

Universidade Federal de Juiz de Fora
Instituto de Ciências Exatas / Faculdade de Engenharia
Programa de Pós-graduação em Modelagem Computacional

Vinicius Schmitz Pereira Nunes

**Análise comparativa das Ecto-NTPDase 1 de
Homo sapiens e *Schistosoma mansoni* por meio de modelagem
tridimensional, dinâmica molecular e *docking* receptor-ligante**

Juiz de Fora
2015

Vinicius Schmitz Pereira Nunes

Análise comparativa das Ecto-NTPDase 1 de *Homo sapiens* e *Schistosoma mansonii* por meio de modelagem tridimensional, dinâmica molecular e *docking* receptor-ligante

Tese apresentada ao Programa de Pós-graduação em Modelagem Computacional, da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de Doutor em Modelagem Computacional.

Orientador: Prof.^a D.Sc. Priscila Vanessa Zabala Capriles Goliatt

Coorientador: Prof. D.Sc. Carlos Cristiano Hasenclever Borges

Juiz de Fora

2015

Ficha catalográfica elaborada através do programa de geração automática da Biblioteca Universitária da UFJF, com os dados fornecidos pelo(a) autor(a)

Nunes, Vinicius Schmitz Pereira.

Análise comparativa das Ecto-NTPDase 1 de Homo sapiens e Schistosoma mansoni por meio de modelagem tridimensional, dinâmica molecular e docking receptor-ligante / Vinicius Schmitz Pereira Nunes. -- 2015.

153 p. : il.

Orientadora: Priscila Vanessa Zabala Capriles Goliatt

Coorientador: Carlos Cristiano Hasenclever Borges

Tese (doutorado) - Universidade Federal de Juiz de Fora, ICE/Engenharia. Programa de Pós-Graduação em Modelagem Computacional, 2015.

1. Schistosoma mansoni. 2. SmATPDase1. 3. Modelagem Comparativa. 4. Dinâmica Molecular. 5. Docking receptor-ligante. I. Goliatt, Priscila Vanessa Zabala Capriles, orient. II. Borges, Carlos Cristiano Hasenclever, coorient. III. Título.

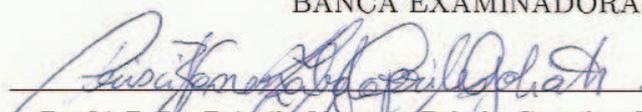
Vinicius Schmitz Pereira Nunes

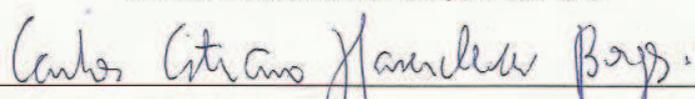
Análise comparativa das Ecto-NTPDase 1 de *Homo sapiens* e *Schistosoma mansoni* por meio de modelagem tridimensional, dinâmica molecular e docking receptor-ligante

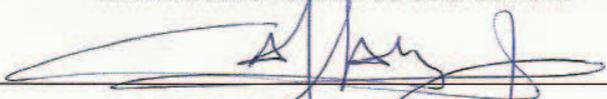
Tese apresentada ao Programa de Pós-graduação em Modelagem Computacional, da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do grau de Doutor em Modelagem Computacional.

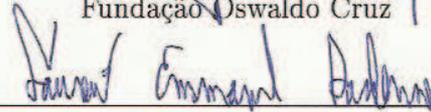
Aprovada em 7 de Dezembro de 2015.

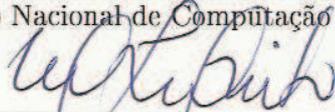
BANCA EXAMINADORA

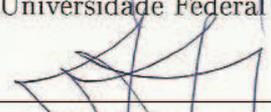

Prof.^a D.Sc. Priscila Vanessa Zabala Capriles Goliatt - Orientadora
Universidade Federal de Juiz de Fora

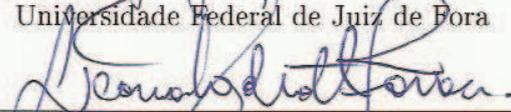

Prof. D.Sc. Carlos Cristiano Hasenclever Borges - Coorientador
Universidade Federal de Juiz de Fora


Prof. D.Sc. Ernesto Raúl Caffarena
Fundação Oswaldo Cruz


Prof. D.Sc. Laurent Emmanuel Dardenne
Laboratório Nacional de Computação Científica


Prof.^a D.Sc. Priscila de Faria Pinto
Universidade Federal de Juiz de Fora


Prof. D.Sc. Helio Ferreira dos Santos
Universidade Federal de Juiz de Fora


Prof. D.Sc. Leonardo Goliatt da Fonseca
Universidade Federal de Juiz de Fora

AGRADECIMENTOS

Agradeço aos meus pais, Jorge e Sandra, e aos meus irmãos, Marcelo e Flávio, pela amizade.

Aos meus orientadores, Priscila Goliatt e Carlos Cristiano Borges, pelo aprendizado.

Ao Programa de Pós-Graduação em Modelagem Computacional da Universidade Federal de Juiz de Fora, pelo suporte material.

A CAPES pelo apoio financeiro.

“O medo de amar é o medo de ser... De a todo momento escolher. Com acerto e precisão a melhor direção.” Beto Guedes.

