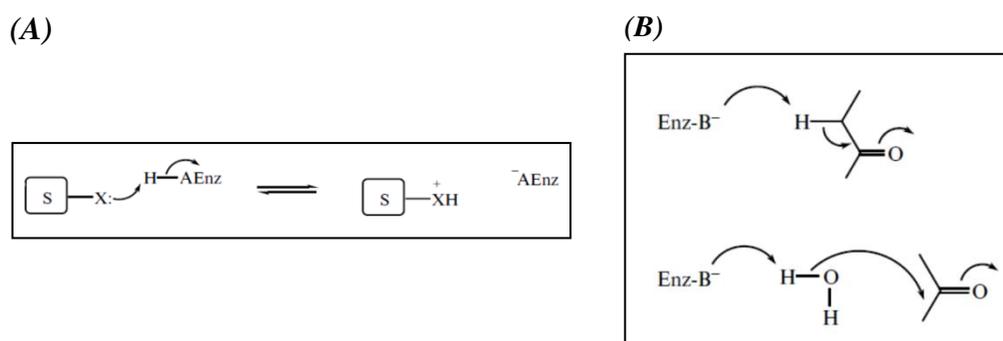


Figura 14- Mecanismo geral da catálise enzimática ácida (a) e básica (b1-2). Adaptado de Bugg (2012).⁶⁴

Apesar do pequeno número de aminoácidos com grupos laterais ionizáveis e de operar em condições fisiológicas, uma ampla faixa de pKas é obtida para os resíduos o sítio ativo graças ao microambiente criado pela estrutura tridimensional da enzima. Assim, por exemplo, a aproximação de um resíduo positivamente carregado a uma amina torna sua forma protonada

termodinamicamente menos favorável e reduz seu pKa, permitindo que esta possa atuar como uma base sobre o substrato, mas como um ácido sobre a água, mediando a transferência de prótons. Ou ainda colocando resíduos cataliticamente ativos em regiões hidrofóbicas, o que desfavorece sua forma eletricamente carregada, alterando drasticamente seu pKa.⁶³

A *catálise covalente* envolve o ataque nucleofílico de um resíduo do sítio ativo ao substrato, formando uma ligação covalente entre a enzima e o substrato, e, portanto, um intermediário covalente no mecanismo da reação, como exemplificado no mecanismo da acetoacetato decarboxilase (Figura 15). O nucleófilo mais poderoso disponível ao sítio ativo enzimático é o grupo tiol da cisteínas, que pode ser desprotonado para formar o ânion tiolato, ainda mais nucleofílico, seguindo do grupo ε-amino de lisinas, muito utilizado para formação de ligações imina com grupos cetona de substratos, além de oxigênios nucleofílicos de serinas, treoninas, tirosinas e os resíduos carboxilados.⁶⁴

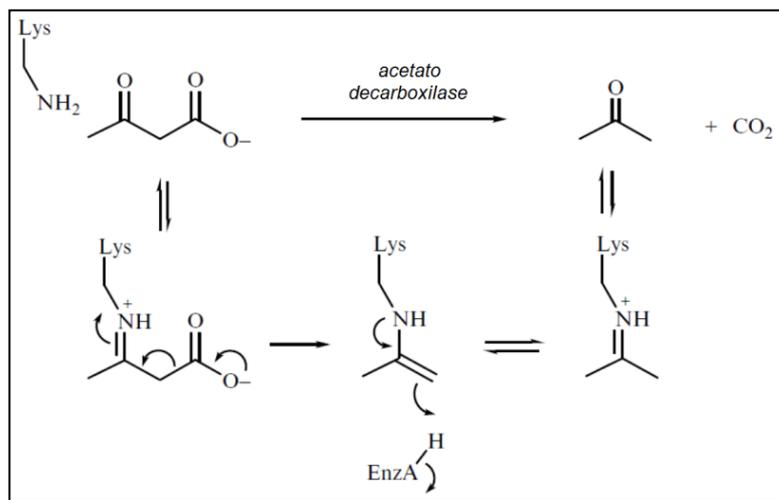


Figura 15 - Mecanismo da enzima acetoacetato decarboxilase ilustrando como se processa a catálise enzimática nucleofílica ou covalente. Adaptado de Bugg (2012).⁶⁴

A *catálise eletrostática* está relacionada com a estabilização de intermediários carregados por meio da formação de ligações iônicas com resíduos eletricamente carregados ou cofatores metálicos como zinco e magnésio. As propriedades de ácidos de Lewis de íons metálicos têm especial aplicação enzimática na medida em que, sendo aceptores de elétrons, permitem a ativação de substratos ou mesmo da água, cujo pKa pode ser reduzido a ponto de atuar como um nucleófilo, como o que ocorre na enzima termolisina (Figura 16).

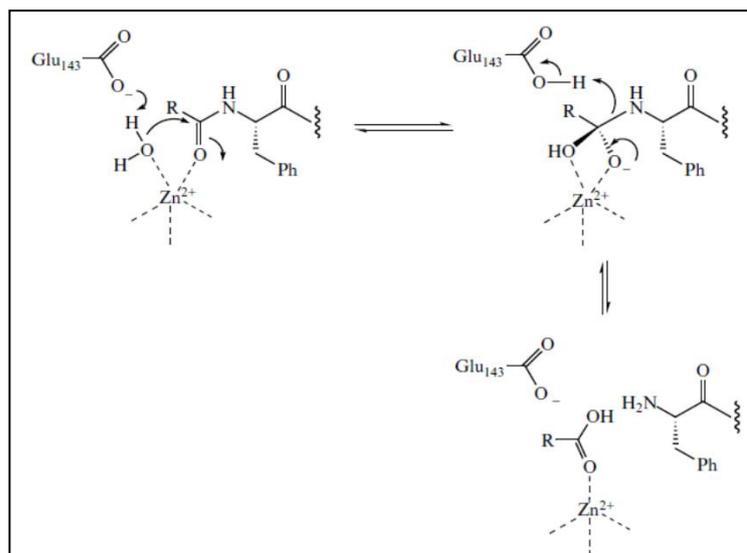


Figura 16 - Mecanismo de atuação de termolisina, demonstrando a atuação de um cofator eletricamente carregado na catálise enzimática. Adaptado de Bugg (2012) ⁶⁴

3.2.5 Inibição Enzimática

A inibição enzimática acontece quando uma molécula diferente do substrato se liga ao sítio ativo e previne a ocorrência da reação enzimática. Há dois tipos principais de inibidores normalmente observados: os *reversíveis*, compostos que se ligam de forma não covalente ao sítio ativo ou outra porção da enzima, impedindo o substrato de se coordenar e os inibidores *irreversíveis*, que se ligam permanentemente à enzima por meio da reação com algum resíduo do sítio ativo formando uma ligação covalente. O sítio ativo é então permanentemente bloqueado e a enzima irreversivelmente desativada. ⁶³

Esse tipo de inibidor geralmente contém grupos funcionais eletrofílicos como substituintes halogênio ou epóxido que sofrem um ataque nucleofílico de algum resíduo do sítio ativo (catálise enzimática nucleofílica) produzindo um complexo energeticamente mais estável e levando à modificação covalente permanente desse resíduo. ⁶³

Os inibidores reversíveis – mais comumente observados – podem ser de três tipos principais: competitivo, anti-competitivo e não competitivo, os quais variam em função da natureza da interação com a enzima e do efeito que produzem na cinética enzimática (Figura 17)

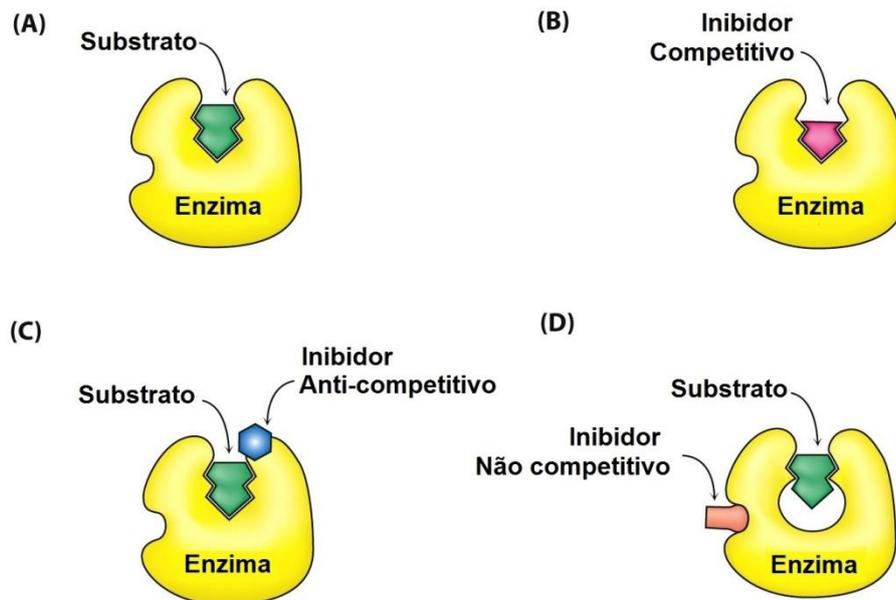


Figura 17 - Esquema da interação enzima-substrato (a) e da ação de inibidores competitivos clássicos (b), anti-competitivos (b) e não-competitivos (c). Adaptado de Berg et al (2010).⁶³

Os *inibidores competitivos* (Figura 17B) são moléculas que se assemelham estruturalmente ao substrato (S) e se que ligam ao sítio ativo da enzima (E) impedindo que o substrato possa interagir e de fato competindo com este pelo sítio ativo. Nessa situação, a enzima pode se ligar ao substrato, formando o complexo ES ou se ligar ao inibidor (complexo EI) mas não pode se ligar aos dois ao mesmo tempo (ESI).⁶⁴

Esses inibidores reduzem a velocidade da reação enzimática por reduzirem a proporção de moléculas de enzima ligadas ao substrato. Assim, a ação desses inibidores pode ser desfeita aumentando-se a concentração de substrato, o que favorece a formação do complexo ES (e conseqüentemente a reação enzimática) por deslocar o equilíbrio termodinâmico a favor dessa reação.⁶⁴

Inibidores anti-competitivos se distinguem pelo fato do inibidor se ligar apenas ao complexo enzima-substrato e não à enzima sozinha. Para esse tipo de inibidor, o sítio de ligação é criado apenas quando a enzima se liga ao substrato (Figura 17C) e a inibição não pode ser superada pela adição de mais substrato.⁶³

Inibidores não-competitivos, por sua vez, podem se ligar tanto à enzima (SI) quanto ao complexo enzima-substrato (ESI), na medida em que o inibidor e o substrato interagem em porções distintas da enzima (Figura 17D). Esse tipo de inibidor atua reduzindo a número geral de moléculas de enzima ativas ao invés de apenas diminuir a proporção de enzima ligada ao substrato e por essa razão a inibição não diminui com o aumento da concentração do substrato.⁶³

Tendo em vista o papel primordial que as enzimas desempenham no controle dos processos metabólicos e nos mecanismos de biossíntese celular, não é surpresa que mais de 47% de todas as drogas atualmente comercializadas tenham como alvo a inibição dessa classe de biomoléculas, segundo Copeland (2013), a exemplo do ibuprofeno e do ácido acetilsalicílico, dois dos medicamentos mais consumidos do mundo os quais atuam na inibição de ciclooxigenase, enzima responsável pelo estabelecimento de respostas inflamatórias.⁶⁷

Inibidores enzimáticos também se destacam pela aplicação como antibióticos, sobretudo devido à existência de enzimas essenciais em bactérias que não possuem homólogos em organismos superiores, limitando os possíveis efeitos colaterais desses compostos.⁶⁷

Um dos primeiros exemplos de inibidor enzimático com ação antibiótica foi a sulfanilamida, composto estruturalmente similar ao ácido *p*-aminobenzoico (PABA), um metabólito requerido por bactérias na biossíntese do ácido fólico, composto essencial em diversas funções bioquímicas. A sulfanilamida atua como inibidor competitivo da enzima que metaboliza PABA, inviabilizando a célula bacteriana, mas sem qualquer efeito sobre humanos que adquirem ácido fólico pela dieta exclusivamente.⁶⁷

3.2.6 Afinidade Enzima-Ligante e Energia Livre

No desenho de fármacos, sobretudo os inibidores enzimáticos, é de extrema importância compreender qual é a afinidade do ligante pela proteína com a qual interage e ter em mente que essa afinidade é definida pelo equilíbrio entre o ligante não-ligado e a proteína não-ligada de um lado e o complexo proteína-ligante do outro, como mostrado na Figura 18.⁶⁸

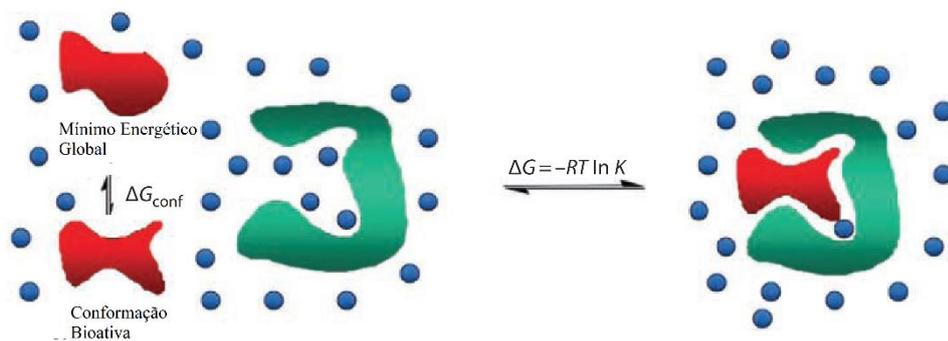


Figura 18 - O equilíbrio químico que determina a afinidade proteína-ligante. Adaptado de Larsen et al. (2010).⁶⁹

Este é, do ponto de vista termodinâmico, o modelo mais simples para a ligação de uma pequena molécula (substrato ou ligante) com um receptor macromolecular biológico (enzima), isto é, uma associação reversível, na qual o ligante e o receptor formam um complexo equimolecular não-covalente que pode ser expressa pela seguinte equação de equilíbrio:



Os termos E, S e ES correspondem à enzima, substrato e complexo respectivamente, k_{dir} é a constante de velocidade de formação e k_{inv} a constante de velocidade de dissociação do complexo.⁷⁰

Além disso, as constantes de velocidade da reação estão relacionadas às constantes termodinâmicas de associação (K_A) e dissociação (K_D) segundo:

$$K_A = \frac{1}{K_D} = \frac{k_{dir}}{k_{inv}} = \frac{[ES]}{[E][S]} \quad Eq. 2$$

Assim, a termodinâmica governa os princípios físico-químicos básicos do reconhecimento molecular. A afinidade do ligante pela proteína é uma função de estado e é dada pela diferença de energia livre (ΔG) entre o complexo proteína-ligante (o lado direito na Figura 18) e o ligante e a proteína "livres", isto é, não-ligados (o lado esquerdo).⁶⁹

Essa diferença de energia livre de Gibbs de ligação está relacionada com as constantes termodinâmicas de associação e dissociação de acordo com:

$$\Delta G^\circ = -RT \ln K_A = RT \ln K_D \quad Eq. 3$$

Nessa equação, R é a constante universal dos gases ($8,314 \text{ J.K}^{-1}.\text{mol}^{-1}$) e T é a temperatura em Kelvin. Uma maior afinidade implica um maior valor positivo de K_A e, portanto, um valor negativo maior de ΔG .

A constante de dissociação K_D é comumente utilizada como uma medida da força de ligação de um determinado ligante, sendo tão menor quanto mais efetiva for a interação desse ligante com o receptor. Quando esse ligante atua como inibidor (I), essa constante de dissociação recebe o nome de constante de inibição (K_I).⁷⁰ Com isso:

$$\Delta G^\circ = RT \ln K_I = RT \ln \left\{ \frac{[E][I]}{[EI]} \right\} \quad Eq. 4$$

Portanto, a energia livre de Gibbs de ligação está relacionada com as concentrações de equilíbrio do inibidor (I), da enzima (E) e do complexo inibido (EI).

Uma afinidade mais elevada (um ΔG mais negativo) corresponde a um valor menor da constante de inibição K_I (mais frequentemente dado em nM ou μM). Um K_I de 1 nM corresponde a um ΔG de $-53,4 \text{ kJ/mol}$ a 310 K e um K_I de 1 μM a $-35,6 \text{ kJ/mol}$. Além disso, usando a Equação 4 pode-se calcular que uma alteração em ΔG em $5,9 \text{ kJ/mol}$ altera K_I por um fator de 10 .⁶⁹

Quando um ligante age como um inibidor competitivo com o substrato natural de uma enzima – sua atividade inibitória é geralmente expressa através do IC_{50} e não mais por meio da constante de inibição K_I .⁷⁰

O IC_{50} é a concentração de inibidor na qual a velocidade da reação de formação do complexo enzima-substrato [ES] se reduz à metade do que seria obtida sem a presença do inibidor, o que indica, na prática, o ponto no qual 50 % da enzima em solução estará ocupada pelo inibidor, portanto indisponível para interagir com o substrato.⁷⁰

A constante de inibição K_I e o valor de IC_{50} estão relacionados por meio da equação de Cheng-Prusoff:⁷¹

$$IC_{50} = K_I \left(1 + \frac{S_0}{K_D} \right) \quad Eq. 5$$

Nessa equação K_I é a constante inibitória do inibidor, S_0 é a concentração inicial do substrato e K_D é a constante de dissociação do complexo enzima-substrato (ES).

A derivação da equação (5) assume que há um excesso de proteína em comparação com o substrato e o inibidor e de que ambos se ligam reversível e estequiometricamente ao mesmo sítio de ligação da enzima, o que é uma aproximação adequada para a maioria dos sistemas.⁷⁰

Na prática, o processo de ligação receptor-ligante tende a ser muito mais complicado devido à existência de efeitos alostéricos ou de múltiplos sítios de ligação, oligomerização dos receptores, formação de ligações covalentes irreversíveis ou mesmo dos mecanismos de reação envolver múltiplas etapas, entre outros. Apesar disso, a caracterização de ligantes por uma única constante K_D é comum na literatura, ainda que a cinética de ligação não seja completamente entendida.⁷⁰

Em geral, os ensaios de inibição são realizados de tal modo que a concentração inicial do substrato seja bem inferior à sua constante de dissociação, de forma que o termo S_0/K_D possa ser desprezado.⁶⁹ Com isso:

$$IC_{50} = K_I \quad \text{Eq. 6}$$

E, conseqüentemente, a energia livre de Gibbs de ligação do inibidor pode ser diretamente calculada a partir do seu IC_{50} , segundo:

$$\Delta G^\circ = RT \ln K_I = RT \ln IC_{50} \quad \text{Eq. 7}$$

Através dessa equação, as energias livres de ligação podem ser estimadas a partir dos resultados experimentais, bem como se pode correlacionar energias livres ou sua versão teórica aproximada, a “afinidade” ou “escore” de *docking* calculadas computacionalmente com o logaritmo do valor de IC_{50} , através de uma relação linear simples.

A energia livre de Gibbs de ligação possui um fator entálpico (ΔH) e um fator entrópico (ΔS) que estão relacionados segundo:

$$\Delta G = \Delta H - T\Delta S \quad \text{Eq. 8}$$

A força motriz no reconhecimento proteína-ligante são as interações eletrostáticas (íon-íon, íon-dipolo e dipolo-dipolo), lipofilicidade/hidrofobicidade e complementaridade geométrica.⁷⁰

Para entender a natureza e as contribuições relativas dessas diferentes forças, é uma útil particionar o ΔG como uma soma de diversas contribuições de energia livre, como mostrado na Equação 9.

$$\Delta G_{lig} = \Delta G_{trans+rot} + \Delta G_{conf} + \Delta G_{polar} + \Delta G_{hidro} + \Delta G_{vdw} \quad \text{Eq. 9}$$

Nessa equação, $\Delta G_{trans+rot}$ contabiliza as restrições de movimentos translacional (movimentos nas direções x, y e z) e restrição de rotações (sobre os eixos x, y e z) das moléculas “inteiras” quando passam do estado livre para o estado ligado; ΔG_{conf} é a diferença nas energias livres conformacionais entre os estados não ligados e ligados devido a restrições conformacionais no complexo ligando-proteína; ΔG_{polar} é a mudança de energia livre devido a interações de grupos

funcionais polares na cavidade de ligação da proteína; ΔG_{hidro} explica a energia livre de ligação devido ao efeito hidrofóbico; e ΔG_{vdw} descreve a diferença na energia livre devido às interações de van der Waals entre os estados ligados e não-ligados.⁶⁹

3.2.6.1 Contribuição da restrição de movimento

Fora da cavidade de ligação na proteína, o ligante se move livremente na solução aquosa através de rotações e translações de toda sua molécula. Entretanto, no interior do sítio ativo, seus três graus de liberdade translacional e os três graus de liberdade rotacional são perdidos na formação do complexo proteína-ligante, que congela esses movimentos.⁶⁹

Termodinamicamente, isso leva a uma diminuição da entropia, resultando em um ΔS mais negativo e conseqüentemente um $T\Delta S$ mais negativo e um ΔG mais positivo. Assim, a perda de liberdade translacional e rotacional se opõe à ligação e $\Delta G_{\text{trans+rot}}$ é um custo de energia livre, que deve ser superado pelas forças de ligação favoráveis para formar o complexo proteína-ligante.⁷⁰

A magnitude desse custo de energia livre varia apenas ligeiramente com o peso molecular, porém é bastante dependente de quão “apertado” for o complexo proteína-ligante. Um complexo mais apertado leva a uma maior perda na liberdade de movimento e, portanto, produz um $T\Delta S$ mais negativo.⁶⁹

A maioria das estimativas de $\Delta G_{\text{trans+rot}}$ varia de 12 kJ/mol para um complexo mais “solto” a 45 kJ/mol para um complexo bem ligado, portanto, é uma contribuição energética desfavorável muito significativa, que está sempre presente e não pode ser reduzida no desenho do ligante.⁶⁹

Para um ligante com K_I de 1 nM, as forças de ligação favoráveis devem liberar não apenas os -53,4 kJ/mol de energia livre de ligação, mas também esse 12 a 45 kJ/mol de energia livre para tornar a própria associação possível.⁶⁹

3.2.6.2 Contribuição da restrição conformacional

A maioria das moléculas ligantes são flexíveis, portanto fora do sítio de ligações e em solução aquosa, o ligante sofre alterações conformacionais por rotação em torno de ligações simples (Figura 19), entretanto, um ligante geralmente se liga a uma proteína em uma única conformação bem definida⁷⁰

Essa conformação posiciona os grupos funcionais usados para ligação em locais apropriados no espaço para que haja interações com seus parceiros de ligação na proteína. Isto implica que os movimentos correspondentes à liberdade conformacional em solução aquosa também estão em grande parte congelados no local de ligação, o que também leva a uma diminuição na entropia (entropia conformacional), com um custo de energia livre para a ligação.⁷⁰

A magnitude de ΔG_{conf} foi estimada em 1-6 kJ/mol por rotação interna restrita e também depende de quão apertado for o complexo proteína-ligante, tal qual para o caso do $\Delta G_{\text{trans+rot}}$.⁶⁹

Além disso, cristalografias de raios-x de complexos proteína-ligante mostra que, em geral, um ligante não se liga à proteína em sua conformação preferida (de mais baixa energia) em

solução aquosa, o que agrega uma penalidade energética adicional à ligação. Essas penalidades, entretanto, são normalmente inferiores a 13 kJ/mol.⁶⁹

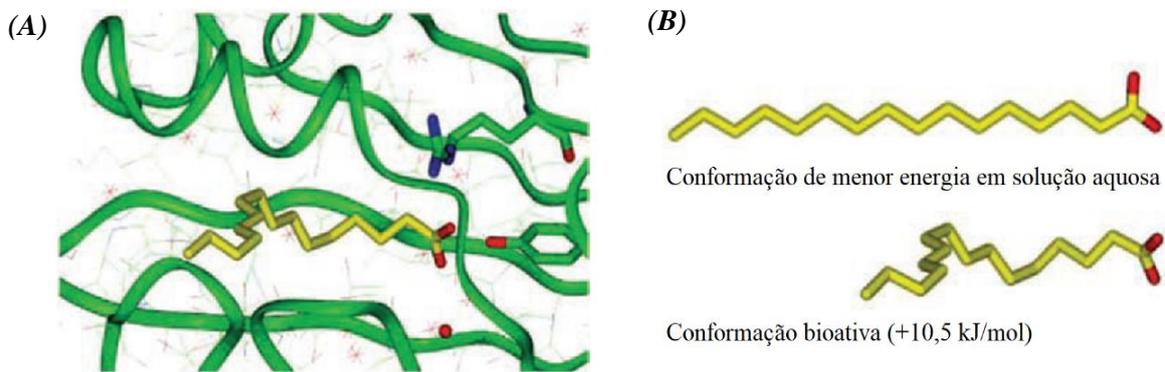


Figura 19 - Ácido palmítico ligado à proteína de ligação de lipídios adipócitos (PDB 1LIE, Figura A) e comparação entre a conformação preferencial em solução e a conformação sítio ativo (B). Adaptado de Larsen et al. (2010).⁶⁹

3.2.6.3 Interações eletrostáticas e ligações de hidrogênio

O termo ΔG_{polar} se refere à mudança de energia livre devido a interações entre grupos funcionais polares no ligante e as cadeias laterais polares dos aminoácidos ou grupos polares do esqueleto polipeptídico (C=O e N-H), bem como interações indiretas de proteína-ligante mediadas por moléculas de água, bastante frequentes no sítio ativo enzimático.⁶⁴

Essas interações incluem as interações íon-íon, íon-dipolo e dipolo-dipolo, todas elas descritas pela Lei de Coulomb:

$$E_{\text{polar}} = \frac{q_i q_j}{\epsilon r_{ij}} \quad \text{Eq. 10}$$

Onde a energia da atração ou repulsão entre as cargas (E_{polar}) depende dos valores q_i e q_j que são as cargas dos íons envolvidos – ou as cargas atômicas parciais para outras interações polares – da constante dielétrica do meio ϵ (que é de 78,4 para a água a 25 °C) e de r_{ij} é a distância que separa as cargas.⁷⁰

Apesar da atração entre as cargas opostas ou dipolos antiparalelos desempenhar um papel importante no reconhecimento proteína-ligante, E_{polar} é difícil de quantificar em proteínas, pois ϵ não é uniforme em toda a proteína, mas depende do microambiente que tende a ser mais lipofílico ($\epsilon \approx 4$) no interior da proteína e na região do sítio ativo.⁶⁹

As *interações íon-íon* ocorrem entre ligantes que contêm grupos ionizáveis eletricamente carregados em pH fisiológico os quais se ligam a resíduos com grupos laterais de carga oposta. É o que ocorre, por exemplo, com ácidos carboxílicos que em pH 7 estão negativamente carregados (ânion carboxilato), podendo interagir com resíduos positivamente carregados nessas condições, como a guanidina de uma arginina ou o ϵ -amino lateral protonado de uma lisina. Esse tipo de interação tem energia de ligação entre 25-50 kJ/mol.⁶⁹

As *Ligação de hidrogênio* (X-H ----Y) são interações de natureza eletrostática formadas entre um átomo aceptor de ligação de hidrogênio (Y) que contém um par de elétrons

isolado (geralmente oxigênio ou nitrogênio) e um doador de ligação (X) de hidrogênio um átomo contendo um hidrogênio ácido, tais como como O–H ou N–H.⁶⁴

Esse tipo de interação é muito utilizado para coordenar ligantes com grupos polares e sua força depende tanto da natureza química quanto do alinhamento geométrico dos grupos interagentes. Uma ligação de hidrogênio típica tem comprimento de ligação entre 2,5 e 3,0 Å (medida entre os átomos pesados X e Y) e é altamente dependente da orientação, com um ângulo ótimo X-H ---- Y de 180°.⁶⁹

A Figura 20 ilustra exemplo de diferentes tipos de ligações de hidrogênio comumente observadas em complexos proteína-ligantes, e que incluem as ligações de hidrogênio neutras (dipolo-dipolo), as ligações assistidas por carga (íon-dipolo) e as pontes salinas, que combinam a formação de ligações de hidrogênio com interações eletrostáticas do tipo íon-íon.⁶⁹

As energias de ligação para aceptores/doadores neutros é da ordem de 2-7,5 kJ/mol enquanto que com aceptores/doadores eletricamente carregados esses valores aumentem para 12,5-25 kJ/mol, sobretudo na formação de pontes salinas.⁶⁴

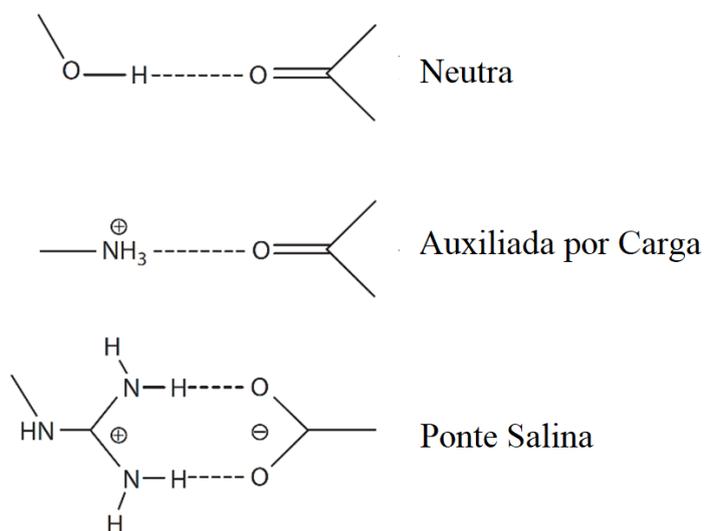


Figura 20 - Exemplos de diferentes tipos de ligações de hidrogênios presentes em sistemas biológicos. Adaptado de Larsen et al. (2010).⁶⁹

Entretanto, a formação de ligação de hidrogênio é um processo de troca, na medida em que antes da formação do complexo proteína-ligante os grupos funcionais polares do ligante e as porções polares dos aminoácidos e de grupos estruturais da proteína estão envolvidos em ligação de hidrogênio com moléculas de água do meio circundante.⁶⁹

Na formação do complexo essas ligações de hidrogênio com o solvente são substituídas pelas ligações de hidrogênio entre o ligante e a proteína, de modo que o efeito líquido dessa troca é a diferença na energia livre entre a formação de ligação de hidrogênio com água ou com a proteína.

Assim, se o ligante possuir um grupoceptor ou doador de hidrogênio que forma ligações de hidrogênio com a água, mas não com a proteína (isto é, se houver ligações de hidrogênios desemparelhadas quando o ligante está “enterrado” na proteína) há uma penalidade energética significativa, de cerca de 4 kJ/mol para um substituinte neutro e de 16 kJ/mol para um substituinte carregado, o que corresponde a uma perda de afinidade por um fator de 5 até 500.⁶⁹

Outros tipos de interações polares frequentemente observadas nos complexos proteína-ligante são interações de π - π , também conhecidas como empilhamentos π , e cátion- π . A natureza exata dessas interações é bastante complexa, mas qualitativamente elas podem ser compreendidas em termos eletrostáticos.⁶⁹

A Figura 21 mostra o potencial eletrostático molecular calculado do benzeno quando um cátion é colocado em diferentes posições ao redor do anel aromático. A cor vermelha indica uma forte atração entre o cátion e o anel aromático e uma cor azul indica uma forte repulsão.⁶⁹

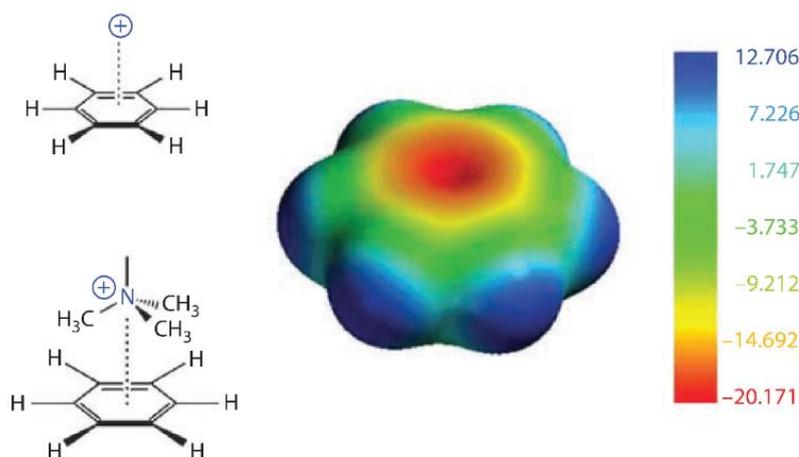


Figura 21- O potencial eletrostático molecular do benzeno na presença de um cátion. Adaptado de Larsen et al. (2010).⁶⁹

Numa interação do tipo cátion- π , o íon amônio, por exemplo, pode interagir favoravelmente com a face do anel de benzeno enquanto que nas interações π - π , a ponta de um anel benzênico interage com a face do outro (empilhamento-T) ou as duas faces interagem de forma alinhada (empilhamento sanduíche).⁶⁹

Outros anéis aromáticos como fenol e indol exibem potenciais eletrostáticos similares como benzeno. Dessa forma, os anéis aromáticos das cadeias laterais de fenilalanina, tirosina e triptofano podem interagir favoravelmente com grupos funcionais carregados positivamente ou com anéis aromáticos do ligante.⁶⁹

Estima-se que uma interação π - π e cátion- π podem contribuir 8-17 kJ/mol para a energia livre de ligação global do ligante, o que equivale a um aumento de 23 a 760 vezes na afinidade receptor-ligante.⁶⁹

3.2.6.4 Contribuição Hidrofóbica

O efeito hidrofóbico descreve a tendência de moléculas orgânicas hidrofóbicas se agregarem e se transferirem para solventes apolares ao invés de permanecer em solução aquosa. No contexto biológico, essa fase orgânica pode ser, por exemplo, uma região lipofílica de uma proteína.⁶⁴

Quando um composto lipofílico é inserido na água, ele muda a rede dinâmica de ligações de hidrogênio que havia entre as moléculas de água puras, criando uma nova interface na qual as moléculas de água precisam se arranjar em torno do composto lipofílico e assumir uma estrutura mais ordenada do que teriam puramente sozinha. Isso resulta em uma diminuição da entropia.⁶⁴

A formação de um complexo ligando-proteína desloca essa água ordenada que circunda o ligante e que também preenche de forma relativamente ordenada o sítio ativo da proteína empurrando-as de volta para o “bulk” da solução, o que aumenta a liberdade de movimento das moléculas resultando em um aumento na entropia, com o consequente um aumento na afinidade receptor-ligante ao contribuir com a redução do ΔG .⁶⁹

A contribuição hidrofóbica tem uma dependência com a área superficial das moléculas em interação com o solvente, também chamada de superfície acessível ao solvente, ou SAS. Com base nas medidas de transferência entre solvente e na energia de ligação proteína-ligante, estima-se que essas contribuições hidrofóbicas variam entre 0,1 e 0,24 kJ/Å²mol. Uma metila, por exemplo, com SAS de cerca de 25 Å² provoca um aumento de afinidade por um fator de 3 a 10, ao passar do meio aquoso para o interior lipofílico do sítio ativo.⁶⁹

3.2.6.5 Contribuição não-polares (Interações de Van der Waals)

As interações não-polares entre os átomos, também chamadas de interações de Van der Waals, decorrem dos contatos interatômicos do ligante com sítio ativo e podem ser atrativas ou repulsivas como mostrado pela curva de energia na Figura 22.

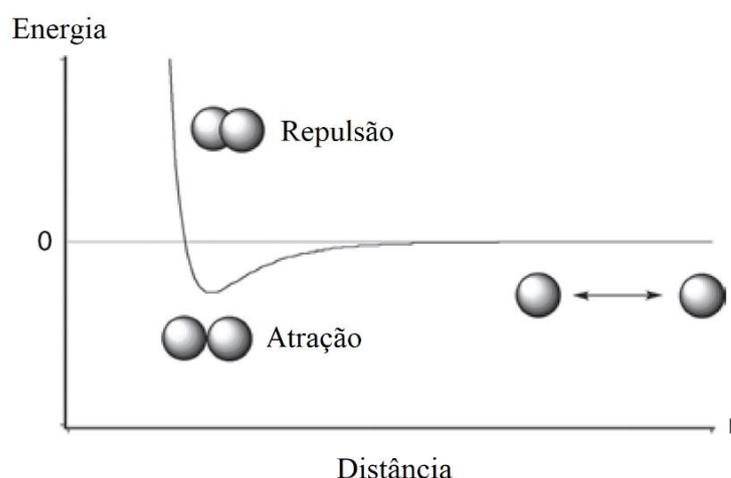


Figura 22 - Curva de energia potencial em função da distância interatômica para interações do tipo Van der Waals. Adaptado de Larsen et al. (2010).⁶⁹

Quando as distâncias que ligam os núcleos dos átomos são muito curtas, a sobreposição das nuvens eletrônicas provoca um aumento abrupto de energia devido à repulsão eletrostática. Assim, quando uma parte do ligante entra em conflito ou colide com átomos no sítio de ligação, essa interação estericamente repulsiva frequentemente leva a uma diminuição na afinidade de ligação receptor-ligante.⁶⁴

A distâncias mais longa, entretanto, existe uma região de atração entre os átomos. Essa atração se deve às chamadas forças de dispersão, que tem natureza basicamente eletrostática e que se devido a interações entre dipolos temporários induzidos em dois átomos adjacentes.⁶⁴

Para um único contato átomo-átomo a força da interação é pequena de cerca de 0,2 kJ/mol. Contudo, como o número total de tais interações pode ser muito grande, a soma das interações de Van der Waals pode ser significativamente substancial (50-100 kJ/mol), sobretudo

se há uma alta complementaridade geométrica entre o ligante e o sítio ativo. Esse encaixe adequado também se faz necessário na medida em que força dessas interações de dispersão varia com $1/r^6$, sendo, portanto de curtíssimo alcance.⁶⁹

3.3 MODELAGEM MOLECULAR

Da mesma forma como a ciência computacional impactou desde os meios de produção e comércio à forma como as pessoas se comunicam e se relacionam, nos últimos anos o aumento da capacidade de processamento e a diminuição de custos promovida pelo desenvolvimento de *hardwares* melhores, como unidades de processamento gráfico e processadores multi-core, aliados ao desenvolvimento de algoritmos mais robustos e da computação em paralelo, tornaram possíveis a modelagem computacional de sistemas macromoleculares biológicos grandes e extremamente complexos, contendo centenas de milhares de átomos.

Deste modo, a modelagem molecular e a simulação de biomoléculas se tornaram ferramentas imprescindíveis tanto para a biologia molecular quanto para a química medicinal e farmacêutica.

Atualmente, ferramentas como o *docking molecular* permitem a predição dos modos de ligação e das energias livres de complexos receptor-ligante, de forma rápida, robusta e relativamente precisa, permitindo a triagem de bibliotecas de moléculas na busca de compostos líderes e no desenho racional de melhores fármacos, ao mesmo tempo em que simulações de *dinâmica molecular* permitem a compreensão da flexibilidade e das alterações conformacionais de complexos proteína-ligante e em que métodos híbridos quanto-mecânicos e de mecânica molecular permitem explorar os mecanismos de reações e a origem do poder catalítico das enzimas. Apenas alguns dos inúmeros exemplos de como a modelagem computacional está revolucionado de forma irreversível, a química, a bioquímica e a compreensão humana sobre a Natureza.⁷²

3.3.1 *Docking* Molecular

O *docking* ou atracamento molecular, em uma tradução literal, é uma metodologia utilizada para a predição computacional da energia ou intensidade de interação e da estrutura de complexos ligante-receptor, no qual o ligante é uma pequena molécula e o receptor é normalmente um alvo macromolecular, como uma proteína ou ácido nucleico, tendo especial aplicação do estudo da forma de atuação e na otimização de inibidores enzimáticos.⁷³

Em linhas gerais - considerando-se o caso enzimático - o *docking* se constitui em dois processos inter-relacionados que são a amostragem ou busca das possíveis conformações do ligante no sítio ativo da enzima e a classificação dessas conformações (em termos de energia livre) por meio de uma função de escore. Idealmente os algoritmos de busca devem ser capazes de reproduzir o modo de ligação experimental e a função de escore deveria classificar esse modo como o melhor entre todas as conformações geradas.^{74,75}

O problema, entretanto, está no fato do *docking* molecular envolver um grande número de graus de liberdade, pois além dos seis graus de liberdade translacional e rotacional de cada molécula os inúmeros graus de liberdade conformacional, sobretudo do receptor macromolecular,

necessitariam ser considerados, gerando um número enorme de possíveis modos de ligação entre duas moléculas e demandando um custo computacional proibitivo para se gerar todas as configurações possíveis.^{72,73}

Diversas metodologias de *docking* distintas têm sido desenvolvidas para abordar esse problema, as quais se distinguem no número de graus de liberdade que ignoram. O caso mais simples - e computacionalmente mais econômico - é considerar o ligante e o receptor com sendo corpos rígidos, resumindo o problema aos seis graus de liberdade translacionais e rotacionais.⁷⁴

Essa ideia está de acordo com o modelo chave-fechadura de Emil Fisher e se fundamenta na premissa de que alta especificidade de uma enzima por seu substrato está relacionada à complementaridade geométrica tal como uma chave é complementar a uma única fechadura. Dessa forma, essas metodologias buscam encontrar moléculas com um alto grau de complementaridade especial com o sítio ativo. A flexibilidade do ligante pode ser considerada por meio do *docking* de um conjunto pré-calculado de conformações distintas, mas não no *docking* em si.⁷⁴

O comportamento real, como visto anteriormente, está mais relacionado ao modelo de encaixe induzido, postulado por Daniel Koshland em 1958, no qual a interação inicial relativamente fraca do substrato com o sítio ativo induz alterações conformacionais tanto na enzima quanto no ligante que provocam o fortalecimento dessa interação (encaixe perfeito) e ajuste dos resíduos do sítio ativo para as posições precisas que possibilitam a catálise enzimática.⁷²

Como a consideração da flexibilidade do ligante e do receptor é custosa computacionalmente, o melhor custo-benefício entre exatidão e tempo computacional é a situação intermediária, amplamente implementada em diversos softwares como AutoDock e FlexX.⁷⁵

Essa situação intermediária - e mais comumente utilizada - diz respeito a considerar o ligante flexível e o receptor rígido, o que representa uma significativa evolução em relação ao modelo chave-fechadura, mas ainda é incapaz de expressar apropriadamente a natureza flexível e dinâmica da interação enzima-ligante em condições fisiológicas.^{74,75}

No outro extremo está a consideração de todos os graus de liberdade do sistema, o que pode ser realizada por meio de simulação de dinâmica molecular (MD, do inglês *molecular dynamics*). Ao mover cada átomo do sistema independentemente dos demais, há a representação completa da flexibilidade do ligante e da proteína. Entretanto, o progresso em passos muito pequenos e a impossibilidade de sobrepular grandes barreiras de energia conformacional – o que pode levar à amostragem inadequada das possíveis conformações – aliado ao elevado custo computacional impossibilita a aplicação dessa técnica no estudo de bibliotecas de moléculas, ainda que pequenas.^{72,75}

Vários outros métodos têm sido implementados para a consideração da flexibilidade do receptor. O mais simples, chamado de “*soft-docking*”, consiste em reduzir a energia de repulsão de Van der Waals na função de escore, permitindo assim certo grau de sobreposição atômica entre o ligante e o receptor, o que não inclui adequadamente a flexibilidade, mas é computacionalmente eficiente e representa alguma melhora em relação à estrutura completamente rígida.⁷⁴

Outra possibilidade é a utilização de bibliotecas de rotâmeros, um conjunto de conformações das cadeias laterais dos aminoácidos determinadas pela análise estatística de dados cristalográficos experimentais, permitindo a consideração da flexibilidade dessas cadeias laterais em uma amostragem relativamente rápida. Diversas cadeias laterais podem ser selecionadas e amostradas simultaneamente com o ligante, enquanto as outras porções do receptor permanecem rígidas.^{72,75}

3.3.1.1 Algoritmos de busca

Diversos algoritmos de busca diferentes foram criados na tentativa de reproduzir satisfatoriamente o modo de ligação experimental, os quais se encontram amplamente implementados em softwares de *docking* molecular, com destaque para os algoritmos de correspondência, construção incremental, Monte Carlo e algoritmos genéticos.^{74,75}

Algoritmos de correspondência (MA, do inglês *matching algorithms*) se baseiam no mapeamento do formato molecular de um ligante no sítio ativo em função de características estruturais e informação química tais como grupos doadores ou aceptores de ligação de hidrogênio. Nessa metodologia, a proteína e o ligante são representados por grupos farmacofóricos - o conjunto de características eletrônicas e estereais que caracterizam subunidades estruturais e grupos funcionais.⁷²

A distância entre grupos farmacofóricos complementares na enzima e no ligante é calculada para cada diferente conformação do ligante, predominando as combinações que maximizem as interações e diminuam as distâncias. Devido à rapidez esse tipo de algoritmo é muito aplicado na triagem computacional de bibliotecas de moléculas na busca por compostos líderes, compostos com possível aplicação farmacêutica em função de alta potência.⁷⁴

Os métodos de construção incremental (IC) o ligante é dividido em diversos fragmentos pela quebra das ligações móveis e um desses fragmentos, chamado de âncora – geralmente o maior ou com o papel funcional mais significativo na interação com a enzima – é docado no sítio ativo primeiro. Os demais fragmentos são então acrescentados um por um em diferentes orientações de forma a se encaixar no sítio ativo da forma mais satisfatória possível.⁷⁵

Os algoritmos genéticos e de Monte Carlo fazem parte dos métodos estocásticos, isto é, cada novo estado do sistema tem origem em um evento aleatório. Assim, o espaço conformacional é examinado pela modificação aleatória da conformação do ligante, possibilitando mudanças grandes e saltos na barreira de energia potencial e permitindo a consideração de configurações que não seriam aceitas em um método sequencial.^{72,75}

Os algoritmos de busca Monte Carlo geram um conjunto aleatório de configurações do ligante por meio de rotações, translações e rotação de ligações. Cada uma dessas conformações é então avaliada por um critério de seleção baseado na energia. Se passar por esse critério, essa configuração será armazenada e modificada para gerar a próxima conformação, repetindo-se esse processo até que um número pré-definido de conformações seja obtido.^{72,75}

Os algoritmos genéticos, por sua vez, foram inspirados na teoria da evolução de Darwin e se baseiam na codificação dos graus de liberdade do ligante em cadeias binárias denominadas "genes". Esses genes se organizam formando "cromossomos", cada um representando uma determinada configuração do ligante. Um operador genético atua promovendo mutações - a troca

aleatória de um desses genes por outro - ou então cruzamentos, no qual porções do gene são trocados entre um cromossomo e outro, o que gera uma nova configuração estrutural para o ligante.

Essas novas estruturas são avaliadas pela função de escore e aquelas que sobreviverem, isto é, superarem a linha de corte, serão usadas para a próxima geração. Essa é a metodologia implementada no software AutoDock, e de forma aprimorada no AutoDock Vina, o programa de *docking* utilizado nesse trabalho.⁷⁶

3.3.1.2 Função de Escore

Os algoritmos de busca são capazes de produzir um conjunto amplo de potenciais soluções, algumas das quais podem ser facilmente desconsideradas em função de sua alta energia por colidirem com a enzima. As outras possibilidades, entretanto, precisam ser classificadas de modo que a orientação do complexo proteína-ligante mais próxima da “verdadeira” seja encontrada e de que seja possível classificar um conjunto de ligantes em razão de suas capacidades de ligação – o que é primordial na triagem virtual de compostos. Essa classificação é feita pela função de escore.⁷⁴

Devido ao grande número de orientações que são geradas durante uma simulação de *docking*, a classificação das soluções necessita ser realizada de forma rápida. Por essa razão, as funções de escore normalmente envolvem uma estimativa ao invés de um cálculo preciso da energia livre de ligação proteína-ligante, adotando diversas suposições e simplificações.^{74,75}

As funções de escore mais amplamente utilizadas são aquelas baseadas em campos de força, que avaliam a energia livre de ligação por meio da soma das interações não-ligadas como interações eletrostáticas e de Van der Waals. As interações eletrostáticas são calculadas por potenciais de Coulomb utilizando uma função dielétrica dependente da distância ao invés de considerar a enzima como um conjunto de cargas pontuais, ao passo que os termos de Van der Waals são descritos por potenciais de Lennard-Jones, ajustados para controlar quão próximo o contato entre os átomos da proteína e do ligante podem ser aceitáveis.^{72,75}

O cálculo dessas funções, que convergem à distância infinita, é computacionalmente custoso. Assim, um raio de corte é utilizado para as interações não ligadas, o que acaba reduzindo a exatidão dos efeitos de longo alcance envolvidos na ligação. A função de escore do software AutoDock é baseada no campo de forças AMBER⁷⁷ e inclui termos de van der Waals, ligações de hidrogênio, interações eletrostáticas, entropia conformacional e dessolvatação, em que cada termo é ponderado por um fator obtido por um conjunto de dados experimentais.^{76,77}

As funções de escore empíricas decompõem a energia livre em diversos componentes, como ligações de hidrogênio, interações iônicas, efeitos hidrofóbicos, entre outras, que são multiplicados por um determinado coeficiente e somadas para produzir o escore final. Esses coeficientes são obtidos pelo ajuste de funções de regressão em conjuntos de complexos proteína-ligante com energias livres de ligação conhecidas.^{72,74}

Há ainda funções de escore baseadas em conhecimentos da estrutura cristalina, na qual as distâncias e frequências de contato interatômico de complexos enzima-ligante conhecidos são utilizadas para parametrizar o campo de forças e depois aplicados a sistemas em geral, além do escore de consenso, que combina diversas metodologias diferentes para avaliar as conformações

de *docking*. Nesse método, uma determinada configuração do complexo enzima-ligante só será aceita quando apresentar um bom escore em todas as diferentes metodologias de escore.⁷²

3.3.2 Dinâmica Molecular

A simulação de dinâmica molecular (MD) é um dos métodos computacionais mais amplamente utilizados no estudo macromoléculas biológicas como enzimas, receptores, ácidos nucleicos até mesmo agregados macromoleculares como membranas celulares.⁷⁴

Tais sistemas, compostos por um número muito grande de átomos, não podem ser tratados quanticamente em função da complexidade matemática das equações envolvidas, sobretudo em função dos graus de liberdade eletrônicos, o que demandaria um elevado custo computacional e um tempo de processamento proibitivo.⁷⁴

Nesses sistemas, efeitos de natureza quântica podem ser incorporados dentro de uma descrição clássica, na medida em que as massas e energias envolvidas são muito maiores que as envolvidas em fenômenos quânticos, permitindo uma descrição contínua das transferências de energia e a aplicação da termodinâmica estatística na determinação das suas propriedades macroscópicas, sem perdas importantes da exatidão frente ao resultado real.^{72,74}

A ideia central da dinâmica molecular é estudar a evolução temporal das partículas de um sistema do qual se conhecem os potenciais de interação entre as partículas e as equações que regem seu movimento. Assim, conhecendo-se as posições e velocidades iniciais das partículas que compõem o sistema pode-se calcular as forças que atuam sobre cada uma delas em função da interação com as demais e determinar novas posições e velocidades em um instante posterior por meio das equações do movimento de Newton.⁷⁸

Essas novas posições são então utilizadas para calcular novas forças, que produzem novas posições e velocidades e assim sucessivamente, gerando-se trajetórias moleculares para todo o sistema. Através das médias dessa trajetória as propriedades macroscópicas podem ser determinadas.⁷⁴

A realização de uma simulação de dinâmica molecular envolve quatro etapas fundamentais, que são a inicialização, o cálculo das forças exercidas sobre cada partícula devido à interação com as demais, a integração das equações do movimento e a amostragem, que envolve a análise e o armazenamento das trajetórias. O cálculo de forças e a movimentação das partículas é realizado recursivamente durante o intervalo de simulação determinado.⁷²

A *inicialização* consiste na determinação da configuração inicial do sistema (as posições de cada átomo) e a das velocidades iniciais, que podem ser nulas para o sistema evoluir a partir do repouso ou determinadas pela temperatura numa distribuição de Maxwell-Boltzmann.⁷²

Em sistemas biológicos, como é o caso de enzimas, as posições iniciais dos átomos e seus modos de ligação são geralmente obtidas a partir da estrutura cristalográfica da macromolécula. Outras moléculas (ligantes) podem já estar previamente na estrutura experimental ou serem colocadas computacionalmente, inclusive por meio de *docking* molecular, em uma posição favorável à interação com a enzima.⁷⁴

O sistema é então centrado em uma caixa de simulação, normalmente cúbica, à qual se adicionam um determinado número de moléculas de solvente tal que a densidade numérica

corresponda à densidade real nas condições desejadas, além de contra íons para neutralizar quaisquer cargas residuais do sistema. Uma minimização de energia é normalmente efetuada para reacomodar todas as moléculas dentro da caixa, eliminando possíveis sobreposições atômicas que prejudicariam o cálculo de forças.⁷²

A fim de se eliminar efeitos relacionados com a borda da caixa de simulação, *condições periódicas de contorno* são definidas. Para isso, cópias idênticas da caixa original são dispostas em todas as direções ao redor da caixa central e suas partículas deve se mover de maneira idêntica à da caixa principal. Dessa forma, se um átomo sai por uma das seis faces da caixa principal, sua imagem periódica entra pela face oposta com velocidade idêntica ao que saiu. Assim, o número de partículas permanece inalterado durante toda a simulação.⁷⁸

Um problema surge ao se utilizar condições periódicas de contorno que é a possibilidade de interação entre uma partícula e diversas imagens periódicas de outra, o que eleva o custo computacional. Quando essa interação é de curto alcance (potencial de Lennard-Jones) utiliza-se a convenção da imagem mínima, segundo a qual uma determinada partícula pode interagir apenas com a imagem mais próxima de outra. Além disso, define-se um raio de corte para esses potenciais, a partir do qual a interação é considerada nula, evitando-se a interação com sua própria imagem ou com a mesma partícula duas vezes.⁷⁸

Para as interações de longo alcance (potenciais de Coulomb), cujas forças decaem lentamente com a distância, limitar um raio de corte compromete severamente os resultados da simulação. Assim, a força de interação entre duas partículas é substituída por uma força efetiva, calculada pelo método de Ewald, que soma as interações de uma partícula com todas as imagens periódicas de outra, convergindo no limite infinito, como é o caso do método PME (do inglês, *Particle Mesh Ewald*), uma das variantes da soma de Ewald, que apresenta melhor precisão e velocidade de processamento.^{72,74}

A etapa seguinte, o *cálculo de forças*, é a etapa mais custosa computacionalmente e inicia-se com a determinação das distâncias entre os átomos, comprimentos e ângulos de ligação, ângulos diedros e dos raios de interações inter e intramoleculares de curto e longo alcance. As forças resultantes sobre cada uma das partículas são então calculadas numericamente segundo um conjunto de equações e parâmetros denominado campo de forças.⁷²

Esse campo descreve as interações intra e intermoleculares como funções potenciais classificadas como potenciais ligados e não ligados.^{72,74}

Os potenciais ligados se referem aos estiramento e deformações angulares das ligações químicas bem como a torção de ângulos diedros próprios e impróprios, movimentos que envolvem dois, três ou quatro átomos, respectivamente. As interações de átomos não ligados são tratados por potenciais efetivos, compostos por uma soma de termos que incluem a repulsão das nuvens eletrônicas e a atração de Van der Waals - comumente descrita pelo potencial de Lennard-Jones - ambos de curto alcance, e o termo eletrostático de Coulomb, de longo alcance.^{74,78}

Com exceção dos átomos vizinhos que estão quimicamente ligados, a interação entre os átomos de uma mesma molécula também é descrita por potenciais não ligados. Essas interações estão intimamente relacionadas com a formação da estrutura tridimensional das macromoléculas biológicas, sendo imprescindível sua parametrização adequada para a simulação satisfatória desse tipo de sistema.⁷²

Diversos campos de força foram desenvolvidos para o estudo dos mais diversos sistemas, com destaque para os campos AMBER⁷⁷, CHARMM⁷⁹ e GROMOS⁸⁰, que apresentam uma boa parametrização para simulações de biomoléculas em geral. O campo de forças AMBER03, utilizados nesse trabalho é do tipo *all-atom*, isto é, todos os átomos do sistema, inclusive hidrogênios apolares são descritos explicitamente, ao contrário dos campos da família GROMOS, que são do tipo *united atom*, não explicitando hidrogênios alifáticos, mas sim os unindo aos átomos pesados visando o aumento do desempenho computacional.^{77,80}

Após o cálculo das forças, a *movimentação das partículas* é realizada por meio da resolução das equações diferenciais do movimento que as controlam. Como o sistema é conservativo, isto é, o potencial interno não depende das velocidades nem do tempo, a força resultante (F_i) atuando sobre uma determinada partícula i (já calculadas pelo campo de força) em um determinado instante de tempo t , é uma função apenas das coordenadas da partícula (r_i) segundo a equação clássica do movimento newtoniano:

$$a_i = \frac{d^2 r_i}{dt^2} = \ddot{r}_i(t) = \frac{F_i(t)}{m_i} \quad \text{Eq. 11}$$

Em que a_i é a aceleração da partícula, m_i é sua massa e r_i sua posição no tempo t . Assim, conhecendo-se a força resultante e a massa das partículas, suas posições em um tempo ($t + \delta t$) podem ser determinadas, em que δt , normalmente da ordem de femtossegundos (10^{-15} s) é chamado de passo de integração. Diversos algoritmos de integração numérica podem ser utilizados na simulação de MD, com destaque para os algoritmos de Verlet e leap-frog.⁷⁴

Após o cálculo das novas posições e velocidades (atualização da trajetória), as forças resultantes são novamente calculadas, produzindo outras posições e velocidades, processo que se repete alguns milhões de vezes, de tal forma que intervalos da ordem de dezenas a centenas de nanossegundos sejam simulados.^{72,74}

Naturalmente, a memória necessária para armazenar as informações de posição, velocidades e aceleração em três dimensões para todas as partículas que compõem o sistema em cada um desses milhões de passos é absolutamente grande. Por essa razão, após um determinado número de etapas ou passos de integração predeterminados, as trajetórias são gravadas, num processo chamado de amostragem. Com as trajetórias do sistema, expressas em função das posições e velocidades atômicas, propriedades termodinâmicas como a energia interna, entalpia e capacidade calorífica, propriedades estruturais e propriedades dinâmicas podem ser determinadas.⁷²

3.3.3 Ensemble docking

Diversas técnicas têm sido propostas para a inserção de flexibilidade em modelos estáticos de proteínas com o propósito de aperfeiçoar o *docking* molecular ao permitir a inclusão de alterações conformacionais de larga escala, variando desde simples modificação das cadeias dos aminoácidos até a inclusão da mobilidade completa do esqueleto polipeptídico.⁸¹

Dentre essas, uma metodologia que se destaca é o *ensemble docking* ou *docking* de grupo, na qual o ligante é docado em um conjunto de conformações do receptor ao invés de

apenas uma. Essas diferentes estruturas permitem, idealmente, a amostragem de grande parte, se não toda a paisagem conformacional do receptor proteico.⁸²

Como essas diferentes conformações são geradas *antes* do *docking* e por meio da consideração da flexibilidade completa da proteína, a metodologia de *ensemble docking* tem uma analogia direta com o modelo de seleção conformacional, uma vez que o *docking* permite a seleção, de dentro desse conjunto de estruturas da proteína, aquela ou aquelas que permitem o melhor encaixe e interação com o ligante.⁸³

Na perspectiva do *docking*, entretanto, a flexibilidade do receptor é considerada apenas de forma implícita, na medida em que a conformação da proteína *não se altera* durante o *docking*, o que faz com que os efeitos dinâmicos e as alterações estruturais relacionado com presença do ligante no sítio de ligação e por ele induzidas não sejam plenamente amostradas.⁸⁴

O *ensemble docking* rapidamente se tornou popular pois permite a inclusão da flexibilidade do receptor sem a necessidade de quaisquer alterações significativa nos programas de *docking* já existentes.⁸⁵

O primeiro passo na aplicação dessa metodologia envolve a amostragem da paisagem conformacional da proteína alvo, a fim de obter novas estruturas distintas da estrutura inicial, geralmente uma estrutura experimental ou construída por homologia.⁸⁵

Uma das formas mais eficientes e também mais populares de se amostrar as conformações da proteína é através da simulação de dinâmica molecular, na qual a flexibilidade completa da proteína pode ser simulada em ambiente solvatado e relativamente realístico, o qual permite grandes rearranjos estruturais.⁸³

Durante a simulação de DM, a evolução temporal da estrutura do receptor, partindo de sua conformação experimental, pode ser acompanhada, produzindo uma trajetória, que nada mais é do que uma espécie de “filme” dos movimentos moleculares da proteína, o qual é composto por milhares de quadros ou frames, que podem ser extraídos para análise e utilizados no *docking* proteína-ligante.⁸⁴

Essa metodologia, entretanto, ainda apresenta um custo computacional significativo, além de que as simulações de DM geralmente não são capazes de cruzar altas barreiras energéticas e acabam explorando apenas mínimos locais em proteínas com paisagens de energia livre conformacional muito rugosas.⁸⁴

O maior desafio, entretanto, é a seleção de um subconjunto de frames da dinâmica molecular que seja realmente representativo das alterações conformacionais pelas quais o receptor passa durante a trajetória, na medida em que seria impossível realizar o *docking* de cada confôrmero obtido em cada um dos frames.⁸⁵

Diversos estudos mostram que o *ensemble docking* proporciona uma melhora na qualidade e confiabilidade dos resultados de *docking* quando comparado com o *docking* em uma única estrutura rígida.⁸⁵

Entretanto, a construção e composição do *ensemble* de estruturas tem um grande impacto nos resultados, sendo que algumas vezes efeitos não cooperativos podem ser observados, nos quais a adição de algumas conformações ao *ensemble* acaba deteriorando a qualidade do *docking* geralmente pela geração de falsos positivos. Isso evidencia a necessidade de novas e melhores estratégias para seleção de estruturas e construção do *ensemble*.⁸⁵

Apesar dessas desvantagens, é indiscutível a necessidade de inclusão da flexibilidade do receptor para modelar de forma mais realística os sistemas macromoleculares biológicos e assim poder inferir sobre os mecanismos bioquímicos que orientam esses sistemas, especialmente aqueles envolvidos nas interações receptor-ligantes, e o *ensemble docking* permite essa inclusão de maneira relativamente rápida e eficaz.⁸⁶

Além disso, a geração do conjunto de estruturas a partir de simulações de DM de complexos proteína-ligante com compostos experimentalmente ativos permite ao *ensemble docking* incluir a ação de ligantes sobre a estrutura proteica e unir os modelos de encaixe-induzido e seleção conformacional, o que não seria acessível por outras metodologias de modelagem molecular, tampouco pelo *docking* “clássico”.⁸⁷

4 METODOLOGIA

4.1 PREPARAÇÃO DAS MOLÉCULAS

4.1.1 Preparação do Receptor

As coordenadas tridimensionais do receptor (AsbF) foram obtidas da única estrutura cristalina de deidroshikimato desidratase de *Bacillus anthracis* depositada até o presente momento no Protein Data Bank.

Essa estrutura cristalográfica (ID: 3DX5, resolução 2,12 Å) foi depositada por Pflieger e colaboradores (2008)⁵⁸ e obtida por meio da superexpressão do gene AsbF do *Bacillus anthracis* cepa Sterne em *Escherichia coli* BL21.

Além da proteína, a estrutura cristalográfica apresentava ainda moléculas de glicerol, ácido 3,4-dihidroxibenzoico (3,4-DBH), 2-amino-2-hidroximetil-propano-1,3-diol (TRIS) além de águas de cristalização e íons cloreto e manganês. Todas essas pequenas moléculas cocrystalizadas foram excluídas da estrutura do receptor, mantendo-se apenas as coordenadas dos aminoácidos e dos íons como arquivo de entrada dos estudos computacionais.

Tendo sido superexpressa em um meio enriquecido em selenometionina com o propósito de incrementar a resolução cristalográfica a proteína apresentava este aminoácido ao invés de metioninas. Fez-se então a substituição direta de selênio por enxofre nos resíduos metionina do arquivo de coordenadas, a fim de se reproduzir as condições biológicas naturais.

Além disso, a estrutura cristalográfica apresentava manganês como cofator metálico. Como os estudos experimentais de inibição enzimática foram realizados utilizando-se magnésio como cofator metálico, fez-se também a substituição de Mn^{2+} por Mg^{+2} de modo a se reproduzir mais fielmente as condições experimentais utilizadas por Koppisch e colaboradores nos ensaios de inibição.⁸⁸

Por fim, em razão de átomos faltando no resíduo Asparagina-274, o último resíduo da porção C-terminal, optou-se por excluir esse aminoácido da estrutura, visto que o mesmo se encontra bem afastado do sítio ativo e provavelmente não desempenha qualquer papel relevante sobre a atividade e catálise enzimática.

Após tais modificações, o padrão de protonação da enzima foi analisado por meio do programa PROPKA.⁸⁹ O pH de estabilidade ótima para a proteína foi calculado como sendo 8,4, para o qual a energia livre de enovelamento atinge o valor mínimo de 36,8 kcal/mol em 298K, valor muito próximo do utilizado nos ensaios experimentais de inibição, que foi de 8,3.

A carga total média foi estimada em -14.42 para a proteína desnaturada e -15.46 para a forma enovelada em pH 7,0 as quais passaram, respectivamente para -17.67 e -17.97 em pH 8.0. Porém, como o único resíduo que apresentou pKa entre 7.0 e 8.3 foi o grupo amino da porção N-terminal (pKa = 7,39) e este não desempenha papel catalítico, pode-se afirmar que o padrão médio de protonação no pH experimental de 8,3 é mesmo que o de pH neutro.

Portanto, os grupos laterais dos ácidos aspárticos e ácidos glutâmicos foram considerados como desprotonados (negativamente carregados), das lisinas e argininas protonados (positivamente carregados) e os grupos laterais de cisteínas, tirosinas e especialmente das

histidinas foram considerados como estando em seu estado padrão de protonação (eletricamente neutros).

4.1.2 Preparação de Ligantes

Os estudos experimentais de inibição foram realizados por Andrew T. Koppisch e colaboradores da Northern Arizona University (Flagstaff, EUA) e os dados gentilmente cedidos para a realização do presente estudo computacional.

Koppisch e colaboradores avaliaram o potencial inibitório *in vitro* de 48 pequenas moléculas sobre deidrosnikotinato desidratase de *Bacillus anthracis* obtida pela clonagem do gene AsbF do *B. anthracis* em *Escherichia coli* BL21, sua superexpressão e posterior purificação e caracterização da enzima obtida. Maiores detalhes sobre os procedimentos experimentais podem ser encontrados em Simon et al (2015).⁸⁸

Dentre os compostos avaliados, incluem-se indolcarboxilatos, hidróxido e aminobenzoatos, benzofuranocarboxilatos além de ácido protocatecuico (produto da reação enzimática) e análogos, entre outros. As estruturas químicas desses compostos e os códigos utilizados nos estudos experimentais estão mostrados no Apêndice A

Dos 48 compostos com potencial ação como inibidores que foram estudados, apenas 25 apresentaram potencial inibitório mensurável e foram escolhidos para o presente estudo computacional. Nesse trabalho manteve-se a numeração original adotada nos ensaios experimentais, por isso há lacunas na lista de inibidores que correspondem aos compostos inativos, que não foram incluídos no estudo teórico. As estruturas dessas moléculas e os respectivos códigos podem ser encontrados na Figura 23, juntamente com a estrutura do substrato natural da enzima.

As estruturas 2D dessas moléculas foram construídas no programa MarvinSketch⁹⁰ onde também se fez o cálculo do pKa e a avaliação do padrão de protonação em função do pH. Para cada composto foi selecionado o estado de protonação mais relevante no pH de 8,3, de modo a se amostrar as moléculas nas mesmas condições utilizadas nos ensaios *in vitro*.

Ainda por meio do MarvinSketch fez-se a construção das estruturas tridimensionais dos compostos (3D *clean-up*) e sua exportação no formato *mol2* para a programa de interfaceamento gráfico GaussView⁹¹, do pacote Gaussian⁹².

Nesse programa, as estruturas de todos os compostos estudados, além do 3-deidrosnikotinato (substrato natural da AsbF) e 3,4-dihidrobencenoato (produto enzimático) foram novamente verificadas, as cargas totais assinaladas e fez-se a otimização geométrica no vácuo, por meio de cálculo quântico *ab initio* Restricted Hartree-Fock (RHF) com o conjunto de base 6-31G(*d,p*) utilizando-se o programa Gaussian09.⁹²

Após a etapa de otimização, os arquivos de saída do Gaussian09 no formato *log* foram convertidos para o formato *mol2* utilizando-se o programa Babel⁹³ e estes, por sua vez, foram utilizados na geração da topologia de entrada para o *docking*.

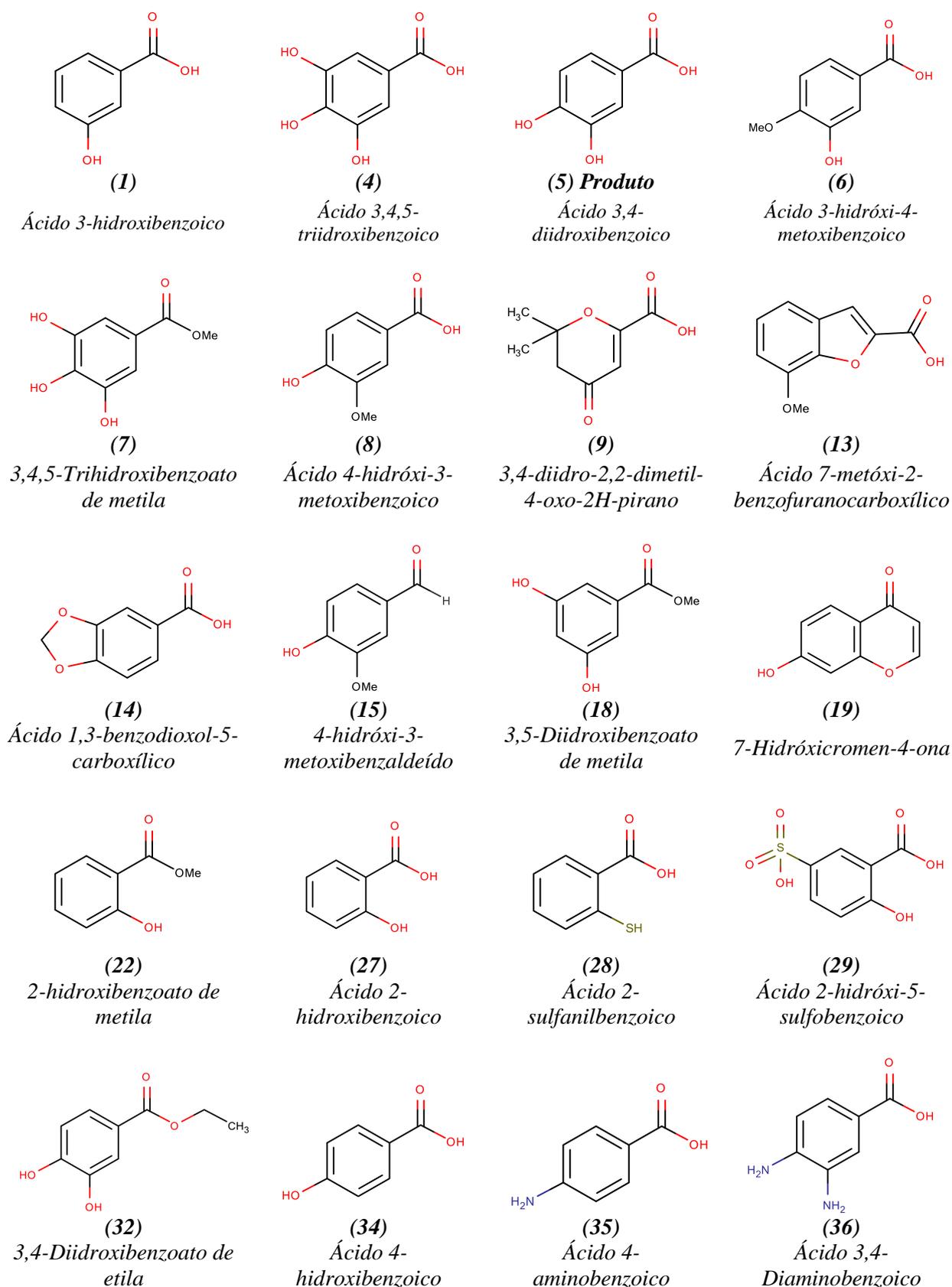


Figura 23 - Estrutura do substrato enzimático e dos compostos com potencial inibitório mensurável experimentalmente utilizados no presente estudo computacional, com respectivos códigos e nomenclatura IUPAC.

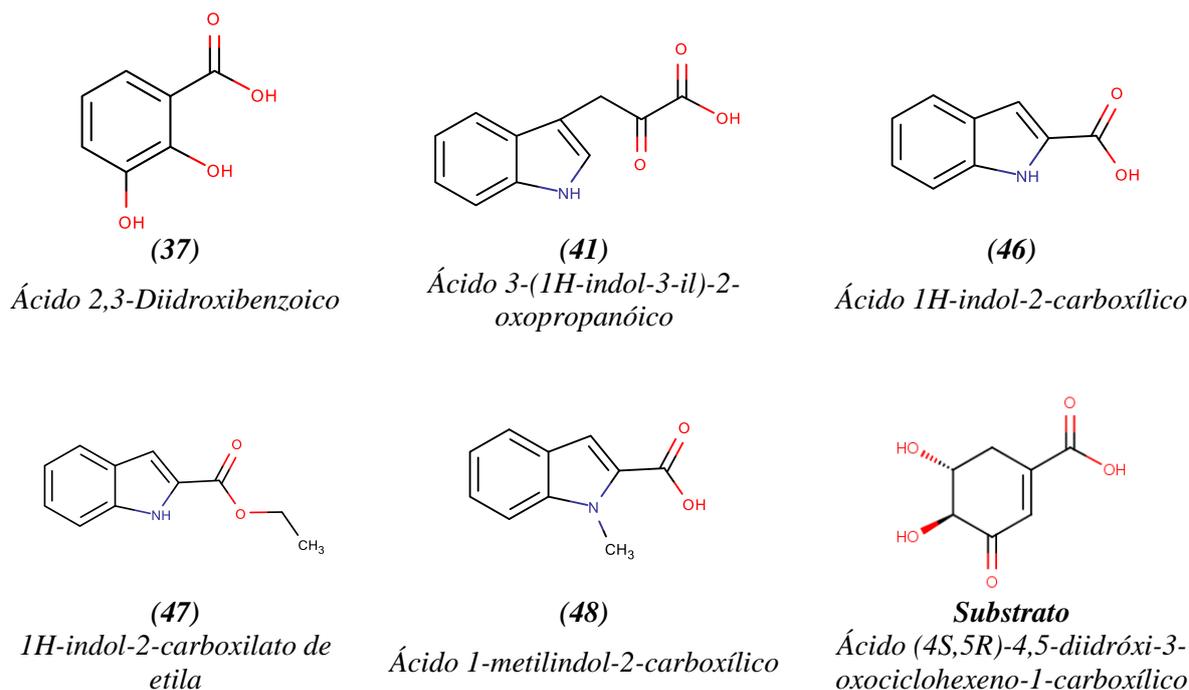


Figura 23 cont. – Estrutura do substrato enzimático e dos compostos com potencial inibitório mensurável experimentalmente utilizados no presente estudo computacional, com respectivos códigos e nomenclatura IUPAC

4.2 DOCKING MOLECULAR

A topologia e as coordenadas de entrada para o *docking* foram geradas a partir das coordenadas 3D da enzima AsbF no formato *pdb* após as modificações anteriormente descritas. Por meio do programa AutoDockTools (ADT)⁹⁴, o arquivo *pdb* foi importado, fez-se a adição de todos os hidrogênios em condições de pH neutro, o cálculo e adição das cargas Gasteiger, seguida da remoção dos hidrogênios não polares e fusão de suas cargas às dos átomos pesados. O programa ADT assinalou uma carga final de -14 para a proteína.

Além disso, os tipos atômicos AD4 foram assinalados e fez-se a exportação das coordenadas e cargas para o formato *pdbqt*, padrão de entrada para o programa Autodock Vina⁷⁶

De maneira análoga gerou-se a topologia de entrada para os ligantes. Os arquivos *mol2* dos ligantes foram importados para o AutoDockTools, os hidrogênios apolares foram unidos aos átomos pesados e os tipos atômicos AD4 assinalados, sendo ao final exportados no formato *pdbqt*.

Inicialmente, realizou-se um *docking* “cego”, isto é, um *docking* com um grid capaz de envolver toda a enzima, de modo a se avaliar a capacidade da metodologia de *docking* em reproduzir o modo de ligação semelhante ao encontrado na estrutura cristalográfico para o 3,4-dihidroxibenzoato (*redocking*).

Isso também permite avaliar a capacidade dos demais ligantes em interagir preferencialmente no sítio ativo em detrimento a outras porções da enzima, inclusive sua

superfície hidrofílica, além de verificar a existência de eventuais sítios de ligação não usuais, fora do sítio ativo.

Para o *docking* cego, utilizou-se um grid cúbico de dimensões 56x56x56 Å e espaçamento de 1,0 Å centrado em $x= 65.4$, $y= 70.43$ e $z= 2.16$, que correspondem ao centro geométrico da proteína, bem como uma *exhaustiveness* de 60, parâmetro de configuração do programa Vina controla quantas vezes os cálculos são repetidos, o que aumenta linearmente o tempo de execução do programa enquanto diminui exponencialmente a probabilidade de não se encontrar o mínimo global da função de escore.⁷⁶

Após o *docking* cego, como o sítio ativo e o centro geométrico da enzima se encontram muito próximos, fez-se a redução do tamanho do grid para 26 Å³ mantendo-o centrado nas mesmas coordenadas que as utilizadas para o *docking* cego e com o mesmo espaçamento, porém garantindo que o mesmo englobasse todos os resíduos que compõem o sítio ativo (Figura 24). Isso permitiu a realização de um *docking* mais refinado, apenas no sítio ativo.

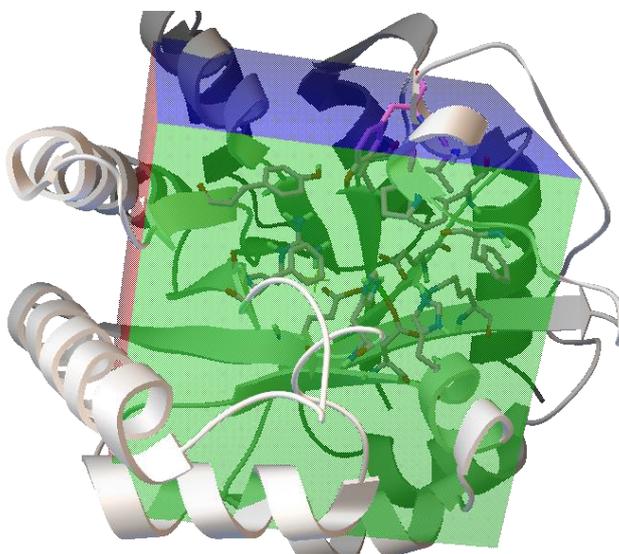


Figura 24- Grid utilizado para a realização do *docking* específico no sítio ativo, com os resíduos que compõem o sítio em destaque

Todos os *dockings* foram realizados em triplicata para cada ligante em seu estado de protonação mais relevante em pH 8,3. A média aritmética da menor energia de ligação obtida nas três rodadas de *docking* independentes foi calculada para todos os compostos, constituindo o escore final, expresso em kcal/mol.

4.3 DINÂMICA MOLECULAR

Todas as simulações de dinâmica molecular foram realizadas utilizando-se programas do pacote Gromacs⁹⁵⁻⁹⁷, versões 4.5.4 e 4.5.7 com o campo de forças Amber03.⁹⁸

Foram simulados oito sistemas: enzima livre em solução, enzima em complexo com produto, enzima em complexo com substrato e enzima em complexos com os cinco inibidores mais relevantes segundo apontado por resultados experimentais, de *docking* ou através da

comparação de ambos. O número de réplicas ou de simulações realizadas para cada sistema, juntamente com a estrutura dos ligantes avaliados estão mostrados na Tabela I.

Para as simulações da enzima livre em solução, a estrutura otimizada do receptor no formato *pdb*, obtida após as modificações anteriormente descritas, foi utilizada para a geração dos arquivos de entrada de topologia e de coordenadas para a dinâmica molecular por meio do programa *pdb2gmx*, do pacote Gromacs.

Como nenhum resíduo apresentou pKa entre 7 e 8,3, segundo análise com o programa Propka⁸⁹, optou-se por realizar a simulação em pH neutro e com a utilização dos estados de protonação pré-definidos pelo programa Gromacs, isto é, com todas as lisinas protonadas (3 hidrogênios), ácidos aspártico e glutâmico desprotonados e seleção automática de protonação para histidinas (protonados no ND1, ND2 ou ambos) seguindo critérios geométricos de otimização das ligações de hidrogênio formadas por esse resíduo.

O sistema foi então centrado em uma caixa cúbica (aprox. 660 nm³), cujas paredes distanciavam-se no mínimo 1,4 nm de qualquer resíduo da enzima a qual foi preenchida com aproximadamente 20.000 moléculas de água do tipo TIP3P.⁹⁹

Após a inserção de água fez-se a adição de íons sódio e cloreto (aproximadamente 80 Na⁺ e 60 Cl⁻) de forma a se neutralizar as cargas e aproximar o sistema de condições fisiológicas, reproduzindo uma concentração salina 0,154 mol.L⁻¹ ou de 0,9% NaCl (m/v).

Em seguida, fez-se uma minimização de energia por meio do algoritmo *steepest descent* utilizando-se como critério de convergência a força máxima de 2000 kJ.mol⁻¹.nm⁻¹ e tamanho de passo inicial de 0,01 nm. A convergência foi obtida em menos de 80 passos.

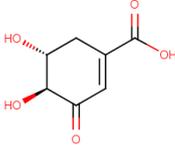
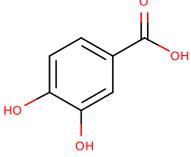
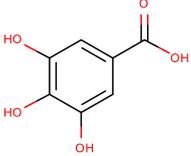
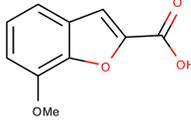
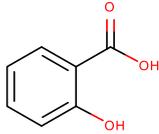
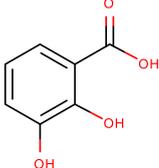
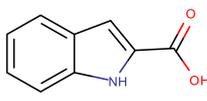
Após e como forma de equilibrar o sistema para a temperatura desejada, fez-se um conjunto de pequenas simulações de dinâmica molecular com restrições de posições, na qual apenas as moléculas do solvente e os íons sódio e cloreto foram permitidas de se moverem. Os átomos do receptor foram fixados em suas posições iniciais, podendo se mover às custas de uma grande penalidade energética, evitando assim um rearranjo drástico da estrutura pela inclusão do solvente e o aumento abrupto de temperatura.

Essa restrição, no entanto, foi sendo removida ao longo de 20 ns de simulação através da redução por um fator de 10 na constante de força isotrópica aplicada na restrição, passando de 1000 kJ.mol⁻¹.nm⁻² nos primeiros 5 ns, para 100 kJ.mol⁻¹.nm⁻² nos 5 ns seguintes, depois 10 kJ.mol⁻¹.nm⁻² e finalizando com 5 ns sem restrições.

Ao final dessas etapas de equilíbrio e termalização, a estrutura obtida foi considerada como o tempo inicial para a etapa de produção propriamente dita e a partir desse ponto realizaram-se quatro simulações utilizando-se diferentes sementes de geração de velocidade, de modo a se amostrar configurações distintas do espaço de fases a partir de pontos de partida ligeiramente diferentes.

A produção foi realizada com integrador MD do programa Gromacs, que utiliza um algoritmo *leap-frog* na integração das equações newtonianas do movimento, em 60 milhões de passos de 2,0 femtossegundos (2x10⁻¹⁵ s) cada, totalizando 120 nanossegundos de trajetória. As posições foram salvas em intervalos de 5 picossegundos.

Tabela I - Número de reproduções e estrutura do ligante para cada sistema simulado por dinâmica molecular

Sistema	Reproduções	Estrutura do Ligante
Enzima Livre	4	<i>Nenhum</i>
Enzima + Substrato	3	 <i>Ácido 3-dehidroshikímico (3-DHS)</i>
Enzima + Produto	3	 <i>Ácido 3,4-dihidroxibenzoico (3,4-DHB ou PCA)</i>
Enzima + Ligante 4	3	 <i>Ácido 3,4,5-triidrozibenzoico (Ácido gálico)</i>
Enzima + Ligante 13	2	 <i>Ácido 7-metóxi-1-benzofurano-2-carboxílico</i>
Enzima + Ligante 27	2	 <i>Ácido 2-hidroxibenzoico (Ácido salicílico)</i>
Enzima + Ligante 37	2	 <i>Ácido 2,3-dihidróxibenzoico (2,3-DHB)</i>
Enzima + Ligante 46	2	 <i>Ácido indol-2-carboxílico</i>

As simulações da enzima em complexo com ligantes foram realizadas de uma forma levemente diferente, sobretudo na etapa de inicialização, em que as coordenadas do complexo foram obtidas a partir do modo de ligação de menor energia apontado por *docking*, as quais, em geral, correspondem aos ligantes posicionados no interior do sítio ativo.

A geração da topologia e das coordenadas iniciais do receptor foi realizada da mesma maneira que para a enzima livre. A geração das coordenadas de entrada e da topologia dos ligantes, entretanto, envolveu a separação do modo de ligação de menor energia dos demais modos apontados pelo *docking* e sua conversão do formato *pdbqt* para *mol2* através do programa *Babel*.⁹³ Nessa conversão os hidrogênios foram reinseridos na estrutura – para o *docking* apenas os hidrogênios polares haviam sido adicionados – respeitando-se o estado de protonação presente em pH 8,3.

Dessa forma, ainda que o pH experimental não seja amostrado explicitamente nas simulações de dinâmica, sua principal consequência, que é estado de protonação dos ligantes – todos ácidos carboxílicos – acaba sendo levada em consideração de maneira indireta.

Após essa conversão, as coordenadas de entrada no formato *gro*, próprio do *Gromacs*, foram geradas pelo programa *AcPype*¹⁰⁰, que também gerou a topologia do ligante utilizando o campo de forças *GAFF* (Generalized Amber Force Field)¹⁰¹.

Esse campo de forças, compatível com os campos de força da família *Amber* – de quem toma emprestado alguns parâmetros como σ_{ij} e ϵ_{ij} do potencial de tipo Lennard-Jones – foi parametrizado a partir de diversos resultados experimentais e de cálculos quânticos de alto nível (MP2 e MP4) gerando um campo de força com parâmetros mais gerais e que pode ser aplicado à maioria dos compostos orgânicos, diferentemente dos campos de força da família *Amber*, que são parametrizados apenas para macromoléculas biológicas.¹⁰¹ Dessa forma, a aplicação do *GAFF* para os ligantes e do *Amber03* para o receptor biológico se mostra uma escolha adequada e coerente.

As cargas dos ligantes também foram calculadas pelo programa *AcPype* e incluídas na topologia. No cálculo das cargas optou-se pelo método semi-empírico *AM1-BCC*¹⁰², que foi parametrizado para reproduzir as cargas *RESP HF/6-31G(d)* e validado em conjunto com o campo de forças *Amber*, reproduzindo diversos resultados experimentais, como as energias livres de solvatação, com coeficientes de correlação superiores a 0,96, e sendo por isso considerado uma alternativa rápida, robusta e precisa ao cálculo quântico de cargas, em meio polar.

Após a geração das topologias do receptor e ligantes, os complexos foram centrados em caixas cúbicas com paredes distando em 1,4 nm dos átomos do receptor (aprox. 660 nm³), a qual foi preenchida com águas *TIP3P* e íons sódio e cloreto, em praticamente o mesmo número e da mesma forma que o realizado para a enzima livre, seguida de uma breve minimização de energia, aplicando-se os mesmos parâmetros de convergência, a qual foi obtida em aproximadamente 150 passos.

Após, fez-se a equilibração dos sistemas por meio de uma dinâmica com restrições de posições (constante de força 1000 kJ.mol⁻¹.nm⁻² para enzima e ligantes), a qual foi realizada em 50.000 passos de 1,0 femtossegundo, totalizando 50 ps de simulação, o que foi julgado suficiente para eliminação ou minimização das interações desfavoráveis soluto-solvente que podem ter sido geradas nas etapas de inicialização do sistema.

Ao final dessa equilibração, a estrutura obtida foi considerada como as coordenadas do tempo inicial para as etapas de produção da trajetória. As simulações foram realizadas em duplicata ou triplicata (cf. Tabela I) mantendo-se a mesma semente de geração de velocidades. O integrador MD foi utilizado na geração dessas trajetórias, em 50 milhões de passos de 2,0 femtossegundos cada, totalizando 100 ns de simulação. As coordenadas foram armazenadas em intervalos de 10 picossegundos.

Todas as simulações foram realizadas no *ensemble* isotérmico-isobárico (NPT), isto é, com número de partículas, pressão e temperaturas constantes. A temperatura do sistema foi ajustada para 303 K (30°C) a fim de reproduzir as condições empregadas nos ensaios experimentais *in vitro* utilizando o termostato V-rescale¹⁰³ ao passo em que a pressão padrão de 100.000 Pa (1 bar) foi ajustada isotropicamente por meio do barostato de Parinello-Rahman¹⁰⁴ com coeficiente de compressibilidade de $4,5 \times 10^{-5} \text{ bar}^{-1}$, valor característico para a água à 30°C.

A atualização da lista de átomos vizinhos para o cálculo das forças de curto alcance foi feita em intervalos de 10 passos, utilizando-se um raio de corte de 1,3 nm, mesmo valor do raio de corte aplicado no cálculo das interações do tipo Van der Waals, calculadas pelo método de cut-off simples. As interações de longo alcance, por sua vez, foram calculadas pelo método Particle-Mesh Ewald (PME)¹⁰⁵ com raio de corte de 1,3 nm, com espaçamento de grid de 0,12 nm e interpolação cúbica.

Um exemplo do arquivo de parâmetros *mdp* utilizado nas etapas de produção das trajetórias desse trabalho está mostrado no Apêndice B, em que podem ser observados aspectos mais detalhados das configurações do Gromacs aplicadas nas simulações.

4.4 ENSEMBLE DOCKING

Diversas estratégias foram avaliadas na seleção das conformações biologicamente mais relevantes da enzima para a construção do conjunto de estruturas e realização do *ensemble docking* a partir das trajetórias de DM, uma vez que seria impossível realizar-se o *docking* de todos os ligantes em todas as estruturas da enzima geradas pelas simulações de dinâmica molecular.

A primeira e mais simples de todas as estratégias foi a extração do último frame de cada trajetória, o que corresponde ao tempo de 120 ns para as simulações da enzima livre em solução e de 100 ns para suas simulações em complexo com os ligantes, assumindo-se que após esse tempo os sistemas estariam plenamente equilibrados e se encontrariam em seus estados mais estáveis nas condições computacionais utilizadas.

A segunda estratégia foi a busca das conformações que maximizassem as interações receptor-ligante com cada um dos compostos simulados, e que está baseada na teoria do encaixe induzido, segundo o qual os ligantes induzem o receptor a conformações que maximizem suas interações com o mesmo, as quais jamais ocorreriam na ausência desses compostos.

Especificamente, buscou-se extrair o frame de cada trajetória no qual houvesse a formação do maior número de ligações de hidrogênio, visto que essas ligações além de serem uma das interações não covalentes mais intensas comumente encontradas em sistemas biológicos também podem ser avaliadas diretamente a partir das trajetórias com programas implementados

no próprio pacote Gromacs. A localização temporal e o número de ligações de hidrogênio formadas em cada um desses frames estão detalhados na Tabela II.

Nos casos em que mais de um frame apresentou o mesmo número máximo de ligações de hidrogênio optou-se pela seleção do frame mais próximo do final da trajetória, pois esse deve corresponder a uma estrutura mais equilibrada e estável quando comparada com estruturas extraídas do início das simulações.

Tabela II - Número de ligações de hidrogênio e localização dos frames em que ocorrem o maior número de ligações de hidrogênio simultâneas para cada réplica de cada ligante.

Simulação	Tempo (ps)	Número de Ligações de Hidrogênio
Enzima + Substrato R1	33330	9
Enzima + Substrato R2	69590	11
Enzima + Substrato R3	84760	8
Enzima + Produto R1	59600	9
Enzima + Produto R2	43280	8
Enzima + Produto R3	7980	9
Enzima + Ligante 04 R1	97680	9
Enzima + Ligante 04 R2	99740	9
Enzima + Ligante 04 R3	17020	10
Enzima + Ligante 13 R1	45440	5
Enzima + Ligante 13 R2	99100	5
Enzima + Ligante 27 R1	97160	4
Enzima + Ligante 27 R2	7620	5
Enzima + Ligante 37 R1	94450	8
Enzima + Ligante 37 R2	61510	9
Enzima + Ligante 46 R1	45000	4
Enzima + Ligante 46 R2	54310	6

Essas duas estratégias, no entanto, compreendem pontos muito específicos do espaço de fases conformacional, que não necessariamente correspondem às conformações biologicamente ativas. Assim, na terceira estratégia buscou-se gerar um *ensemble* de estruturas significativamente representativas da mobilidade estrutural do receptor considerando-se a totalidade das trajetórias. Para isso aplicou-se a análise de clusters usando-se o método *single-linkage* como implementado na ferramenta *g_cluster* do Gromacs.

Por esse método, uma dada conformação é adicionada a um cluster sempre que sua distância média (avaliada por meio do desvio quadrático médio das posições atômicas, o RMSD) a qualquer outra estrutura do cluster for inferior a um dado raio de corte. Uma vez formado um cluster, as conformações da proteína que o compõem são "removidas" da trajetória e o processo é repetido para as estruturas restantes até que todas as estruturas sejam incluídas em algum cluster. Após todos os clusters serem formados, a conformação com o menor RMSD, portanto com a menor distância média em relação às outras estruturas que compõem o seu cluster é selecionada para representar aquele cluster.¹⁰⁶

Na geração de clusters, duas formas de se medir as distâncias entre as estruturas foram consideradas. A primeira delas foi incluir no cálculo da matriz de RMSD todos os carbonos- α (CA), como forma de se considerar as alterações conformacionais globais do receptor enzimático enquanto que na segunda apenas os resíduos do sítio ativo (SA) foram utilizados nesse cálculo,

porém incluindo-se todos os átomos que os compõem, de forma que alterações locais e sobretudo alterações das cadeias laterais dos resíduos do sítio, pudessem ser avaliadas.

Os raios de corte utilizados foram de 0,5 Å para a clusterização do sítio ativo e de 0,6 Å na clusterização dos carbonos- α . Esses valores significativamente baixos foram escolhidos para que pudessem ser gerados de algumas dezenas até centenas de clusters. Isso permite separar de dentro das trajetórias, em geral formadas de 10.000 estruturas cada, as conformações mais frequentes e, conseqüentemente, mais estáveis, daquelas que ocorrem poucas vezes ao longo das simulações.

Supõem-se que, ainda que não sejam necessariamente mínimos globais, as estruturas biologicamente ativas devem ser mínimos locais na superfície de energia livre potencial, sobretudo se de fato há a seleção e estabilização das mesmas pela ação dos ligantes, como pressupõem os modelos mais modernos de racionalização da formação de complexos proteína-ligante.

Dessa forma, a seleção dos clusters mais frequentes e com maiores números de estruturas (mais populosos) se mostra adequada na busca das conformações enzimáticas ativas biologicamente, pois apesar de se gerar um grande número de clusters, em geral se obtém um número pequeno de estruturas estáveis, como está exemplificado na Figura 25, que mostra um dos gráficos do número de clusters formados em função do tempo de simulação, isto é, a identidade de cada cluster ao longo da trajetória, com os patamares indicando os clusters que são recorrentes (e, portanto, mais populosos).

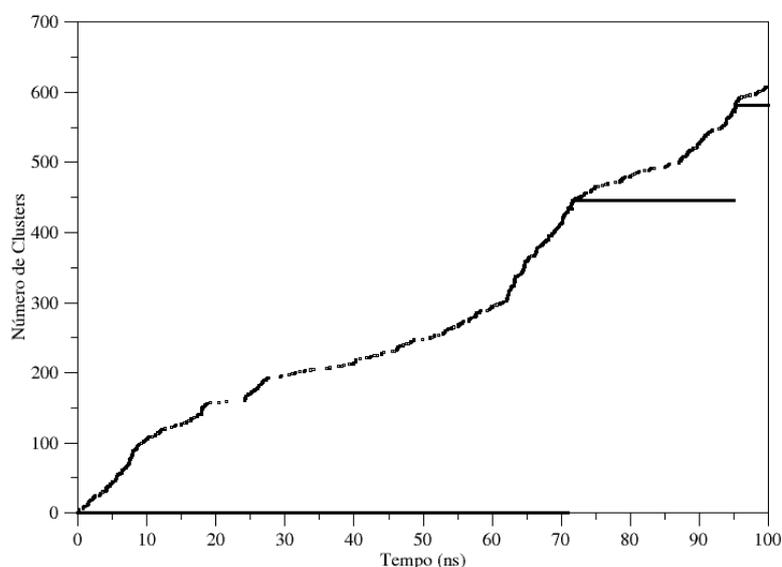


Figura 25 - Número de clusters formados em função do tempo de simulação para a réplica 1 das simulações enzima-substrato. Clusterização por RMSD dos carbonos- α .

Pode-se observar na figura que, ainda que mais de 600 clusters tenham sido formados, apenas três têm populações significativas de estruturas e somente estes foram selecionados para compor o *ensemble*. Pode-se perceber ainda como essas outras inúmeras conformações são apenas oscilações momentâneas que rapidamente retornam a um dos três mínimos, esses sim, representativos de três estados conformacionais pelos quais a enzima passa ao longo da simulação, o último deles nos 10 nanossegundos finais.

Na Tabela III estão discriminados o número total de clusters formados, tanto pela clusterização com carbonos- α quanto com o sítio ativo, bem como o número de clusters principais, que foram incluídos no *ensemble*. No Apêndice C estão discriminadas as identidades de cada um desses clusters principais.

Tabela III - Número total de cluster formados e número de cluster principais em cada metodologia de clusterização aplicada para cada simulação.

Simulação	Clusterização Carbonos- α		Clusterização Sítio Ativo	
	Clusters	Principais	Clusters	Principais
Enzima Livre S1	1449	3	527	5
Enzima Livre S2	933	4	988	4
Enzima Livre S3	1003	4	311	3
Enzima Livre S4	807	3	675	3
Enzima + Substrato R1	608	3	692	6
Enzima + Substrato R2	903	6	1611	4
Enzima + Substrato R3	721	2	384	2
Enzima + Produto R1	353	2	827	6
Enzima + Produto R2	272	2	1820	5
Enzima + Produto R3	947	3	1436	4
Enzima + Ligante 04 R1	1022	6	1581	3
Enzima + Ligante 04 R2	450	4	797	2
Enzima + Ligante 04 R3	341	5	750	5
Enzima + Ligante 13 R1	1112	4	346	5
Enzima + Ligante 13 R2	440	4	389	8
Enzima + Ligante 27 R1	634	2	218	3
Enzima + Ligante 27 R2	849	2	305	3
Enzima + Ligante 37 R1	276	2	401	4
Enzima + Ligante 37 R2	345	3	381	4
Enzima + Ligante 46 R1	853	5	950	5
Enzima + Ligante 46 R2	347	3	751	7

Para cada cluster principal, extraiu-se da respectiva trajetória a estrutura média, no formato *pdb*. Essas estruturas em conjunto com estruturas finais e aquelas em que se formam os maiores números de ligações de hidrogênio simultâneas constituíram um *ensemble* com cerca de 200 estruturas da enzima.

A preparação dessas estruturas para o *docking* envolveu a remoção do ligante, quando presente, seguida conversão do arquivo *pdb* para *pdbqt* utilizando-se o plug-in “prepare_receptor4.py” do PyMOL, que é parte do pacote de programas AutoDockTools. Nessa conversão há também a adição de cargas Gasteiger, remoção dos hidrogênios não-polares e fusão das cargas dos hidrogênios não-polares aos átomos pesados.

Apesar de se fazer o ajuste das trajetórias de DM à estrutura inicial, com correção de rotação e translação, as coordenadas das estruturas extraídas das dinâmicas não correspondem necessariamente às mesmas da estrutura inicial. Por essa razão, fez-se o cálculo do centro geométrico de cada estrutura, ao redor do qual foi definido o volume de busca do programa de *docking* AutoDock Vina, constituído de um *grid* cúbico de dimensões 26 Å x 26 Å x 26 Å e espaçamento de 1,0 Å.

Como essa enzima apresenta um sítio ativo centralizado, muito próximo de seu centro geométrico, as dimensões de *grid* e seu posicionamento foram adequadas para envolver toda a

porção central da enzima e o sítio ativo, porém desprezando sua superfície, como forma de se evitar falsos-positivos e se obter resultados mais precisos.

Para cada um dos 25 ligantes estudados fez-se o *docking* em triplicata com cada estrutura, totalizando mais de 15.000 rodadas de *docking* individuais. O parâmetro de *exhaustiveness* foi ajustado em 60 e as 10 poses energeticamente mais favoráveis foram geradas como arquivo de saída do *docking*.

A média aritmética da pose de menor energia obtida nas três rodadas de *docking* foi calculada, compondo o escore final em kcal/mol. A correlação linear entre as médias dos escores de *docking* e os valores de energia livre experimentais dos ligantes foram computadas para cada estrutura do *ensemble*. Os parâmetros coeficiente de correlação de Pearson, coeficiente de determinação (COD ou r^2), erro quadrático médio (RMSE) e a inclinação da curva de correlação foram utilizados como critério de qualidade dos modelos gerados.

5 RESULTADOS E DISCUSSÃO

5.1 DOCKING MOLECULAR

Na Tabela IV estão mostrados os valores experimentais de IC₅₀ gentilmente fornecidos por Koppisch e colaboradores bem como os valores energia livre calculados segundo a equação (7), na temperatura experimental de 303 K. Dentre esses valores, destacam-se os ligantes 4 (ácido gálico) e 7 (galato de metila), bem como o ligante 5 (ácido 3,4-dihidroxibenzoico) que é o produto da reação enzimática.

De fato, na estrutura cristalográfica depositada por Pflieger e colaboradores⁵⁸ se observa o ácido 3,4-DHB coordenado ao sítio ativo da enzima, porém esta é a primeira comprovação experimental de um mecanismo de retro-inibição para a enzima AsbF.

Tabela IV - Valores de IC₅₀ determinados experimentalmente por Koppisch e colaboradores, em ordem ascendente, e energia livre de Gibbs de ligação calculadas segundo equação (7) a 303 K.

Código	IC ₅₀ (μmol.L ⁻¹)	ΔG° (kcal.mol ⁻¹)	Código	IC ₅₀ (μmol.L ⁻¹)	ΔG° (kcal.mol ⁻¹)
04	4.36	-7.43	36	273.0	-4.94
07	37.5	-6.14	14	286.7	-4.91
05	38.0	-6.13	35	304.4	-4.88
22	56.7	-5.89	18	350.1	-4.79
13	68.5	-5.77	48	355.8	-4.78
01	71.5	-5.75	41	386.5	-4.73
46	90.0	-5.61	27	423.7	-4.68
06	92.1	-5.60	08	440.8	-4.65
28	108.2	-5.50	47	480.1	-4.60
09	190.7	-5.16	34	536.4	-4.53
29	221.3	-5.07	37	542.8	-4.53
32	225.0	-5.06	15	857.7	-4.25
19	246.4	-5.00			

Dos 25 compostos ativos, 19 são derivados mono-, di- ou tri-substituídos do ácido benzoico ou apresentam alguma similaridade estrutural com este, indicando a existência de um mecanismo de inibição enzimática comum a todos esses ligantes. Merecem destaque, no entanto, exceções como os ligantes 13 (ácido 7-metóxi-2-benzofuranocarboxílico), 46 (ácido indole-2-carboxílico), e 19 (7-hidroxicromona), que se mostraram relativamente potentes, apesar da estruturalmente diferentes do demais compostos do conjunto.

A compreensão dos detalhes acerca deste mecanismo comum de inibição, bem como acerca das razões que justificam a potência dos ligantes *outliers* é uma das motivações primordiais do presente estudo computacional.

Esse estudo iniciou-se com a realização do *docking* “cego” dos 25 ligantes na estrutura cristalográfica da AsbF, ajustada para reproduzir as condições experimentais utilizadas por Koppisch et al.⁸⁸

Na Figura 26, estão sobrepostos o melhor modo de ligação para o 3,4-DHBA gerado pelo *docking* com programa Vina, utilizando um grid envolvendo toda a enzima, e o modo de ligação observado estrutura cristalográfica conforme depositada no PDB.⁵⁸



Figura 26 - Sobreposição do melhor modo de ligação do 3,4-dihidroxibenzoato obtido no *docking* ampliado (rosa) com a estrutura cristalográfica depositada no PDB (verde). Em destaque o cofator magnésio.

Pode-se observar que o modo de ligação mais favorável segundo o *docking* é coincidente com o resultado experimental, com um desvio quadrático médio (RMSD) de 0,95 Å, o que pode ser considerado satisfatório, na medida em que este valor é inferior à própria resolução da estrutura cristalográfica, que é de 2,1 Å.

O principal ponto de divergência entre os modos de ligação é a conformação do grupo carboxilato, que se mostra levemente torcido em relação ao anel aromático, ao contrário do que seria esperado pela existência de conjugação com o anel. Isso evidencia uma falha do algoritmo de busca do programa de *docking*, ao considerar a ligação entre o carbono carboxílico e o anel como passível de rotação.

Essa rotação “artificial” produzida pelo algoritmo acaba sendo responsável pela obtenção de um padrão de interações levemente diferentes entre o modo de ligação experimental e o modo apontado por *docking*, como pode ser verificado na Figura 27.

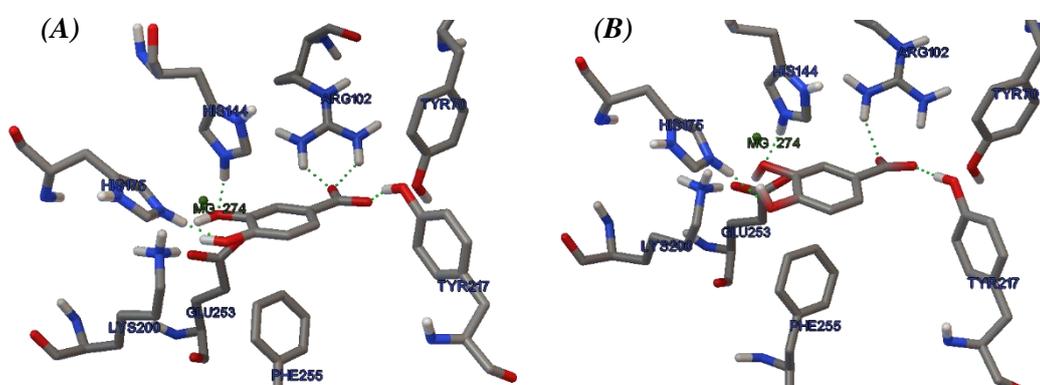


Figura 27- Padrão de interações determinado pelo programa ADT para o 3,4-DHB conforme estrutura cristalográfica (A) e o melhor modo de ligação do *docking* (B). As ligações de hidrogênio estão pontilhadas em verde.

Utilizando a ferramenta de visualização de interações do programa AutoDockTools com fator de escala de Van der Waals de 1 Å, se pode observar que no modo de ligação cristalográfico (Figura 27A), há a formação de 5 ligações de hidrogênio, sendo duas entre os resíduos His144 e His175 e as hidroxilas do ligante e três entre a porção carboxilato do 3,4-DHB e os resíduos Tyr217 e Arg102, essa última interação com características de ponte salina, agregando duas ligações de hidrogênio e uma interação eletrostática entre o grupamento guanidínico positivamente carregado e o ânion carboxilato.

Além dessas interações, há o contato próximo com a cadeia lateral da Lys200 e Glu253 e sobretudo com o anel aromático da Phe255, que poderiam auxiliar na ligação por meio de interações de Van der Waals ou mesmo através da formação de uma “bolsa” hidrofóbica para a porção aromática do ligante.

É interessante notar que esse padrão é exatamente o mesmo que o obtido por meio do *docking* (Figura 27B), sendo a única diferença a interação com a arginina, com a qual ocorre a formação de apenas uma ligação de hidrogênio em virtude da rotação artificial do grupo carboxilato, como descrito anteriormente.

Esse resultado, ao lado do RMSD relativamente baixo entre o modo de ligação teórico e o experimental, demonstra que a metodologia de *docking* é adequada para reproduzir as interações receptor-ligante da enzima AsbF.

Diante disso, a metodologia poderia ser utilizada para a obtenção de informações inacessíveis experimentalmente, como a observação do padrão de interações e do modo de ligação do substrato com o sítio ativo, que é a primeira etapa envolvida no mecanismo da reação enzimática.

Por apresentar um anel alifático, porém com uma insaturação, o substrato 3-deidroschikimato possui, em princípio, duas conformações em equilíbrio, que correspondem a mínimos energéticos na superfície de energia potencial conformacional. Na primeira delas, os substituintes hidroxila estão em posições axiais (Figura 28A) enquanto que no segundo mínimo energético tais substituintes encontram-se em posição equatorial (Figura 28B).

As energias para essas duas conformações foram avaliadas por cálculo quântico *ab initio* Hartree-Fock Restrito com conjunto de base 6-31G(d,p), obtendo-se para a conformação (A) uma energia de -643,4947873 Ha ao passo em que para conformação com os grupos hidroxila em posição equatorial (B) obteve-se -643,518046 Ha, portanto com uma diferença energética no vácuo de 0,0232587 Ha ou 14,60 kcal/mol. Com isso, a razão entre as populações de Boltzmann das conformações é 2×10^{11} mais favorável àquela com grupos OH em posição equatorial (B), na temperatura de 298 K.

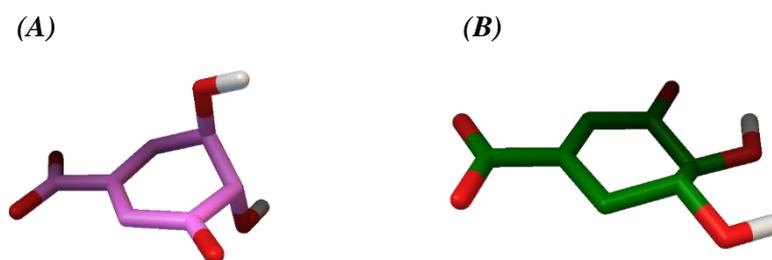


Figura 28 - Mínimos conformacionais possíveis para o 3-deidroschikimato, com os grupos hidroxila em posição axial (A) ou equatorial (B).

De fato, o confômero A não pode ser docado adequadamente no sítio ativo da enzima, como mostra a Figura 29, onde se observa que os 9 modos de ligação apontados pelo programa Vina encontram-se na superfície enzimática e significativamente distantes do sítio principal de ligação.

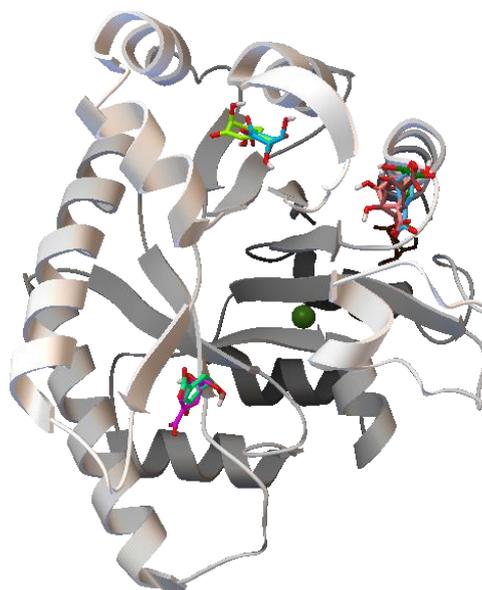


Figura 29 - Modos de ligação do substrato na conformação axial obtidos por *docking* cego. O cofator magnésio em destaque aponta a posição aproximada do sítio ativo da AsbF.

Por essas razões, a conformação axial não foi mais considerada nesse estudo computacional – visto que provavelmente não tem relevância do ponto de vista experimental – e todas as referências ao substrato, a partir desse ponto, referem-se a sua forma conformacional mais abundante, que é a equatorial.

Na Figura 30 estão mostradas as principais interações do substrato presentes no modo de ligação de menor energia segundo o Vina. Esse modo de ligação foi apontado pelo programa como tendo a menor energia em duas das três réplicas de *docking*.

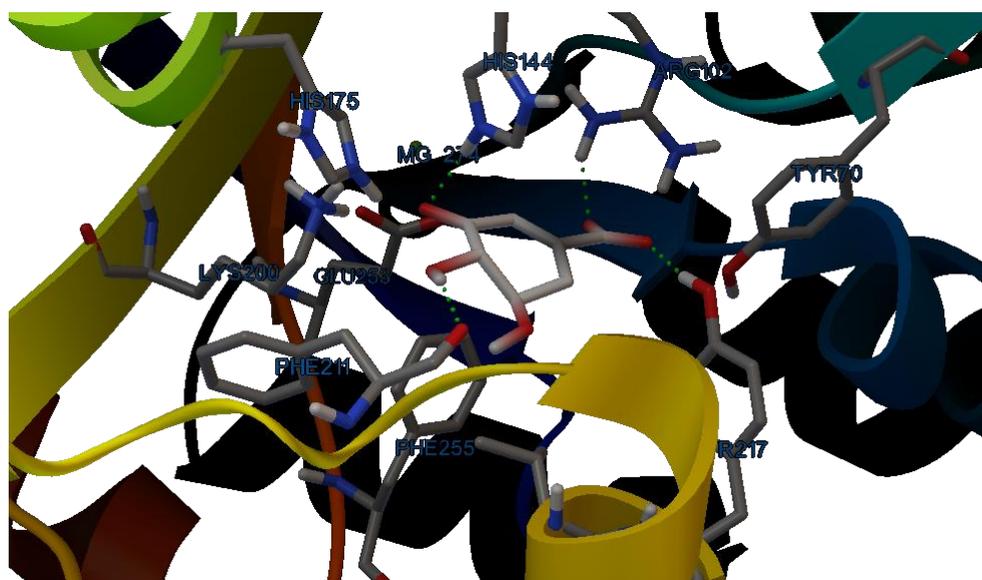


Figura 30 - Principais interações do substrato com o sítio ativo da enzima AsbF para a modo de ligação de menor energia segundo o *docking*.

Pode-se observar que o padrão de ligações de hidrogênio é muito similar àquele observado para o produto (cf. Figura 27), com a adoção de uma conformação praticamente idêntica a este, o que permite, inclusive, especular sobre o mecanismo da reação enzimática por meio da comparação entre os modos de ligação do substrato e do produto.

Observa-se a formação de quatro ligações de hidrogênio, uma das quais entre a porção carboxilato do substrato e a Tyr217 e, novamente, uma ponte salina entre essa porção do ligante e a Arg102. Dessa vez, no entanto, uma das hidroxilas aparece interagindo com o oxigênio do esqueleto peptídico (Phe211) ao invés da His175, como apontado no *docking* do produto.

O mais interessante, no entanto, é a formação de uma ligação de hidrogênio entre a His144 e o oxigênio carbonílico do anel, interação que é mantida após sua conversão a hidroxila. É desta histidina que provavelmente advém o hidrogênio da hidroxila, numa etapa final do mecanismo.

Imaginando-se que, inicialmente, a His144 se encontre desprotonada – o que seria factível no pH experimental, segundo pontado pelo programa PROPKA⁸⁹. A aproximação do substrato e seu posicionamento se daria de tal modo que houvesse a formação de um complexo entre sua carbonila e o magnésio, típicos base e ácido de Lewis “duros”, formando um complexo relativamente estável.

Essa complexação produziria dois efeitos: o primeiro seria um rearranjo de cargas no substrato, eventualmente aumentando a eletrofilicidade do hidrogênio axial na posição 4 do anel. O segundo efeito estaria relacionado ao uma leve alteração conformacional da His144 gerado pela movimentação do resíduo vizinho Glu142, que é parte do sítio de ligação do magnésio (Figura 31). O complexação do substrato poderia enfraquecer a ligação do Mg²⁺ a este ácido glutâmico, induzindo um leve rearranjo da estrutura secundária, que favoreceria o posicionamento adequado para o ataque nucleofílico da His144 ao hidrogênio axial do C4, levando à segunda insaturação do anel.

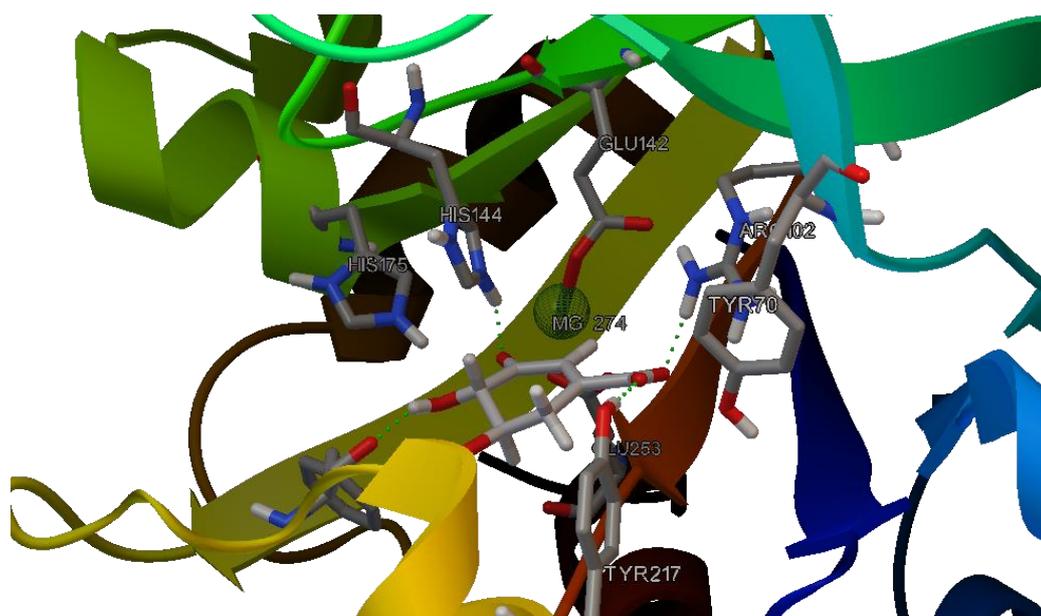


Figura 31- Modo de ligação mais favorável do substrato ao sítio ativo, com destaque para o sítio de ligação do magnésio, os ácidos glutâmicos 142 e 253. O ataque nucleofílico da His144 ao hidrogênio axial vizinho à carbonila levaria ao surgimento da segunda insaturação do anel do substrato.

Os resultados do *docking*, portanto, reforçam a primeira etapa do mecanismo proposto por Pflieger et al.⁵⁸, que sugere a participação da His144 na formação da segunda insaturação do anel. Todavia, a formação da terceira insaturação, que leva à aromatização, é um pouco mais complexa.

Pflieger e colaboradores sugerem que, sob ação de uma base “desconhecida”, ocorra a captura de um dos hidrogênios da posição 6 e o rearranjo eletrônico do sistema, com expulsão da hidroxila em posição 5 na forma de água e consequente aromatização do anel. A partir da observação dos resíduos próximos e das interações de *docking*, pode-se especular que a base envolvida no ataque nucleofílico a este hidrogênio seja a própria Tyr217 (Figura 32).

Apesar do pKa calculado para esse resíduo ser bem acima do esperado para o aminoácido livre (13,06 contra o 10,0 para o aminoácido) em razão da sua localização interna na enzima e proximidade a resíduos hidrofóbicos, é possível que as alterações estruturais geradas pela complexação do substrato induzam rapidamente uma diminuição do pKa desse resíduo e sua desprotonação, propiciando a atuação deste como uma base.

Naturalmente, a atuação desse resíduo como base é tão melhor quanto mais alto for o pH do meio. Isso justifica porque a atividade da enzima aumenta como o aumento do pH, até o limite de pH 8,3, após o qual ocorreria a desnaturação da enzima e queda abrupta da atividade, como mostrado por Fox e colaboradores⁵⁷, indicando que essa é a etapa lenta do mecanismo.

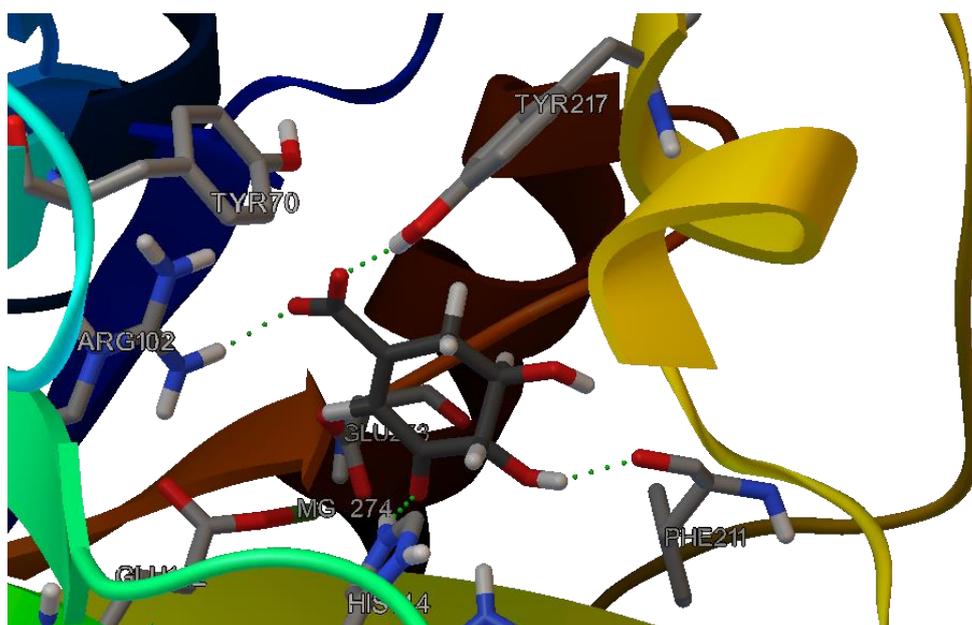


Figura 32- Posicionamento da Tyr217 (topo) em proximidade com os hidrogênios do carbono-5 do produto na estrutura docada indicando sua provável atuação como base. Interações do substrato com o esqueleto peptídico da Phe211 (direita), possivelmente auxiliando na desidratação do substrato.

O ataque da Tyr217 remove então um dos hidrogênios na posição 6, produzindo a terceira insaturação e formação do anel aromático com a expulsão de água na posição 5 (desidratação). Essa expulsão seria favorecida por meio da atuação do esqueleto peptídico da Phe211, inclusive através de um equilíbrio amida-iminol conforme sugerido por Pflieger et al. Ainda que esse mecanismo seja pouco usual em solução, ele é plenamente factível no

microambiente do sítio ativo graças aos efeitos de proximidade, hidrofobicidade e até mesmo influência de alterações conformacionais.

Finalmente, a His144 devolve o hidrogênio que havia capturado, porém o faz para o oxigênio em posição C3 do recém-formado produto, que pode então se descomplexar do magnésio e deixar o sítio ativo ou então interagir com esse, inibindo-o. A interação do grupo 3-hidróxi do produto com essa histidina é observada no *docking* do produto e na estrutura cristalográfica (cf. Figura 27).

Ainda que o *docking* possa auxiliar na proposição de um mecanismo ou mesmo reproduzir modos de ligação experimental, como visto aqui para o 3,4-DHB, há uma grande dificuldade na estimativa da intensidade dessas interações, especialmente sua avaliação por meio de uma função de escore. Esse obstáculo, entretanto, não é exclusividade da função de escore do programa Vina, tampouco é uma falha unicamente das metodologias de *docking*.¹⁰⁷

A função de escore do programa AutoDock Vina, ao contrário, é reconhecida por seu grau de precisão elevado quando comparado com outros softwares, comerciais ou livres, conforme demonstrado por Wang e colaboradores (2016), por meio da avaliação de 2002 complexos receptor-ligantes com afinidades de ligação conhecidas, obtidos da literatura.¹⁰⁸

Ao comparar as energias de ligação fornecidos pelo Vina com os dados experimentais, Wang et al. obtiveram os valores de 0.564 e 0.580 para os coeficientes de correlação de Pearson e Spearman, respectivamente. Esses valores, apesar de baixos, foram significativamente melhores que os obtidos utilizando-se outros 9 programas de *docking*, incluindo Glide, GOLD e Surflex-Dock.¹⁰⁸

Na Tabela V estão mostrados os valores de energia de ligação obtidos pelo Vina para os 25 inibidores da AsbF aqui estudados, considerando-se um grid envolvendo toda a enzima (*docking* cego, DC) e um grid envolvendo apenas a porção do sítio ativo (DSA). Como o programa Vina estima as energias de ligação em 300 K, as energias livres experimentais foram ajustadas para essa mesma temperatura por meio da equação 7.

Os escores de *docking* oscilaram entre -7,5 e -5,0 kcal/mol para as duas metodologias aplicadas, com um desvio em relação à energia de ligação experimental da ordem de 23%, em média, tanto para o *docking* cego quanto para o *docking* de sítio ativo, o que é bastante elevado.

Ainda assim, observa-se que o *docking* foi capaz de discriminar satisfatoriamente o inibidor mais potente (ligante 4) e o menos potente (ligante 15), inclusive produzindo para 4 uma energia de ligação com apenas 0,5% de desvio do valor experimental, sobretudo com o *docking* mais refinado.

Também é interessante notar que o produto enzimático (ligante 5) foi ranqueado com a terceira energia mais negativa, a mesma posição do ranking experimental, ainda que apresente um desvio de 20% em relação ao experimento. Para esse ligante em específico, a metodologia de *docking* foi capaz de prever com relativa satisfação tanto o modo como a intensidade de ligação, o que é surpreendente.

Houve, porém, diversos falsos positivos. Ligantes com energia livre experimental bem altas, mas que tiveram sua interação superestimada pelo *docking*, resultando em desvios elevados, com destaque para os ligantes 27 (o sexto menos ativo) e sobretudo para o 37, que mesmo tendo a segunda menor atividade experimental, obteve a energia livre de interação mais negativa de todas (-7,5 kcal/mol) e conseqüentemente um desvio de mais de 67%.

Tabela V - Afinidades de ligação estimadas pelo *docking* com Vina utilizando-se grid envolvendo toda a enzima (DC) ou restrito ao sítio ativo (DSA) em comparação com valores de ΔG experimentais (a 300 K).

Código	ΔG Exp. (kcal/mol)	Escore DC (kcal/mol)	Desvio DC (%)	Escore DSA (kcal/mol)	Desvio DSA (%)
4	-7.36	-7.3	0.8	-7.4	-0.5
7	-6.07	-6.6	-8.7	-6.9	-13.7
5	-6.07	-6.9	-13.7	-7.3	-20.3
22	-5.83	-6.3	-8.1	-6.6	-13.2
13	-5.72	-6.0	-4.9	-6.0	-4.9
1	-5.69	-6.8	-19.5	-6.8	-19.5
46	-5.55	-6.4	-15.3	-6.4	-15.3
6	-5.54	-5.7	-2.9	-5.4	2.5
28	-5.44	-5.2	4.4	-5.2	4.4
9	-5.11	-5.6	-9.6	-5.4	-5.7
29	-5.02	-6.7	-33.5	-6.0	-19.5
32	-5.01	-6.8	-35.7	-7.0	-39.7
19	-4.95	-6.2	-25.3	-6.2	-25.3
36	-4.89	-5.8	-18.6	-5.6	-14.5
14	-4.86	-5.8	-19.3	-6.1	-25.5
35	-4.83	-5.8	-20.1	-5.8	-20.1
18	-4.74	-5.9	-24.5	-6.4	-35.0
48	-4.73	-6.1	-29.0	-6.1	-29.0
41	-4.68	-6.9	-47.4	-6.9	-47.4
27	-4.63	-6.9	-49.0	-6.9	-49.0
8	-4.61	-5.5	-19.3	-5.4	-17.1
47	-4.56	-6.6	-44.7	-6.6	-44.7
34	-4.49	-6.1	-35.9	-6.1	-35.9
37	-4.48	-7.5	-67.4	-7.5	-67.4
15	-4.21	-5.0	-18.8	-5.0	-18.8
Substrato	N.D.	-6.7	N.D.	-6.7	N.D.

*N.D = Não disponível

Na Figura 33 está mostrado o gráfico entre os escores de *docking* das duas metodologias aplicadas e as energias livres de ligação experimentais. Pode-se observar, por comparação com a função canônica $y=x$ (linha pontilhada), que as duas metodologias de *docking* aplicadas em geral superestimam (no sentido de ‘tornam mais negativas’) as energias de interação e que o *docking* mais refinado apesar de produzir um leve aumento na dispersão dos pontos, também tende a aproximá-los da função canônica.

De fato, esse comportamento pode ser comprovado por meio das equações de regressão linear que correlacionam o ΔG experimental com as energias de interação do *docking* cego (linha preta) e do *docking* de sítio ativo (linha vermelha), também mostradas no gráfico da Figura 33.

Para o *docking* cego, obteve-se para a curva de correlação uma inclinação de 0,3021, com erro quadrático médio (RMSE) de 0,62 kcal/mol ao passo em que para o *docking* com sítio

ativo, ainda que o RMSE aumente para 0,67 kcal/mol, indicando um aumento da dispersão dos dados, a inclinação da curva de regressão aumenta para 0,3640, tornando-se mais próxima do ideal que é de uma inclinação igual a 1,0 no qual as energias livres experimentais de ligação e as afinidades de *docking* seriam iguais.

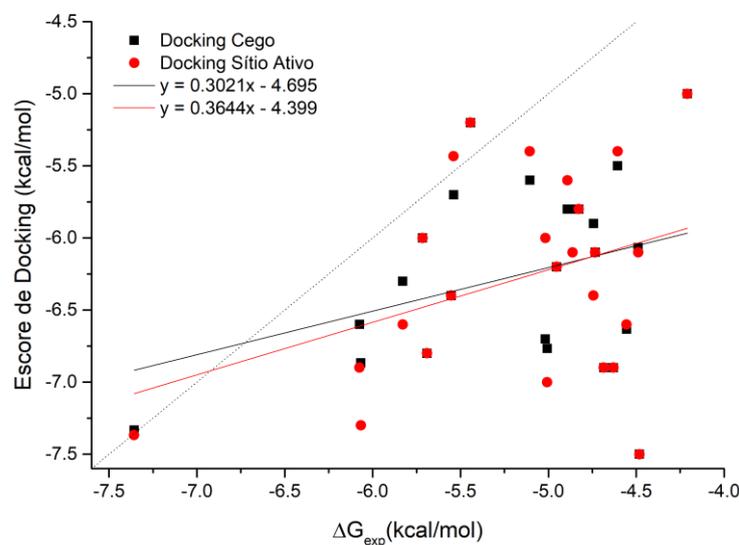


Figura 33 - Energias de interação do *docking* cego (●) e do *docking* de sítio ativo (■) em comparação com ΔG experimental a 300K e respectivas equações de regressão linear para *docking* cego (—) e *docking* de sítio ativo (—). Em pontilhado está a função canônica $y=x$.

Além disso, os parâmetros de qualidade da correlação linear são maiores para o *docking* mais refinado, com o coeficiente de determinação (R^2) passando de 0,1078 no *docking* cego para 0,1295 no *docking* de sítio ativo e o coeficiente de correlação de Pearson (R_P) passando de 0,3284 para 0,3599 para cada metodologia de *docking*, respectivamente.

Os valores extremamente baixos de R^2 , entretanto, evidenciam que o modelo linear obtido é incapaz de replicar satisfatoriamente a relação existente entre os resultados teóricos e os dados experimentais, ainda que essa relação possa ser considerada linear, como indicado pelo coeficiente de correlação de Pearson da ordem de 0,36, característico de uma correlação linear positiva fraca. Ou seja, esse modelo é capaz de expressar a tendência experimental, ainda que não permita a predição quantitativa das energias de ligação de forma adequada.

Dentre as razões que justificam essa dificuldade, pode-se citar a o fato de estrutura do receptor utilizada no *docking* ter sido gerada a partir da estrutura cristalográfica. É provável que as interações e a conformação adotada pela enzima num sólido cristalino sejam muito diferentes daquelas existentes nas condições biológicas ou nos ensaios *in vitro*.

Além disso, se pode citar o fato de que as metodologias de *docking* em geral não amostram os graus de liberdade do receptor, portanto a enzima é considerada como um corpo rígido à qual se complexa um ligante flexível.

Porém, sabe-se a aproximação e a interação do substrato induz alterações conformacionais no sítio ativo ou até mesmo em outras porções relativamente longínquas da enzima capazes de aumentar a interação receptor-ligante e alterar o modo de ligação.

Assim, o substrato (ou inibidores análogos a ele) não se ligam em um sítio ativo rígido, mas sim o moldam de modo a favorecer e amplificar as interações receptor-ligante, as quais então

promovem a redistribuição de cargas e a distorção conformacional do ligante levando à estabilização do estado de transição e consequente diminuição da energia de ativação.

Dessa forma, qualquer tentativa de quantificar as interações de um conjunto de ligantes através do *docking* em uma *única* estrutura do receptor jamais será plenamente válida, na medida em que cada ligante impõe ao receptor um mecanismo de interação próprio que não pode ser amostrado por *docking*.

Alternativamente, simulações por dinâmica molecular, que levam em consideração todos os graus de liberdade – tanto da enzima quanto dos ligantes – e que, portanto, consideram a flexibilidade do sistema às custas de maior demanda computacional podem ser utilizadas para se gerar um conjunto de estruturas diferentes da enzima permitindo a obtenção de resultados mais confiáveis por *docking*, em uma metodologia conhecida como *ensemble docking*.

Retornando à análise qualitativa dos resultados de *docking*, na Figura 34 estão mostradas as principais interações do ligante 4, o ligante experimentalmente mais potente, com o sítio ativo enzimático.

Pode-se observar a formação de pelo menos 4 ligações de hidrogênio entre o ligante e os resíduos His144, His175, Tyr217 e Arg102, essa última com características de ponte salina, além de uma interação importante de natureza hidrofóbica com a Phe255.

Essas cinco interações, comuns tanto ao ligante 4 e quanto ao produto (ligante 5), aparentam serem as mais importantes para a inibição enzimática, com destaque para a interação com a His175, que não ocorre com o substrato, mas que está presente nos inibidores graças à planaridade de seus anéis, que possibilita e maximiza essa interação.

Ainda na figura, pode-se observar que há uma grande proximidade entre o grupo 5-hidróxi do ligante e o oxigênio carboxílico da Phe255, o que permitiria a formação de uma quinta ligação de hidrogênio, não retratado na figura em razão dos raios e ângulos do corte aplicados pelo programa de visualização. Essa quinta ligação poderia produzir uma estabilização adicional a este complexo, justificando porque o ligante 4 tem uma potência aproximadamente 10 vezes superior ao ligante 5, ainda que compartilhem as mesmas interações principais.

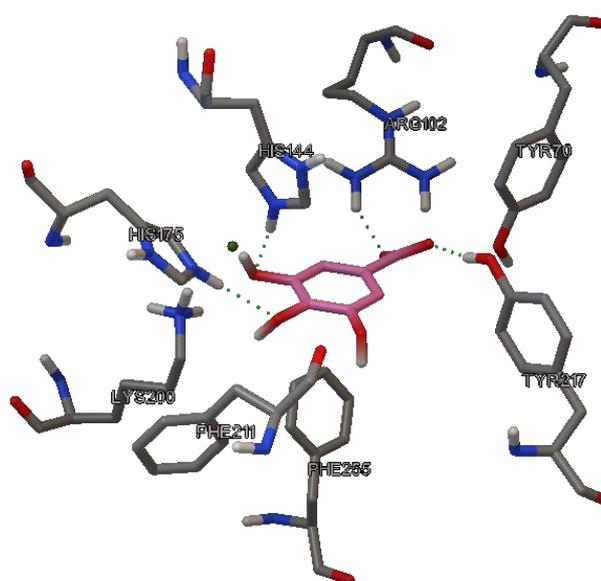


Figura 34- Padrão de interações do ligante 4 obtido pelo *docking* restrito ao sítio ativo. As ligações de hidrogênio estão marcadas pelas linhas pontilhadas em verde.

Comparando os três ligantes até aqui detalhados, observa-se que a Tyr217 (e a Tyr70 eventualmente) em conjunto com a Arg102 formam um sítio de ligação muito específico para o grupamento carboxilato dos ligantes, inclusive pela estabilização eletrostática obtida por ponte salina com a arginina.

Isso indica que este é um sítio de ligação importante e que deve ser considerado no desenho racional de inibidores mais potentes, que devem manter um grupo carboxilato nessa posição ou pelo menos um bioisómero de ácido carboxílico, como ácidos sulfônicos e fosfônicos, sulfonamidas ou mesmo tetrazóis, entre outros.

A formação de ligações de hidrogênio com as histidinas 144 e sobretudo com a 175 também demonstram ter um papel fundamental na inibição enzimática, como pode ser constatado na Tabela VI, que detalha as principais interações de todos os ligantes gerada pelo programa Binding Analyser (BINANA)¹⁰⁹ a partir dos arquivos *pdqbt* de saída do *docking* de sítio ativo.

Tabela VI – Principais interações entre os ligantes e a enzima obtidos pelo programa BINANA¹⁰⁹ após *docking* restrito ao sítio ativo

Ligante	IC50 ($\mu\text{mol.L}^{-1}$)	Ligação de Hidrogênio							Ponte Salina				Empilhamento			Contatos hidrofóbicos	
		Arg-102	His-144	His-175	Tyr-217	Gln-107	Glu-253	Gly-106	Arg-102	Mg-274	His-144	Glu-253	His-144	Phe-255	Phe-104		Tyr-70
SB	ND																15
04	4.36																17
07	37.5																18
05	38.0																17
22	56.7																23
13	68.5																19
01	71.5																16
46	90.0																22
06	92.1																28
28	108.0																11
09	190.7																25
29	221.3																15
32	225.0																24
19	246.4																15
36	273.0																19
14	286.7																26
35	304.4																17
18	350.1																19
48	355.8																16
41	386.5																24
27	423.7																21
08	440.8																22
47	480.1																29
34	536.4																16
37	542.8																21
15	857.7																19

*SB = Substrato *ND = não disponível

Pode-se observar que os principais inibidores (4, 5 e 7) compartilham dessas 4 ligações de hidrogênio que promovem uma ancoragem forte no sítio ativo, além de interações do tipo empilhamento-T com a His144.

Esse padrão básico de interações pode ser observado em outros inibidores um menos efetivos, como o 1, 6, 9 e 32, porém não se pode estabelecer claramente um comportamento geral de justificasse as diferentes potências observadas experimentalmente, possivelmente porque o *docking* na estrutura cristalográfica acaba por superestimar não apenas as energias, mas também o número de interações dos ligantes menos ativos.

Dois ligantes merecem destaque, no entanto, em razão dos elevados desvios que seus resultados de *docking* apresentaram frente aos dados experimentais, que são o ligante 27, com desvio de 49% e o sobretudo o ligante 37, com um desvio de quase 70%, nos quais houve uma superestimação das energias de ligação do *docking*, indicando para esses compostos um potencial inibitório muito superior ao apontado nos ensaios experimentais.

Na Figura 35A estão mostradas o perfil de ligações de hidrogênio apontadas pelo *docking* para o ligante 27, em que se observa a formação de 4 ligações de hidrogênio, três envolvendo a porção carboxílica do composto, sendo duas com a Arg102, uma com a Tyr217, além de uma ligação entre o grupo α -hidroxila e a His144, quatro interações que também são observadas para o substrato, produto e o ácido gálico.

Isso indica que a essa superestimação da energia de ligação não está relacionada com a capacidade de formação de ligações de hidrogênio, mas sim com um maior número de interações de natureza hidrofóbica – 21 interações contra cerca de 17 dos ligantes experimentalmente mais potentes, segundo análise com o Binana – cujas energias estão sendo superestimadas pela metodologia de *docking* ao mesmo tempo em que pode haver uma menor penalização estérea do ligante 27 comparado aos demais em virtude de suas dimensões reduzidas.

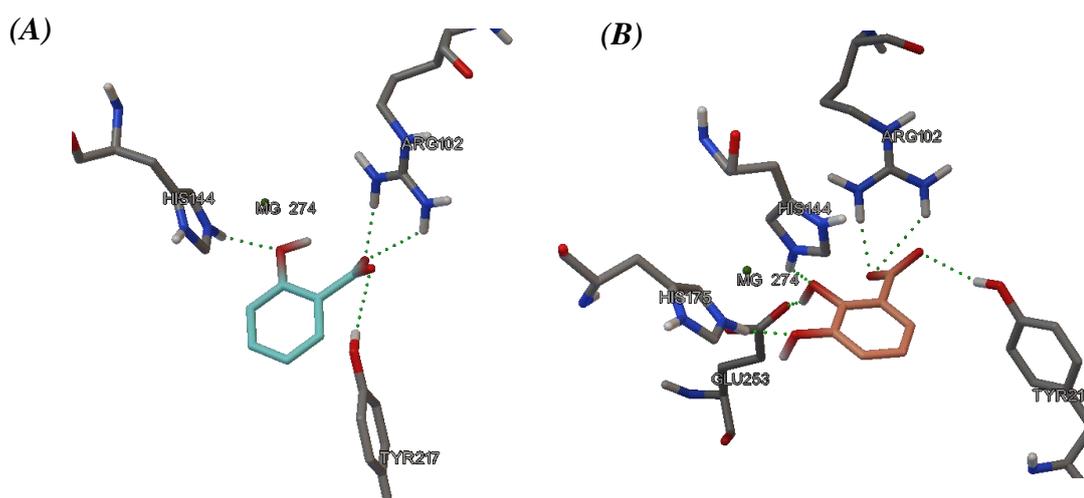


Figura 35- Padrão de ligações de hidrogênio apontado pelo *docking* para os ligantes 27 (Figura A) e 37 (Figura B). As ligações estão pontilhadas em verde.

Na Figura 35B pode-se ver as ligações de hidrogênio apontadas por *docking* para o ligante 37, para o qual se obteve o valor mais negativo de energia de ligação ($-7.5 \text{ kcal.mol}^{-1}$) e, conseqüentemente, o maior desvio em relação aos resultados experimentais, visto que este é um dos ligantes menos ativos *in vitro*.

Percebe-se pela figura que, assim como para substrato, produto e ácido gálico, o ligante 37 forma ligações de hidrogênio com a Arg102, Tyr217, His144 e His175, porém também aparece formando uma ligação entre o grupo α -hidroxila e a Glu253, que é parte do sítio de complexação do cofator magnésio.

Isso mostra que o desvio na energia de ligação no *docking* desse composto pode estar relacionada à superestimação da energia das interações hidrofóbicas (novamente há 21 contatos hidrofóbicos) aliada à possível formação de ligações de hidrogênio espúrias, como essa que ocorre com Glu253, um resíduo que provavelmente não está disponível para esse tipo de interação em função de sua atuação na formação de um complexo com um íon metálico altamente carregado eletricamente.

É importante salientar que programa Vina desconsidera cargas na geração dos escores de *docking*, portanto, ainda que o BINANA aponte a possível formação de ponte salina/interação coulômbica entre os ligantes 27 e 37 e o cofator Mg^{2+} , essa interação de natureza eletrostática não está sendo contabilizada na energia de ligação do *docking*, não sendo responsável pelo desvio observado.

Outro aspecto que deve ser salientado é torção apresentada pelos grupos carboxílicos desses dois ligantes em relação ao plano do anel aromático. Sabe-se que, em função da aromaticidade e deslocalização eletrônica há, na verdade, um favorecimento energético na manutenção de planaridade entre o COO^- e o anel, o que não é levado em consideração pela função de escore do *docking*.

A manutenção da planaridade dificultaria a formação de algumas ligações de hidrogênio com as que ocorrem com a Arg102 e a Tyr217, além de permitir que haja a formação de ligações de hidrogênio intramoleculares entre a α -hidroxila e o grupo carboxilato, interação que, inclusive, seria favorecida pelo grupo ácido se encontrar desprotonado nas condições experimentais, comprometendo a participação dessa hidroxila na formação de ligações de hidrogênio com o sítio ativo.

Esses dois fenômenos que certamente impactam o potencial inibitório *in vitro* dos ligantes e que não são considerados apropriadamente pelo *docking* podem justificar os desvios extremamente elevados entre os resultados teóricos e experimentais.

A comprovação, ao menos em parte, das hipóteses aqui levantadas e a racionalização das divergências observadas entre os dados experimentais e os resultados de *docking* foram as motivações para o estudo por dinâmica molecular desses dois ligantes

Além desses dois ligantes *outliers*, chama a atenção os resultados obtidos para os ligantes 13 e 46, que são dois inibidores muito ativos experimentalmente, mas que apresentam pouca ou até mesmo nenhuma interação importante com a enzima nos resultados de *docking*. As únicas interações possíveis apontadas seriam contatos hidrofóbicos e interações do tipo Van der Waals.

A análise mais detalhada desses dois ligantes mostra que ambos se ligam preferencialmente na mesma região, que corresponde à entrada do sítio ativo (Figura 36) e esse resultado se observa tanto no *docking* cego quanto no *docking* restrito ao sítio ativo, não sendo, portanto, um artifício do *docking* mais restrito.

Apesar de poderem se ligar em qualquer ponto da superfície hidrofílica da enzima no *docking* cego, há aparentemente uma preferência pela ligação nessa posição que corresponderia ao canal de entrada do sítio ativo.

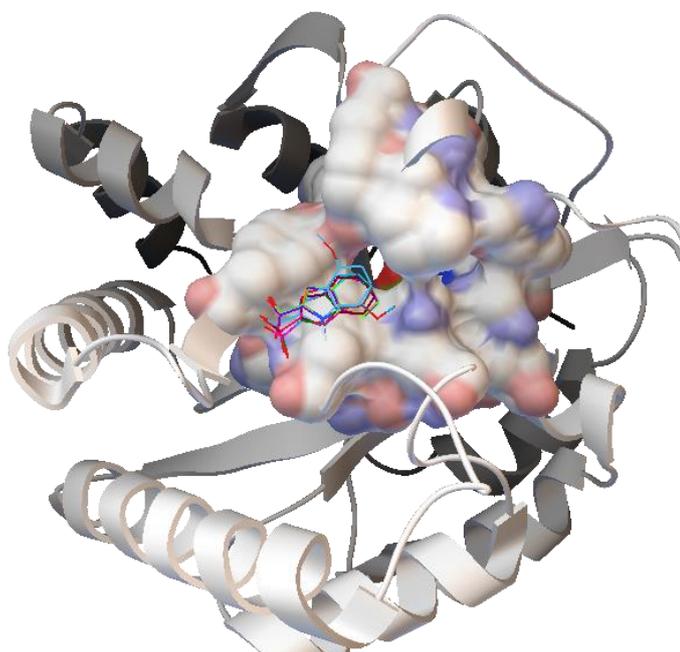


Figura 36– Sobreposição dos modos de ligação mais favoráveis obtidos tanto no *docking* restrito quanto no *docking* cego para os ligantes 13 e 46. Os resíduos do sítio ativo estão mostrados como superfície e coloridos segundo a polaridade das cadeias laterais.

Na Figura 37 estão mostrados os modos de interação de melhor escore para esses dois ligantes obtidos após *docking* envolvendo toda a enzima. Pode-se observar que as principais interações são de natureza hidrofóbica e envolvem resíduos como a Pro213, Tyr217, Gly106 e sobretudo Phe104, sendo que ambos poderiam interagir também por empilhamento- π com esse resíduo.

Além disso, interações de natureza eletrostática como ponte salina também poderiam ocorrer com a glutamina-107, com o ligante 13 inclusive formando uma ligação de hidrogênio com o esqueleto peptídico desse resíduo.

Ao contrário dos demais ligantes ativos, esses dois não são derivados ou análogos diretos do ácido benzoico, mas sim um benzofuranocarboxilato (13) e um indolecarboxilato (46), tendo em comum o grupo carboxilato na posição 2, ao lado do heteroátomo no anel de cinco membros, o que levanta duas hipóteses: a primeira seria de que, em razão do maior volume gerado pelo sistema bi-anelar esses ligantes não caberiam adequadamente no sítio ativo e *docking* não conseguiria amostrar esse modo de ligação pelo impedimento estérico elevado e consequente penalidade no escore.

A segunda hipótese, por sua vez, estaria relacionada à possibilidade do *docking* realmente estar amostrando um modo de ligação válido, na entrada do sítio ativo, que permitiria uma ação alostérica e cujo resultado seria o bloqueio da entrada do sítio, com consequente impedimento da entrada do substrato ou da saída do produto após a reação, levando à inibição enzimática experimentalmente observada para essas moléculas.

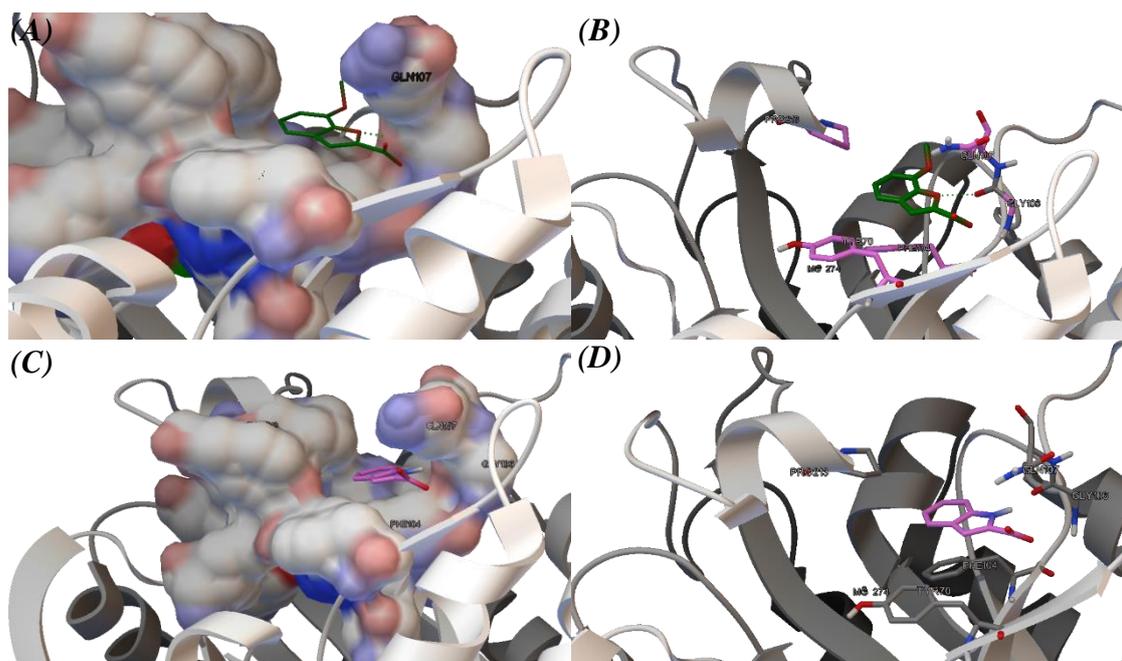


Figura 37 – Modo de interação principal do ligante 13 (A e B) e do ligante 46 (C e D) após *docking* sem restrição. A entrada e os resíduos principais do sítio ativo estão mostradas como superfícies em A e C.

Uma das formas de se avaliar essa hipótese seria a dinâmica molecular desses ligantes partindo desse modo de ligação e avaliando-se se os mesmos permanecem ligados durante a simulação ou se ocorre a migração para outras regiões da enzima ou mesmo para o meio aquoso. O teste dessa hipótese foi a motivação para o estudo dos ligantes 13 e 46 em complexo com a AsbF por meio de dinâmica molecular.

5.2 DINÂMICA MOLECULAR

A partir dos resultados de *docking*, optou-se por simular por dinâmica molecular a enzima livre em solução e seus complexos com substrato, produto e com os ligantes 4, 13, 27, 37 e 46.

O ligante 4 (ácido gálico) foi simulado por ser experimentalmente o mais potente e a compreensão os padrões de interação e fatores dinâmicos que determinam essa elevada potência é fundamental para o desenho de inibidores melhores e mais seletivos.

Já os ligantes 27 (ácido salicílico) e 37 (ácido 2,3-dihidróxibenzoico) foram simulados por apresentarem os maiores desvios entre os valores teóricos de afinidade e os resultados experimentais, sobretudo em função da superestimação de seus escores de *docking*. Através da dinâmica, espera-se compreender porque tais compostos se ligam de forma tão fraca experimentalmente e entender as razões que produzem os valores tão negativos de energia observados no *docking*.

Os outros dois ligantes, 13 (ácido 7-metóxi-2-benzofuranocarboxílico) e 46 (ácido indol-2-carboxílico), foram estudados em razão de apresentarem um modo de ligação diferente dos demais inibidores nos resultados do *docking*, na entrada do sítio ativo. Espera-se que dinâmica molecular seja capaz de elucidar se tal modo de ligação é factível, havendo eventualmente um

sítio de ligação alostérico ou se esse resultado se trata apenas um artifício produzido por *docking* em razão de impedimentos estéreos elevados no sítio de ligação principal.

Além disso, todas essas simulações serão usadas para a obtenção de estruturas distintas do receptor, que possam abranger diversos pontos no espaço conformacional da enzima e que assim possam gerar uma amostragem mais completa das diversas formas que o receptor pode assumir em solução, permitindo um refino do *docking* (*ensemble docking*) e eventualmente a obtenção de um modelo que correlacione melhor as energias de ligação teóricas e experimentais, de tal forma que se possa prever não apenas o modo mas também a intensidade de ligação de novos compostos.

5.2.1 Enzima livre em solução

Com o intuito de se explorar ao máximo a superfície de energia livre conformacional da enzima em solução, foram realizadas 4 simulações do receptor em solução salina, de 120 ns cada e com diferentes “sementes” de geração de velocidades.

O comportamento dinâmico básico dessas quatro simulações pode ser constatado na Figura 38, em que está mostrado o deslocamento quadrático médio (RMSD, do inglês *Root Mean Square Deviation*) dos carbonos-alfa em relação à estrutura inicial, juntamente com seus respectivos histogramas.

Pode-se observar que, de uma forma geral, os quatro sistemas simulados evoluem ao longo do tempo para um patamar, oscilando em torno de um RMSD médio de 0,1323 nm, o que indica que os mesmos atingiram conformações de equilíbrio para essa temperatura, com uma pequena flutuação das posições atômicas, que é característica de sistemas dessa natureza em solução aquosa. Apenas a simulação 2 apresentou alguma divergência em relação a esse comportamento, com oscilações maiores que variam de 0,08 a 0,22 nm, porém ainda centradas em torno de 0,13 nm aproximadamente.

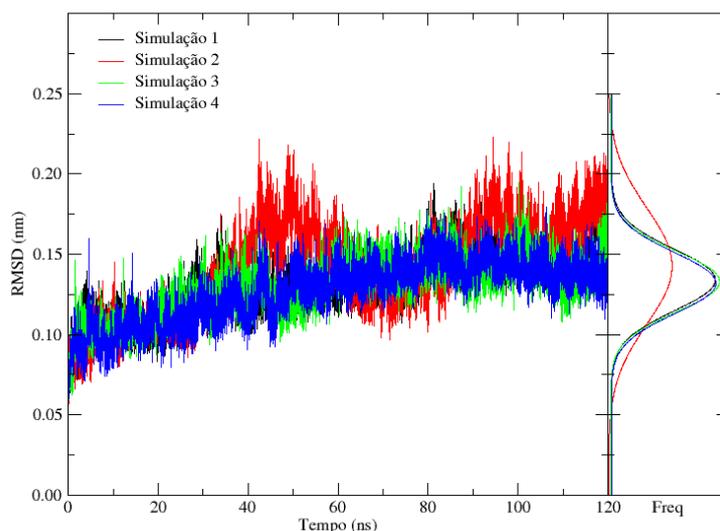


Figura 38 - Deslocamento quadrático médio das posições dos carbonos-alfa (em nanômetros) em relação à estrutura inicial, em função do tempo de simulação, para as quatro simulações da enzima livre em solução. Ao lado, histograma desses deslocamentos.

Portanto, ainda que formalmente se tratem de simulações distintas pelo uso de diferentes sementes de geração de velocidade, aparentemente os quatro sistemas evoluem para um comportamento de equilíbrio muito similar, com pouca mobilidade dos átomos, o que pode estar relacionado com a estrutura em forma de barril da enzima, formada por um conjunto de folhas- β centrais cercado por hélices- α e com um número relativamente pequeno de alças e outros motivos mais flexíveis, gerando uma estrutura significativamente rígida e com reduzida mobilidade intrínseca dos resíduos.

Esse comportamento é reforçado por meio da análise de mobilidade estrutural realizada com o programa MDLovoFit¹¹⁰, que consistiu na separação de cada trajetória em 120 frames, a cada nanossegundo de simulação, seguida do alinhamento e sobreposição de diferentes subconjuntos da proteína correspondendo a diferentes frações do número total de carbonos-alfa. O algoritmo do programa detecta automaticamente o subconjunto de átomos com o menor RMSD para cada uma dessas frações, permitindo identificar, quantificar e separar as porções mais móveis daquelas com menor flexibilidade.

A primeira das análises realizada com esse programa está mostrada na Figura 39, em que se pode observar o RMSD acumulado dos carbonos-alfa em função da fração de átomos incluída na análise.

Observa-se que, para os 4 sistemas simulados, é possível alinhar em torno de 80% dos carbonos- α mantendo um desvio quadrático médio inferior a 1,0 Å em relação a estrutura de partida, um valor significativamente elevado e que evidencia a grade rigidez estrutural dessa enzima. Isso também mostra que apenas 20% da estrutura é responsável pela maior parte do RMSD observado na Figura 38.

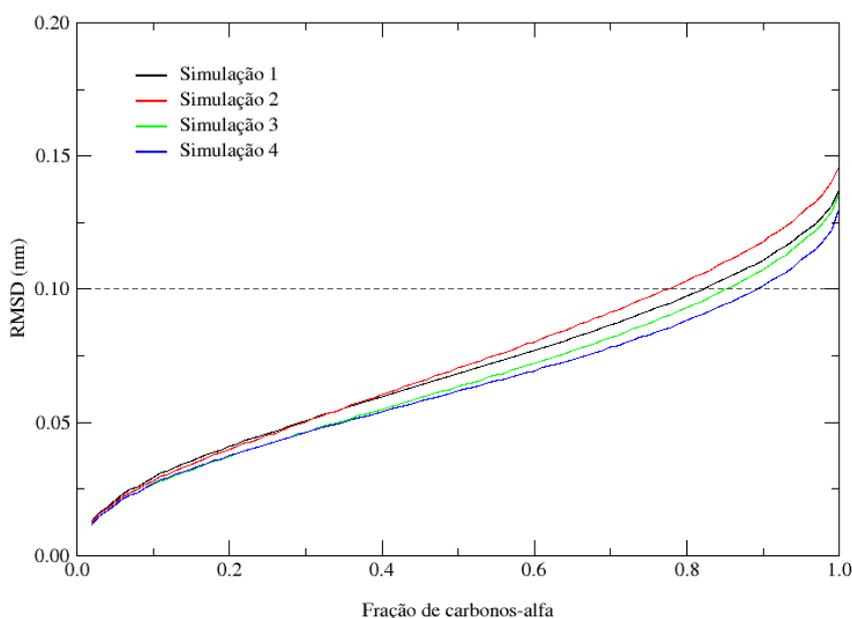


Figura 39 - RMSD acumulado dos carbonos-alfa, em comparação com a estrutura inicial, em função da fração do número de carbonos- α considerado na análise.

Naturalmente, como já poderia ser previsto pelo gráfico de evolução temporal do RMSD, a estrutura com maior flexibilidade foi gerada durante a simulação 2, na qual foi possível alinhar apenas 78% dos carbonos- α a menos de 1,0 Å de RMSD, enquanto que a mais rígida foi

obtida pela simulação 4, em que foi possível alinhar 90% dos átomos dentro desse limite. Para as simulações 1 e 3 obteve-se os valores intermediários de 83% e 86 % de alinhamento, respectivamente.

Utilizando-se esses percentuais como valores de corte, foi possível separar os RMSD em dois grupos e acompanhar a evolução temporal das frações mais móveis e das mais rígidas separadamente, como é mostrado na Figura 40.

Pode-se observar pela figura que durante os 5 ns iniciais há um incremento abrupto de mobilidade das duas frações, na medida em que o sistema relaxa as interações desfavoráveis existentes na estrutura cristalográfica, seguida de uma rápida estabilização no deslocamento da maior parte dos resíduos, que leva a formação de um patamar em torno de 1,0 Å, ao redor do qual há pouca (no caso da simulação 2) ou praticamente nenhuma oscilação de RMSD (simulação 3).

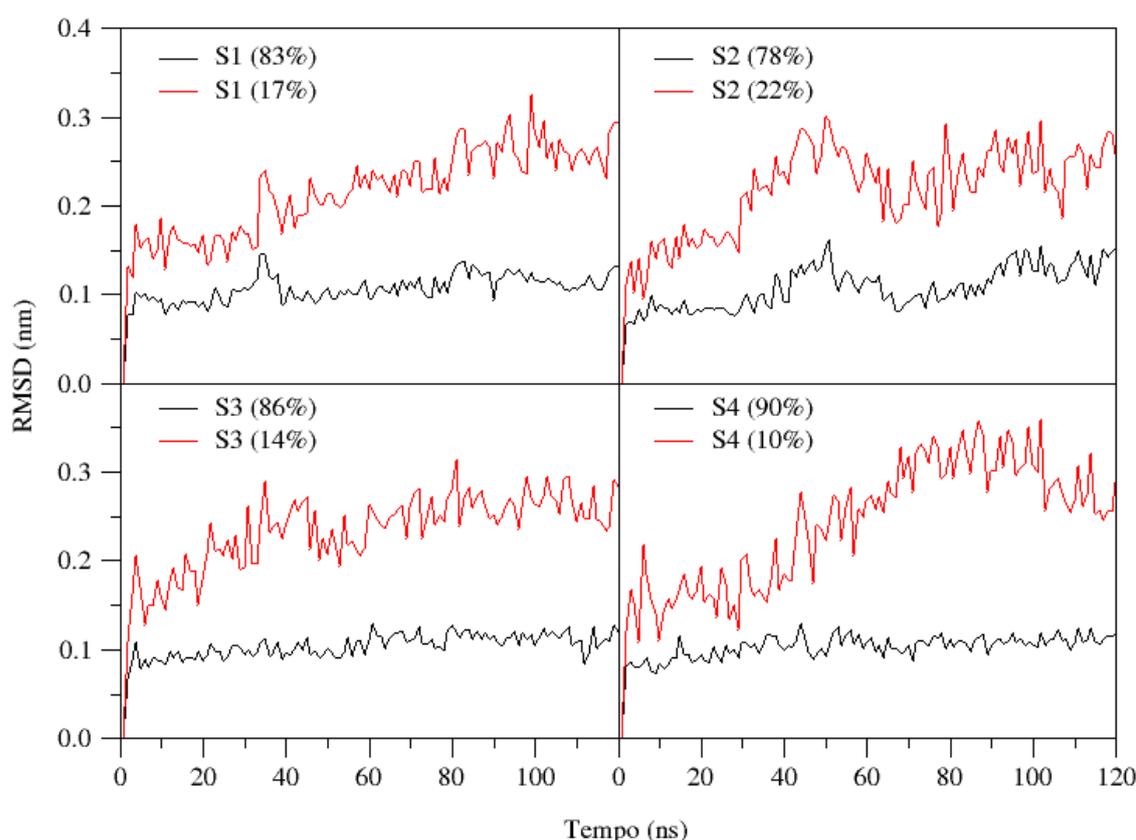


Figura 40 – RMSD das frações mais móveis (vermelho) e das mais rígidas (preto) em relação a estrutura inicial para as quatro simulações da enzima livre em solução (S1 a S4).

Por outro lado, a fração mais móvel da estrutura continua a aumentar seu deslocamento médio com o andamento da simulação, até atingir um RMSD relativamente constante de 2,5 Å. Uma vez atingido esse patamar não ocorrerem mais alterações significativas no RMSD dessas frações, apenas a oscilação em torno desse valor até o fim da simulação. O alcance desse patamar pode ocorrer de maneira mais imediata, como para os sistemas 2 e 3, ou mais tardiamente, como para os sistemas 1 e 4, porém acaba presente ao final de todas as dinâmicas, indicando que essa é uma evolução natural da enzima, não um artifício das simulações.

Ainda por meio do MDLovoFit foi possível gerar imagens sobrepostas das estruturas em diferentes momentos das simulações (120 frames, em passos de 1 ns) e colorir-las em função das maiores e menores mobilidades em relação à estrutura inicial (Figura 41).

A predominância da coloração azul nas imagens evidencia que toda a porção interna da enzima, composta principalmente por folhas- β , apresenta mobilidades inferiores à 1 Å. O mesmo acontece com grande parte da sua superfície, constituída majoritariamente de hélices- α cuja rigidez pode ser facilmente associada à existência de muitas ligações de hidrogênio entre os *loops* que as constituem. Uma exceção, no entanto, é a α -hélice 2, que apresentou uma elevada mobilidade nas quatro simulações realizadas.

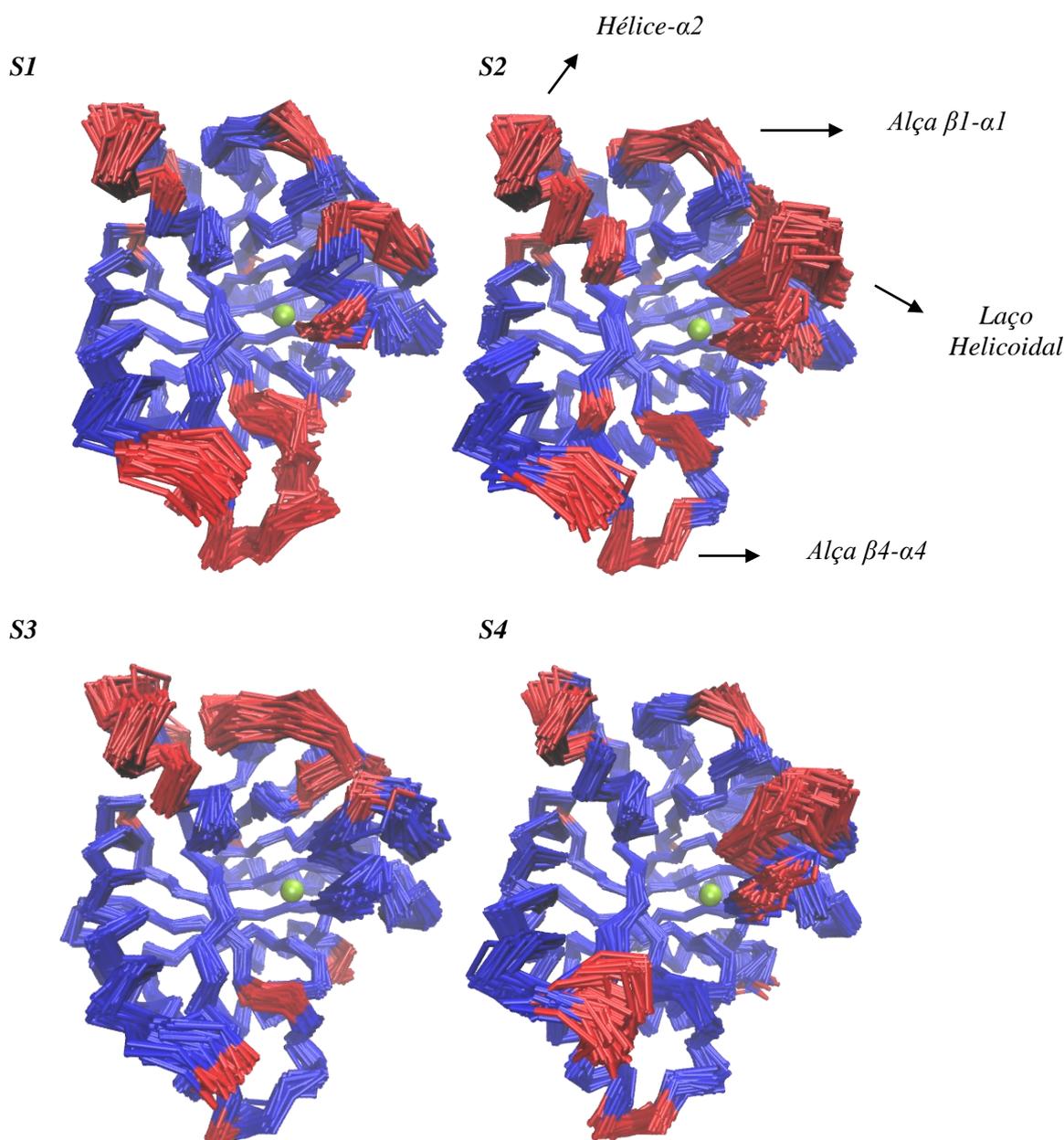


Figura 41 - Sobreposição de 120 frames extraídos a cada nanossegundo das quatro simulações da enzima livre em solução (S1-S4). As cores distinguem os resíduos mais móveis (vermelho) dos de menor mobilidade (azul). Em verde está destacado o cofator Mg.

As demais partes móveis se encontram totalmente na superfície e são compostas principalmente por alças, como a alça que conecta a folha- β 4 à α -hélice 4 e a alça que conecta a folha- β 1 com a α -hélice 1.

O destaque, entretanto, fica com uma região conhecida como *laço helicoidal*, que forma parte da parede externa e da entrada do sítio ativo, atuando como uma tampa sobre ele. Em todas as simulações essa região apresentou mobilidade significativamente elevada, em alguns casos a mais elevada de toda a proteína, como nas simulações 2 e 4.

Compreendendo do resíduo 206 ao 224, essa porção da enzima engloba diversos resíduos cataliticamente importante como Phe211, Pro213 e sobretudo a Tyr217, apontada pelo *docking* como um resíduo com provável ação direta no mecanismo de catálise ácido-base da enzima.

Os grandes valores de RMSD observados para essa porção ao longo das simulações reforçam a possibilidade da cavidade do sítio ativo tal como obtida da estrutura cristalográfica não se encontrar em seu estado mais relaxado e natural – que eventualmente corresponderia ao estado biologicamente ativo – possivelmente pela presença do produto 3,4-DHB coordenado ao sítio na estrutura experimental.

Sabendo que esse composto atua como um inibidor relativamente potente, é provável que sua complexação e interação intensa com o sítio ativo o force a uma conformação mais justa, reduzindo seu volume interno. Ao utilizar-se essa estrutura como ponto de partida para o *docking*, os compostos com características estruturais mais próxima do 3,4-DHB foram docados adequadamente, enquanto outros compostos maiores, como os ligantes 46 e 13 possivelmente encontraram um elevado impedimento estérico, o que contribuiu para a elevada discrepância entre as potências experimentais e as energias de interação do *docking*.

A flexibilização do laço helicoidal aparenta ser uma evolução natural da enzima em solução, fato que é reforçado pela análise das flutuações quadráticas médias (RMSF) de todos os resíduos ao longo das simulações, que está mostrada na Figura 42 e que também foi realizada com o programa MDLovoFit.

Pode-se observar que em todas as simulações a porção que compreende os resíduos 206 a 224 apresenta flutuações da ordem de 3 a 4 Å, sendo a porção contígua mais longa a apresentar mobilidade elevada. Para as simulações 1 e 2, inclusive, os resíduos com as maiores flutuações são a Gly221 e Ala219, respectivamente, ambos resíduos que compõem esse laço.

Observa-se também que outras porções de RMSF elevado são comuns a todas as simulações e já foram descritas na análise de RMSD, como as alças β 1- α 1 e β 4- α 4 e a α -hélice 4, indicando que estas flutuações fazem parte de um estado natural da enzima em solução, o qual eventualmente impacta os modos e intensidades de ligação dos inibidores.

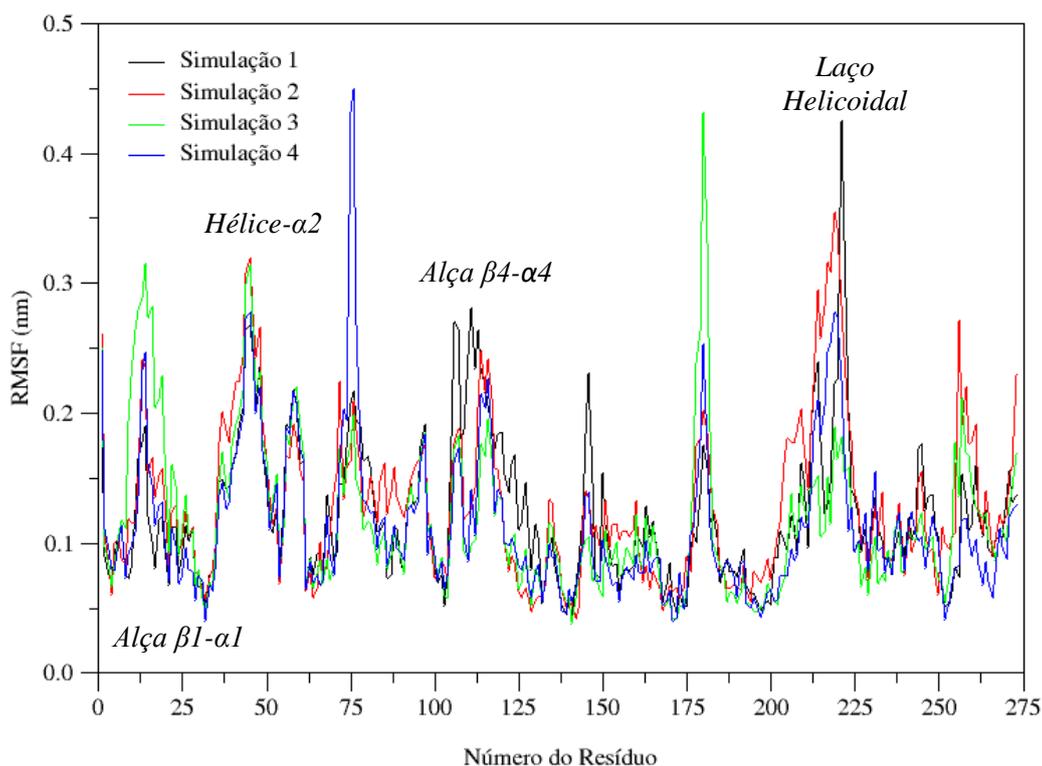


Figura 42 - Flutuação quadrática média (RMSF) em nanômetros, em relação a estrutura inicial para todos os resíduos das quatro simulações da enzima livre em solução. As porções mais móveis estão identificadas.

Para a simulação 3 se observa um pico de RMSF para a Gly180, enquanto que para a simulação 4 esse pico é visto para a Leu75 e Ser76, todos resíduos que participam de alças que conectam as folhas- β internas às α -hélices superficiais, tendo, portanto, uma flexibilidade intrínseca e não sendo surpreendente que apresentem RMSF elevado.

Na Figura 43 está mostrada a sobreposição da estrutura inicial e da estrutura obtida após 120 ns de dinâmica molecular para a simulação 2, com os resíduos coloridos de acordo com seus valores de RMSF. Essa simulação foi escolhida para geração da imagem ter apresentado as maiores alterações estruturais, porém as modificações observadas são as mesmas que ocorrem nas demais simulações, ainda que com menor intensidade, por isso apenas essa estrutura é mostrada.

Observa-se que, à exceção das folhas- β internas, que permanecem praticamente imóveis, todos os demais resíduos se alteram, inclusive as α -hélices, e em geral esse movimento pode ser descrito como uma expansão ou afastamento dos resíduos mais externos para longe do centro enzimático. Inclusive, pode-se observar que as porções que compõem as paredes do sítio ativo – e em especial o laço helicoidal – aparentam estar mais longes do cofator magnésio, mostrando que ocorre uma aparente abertura da cavidade do sítio ativo.

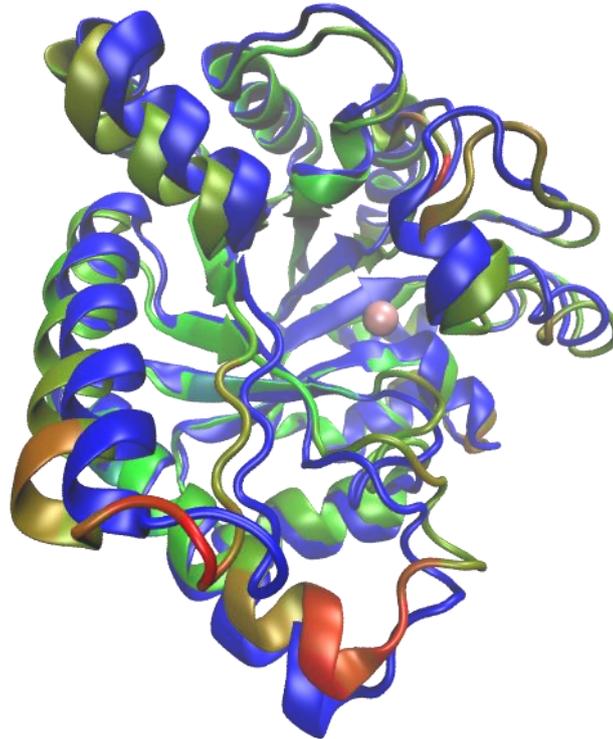


Figura 43 - Sobreposição da estrutura inicial (azul) e da estrutura obtida após 120 ns para a simulação 2 (colorido). As cores indicam o grau de flutuação dos resíduos, sendo maior quanto mais avermelhado. O cofator magnésio está destacado para indicar a localização do sítio ativo.

Como os resíduos do sítio ativo fazem parte de diferentes domínios da enzima, localmente próximos, porém sequencialmente distantes, é interessante analisar suas flutuações em separado dos demais resíduos, o que está mostrado na Figura 44.

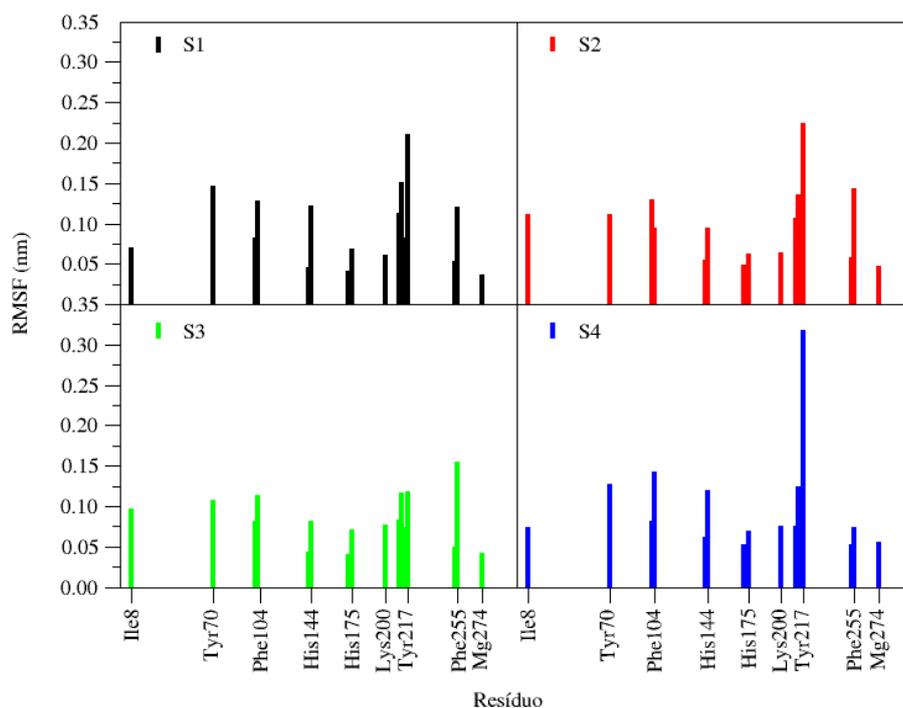


Figura 44 – Flutuação Quadrática Média (RMSF) em relação à estrutura inicial para os resíduos que compõem o sítio ativo para cada simulação da enzima livre em solução.

Observa-se que, em geral, os resíduos do sítio flutuam pouco, com RMSF da ordem de 1 Å ou menos. Uma exceção interessante, no entanto, é a Tyr217, que apresenta elevados valores de RMSF para três das quatro simulações, atingindo valores que chegam a 3 Å, como é o caso da simulação 4. Outros resíduos com RMSF elevado são a Phe104 e a Phe255, cuja função é a formação uma espécie de gaiola hidrofóbica para complexação da porção aromática do substrato/ligantes, bem como a Tyr70, que estaria envolvida na coordenação do grupamento carboxilato do ligante.

A flutuação elevada desses quatro resíduos reforça a ideia de que o 3,4-DHB atua restringido o sítio ativo na estrutura cristalográfica, a qual pode evoluir livre em solução para um estado mais ou até mesmo menos ativo biologicamente, porém em uma conformação diferente do estado inicial, o qual aparenta ser uma evolução natural da enzima em solução, visto que os padrões de RMSD e RMSF são muito similares nas quatro simulações realizadas.

5.2.2 Enzima em complexo com substrato

Foram realizadas 3 simulações da enzima em complexo com o substrato, a partir do melhor modo de ligação apontado por *docking*. Como utilizou-se a mesma semente de geração de velocidade e as mesmas configurações iniciais de geometria e topologia, essas três simulações constituem formalmente triplicatas do mesmo sistema, de modo que se pode analisar os resultados das dinâmicas através das médias dessas triplicatas.

Na Figura 45 está mostrado o comportamento temporal do RMSD dos carbonos- α para as três réplicas e para a média. Pode-se observar que, apesar das simulações se tratarem do mesmo sistema utilizando-se as mesmas condições iniciais, os sistemas apresentam evoluções temporais significativamente diferentes.

É sabido, porém, que simulações de dinâmica molecular são plenamente reprodutíveis apenas se executadas a partir do mesmo arquivo binário e no mesmo computador ou em outro computador com processador e sistema operacional idênticos – o que não é factível quando se utiliza computação paralela, como no caso do presente estudo – e mesmo assim, as simulações não seriam exatamente iguais em função da acumulações e propagação de erros de arredondamento e de peculiaridades dos termostatos utilizados.¹⁰⁶

Por outro lado, o teorema do limite central afirma que para simulações suficientemente longas (ou para um número suficientemente grande de simulações) todos os observáveis físicos convergem para seus valores de equilíbrio, na medida em que a soma de um número suficientemente grande de interações de variáveis aleatórias independentes tende a adotar uma distribuição normal centrada no valor de equilíbrio.¹¹¹

Além disso, o teorema da ergodicidade afirma que todos os microestados acessíveis de um sistema são igualmente prováveis em uma amostragem suficientemente longa, de tal forma que a média temporal e a média de *ensemble* são iguais após tempos suficientes longos.¹¹¹

Isso significa que, embora as simulações apresentem evoluções temporais levemente distintas, a média temporal de suas propriedades se iguala à média de *ensemble* e esse valor converge para o valor de equilíbrio à medida em que a amostragem (o número de réplicas) aumenta. Por essa razão, optou-se pela realização de pelo menos três reproduções longas dos principais sistemas da enzima em complexo com ligantes.

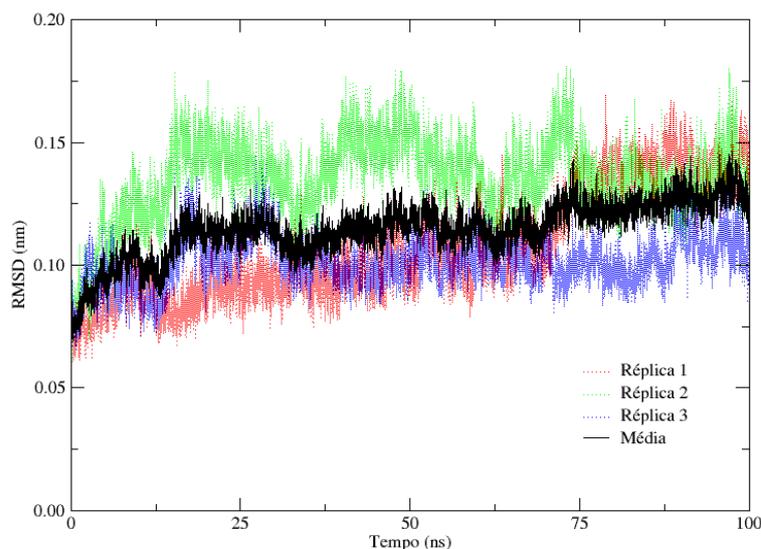


Figura 45 – Evolução temporal do desvio quadrático médio (RMSD) das posições dos carbonos- α em relação à estrutura inicial para as três réplicas do sistema enzima-substrato e média das reproduções.

Observa-se na figura que o RMSD apresenta uma flutuação maior para a réplica 2, variando entre 0,10 e 0,18 nm enquanto se mantém mais estável para a réplica 3, com variações mais leves e suaves ao longo do tempo. Em média, o RMSD do sistema flutua em torno de 0,1145 nm, portanto 0,18 Å abaixo do que foi observado para a enzima livre, indicando que a complexação do substrato impacta a mobilidade estrutural da enzima.

Esse impacto pode ser constatado também pela análise fracionada de mobilidade gerada com o MDLovoFit (Figura 46), na qual se observa que mais de 90% dos carbonos- α da estrutura podem ser sobrepostos com RMSD inferior a 1,0 Å. De fato, em média, essa sobreposição chega a 96% enquanto que na enzima livre esse valor não passava de 85%, mostrando como a estrutura tornou-se mais rígida pela interação com o substrato, porém eventualmente atingindo/mantendo uma conformação mais adequada para a interação com o mesmo, o que não pode ser descartado.

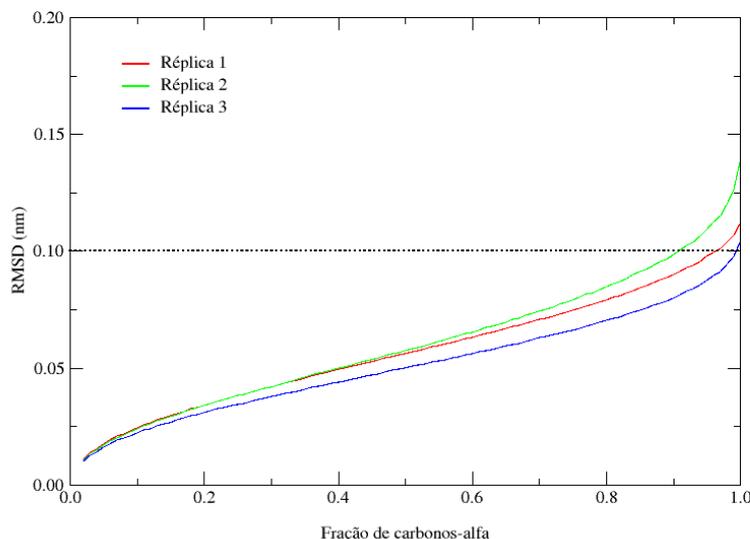


Figura 46 - RMSD dos carbonos- α em função da fração de átomos consideradas no ajuste da estrutura para as três réplicas da simulação enzima-substrato.

A análise visual do RMSD separada em função das frações com maiores e menores mobilidades (Figura 47) mostra que os grupos de maior variação não são os mesmo nas três réplicas, porém eles compartilham a mesma função estrutural – são todos alças que ligam as folhas- β às α -hélices externas, com destaque novamente para a alça $\beta 1-\alpha 1$ e para o laço helicoidal, que participam da formação do sítio ativo e para a alça $\beta 3-\alpha 3$, que apresenta alta mobilidade em pelo menos duas das três réplicas.

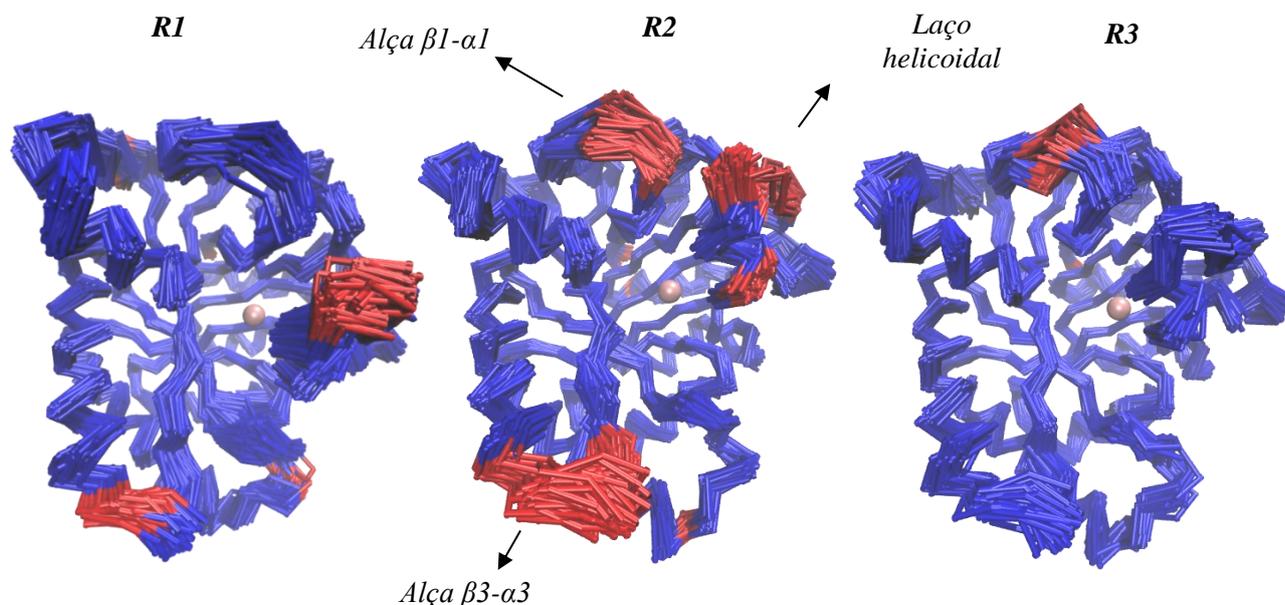


Figura 47 - Sobreposição de 100 frames extraídos a cada nanossegundo para as três reproduções da enzima em complexo com o substrato (R1-R3) coloridos de acordo com menores (azul) e maiores de RMSD. Em destaque o cofator Mg.

Observa-se ainda que, à exceção do resíduo Gly211 na réplica 2, todos os resíduos apresentam flutuação quadrática média inferior ao observado para a enzima livre (Figura 48), com o RMSF médio passando de 0,12 nm para os sistemas sem ligante para 0,10 nm para a enzima em complexo com o substrato.

A presença do ligante também trouxe algumas alterações do padrão de RMSF, que passa a ter uma contribuição maior da alça $\beta 3-\alpha 3$ e uma contribuição menor do laço helicoidal, assim como não há mais flutuações significativas da hélice- $\alpha 2$, como observadas no sistema livre, eventualmente em função da participação de algum resíduo desse grupo na interação com o substrato, que acaba propagando essa rigidez ao longo das hélices, como uma alteração estrutural de natureza alostérica.

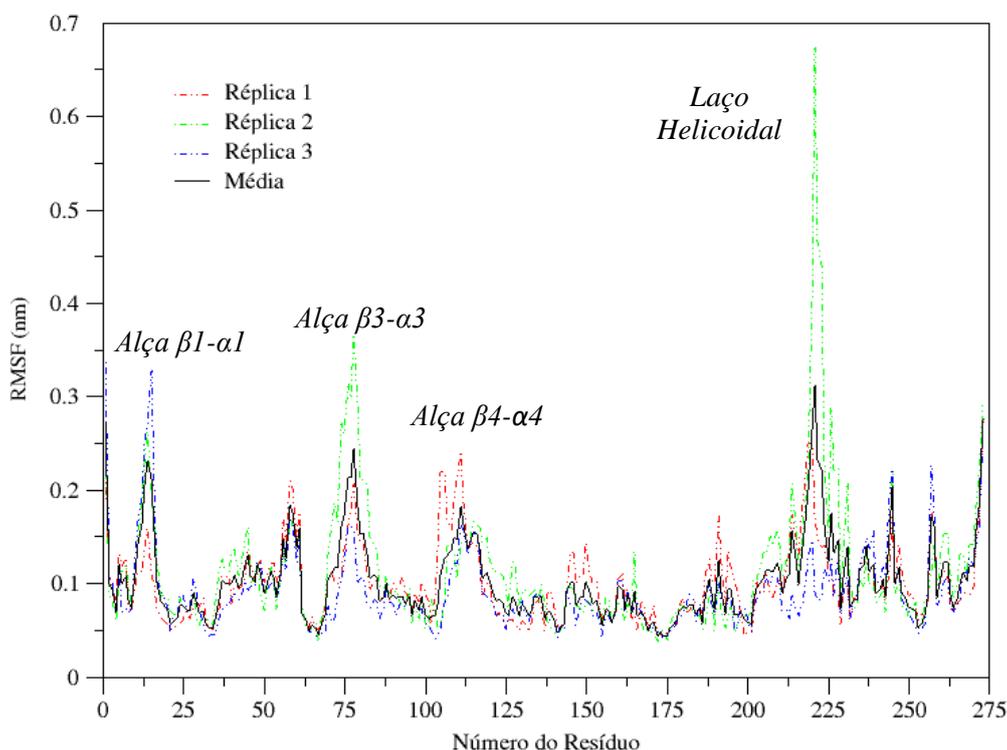


Figura 48 - Flutuação quadrática média (RMSF) de todos os resíduos em relação à estrutura inicial para as réplicas do sistema enzima-substrato e média das triplicatas.

Em relação especificamente ao sítio ativo, pode-se observar que houve também uma significativa diminuição do RMSF dos resíduos que o compõem quando comparado com a enzima livre (Figura 49), com o valor médio passando de 0,096 nm no sistema sem ligante para 0,066 nm no complexo com o substrato, considerando todas as simulações de cada sistema.

A Tyr217 ainda se mostra como o resíduo com as maiores flutuações, ainda que seu RMSF tenha sido reduzido em mais de 0,7 Å em relação a enzima livre, passando de 0,22 nm para 0,15 nm de flutuação, em média. Em segundo lugar entre as maiores flutuações e contramão à Tyr217 está a Tyr70, cujo RMSF passou de 0,12 para 0,13 nm no comparativo entre as médias das simulações, seguida da Phe104 e da Ile8.

A diminuição abrupta da mobilidade da Tyr217 para o sistema em complexo com o substrato, na comparação com a simulação sem ligante, indica que este resíduo está envolvido em uma interação importante entre o sítio ativo enzimático e esse composto, reforçando a ideia de que essa tirosina desempenha um papel catalítico importante.

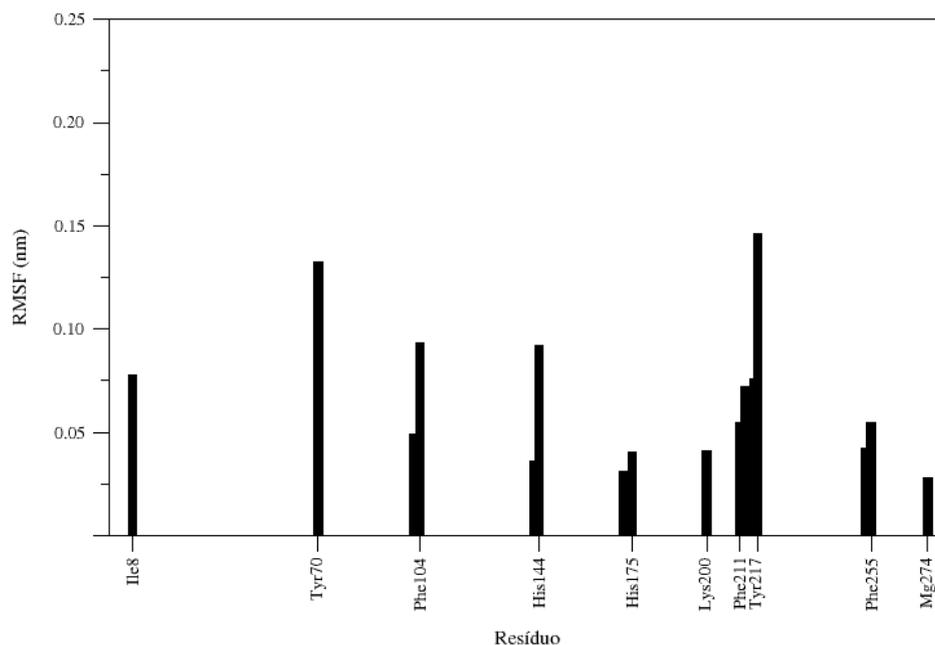


Figura 49 - RMSF médio dos resíduos do sítio ativo em relação à estrutura inicial das simulações enzima-substrato.

Por outro lado, o leve aumento de RMSF da Tyr70 na simulação com ligante mostra que esse resíduo apresenta um papel mais secundário, ao contrário do observado através do *docking* do substrato, que mostrava a participação desse resíduo em uma ligação de hidrogênio importante com a porção carboxílica do 3-DHS.

Uma das interações mais significativas de natureza não-covalente que podem ocorrer entre uma macromolécula biológica e um ligante, seja ele substrato ou um inibidor, são as ligações de hidrogênio. Simulações de dinâmica molecular possibilitam a análise do comportamento dinâmico e da evolução temporal dessas ligações ao longo do tempo, o que está mostrado na Figura 50 para o complexo da AsbF com seu substrato.

Observa-se que nas três replicatas há um número elevado de ligação de hidrogênio entre a enzima e o substrato, as quais variam entre o valor mínimo de 1 e máximo de 11 ligações simultâneas, ambos muito pouco frequentes. Em média, considerando-se a totalidade dos resultados, há formação de 5,02 ligações de hidrogênio simultaneamente.

A réplica 2 apresentou o maior valor médio de ligações de hidrogênio simultâneas (6,27) enquanto a réplica 3 apresentou o menor valor, de 4,31, ao passo em que a primeira simulação produziu um valor intermediário, de 4,47 ligações simultâneas.

Além do acompanhamento do número de ligações formadas, é interessante o conhecimento de quais são as ligações de hidrogênio mais comuns e mais duradouras e com que frequência cada uma dessas ligações está ocorrendo.

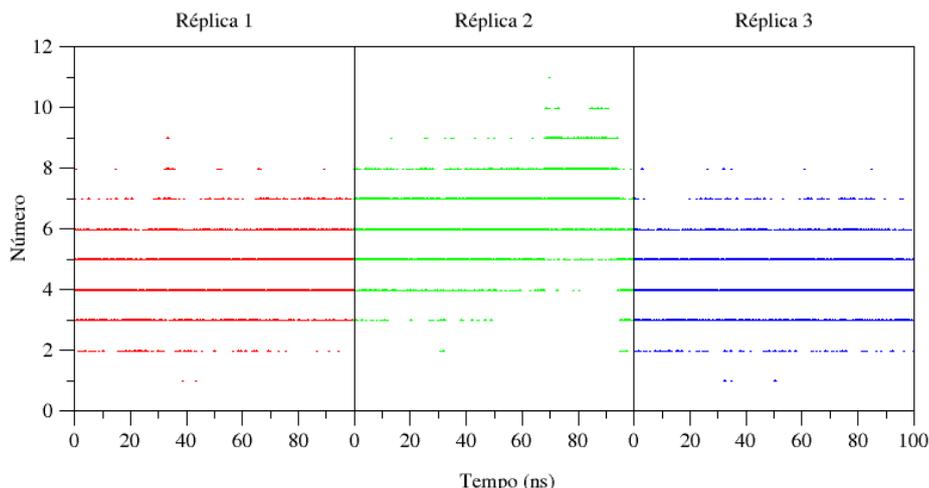


Figura 50 – Evolução temporal do número de ligações de hidrogênio que ocorrem entre a enzima e o substrato para cada réplica realizada.

Para isso, analisa-se a *ocupação* das ligações de hidrogênio, que é definida como a razão entre o tempo em que a ligação de hidrogênio está presente em relação ao tempo total da trajetória considerada, de forma que uma ligação de hidrogênio que nunca se rompa ao longo de toda a simulação apresenta uma ocupação unitária. Em geral, se define ainda um limite mínimo de ocupação, geralmente de 10%, abaixo do qual se consideram as ligações pouco importantes ou até mesmo desprezíveis, visto que tem um caráter muito mais aleatório do que de uma interação verdadeira entre enzima e ligante.

Na Figura 51 está mostrado o padrão de ocupação das ligações de hidrogênio gerados em cada réplica da simulação da enzima em complexo com o substrato, gerado por meio do script ReadHBMap¹¹² com limite de 10%.

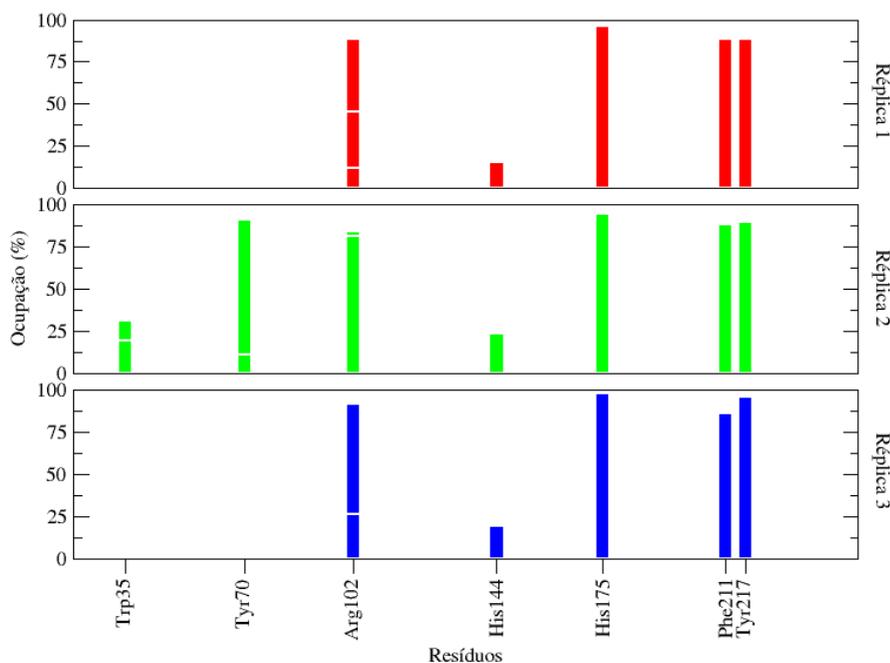


Figura 51 - Ocupação dos resíduos em ligações de hidrogênio, acima do limite de 10%, para cada réplica da simulação enzima-substrato. Os cortes nas barras indicam a ocupação de duas ou mais ligações do mesmo resíduo.

Pode-se observar na figura que há pelo menos 4 ligações de hidrogênio que ocorrem o tempo todo entre a enzima e o substrato, em todas as simulações, e envolvem os resíduos His175, Phe211 (esqueleto peptídico), Tyr217 e Arg102, este último podendo formar até três ligações simultâneas e possivelmente a razão do número médio de ligações ser 5 e não 4.

É interessante notar que esse padrão é muito similar ao observado pelo *docking* na estrutura cristalográfica, mostrando que ao longo de toda a simulação o substrato permanece em uma posição razoavelmente fixa, interagindo com praticamente os mesmos resíduos com os quais interage na configuração inicial.

Uma diferença importante, entretanto, é a maior participação da His175 em detrimento da His144, ao contrário do observado no *docking*. De fato, a His144 forma ligações de hidrogênio em menos de 25% do tempo para todas as réplicas, mostrando que esse resíduo de fato permanece livre para participar do ataque nucleofílico ao hidrogênio anelar em posição 4, como proposto por Pflieger e colaboradores para a primeira etapa do mecanismo de catálise enzimática.

Observa-se ainda que, para a réplica 2, além do padrão de interação comum, há formação de duas ligações de hidrogênio adicionais, envolvendo o Trp35 e a Tyr70. Nesse sistema, a tirosina-70 apresenta ocupação de mais de 90% e inclusive a possibilidade de formar mais de uma ligação de hidrogênio simultaneamente. Esse resíduo interage na porção carboxílica do substrato, juntamente com a Tyr217 e a Arg102, o que sugere que a interação da Tyr70 com o ligante poderia liberar, em alguns momentos, a Tyr-217 para atuação como base na etapa final do mecanismo, que é a desidratação.

O triptofano-35, por sua vez, é um resíduo que compõem o início da hélice- α 2, sendo o resíduo dessa estrutura mais próximo do sítio ativo. Isso justifica porque a mobilidade dessa hélice é reduzida no complexo com o substrato na comparação com a enzima livre.

É possível que essa interação esteja presente nas demais réplicas – justificando porque a mobilidade dessa hélice também é reduzida nas demais simulações – mas não tenha sido contabilizada em função do raio de corte de 3,5 Å utilizando na geração do mapa de ligações de hidrogênio. Em sistemas biológicos é usual definir-se o comprimento máximo de ligações de hidrogênio como até 4,0 Å, porém optou-se por uma abordagem mais conservadora a fim de se filtrar as interações mais essenciais e diminuir a complexidade da análise.¹¹³

Observando a estrutura do complexo para a segunda réplica no tempo 69590 ps – quando ocorrerem 11 ligações de hidrogênio simultâneas, o máximo absoluto (Figura 52) – é possível perceber a formação de uma intrincada rede de ligações de hidrogênio, sobretudo entre a porção carboxílica e resíduos de diversos domínios, inclusive o Trp35, da hélice- α 2, que atua restringindo a movimentação dessa estrutura inteira, como descrito anteriormente.

Na Figura 52 estão discriminadas apenas 8 das 11 ligações de hidrogênio apontadas pelo Gromacs. Isso ocorre pois o programa ADT não sinaliza mais do que uma ligação por resíduo, ainda que alguns resíduos claramente façam mais de uma ligação simultânea, como é o caso da Arg102.

A porção carboxílica do ligante faz ao todo 7 ligações de hidrogênio com o sítio ativo, sendo três ligações com a Arg102 (ponte salina), duas com o Trp35 (o ângulo para ligação com os dois oxigênios carboxílicos é adequado, ainda que a figura não mostre), uma com a Tyr70 e uma com a Tyr217, enquanto que as hidroxilas do anel produzem as quatro ligações restantes com os resíduos His175, Lys200, His144 e o esqueleto peptídico da Phe211.

Observa-se ainda que a His144 se deslocou bastante em relação à posição inicial, participando de uma ligação de hidrogênio com a hidroxila da posição 5, e não mais com a carbonila do substrato, como apontado pelo *docking*. Isso evidencia como a presença do substrato gerou alterações conformacionais no sítio ativo, que corroboram os aspectos mecanísticos anteriormente descritos.

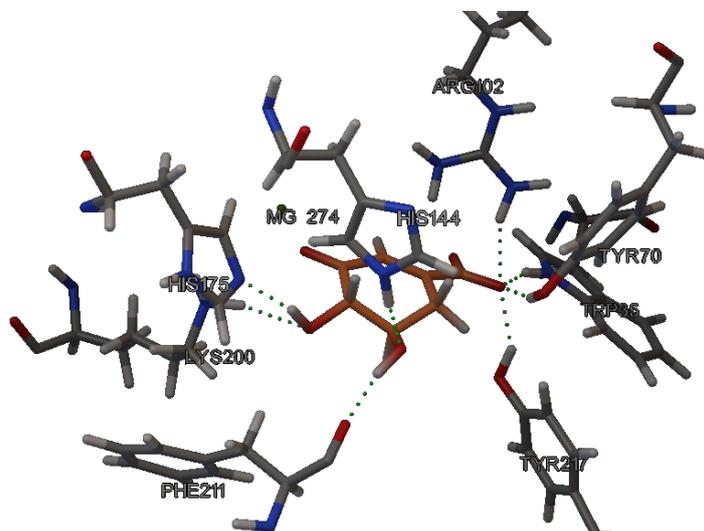


Figura 52 – Padrão de ligações de hidrogênio apontado pelo programa ADT para o quadro de 69590 ps da réplica 2 da simulação enzima-substrato. As ligações de hidrogênio estão pontilhadas em verde.

Um desses aspectos é a atuação da His144 no ataque nucleofílico ao hidrogênio axial em posição 4, o que é factível na medida em que houve de fato um deslocamento desse resíduo por sobre este hidrogênio ao longo da simulação promovido, provavelmente, pela complexação do ligante no sítio.

Outro aspecto interessante é a observação de que a Tyr217 se encontra em uma posição favorável para remover o hidrogênio-6 em posição equatorial no anel, o qual também está em posição *anti* ao grupo OH de saída (posição 5), num forte indício de mecanismo de eliminação E2. Esse grupo de saída é, inclusive, estabilizado não apenas por ligação de hidrogênio com o esqueleto peptídico da Phe-211, mas eventualmente também pela His144 – que estaria protonada após remover o hidrogênio axial 4 – favorecendo essa eliminação, que é a etapa lenta da reação.

É possível ainda acompanhar a evolução temporal de formação e extinção das ligações por meio do gráfico da ocupação parcial – a ocupação a cada passo de tempo ou número de quadros pré-definido – em função do tempo total de simulação. Esse resultado é mostrado na Figura 53 e foi gerado a partir da réplica 1, pois essa apresenta um número de ligações médio mais próximo da média geral, utilizando-se passo de 100 quadros (0,5 ns) e um limite mínimo de ocupação total de 10% para uma ligação ser considerada na análise.

Pode-se observar pela figura que há três interações que são constantes ao longo do tempo de simulação, mantendo uma ocupação acima de 80% em toda a trajetória e envolvem os resíduos Phe211 e His175, que se ligam à porção anelar do ligante, e Arg102, que se liga ao grupo carboxilato. É interessante observar que a ocupação dessas ligações é intensificada ao longo da trajetória, em especial para a arginina, que passa a realizar até 3 ligações de hidrogênio simultâneas

com o substrato nos 20 ns finais da simulação – ainda que não permaneçam até o final de simulação.

A constância e eventualmente a intensificação dessas três interações também foi observada nas outras reproduções do sistema indicando que esses três resíduos são os mais importantes para o ancoramento do substrato no sítio ativo e para fixá-lo em uma conformação adequada para a reação enzimática.

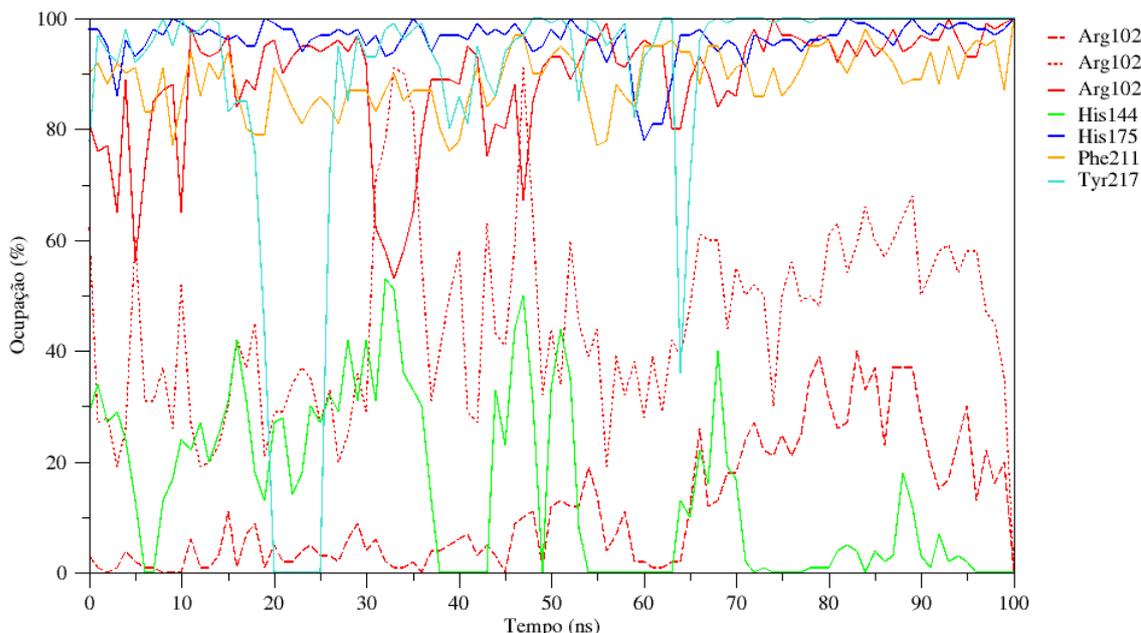


Figura 53 - Ocupação parcial em função do tempo de simulação para o sistema enzima-substrato (réplica 1)

É interessante observar ainda que não há um comprometimento significativo da His144 na formação de ligações de hidrogênio. De fato, tanto nesta como nas demais reproduções do complexo enzima-substrato a ocupação desse resíduo oscila abaixo de 40%, o que permitiria a atuação catalítica desse resíduo nos seus momentos livres, como proposto no modelo mecanístico.

Analogamente, pode-se observar que há momentos ao longo da simulação em que a Tyr217 se desliga da porção carboxílica do ligante, como ocorre entre 20 e 25 ns no gráfico mostrado. Isso reforça a hipótese de que essa tirosina possa atuar diretamente no mecanismo de catálise enzimática, na medida em que mostra que esse resíduo não está plenamente imobilizado por interação com o substrato e que pode, como mostrado anteriormente, adotar em uma configuração geométrica adequada para atuar na desidratação.

Além dessas interações principais, que são comuns às réplicas 1 e 3, a réplica 2 apresentou ainda a formação de ligações de hidrogênio com a Tyr70 e com o Trp35, ambos resíduos que interagem na porção carboxílica do ligante. Essas interações, como descrito anteriormente, podem estar relacionadas à estabilização de porções da estrutura secundária e eventualmente não estão sendo observadas nas demais réplicas em razão dos raios de corte aplicados.

De fato, pode-se observar essa interação em alguns momentos das outras simulações do sistema, como no quadro de 33330 ps da réplica 1, que representam um dos máximos de ligações de hidrogênio simultâneas para essa simulação (Figura 54) e na qual se observa a atuação da Trp35 na interação com a porção carboxílica do substrato, assim como observado na réplica 2 (Figura 52), inclusive observando-se o mesmo padrão de ligações de hidrogênio entre as duas réplicas.

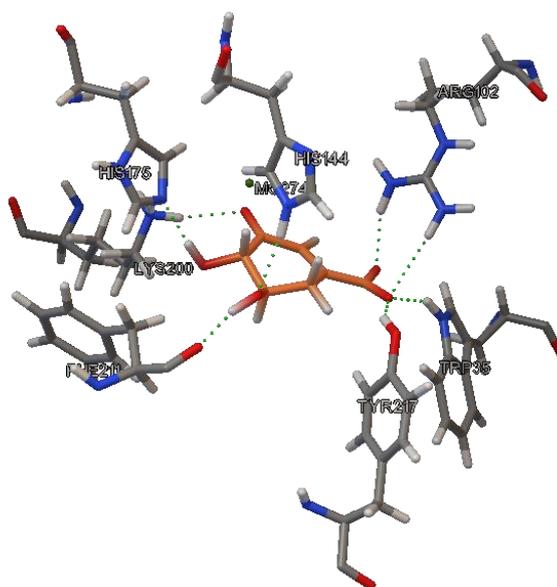


Figura 54 – Padrão de ligações de hidrogênio entre enzima e substrato no tempo de 33330 ps da réplica 2. As ligações de hidrogênio encontram-se pontilhadas em verde.

Isso reforça a atuação do resíduo Trp35, que não é considerado como parte do sítio ativo, na interação com o substrato e mostra que essa interação pode estar relacionada com alterações conformacionais globais do receptor (redução da mobilidade da hélice $\alpha 2$, por exemplo) as quais podem impactar o mecanismo de catálise enzimática, que depende do posicionamento adequado dos resíduos cataliticamente ativos.

5.2.3 Enzima em complexo com produto

Através da cristalografia de raios-X, Pflieger e colaboradores observaram que a enzima AsbF havia se cristalizado complexando o produto, ácido 3,4-dihidroxibenzoato (DHB).⁵⁸ Porém, apenas através do presente trabalho e das investigações experimentais que o precederam foi possível a constatação e comprovação que esse composto de fato atua como um inibidor enzimático razoavelmente potente, sinalizando um mecanismo de retro-inibição.

Por essa razão, é interessante compreender a evolução dinâmica da interação deste composto com o sítio ativo enzimático com vistas não apenas no desenvolvimento de inibidores mais eficazes, mas também no aprimoramento da enzima por meio de engenharia genética, na medida em que o ácido 3,4-DHB é um insumo industrial importante e essa enzima é apontada como muito promissora na biossíntese industrial desse composto.¹¹⁴

Novamente foram realizadas 3 reproduções da simulação da enzima em complexo com o produto de 100 ns cada. Na Figura 55 está mostrado a evolução temporal do RMSD dos

carbonos- α para cada uma dessas réplicas e para sua média, em relação à estrutura inicial. Pode-se observar que o desvio quadrático médio apresenta uma oscilação razoável, sobretudo para a réplica 1, porém tende a uma estabilização nos tempos finais de simulação.

Em média, o RMSD oscila em torno de 0,1144 nm, praticamente o mesmo valor observado para a simulação enzima-substrato. Percebe-se ainda que após um rápido crescimento nos 10 ns iniciais, um patamar é obtido e se mantém relativamente constante ao longo de as simulações, indicando que um estado de equilíbrio foi atingido.

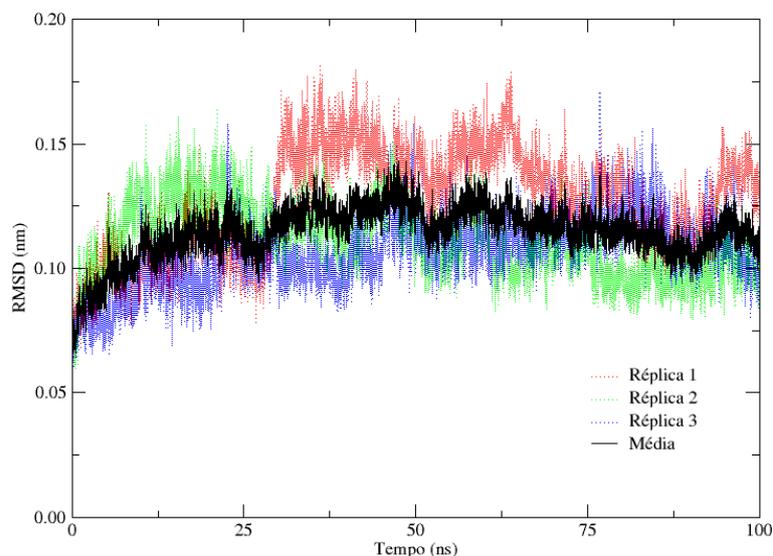


Figura 55 – Evolução temporal do RMSD dos carbonos- α em relação a estrutura inicial para as 3 reproduções da simulação enzima-produto e sua média.

Pela análise da mobilidade dos carbonos- α realizada com o MDLovoFit (Figura 56) percebe-se que as réplicas 2 e 3 apresentam um padrão praticamente idêntico de RMSD acumulado em função da fração de carbonos- α considerada, sendo que 97% dos átomos para a réplica 2 e 98% para a réplica 3 apresentam desvio quadrático médio inferior a 1,0 Å. Apenas a réplica 1 destoou desse padrão, porém, ainda assim 91% dos seus carbonos- α podem ser alinhados com um RMSD inferior a 1,0 Å.

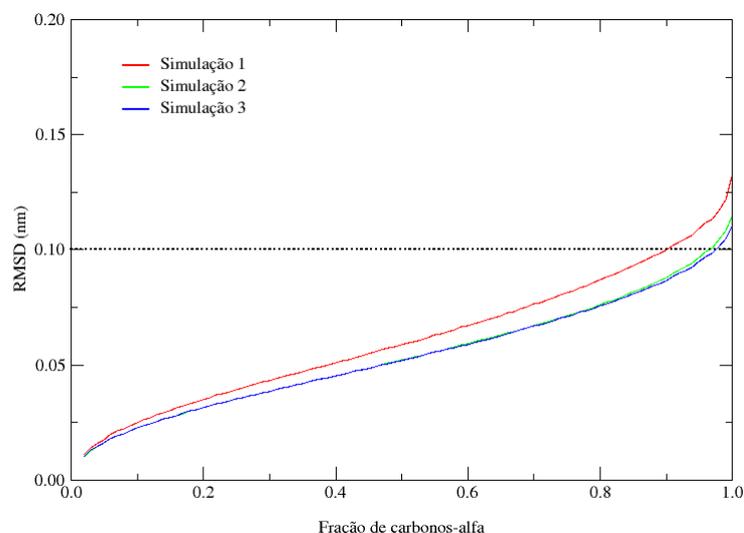


Figura 56 - RMSD acumulado dos carbonos- α em função da fração de átomos consideradas no ajuste para as três réplicas da simulação enzima-produto.

Esse comportamento é muito similar ao observado na simulação enzima-substrato, podendo-se alinhar, em média, 95% dos carbonos- α a menos de 1,0 Å de RMSD, mostrando que a interação com esses ligantes induz alterações em outras porções da enzima, que impactam a mobilidade de toda a estrutura.

Pela análise visual do RMSD dos carbonos- α (Figura 57) observa-se que não há um padrão tão claro de mobilidade como o observado nas simulações enzima-substrato. É interessante notar, no entanto, que regiões mais próximas do sítio ativo, como o laço helicoidal e a alça $\beta 1-\alpha 1$ se mostram bem menos flexíveis com o produto do que com o substrato, indicando a existência de uma interação mais intensa, que resulta em uma restrição maior do sítio ativo.

Novamente se observa que a hélice- $\beta 2$ apresenta baixos valores de RMSD nas três simulações, evidenciando que, assim como com o substrato, há interação entre resíduos dessa região e o ligante a qual impacta a mobilidade dessa porção enzimática.

De fato, a restrição da mobilidade dos resíduos do sítio ativo pela interação com ligantes aparentemente se propaga ao longo de toda a estrutura visto que tanto na simulação da enzima com substrato, mas especialmente na simulação com o produto, as porções de maior mobilidade são alças, como a que liga a hélice- $\beta 3$ à folha- $\alpha 3$ ou então as porções N-terminal (hélice $\alpha 1$) ou C-terminal, que são as porções mais afastadas do sítio ativo.

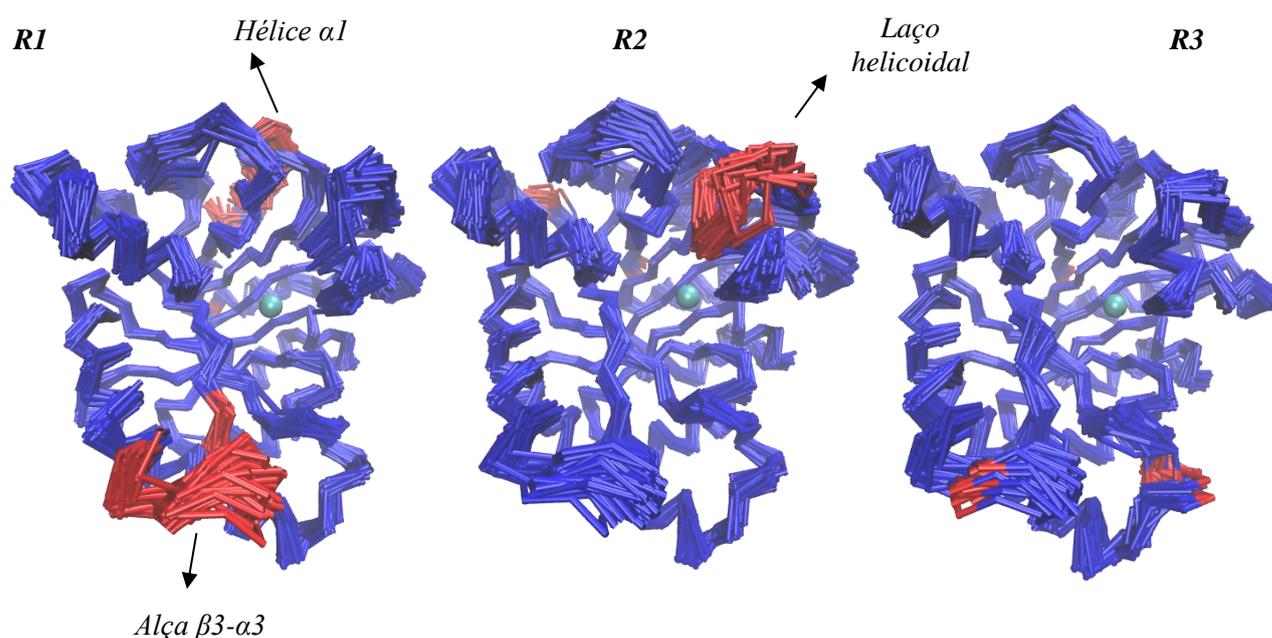


Figura 57 – Sobreposição de 100 frames extraídos a cada nanossegundo para as três reproduções da simulação enzima-produto (R1-R3) coloridos segundo RMSD dos carbonos- α . Em vermelho as porções mais móveis e em azul as mais rígidas. O cofator magnésio está destacado.

Uma exceção, no entanto, seria o laço helicoidal que compõe a parede externa do sítio ativo e que apresentou uma mobilidade significativa nas simulações enzima-substrato, porém valores de RMSD menores na simulação com o 3,4-DHB. Possivelmente é a interação com os resíduos dessa porção o fator determinante para a atividade inibitória dos ligantes.

Essa diminuição da mobilidade se faz mais clara quando se observa a flutuação quadrática média dos resíduos (Figura 58), em que se pode perceber uma diminuição significativa

na mobilidade do laço helicoidal, que passa de 0,3 nm no complexo com substrato para pouco mais de 0,15 nm na simulação com produto.

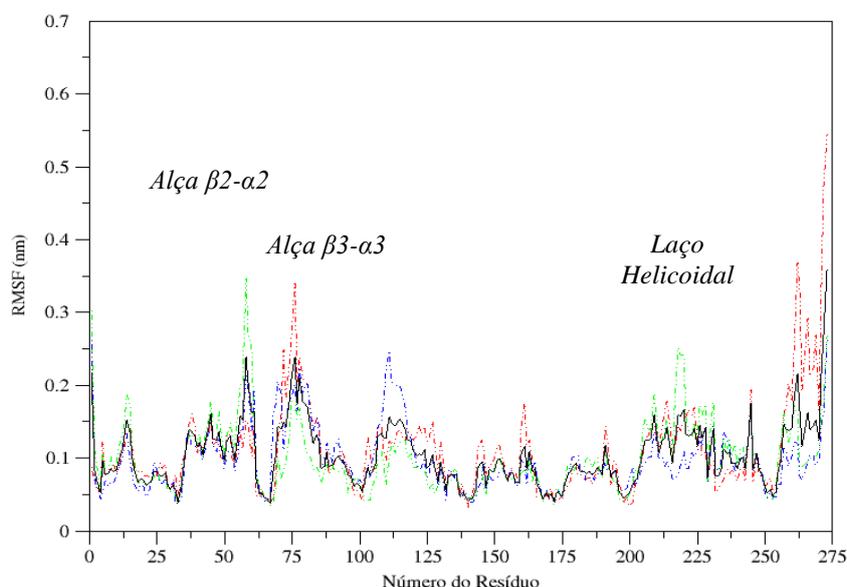


Figura 58 – Flutuação quadrática média de todos os resíduos para as três réplicas da simulação enzima-produto e para a média.

De fato, as regiões com maiores valores de RMSF são as porções nitrogênio e carbono-terminais, compostas por resíduos naturalmente mais móveis por estarem nas pontas livres da cadeia peptídica, bem como algumas alças, como a que liga a hélice- $\alpha 2$ à folha- $\beta 2$ e a alça $\alpha 3$ - $\beta 3$, que mantém o mesmo RMSF que o observado para o complexo com o substrato.

Especificamente para o sítio ativo, se observa um RMSF médio de 0,074 nm (Figura 59), um pouco maior do que observado para a simulação com o substrato (0,066nm), possivelmente em função do aumento da mobilidade da Tyr217, cujo RMSF passou de 0,15 nm na simulação com o substrato para 0,18 nm com o produto, porém ainda abaixo do observado para a enzima livre (0,22 nm) e da Pro213, que passou de 0,072 para 0,093 nm de RMSF, mas que também se manteve abaixo do observado na ausência de ligantes.

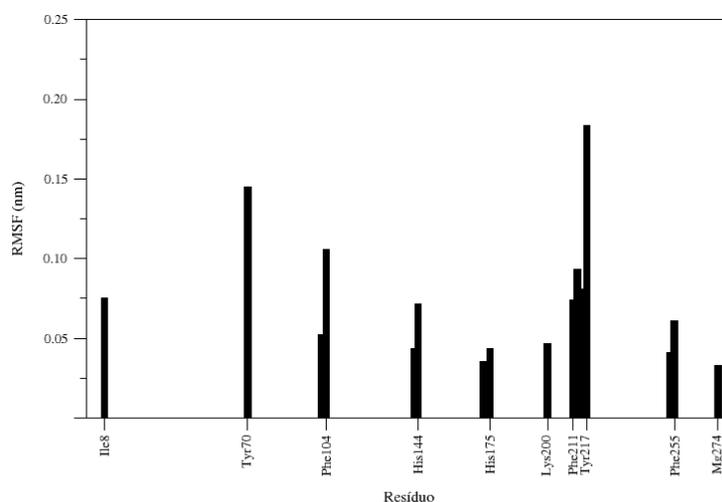


Figura 59- RMSF médio dos resíduos do sítio ativo em relação à estrutura inicial das simulações da enzima em complexo com seu produto.

Em geral, os valores médios de RMSF obtidos para esse sistema são muito próximos dos observado para a enzima com o substrato, indicando a existência de um padrão de interações similar entre os compostos. A maioria dos resíduos manteve o RMSF estável no comparativo dos dois sistemas. A única divergência, além da Tyr217 e da Pro213, é a His144 que apresentou uma diminuição de 0,92 Å para 0,71 Å no comparativo entre as simulações, sendo o único resíduo do sítio ativo que apresentou diminuição significativa de RMSD quando se compara os dois ligantes.

Considerando que para a enzima livre o RMSF médio desse resíduo é 1,0 Å, percebe-se que houve uma interação mais significativa deste com o produto do que com o substrato, razão pela qual há essa diminuição significativa da mobilidade. É possível, inclusive, que essa interação tenha grande responsabilidade na atuação do produto como um inibidor, especialmente quando se considera o papel fundamental desse resíduo no mecanismo de catálise enzimática.

Analisando-se o padrão temporal de ligação de hidrogênio entre enzima e produto ao longo das três triplicatas da simulação (Figura 60), observa-se que são formadas até 9 ligações de hidrogênio simultâneas, portando duas ligações a menos que o valor máximo que substrato foi capaz de formar. Em média, o produto forma 4,84 ligações de hidrogênio simultâneas com o produto, número muito próximo do observado para o substrato, que foi de 5,02 ligações, indicando que a inibição enzimática não está relacionada com a quantidade, mas sim com a intensidade e natureza dessas interações.

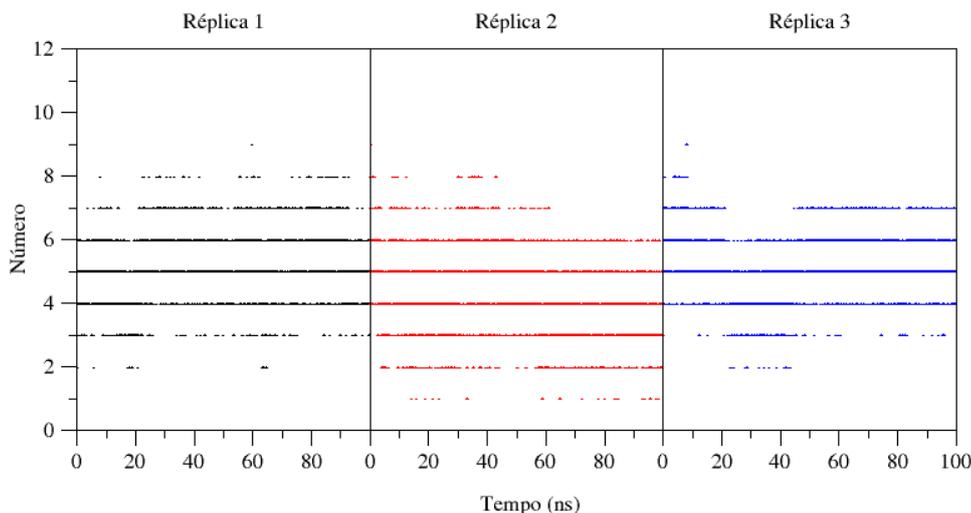


Figura 60 – Evolução temporal do número de ligações de hidrogênio entre enzima e produto para as três réplicas da simulação.

A identidade dessas ligações pode ser constatada no gráfico de ocupação das ligações de hidrogênio (Figura 61), em que se observa uma menor participação da Tyr217 e da Phe211 em ligações de hidrogênio, o que é coerente com o observado na análise de RMSF, e mostra que esses resíduos possivelmente não desempenham um papel tão fundamental na inibição enzimática.

Por outro lado, a Tyr70 apresentou incremento significativo formação de ligações de hidrogênio, com ocupação média passando de 34% na simulação com substrato para mais de 65% na simulação com produto. De fato, com o substrato houve formação de ligações de hidrogênio

com esse resíduo em apenas uma simulação, enquanto que com o produto essa interação esteve presente em todas as réplicas.

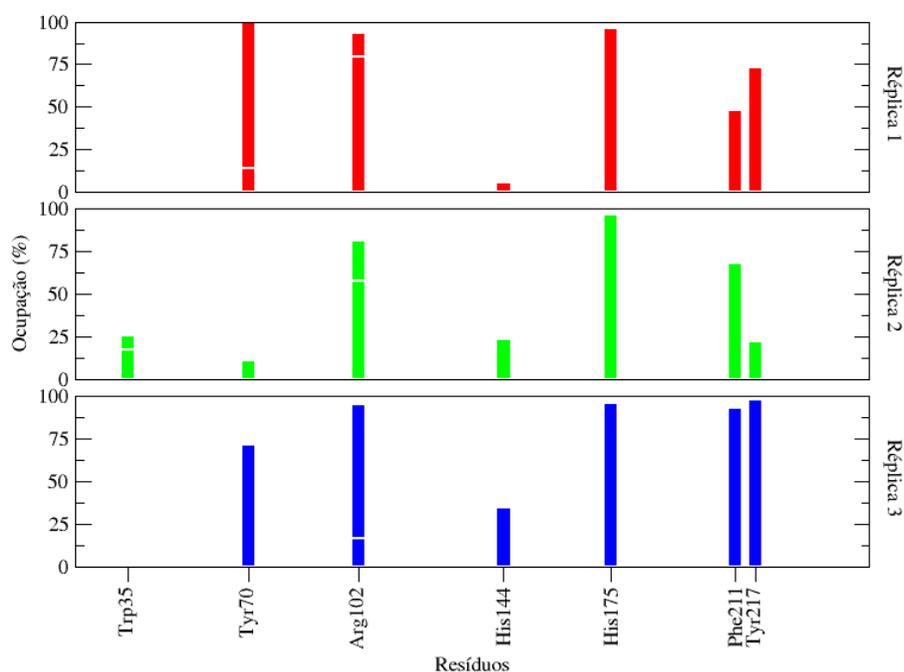


Figura 61 – Ocupação dos resíduos em ligações de hidrogênio para as três réplicas da simulação enzima-produto ao longo dos 100 ns das trajetórias.

Quando comparadas com a simulação com substrato, a Arg102 e a His175 não apresentaram diferenças significativas de ocupação em ligações de hidrogênio, reforçando a ideia de que esses resíduos, que interagem em extremos opostos do ligante, sejam as “âncoras” principais de fixação dos ligantes no sítio ativo enzimático.

Na réplica 2 houve ainda uma fraca ocupação do Trp35, que assim como descrito anteriormente para o substrato, pode não ter sido observado nas outras simulações em função dos raios de corte aplicados, mas que justificaria a diminuição de mobilidade da hélice- α 2 observada para ambos os ligantes, com a consequente estabilização estrutural global da enzima.

Outro aspecto interessante, é a atuação da His144 cuja ocupação passa de 19,4% na simulação com o substrato para 21,3% na simulação com produto. Um leve incremento, porém, em concordância com a diminuição de RMSF observada. Interações de outra natureza também podem ter contribuído para a diminuição de mobilidade observada, em especial empilhamento- π , que já havia sido apontado pelo programa BINANA como um diferencial entre substrato e inibidores na interação com esse resíduo.

Utilizando-se como referência novamente a réplica com um número médio de ligação de hidrogênio intermediário (réplica 3, com 5,09 ligações simultâneas) acompanhou-se a evolução temporal da ocupação, resultado que é mostrado na Figura 62.

Pode-se observar, sobretudo na comparação com o substrato (cf. Figura 51), que há realmente um maior envolvimento da His144 em ligações de hidrogênio, sobretudo na metade final da simulação ao mesmo tempo em que a Tyr70 também passa a atuar nesse tipo de interação – ainda que por vezes se desligue do produto – o que não foi observado para o substrato com a mesma frequência.

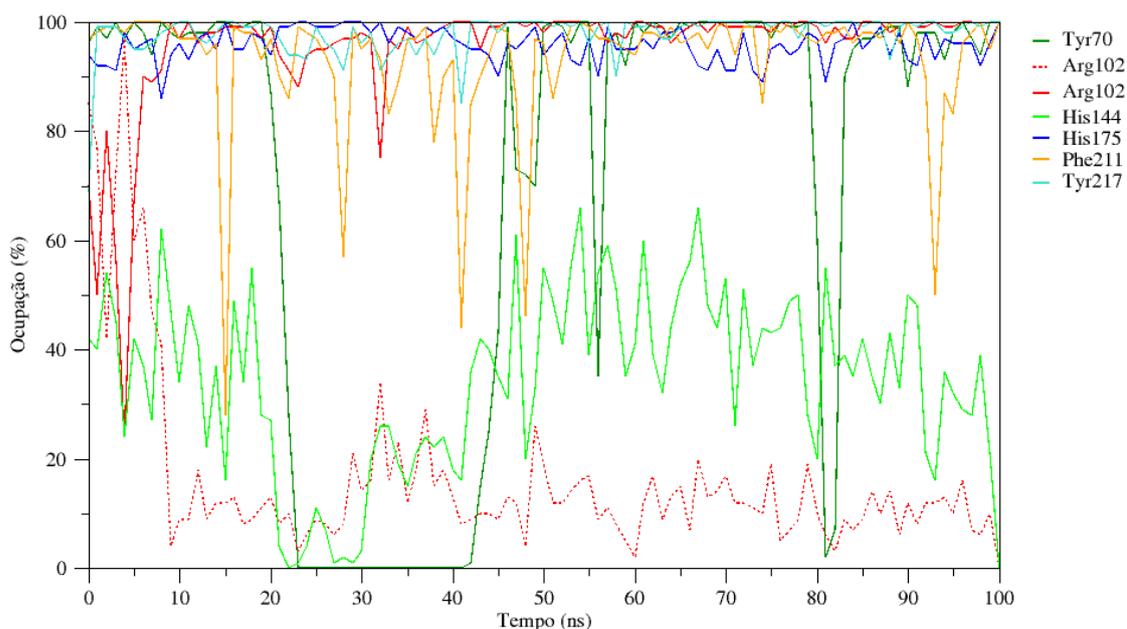


Figura 62 – Evolução temporal da ocupação dos resíduos em ligações de hidrogênio para o complexo enzima-produto.

Novamente se observa que o par Arg102-His175 mantém uma forte interação com o ligante, sustentando uma ocupação superior a 90% ao longo de toda a simulação, mesmo comportamento observado para a Tyr217.

Na Figura 63 está mostrado o padrão de ligações de hidrogênio obtido no frame de 59600 ps da primeira réplica, momento em que ocorrem 9 ligações de hidrogênio simultâneas, o máximo absoluto para esse sistema.

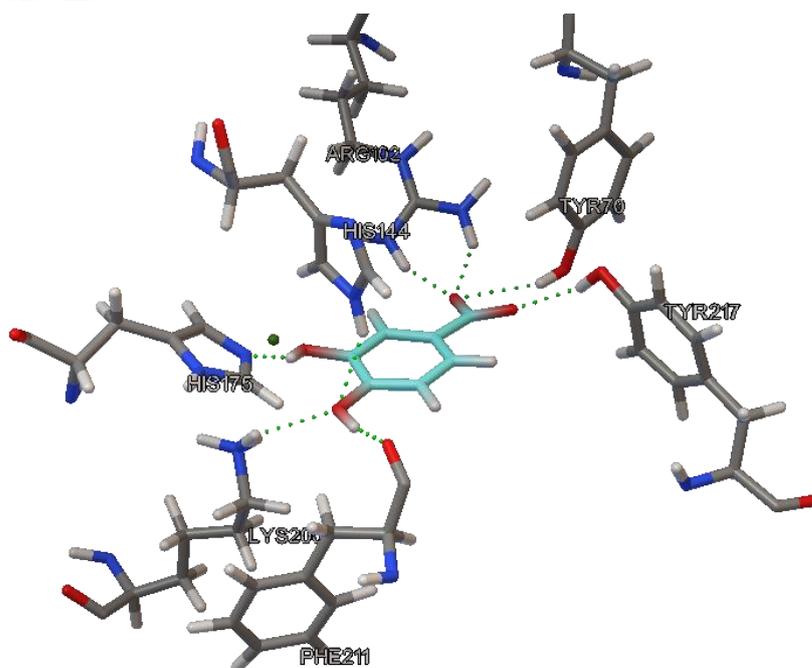


Figura 63 – Padrão de ligação de hidrogênio entre enzima e produto para o frame de 59600 ps da réplica 1, que representam o máximo de ligações simultâneas.

É interessante observar que houve uma mudança significativa no arranjo dos resíduos ao redor do ligante e no próprio padrão de interações quando se compara esse resultado com o obtido por *docking* e com o próprio resultado cristalográfico (cf. Figura 27).

Em primeiro lugar, percebe-se um reposicionamento da Arg102, que adota uma conformação muito mais propícia para a formação de ponte salina com o grupo carboxilato do ligante, ao mesmo tempo em que há uma aproximação da Tyr217 e especialmente da Tyr70 (com quem o grupo COO⁻ provavelmente mantém duas ligações de hidrogênio, ainda que apenas uma seja mostrada na figura), favorecendo a retenção do grupo carboxilato através de uma rede de ligações de hidrogênio.

A formação dessa rede de ligações também ocorre com os grupos hidroxila, sobretudo para o OH em posição *para*, que passa a interagir com a Phe211 e a His144, resíduos que inicialmente estavam envolvidos com a eliminação da hidroxila vizinha e que após a conversão de substrato a produto se rearranjam para formar essa rede, justificando a atuação do produto como um inibidor.

Outro aspecto interessante na comparação do resultado de dinâmica com o dado cristalográfico é a inversão das interações realizadas pelas histidinas 144 e 175. Inicialmente, a His175 estava interagindo com o grupo 4-OH enquanto a His144 interagia com o 3-OH, situação de se reverte durante a dinâmica, indicando que a estrutura cristalográfica realmente apresenta um comportamento distinto do que seria esperado em solução, na medida em que não houve um giro do ligante – as interações com a porção carboxilato se mantem inalteradas – mas sim houve um rearranjo dos resíduos ao redor do mesmo durante a dinâmica.

5.2.4 Enzima em complexo com ácido gálico (ligante 4)

Nos ensaios experimentais de inibição enzimática, o ligante 4 (ácido gálico) se mostrou o mais potente dentre todos os compostos estudados, com IC₅₀ estimado em 4,36 μmol.L⁻¹. Compreender de que forma esse composto atua na inibição enzimática é de fundamental importância no desenho de moléculas mais potentes e sobretudo mais seletivas na interação e inativação da desidroshikimato desidratase de *B. anthracis*.

Por essa razão, a evolução temporal da enzima em complexo com esse ligante também foi estudada por meio de simulação de dinâmica molecular, realizada em triplicada e tomando-se como ponto de partida o modo de ligação de menor energia apontado por *docking*.

A primeira propriedade avaliada nessas simulações foi a evolução temporal do RMSD dos carbonos-α, resultado que está mostrado na Figura 64, em que se pode observar que há um leve aumento do RMSD ao longo das simulações, até aproximadamente 75 ns, quando ocorre uma relativa estabilização e o alcance de um patamar em torno de 0,14 nm. O RMSD médio foi de 0,1215 nm, um pouco acima do valor observado nas simulações com substrato e com produto, mas ainda abaixo do observado para a enzima livre, sugerindo que esse ligante também atua restringindo a mobilidade estrutural da enzima por meio de sua interação com o sítio ativo.

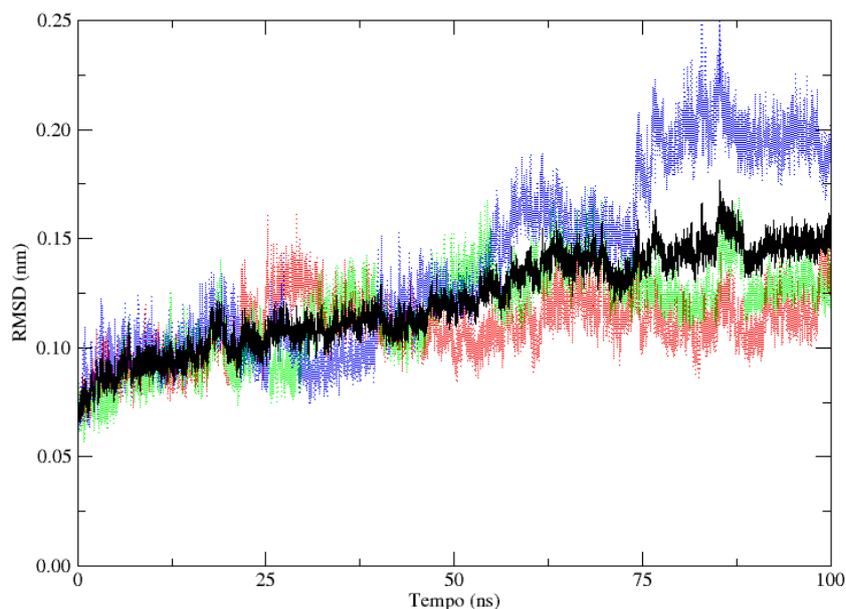


Figura 64 – Desvio quadrático das posições atômicas do carbonos- α para as três reproduções da simulação da enzima em complexo com o ácido gálico.

Grande parte desse aumento da mobilidade média se deve ao comportamento da terceira réplica, que apresentou RMSD médio de 0,1387 nm, sobretudo em função de um rearranjo configuracional que ocorre por volta de 75 ns. Para as demais simulações o RMSD médio oscilou em torno de 0,11 nm, mesmo valor observado para os sistemas estudados anteriormente.

Entretanto, quando se avalia a mobilidade acumulada por meio do MDLovoFit percebe-se que o comportamento das reproduções é muito semelhante (Figura 65) e mesmo a terceira réplica pode ter seus 94 % de seus carbonos- α alinhados com RMSD inferior a 1,0 Å, valor muito próximo do observado para as demais reproduções, que é de 97%.

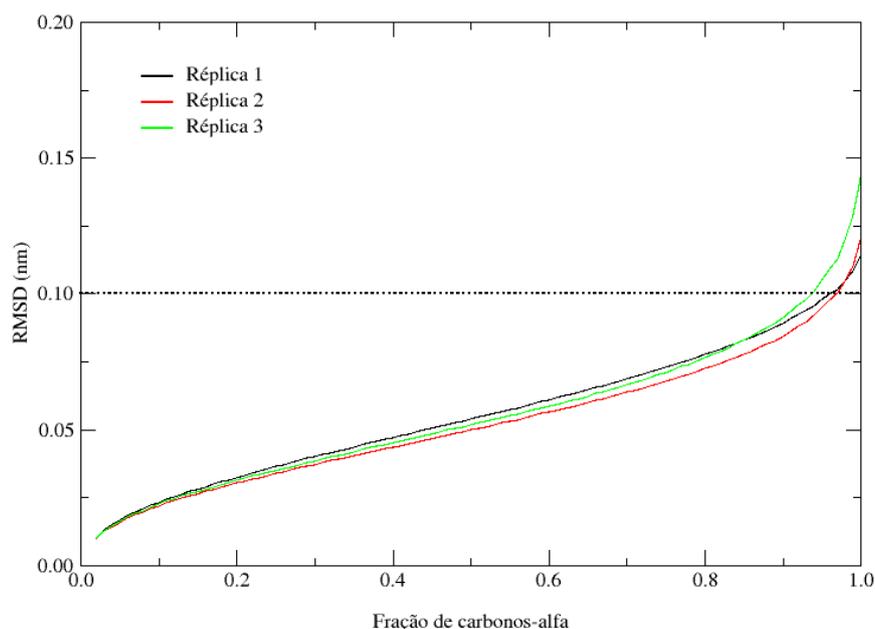


Figura 65 – RMSD dos carbonos- α das três réplicas da simulação enzima-ligante 4 em função da fração de átomos considerada no alinhamento.

Isso indica que um pequeno grupo de átomos é responsável pelo grande incremento de mobilidade observado na terceira réplica desse sistema, enquanto que a maior parte da estrutura não sofreu alterações significativas ao longo das simulações, o que pode ser comprovado visualmente pela sobreposição de frames das três trajetórias geradas e sua coloração de acordo com os percentuais de mobilidade observados (Figura 66).

Percebe-se que os maiores valores de RMSD da réplica 3 ocorrem na região do laço helicoidal, que engloba muito resíduos cataliticamente importantes, enquanto que praticamente a totalidade da estrutura apresenta valores muito baixos de mobilidade, possivelmente até inferiores aos observados nas demais reproduções, sobretudo para as hélices como as hélices $\alpha 2$ e $\alpha 3$. Além do laço helicoidal, outras partes que apresentaram valores de RMSD significativos foram regiões de alça, como as alças $\beta 3-\alpha 2$ e $\beta 4-\alpha 4$.

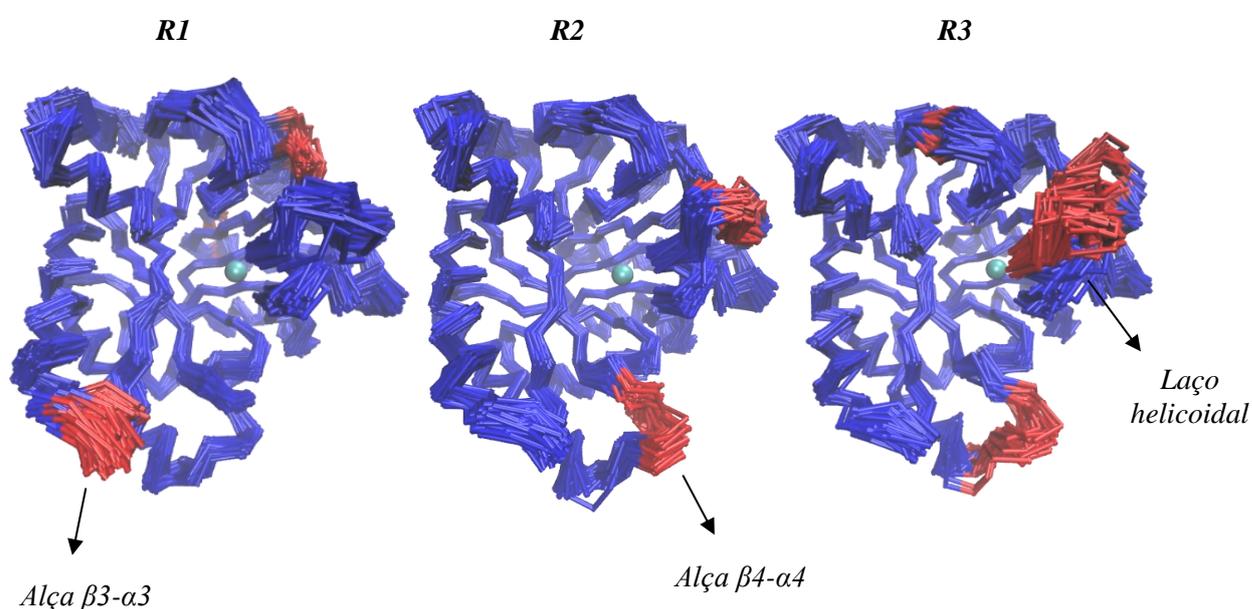


Figura 66 – Sobreposição das estruturas geradas por dinâmica em 100 frames de 1 ns para os sistema enzima-ligante 4 para as três réplicas (R1-R3). A cor vermelha indica porções com RMSD superior a 1,0 Å.

Essa maior mobilidade para o laço helicoidal também pode ser observada nas demais reproduções, ainda que não de forma tão intensa, indicando que a presença do ligante 4 no sítio ativo ao invés de produzir uma restrição na mobilidade do sítio, como observado para o produto e em menor grau para o substrato, gerou, na verdade, um rearranjo estrutural do mesmo.

Eventualmente, esse rearranjo pode estar relacionado a uma relaxação do sítio ativo, que inicialmente estava “ajustado” para a interação com o produto e que foi forçado, por meio do *docking*, a se ligar a um composto mais volumoso e significativamente rígido, o qual, ao contrário do substrato que pode ser moldar mais facilmente ao sítio, obriga que o sítio se molde a ele.

Isso reforça a hipótese de que o sítio ativo da estrutura cristalográfica é excessivamente restrito, dificultando o encaixe dos ligantes por *docking* e contribuindo para a baixa correlação entre as energias de ligação do *docking* e os resultados experimentais. Induzir o alargamento do sítio por meio da dinâmica molecular com ligantes mais volumosos pode permitir a obtenção de

conformações estruturais mais próximas das condições biológicas, seguindo o modelo de encaixe induzido por ligante, e, portanto, mais relevantes para realização do *ensemble docking*.

A existência desse comportamento de relaxação/alargamento do sítio ativo é reforçada pelos dados de flutuação quadrática média dos resíduos, que estão mostrados na Figura 67, em que se observa valores significativamente elevados de mobilidade entre os resíduos 200 e 225, atingindo um valor médio de 0,4 nm, superiores inclusive, ao observado para a enzima livre em solução.

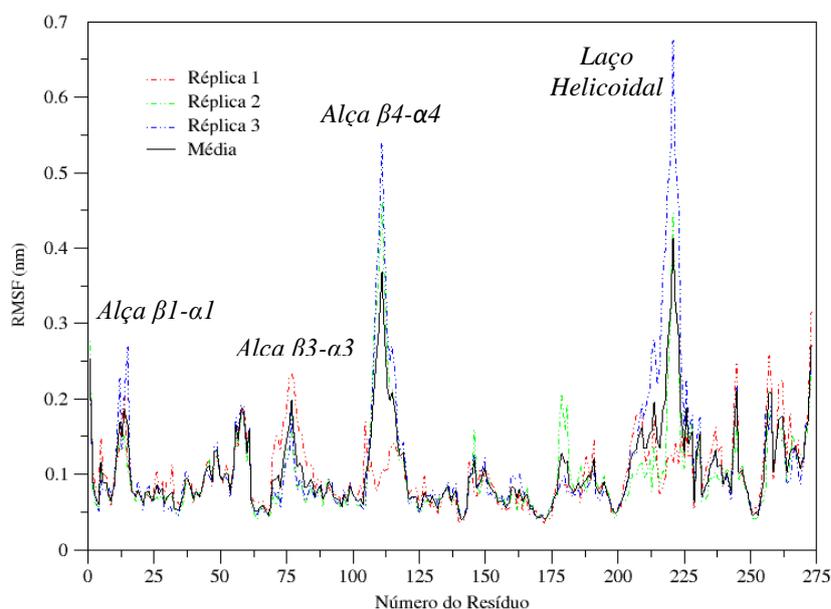


Figura 67 – Flutuação quadrática média (RMSF) de todos os resíduos para as três reproduções do sistema AsbF-ligante 4 e valor médio.

As demais porções da enzima, no entanto, permanecem com RMSF significativamente baixo, em um comportamento comparável ao observado no complexo com o produto (cf. fig. 5.33). À exceção estão algumas regiões de alças, sobretudo a alça $\beta 4-\alpha 4$, porém com impacto muito pequeno sobre o sítio de ligação, uma vez que se encontra significativamente distante deste.

Em relação especificamente aos resíduos que compõem o sítio ativo (Figura 68), observa-se um RMSF médio de 0,079 nm, muito próximo do observado para o produto (0,074 nm) e naturalmente acima do observado para o substrato (0,066 nm).

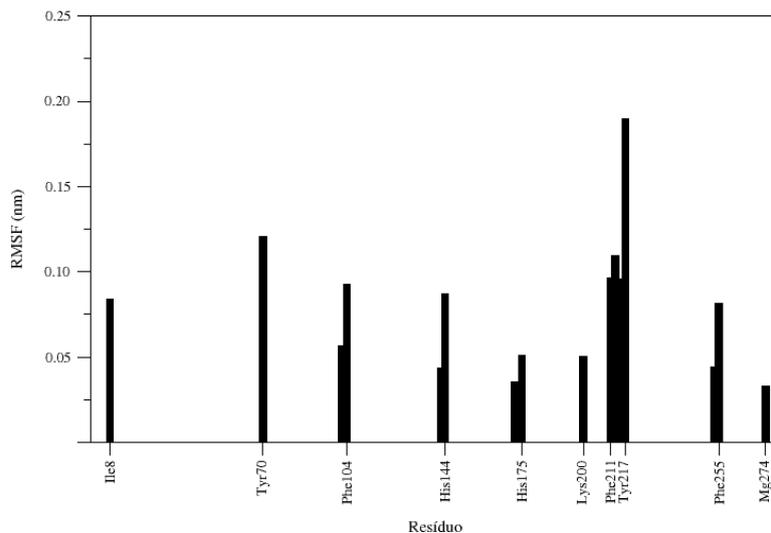


Figura 68 – RMSF médio dos resíduos do sítio ativo das três reproduções do complexo enzima-ligante 4.

Na comparação com o complexo enzima-produto (cf. Figura 59), houve, como esperado, um aumento das mobilidades de todos os resíduos catalíticos localizados no laço helicoidal, com destaque para a Phe211 e Pro213, que passaram de 0,074 nm para 0,096 e de 0,093 para 0,110 nm, respectivamente, na comparação entre os complexos, porém ainda abaixo do observado na ausência de ligante, o que foi compensado com uma diminuição da mobilidade da Tyr70 e Phe104.

Analisando-se o comportamento temporal do número de ligações de hidrogênio entre o ligante 4 e o receptor (Figura 69), observa-se que há, em média, a formação de 4,49 ligações de hidrogênio simultâneas, valor menor do observado para o produto (4,84 ligações) sobretudo por influência da réplica 3, que produziu um valor médio de 3,68 ligações de hidrogênio simultâneas, enquanto que a primeira e a segunda reproduções da simulação levaram aos valores de 4,88 e 4,92 ligações de hidrogênio simultâneas.

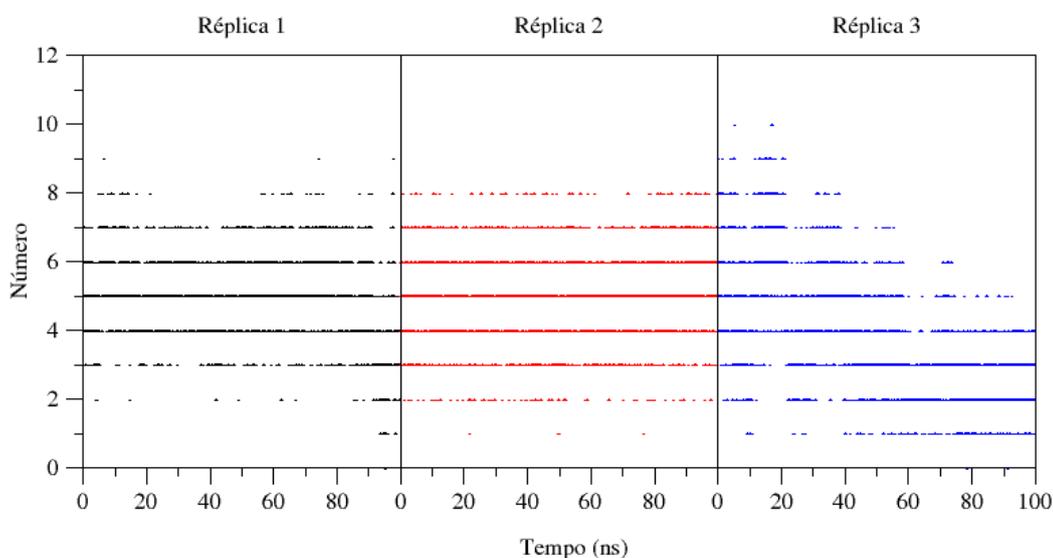


Figura 69 – Evolução do número de ligações de hidrogênio simultâneas formadas entre o ligante 4 e o receptor para as três reproduções desse complexo.

É interessante observar que a réplica 3 também foi a que produziu o maior número de ligações de hidrogênio simultâneas entre ligante e receptor (10 ligações), em dois momentos nos 20 ns iniciais da simulação, ao mesmo tempo em que produziu os menores valores em seus 20 ns finais, mostrando que houve um desligamento do ligante ao longo da simulação, possivelmente induzido pelo rearranjo do sítio ativo anteriormente descrito. É possível inclusive acompanhar a diminuição do número de ligações de hidrogênio máximo formado nessa reprodução ao longo da trajetória.

Porém, tão interessante quanto isso é o fato de que ao longo dessa simulação as alterações estruturais produziram uma maior organização de porções importantes enzima, como pode ser observado na Figura 70, que sobrepõem o primeiro e o último frame da trajetória gerada.

Pode-se observar, por exemplo, que a região do laço helicoidal (em amarelo na figura) passou de uma estrutura secundária disforme para uma forma mais rica em hélices- α , que são, conseqüentemente mais estáveis, indicando que, apesar de haver uma menor interação com o ligante, sua presença induziu alterações conformacionais que agregam uma maior estabilidade estrutural ao sistema, o que possivelmente acontece também nas condições experimentais.

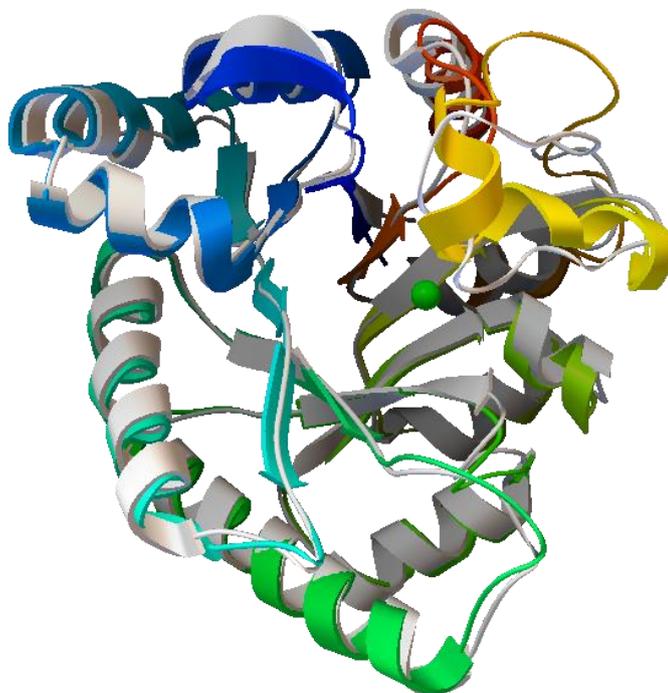


Figura 70 – Sobreposição do primeiro frame (estrutura clara) e do último (estrutura colorida) para a terceira réplica do sistema enzima-ligante 4.

Em relação a ocupação dos resíduos em ligações de hidrogênio, pode-se observar pela Figura 71 que novamente as maiores ocupações pertence aos resíduos Arg102 e His175, os dois ganchos principais de interação entre a enzima e os ligantes, como já foi observado anteriormente para o substrato e produto.

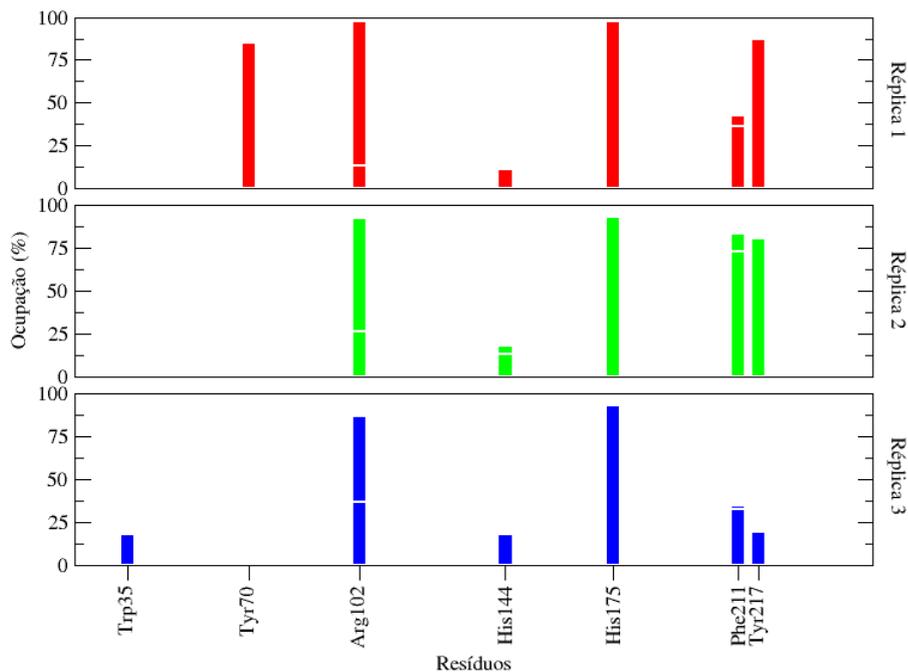


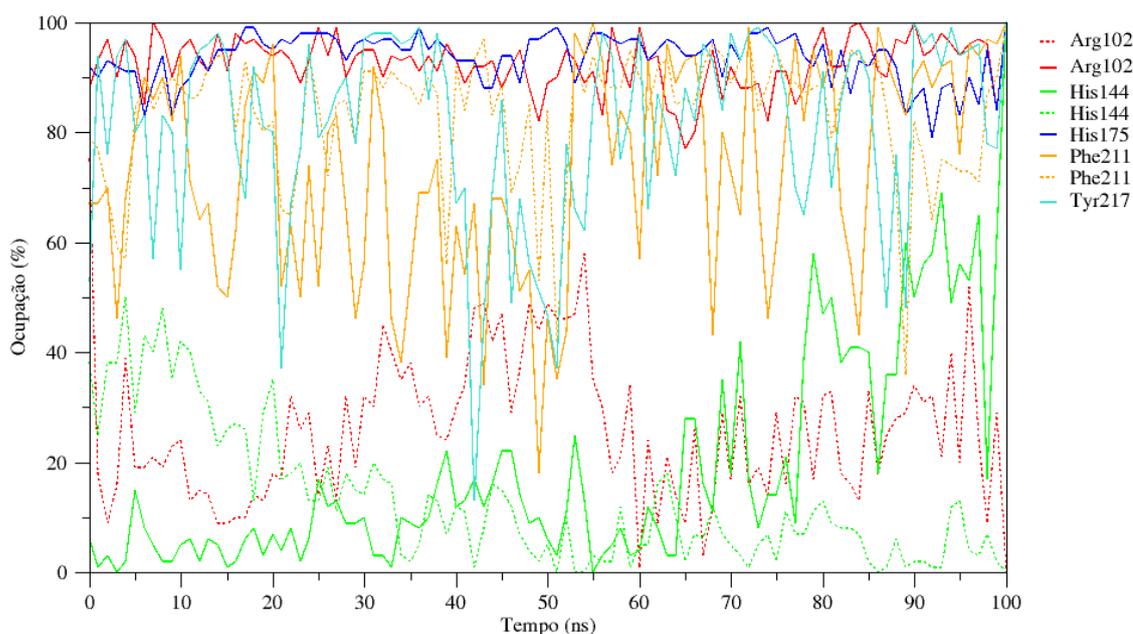
Figura 71– Ocupação dos resíduos em ligação de hidrogênio com o ácido gálico para cada réplica da simulação desse sistema.

Além destas, se destacam as interações estabelecidas pela Tyr217 com ocupação média de 62 %, valor próximo do observado para o produto (64%) e da Phe211, cuja ocupação foi em média de 53%, abaixo do observado para o produto (70%), no entanto, em todas as réplicas esses resíduos é capaz de interagir com duas ligações de hidrogênio – graças à hidroxila extra do ácido gálico comparada com o 3,4-DHB – o que pode estar relacionado com a maior potência do ácido gálico.

Assim como para o produto, a His144 atua em menos de 25% do tempo de trajetória na formação de ligações de hidrogênio, entretanto, conforme mostram os dados da réplica 2, esta pode formar mais de uma ligação simultânea com o inibidor, o que também pode ter relação com a potência do composto, sobretudo diante do papel primordial que esse resíduo detém no mecanismo de catálise enzimática.

A Tyr70 e o Trp35 apresentam ocupação relevante em ligações de hidrogênio em apenas uma das três réplicas cada, indicando que esses detêm um papel secundário ao menos nessa forma de interação. Além disso, a redução observada na mobilidade da Tyr70 não aparenta estar relacionada diretamente com sua interação com o ligante.

Na Figura 72 está mostrada a evolução temporal da ocupação dos resíduos em ligações de hidrogênio ao longo da réplica 2, onde se observa que ao longo de toda a trajetória os resíduos Arg102 e His175 mantém uma ocupação superior a 90%, em semelhança ao que ocorre na simulação com o produto. Em compensação, a interação com a Tyr217 se mostra mais intermitente, ainda que se mantenha em um nível superior a 50% em praticamente todos os momentos.



0

Figura 72 – Evolução temporal da ocupação em ligações de hidrogênio ao longo da réplica 2 do complexo enzima-ligante 4

Destacam-se ainda a interação com a Phe211, que apesar de inconstante, ocorre por meio de duas ligações de hidrogênio simultâneas com o ácido gálico, ambas com ocupação superior a 40%, bem como a interação com a His144, que também ocorre por meio de duas ligações, uma das quais tem sua ocupação aumentada ao longo da simulação, atingindo 100% nos instantes finais de simulação, dois comportamentos que não foram observados em nenhuma das simulações com o produto.

Na Figura 73 está mostrado o padrão de ligações de hidrogênio dos frames de 97680 ps (réplica 1, figura A) e 17020 ps (réplica 3, figura B), que correspondem aos valores máximos de 9 e 10 ligações de hidrogênio simultânea entre a enzima e o ácido gálico, respectivamente.

Pode-se observar que assim como para o produto e o substrato, a porção carboxílica do ligante forma duas ligações de hidrogênio com a Arg102, com características de ponte salina e uma com a Tyr217. No frame da réplica 3 há também a formação de ligação de hidrogênio entre essa porção e o Trp35.

A interação com o Trp35 também foi observada em algumas das reproduções para produto e substrato, não sendo, entretanto, tão frequente e perene quanto as outras duas interações e sendo comumente substituída pela interação com Tyr70, o que ocorre frequentemente no produto.

Isso mostra que tanto a interação com Tyr70 quanto com o Trp35 não detém funções essenciais na potência dos ligantes, mas poderiam ser utilizados para aprimorar a seletividade dos mesmos pela inserção de grupos capazes de interagir com ambos resíduos simultânea e mais frequentemente – o que poderia ser feito por uma reação de esterificação com um álcool de cadeia curta, visto que essa porção do sítio se mostra significativamente impedida.

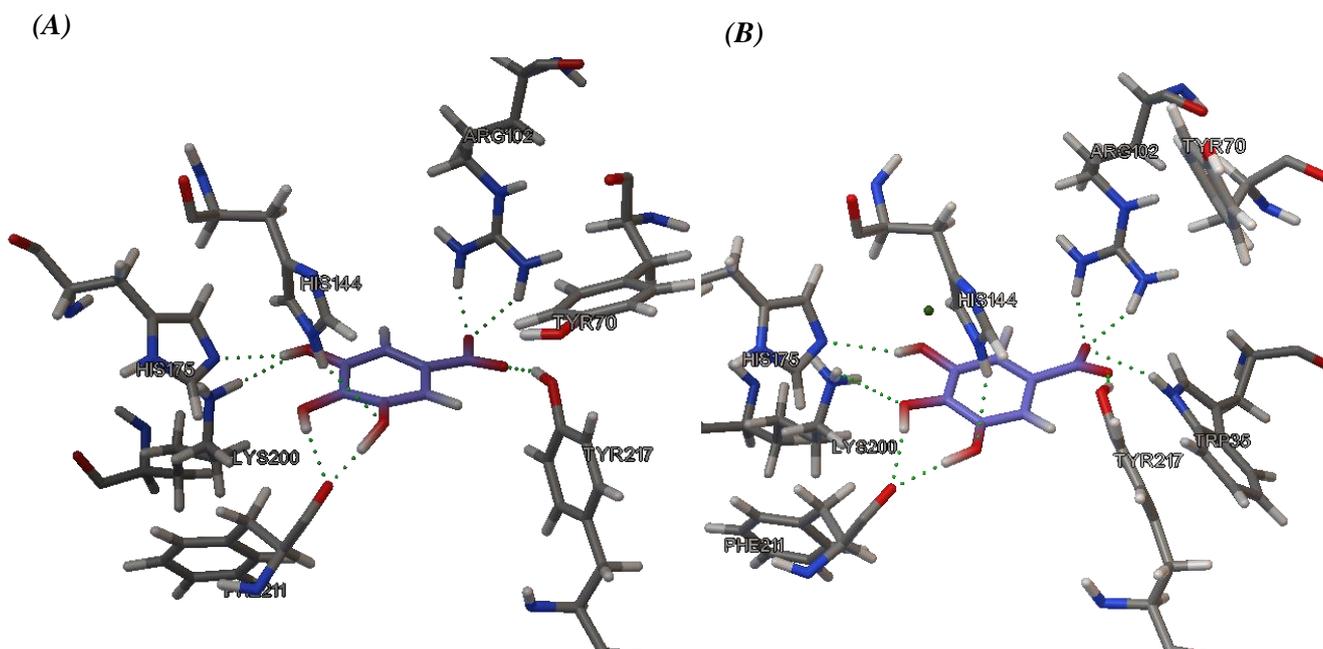


Figura 73 – Padrão ligações de hidrogênio que ocorrem entre o ligante 4 e o sítio ativo enzimático no frame de 97680 os da réplica 1 (figura A) e de 17020 ps da réplica 3 (figura B) e que correspondem aos máximos de LHs simultâneas de suas trajetórias. As ligações estão pontilhadas em verde.

Certamente determinante na ação inibitória, no entanto, são as interações que ocorrem com os grupos hidroxila do ligante e os resíduos His144, His175, Lys200 e Phe211, esse último integrante do laço helicoidal. Pode-se observar que há a formação de uma rede de ligações de hidrogênio e que envolve inclusive a formação de duas ligações de hidrogênio com a Phe211 e possivelmente também duas com a His144, o que não é de fácil ocorrência com o produto e provavelmente contribui para o potencial inibitório do ácido gálico ($4.36 \mu\text{mol.L}^{-1}$) ser quase 10 vezes maior do que o do produto ($38 \mu\text{mol.L}^{-1}$).

Assim como observado para o produto e diferindo dos resultados de *docking* (cf. Figura 34) a His175 atua formando uma ligação de hidrogênio com a hidroxila em posição 3, a qual mantém-se durante todas as simulações, como mostrado pela evolução temporal da ocupação, enquanto que a His144 prefere interagir com as hidroxilas 4 e 5, ao contrário do apontado por *docking*, em que mostrou a interação deste resíduo com a hidroxila em posição 3.

Isso indica que, assim como para o produto, ao longo das simulações houve uma alteração conformacional do sítio ativo induzida pela presença dos ligantes levando os resíduos a orientações diferentes daquelas apresentadas na estrutura cristalográfica e que certamente impactam os resultados de *docking* e sua correlação com os dados experimentais.

5.2.5 Enzima em complexo com ligantes 27 e 37

Em meio à baixa correlação observada entre os resultados de *docking* e os dados experimentais, dois ligantes se destacaram em virtude de seus elevados desvios frente aos dados de potencial inibitório *in vitro*, que foram o ligante 27 (desvio de 49%) e especialmente o ligante 37, que apresentou um desvio de quase 70% sobretudo em razão da superestimação das energias de ligação pelo *docking*.

Para compreender as razões de justificam esses desvios e racionalizar eventuais falhas na metodologia de *docking* empregada, realizaram-se simulações de dinâmica molecular desses ligantes a partir de seus modos de ligação de menor energia apontados por *docking* e em duplicata para cada sistema.

Novamente, primeira propriedade avaliada é o desvio quadrático das posições atômicas dos carbonos- α , que permite um acompanhamento geral das simulações e a verificação de que um estado de equilíbrio dinâmico foi atingido, resultado que está mostrado na Figura 74, para os dois sistemas.

Pode-se observar que há um aumento do RMSD nos primeiros 25 ns das simulações, seguido de uma relativa estabilização, sobretudo a partir da metade final, quando um patamar é obtido, indicando que os sistemas passam a oscilar em torno de conformações de equilíbrio.

Em média, se observa um RMSD de 0,1099 nm para o complexo com ligante 27 e de 0,1144 nm para o complexo com o ligante 37, na prática o mesmo valor observado para as simulações com o substrato e com produto, mostrando que esses ligantes não induzem alterações estruturais significativas à enzima, o que é coerente com seu volume relativamente reduzido, comparado ao ácido gálico, por exemplo.

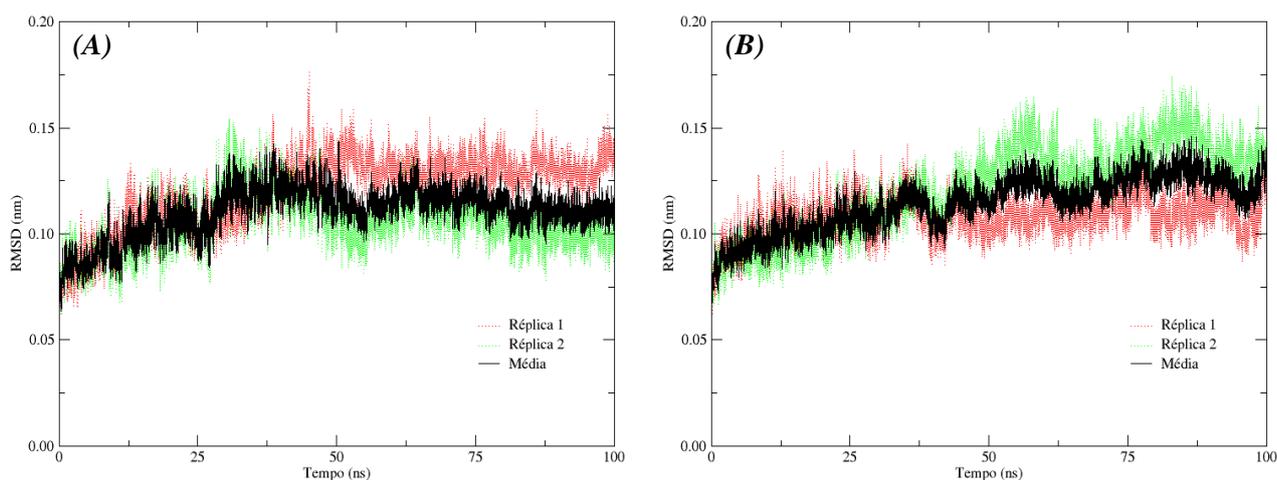


Figura 74 – RMSD das posições dos carbonos- α ao longo das simulações da enzima em complexo com ligante 27 (A) e em complexo com ligante 37 (B) e suas respectivas médias.

Essa constatação é comprovada pela análise acumulada de RMSD dos carbonos- α em função da fração de átomos considerada, gerada pelo MDLovoFit (Figura 75), em que se pode

observar que, nas quatro simulações realizadas, mais de 96% dos carbonos- α podem ser alinhados, em média, com desvio quadrático médio inferior a 1,0 Å.

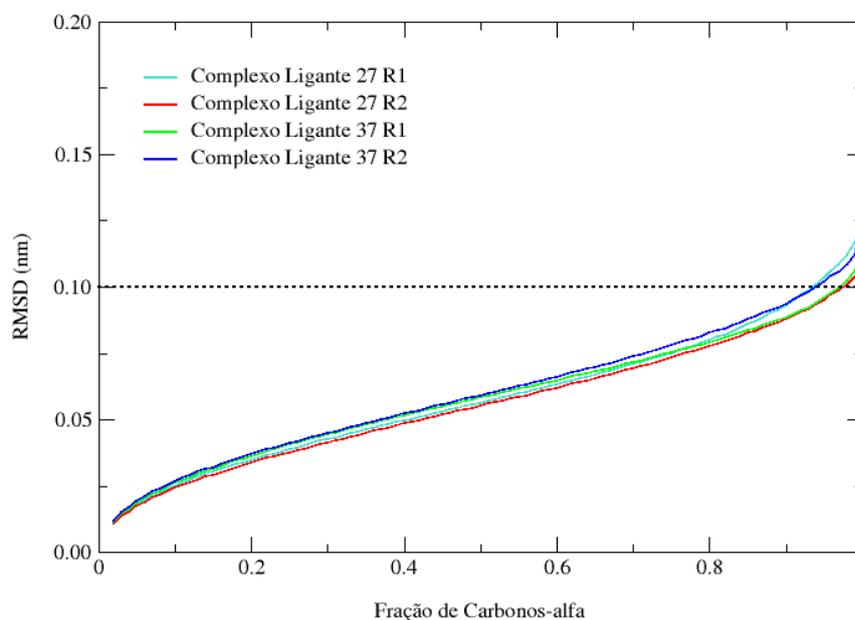


Figura 75 – RMSD acumulado em função da fração de carbonos- α da estrutura considerada na análise para as duas réplicas das simulações da enzima em complexo com os ligantes 27 e 37.

Através da sobreposição de 100 frames gerados para cada uma dessas trajetórias e coloridos de acordo com as mobilidades quadráticas médias dos resíduos (Figura 76) pode-se observar que as maiores mobilidades ocorram com porções de alça da enzima, sobretudo a alça $\beta 3-\alpha 3$, que tem porções de RMSD elevado para os dois sistemas e em todas as simulações, sendo relativamente maior para o complexo com o ligante 37.

Para o complexo com o ligante 27, ainda que haja uma redução de mobilidade dessa alça, se observa uma maior movimentação de resíduos próximos ao sítio ativo, inclusive do próprio laço helicoidal, como ocorre na réplica 1.

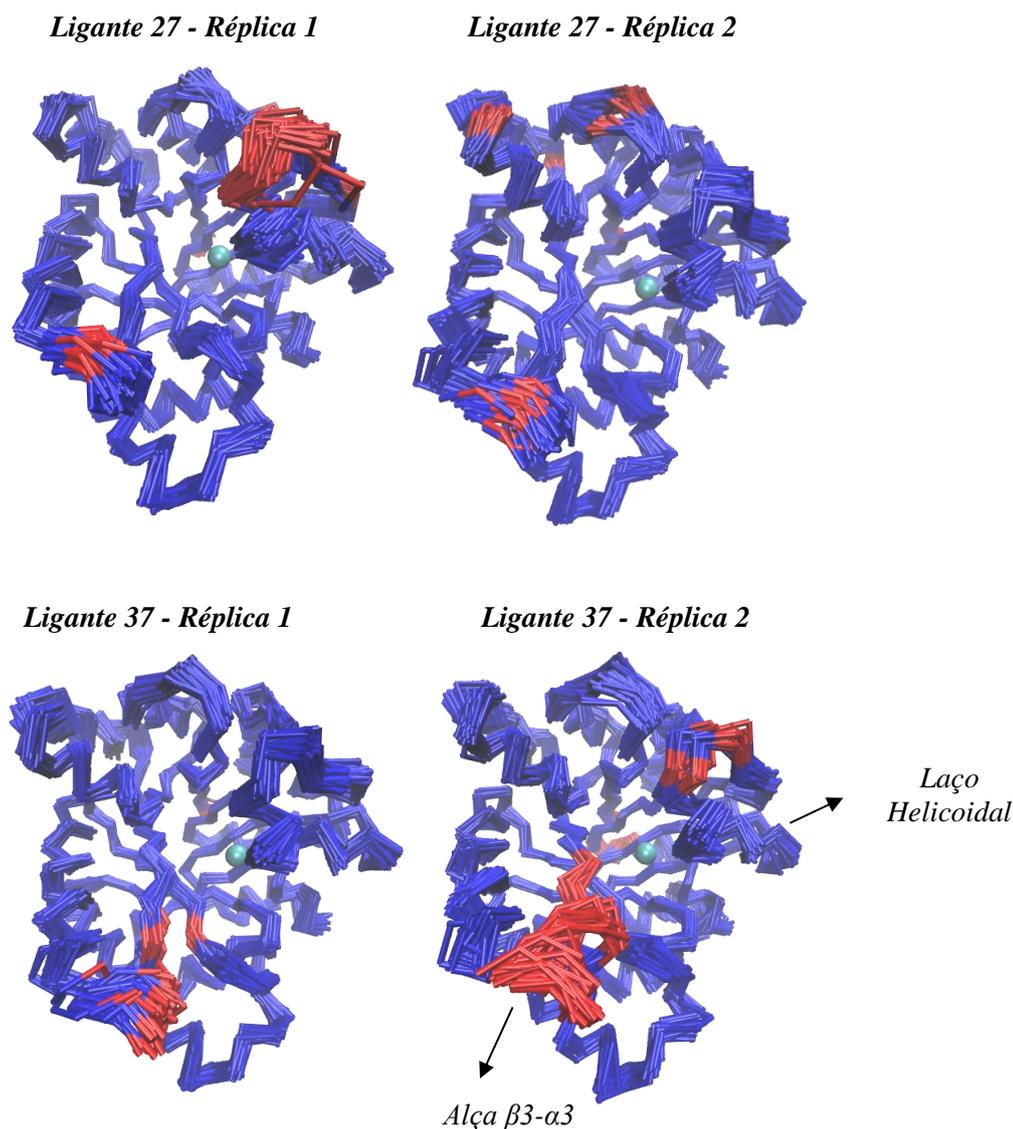


Figura 76 – Sobreposição de 100 frames ao longo das trajetórias geradas para a enzima em complexos com o ligante 27 (acima) e com o ligante 37 (abaixo). Em vermelho estão destacados os resíduos com RMSF dos carbonos- α superiores a 1 Å.

Essa constatação fica ainda mais evidente através da análise de RMSF dos resíduos de cada um desses sistemas, resultado que é mostrado na Figura 77, em que se pode observar que, à exceção das porções carbono e nitrogênio terminais, grande parte dos resíduos desses sistemas tem flutuações inferiores a 2,0 Å. Para o complexo com o ligante 27, em média nenhum resíduo apresenta RMSF inferior a 3 Å, ao passo em que para o ligante 37 apenas a alça β 3- α 3 ultrapassa esse marco.

Convém lembrar que nas simulações com ligante anteriores apenas o complexo enzima-produto apresentou resíduos com flutuação superior a 3 Å, nas demais, tanto da enzima com seu substrato quanto em complexo com o ligante 4 porções importantes como o laço helicoidal apresentaram valores de RMSF muito significativos, inclusive ultrapassando a marca de 4,0 Å com o ácido gálico.

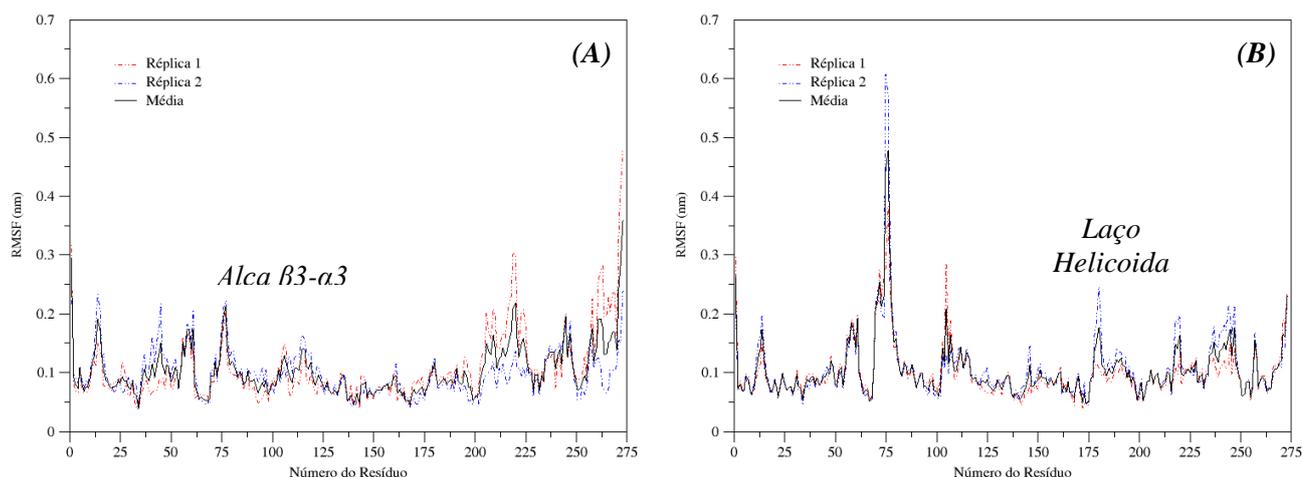


Figura 77 – Flutuação quadrática média dos resíduos ao longo das simulações da enzima em complexo com ligante 27 (A) e com ligante 37 (B) e respectivas médias.

Lembrando que a estrutura cristalográfica, ponto de partida dessas simulações, é o complexo enzima-produto, os baixos valores de RMSF dos resíduos nos complexos com os ligantes 27 e com o 37 apenas mostram que o volume diminuto desses compostos não produz alterações estruturais significativas no sítio ativo, da mesma maneira que o produto não induz mudanças uma vez que a estrutura cristalográfica já está adaptada para o receber.

Esse resultado é um forte indicativo de que as interações (ligações de hidrogênio, interações hidrofóbicas, etc.) desses compostos com o receptor enzimático não são as verdadeiras responsáveis pelo elevado desvio observado entre os escores de *docking* e as energias livres experimentais destes, mas sim é provável que o *docking* não esteja penalizando o impedimento estéreo do sítio na mesma intensidade que faz com os demais compostos, o que na prática subestima as energias de interação dos demais ligantes e produz a baixa correlação entre os dados observada.

Especificamente para os resíduos do sítio ativo (Apêndice D e Apêndice E), observou-se um RMSF médio de 0.0736 nm para o complexo com ligante 27 e de 0.0598 nm para o complexo com o ligante 37, valores inferiores ao registrado tanto para o produto (0.0740 nm) quanto para o ácido gálico (0.0795 nm). Para o ligante 37, inclusive, o RMSF médio do sítio ativo foi inferior ao observado para o próprio substrato (0.0667 nm).

No caso do substrato, ainda que o RMSF médio dos resíduos do sítio ativo seja pequeno – graças a interações fortes e efetivas – porções da enzima próximas ganharam mobilidade induzidas pelo maior volume do substrato, enquanto que nessas simulações, a presença do ligante teve pouco ou nenhum efeito tanto no sítio ativo quando na estrutura de uma forma global.

Em relação às ligações de hidrogênio (Figura 78), pode-se observar que ocorrem, em média, 1,28 ligações de hidrogênio entre a enzima e o ligante 27 e 5.68 ligações simultâneas com o ligante 37.

O número máximo de ligações simultâneas foi observado em 7620 ps da réplica 2 para o ligante 27 (5 ligações simultâneas) e em 61510 ps também na segunda réplica para o ligante 37, quando houve a formação de 9 ligações de hidrogênio simultâneas.

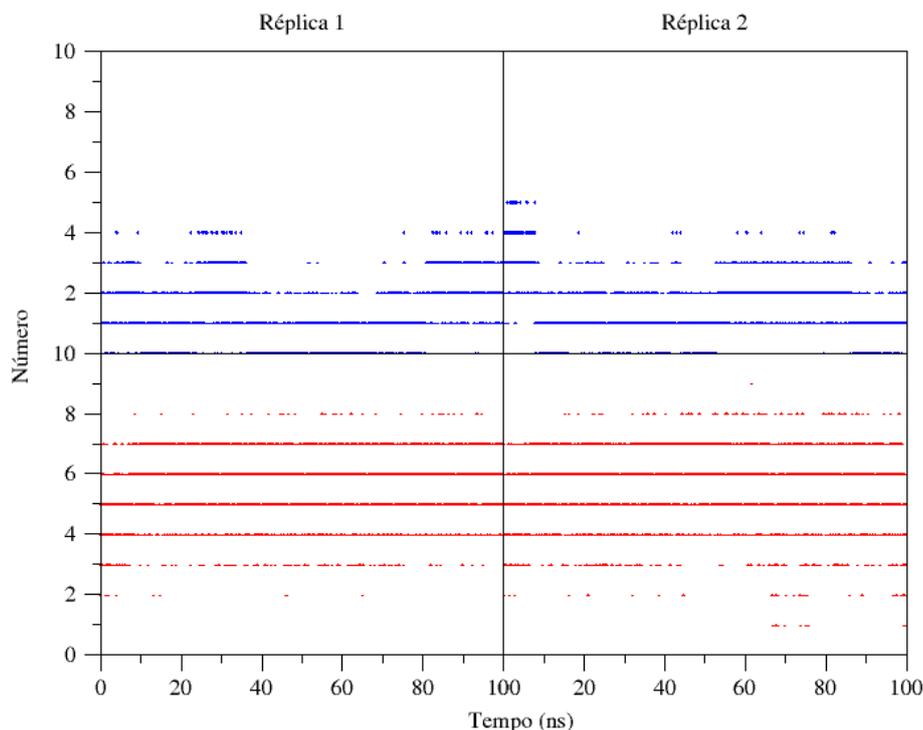


Figura 78 – Evolução temporal do número de ligações de hidrogênio simultâneas para cada réplica da simulação da enzima em complexo com ligante 27 (superior, em azul) e em complexo com ligante 37 (inferior, em vermelho).

Facilmente se percebe que o ligante 27, em razão do limitado número de grupos hidroxilas e do posicionamento inadequado de seus doadores e aceptores de ligação de hidrogênio é de fato incapaz de formar e manter um número significativo dessas interações. Enquanto o *docking* havia indicado a formação de 4 ligações de hidrogênio para esse resíduo, percebe-se que este forma, na prática, menos de duas dessas ligações, em média, evidenciando mais uma das formas com que o *docking* pode estar superestimando a energia de ligação desse composto.

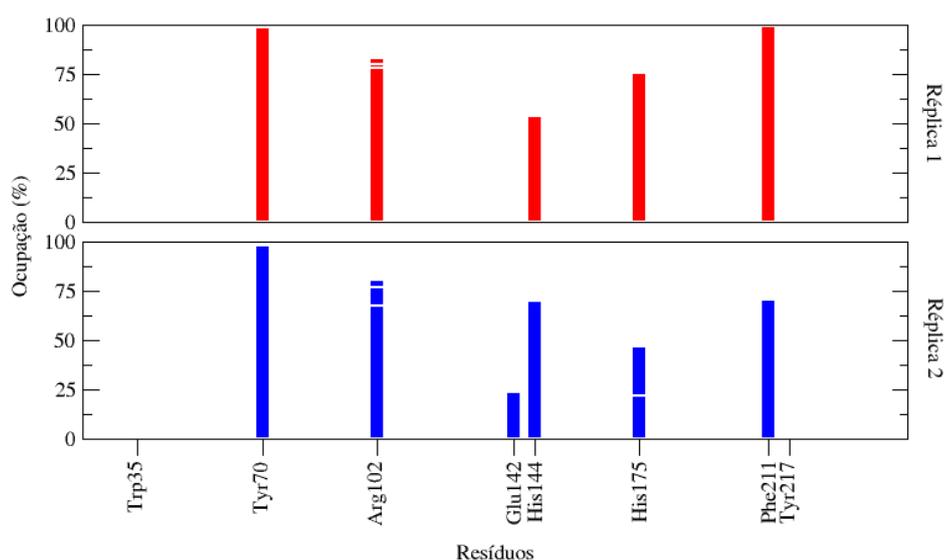


Figura 79 – Ocupação dos resíduos em ligações de hidrogênio para as duas réplicas da simulação da enzima em complexo com o ligante 37.

Por outro lado, o ligante 37 forma entre 5 e 6 ligações de hidrogênio, em média, o que estaria de acordo com o apontado por *docking*, não fosse a identidade dessas ligações, que sofre uma alteração significativa ao longo da dinâmica, como pode ser observado no padrão de ocupação dos resíduos em ligações de hidrogênio para esse sistema (Figura 79).

Em primeiro lugar, chama à atenção que o ligante 37 não forma, em nenhuma das reproduções, ligações de hidrogênio com a Tyr217, ao contrário do observado para tanto para substrato e produto quanto para o ácido gálico e contrariando, inclusive seus resultados *docking*.

É interessante notar também que os resíduos com as maiores ocupações foram a Tyr70 e Phe211, os quais foram apontados para os demais ligantes como tendo um papel secundário na interação com o sítio ativo. O *docking* inicial considerou em seu lugar a interação com a Tyr217 e com a His175.

Outro ponto importante e que merece destaque é a redução da ocupação da His175, que para os demais inibidores nunca esteve abaixo de 90%, mas que para o ligante 37 chega a ser reduzida para menos de 50%, como ocorre na réplica 2.

Como discutido anteriormente, a His175 em conjunto com a Arg102 atua como as principais âncoras dos ligantes no sítio ativo, sendo imprescindíveis na inibição enzimática. Portanto, ao mostrar que esse ligante não é eficiente na manutenção simultânea dessas ligações, os resultados de dinâmica molecular reforçam que o *docking* superestimou a contribuição dessas interações para a energia de ligação desse composto.

A ineficiência desse composto em se manter fixo e plenamente ancorado no sítio ativo fica ainda mais evidente quando se observa a evolução temporal da ocupação em ligações de hidrogênio (Figura 80), aqui mostrada para a réplica 2.

Pode-se perceber que à exceção da Tyr70, que permanece interagindo com o grupo carboxilato do ligante ao longo de toda a simulação, os demais resíduos sofrem alteração em sua ocupação, desde uma diminuição à troca dos átomos interagentes. Por exemplo, há inicialmente a formação de ligações de hidrogênio com a His175 e a Phe211, porém, a partir de 65 ns, a ocupação da Phe211 é drasticamente reduzida e acompanhada de um aumento abrupto da ocupação do Glu142. Da mesma maneira, há nesse mesmo momento uma troca entre os grupos do ligante que estão interagindo com a His175, mostrando que o ligante sofre uma leve rotação dentro do sítio ativo, o que evidencia a fraqueza de suas interações.

O glutamato-142, assim como o Glu253, compõe o sítio de complexação do cofator magnésio. O fato da dinâmica mostrar uma interação com a Glu142 enquanto o *docking* mostrou a interação com a Glu253 mostra que houve um leve reposicionamento dos resíduos ao longo da dinâmica, possivelmente induzido pela presença e pelas interações do ligante.

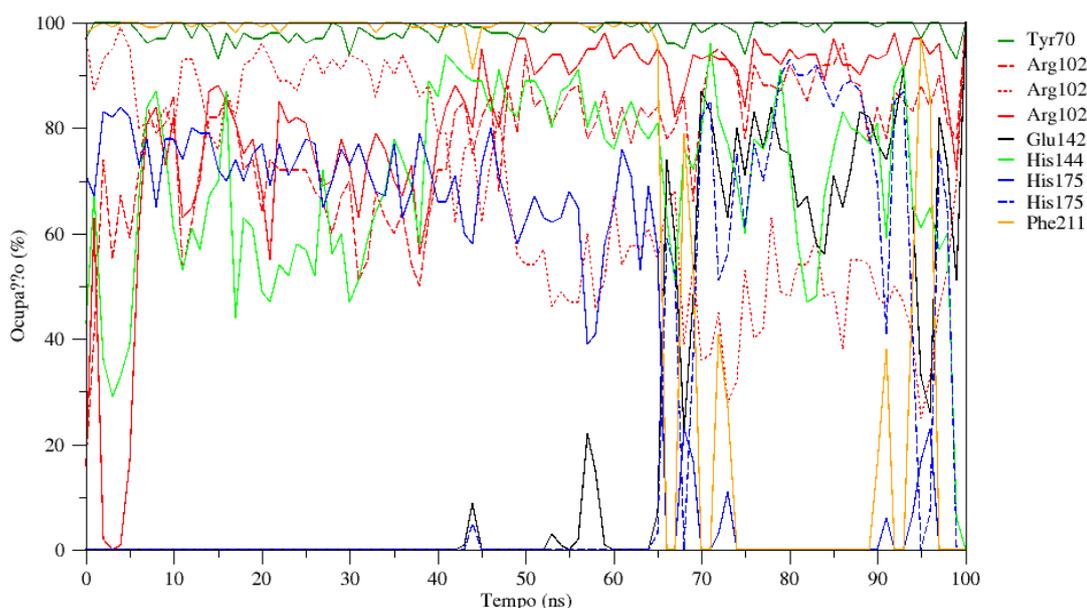


Figura 80 – Evolução temporal da ocupação dos resíduos em ligações de hidrogênio para a réplica 2 do complexo enzima-ligante37.

Quando se analisa o padrão de ligações de hidrogênio nos instantes finais das trajetórias desse complexo (Figura 81) percebe-se como o ligante é capaz de se movimentar dentro do sítio ativo, produzindo padrões de ligação significativamente diferentes entre os dois frames.

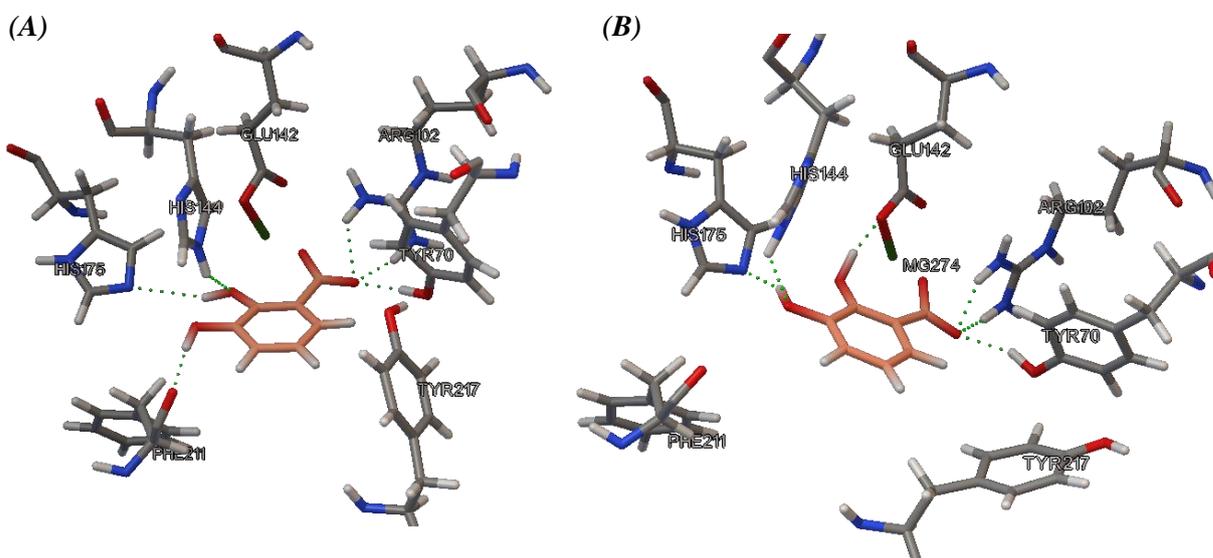


Figura 81- Ligações de hidrogênio formadas entre o ligante 37 e os resíduos do sítio ativo no último frame (100 ns) da réplica 1 (A) e da réplica 2 (B) desse sistema. As ligações de hidrogênio estão pontilhadas em verde

Chama a atenção que, além de nunca haver a interação com a Tyr217 – ao contrário do apontado por *docking* – para maximizar o número de interações das hidroxilas a porção carboxílica do ligante acaba interagindo por apenas um de seus oxigênios e não com ambos, como ocorre com os ligantes mais potentes, o que prejudica a fixação do ligante, permitindo essa movimentação ao longo do sítio ativo e que também evidencia como esse composto se liga de

forma menos efetiva ao sítio ativo enzimático e justifica a sua baixa potência inibitória experimental.

Para o ligante 27, observa-se que além do número de ligações de hidrogênio apontado pelo *docking* na estrutura cristalográfica (4 interações) ser um evento muito raro (cf. Apêndice F), a porção carboxílica do ligante – um ponto de fixação primordial no ancoramento dos inibidores no sítio ativo – raramente interage por ligações de hidrogênio, como pode ser constatado na Figura 82, em que estão mostrados o padrão de ligações de hidrogênio para o último frame, de 100 ns, de cada réplica desse sistema.

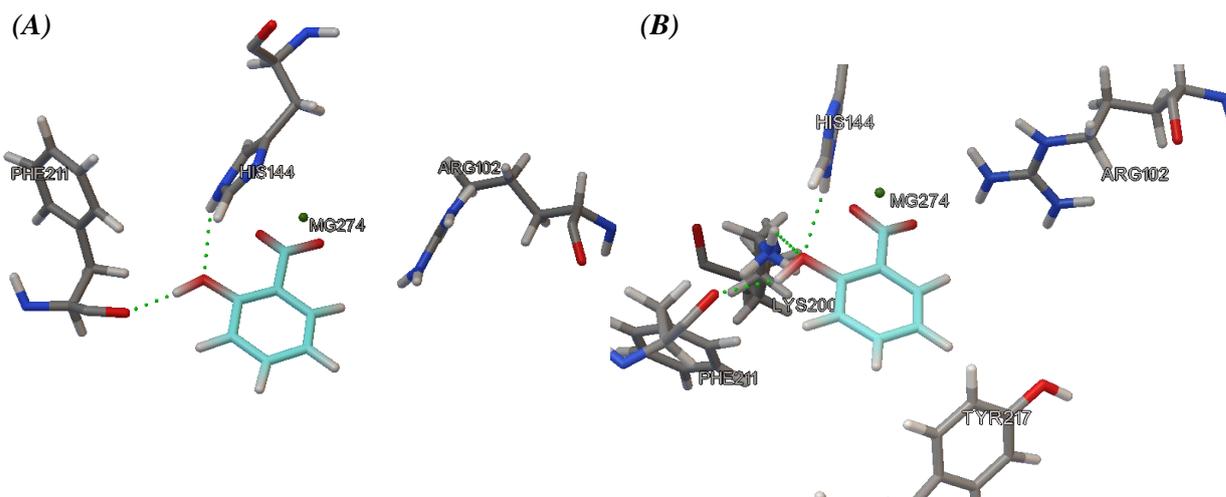


Figura 82 - Ligações de hidrogênio entre o ligante 27 e o sítio ativo enzimático no último frame (100 ns) da réplica 1 (A) e da réplica 2 (B) desse sistema. As ligações de hidrogênio estão pontilhadas em verde.

Pode-se observar que há geralmente duas ligações formadas, uma com His144 e uma com a porção peptídica da Phe211, essa última que não estava presente no *docking*. Em seu lugar, o *docking* considerou a formação de uma ligação com a His175, a qual não ocorre em nenhum momento das trajetórias de dinâmica, como pode ser constatado no gráfico da ocupação dos resíduos em ligações de hidrogênio, mostrado no.

A evolução temporal da ocupação dos resíduos em ligações de hidrogênio (Figura 83) mostra que, em alguns breves momentos há a interação do grupo carboxílico com a Arg102 para formação de ligações de hidrogênio. Porém, percebe-se também que a fixação ineficiente desse composto provoca frequentemente o rompimento não apenas desta, mas de todas as ligações de hidrogênio entre a enzima e o ligante, mostrando como o mesmo interage fracamente e evidenciando como o modo de ligação apontado por *docking* não condiz com a condição mais natural ou pelo menos mais frequente para essa interação receptor-ligante, justificando o elevado desvio observado entre os dados de *docking* e os resultados experimentais.

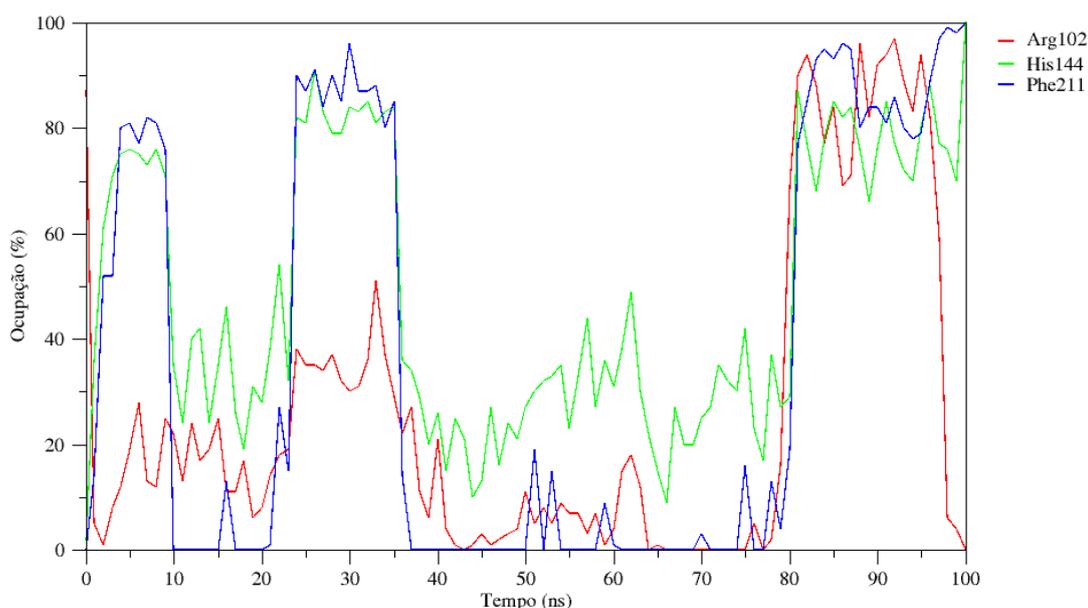


Figura 83 – Temporal da ocupação dos resíduos em ligações de hidrogênio com o ligante 27 na réplica 1 desse sistema.

Outro ponto que deve ser destacado, não apenas para o ligante 27 como também para o 37, é que a torção do grupo carboxílico em relação ao plano do anel aromático foi significativamente reduzida na dinâmica, sendo uma das principais razões pelas quais o número de ligações de hidrogênio entre o carboxilato dos ligantes e a enzima foi diminuído.

Esse resultado mostra como a conformação e os modos de ligação apontados por *docking* para esses ligantes não correspondem a uma condição realística, evidenciando, juntamente com todas as razões anteriormente descritas, como a metodologia de *docking* superestimou o número e a intensidade das interações da enzima com esses ligantes, o que ao final se traduz em um elevado desvio entre as energias livres de ligação desses compostos, inibidores ineficientes, e os escores de *docking*.

5.2.6 Enzima em complexo com ligantes 13 e 46

À exceção dos ligantes até aqui avaliados, os ligantes 13 (ácido 7-metóxi-2-benzofuranocarboxílico) e 46 (ácido indol-2-carboxílico) foram os únicos compostos relativamente potentes segundo os dados experimentais que não interagiram com o sítio ativo segundo o *docking*.

Ainda que seja provável que essa dificuldade de interagir com o sítio ativo esteja relacionada ao volume relativamente maior desses compostos, que em função dos impedimentos estéreos provoca uma penalização grande para os modos de ligação no sítio ativo, não se pode descartar *à priori* que possa haver uma interação de natureza alostérica, em que esses ligantes se ligam à região de entrada do sítio, como discutido na seção anterior. Essa possibilidade é

reforçada pelo fato dos ligantes 13 e 46 serem também os mais distintos quimicamente dos compostos de que coordenam ao sítio ativo.

Dessa forma, realizaram-se simulações de dinâmica molecular desses compostos partindo-se do modo de ligação “alostérico” apontado pelo *docking*. A manutenção dos ligantes coordenados nessa região ao longo da dinâmica pode então comprovar que há interações significativamente fortes atuando sobre os mesmos, sugerindo um sítio de ligação alostérico, ao passo que seu desligamento e migração reforçaria a hipótese de que um sítio ativo da estrutura cristalográfica é demasiadamente restrito.

O primeiro ponto a ser destacado é que, ao contrário das simulações que partem de ligantes no sítio ativo, o RMSD médio dessas simulações foi muito mais elevado (Figura 84), sendo 0,1312 nm para o complexo com ligante 13 e de 0.1560 nm para o complexo com ligante 46, valores comparáveis aos da enzima livre, e no caso do ligante 46, inclusive, superiores a essa, que foi de 0,1323 nm.

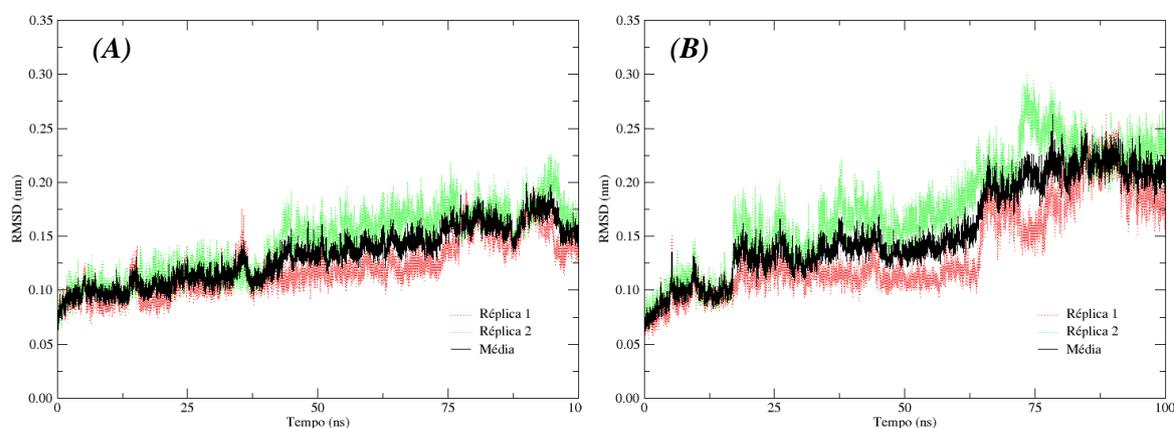


Figura 84 – Evolução temporal do RMSD dos carbonos- α em relação à estrutura inicial para as duas réplicas da simulação da enzima em complexo com o ligante 13 (figura A) e em complexo com o ligante 46 (figura B) e respectivas médias.

Pode-se observar que há, para o ligante 13, um aumento constante do RMSD até aproximadamente 75 ns, a partir de quando uma certa estabilização é atingida, porém num padrão oscilatório, evidenciando um sistema mais flexível do que observado nas simulações anteriores.

Para o complexo com o ligante 46, observa-se uma evolução do RMSD um pouco mais drástica. Nos primeiros 5 ns há um rápido incremento do desvio quadrático médio, que ligeiramente se estabiliza, porém em torno de 20 ns há um salto abrupto seguido por uma nova estabilização e o alcance de um patamar centrado em 0,14 nm, o qual se mantém relativamente estável.

Entretanto, em torno de 60 ns ocorre um novo incremento, a partir do qual o patamar final é atingido, se mantendo estável em torno de 0,22 nm até o final da trajetória. Na verdade, esses saltos mostram a ocorrência de dois eventos importantes, o primeiro em 17 ns da réplica 2 e o segundo em 64 ns da réplica 1, que impactam a estabilidade global da enzima.

O incremento de mobilidade da enzima nesses sistemas, na comparação com as simulações com os ligantes no sítio ativo pode ser comprovado por meio do RMSD acumulado em função da fração de carbonos- α considerada (Figura 85), em que se observa que, em média

87% dos átomos do sistema enzima-ligante 13 e 88% dos do sistema enzima-ligante 46 podem ser alinhados com menos de 1,0 Å de desvio quadrático médio, enquanto que para as simulações anteriores esses valores foram sempre superiores a 95%.

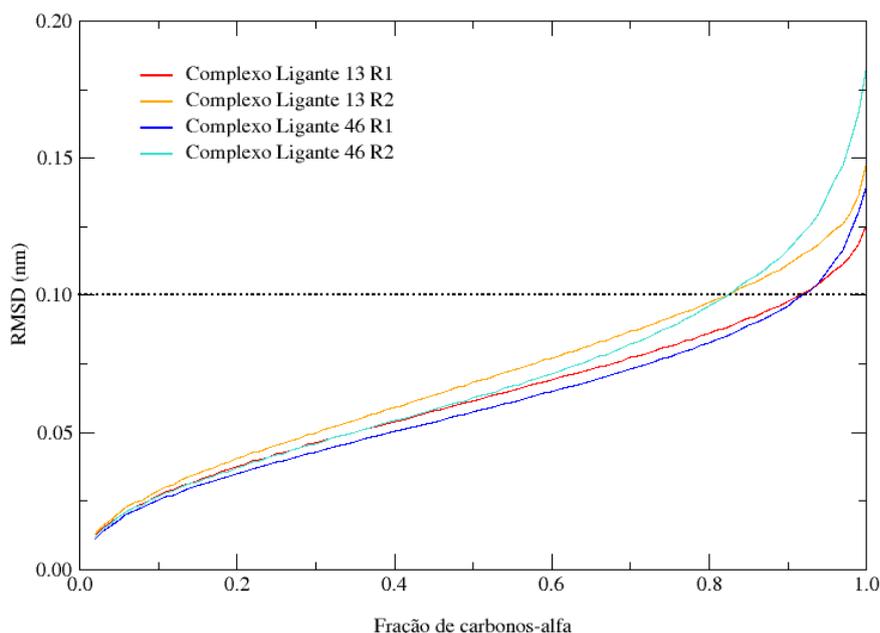


Figura 85 – RMSD acumulado em função da fração de carbonos- α consideradas na análise para cada réplica das simulações da enzima em complexo com os ligantes 13 e 46.

Chama a atenção, no entanto, que esses 12 a 13% dos resíduos que são responsáveis pela maior parte do RMSD se encontram principalmente no laço helicoidal e em outras regiões próximas do sítio ativo, como pode ser constatado na sobreposição de 100 frames ao longo de cada uma dessas trajetórias (Figura 86) e na flutuação quadrática média de todos os resíduos ao longo das simulações, mostradas no Apêndice G.

Ainda que haja, como sempre, valores mais elevados de mobilidade para regiões de alças, como as alças $\beta 2-\alpha 2$ e $\beta 4-\alpha 4$, a região próxima do sítio ativo e em especial o laço helicoidal apresentou altíssima mobilidade em todas as réplicas e em ambos os sistemas, em especial no complexo com o ligante 46, em que essa região chega a apresentar um RMSF superior a 0,7 nm, superior aos observados para a enzima livre.

Por outro lado, ao contrário das simulações da enzima livre, em que diversas regiões da enzima apresentam alta mobilidade, incluindo algumas α -hélices e folhas- β , como hélice $\alpha 2$, nas simulações dos ligantes 13 e 46 o incremento de mobilidade ocorre em regiões bem mais específicas, mostrando que apesar desses compostos não estarem auxiliando na manutenção da estabilidade estrutural, como fazem os ligantes anteriormente estudados, também não estão permitindo a evolução natural do sistema, portanto há alguma interação com enzima.

Uma das hipóteses levantadas inicialmente foi a da existência de um sítio de ligação alostérico na região que corresponde à entrada do sítio ativo, como apontado por *docking*. Tomando-se como referência o cofator magnésio, que em função da complexação com os glutamatos 142 e 253 se mantém praticamente imóvel ao longo das simulações, calculou-se a distância entre os ligantes e o Mg^{2+} , resultado que está mostrado na Figura 87, com o intuito de se testar essa hipótese.

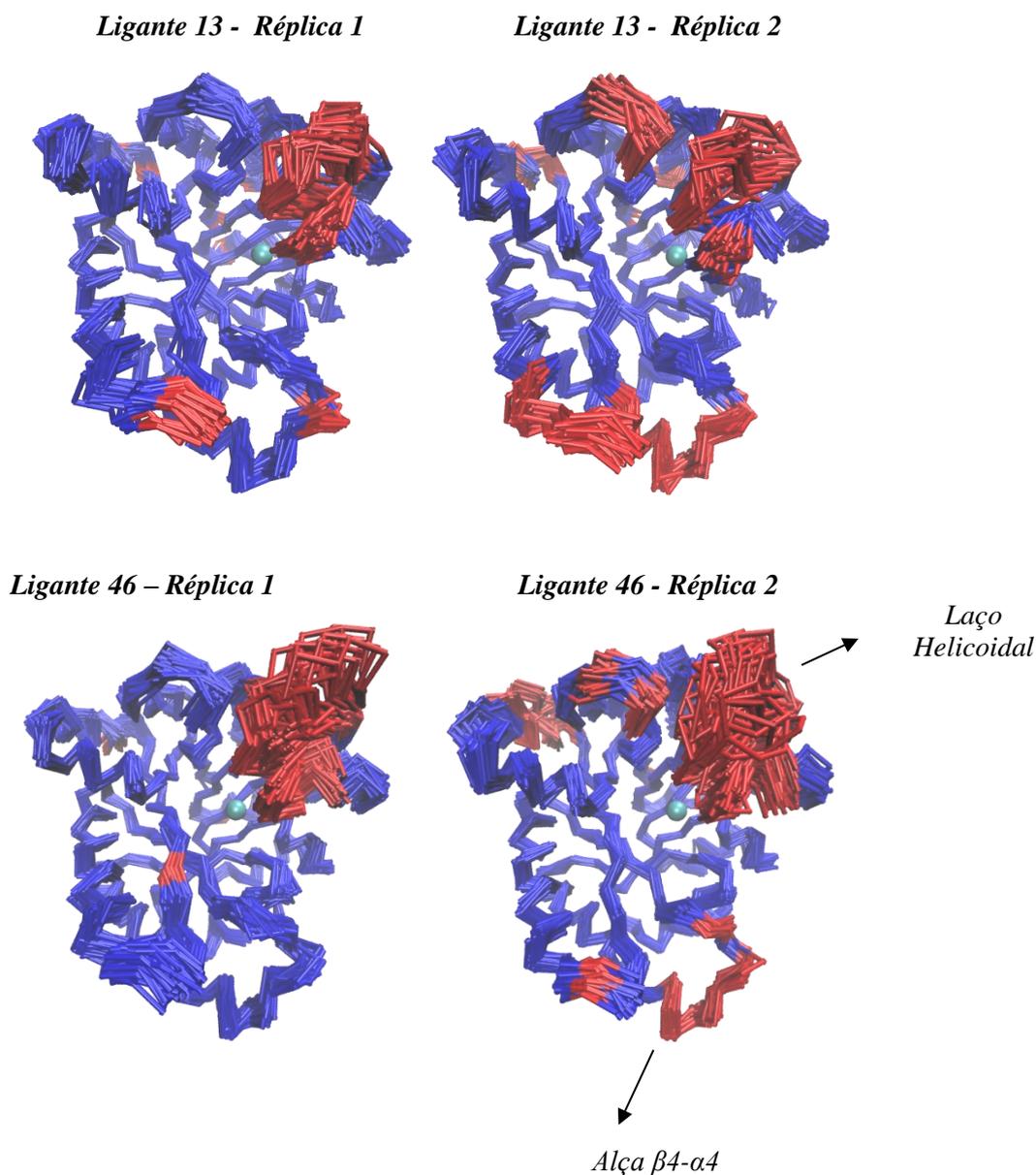


Figura 86 - Sobreposição de 100 frames, em cada nanossegundos de trajetória, para os sistemas enzima em complexo com ligante 13 (acima) e enzima em complexo com ligante 46 (abaixo). Em vermelho estão destacados os resíduos com RMSD superior a 1,0 Å.

Pode-se observar para o ligante 13 (Figura 87A), que os sistemas partem de uma distância inicial de 1,2 nm, que chega a diminuir para menos de 1 nm para a réplica 1 até os 37 ns, quando ocorre o afastamento abrupto do ligante para uma distância praticamente invariável em torno de 1,6 nm. Para a réplica 2, há uma variabilidade muito maior e mais abrupta dessa distância, que oscila entre 0,9 e 2,5 nm ao longo da trajetória, mas que ao final aparenta se estabilizar à mesma distância que o observado ao final de réplica 1.

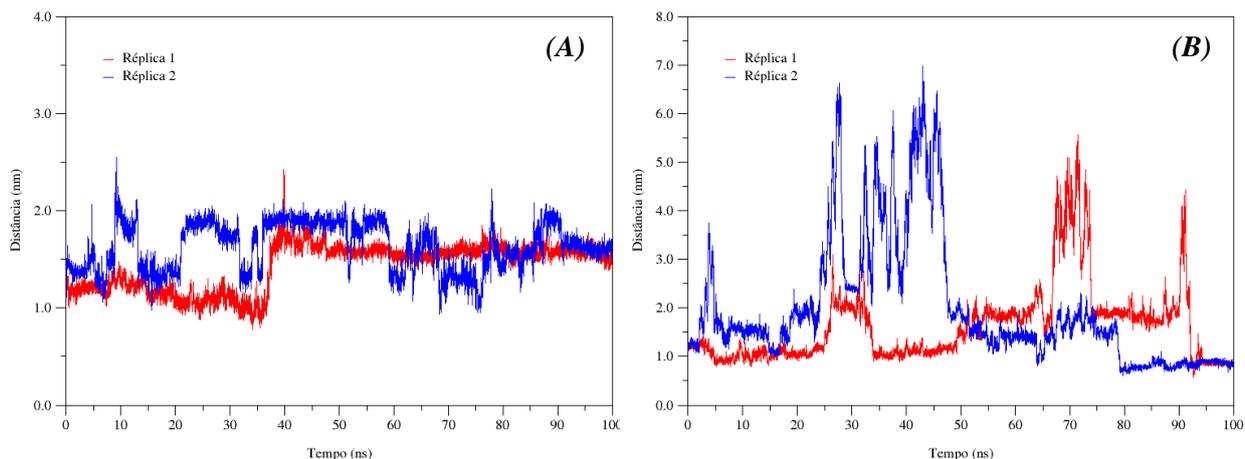


Figura 87 - Distância entre o cofator Magnésio e os ligantes 13 (figura A) e 46 (figura B) ao longo das trajetórias para cada réplica das simulações.

Apesar de que nas duas reproduções o ligante 13 terminar as duas reproduções à mesma distância do cofator Magnésio, não se encontra na mesma posição, como pode ser constatado na Figura 88, na qual estão sobrepostas as estruturas do ligante no início e ao final de cada trajetória. Na réplica 1 a estrutura final se encontra muito distante do modo de ligação do *docking*, ao passo em que para a segunda reprodução há uma maior proximidade, porém, adotando uma orientação significativamente diferente da inicial.

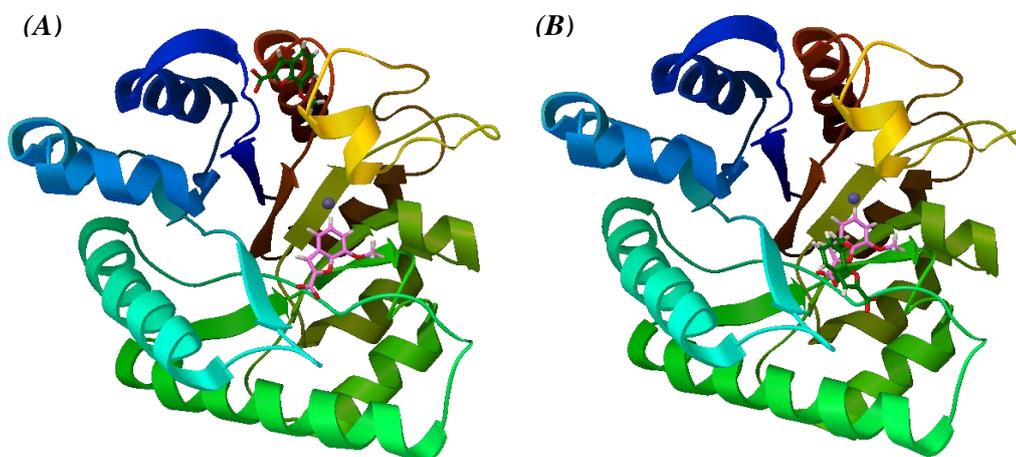


Figura 88 - Sobreposição das estruturas inicial (em rosa) e após 100 ns de simulação (em verde) do ligante 13 para a réplicas 1 (figura A) e a réplica 2 (figura B).

Para o ligante 46 (Figura 87B) há um comportamento ainda mais distinto entre as réplicas, que também partem de uma distância inicial de 1,2 nm.

Na réplica 1 ocorrem inicialmente uma aproximação do ligante para menos de 1,0 nm durante os 25 ns iniciais da dinâmica, seguido de afastamento e reaproximações até que em 67 ns ocorre um distanciamento drástico para quase 6 nm de distância do cofator Mg, o que é formalmente fora da enzima. É essa saída rápida do ligante a responsável pelo aumento do RMSD dos carbonos- α observado nesse mesmo instante. Supreendentemente, entretanto, há o retorno do ligante, que chega ao final da trajetória a uma distância estável de 0,9 nm.

Na réplica 2, por outro lado, há o afastamento do ligantes desde os primeiros nanossegundos, seguida de uma breve reaproximação em torno de 15 ns, até que a partir de 17 ns – mesmo tempo em que há um salto no RMSD dos carbonos- α – o ligante começa a se afastar completamente, atingindo distâncias superiores a 7 nm a partir de 25 ns, tendo nesse momento migrado para fora da enzima.

De forma análoga à replica 1 e igualmente surpreendente, há o retorno o ligante a partir da metade final de trajetória, se mantendo à mesma distância em que se encontrava no início, com pequenas oscilações em torno de 1,0 nm até os 20 ns finais, quando ocorre uma aproximação ainda maior do ligante que chega a menos de 0,6 nm do magnésio em alguns momentos, terminando a trajetória a uma distância de 0,9 nm tal qual observado para réplica 1.

Quando se observa o último frame de cada réplica (Figura 89) em comparação com a estrutura inicial, percebe-se que o ligante 46 termina as duas trajetórias praticamente na mesma posição, ainda que elas tenham tido evoluções temporais completamente diferentes.

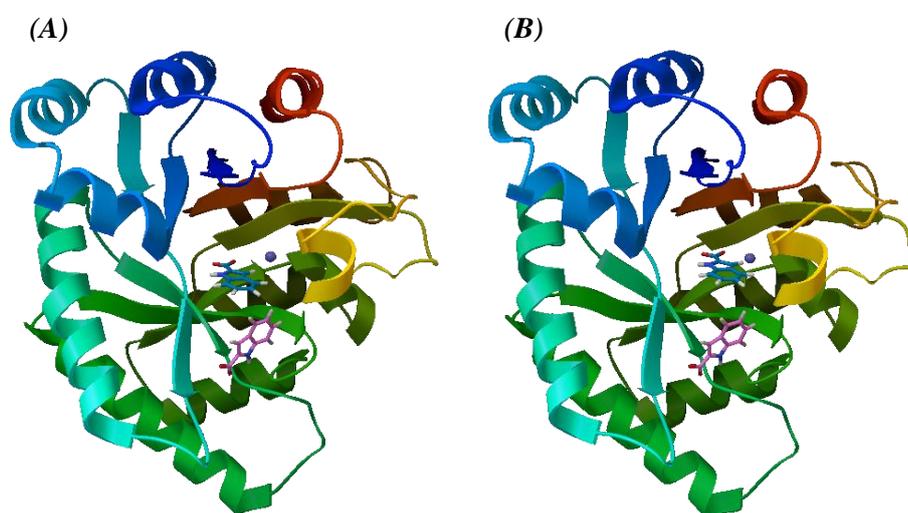


Figura 89- Sobreposição das estruturas do ligante 46 no tempo inicial (em rosa) e ao final de 100 ns de simulação (em azul) para a primeira (figura A) e a segunda réplica do sistema (figura B).

Esse modo final de ligação não corresponde mais à entrada do sítio, mas sim houve a entrada do ligante, sua aproximação do magnésio e inclusive a interação com resíduos importantes do sítio ativo, como pode ser observado no gráfico da ocupação dos resíduos em ligações de hidrogênio (Figura 90), em que se observa que ao final das duas reproduções do sistema começa a haver a formação de uma ligação de hidrogênio entre o inibidor 46 e a Arg102, o principal resíduo responsável pelo ancoramento da porção carboxílica dos ligantes no sítio ativo. Além disso, é interessante observar o aumento da ocupação da Arg102 em ligações de hidrogênio ao longo das simulações, atingindo 100% ao final das mesmas.

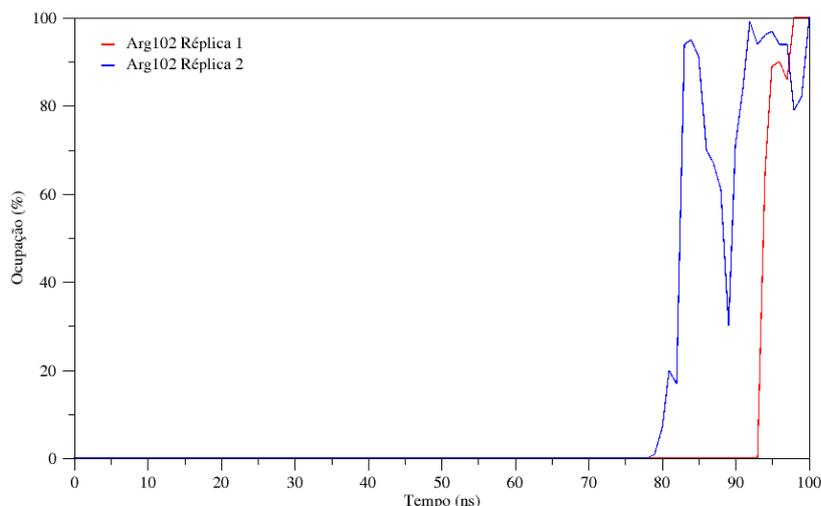


Figura 90 – Evolução temporal da ocupação dos resíduos em ligações de hidrogênio com o ligante 46, na primeira (vermelho) e segunda reprodução (azul).

Observando mais detalhadamente a orientação do ligante ao final da trajetória da réplica 2 (Figura 91) percebe-se que ele chega a formar duas ligações de hidrogênio simultâneas com a Arg102, além de uma ligação com a Ser68. Porém, observa-se também que o composto não encontra-se ainda completamente inserido do sítio ativo, possivelmente em razão de não ter havido tempo suficiente de simulação para que o ligante pudesse entrar por completo.

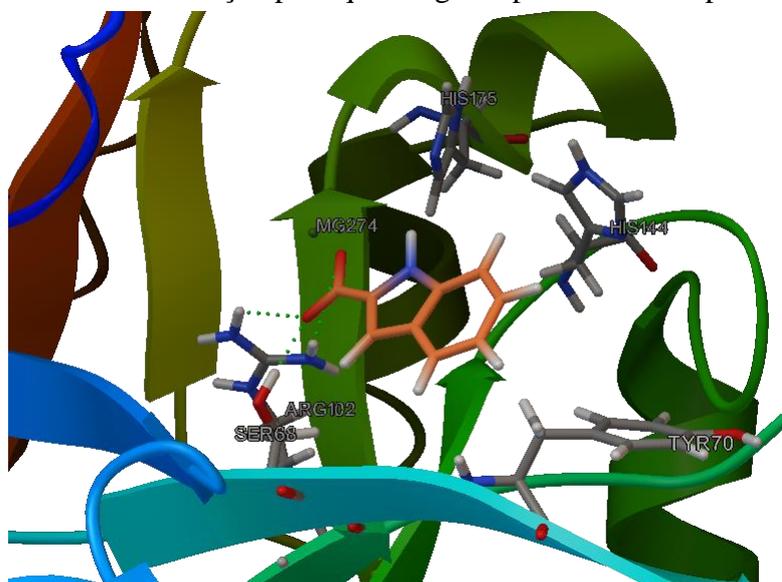


Figura 91 - Ligações de hidrogênio entre a enzima e o ligante 46 ao final de 100 ns de simulação, mostrando que o inibidor entra parcialmente no sítio ativo. As ligações de hidrogênio estão pontilhadas em verde e os principais resíduos do sítio ativo encontram-se destacados.

Apesar de ligações de hidrogênio contínuas e estáveis terem sido obtidas apenas ao final das simulações, em média ocorre a formação de 1,13 ligações de hidrogênio para o composto 13 e de 0,61 ligações para o ligante 46, atingindo inclusive valores como 5 e até 6 ligações simultâneas em alguns momentos das trajetórias, como pode ser constatado na Figura 92, em que está mostrada a evolução temporal do número de ligações de hidrogênio de cada ligante com a enzima ao longo de cada reprodução.

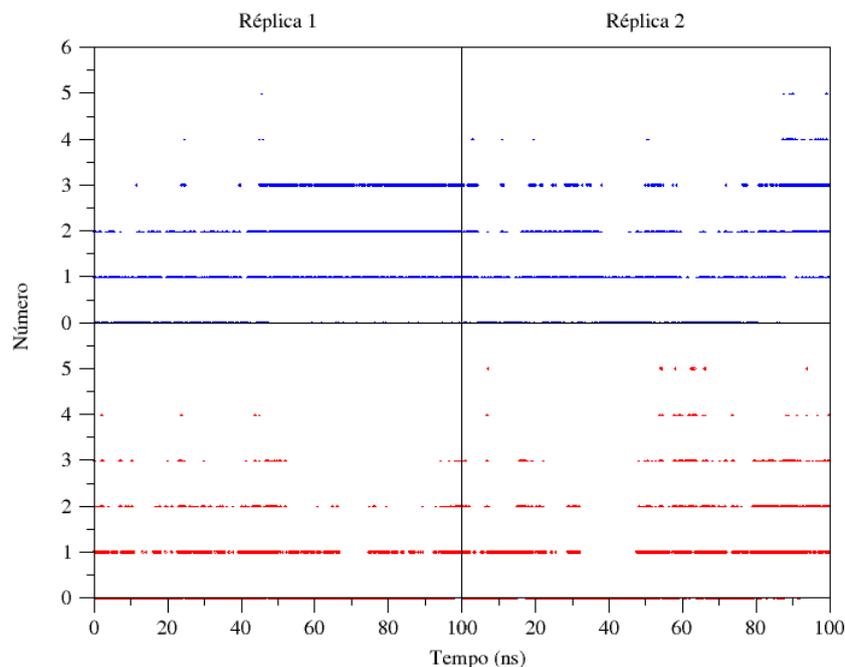


Figura 92 – Evolução temporal do número de ligações de hidrogênio entre o ligante 13 (azul) e ligante 46 (vermelho) e o receptor, para cada réplica realizada.

Isso mostra que mesmo durante suas derivas ao redor da enzima os ligantes interagem com o receptor, sendo os grandes responsáveis pelas alterações estruturais observadas em regiões específicas, como o laço helicoidal, sobretudo as que ocorram nas simulações com o ligante 46 e que culminaram em uma abertura do sítio ativo, permitindo que esse composto o adentrasse, ainda que não completamente em virtude do tempo de simulação insuficiente.

Esse resultado, em conjunto com as observações feitas para o ligante 13, permite afirmar que os modos de ligação apontado por *docking* para esses ligantes são artifícios criados pela dificuldade de os ancorar no sítio ativo devido aos seus elevados volumes, o que causa um grande impedimento estérico e acaba penalizando o modo de ligação natural.

Convém salientar que esses dois compostos são inibidores potentes experimentalmente, com valores de IC_{50} de $68,5 \mu\text{mol.L}^{-1}$ e $90,0 \mu\text{mol.L}^{-1}$ para os ligantes 13 e 46, respectivamente. Observar por meio da dinâmica molecular que o composto 46 tende a se dirigir espontaneamente para o sítio ativo, evidencia que há coerência entre os resultados teóricos de dinâmica e os dados experimentais. É muito provável que o ligante 13 também se dirigiria ao sítio ativo caso as simulações fossem suficientemente longas, pois, tal qual o ligante 46, sua presença também produziu mudanças conformacionais na enzima, sobretudo na região do sítio ativo.

Essas fortes alterações estruturais do receptor na presença desses compostos mostram a validade do modelo de encaixe induzido, que nesse caso se evidencia por meio da expansão e abertura do sítio ativo induzida pela aproximação dos ligantes, a qual, no limite de simulações suficientemente longas, propicia sua interação de maneira apropriada no sítio ativo.

Isso mostra o quanto a estrutura cristalográfica é limitada para racionalizar por meio de *docking* os modos, padrões de interação e principalmente as energias de ligação dos diversos compostos estudados e evidencia como a flexibilidade do receptor necessita ser levada em conta

para obtenção de um modelo que permita uma melhor correlação entre os resultados e sobretudo permita a predição minimamente confiável do potencial de inibição *in vitro* de novos compostos a partir de resultados puramente teóricos, o grande objetivo desse trabalho e a motivação do próximo passo, que é o *ensemble docking* a partir das diversas estruturas da enzima obtidas em todas essas simulações de dinâmica molecular.

5.3 ENSEMBLE DOCKING

A partir das trajetórias das simulações de dinâmica molecular, diversas estruturas do receptor foram selecionadas e gerou-se um conjunto (ou *ensemble*) dessas estruturas para novos *dockings*.

Idealmente, essas estruturas abrangeriam diferentes conformações que a enzima adota em solução e sobretudo conformações que adota na presença dos ligantes, os quais podem tornar o receptor mais propício à interação na medida em que cada ligante produziria um espectro de conformações ativas que são por ele estabilizadas (seleção conformacional) ou então induziria conformações do receptor nunca seriam encontradas na sua ausência (encaixe induzido).

Dessa forma, a ideia do principal ao se realizar o *ensemble docking* é selecionar, dentro dessas inúmeras conformações da enzima obtidas das simulações, aquela ou aquelas que melhor encaixam aos ligantes e que corresponderiam aos estados biologicamente ativos. Para isso, se realizam *dockings* com essas múltiplas estruturas da proteína em busca daquelas que geram os melhores escores ou que melhor correlacionam os escores de *docking* com os resultados experimentais.

Conhecendo-se essas estruturas biologicamente ativas, é possível não apenas se obter um entendimento mais aprofundado do mecanismo de ligação – o que é imprescindível no desenho racional de inibidores mais eficientes - mas também utilizar o *docking* para estimar de forma rápida e relativamente confiável a potência inibitória de novos compostos, evitando assim o dispêndio de tempo e de recursos na síntese e no teste de moléculas que não seriam experimentalmente ativas, um processo denominado de triagem virtual.

Através das simulações de dinâmica molecular é possível se obter um grande número de estruturas distintas do receptor enzimático. Entretanto, realizar o *docking* dos 25 compostos experimentalmente ativos para cada uma das conformações possíveis demandaria não apenas muito tempo, mas também um volume gigantesco de recursos computacionais, sobretudo de memória de armazenamento.

Além disso, é provável que a maioria dessas estruturas apresentaria pouca diferença entre si, tendo um impacto negligenciável sobre os resultados de *docking* quando comparado com a estrutura inicial, isto porque as proteínas apresentam níveis de flexibilidade, que vão desde rearranjos de cadeias laterais até movimentações globais, e que ocorrem em uma grande variedade de escalas de tempo e de distâncias.

Assim, não sendo possível “docar” todas as conformações geradas por dinâmica molecular, é necessária uma estratégia para a seleção das conformações mais relevantes dentro das trajetórias, de forma que se possa incorporar a flexibilidade da proteína sem comprometer

sua eficiência do *docking* e sobretudo que possa incluir as estruturas mais relevantes biologicamente, que corresponderiam às conformações ativas.

Diante disso, diversas estratégias foram avaliadas na seleção dessas conformações e construção do *ensemble*. A primeira foi a extração do último frame de cada trajetória, o que corresponde ao tempo de 120 ns para as simulações da enzima livre e solução e de 100 ns para suas simulações em complexo com os ligantes.

Conforme detalhado na seção anterior, ao final das trajetórias os sistemas, em geral, atingem conformações de equilíbrio – o que pode ser comprovado por meio da análise temporal do RMSD, que atinge um patamar ao final das simulações, com pouca oscilação em torno de um valor médio – justificando a relevância de se incluir a última estrutura de cada trajetória no *ensemble*.

Na Tabela VII estão mostrados os parâmetros da correlação linear entre os resultados experimentais e os escores de *docking* obtidos com cada uma dessas estruturas finais. Os valores obtidos pelo *docking* na estrutura cristalográfica estão mostrados como referência.

Tabela VII – Coeficientes de correlação de Pearson, coeficiente de determinação (COD), inclinação da reta de correlação e erro quadrático médio (RMSE) entre as energias livres experimentais e os resultados de *docking* com cada estrutura final gerada.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima + Ligante 37 R2	-0.2126	0.0452	-0.168	0.547
Enzima Livre S2	-0.0679	0.0046	-0.025	0.256
Enzima + Ligante 46 R1	-0.0637	0.0041	-0.041	0.452
Enzima + Ligante 13 R1	-0.0622	0.0039	-0.036	0.408
Enzima + Ligante 13 R2	0.0135	0.0002	0.009	0.466
Enzima Livre S3	0.0252	0.0006	0.018	0.494
Enzima + Produto R1	0.0330	0.0011	0.016	0.344
Enzima + Ligante 37 R1	0.1285	0.0165	0.094	0.515
Enzima Livre S4	0.1363	0.0186	0.078	0.403
Enzima Livre S1	0.1458	0.0213	0.083	0.398
Enzima + Substrato R3	0.2152	0.0463	0.170	0.548
Enzima + Ligante 27 R1	0.2155	0.0465	0.142	0.459
Enzima + Produto R3	0.2653	0.0704	0.357	0.921
Enzima + Ligante 04 R3	0.2812	0.0790	0.189	0.458
Enzima + Substrato R1	0.2922	0.0854	0.308	0.716
Enzima + Ligante 04 R1	0.3180	0.1011	0.270	0.572
Enzima + Produto R2	0.3187	0.1016	0.298	0.630
Enzima + Ligante 04 R2	0.3411	0.1163	0.327	0.641
Enzima + Ligante 46 R2	0.3502	0.1226	0.268	0.509
Enzima Inicial	0.3599	0.1295	0.364	0.671
Enzima + Ligante 27 R2	0.3753	0.1409	0.296	0.519
Enzima + Substrato R2	0.4125	0.1702	0.318	0.498

Pode-se observar que, dentro do conjunto de 21 estruturas avaliado, apenas duas foram capazes de produzir resultados mais correlacionados com os dados experimentais do que a estrutura cristalográfica, a saber, a réplica 2 da simulação com o ligante 27 e a segunda réplica

da simulação com substrato. Esta última permitiu um incremento no R^2 de 0,13 para 0,17, porém às custas de uma diminuição na inclinação da reta de correlação, indicando que embora o modelo obtido consiga explicar melhor a relação existente entre dados, o faz comprometendo a precisão dos resultados frente aos valores experimentais.

O fato da maioria das estruturas avaliadas apresentar coeficiente de correlação de Pearson inferior ao da estrutura de partida revela que a escolha do último frame das trajetórias não é adequada para a construção do *ensemble*, pois mesmo que ao final das simulações os sistemas possam ser considerados estáveis, atingindo uma conformação global de equilíbrio – como demonstrado pela evolução temporal do RMSD dos carbonos- α , por exemplo – alterações estruturais instantâneas continuam ocorrendo, bem como há o reposicionamento de porções flexíveis e das próprias cadeias laterais as quais, *em média*, correspondem a uma estrutura equilibrada.

Dessa forma, o frame final da trajetória não corresponde necessariamente à estrutura final equilibrada da simulação, e esta, por sua vez, também pode não corresponder à estrutura biologicamente ativa, não sendo surpreendente que a seleção arbitrária desse ponto acarrete em baixos valores de correlação frente aos dados experimentais.

Apesar disso, se pode perceber alguns aspectos comuns entre as réplicas dos mesmos sistemas, os quais permitem inferir sobre o comportamento médio final da enzima em cada condição estudada. Um desses aspectos diz respeito às estruturas retiradas das simulações da enzima sem ligante, as quais estão dentre as 10 com os piores desempenhos, uma das quais produziu, inclusive, uma anti-correlação com os dados experimentais.

Isso mostra que na ausência de ligantes a enzima adota uma estrutura mais relaxada e mais aberta – como verificado por meio das simulações de dinâmica molecular – a qual é adequada para facilmente receber ligantes, porém não para interagir efetivamente com eles. Na presença do ligantes, entretanto, a correlação com os valores experimentais aumenta, indicando que a interação com tais compostos conserva ou induz o sítio de ligação em uma conformação adequada ao seu encaixe.

Dessa forma se pode imaginar que, livre em solução, a enzima adota uma conformação mais aberta justamente para permitir a entrada do substrato/ligantes na cavidade do sítio ativo. À medida em que estes entram e começam a interagir com o sítio, provocam nele alterações conformacionais que fortalecem as interações já formadas ao mesmo tempo em que possibilitam que novas interações sejam realizadas.

É isso que se observa quando ligantes como o ácido gálico, produto e especialmente o substrato estão presentes na estrutura, o que faz com que haja pouca alteração nos parâmetros de correlação quando comparados com a estrutura inicial ou até mesmo haja uma melhora nos mesmos, como ocorre na segunda réplica da simulação com substrato.

Frente à essa constatação, desenvolveu-se a segunda estratégia de seleção de estruturas, baseada explicitamente no modelo do encaixe induzido.

Esse modelo sugere que os ligantes induzem alterações conformacionais nas porções da enzima com a qual interagem, as quais podem ser transmitidas às subunidades vizinhas, levando à modificação de sua estrutura tridimensional. Essas modificações, de diferentes graus de amplitude e complexidade, induzem o receptor a uma conformação que favoreça o

reconhecimento e *maximize* as interações com os ligantes, as quais jamais seriam obtidas na ausência dos mesmos.

Assim, buscou-se nas trajetórias as conformações da enzima que maximizassem as interações e que, portanto, deveriam corresponder à melhor estrutura para encaixe de cada um dos ligantes simulados. Especificamente, buscou-se pelos frames em que houvesse a formação do maior número de ligações de hidrogênio simultâneas entre a enzima e cada um dos compostos estudados, não apenas por esse ser o tipo de interação receptor-ligante mais facilmente avaliada por meio das ferramentas de simulação de dinâmica molecular utilizadas, mas também porque a presença de grupos doadores e aceptores de ligação de hidrogênio aparenta ser essencial para a atuação dos compostos como inibidores, mostrando o papel primordial que essas interações detêm no mecanismo de inibição enzimática.

Na Tabela VIII estão mostrados os parâmetros de qualidade e a inclinação das curvas de correlação entre os resultados de *docking* com cada uma dessas estruturas e os dados experimentais.

Pode-se observar que, dentre as 17 estruturas geradas, praticamente a metade foi capaz de produzir resultados com melhor correlação com os dados experimentais do que a estrutura de partida, com destaque mais uma vez para as simulações com o substrato, para as quais se obteve incremento nos parâmetros de qualidade e na inclinação da curva de correlação para as três reproduções, e notadamente para a réplica 3, que produziu um R-Pearson de 0,47 e inclinação de 0,42. Portanto, o modelo obtido não apenas explica melhor os dados experimentais como também o faz com uma precisão melhor do que a obtida pelo *docking* na estrutura cristalográfica.

Tabela VIII – Coeficientes de correlação de Pearson, coeficiente de determinação (COD), inclinação da reta de correlação e erro quadrático médio (RMSE) entre as energias livres experimentais e os resultados de *docking* obtidos com a estruturas que maximizam o número de ligações de hidrogênio simultâneas enzima-ligante.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima + Ligante 13 R1	0.0210	0.0004	0.012	0.413
Enzima + Ligante 46 R1	0.0367	0.0014	0.024	0.457
Enzima + Ligante 27 R1	0.0791	0.0063	0.039	0.349
Enzima + Ligante 27 R2	0.1077	0.0116	0.119	0.782
Enzima + Produto R3	0.1546	0.0239	0.142	0.644
Enzima + Ligante 13 R2	0.1645	0.0271	0.109	0.462
Enzima + Produto R1	0.2924	0.0855	0.258	0.601
Enzima + Ligante 46 R2	0.3177	0.1009	0.191	0.405
Enzima + Ligante 04 R3	0.3496	0.1222	0.356	0.678
Enzima Original	0.3599	0.1295	0.364	0.671
Enzima + Ligante 37 R1	0.3645	0.1329	0.514	0.933
Enzima + Substrato R2	0.3654	0.1335	0.653	1.182
Enzima + Produto R2	0.3669	0.1346	0.281	0.506
Enzima + Ligante 04 R2	0.3769	0.1420	0.503	0.878
Enzima + Ligante 37 R2	0.3827	0.1464	0.444	0.763
Enzima + Substrato R1	0.3828	0.1466	0.429	0.735
Enzima + Ligante 04 R1	0.4186	0.1752	0.397	0.612
Enzima + Substrato R3	0.4708	0.2216	0.421	0.560

No sentido oposto, das estruturas que produziram os piores parâmetros de correlação, encontram-se aquelas que foram geradas nas simulações com os ligantes 13 e 46, que chegam a apresentar coeficiente de determinação próximo de zero, como observado na primeira réplica de cada um desses sistemas.

Convém lembrar que tais ligantes não puderam ser docados no sítio ativo da estrutura cristalográfica e que as simulações de dinâmica molecular partem dos mesmos ligados fora do sítio, havendo inclusive a saída de tais ligantes ao longo da simulação, os quais estão interagindo na superfície enzimática no momento em que ocorrem a formação do número máximo de ligações de hidrogênio simultâneas (Figura 93), o que justifica os baixos valores de correlação observados.

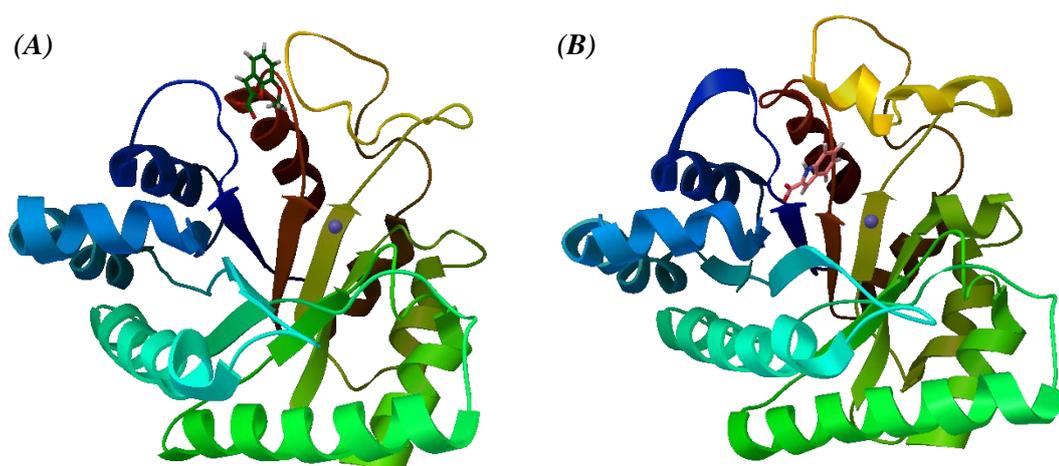


Figura 93- Complexos enzima-ligante 13 (figura A, com ligante em verde) e enzima-ligante 46 (figura B, com ligante em rosa) nos tempos de 45440 ps e 45000 ps, respectivamente, momento em que ocorre a formação do maior número de ligações de hidrogênio simultâneas com a enzima.

Apesar disso, no entanto, esses resultados reforçam como a presença dos ligantes no sítio ativo induz a adoção de conformações mais propícias seu encaixe e interação com a enzima, ao mesmo tempo em que reforça que na ausência destes, essas conformações mais ativas biologicamente não são alcançadas. Ao invés disso, quando os ligantes estão fora do sítio ou ausentes, se observa que há a abertura do mesmo, porém com a diminuição de sua capacidade de interação.

Por outro lado, essa abertura é o que efetivamente permite a entrada dos ligantes no sítio, o que foi de fato observado nas simulações com esses compostos (cf. Figura 87), em que há o retorno dos ligantes e sua entrada no sítio ativo nos instantes finais das simulações.

Provavelmente o tempo de simulação não foi suficiente para que esses ligantes se acomodassem plenamente no sítio ativo e pudessem ajustá-lo para maximizar suas interações, motivo pelo qual as estruturas finais das simulações também não apresentaram resultados bem correlacionados com os dados experimentais.

Estender as simulações desses compostos até que a estabilidade estrutural plena seja atingida pode comprovar essa hipótese e é uma das perspectivas futuras do presente estudo.

A terceira estratégia empregada na construção do *ensemble* de estruturas é baseada no modelo de seleção conformacional, segundo o qual proteínas em solução existem como um conjunto de conformações diferentes, que correspondem a mínimos em sua superfície de energia potencial e cujas populações são determinadas pela distribuição de Boltzmann com base na

diferença de energia entre esses mínimos. Nesse modelo, o ligante apresentaria uma afinidade maior por uma dessas conformações, estabilizando-a através de suas interações e fazendo com que sua frequência no conjunto aumente.

Dessa forma, gerou-se por meio do programa *g_cluster* do pacote Gromacs um conjunto de conformações distintas da enzima para cada trajetória de dinâmica molecular, agrupadas por meio do desvio quadrático médio das posições atômicas de todos os carbonos- α ou apenas dos resíduos do sítio ativo, a fim de se amostrar tanto alterações conformacionais globais quanto modificações específicas no sítio ativo.

Como as populações desses conjuntos de estruturas estão relacionadas com sua estabilidade energética e assumindo que a conformação nativa e as conformações biologicamente ativas são pontos de mínimo (globais ou locais) da superfície de energia potencial – sendo inclusive estabilizadas pela interação com os ligantes – dentre os inúmeros conjuntos gerados apenas os conjuntos mais populosos foram levados em consideração na composição do *docking*.

Ainda assim, 168 estruturas distintas foram selecionadas dentro das 21 trajetórias geradas, sendo 74 pela clusterização por RMSD dos carbonos- α e 94 pela clusterização por RMSD do sítio ativo. Nas tabelas Tabela IX e Tabela X estão mostrados os parâmetros de qualidade das curvas de correlação entre os resultados de *docking* e as energias livres de ligação experimentais, as quais correspondem à clusterização por RMSD dos carbonos- α e do sítio ativo, respectivamente. Apenas os sistemas com coeficiente de correlação de Pearson superior ao obtido pelo *docking* na estrutura cristalográfica são mostrados nas tabelas. Os resultados completos podem ser observados nos Apêndice H e Apêndice I.

Tabela IX – Parâmetros de qualidade dos modelos de correlação linear entre as energias livres experimentais e os resultados de *docking* com as estruturas principais obtidas pela clusterização por RMSD dos carbonos- α .

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima Inicial	0.3599	0.1295	0.364	0.671
Enzima + Ligante 04 R3 CLS-1	0.3646	0.1330	0.273	0.495
Enzima + Ligante 04 R3 CLS-52	0.3659	0.1339	0.383	0.692
Enzima + Ligante 13 R2 CLS-1	0.3728	0.1390	0.284	0.503
Enzima + Ligante 04 R2 CLS-255	0.3773	0.1423	0.390	0.680
Enzima + Substrato R1 CLS-446	0.3811	0.1452	0.364	0.628
Enzima + Substrato R1 CLS-1	0.3854	0.1486	0.339	0.576
Enzima + Ligante 27 R2 CLS-33	0.3909	0.1528	0.414	0.693
Enzima + Substrato R3 CLS-166	0.4006	0.1605	0.420	0.683
Enzima + Produto R3 CLS-1	0.4060	0.1648	0.506	0.809
Enzima + Ligante 46 R2 CLS-189	0.4234	0.1793	0.290	0.442
Enzima + Ligante 13 R2 CLS-415	0.4309	0.1857	0.245	0.365
Enzima + Ligante 37 R1 CLS-2	0.4386	0.1924	0.403	0.586

Nas duas metodologias de clusterização avaliadas, cerca de 15% das estruturas produziu resultados de *docking* com melhor correlação linear com os dados experimentais. Em média, obteve-se um coeficiente de correlação de Pearson de 0,21 pela clusterização por RMSD dos

carbonos- α e de 0,20 para clusterização por RMSD do sítio ativo, bem como uma inclinação média das retas de correlação de 0,19 e de 0,18 para cada metodologia, respectivamente.

A proximidade entre esses valores sugere que boa parte dos clusters principais gerados são os mesmos nas duas metodologias, o que é coerente na medida em que alguns resíduos do sítio ativo, sobretudo da parede externa e do laço helicoidal estão entre os que apresentam as maiores mobilidades na maioria das simulações, tendo um impacto grande no RMSD total. Portanto, ao se clusterizar através da mobilidade de todos os carbonos- α também se está clusterizando indiretamente pela mobilidade do sítio ativo, em função da contribuição expressiva desses resíduos têm na flexibilidade total.

A clusterização por meio do RMSD do sítio ativo, no entanto, se mostra mais interessante na medida em que pode capturar alterações pequenas, possivelmente rearranjos das cadeias laterais dos aminoácidos, que acabam tendo um impacto significativo na capacidade de ligação dos inibidores. Por essa razão é que se obtém em média 4,3 clusters principais em cada trajetória aplicando-se essa metodologia, contra 3,4 da clusterização pelo RMSD dos carbonos- α e também o motivo de se obter os maiores valores de correlação de Pearson (0,53 e 0,52) para estruturas extraídas através desse método.

Tabela X – Parâmetros de qualidade dos modelos de correlação linear entre as energias livres experimentais e os resultados de *docking* com a estruturas principais obtidas pela clusterização por RMSD do sítio ativo.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima Inicial	0.3599	0.1295	0.364	0.671
Enzima + Produto R2 CLS-383	0.3650	0.1332	0.292	0.529
Enzima + Ligante 04 R3 CLS-143	0.3665	0.1343	0.217	0.392
Enzima + Produto R1 CLS-654	0.3716	0.1381	0.386	0.685
Enzima + Ligante 04 R3 CLS-55	0.3722	0.1385	0.279	0.494
Enzima + Substrato R2 CLS-18	0.3754	0.1410	0.461	0.809
Enzima + Produto R1 CLS-817	0.3757	0.1412	0.375	0.658
Enzima + Substrato R2 CLS-784	0.3773	0.1423	0.564	0.984
Enzima + Ligante 27 R2 CLS-6	0.3815	0.1455	0.596	1.026
Enzima + Ligante 04 R3 CLS-203	0.3940	0.1553	0.324	0.537
Enzima + Produto R2 CLS-10	0.4064	0.1651	0.550	0.879
Enzima + Substrato R1 CLS-57	0.4108	0.1688	0.395	0.623
Enzima + Ligante 04 R2 CLS-19	0.4132	0.1707	0.361	0.565
Enzima + Ligante 04 R1 CLS-19	0.4151	0.1723	0.371	0.578
Enzima + Ligante 04 R3 CLS-49	0.4369	0.1909	0.403	0.590
Enzima + Ligante 37 R1 CLS-103	0.5222	0.2727	0.345	0.400
Enzima + Ligante 27 R2 CLS-150	0.5302	0.2811	0.551	0.627

Novamente, os destaques negativos ocorrem para as simulações da enzima livre em solução, como pode ser constatado na Tabela XI, em que são mostrados os valores médios dos parâmetros de qualidade e a inclinação das curvas de correlação para cada sistema, avaliadas para todas as estruturas obtidas pelas duas metodologias de clusterização e sobre todas as réplicas.

Pode-se observar que o coeficiente de correlação e Pearson e o coeficiente de determinação obtidos para enzima livre são próximos de zero, havendo para algumas estruturas

até mesmo valores negativos de correlação (cf. Apêndice H e Apêndice I), comprovando a hipótese de que na ausência de ligantes que estabilizem o sítio ativo em uma conformação adequada para interação, há a perda da capacidade de ligação efetiva.

Por essa mesma razão os sistemas da enzima com ligantes 13 e 46 vêm logo em seguida na lista dos piores valores de correlação entre os resultados de *docking* e os dados experimentais, visto que tais na maior parte do tempo simulado tais ligantes encontram-se fora do sítio de ligação e em geral muito distantes desse.

Tabela XI - Valores médios dos parâmetros de qualidade e da inclinação das curvas de correlação considerando-se todas as estruturas geradas de cada sistema, em ordem crescente de correlação de Pearson.

Sistema	R-Pearson	COD	Inclinação	RMSE
Enzima Livre	0.0498	0.0228	0.031	0.422
Enzima + Ligante 46	0.1063	0.0347	0.068	0.415
Enzima + Ligante 13	0.1460	0.0401	0.097	0.452
Enzima + Ligante 37	0.1653	0.0593	0.142	0.646
Enzima + Ligante 27	0.2612	0.0865	0.248	0.584
Enzima + Produto	0.2991	0.0949	0.310	0.655
Enzima + Ligante 04	0.3049	0.0991	0.265	0.586
Enzima + Substrato	0.3282	0.1103	0.371	0.751

No sentido oposto, estão as estruturas geradas pelas simulações da enzima com seu substrato natural, seguidas pelas do ligante 04 (ácido gálico), que é composto experimentalmente mais ativo e também é o composto que mais mimetiza o substrato, indicando que a indução conformacional que o substrato provoca no sítio ativo a fim de maximizar suas interações e que culminam com sua conversão à produto também é realizada pelo ácido gálico, porém ao maximizar essas interações esse ligante induz a inibição enzimática.

Outro ligante importante e que também atua como inibidor é o produto (ou ligante 05), e que vem logo em sequência na lista das maiores correlações médias, seguida dos inibidores menos potentes, ligantes 27 e 37.

Portanto, tirando-se os ligantes 13 e 46 que não se encontram apropriadamente ligados no sítio ativo durante as simulações, a correlação de Pearson média de cada sistema respeita a mesma ordem que a atividade inibitória experimental, o que evidencia como os ligantes produzem de fato alterações no sítio ativo que afetam sua capacidade de ligação e que impactam a atividade inibitória *in vitro*, alterações essas que devem ser levadas em consideração para a compreensão dos mecanismo de inibição enzimática e sobretudo para o desenho de novos compostos, melhores e plenamente efetivos como inibidores.

Ao todo, combinando-se as estruturas geradas pelas três estratégias de clusterização empregadas, foi possível selecionar 38 conformações que apresentaram um desempenho melhor que a estrutura cristalográfica na correlação entre os resultados de *docking* e os dados experimentais (Tabela XII).

Tabela XII – Parâmetros de qualidade e da inclinação das retas de correlação em ordem crescente de R-Pearson para as estruturas com desempenho de *docking* superior à estrutura cristalográfica geradas pelas três estratégias de clusterização empregadas.

Estrutura	R-Pearson	COD	Inclinação	RMSE	Ordem
Enzima + Ligante 27 R2 SA-150	0.5302	0.2811	0.551	0.627	1
Enzima + Ligante 37 R1 SA-103	0.5222	0.2727	0.345	0.400	2
Enzima + Substrato R3 LH	0.4708	0.2216	0.421	0.560	3
Enzima + Ligante 37 R1 CA-2	0.4386	0.1924	0.403	0.586	4
Enzima + Ligante 04 R3 SA-49	0.4369	0.1909	0.403	0.590	5
Enzima + Ligante 13 R2 CA-415	0.4309	0.1857	0.245	0.365	6
Enzima + Ligante 46 R2 CA-189	0.4234	0.1793	0.290	0.442	7
Enzima + Ligante 04 R1 LH	0.4186	0.1752	0.397	0.612	8
Enzima + Ligante 04 R1 SA-19	0.4151	0.1723	0.371	0.578	9
Enzima + Ligante 04 R2 SA-19	0.4132	0.1707	0.361	0.565	10
Enzima + Substrato R2 FN	0.4125	0.1702	0.318	0.498	11
Enzima + Substrato R1 SA-57	0.4108	0.1688	0.395	0.623	12
Enzima + Produto R2 SA-10	0.4064	0.1651	0.550	0.879	13
Enzima + Produto R3 CA-1	0.4060	0.1648	0.506	0.809	14
Enzima + Substrato R3 CA-166	0.4006	0.1605	0.420	0.683	15
Enzima + Ligante 04 R3 SA-203	0.3940	0.1553	0.324	0.537	16
Enzima + Ligante 27 R2 CA-33	0.3909	0.1528	0.414	0.693	17
Enzima + Substrato R1 CA-1	0.3854	0.1486	0.339	0.576	18
Enzima + Substrato R1 LH	0.3828	0.1466	0.429	0.735	19
Enzima + Ligante 37 R2 LH	0.3827	0.1464	0.444	0.763	20
Enzima + Ligante 27 R2 SA-6	0.3815	0.1455	0.596	1.026	21
Enzima + Substrato R1 CA-446	0.3811	0.1452	0.364	0.628	22
Enzima + Substrato R2 SA-784	0.3773	0.1423	0.564	0.984	24
Enzima + Ligante 04 R2 CA-255	0.3773	0.1423	0.390	0.680	23
Enzima + Ligante 04 R2 LH	0.3769	0.1420	0.503	0.878	25
Enzima + Produto R1 SA-817	0.3757	0.1412	0.375	0.658	26
Enzima + Substrato R2 SA-18	0.3754	0.1410	0.461	0.809	27
Enzima + Ligante 27 R2 FN	0.3753	0.1409	0.296	0.519	28
Enzima + Ligante 13 R2 CA-1	0.3728	0.1390	0.284	0.503	29
Enzima + Ligante 04 R3 SA-55	0.3722	0.1385	0.279	0.494	30
Enzima + Produto R1 SA-654	0.3716	0.1381	0.386	0.685	31
Enzima + Produto R2 LH	0.3669	0.1346	0.281	0.506	32
Enzima + Ligante 04 R3 SA-143	0.3665	0.1343	0.217	0.392	33
Enzima + Ligante 04 R3 CA-52	0.3659	0.1339	0.383	0.692	34
Enzima + Substrato R2 LH	0.3654	0.1335	0.653	1.182	35
Enzima + Produto R2 SA-383	0.3650	0.1332	0.292	0.529	36
Enzima + Ligante 04 R3 CA-1	0.3646	0.1330	0.273	0.495	37
Enzima + Ligante 37 R1 LH	0.3645	0.1329	0.514	0.933	38
Enzima Inicial	0.3599	0.1295	0.364	0.671	--

*SA = clusterização por RMSD do sítio ativo, CA = clusterização por RMSD dos carbonos- α , FN = estrutura do último frame da respectiva trajetória, LH = estrutura com maior número de ligações de hidrogênio simultâneas enzima-ligante.

Contrariando o resultado médio, entretanto, duas estruturas extraídas das simulações enzima-ligante 13 e enzima-ligante 46 apresentaram a sexta e a sétima melhor correlação com os dados experimentais, respectivamente. Uma análise mais criteriosa mostra que essas duas conformações foram geradas ao final de suas trajetórias, nos tempos de 97230 ps e de 97060 ps, para as simulações com os ligantes 13 e 46, respectivamente, momento em que tais ligantes já haviam retornado e entrado, ainda que parcialmente, no sítio ativo.

Isso reforça como os ligantes afetam o sítio ativo e o induzem a adotar conformações que favoreçam e fortaleçam suas interações, comprovando a validade do modelo de encaixe-induzido e sua aplicabilidade no estudo da presente enzima, bem como mostra que sem o ligante essas conformações mais ativas não seriam obtidas.

Além disso, isso comprova como a seleção arbitrária do último frame não é adequada para amostrar a conformação da proteína ao final da simulação, uma vez que o *docking* com tais estruturas produziu correlações significativamente mais baixas (R-Pearson de 0.1645 e 0.3177 para os ligantes 13 e 46, respectivamente), mostrando que flutuações instantâneas podem fazer com que a estrutura desse frame não condiga com a estrutura final equilibrada.

Outro aspecto interessante é que os dois melhores resultados de correlação ocorrem com estruturas geradas nas simulações com os ligantes 27 e 37, que não são inibidores experimentalmente eficazes, mas que ainda assim produziram R^2 de 0,28 e 0,27, respectivamente, valores muito superiores ao observado para a estrutura inicial ($R^2 = 0,13$).

É possível que seja justamente a ineficácia desses compostos a responsável pela maior correlação entre os dados de *docking* e os resultados experimentais observadas para essas estruturas, na medida em que ao não interagirem tão bem com os resíduos do sítio ativo, também não restringem sua movimentação de forma tão intensa como fazem ligantes potentes como ácido gálico e próprio produto enzimático.

Isso permite rearranjos das cadeias laterais dos aminoácidos do sítio, porém sem que haja uma alteração tão drástica da estrutura total do sítio ativo, como a observada nas simulações da enzima livre e que culminam na perda da capacidade de ligação efetiva com os compostos. Esses rearranjos possivelmente levam a um aumento momentâneo do volume da cavidade interna do sítio ativo, permitindo o *docking* adequado dos ligantes e a redução da penalização por impedimentos estéreo.

Ou seja, esses ligantes pouco efetivos fazem com que para essas simulações haja uma situação intermediária entre a ausência total de ligantes, com perda da capacidade de interação do sítio, e o outro extremo que é a presença de ligantes extremamente potentes, que constroem o sítio e prejudicam os resultados de *docking* ao induzir conformações extremamente impedidas.

Essa situação, no entanto, é muito efêmera, como comprova o RMSF dos carbonos- α dessas simulações (cf. Figura 76), que mostram a reduzida mobilidade do sítio ativo para esses sistemas, razão pela qual poucas estruturas extraídas dessas trajetórias são realmente efetivas no *docking*, produzindo os baixos valores médio de correlação mostrados na Tabela XI.

Aliás, provavelmente é a constrição do sítio ativo na presença dos ligantes a responsável para que mais de 78% das estruturas obtidas das trajetórias tenham um desempenho de *docking* inferior ao obtido para a estrutura cristalográfica.

Essa hipótese é reforçada pela observação de que a presença de ligantes mais volumosos do que o produto, como o ácido gálico e o substrato levaram a um aumento na mobilidade do

sítio ativo, como observado no comportamento do RMSD dos carbonos- α desses sistemas (c.f. Figura 47e Figura 66), levando-os a romper parcialmente o estreitamento do sítio que havia na estrutura cristalográfica, diferentemente do que fazem os ligantes 27, 37 e o produto (cf. Figura 57Figura 76), o que faz com que desempenho médio de *docking* das estruturas geradas na presença do substrato e do ligante 04 sejam significativamente superiores às dos demais ligantes.

Lembrando que soma de um número suficientemente grande de interações de variáveis aleatórias independentes tende a uma distribuição normal centrada no valor de equilíbrio, segundo o teorema do limite central, fez-se a média dos resultados de *docking* das estruturas que produziram individualmente os melhores resultados de correlação frente aos dados experimentais, adicionando-se uma a uma até que todas as 38 estruturas tenham sido contabilizadas.

Na Figura 94 está mostrado o gráfico do coeficiente de Pearson e da inclinação da curva de correlação em função do número de estruturas adicionadas no cálculo dessas médias. Pode-se observa que tanto a correlação como a inclinação da curva de correlação atingem o valor máximo quando os resultados das duas melhores estruturas são somados, para os quais se obtém um R-Pearson de 0,5834 e inclinação de 0,448.

A consequente inclusão de mais estruturas tende a diminuir a correlação, até a sexta estrutura ser adicionada, quando há um novo incremento no coeficiente de Pearson, porém acompanhando deste ocorre uma diminuição importante na inclinação da curva, indicando um declínio da precisão. De fato, se os resultados dessa estrutura são adicionados aos das duas melhores obtém uma correlação de 0,6022, porém a inclinação da curva de correlação se reduz para 0,380, reforçando que essa melhora do modelo está de fato condicionada à perda de precisão.

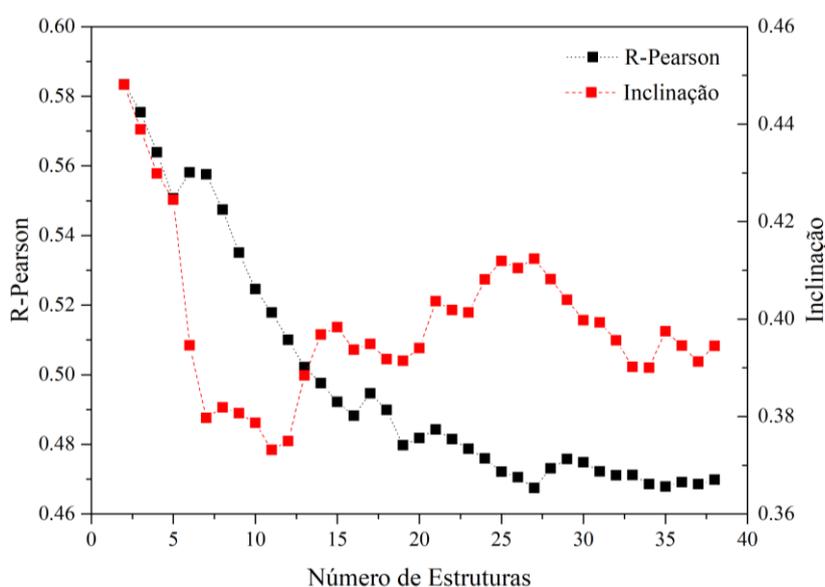


Figura 94 – Coeficientes de correlação de Pearson e inclinação da curva de correlação entre a média dos resultados de *docking* e as energias livres experimentais em função do número de estruturas consideradas na composição da média.

Além desse ponto, a inclusão de mais estruturas não é capaz de incrementar nem a correlação nem a inclinação da curva. Ainda que algumas, com a 17 e a 28 tenham conseguido aumentar levemente o R-Pearson, esse valor jamais volta a superar o obtido para as duas

melhores. Ao final, quando todas as 38 melhores estruturas são adicionadas no cálculo das médias obtém-se um coeficiente de correlação de 0.4699 e uma inclinação de 0.394.

Essas observações estão em acordo com outros estudos de *ensemble docking*, que mostram que conjuntos com algumas poucas conformações podem melhorar a correlação frente aos dados experimentais, enquanto que conjuntos maiores tendem a comprometer não apenas o desempenho como também a precisão dos modelos gerados. Essa diminuição de qualidade se deve tanto à imprecisão intrínseca das funções de escore implementadas nos programas de *docking* quanto ao potencial de se gerar mais falsos positivos à medida em que o número de conformações da proteína aumenta, sobretudo se estruturas não-nativas estão sendo incluídas na composição do *ensemble*.^{83,84}

Tabela XIII - Escores de *docking* e desvio percentual em relação aos valores experimentais dos dois *ensembles* finais gerados a partir dos melhores resultados das estruturas individuais. Resultados em ordem crescente de ΔG de ligação experimental.

Ligante	ΔG exp. (kcal/mol)	Ensemble 1		Ensemble 2	
		Escore (kcal/mol)	Desvio (%)	Escore (kcal/mol)	Desvio (%)
4	-7.36	-8.0	-8.7	-7.4	-0.6
7	-6.07	-6.8	-11.1	-6.4	-5.5
5	-6.07	-7.4	-21.7	-6.9	-13.5
22	-5.83	-6.3	-7.2	-5.9	-1.8
13	-5.72	-6.1	-5.8	-5.9	-3.6
1	-5.69	-6.7	-16.9	-6.1	-8.0
46	-5.55	-6.1	-8.9	-5.8	-5.0
6	-5.54	-6.0	-8.9	-5.9	-6.3
28	-5.44	-6.1	-11.5	-5.7	-5.3
9	-5.11	-6.2	-20.5	-5.9	-15.6
29	-5.02	-6.3	-25.6	-6.2	-22.9
32	-5.01	-6.5	-29.2	-6.2	-23.4
19	-4.95	-6.0	-20.1	-5.9	-19.1
36	-4.89	-6.3	-28.1	-5.9	-20.2
14	-4.86	-5.8	-18.3	-5.5	-13.8
35	-4.83	-6.1	-25.3	-5.6	-15.3
18	-4.74	-6.9	-45.5	-6.4	-34.9
48	-4.73	-5.6	-17.2	-5.5	-16.2
41	-4.68	-6.4	-35.6	-6.2	-32.8
27	-4.63	-6.6	-42.6	-6.1	-32.5
8	-4.61	-6.5	-41.1	-6.1	-33.2
47	-4.56	-6.0	-32.5	-5.9	-28.8
34	-4.49	-6.4	-41.8	-5.9	-30.9
37	-4.48	-6.9	-53.6	-6.6	-47.0
15	-4.21	-5.8	-36.6	-5.5	-29.9

Diante disso, avaliou-se comparativamente os dois *ensembles* que produziram os melhores resultados na comparação com os dados experimentais. O *Ensemble 1* é composto por duas estruturas obtidas pela clusterização de sítio ativo, uma da simulação com o ligante 27 (cluster 150, réplica 2) e outra da simulação com o ligante 37 (cluster 103, réplica 1), as quais geram individualmente coeficientes de correlação de 0.5302 e 0.5222, respectivamente, mas combinadas esse valor chega a 0,5834.

O *Ensemble 2*, por sua vez, é composto por essas duas estruturas e pela sexta com melhor desempenho individual (cluster 415 da simulação enzima-ligante 13, réplica 2), que foi obtida pela clusterização por RMSD dos carbonos- α . Individualmente essa estrutura gerou um R-Pearson de 0.4309, entretanto sua inclusão junto às estruturas do *Ensemble 1* promoveu um incremento no coeficiente de correlação, que passou para 0,6022.

Apesar dessa inclusão levar à uma diminuição na inclinação da curva de correlação, que passa de 0,448 no *Ensemble 1* para 0.394 no *Ensemble 2*, o que formalmente indica uma perda na precisão do modelo, quando se avalia o desvio dos resultados de *docking* em relação aos valores experimentais (Tabela XIII) obtém-se para o *Ensemble 1* um desvio médio de 25%, comparável à estrutura inicial, que foi de 23%, enquanto para o *Ensemble 2* esse desvio é reduzido para pouco mais de 18%.

A melhora no coeficiente de correlação aliada à diminuição do desvio percentual médio frente aos valores experimentais motivou a escolha deste *Ensemble* como o modelo final. A curva que correlaciona os resultados desse *ensemble* de estruturas com os dados experimentais está mostrada na Figura 95, enquanto que a curva obtida para o *Ensemble 1* é mostrada no Apêndice J.

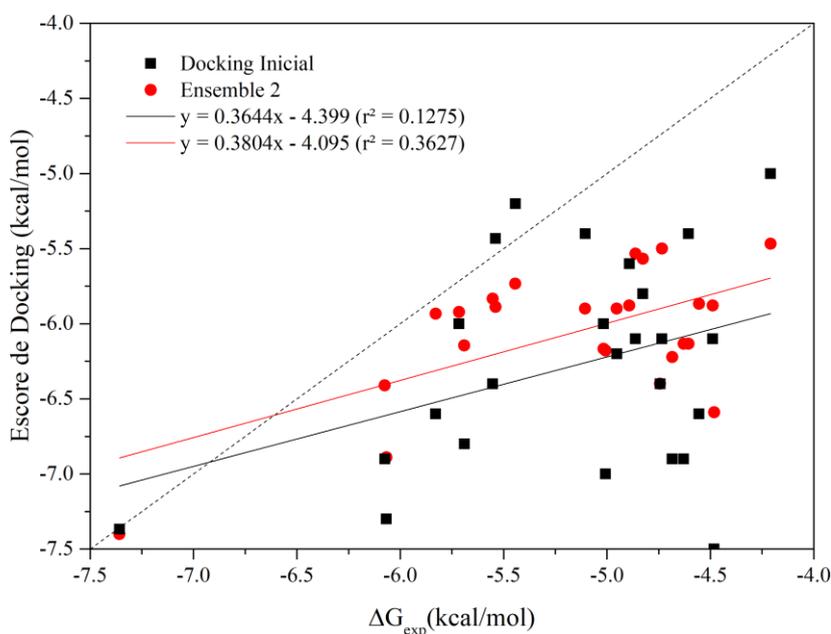


Figura 95 – Relações entre os resultados de *docking* e os valores experimentais de energia livre de ligação e curva de correlação obtidas para o *Ensemble 2* (vermelho) em comparação com o obtido pelo *docking* na estrutura cristalográfica (preto).

Observa-se na figura que a inclinação da curva de correlação do *Ensemble 2* é muito próxima daquela observada para a curva de correlação do *docking* na estrutura cristalográfica, ambas distantes do comportamento canônico, porém a dispersão dos pontos no *Ensemble 2* é significativamente menor que para o *docking* inicial, por isso a melhora tão significativa na correlação entre as variáveis. Além disso, se percebe o deslocamento da curva de correlação do *Ensemble 2* em direção à curva canônica, o que se traduz na melhora do desvio em relação aos valores experimentais na comparação com o *docking* na estrutura cristalográfica.

Quando se sobrepõem os modos de ligação de menor energia de cada ligante nas estruturas que compõem esse *ensemble* (Figura 96) observa-se que apenas na estrutura extraída da simulação com ligante 27 há compostos que ficam fora do sítio ativo, e justamente os compostos mais volumosos (13, 41, 46, 47 e 48). Para as demais simulações todos os compostos foram docados adequadamente no sítio.

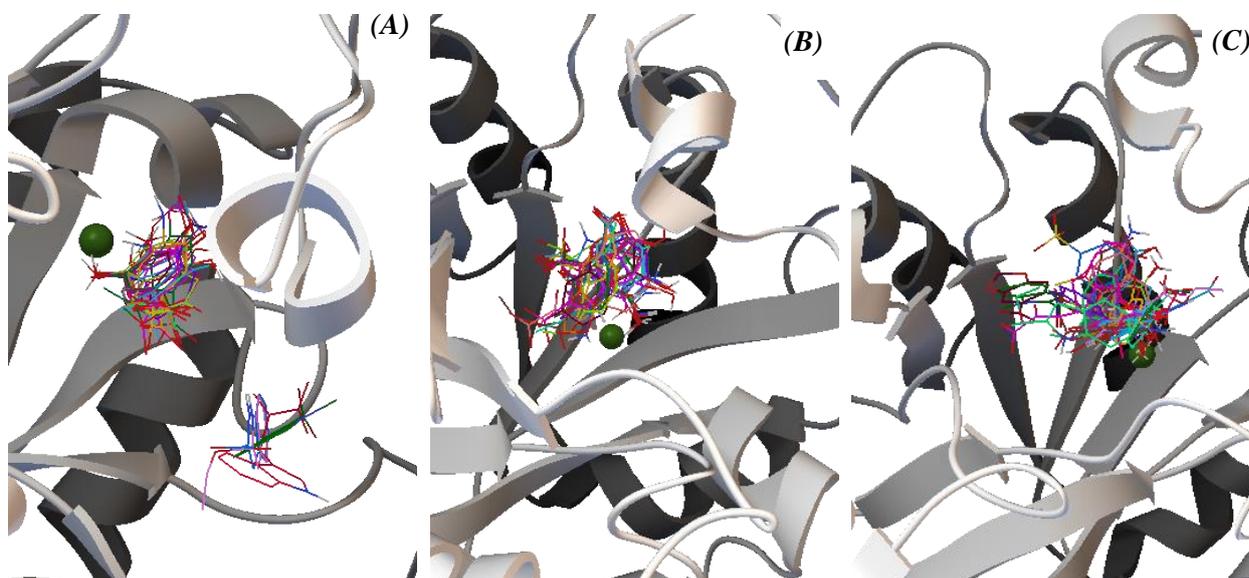


Figura 96 – Sobreposição do modo de ligação de menor energia nas estruturas que compõem o *Ensemble 2*, extraídas das simulações enzima-ligante 27 (A), enzima-ligante 37 (B) e enzima-ligante 13 (C). O cofator magnésio está destacado em verde.

Os modos de ligação de menor energia não são estritamente iguais para as três estruturas – possivelmente a razão pela qual a correlação é melhor combinando-se as três estruturas do que as considerando individualmente - porém tendem a ser muito parecidos, como pode ser constatado na Figura 97, em que está mostrado o melhor modo de ligação para o ácido gálico (ligante 04) no *docking* com as estruturas extraídas das simulações com os ligantes 27 e 37.

Chama a atenção que esses modos de ligação são muito diferentes daqueles observados pelo *docking* na estrutura cristalográfica (cf. Figura 34) e também divergem do padrão principal de interações verificado nas simulações de dinâmica molecular desse composto, não havendo mais, por exemplo, a formação de ponte salina entre a Arg-102 e a porção carboxílica do ligante, tampouco interação significativa de seus grupos hidroxila com a His175, que foram apontados pelas dinâmicas como as âncoras principais de complexação do ligante 04 no sítio ativo enzimático. Ao invés dessas se observam interações com os glutamatos 142 e 253, que fazem parte do sítio de complexação do magnésio.

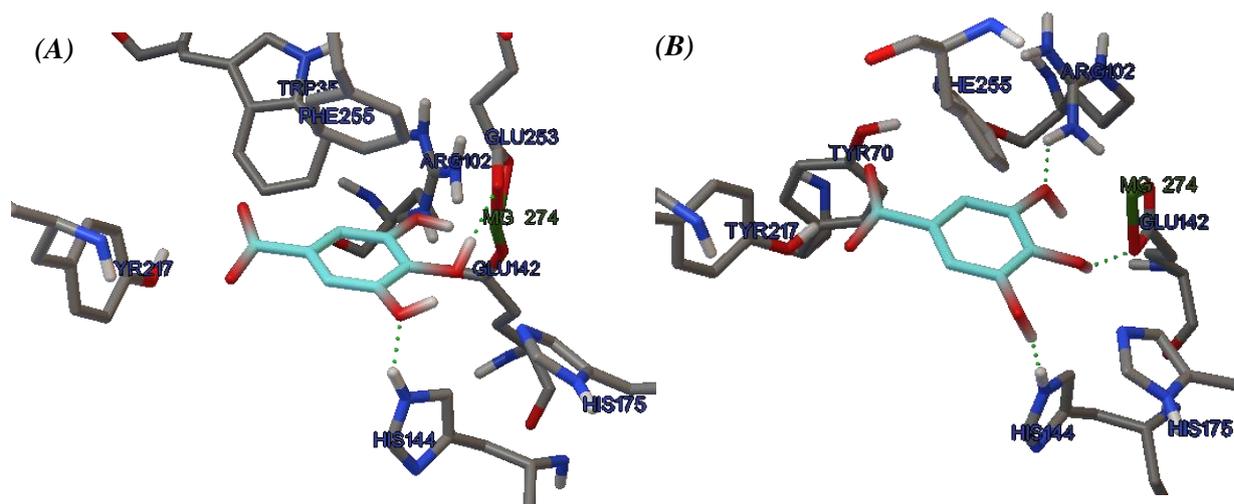


Figura 97 - Modo de ligação de menor energia para o ligante 04 obtido por *docking* nas estruturas do *Ensemble 2* obtidas das simulações com ligante 27 (A) e com o ligante 37 (B).

De forma análoga, os modos de ligação apontados para o produto divergem não apenas do observado por *docking* e nas simulações de dinâmica molecular, mas também do próprio resultado cristalográfico (cf. Figura 27), novamente apontado interações importante com os glutamatos, bem como a atuação da Arg102 na interação com os grupos hidroxila e não mais com a porção carboxílica do ligante, como pode ser constatado na Figura 98.

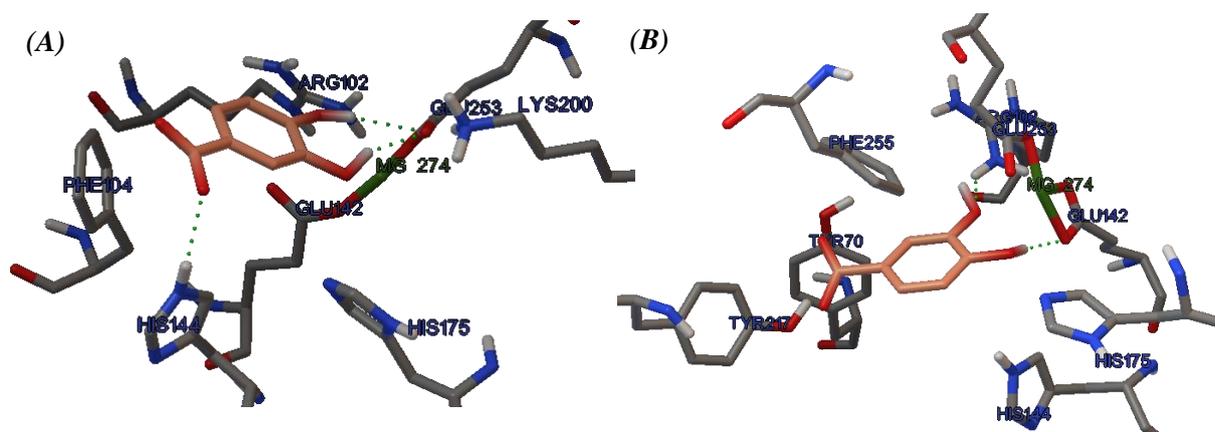


Figura 98 – Modo de ligação de menor energia para o produto segundo *docking* nas estruturas do *Ensemble 2* obtidas das simulações com ligante 13 (A) e com o ligante 37 (B)

Isso indica que tais estruturas não correspondem ao melhor modo de ligação possível para esses ligantes, mas sim são conformações capazes de receber e de ligar relativamente bem todos os ligantes experimentalmente ativos, motivo pelo qual se consegue boa correlação entre os resultados experimentais e teóricos, porém com baixa precisão. Ou seja, são estruturas mais versáteis e certamente com um sítio ativo menos restrito, o que se constata pelo afastamento apresentado por resíduos como a Arg102, Phe104, His175 e Phe255 em comparação com a estrutura de partida (cf. Figura 27).

É possível que tais estruturas correspondam a um mínimo energético local comum a todos os ligantes, o qual pode evoluir para o mínimo global de cada complexo no qual interações receptor-ligantes seriam maximizadas pela indução de alterações conformacionais no sítio ativo que as favorecem. Diante disso é possível que estrutura cristalográfica, na qual o produto

encontra-se complexado ao sítio ativo, seja o mínimo ou esteja perto do mínimo energético global desse complexo, razão pela qual esse modo de ligação se mantém inalterado ao longo das simulações de dinâmica molecular.

Por sua similaridade estrutural com o produto, ligantes com o ácido gálico e o próprio substrato possivelmente tem os mínimos energético globais de seus complexos muito próximos ao mínimo do produto, por isso os padrões de interações desses compostos são muito semelhantes entre si.

Por outro lado, compostos mais volumosos ou que detém pouca similaridade estrutural esses análogos de ácido 3,4-dihidroxibenzoico, como os indoles e benzofuranos não puderam ser adequadamente docados no sítio ativo cristalográfico pois provavelmente os mínimos energéticos absolutos de seus complexos não correspondem ao mínimo cristalográfico, motivo que leva à baixa correlação entre os resultados de *docking* na estrutura inicial e os valores experimentais de energia livre de ligação.

Na estruturas do *Ensemble 2*, entretanto, tais compostos conseguem ser adequadamente ligados, como pode ser observado na figura Figura 99, que mostra como os modos de ligação de menor energia dos ligantes 13 e 46, que são o quinto e o sétimo inibidores mais ativos, em uma das estruturas desse *Ensemble*.

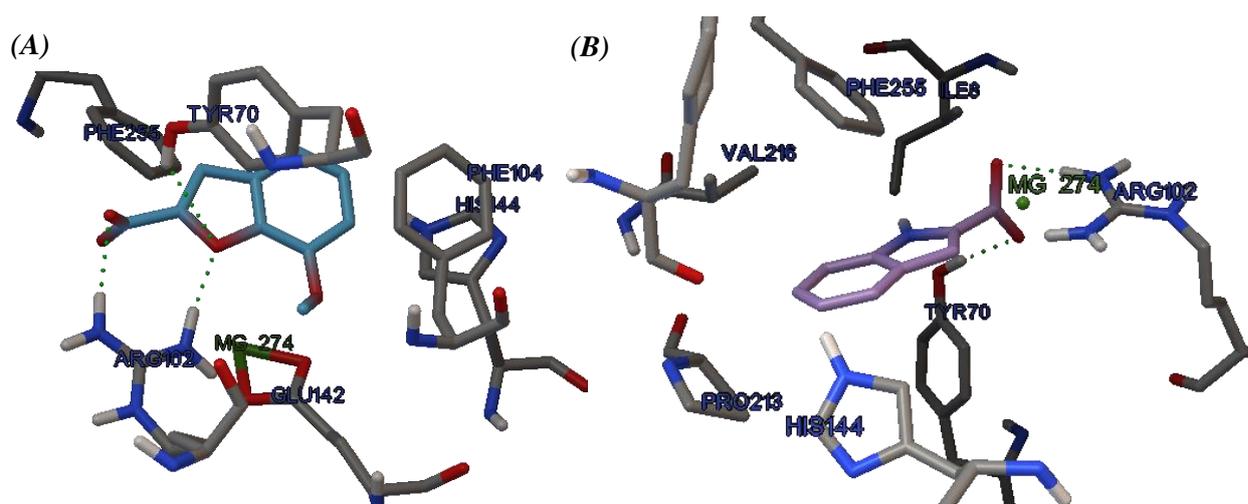


Figura 99 – Modo de ligação de menor energia para o ligante 13 (A) e o ligante 46 (B) segundo *docking* nas estruturas do *Ensemble 2* obtida a partir da simulação com o ligante 37.

Observa-se na figura que o padrão de interação desses dois compostos na estrutura é muito semelhante, com interações importantes entre seus grupos carboxílicos e resíduos como a Arg102 e a Tyr70, com as quais há formação de ligações de hidrogênio, bem como contatos hidrofóbicos com resíduos como a Phe104 e a Phe255, com capacidade de produzirem empilhamentos π - π e até mesmo formar uma gaiola aromática ao redor desses ligantes.

Essas interações, entretanto, provavelmente serão plenamente observadas apenas quando tais complexos convergirem para seus mínimos globais, hipótese que pode e será testada como uma etapa futura do presente trabalho, por meio da realização de novas simulações de dinâmica molecular partindo-se dos complexos nas estruturas do *Ensemble 2*, e verificando-se se os complexos com os análogos de 3,4-DHB e com ele próprio evoluem para o modo de ligação

cristalográfico, bem como se os demais ligantes são capaz de alterar o sítio ativo a partir dessa conformação inicial a fim de maximizar suas interações com a enzima.

Por outro lado, o modelo que correlaciona as energias livres de ligação e os resultados de *docking* obtido a partir desse *Ensemble*, já é suficiente para estimar com relativa confiabilidade a potência inibitória *in vitro* de novos compostos e realizar a triagem virtual de bibliotecas de moléculas em busca de novos *hits*, sobretudo quando o modelo é construído a partir dos 10 ligantes com maior atividade experimental (Figura 100), para o qual se obtém um coeficiente de correlação de Pearson de 0,9031, inclinação da curva de 0,80, RMSE de 0,25 kcal/mol e um coeficiente de determinação de 0,8157. Quando se correlaciona os dados desses mesmos ligantes com os resultados de *docking* na estrutura cristalografica se obtém um R² de 0,5683 e um R-Pearson de 0,7538 com RMSE de 0,55 kcal/mol, evidenciando a melhora gerada por se incluir a flexibilidade do receptor no *docking*.

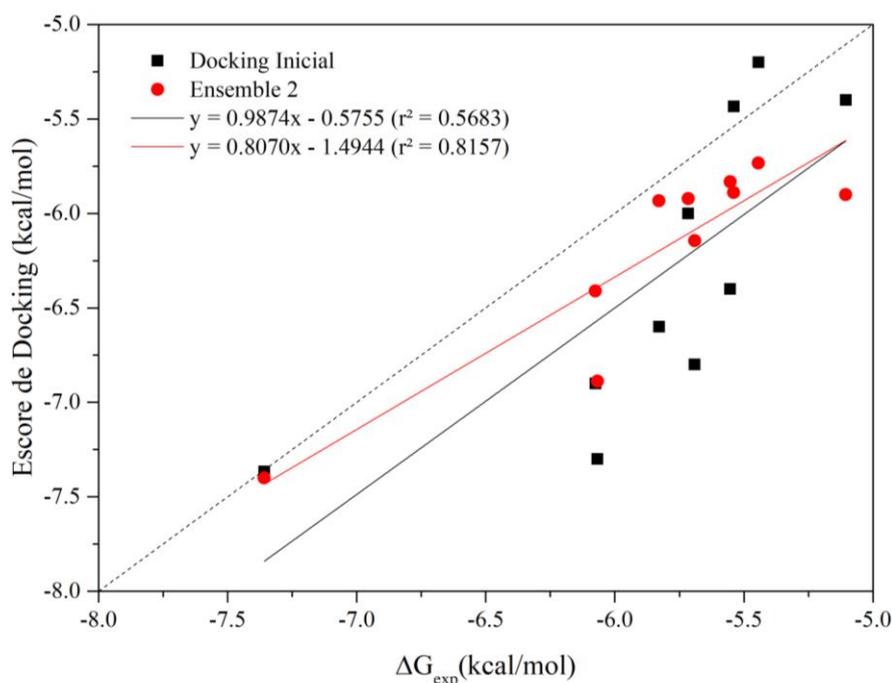


Figura 100 – Relações entre os resultados de *docking* e os valores experimentais de energia livre de ligação e curva de correlação obtidas para o *Ensemble 2* (vermelho) em comparação com o obtido pelo *docking* na estrutura cristalográfica (preto) considerando-se os 10 ligantes mais ativos experimentalmente na construção do modelo.

A construção do modelo a partir dos 10 inibidores mais ativos, cujos potenciais inibitórios em termos de IC₅₀ variam entre 4,36 μmol.L⁻¹ e 191 μmol.L⁻¹, é adequada na medida em que a triagem virtual se destina a encontrar compostos ou testar moléculas desenhados racionalmente que sejam inibidores comparáveis ou, se possível, melhores que o ácido gálico. Assim, compostos que apresentem IC₅₀ teórico superior 200 μmol.L⁻¹ seriam descartados, de qualquer maneira, não sendo, portanto, necessário incluir na construção do modelo os compostos com concentração inibitória superior a este limite.

6 CONCLUSÕES E PERSPECTIVAS

O surgimento de cepas bacterianas super-resistentes e capazes de trocar facilmente essa resistência entre si por meio de plasmídeos, e sobretudo a descoberta recente de bactérias capazes de resistir a *todos* os antibióticos atualmente disponíveis alertam para a necessidade de maiores investimentos na pesquisa e no desenvolvimento de novas terapias antimicrobianas, focadas sobretudo na busca de novos alvos para a ação desses fármacos a fim de burlar as defesas bacterianas.

Dentre os microrganismos que se destacam, não apenas por já se ter encontrado cepas resistente aos principais antibióticos, mas também por sua elevada virulência e alta letalidade está o *Bacillus anthracis*, a agente etiológico do antraz, uma doença conhecida desde os tempos bíblicos e que produz um quadro infeccioso agudo e extremamente grave, com uma taxa de mortalidade superior a 90% em sua forma respiratória.

Não bastasse a alta letalidade, em condições adversas à sua sobrevivência o *B. anthracis* produz esporos facilmente dispersáveis no ambiente e capazes de resistir a agentes físicos e químicos e de se manterem viáveis por vários anos. Assim, à medida em que o aquecimento global se intensifica, esporos de antraz presos no gelo em regiões polares e subpolares estão sendo liberados e surtos da doença – que pensava-se estar controlada – estão voltando a ocorrer. Além disso, esses esporos também conferem ao *B. anthracis* um grande potencial como arma biológica e agente de bioterrorismo, propósito para o qual já foi utilizando em 2001, nos Estados Unidos.

Uma das estratégias mais promissoras que surgiram recentemente na busca de novas terapias antimicrobianas é a inibição dos mecanismos bioquímicos de aquisição de nutrientes pelas bactérias, sendo a aquisição de ferro um dos alvos principais. No *Bacillus anthracis* essa aquisição é mediada pelos sideróforos bacilobactina e petrobactina, sendo este último essencial para o estabelecimento pleno de virulência no hospedeiro.

A recente descoberta do mecanismo de biossíntese da petrobactina e a constatação que a enzima AsbF, uma deidrosikimato desidratase, detém um papel fundamental nesse mecanismo, aliadas à elucidação da estrutura cristalográfica dessa enzima lançaram os alicerces para utilização da AsbF como alvo de novas terapias anti-antraz, sobretudo porque essa rota é única e não tem análoga em organismos superiores.

Dessa forma, nesse trabalho aplicou-se as metodologias computacionais de *docking* molecular e simulação de dinâmica molecular no estudo da interação de 25 compostos com ação inibitória experimental conhecida contra a AsbF, a fim de se compreender os padrões de ligação responsável pela atividade dessas moléculas, tanto de forma estática quanto dinâmica, e de se agregar informações a respeito dos mecanismos de inibição enzimática com foco no desenvolvimento racional de compostos com maiores atividades e seletividades e com potencial real de aplicação como antibióticos contra antraz.

O *docking* na estrutura cristalográfica, realizado com o programa AutoDock Vina, foi capaz de reproduzir o modo de ligação do ácido 3,4-dihidroxibenzoico – o produto da reação enzimática – tal qual observado na estrutura experimental, com um RMSD de 2,2 Å, valor próximo da resolução da estrutura cristalográfica, indicando que a metodologia de *docking*

utilizada é adequada para a predição, ao menos qualitativa, das interações receptor-ligante nesse sistema.

O *docking* do substrato mostrou esse composto em interação com resíduos importantes do sítio ativo como a Arg102 e a His144, com as quais forma ligações de hidrogênio, em concordância com as observações de Pflieger e colaboradores (2008), de que enzimas mutantes em que esse ligantes são trocados por alanina são completamente inativas. Além dessas, ocorre a formação de ligações e hidrogênio com os resíduos Tyr217 e com o esqueleto polipeptídico da Phe211.

A comparação entre os modos de ligação do substrato e do ligante permitiu inferir que o mecanismo, tal como sugerido por Pflieger, envolve o ataque nucleofílico da His144 ao hidrogênio axial do carbono C4 do substrato, formando a segunda insaturação do anel aromático. A etapa seguinte, entretanto, possivelmente envolve a ação da Tyr217 que, atuando como base, abstrai um dos hidrogênios na posição C6 (possivelmente o equatorial, num mecanismo de eliminação E1cB), o que justificaria o comportamento ótimo da enzima em condições básicas (pH 8.5), com a consequente expulsão de água assistida pela ligação de hidrogênio com a Phe211, completando a aromatização do anel.

Evidentemente, nenhum método computacional baseado puramente em mecânica molecular (MM) – que, em geral, despreza a natureza quântica (QM) das ligações químicas, bem como a quebra e formação de ligações – muito menos o *docking* é adequado para investigar o mecanismo de ação enzimática. Por essa razão, a investigação desse processo por métodos híbridos QM/MM, partindo dos resultados aqui expostos, é uma das perspectivas futuras desse trabalho.

Quantitativamente, os escores de *docking* apresentaram um desvio da ordem de 23%, em média, em relação aos valores de energia livre experimentais, com destaque para os ligantes 27 e 37, falsos positivos com desvios de 49% e de 67%, respectivamente.

Para o *docking* restrito ao sítio ativo, obteve-se um coeficiente de correlação de Pearson positivo de 0,360 e um R² de 0,1295 entre os resultados teóricos e os dados experimentais, indicando que o modelo linear obtido expressa a tendência experimental, porém não é adequado para a predição da energia livre de ligação.

Quanto aos modos de ligação, todos os ligantes com esqueleto estrutural semelhantes ao 3,4-DHBA foram docados adequadamente no sítio ativo enzimático, com o qual formam frequentemente ligações de hidrogênio com os resíduos His144, His175, Tyr217 e Arg102, essa última com características de ponte salina. Esse é o caso, por exemplo do ligante 04 (ácido gálico), o experimentalmente mais ativo, que ainda forma ligação de hidrogênio com a Phe211 (esqueleto) e possíveis interações π - π com a Phe255 e com a própria His144. Entretanto, não foi possível estabelecer um padrão claro entre o número de interações a atividade inibitória dos ligantes.

Ligantes mais volumosos, tais como os indolcarboxilatos (e.g. ligante 46) e benzofuranocarboxilatos (e.g. ligante 13) não foram docados adequadamente no sítio ativo cristalográfico, mas sim permaneceram em interação no canal de entrada do mesmo, sugerindo a existência de um modo de ligação de natureza alostérica ou então que o volume reduzido do sítio impede o *docking* adequado dos mesmo por impedimentos estéreos, hipótese de foi avaliada por meio da simulação de dinâmica molecular.

Com base nesses resultados, fez-se simulações de dinâmica molecular da enzima livre (sem ligante) e em complexo com substrato, produto e com os inibidores 04 (o mais potente), 27 e 37 (os falsos positivos), bem como os ligantes 13 e 46, compostos significativamente ativos, mas que apresentaram esses modos de ligação “alostéricos”.

Em linhas gerais, se observou que mesmo para as simulações da enzima livre, a estrutura se mantém significativamente estável, com um RMSD médio dos carbonos- α de cerca de 0,13 nm, valor que se torna ainda menor quando há ligantes no sítio ativo (RMSD médio de 0.11 nm).

Além disso, em todas as simulações foi possível alinhar mais de 85% do carbonos- α à um RMSD inferior a 1 Å, o que a demonstra a grande rigidez estrutural intrínseca dessa proteína, relacionada à sua forma de barril, composta por um conjunto de folhas- β internas cercada de hélices- α externas que são unidas por alças de 10 a 15 resíduos.

Em geral, são essas regiões de alça que compõem à maior parte da mobilidade observadas, bem como o laço helicoidal, uma região que une a folha $\beta 7$ à hélice $\alpha 7$ e que também forma uma pequena hélice, funcionando como “tampa” para o sítio ativo e compondo parte de sua parede externa e canal de entrada, além de agregar resíduos importantes como Phe211, Pro213 e Tyr217.

Em média, essa região apresentou RMSF de 0.2 nm para o produto, 0.3 nm para o substrato, 0.4 nm para o ligante 4, mostrando que sua mobilidade está relacionada com o aumento do volume do ligante, indicando que a presença de ligantes maiores que o cristalográfico (produto) promove um alargamento do sítio. Por essa razão, os ligantes 27 (com apenas um grupo OH) e o 37 (com dois grupos OH, assim como o produto) causam pouca alteração nessa porção, produzindo RMSF abaixo de 0.2 nm.

Isso evidencia que a estrutura do sítio cristalográfico é muito restrita devido à forte interação com o produto enzimático – que atua como um inibidor relativamente potente, segundo os dados experimentais – o que fez com que os ligantes mais volumosos não possam ser docados adequadamente no sítio ativo, produzindo o grande desvio observado entre os resultados de *docking* e as energias livres de ligação experimentais.

Essa observação é comprovada pelas simulações dos ligantes 13 e 46, significativamente mais volumosos que os demais, e que não permanecem ligando à entrada do sítio ativo, como inicialmente proposto (existência de um sítio de ligação “alostérico” na entrada do sítio principal) mas sim se desligam da proteína ao longo da trajetória.

Durante esse desligamento, esses compostos interagem com resíduos da superfície enzimática, provocando alterações estruturais significativas (RMSD médio dos carbonos- α de 0.13 nm para o ligante 13 e de 0.16 nm para o ligante 46), sobretudo no laço helicoidal, que chega a apresentar RMSF superior a 0.7 nm. Essas alterações levam à abertura do sítio ativo e permitem o retorno desses ligantes para dentro dele, evidenciando como o tamanho restrito do sítio cristalográfico impediu seu encaixe no *docking* inicial e mostrando como ligantes podem de fato induzir a alterações conformacionais para otimizar e intensificar suas interações com a enzima.

Possivelmente, os tempos de simulação não foram suficientes para permitir a entrada e o encaixe completos desses ligantes no sítio, de modo que estender essas simulações para comprovar se os mesmo poderão ser encaixados completamente na cavidade é uma das perspectivas desse trabalho, avaliando-se os padrões de interação e eventualmente até mesmo

estimando a energia livre de ligação por meio do Potencial de Força Média com *umbrella sampling* a partir dessas trajetórias.

A atividade inibitória, por sua vez, aparenta estar relacionada não apenas com o número de interações, especialmente ligações de hidrogênio, formadas entre o ligante e resíduos do sítio ativo, mas sobretudo com a identidade dessas interações.

É imprescindível que os inibidores formem ligação de hidrogênio com a His175 e com a Arg102, que são as duas ancoras principais para a fixação de ligantes no sítio ativo enzimático, a primeira relacionada com a complexação de grupos hidroxila na posição 3 do anel aromático dos inibidores, enquanto que a Arg102 atuando na complexação e estabilização eletrostática do grupo carboxilato.

É a impossibilidade de formar essas duas ligações de hidrogênio simultaneamente que tornam o ligante 37 um inibidor pouco eficaz, ainda que esse possa fazer em média 5,7 ligações de hidrogênio simultâneas, mais do que o ácido gálico (4,5 ligações), produto (4,8 ligações) e o próprio substrato (5,0 ligações).

A importância desses dois resíduos se comprova pelos resultados experimentais de mutagênese sítio específica, nos quais a substituição da His175 e da Arg102 por alanina produz enzimas mutantes totalmente inativas, tal como ocorre com a mutação da His144. Os resultados teóricos indicam, entretanto, que importância desses resíduos está relacionada com a complexação e posicionamento adequado do substrato para interação, ao invés de participação no mecanismo enzimático, como é o caso da His144 e da Tyr217.

Além da interação simultânea com a His175 e a Arg102, a formação de ligação de hidrogênio com a His144 aparenta ser determinante para a potência dos inibidores, mantendo-se ocupada em ligações de hidrogênio por cerca de 20% do tempo de simulação para o produto e o ácido gálico, entretanto, com possibilidade de formação de duas ligações de hidrogênio simultâneas com o inibidor mais potente, com o qual intensifica sua ocupação ao longo da simulação.

Destaca-se ainda, a formação de duas ligações de hidrogênio simultâneas com o esqueleto peptídico da Phe211, que ocorre com o ácido gálico, mas não com o produto, que pode formar apenas uma dessas ligações, contribuindo para a maior potência do ácido gálico como inibidor quando comparado ao produto.

A formação de outras ligações de hidrogênio com resíduos como o Trp35, Tyr70 e Tyr217 não aparenta ser determinante para o potencial inibitório dos ligantes, mas é um recurso que pode ser explorado no desenho de novos inibidores, visando uma melhora na seletividade e eventualmente tendo um efeito sinérgico sobre afinidade receptor-ligante, aumentando-a.

De fato, é a formação dessa rede de ligações de hidrogênio estáveis e sobretudo constantes com esses resíduos principais o que distingue os inibidores muito ativos dos falsos positivos de *docking*, que formam poucas ligações (caso do ligante 27) ou muitas ligações lábeis (caso do ligante 37).

Isso mostra que a metodologia de *docking* superestima a interação desses ligantes, ou em outras palavras, “força” algumas interações, ao permitir torções inadequadas de algumas ligações, como, por exemplo, a que resulta na perda de planaridade entre o grupo carboxílico e o anel aromático, o que não é energeticamente favorável na realidade em função da ressonância e deslocalização eletrônica entre esses grupos.

Além disso, esses falsos positivos são ligantes relativamente pequenos e estruturalmente muito similares ao produto, sendo melhores encaixados pelo *docking* no sítio ativo cristalográfico do que os demais ligantes, pois se adequam melhor à “fôrma” criada pelo produto enquanto os demais recebem uma penalização energética maior por colisões estéricas nesse sítio de ligação restrito – alguns nem mesmo cabendo nesse sítio – o que gera o elevado desvio e baixa correlação dos resultados de *docking* frente aos dados experimentais.

Diante disso, a fim de se aprimorar os resultados de *docking* e de se incluir a flexibilidade natural do receptor, fez-se o *ensemble docking* com cerca de 200 estruturas extraídas das simulações de dinâmica molecular, com diferentes estratégias de geração do *ensemble*.

Comparando-se os parâmetros de qualidade das curvas de correlação entre os dados experimentais e os resultados do *docking* dos 25 compostos em cada uma dessas conformações, mostrou que construir o *ensemble* pela escolha arbitrária do último frame das trajetórias pois ainda que os sistemas estejam equilibrados, alterações estruturais instantâneas levam a estrutura obtida a se desviar da conformação nativa, prejudicando os resultados de *docking*, visto que apenas 10% das estruturas assim geradas conseguiram produzir modelos lineares de correlação superiores aos da estrutura de partida.

A escolha das estruturas em que ocorrem a formação do maior número de ligações de hidrogênio simultâneas – em concordância com o modelo de encaixe-induzido – foi mais eficiente, com cerca de 50% das estruturas obtidas gerando parâmetros de qualidade superiores na correlação dos escores de *docking* com os dados experimentais.

A estratégia exaustiva, de seleção das estruturas mais relevantes através da clusterização completa das trajetórias por RMSD dos carbonos- α e RMSD do sítio ativo, se mostrou menos eficiente, na medida em que apenas 15% das estruturas produziram parâmetros de qualidade de correlação superiores ao cristalográfico, porém permitiu a obtenção das melhores estruturas, com coeficientes de correlação de Pearson de 0,53 e de 0,52, obtidas das simulações com os ligantes 27 e 37, respectivamente.

Esses ligantes, por não interagirem de forma tão efetiva com o sítio ativo, ao mesmo tempo em que também não alteram de forma tão drástica sua estrutura como ligantes muito volumosos ou a ausência total de ligantes, possivelmente permitem rearranjos das cadeias laterais dos aminoácidos do sítio, possibilitando o encaixe adequado de todos, ou quase todos os ligantes por *docking* e levando à melhora de correlação observada.

Os modos de ligação observados pelo *docking* nessas estruturas, semelhantes entre si, porém diferentes dos apontados pelo *docking* na estrutura cristalográfica sugerem que essas conformações correspondem a mínimos energéticos locais, suficientemente amplos para receber todos os ligantes e suficientemente estáveis para ter uma população significativa em solução, mas não a melhor conformação possível para cada ligante.

Assim, é possível que a partir dessas estruturas cada ligante induza o sítio a uma conformação que favoreça sua ligação, maximizando e intensificando suas interações, cada um evoluindo para um mínimo global distinto a partir dessa estrutura comum.

A verificação dessa hipótese também é uma das perspectivas futuras desse trabalho, e será feita por meio de novas simulações dos principais ligantes, partindo-se dos modos de ligação apontado pelo *docking* nessas estruturas e verificando-se a evolução de suas interações e da

estrutura do sítio, sobretudo para o substrato, produto e ácido gálico, cujo modo de ligação mais provável já é conhecido.

Finalmente, foi possível gerar um *ensemble* composto por três estruturas e que correlaciona de forma significativamente mais adequada as energias livres experimentais com a média dos escores de *docking* nessas estruturas.

Considerando-se todos os ligantes, foi possível aprimorar o coeficiente de correlação de Person de 0,36 na estrutura inicial para 0,60, bem como passar de um coeficiente de determinação 0,13 para 0,36 com significativa redução no desvio percentual entre os dois conjuntos de dados. Quando o número de ligantes é reduzido para os 10 mais ativos, a correlação aumenta para 0,90, com R^2 de 0,82, inclinação da curva de correlação de 0,8, e, portanto, muito próximo do ideal que é o valor unitário, com redução significativa do erro quadrático médio (0,25 kcal/mol).

Assim, com base em todas as conclusões aqui descritas, novas moléculas estão sendo desenhadas em busca de maior seletividade e potência na inibição de desidroshikimate desidratase, as quais serão testadas por meio de *docking* nesse *ensemble* de estruturas e terão seus potenciais inibitórios avaliados por meios desses modelos de correlação linear obtidos.

Através disso espera-se guiar a síntese de novos compostos com potencial ação *in vitro* na inibição dessa enzima e potencial aplicação *in vivo* contra o *Bacillus anthracis*, auxiliando no desenvolvimento de uma nova terapia contra esse patógeno extremamente ameaçador e mortal, que já ceifou muitas vidas no passado e poderá voltar a ceifar se nada for feito para detê-lo.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- (1) Organization, W. H. *Tackling drug-resistant infections globally: final report and recommendations. The review on antimicrobial resistance.*; Geneva, **2016**.
- (2) U.S. Centers for Disease Control and Prevention. *Antibiotic resistance threats in the United States, 2013*; Atlanta, **2013**.
- (3) Organization, W. H. *Antimicrobial resistance: global report on surveillance*; Geneva, **2014**.
- (4) Guimarães, K. Superbactérias avançam no Brasil e levam autoridades de saúde a correr contra o tempo. *BBC Brasil*. São Paulo julho 11, **2017**.
- (5) Ministério da Saúde (Brasil). Agência Nacional de Vigilância Sanitária (ANVISA). Boletim de Segurança do Paciente e Qualidade em Serviços de Saúde nº 14: Avaliação dos indicadores nacionais das Infecções Relacionadas à Assistência à Saúde (IRAS) e Resistência microbiana do ano de 2015. **2016**, 83.
- (6) Matos, E. Suspeita de bactéria fecha mais uma UTI Neonatal. *Zero Hora Edição Digital*. Porto Alegre maio 25, **2015**.
- (7) Hospital de Americana confirma fim de infecções por superbactéria KPC. *Portal G1*. Campinas março 14, **2017**.
- (8) A um mês da Olimpíada, CNN faz reportagem sobre superbactéria na orla. *O Globo*. Rio de Janeiro julho 5, **2016**.
- (9) Liu, Y. Y.; Wang, Y.; Walsh, T. R.; et al. Emergence of plasmid-mediated colistin resistance mechanism MCR-1 in animals and human beings in China: A microbiological and molecular biological study. *Lancet Infect. Dis.* **2016**, *16* (2), 161–168 DOI: 10.1016/S1473-3099(15)00424-7.
- (10) Skov, R. L.; Monnet, D. L. Plasmid-mediated colistin resistance (*mcr-1* gene): Three months later, the story unfolds. *Eurosurveillance* **2016**, *21* (9), 1–6 DOI: 10.2807/1560-7917.ES.2016.21.9.30155.
- (11) McGann, P.; Snesrud, E.; Maybank, R.; Corey, B.; Ong, A. C.; Clifford, R.; Hinkle, M.; Whitman, T.; Lesho, E.; Schaecher, K. E. *Escherichia coli* Harboring *mcr-1* and *bla*_{CTX-M} on a Novel IncF Plasmid: First report of *mcr-1* in the USA. *Antimicrob. Agents Chemother.* **2016**, *60* (May), AAC.01103-16 DOI: 10.1128/AAC.01103-16.
- (12) Fernandes, M. R.; McCulloch, J. A.; Vianello, M. A.; Moura, Q.; Pérez-Chaparro, P. J.; Esposito, F.; Sartori, L.; Dropa, M.; Matté, M. H.; Lira, D. P. A.; et al. First Report of the Globally Disseminated IncX4 Plasmid Carrying the *mcr-1* Gene in a Colistin-Resistant *Escherichia coli* ST101 isolated from a Human Infection in Brazil. *Antimicrob. Agents Chemother.* **2016**, *60* (August), AAC.01325-16 DOI: 10.1128/AAC.01325-16.
- (13) Hiller, S. Pan-Resistant New Delhi Metallo-Beta- Lactamase-Producing *Klebsiella pneumoniae* — Washoe County, Nevada, 2016. *Morb. Mortal. Wkly. Rep.* **2017**, *66* (1), 33 DOI: 10.15585/mmwr.mm6601a7.
- (14) Kontra, J. M. Zombie infections and other infectious disease complications of Global Warming. *J. Lancaster Gen. Hosp.* **2017**, *12* (1), 12–16.
- (15) Derretimento de solo congelado expõe ameaça de vírus e bactérias. *Portal G1*. agosto 13, **2016**.
- (16) Luhn, A. Anthrax outbreak triggered by climate change kills boy in Arctic Circle. *The Guardian*. Moscow agosto 1, **2016**.
- (17) Fasanella, A.; Galante, D.; Garofolo, G.; Jones, M. H. Anthrax undervalued zoonosis. *Vet. Microbiol.* **2010**, *140* (3–4), 318–331 DOI: 10.1016/j.vetmic.2009.08.016.
- (18) Spencer, R. C. *Bacillus anthracis*. *J. Clin. Pathol.* **2003**, *56* (1), 182–187 DOI: 10.1136/jcp.56.3.182.
- (19) Kyriacou, D. N.; Adamski, A.; Khardori, N. Anthrax: From Antiquity and Obscurity to a

- Front-Runner in Bioterrorism. *Infect. Dis. Clin. North Am.* **2006**, *20* (2), 227–251 DOI: 10.1016/j.idc.2006.03.009.
- (20) O Livro de Êxodo. In *A Bíblia Sagrada: Antigo e Novo Testamento. Tradução de João Ferreira de Almeida*; Sociedade Bíblica do Brasil: Brasília, **1969**.
- (21) Hugh-Jones, M. E.; Vos, V. de; de Vos, V. Anthrax and wildlife. *Rev. Sci. Tech. - Off. Int. des Epizoot.* **2002**, *21* (2), 359–383.
- (22) Schild, A. L.; Sallis, E. S. V.; Soares, M. P.; Ladeira, S. R. L.; Schramm, R.; Priebe, A. P.; Almeida, M. B.; Riet-Correa, F. Anthrax in cattle in southern Brazil: 1978-2006. *Pesqui. Vet. Bras.* **2006**, *26* (4), 243–248 DOI: 10.1590/S0100-736X2006000400009.
- (23) Langenegger, J. Ocorrência do carúnculo hemático em animais no Brasil. *Pesq. Vet. Bras* **1994**, *14* (4), 135–136.
- (24) Turner, M. Anthrax in Humans in Zimbabwe. *Cent. Afr. J. Med.* **1980**, *26* (7), 161–162.
- (25) Cieslak, T. J.; Eitzen, E. M. Clinical and epidemiologic principles of anthrax. *Emerg. Infect. Dis.* **1999**, *5* (4), 552–555 DOI: 10.3201/eid0504.990418.
- (26) Powell, A. G. M. T.; Crozier, J. E. M.; Hodgson, H.; Galloway, D. J. A case of septicaemic anthrax in an intravenous drug user. *BMC Infect. Dis.* **2011**, *11* (21), 6 DOI: 10.1053/j.gastro.2011.02.076.
- (27) Hicks, C. W.; Clinic, C.; College, L.; Sweeney, D. a. An overview of anthrax infection including the recently identified form of disease in injection drug users. *Intensive Care Med* **2012**, *38* (7), 1092–1104 DOI: 10.1007/s00134-012-2541-0.
- (28) Gould, G. W. Recent Advances in the Understanding of Resistance and Dormancy in Bacterial Spores Characteristics of Bacterial Endospores. *J. Appl. Bacteriol.* **1977**, *42*, 297–309.
- (29) Russell, A. D. Bacterial spores and chemical sporicidal agents. *Clin. Microbiol. Rev.* **1990**, *3* (2), 99–119 DOI: 10.1016/0195-6701(83)90071-3.
- (30) Sweeney, D. A.; Hicks, C. W.; Cui, X.; Li, Y.; Eichacker, P. Q. Anthrax infection. *Am. J. Respir. Crit. Care Med.* **2011**, *184* (12), 1333–1341 DOI: 10.1164/rccm.201102-0209CI.
- (31) Bergman, N. H. *Bacillus anthracis and Anthrax*, 1st ed; John Wiley & Sons: Hoboken, **2011**.
- (32) Beierlein, J. M.; Anderson, A. C. New developments in vaccines, inhibitors of anthrax toxins, and antibiotic therapeutics for *Bacillus anthracis*. *Curr. Med. Chem.* **2011**, *18* (33), 5083–5094 DOI: 10.2174/092986711797636036.
- (33) Langston, S. M. The Threat Of Domestic Bioterroism: Fact Or Fiction?, The University of Texas at El Paso, **2013**.
- (34) Jernigan, D. B.; Raghunathan, P. L.; Bell, B. P.; et al. Investigation of bioterrorism-related anthrax, United States, 2001: Epidemiologic findings. *Emerg. Infect. Dis.* **2002**, *8* (10), 1019–1028 DOI: 10.3201/eid0810.020353.
- (35) World Health Organization. *Health Aspects of Chemical and Biological Weapons*; Geneva, **1970**.
- (36) Guidi-Rontani, C. Challenges of Anthrax Drug Design: In Search of “Achilles’ heel” of *Bacillus anthracis*. *Lett. Drug Des. Discov.* **2004**, *1* (2), 110–114 DOI: 10.2174/1570180043485590.
- (37) Helgason, E.; Økstad, O. A.; Dominique, A.; Johansen, H. et al. *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis* — One Species on the Basis of Genetic Evidence One Species on the Basis of Genetic Evidence. *Appl. Environ. Microbiol.* **2000**, *66* (6), 2627–2630 DOI: 10.1128/AEM.66.6.2627-2630.2000.
- (38) Moayeri, M.; Leppla, S. H. The roles of anthrax toxin in pathogenesis. *Curr. Opin. Microbiol.* **2004**, *7* (1), 19–24 DOI: 10.1016/j.mib.2003.12.001.
- (39) Negus, D.; Burton, J.; Sweed, A.; Gryko, R.; Taylor, P. W. Poly- γ -D-glutamic acid capsule interferes with lytic infection of *Bacillus anthracis* by *B. anthracis*-specific bacteriophages. *Appl. Environ. Microbiol.* **2013**, *79* (2), 714–717 DOI: 10.1128/AEM.02682-12.

- (40) Feld, G. K.; Thoren, K. L.; Kintzer, A. F.; Sterling, H. J.; Tang II; Greenberg, S. G.; Williams, E. R.; Krantz, B. A. Structural basis for the unfolding of anthrax lethal factor by protective antigen oligomers. *Nat Struct Mol Biol* **2010**, *17* (11), 1383–1390 DOI: 10.1038/nsmb.1923.
- (41) Coggeshall, K. M.; Lupu, F.; Ballard, J.; Metcalf, J. P.; James, J. A.; Farris, D.; Kurosawa, S. The sepsis model: An emerging hypothesis for the lethality of inhalation anthrax. *J. Cell. Mol. Med.* **2013**, *17* (7), 914–920 DOI: 10.1111/jcmm.12075.
- (42) Friedlander, A. M.; Collins, T. Tackling anthrax Green chemistry through the mill. *Nature* **2001**, *414* (November), 160–161 DOI: 10.1038/35102660.
- (43) Welkos, S. L. Plasmid-associated virulence factors of non-toxigenic (pX01-) *Bacillus anthracis*. *Microb. Pathog.* **1991**, *10* (3), 183–198 DOI: 10.1016/0882-4010(91)90053-D.
- (44) Athamna, A.; Athamna, M.; Abu-Rashed, N.; Medlej, B.; Bast, D. J.; Rubinstein, E. Selection of *Bacillus anthracis* isolates resistant to antibiotics. *J. Antimicrob. Chemother.* **2004**, *54* (2), 424–428 DOI: 10.1093/jac/dkh258.
- (45) Lamb, A. L. Breaking a pathogen's iron will: Inhibiting siderophore production as an antimicrobial strategy. *Biochim. Biophys. Acta* **2015**, *1854* (8), 1054–1070 DOI: 10.1016/j.bbapap.2015.05.001.
- (46) Andrews, S. C.; Robinson, A. K.; Rodríguez-Quiñones, F. Bacterial iron homeostasis. *FEMS Microbiology Reviews*. **2003**, p 215–237.
- (47) Boukhalfa, H.; Crumbliss, A. L. Chemical aspects of siderophore mediated iron transport. *BioMetals* **2002**, *15* (4), 325–339 DOI: 10.1023/A:1020218608266.
- (48) Wilson, M. K.; Abergel, R. J.; Raymond, K. N.; Arceneaux, J. E. L.; Byers, B. R. Siderophores of *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. *Biochem. Biophys. Res. Commun.* **2006**, *348* (1), 320–325 DOI: 10.1016/j.bbrc.2006.07.055.
- (49) Cendrowski, S.; MacArthur, W.; Hanna, P. *Bacillus anthracis* requires siderophore biosynthesis for growth in macrophages and mouse virulence. *Mol. Microbiol.* **2004**, *51* (2), 407–417 DOI: 10.1046/j.1365-2958.2003.03861.x.
- (50) Koppisch, A. T.; Browder, C. C.; Moe, A. L.; Shelley, J. T.; Kinkel, B. A.; Hersman, L. E.; Iyer, S.; Ruggiero, C. E. Petrobactin is the primary siderophore synthesized by *Bacillus anthracis* str. Sterne under conditions of iron starvation. *BioMetals* **2005**, *18* (6), 577–585 DOI: 10.1007/s10534-005-1782-6.
- (51) Zhang, G.; Amin, S. A.; Küpper, F. C.; Holt, P. D.; Carrano, C. J.; Butler, A. Ferrie stability constants of representative marine siderophores: Marinobactins, aquachelins, and petrobactin. *Inorg. Chem.* **2009**, *48* (23), 11466–11473 DOI: 10.1021/ic901739m.
- (52) Abergel, R. J.; Zawadzka, A. M.; Raymond, K. N. Petrobactin-mediated iron transport in pathogenic bacteria: Coordination chemistry of an unusual 3,4-catecholate/citrate siderophore. *J. Am. Chem. Soc.* **2008**, *130* (7), 2124–2125 DOI: 10.1021/ja077202g.
- (53) Barbeau, K.; Zhang, G.; Live, D. H.; Butler, A. Petrobactin, a photoreactive siderophore produced by the oil-degrading marine bacterium *Marinobacter hydrocarbonoclasticus*. *J. Am. Chem. Soc.* **2002**, *124* (3), 378–379 DOI: 10.1021/ja0119088.
- (54) Hotta, K.; Kim, C.; Fox, D. T.; Koppisch, A. T. Siderophore-mediated iron acquisition in *Bacillus anthracis* and related strains. *Microbiology* **2010**, *156*, 1918–1925 DOI: 10.1099/mic.0.039404-0.
- (55) Fischbach, M. A.; Lin, H.; Liu, D. R.; Walsh, C. T. How pathogenic bacteria evade mammalian sabotage in the battle for iron. *Nat. Chem. Biol.* **2006**, *2* (3), 132–138 DOI: 10.1038/nchembio771.
- (56) Lee, J. Y.; Janes, B. K.; Passalacqua, K. D.; Pflieger, B. F.; Bergman, N. H.; Liu, H.; Håkansson, K.; Somu, R. V.; Aldrich, C. C.; Cendrowski, S.; et al. Biosynthetic Analysis of the Petrobactin Siderophore Pathway from. **2007**, *189* (5), 1698–1710 DOI: 10.1128/JB.01526-06.
- (57) Fox, D. T.; Hotta, K.; Kim, C. Y.; Koppisch, A. T. The missing link in petrobactin

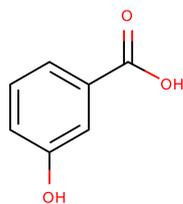
- biosynthesis: asbF encodes a (-)-3-dehydroshikimate dehydratase. *Biochemistry* **2008**, *47* (47), 12251–12253 DOI: 10.1021/bi801876q.
- (58) Pflieger, B. F.; Kim, Y.; Nusca, T. D.; Maltseva, N.; Lee, J. Y.; Rath, C. M.; Scaglione, J. B.; Janes, B. K.; Anderson, E. C.; Bergman, N. H.; et al. Structural and functional analysis of AsbF: origin of the stealth 3,4-dihydroxybenzoic acid subunit for petrobactin biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (44), 17133–17138 DOI: 10.1073/pnas.0808118105.
- (59) Voet, D.; Voet, J. *Fundamentos de Bioquímica - A Vida em Nível Molecular*, 4^o ed; Artmed: Porto Alegre, **2013**.
- (60) Devlin, T. M. *Textbook of Biochemistry with Clinical Correlations*, 5^o ed; John Wiley & Sons: New York, **2002**.
- (61) Nelson, D. L.; Cox, M. M. *Princípios de Bioquímica de Lehninger*, 6^o ed; Artmed: Porto Alegre, **2014**; Vol. 0.
- (62) Smith, C.; Marks, A. D.; Lieberman, M. *Marks' basic medical biochemistry: A clinical approach.*, 2^o ed; Lippincotts Williams & Wilkins: Baltimore, **2006**.
- (63) Berg, J. M.; Tymoczko, J. L.; Stryer, L. *Biochemistry A Short Course*, 2^o ed; W. H. Freeman and Company: New York, **2010**.
- (64) Bugg, T. D. H. *Introduction to Enzyme and Coenzyme Chemistry: Second Edition*, 3^o ed; John Wiley & Sons: West Sussex, **2012**.
- (65) Verli, H.; Barreiro, E. J. Um paradigma da química medicinal: A flexibilidade dos ligantes e receptores. *Quim. Nova* **2005**, *28* (1), 95–102 DOI: 10.1590/S0100-40422005000100018.
- (66) Merz, K. M.; Ringe, D.; Reynolds, C. H. *Drug Design: Structure and Ligand-Based Approaches*, 1^o ed; Cambridge University Press: New York, **2010**.
- (67) Copeland, R. A. *Evaluation of enzyme inhibitors in drug discovery: a guide for medicinal chemists and pharmacologists*, 2^o ed; John Wiley & Sons: Hoboken, **2013**.
- (68) Wood, E. J. *Marks' basic medical biochemistry: A clinical approach (second edition).*; **2006**; Vol. 34.
- (69) Krogsgaard-Larsen, P.; Strømgaard, K.; Madsen, U. *Textbook of drug design and discovery.*, 4^o ed; CRC Press: Boca Raton, **2002**
- (70) De Ruiter, A.; Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* **2011**, *15* (4), 547–552 DOI: 10.1016/j.cbpa.2011.05.021.
- (71) Yung-Chi, C.; Prusoff, W. H. Relationship between the inhibition constant (KI) and the concentration of inhibitor which causes 50 per cent inhibition (I50) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22* (23), 3099–3108 DOI: 10.1016/0006-2952(73)90196-2.
- (72) Morgon, N.; Coutinho, K. *Métodos de Química Teórica e Modelagem Molecular*, 1^o ed; Livraria da Física: São Paulo, **2007**.
- (73) Meng, X.-Y.; Zhang, H.-X.; Mezei, M.; Cui, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **2011**, *7* (2), 146–157 DOI: 10.2174/157340911795677602.
- (74) Leach, A. R. *Molecular Modelling: Principles and Applications*, 2^o ed; Prentice Hall: Harlow, **2001**.
- (75) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3* (11), 935–949 DOI: 10.1038/nrd1549.
- (76) Trott, O.; Olson, A. J. AutoDock Vina. *J. Comput. Chem.* **2010**, *31*, 445–461 DOI: 10.1002/jcc.21334.
- (77) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem.*

- Soc.* **1995**, 117 (19), 5179–5197 DOI: 10.1021/ja00124a002.
- (78) Frenkel, D.; Smit, B. *Understanding molecular simulation: from algorithms to applications*, 2^o ed; Academic Press: New York, **2002**.
- (79) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, 4 (2), 187–217 DOI: 10.1002/jcc.540040211.
- (80) Oostenbrink, C.; Villa, A.; Mark, A. E.; Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *J. Comput. Chem.* **2004**, 25 (13), 1656–1676 DOI: 10.1002/jcc.20090.
- (81) Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining *docking* and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, 26 (5), 531–568 DOI: 10.1002/med.20067.
- (82) Huang, S.; Zou, X. *Ensemble docking* of Multiple Protein Structures : Considering Protein Structural Variations in Molecular *Docking*. **2007**, 421 (August 2006), 399–421 DOI: 10.1002/prot.
- (83) Carlson, H. a; McCammon, J. a. Accommodating protein flexibility in computational drug design. *Mol. Pharmacol.* **2000**, 57 (2), 213–218 DOI: 0026-895X/00/020213-06\$3.00/0.
- (84) Huang, S. Y.; Zou, X. *Ensemble docking* of multiple protein structures: Considering protein structural variations in molecular *docking*. *Proteins Struct. Funct. Genet.* **2007**, 66 (2), 399–421 DOI: 10.1002/prot.21214.
- (85) Baron, R. Computational Drug Discovery and Design. *Methods Mol. Biol.* **2012**, 819 (2), 3–12 DOI: 10.1007/978-1-61779-465-0_20.
- (86) Gohlke, H. Target Flexibility: An Emerging Consideration in Drug Discovery and Design †. *J. Med. Chem.* **2008**, 51 (20), 6237–6255.
- (87) Gohlke, H. Target Flexibility: An Emerging Consideration in Drug Discovery and Design. *J. Med. Chem.* **2008**, 51 (20), 6237–6255.
- (88) Simon, I. A. Desenvolvimento de inibidores da biossíntese de petrobactina: deidrosnikimato desidratase como alvo de terapias anti-antraz, Universidade Federal do Rio Grande do Sul, **2015**.
- (89) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, 7 (2), 525–537 DOI: 10.1021/ct100578z.
- (90) ChemAxon. Marvin Beans Suite. ChemAxon Ltd.: Budapest **2015**.
- (91) Dennington, R.; Keith, T.; Millam, J. GaussView 5. Semichem Inc.: Shawnee Mission KS **2009**.
- (92) Frisch M. J., Trucks G. W., Schlegel H. B., Scuseria G. E., Robb M. A., C. J.; R., S. G. et al. Gaussian 09. Gaussian, Inc.: Wallingford, CT **2009**.
- (93) O’Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open chemical toolbox. *J. Cheminform.* **2011**, 3 (10) DOI: 10.1186/1758-2946-3-33.
- (94) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. Software news and updates AutoDock4 and AutoDockTools4: Automated *docking* with selective receptor flexibility. *J. Comput. Chem.* **2009**, 30 (16), 2785–2791 DOI: 10.1002/jcc.21256.
- (95) Berendsen, H. J. C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, 91 (1–3), 43–56 DOI: 10.1016/0010-4655(95)00042-E.
- (96) Pronk, S.; Páll, S.; Schulz, R.; Larsson, P.; Van Der Spoel, D.; et al. GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **2013**, 29 (7), 845–854 DOI: 10.1093/bioinformatics/btt055.

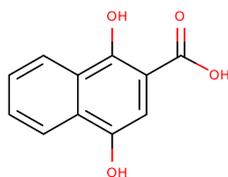
- (97) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*. **2005**, p 1701–1718.
- (98) Duan, Y.; Wu, C.; Chowdhury, S. S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; et al. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24* (16), 1999–2012 DOI: 10.1002/jcc.10349.
- (99) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105* (43), 9954–9960 DOI: 10.1021/jp003020w.
- (100) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interface. *BMC Res. Notes* **2012**, *5* (1), 367 DOI: 10.1186/1756-0500-5-367.
- (101) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25* (9), 1157–1174 DOI: 10.1002/jcc.20035.
- (102) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J. Comput. Chem.* **2002**, *23* (16), 1623–1641 DOI: 10.1002/jcc.10128.
- (103) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126* (1) DOI: 10.1063/1.2408420.
- (104) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Chem. Phys.* **2002**, *71*82 (May 2015) DOI: 10.1063/1.328693.
- (105) Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98* (12), 10089 DOI: 10.1063/1.464397.
- (106) Hess, B.; Spoel, D. van der; Lindahl, E. Gromacs User Manual. *Dep. Biophys. Chem.* ... **2014**, 312 DOI: 10.1006/imms.1993.1022.
- (107) Steinbrecher, T.; Labahn, A. Towards accurate free energy calculations in ligand protein-binding studies. *Curr. Med. Chem.* **2010**, *17* (8), 767–785 DOI: 10.2174/092986710790514453.
- (108) Wang, Z.; Sun, H.; Yao, X.; Li, D.; Xu, L.; Li, Y.; Tian, S.; Hou, T. Comprehensive evaluation of ten docking programs on a diverse set of protein-ligand complexes: the prediction accuracy of sampling power and scoring power. *Phys. Chem. Chem. Phys.* **2016**, *18*, 12964–12975 DOI: 10.1039/c6cp01555g.
- (109) Durrant, J. D.; McCammon, J. A. BINANA: A novel algorithm for ligand-binding characterization. *J. Mol. Graph. Model.* **2011**, *29* (6), 888–893 DOI: 10.1016/j.jmgm.2011.01.004.
- (110) Martinez, L. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. *PLoS One* **2015**, *10* (3), 1–10 DOI: 10.1371/journal.pone.0119264.
- (111) Leimkuhler, B.; Matthews, C. *Molecular Dynamics with Deterministic and Stochastic Numerical Methods*, 1^o ed; Springer: Cham, **2015**.
- (112) Soares, R. ReadHBMap. Faculty of Pharmaceutical Sciences at Ribeirao Preto - University of Sao Paulo - Brazil: São Paulo **2014**.
- (113) Hubbard, R. E.; Kamran Haider, M. Hydrogen bonds in proteins: role and strength. *eLS* **2001**, *1* (February), 1–6 DOI: 10.1002/9780470015902.a0003011.pub2.
- (114) Harrington, L. B.; Jha, R. K.; Canales, G. M.; Finney, K. B.; Koppisch, A. T.; Fox, D. T. et al. Rapid Thermostabilization of *Bacillus thuringiensis* Serovar Konkukian 97-27 Dehydroshikimate Dehydratase through a Structure-Based Enzyme Design and Whole Cell Activity Assay. *ACS Synth. Biol.* **2017**, *6* (1), 120–129 DOI: 10.1021/acssynbio.6b00159.

APÊNDICES E ANEXOS

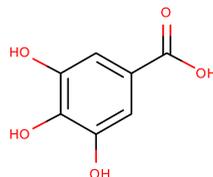
Apêndice A - Estrutura química e códigos adotados para os potenciais inibidores estudados experimentalmente na inibição de deidrosikimato desidratase de *B. anthracis*.



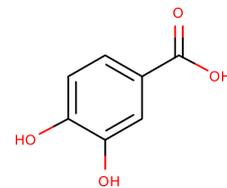
(1)



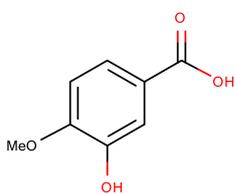
(2)



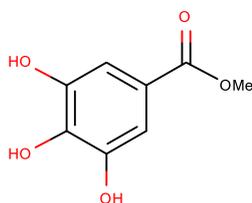
(4)



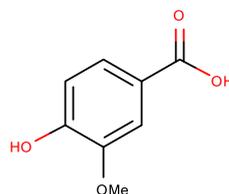
(5)



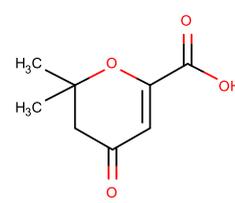
(6)



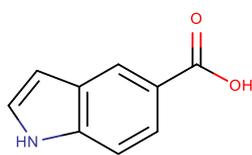
(7)



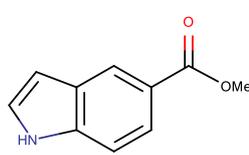
(8)



(9)



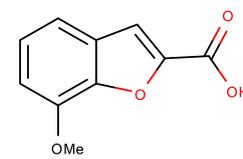
(10)



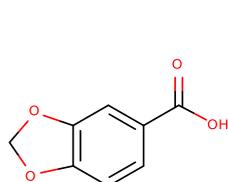
(11)



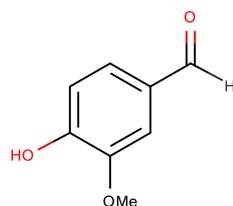
(12)



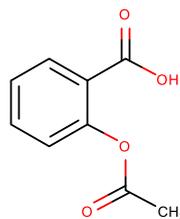
(13)



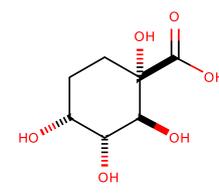
(14)



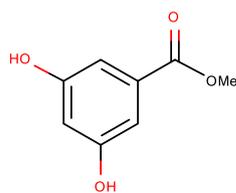
(15)



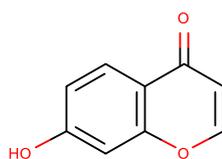
(16)



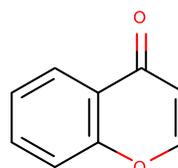
(17)



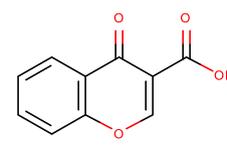
(18)



(19)

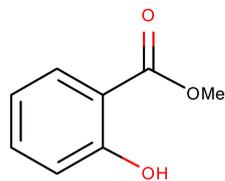


(20)

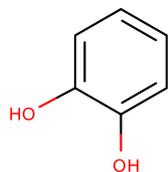


(21)

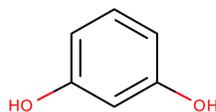
Apêndice A **Cont.** - Estrutura química e códigos adotados para os potenciais inibidores estudados experimentalmente na inibição de deidroschikimato desidratase de *B. anthracis*.



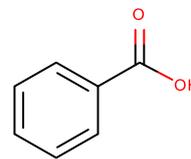
(22)



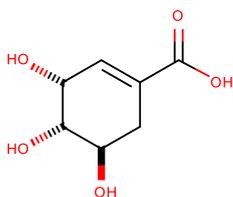
(23)



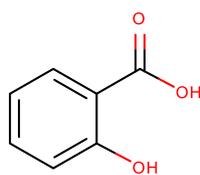
(24)



(25)



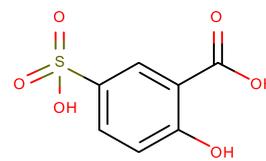
(26)



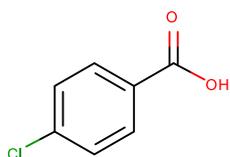
(27)



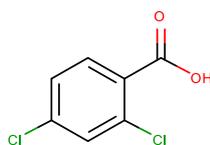
(28)



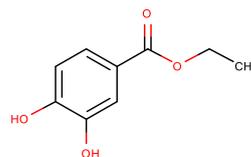
(29)



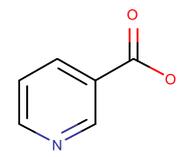
(30)



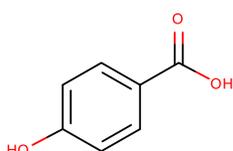
(31)



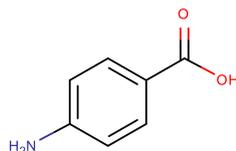
(32)



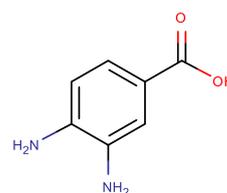
(33)



(34)



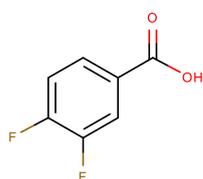
(35)



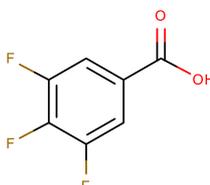
(36)



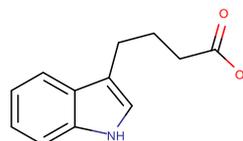
(37)



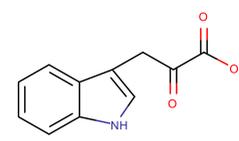
(38)



(39)

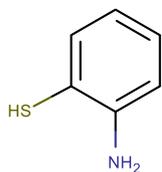


(40)

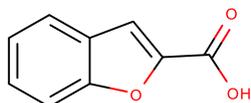


(41)

Apêndice A **Cont.** - Estrutura química e códigos adotados para os potenciais inibidores estudados experimentalmente na inibição de deidroshikimato desidratase de *B. anthracis*.



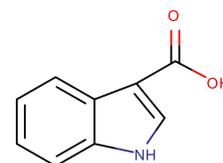
(42)



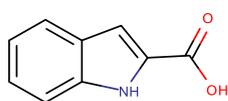
(43)



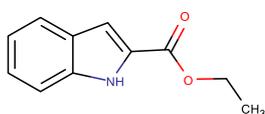
(44)



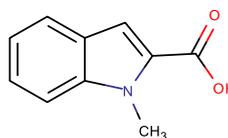
(45)



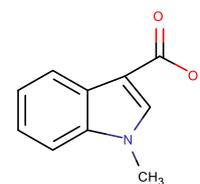
(46)



(47)



(48)



(49)

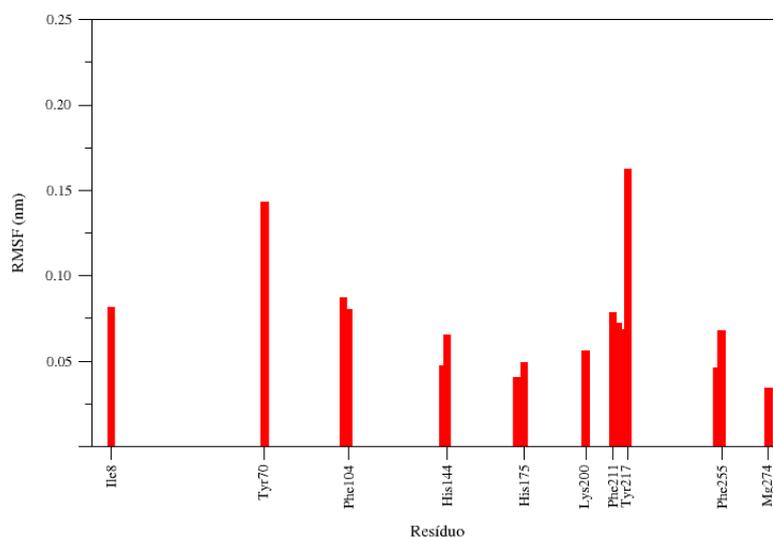
Apêndice B - Parâmetros detalhados aplicados nas etapas de produção das trajetórias de dinâmica molecular

```
cpp = /usr/bin/cpp
constraints = all-bonds
integrator = md
tinit = 0
dt = 0.002
nsteps = 50000000 ; total 100ns.
nstcomm = 10
nstxout = 5000
nstvout = 0
nstfout = 0
nstlog = 500
nstenergy = 100
energygrps = Protein SOL LIG MG NA CL
nstlist = 10
ns_type = grid
rlist = 1.3
coulombtype = PME
rcoulomb = 1.3
vdwtype = cut-off
rvdw = 1.3
fourierspacing = 0.12
fourier_nx = 0
fourier_ny = 0
fourier_nz = 0
pme_order = 4
ewald_rtol = 1e-5
optimize_fft = yes
; V-rescale temperature coupling is on three groups
Tcoupl = V-rescale
tau_t = 0.1 0.1 0.1 0.1 0.1 0.1
tc_grps = Protein SOL LIG MG NA CL
ref_t = 303 303 303 303 303 303
; Pressure coupling is on
Pcoupl = parrinello-rahman
pcoupltype = isotropic
tau_p = 0.5
compressibility = 4.5e-5
ref_p = 1.0
; Generate velocities is on at 303 K.
gen_vel = yes
gen_temp = 303.0
gen_seed = 173529
```

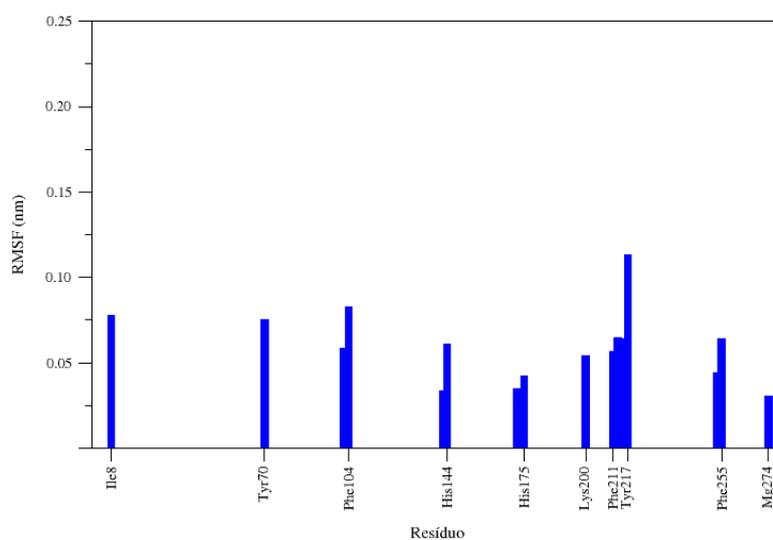
Apêndice C - Identidade dos clusters principais gerados considerando-se os carbonos- α (CA) e os resíduos do sítio ativo (SA) na construção da matriz de RMSD.

Sistema	Clusters Principais CA	Clusters Principais SA
Enzima Livre S1	109, 627, 1205	6, 20, 81, 408, 461
Enzima Livre S1	1, 413, 593, 870	25, 321, 802, 973
Enzima Livre S1	54, 541, 836, 948	1, 125, 264
Enzima Livre S1	1, 85, 424	1, 52, 366
Enzima + Substrato R1	1, 446, 582	57, 144, 268, 431, 600, 649
Enzima + Substrato R2	1, 139, 233, 613, 671, 702	18, 141, 764, 784
Enzima + Substrato R3	2, 166	43, 59
Enzima + Produto R1	1, 64	164, 187, 499, 654, 657, 817
Enzima + Produto R2	71, 186	10, 25, 383, 1440, 1682
Enzima + Produto R3	1, 258, 727	11, 772, 830, 1156
Enzima + Ligante 04 R1	149, 344, 664, 794, 917, 995	19, 284, 1014
Enzima + Ligante 04 R2	1, 158, 255, 342	19, 119
Enzima + Ligante 04 R3	1, 52, 167, 252, 287	49, 55, 143, 203, 474, 613
Enzima + Ligante 13 R1	1, 100, 274, 628	25, 40, 94, 194, 232
Enzima + Ligante 13 R2	1, 199, 314, 415	3, 9, 13, 31, 205, 264, 329, 375
Enzima + Ligante 27 R1	1, 219	6, 192, 212
Enzima + Ligante 27 R2	33, 387	6, 34, 150
Enzima + Ligante 37 R1	2, 56	58, 103, 304, 314
Enzima + Ligante 37 R2	1, 116, 183	21, 211, 340, 364
Enzima + Ligante 46 R1	1, 258, 386, 607, 745	14, 91, 512, 866, 903
Enzima + Ligante 46 R2	1, 75, 189	10, 74, 159, 329, 429, 511, 648

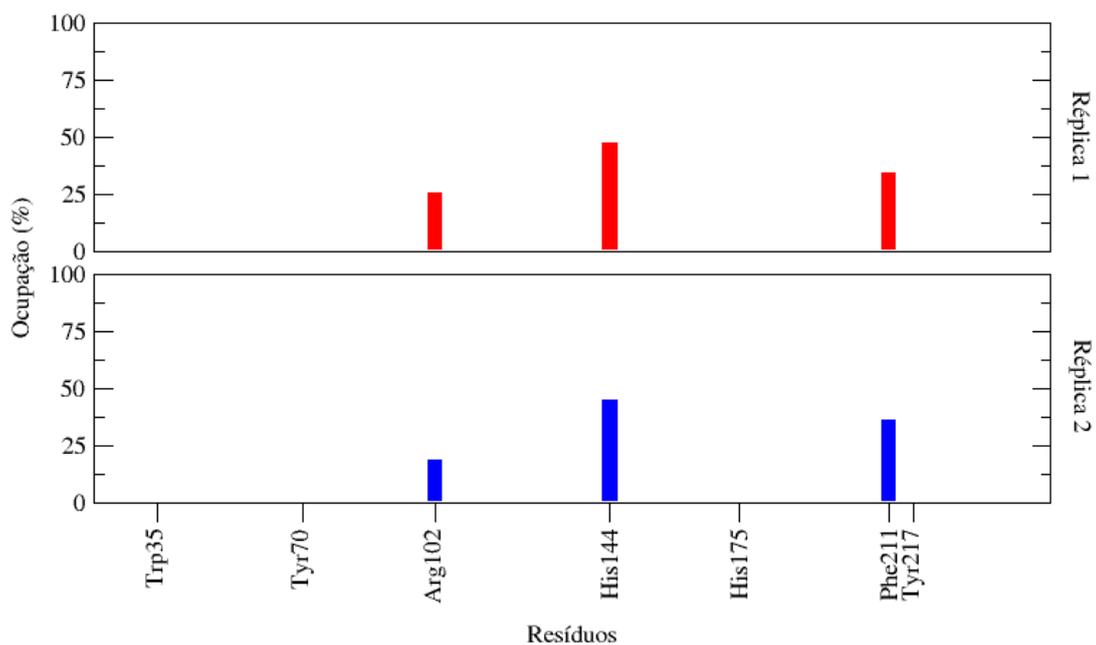
Apêndice D - Flutuação quadrática média dos resíduos do sítio ativo para o complexo enzima-ligante 37. Valor médio das duas réplicas.



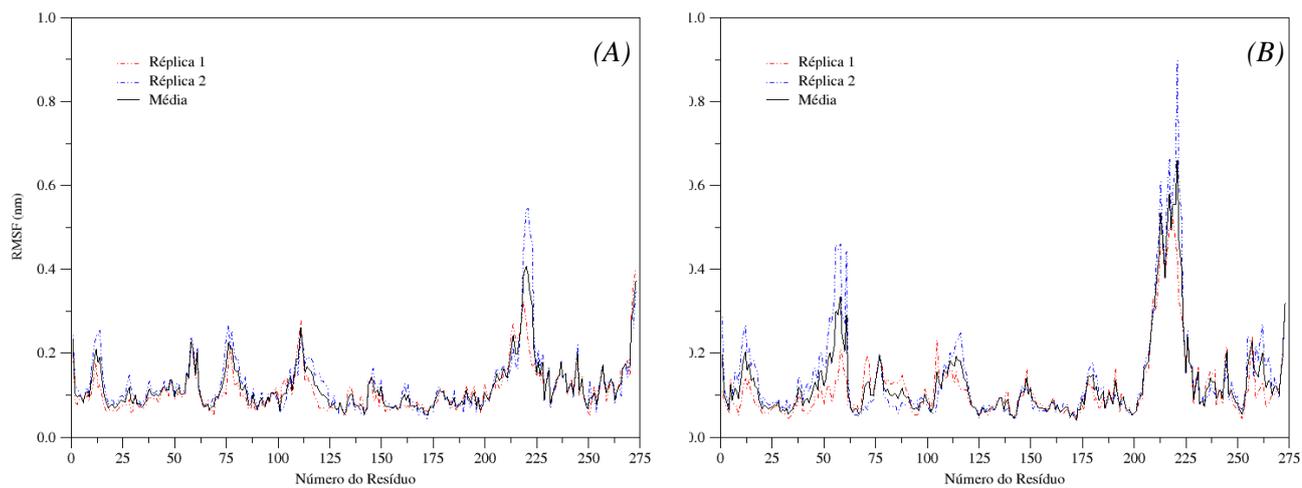
Apêndice E - Flutuação quadrática média dos resíduos do sítio ativo para o complexo enzima-ligante 37. Valor médio das duas réplicas.



Apêndice F - Ocupação dos resíduos em ligações de hidrogênio para as duas réplicas da simulação da enzima em complexo com o ligante 27.



Apêndice G – Flutuação quadrática média dos resíduos ao longo das simulações da enzima em complexo com ligante 13 (A) e com ligante 46 (B) e respectivas médias.



Apêndice H - Parâmetros da correlação linear entre as energias livres experimentais e os resultados de *docking* com a estruturas principais obtidas pela clusterização por RMSD dos Carbonos- α , em ordem crescente de coeficiente de correlação de Pearson.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima Livre S3 CLS-836	-0.2204	0.0486	-0.136	0.428
Enzima + Ligante 46 R1 CLS-607	-0.0464	0.0022	-0.024	0.371
Enzima + Ligante 46 R2 CLS-1	-0.0422	0.0018	-0.026	0.430
Enzima + Ligante 13 R2 CLS-199	-0.0242	0.0006	-0.015	0.440
Enzima Livre S3 CLS-948	-0.0138	0.0002	-0.010	0.498
Enzima Livre S2 CLS-1	-0.0109	0.0001	-0.007	0.482
Enzima Livre S3 CLS-54	-0.0076	0.0001	-0.005	0.422
Enzima + Ligante 37 R2 CLS-183	-0.0063	0.0000	-0.007	0.768
Enzima Livre S2 CLS-870	-0.0061	0.0000	-0.004	0.431
Enzima + Ligante 13 R1 CLS-1	0.0144	0.0002	0.012	0.578
Enzima + Ligante 46 R1 CLS-386	0.0181	0.0003	0.010	0.394
Enzima Livre S4 CLS-85	0.0354	0.0013	0.026	0.524
Enzima Livre S3 CLS-541	0.0455	0.0021	0.024	0.370
Enzima + Ligante 37 R2 CLS-1	0.0594	0.0035	0.080	0.962
Enzima Livre S1 CLS-627	0.0686	0.0047	0.050	0.519
Enzima + Ligante 37 R1 CLS-56	0.0693	0.0048	0.046	0.468
Enzima + Ligante 27 R1 CLS-219	0.0775	0.0060	0.060	0.545
Enzima + Ligante 13 R1 CLS-628	0.0893	0.0080	0.052	0.413
Enzima + Ligante 13 R1 CLS-100	0.1008	0.0102	0.054	0.379
Enzima Livre S2 CLS-413	0.1104	0.0122	0.059	0.375
Enzima + Ligante 46 R2 CLS-75	0.1114	0.0124	0.064	0.406
Enzima + Ligante 04 R2 CLS-1	0.1225	0.0150	0.073	0.419
Enzima Livre S1 CLS-109	0.1234	0.0152	0.074	0.423
Enzima + Ligante 27 R1 CLS-1	0.1493	0.0223	0.117	0.552
Enzima Livre S4 CLS-424	0.1542	0.0238	0.070	0.319
Enzima + Ligante 46 R1 CLS-745	0.1561	0.0244	0.084	0.376
Enzima + Ligante 13 R1 CLS-274	0.1768	0.0313	0.128	0.504
Enzima + Ligante 04 R1 CLS-149	0.1794	0.0322	0.171	0.668
Enzima + Ligante 04 R1 CLS-917	0.1797	0.0323	0.167	0.651
Enzima + Ligante 46 R1 CLS-258	0.1858	0.0345	0.117	0.439
Enzima + Substrato R3 CLS-2	0.1865	0.0348	0.142	0.530
Enzima Livre S4 CLS-1	0.2118	0.0449	0.105	0.343
Enzima + Ligante 04 R1 CLS-344	0.2334	0.0545	0.277	0.820
Enzima + Produto R1 CLS-1	0.2355	0.0555	0.315	0.923
Enzima + Ligante 46 R1 CLS-1	0.2392	0.0572	0.162	0.467
Enzima + Produto R2 CLS-186	0.2397	0.0575	0.202	0.581
Enzima + Ligante 27 R2 CLS-387	0.2438	0.0594	0.163	0.460
Enzima + Ligante 04 R3 CLS-252	0.2515	0.0633	0.163	0.446
Enzima + Ligante 04 R1 CLS-995	0.2629	0.0691	0.182	0.475
Enzima + Produto R2 CLS-71	0.2638	0.0696	0.169	0.439
Enzima Livre S2 CLS-593	0.2691	0.0724	0.104	0.264
Enzima + Ligante 37 R2 CLS-116	0.2701	0.0730	0.326	0.827
Enzima + Substrato R2 CLS-671	0.2795	0.0781	0.291	0.711
Enzima + Ligante 04 R2 CLS-342	0.2817	0.0793	0.290	0.701
Enzima + Ligante 04 R1 CLS-664	0.2840	0.0807	0.350	0.841
Enzima + Ligante 04 R3 CLS-167	0.2924	0.0855	0.173	0.402
Enzima + Substrato R2 CLS-139	0.2946	0.0868	0.268	0.618
Enzima + Produto R1 CLS-64	0.2949	0.0870	0.281	0.647
Enzima + Ligante 04 R2 CLS-158	0.2996	0.0898	0.255	0.576
Enzima + Ligante 04 R1 CLS-794	0.3042	0.0926	0.371	0.827
Enzima + Substrato R2 CLS-702	0.3101	0.0962	0.476	1.038
Enzima + Substrato R2 CLS-613	0.3107	0.0965	0.392	0.853
Enzima + Substrato R2 CLS-1	0.3130	0.0980	0.322	0.694

Apêndice H cont. – Parâmetros da correlação linear entre as energias livres experimentais e os resultados de *docking* com a estruturas principais obtidas pela clusterização por RMSD dos Carbonos- α , em ordem crescente de coeficiente de correlação de Pearson.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima + Ligante 13 R2 CLS-314	0.3140	0.0986	0.202	0.434
Enzima Livre S1 CLS-1205	0.3210	0.1030	0.235	0.492
Enzima + Produto R3 CLS-258	0.3266	0.1067	0.611	0.611
Enzima + Substrato R1 CLS-582	0.3325	0.1105	0.472	0.952
Enzima + Produto R3 CLS-727	0.3446	0.1187	0.459	0.888
Enzima + Substrato R2 CLS-233	0.3451	0.1191	0.418	0.809
Enzima + Ligante 04 R3 CLS-287	0.3524	0.1242	0.243	0.459
Enzima Inicial	0.3599	0.1295	0.364	0.671
Enzima + Ligante 04 R3 CLS-1	0.3646	0.1330	0.273	0.495
Enzima + Ligante 04 R3 CLS-52	0.3659	0.1339	0.383	0.692
Enzima + Ligante 13 R2 CLS-1	0.3728	0.1390	0.284	0.503
Enzima + Ligante 04 R2 CLS-255	0.3773	0.1423	0.390	0.680
Enzima + Substrato R1 CLS-446	0.3811	0.1452	0.364	0.628
Enzima + Substrato R1 CLS-1	0.3854	0.1486	0.339	0.576
Enzima + Ligante 27 R2 CLS-33	0.3909	0.1528	0.414	0.693
Enzima + Substrato R3 CLS-166	0.4006	0.1605	0.420	0.683
Enzima + Produto R3 CLS-1	0.4060	0.1648	0.506	0.809
Enzima + Ligante 46 R2 CLS-189	0.4234	0.1793	0.290	0.442
Enzima + Ligante 13 R2 CLS-415	0.4309	0.1857	0.245	0.365
Enzima + Ligante 37 R1 CLS-2	0.4386	0.1924	0.403	0.586

Apêndice I – Parâmetros da correlação linear entre as energias livres experimentais e os resultados de *docking* com a estruturas principais obtidas por clusterização por RMSD do sítio ativo.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima Livre S2 CLS-802	-0.2324	0.0540	-0.098	0.290
Enzima Livre S1 CLS-6	-0.2297	0.0528	-0.158	0.476
Enzima + Ligante 37 R2 CLS-211	-0.1626	0.0265	-0.128	0.551
Enzima + Ligante 46 R1 CLS-903	-0.1534	0.0235	-0.095	0.434
Enzima + Ligante 46 R1 CLS-512	-0.1261	0.0159	-0.076	0.427
Enzima Livre S2 CLS-973	-0.0717	0.0051	-0.028	0.272
Enzima Livre S1 CLS-461	-0.0664	0.0044	-0.054	0.571
Enzima + Ligante 46 R2 CLS-429	-0.0462	0.0021	-0.029	0.441
Enzima Livre S1 CLS-20	-0.0442	0.0020	-0.027	0.439
Enzima + Ligante 13 R1 CLS-25	-0.0368	0.0014	-0.017	0.327
Enzima + Ligante 46 R2 CLS-159	-0.0163	0.0003	-0.008	0.334
Enzima + Ligante 13 R1 CLS-40	-0.0088	0.0001	-0.005	0.415
Enzima Livre S3 CLS-264	-0.0020	0.0000	-0.001	0.347
Enzima + Ligante 13 R2 CLS-9	-0.0013	0.0000	-0.001	0.581
Enzima Livre S4 CLS-52	0.0156	0.0002	0.011	0.492
Enzima Livre S2 CLS-321	0.0165	0.0003	0.009	0.398
Enzima + Ligante 46 R2 CLS-329	0.0188	0.0004	0.010	0.393
Enzima + Ligante 13 R1 CLS-194	0.0192	0.0004	0.014	0.520
Enzima + Ligante 46 R1 CLS-866	0.0252	0.0006	0.012	0.334
Enzima + Ligante 13 R1 CLS-94	0.0252	0.0006	0.014	0.391
Enzima + Ligante 37 R2 CLS-340	0.0256	0.0007	0.030	0.825
Enzima Livre S3 CLS-125	0.0288	0.0008	0.014	0.356
Enzima Livre S4 CLS-366	0.0307	0.0009	0.015	0.340
Enzima Livre S4 CLS-1	0.0581	0.0034	0.045	0.551
Enzima + Ligante 13 R2 CLS-31	0.0880	0.0077	0.046	0.368
Enzima + Ligante 37 R1 CLS-304	0.1065	0.0114	0.107	0.711
Enzima + Ligante 27 R1 CLS-6	0.1099	0.0121	0.082	0.527
Enzima + Produto R3 CLS-11	0.1143	0.0131	0.101	0.625
Enzima + Ligante 13 R2 CLS-375	0.1159	0.0134	0.078	0.477
Enzima Livre S3 CLS-1	0.1287	0.0166	0.062	0.337
Enzima + Ligante 46 R2 CLS-511	0.1296	0.0168	0.085	0.464
Enzima + Ligante 37 R2 CLS-21	0.1324	0.0175	0.157	0.834
Enzima + Ligante 13 R1 CLS-232	0.1430	0.0205	0.091	0.447
Enzima + Produto R2 CLS-1682	0.1520	0.0231	0.141	0.650
Enzima Livre S1 CLS-81	0.1617	0.0261	0.102	0.444
Enzima + Ligante 37 R1 CLS-58	0.1630	0.0266	0.119	0.511
Enzima + Ligante 13 R2 CLS-329	0.1638	0.0268	0.108	0.460
Enzima + Ligante 27 R1 CLS-192	0.1683	0.0283	0.118	0.491
Enzima + Ligante 46 R1 CLS-91	0.1880	0.0354	0.100	0.373
Enzima + Ligante 46 R2 CLS-10	0.1954	0.0382	0.141	0.501
Enzima + Ligante 13 R2 CLS-264	0.2172	0.0472	0.153	0.489
Enzima + Produto R1 CLS-499	0.2202	0.0485	0.193	0.608
Enzima + Ligante 04 R1 CLS-1014	0.2402	0.0577	0.235	0.676
Enzima Livre S1 CLS-408	0.2443	0.0597	0.175	0.495
Enzima + Ligante 04 R1 CLS-284	0.2445	0.0598	0.312	0.878
Enzima + Ligante 37 R2 CLS-364	0.2477	0.0614	0.165	0.457
Enzima + Ligante 04 R3 CLS-474	0.2697	0.0727	0.163	0.414
Enzima + Produto R3 CLS-772	0.2699	0.0728	0.230	0.584
Enzima + Substrato R3 CLS-59	0.2706	0.0732	0.255	0.644
Enzima + Produto R2 CLS-25	0.2710	0.0734	0.253	0.638
Enzima + Produto R1 CLS-657	0.2712	0.0735	0.218	0.549

Apêndice I cont. – Parâmetros da correlação linear entre as energias livres experimentais e os resultados de *docking* com a estruturas principais obtidas pela clusterização por RMSD do sítio ativo, classificados segundo R-Pearson.

Estrutura	R-Pearson	COD	Inclinação	RMSE
Enzima + Ligante 27 R2 CLS-34	0.2736	0.0748	0.181	0.452
Enzima + Substrato R2 CLS-141	0.2812	0.0791	0.340	0.826
Enzima + Ligante 37 R1 CLS-314	0.2828	0.0800	0.205	0.494
Enzima + Ligante 46 R2 CLS-74	0.2839	0.0806	0.210	0.505
Enzima + Ligante 13 R2 CLS-3	0.2849	0.0812	0.202	0.484
Enzima + Substrato R1 CLS-144	0.2857	0.0816	0.197	0.470
Enzima + Substrato R3 CLS-43	0.2857	0.0816	0.270	0.643
Enzima + Ligante 04 R3 CLS-613	0.2862	0.0819	0.201	0.478
Enzima + Ligante 27 R1 CLS-212	0.2866	0.0822	0.195	0.464
Enzima + Ligante 13 R2 CLS-205	0.2887	0.0834	0.190	0.448
Enzima + Ligante 46 R2 CLS-648	0.2907	0.0845	0.172	0.402
Enzima + Ligante 46 R1 CLS-14	0.2908	0.0845	0.158	0.368
Enzima + Ligante 13 R2 CLS-13	0.2926	0.0856	0.206	0.479
Enzima + Produto R3 CLS-1156	0.3139	0.0985	0.372	0.799
Enzima + Substrato R1 CLS-431	0.3148	0.0991	0.376	0.806
Enzima + Produto R1 CLS-164	0.3162	0.1000	0.193	0.412
Enzima + Produto R2 CLS-1440	0.3215	0.1034	0.303	0.634
Enzima Livre S2 CLS-25	0.3257	0.1061	0.254	0.524
Enzima + Ligante 04 R2 CLS-119	0.3387	0.1147	0.251	0.495
Enzima + Substrato R2 CLS-764	0.3401	0.1157	0.595	1.168
Enzima + Produto R3 CLS-830	0.3438	0.1182	0.293	0.568
Enzima + Substrato R1 CLS-649	0.3525	0.1243	0.407	0.767
Enzima + Substrato R1 CLS-600	0.3548	0.1259	0.386	0.723
Enzima + Produto R1 CLS-187	0.3559	0.1266	0.369	0.689
Enzima + Substrato R1 CLS-268	0.3596	0.1293	0.387	0.714
Enzima Inicial	0.3599	0.1295	0.364	0.671
Enzima + Produto R2 CLS-383	0.3650	0.1332	0.292	0.529
Enzima + Ligante 04 R3 CLS-143	0.3665	0.1343	0.217	0.392
Enzima + Produto R1 CLS-654	0.3716	0.1381	0.386	0.685
Enzima + Ligante 04 R3 CLS-55	0.3722	0.1385	0.279	0.494
Enzima + Substrato R2 CLS-18	0.3754	0.1410	0.461	0.809
Enzima + Produto R1 CLS-817	0.3757	0.1412	0.375	0.658
Enzima + Substrato R2 CLS-784	0.3773	0.1423	0.564	0.984
Enzima + Ligante 27 R2 CLS-6	0.3815	0.1455	0.596	1.026
Enzima + Ligante 04 R3 CLS-203	0.3940	0.1553	0.324	0.537
Enzima + Produto R2 CLS-10	0.4064	0.1651	0.550	0.879
Enzima + Substrato R1 CLS-57	0.4108	0.1688	0.395	0.623
Enzima + Ligante 04 R2 CLS-19	0.4132	0.1707	0.361	0.565
Enzima + Ligante 04 R1 CLS-19	0.4151	0.1723	0.371	0.578
Enzima + Ligante 04 R3 CLS-49	0.4369	0.1909	0.403	0.590
Enzima + Ligante 37 R1 CLS-103	0.5222	0.2727	0.345	0.400
Enzima + Ligante 27 R2 CLS-150	0.5302	0.2811	0.551	0.627