RESUMO

A Esquistossomose é uma doença negligenciada causada por parasitas do gênero *Schistosoma*. Segundo a Organização Mundial da Saúde, mais de 200 milhões de pessoas no mundo estão infectadas, sendo de 4 a 6 milhões de pessoas somente no Brasil. A principal forma de tratamento da doença é o uso do medicamento praziquantel, porém, há relatos na literatura de resistência do parasita ao medicamento. Tal situação levantou a necessidade pela busca de novos alvos moleculares e novos medicamentos para o tratamento da doença. Um grupo de proteínas apresentadas como potenciais alvos moleculares de esquistomicidas é o das NTPDases. Estas enzimas hidrolisam nucleotídeos di e trifosfatados (*e.g.* ADP e ATP) em presença de cátions bivalentes, e estão descritas em diferentes organismos. No parasita *Schistosoma mansoni* foi descrita uma isoforma de NTPDase ancorada na membrana plasmática das células do tegumento do verme adulto e conhecida como SmATPDase1 ou Sm1. Segundo a literatura, esta isoforma é a segunda proteína mais expressa no tegumento dos vermes adultos. Devido à localização e a importância dos nucleotídeos di e trifosfatos na ativação hemostática e de células do sistema imunológico, foi sugerido que a Sm1 estaria envolvida na regulação das concentrações de nucleotídeos em torno do parasita, o que contribuiria com sua evasão. Baseado nos pressupostos acima, tem sido proposto o uso da Sm1 como possível alvo molecular de novos esquistomicidas. Nos mamíferos foram descritas oito isoformas de NTPDases. A isoforma 1 de humanos (HsNTPDase1), mais conhecida como CD39, está presente nas células do endotélio dos vasos sanguíneos, sendo responsável pela regulação das concentrações extracelulares de ATP e ADP. No presente trabalho, apresentamos os modelos 3D da Sm1 e CD39, construídos por meio da técnica de modelagem por homologia. Devido a semelhança da estrutura 3D de ambos os modelos, foi necessário realizar uma análise estrutural comparativa entre as duas enzimas, por meio de simulações de dinâmica molecular e estudos de *docking* receptor-ligante. Predição de cavidades foram feitas no modelo da Sm1 a fim de detectar cavidades diferentes do sítio-ativo que pudessem ser usadas em estudos de *docking*. Dessa etapa resultou a seleção de duas cavidades, o sítio-ativo e um sítio alternativo. Em seguida foram realizadas simulações de dinâmica molecular envolvendo os modelos da Sm1 e CD39, e os moldes PDB3ZX3 e PDB3CJA. Para cada modelo foram feitas duas simulações, uma com a presença do ANP (análogo do ATP) no sítio-ativo e uma

na ausência do ANP. Em todas as quatro simulações os modelos estão acoplados a uma membrana do tipo POPC, e o tempo de simulação foi de 32ns. Para os moldes o tempo de simulação foi 40ns. As simulações mostraram que a região do sítio alternativo apresentou, no modelo da Sm1 sem o ligante, mudanças na conformação que não foram observadas na Sm1 com ligante e nas duas simulações envolvendo o modelo da CD39. Isso sugere que o ligante contribui na estabilização da estrutura da Sm1 e CD39. Também foram feitas análises de drogabilidade e de agrupamento das conformações geradas ao longo das simulações envolvendo os modelos. Para as análises de agrupamento foi proposto o uso da metodologia GLCM, juntamente com o algoritmo *K-means*. Essas análises tinham como objetivo selecionar conformações das trajetórias que pudessem ser usadas nos estudos de *docking*. Para os estudos de *docking* foram analisadas a afinidade de três ligantes (ANP, ARL67153 e praziquantel) no sítio ativo e no sítio alternativo, de ambas as enzimas. Para o sítio ativo, os melhores resultados foram obtidos com os ligantes ANP e ARL67153, em todas as simulações. Foi observado que o ARL67153 apresentou resultados similares com os do ANP. No sítio alternativo os melhores resultados foram obtidos com as conformações da simulação da Sm1 sem o ANP no sítio ativo, sendo que na CD39 nenhum dos ligantes foram atracados dentro do sítio. Com relação ao praziquantel, os melhores resultados foram observados no sítio ativo da Sm1 e CD39. É possível concluir que o sítio alternativo da Sm1 é uma região que pode ser usada para o atracamento de possíveis inibidores dessa enzima.

Palavras-chave: *Schistosoma mansoni*. Esquistossomose. Ecto-NTPDases. SmATPDase1. CD39. Modelagem Comparativa. Dinâmica Molecular. *Docking* receptor-ligante.

ABSTRACT

Schistosomiasis is a neglected disease caused by parasites of the genus *Schistosoma*. According to the World Health Organization, over 200 million people worldwide were infected, and 4 to 6 million people only in Brazil. The main treatment is the use of the drug praziquantel, but there are reports in the literature suggesting the parasite's resistance to praziquantel. This situation raised the need for search new molecular targets and new drugs for treating this disease. A group of proteins as potential targets is NTPDase. These enzymes hydrolyse nucleotides (*e.g.*, ADP and ATP) in the presence of bivalent cations, and they are described in different organisms. In the parasite *Schistosoma mansoni*, an isoform of NTPDase, named SmATPDase1 or Sm1, was described as a protein anchored in the plasma membrane of cells in the seed coat of the adult worm. According to the literature, this isoform is the most expressed enzyme in the tegument of adult worms. Due to the location and importance of nucleotide triphosphates in the hemostatic and activation of immune cells, Sm1 has been suggested to be involved in the regulation of nucleotide concentrations around the parasite, which could be a mechanism of immune evasion in schistosomiasis. Based on the above assumptions, it has been proposed the use of Sm1 as a possible target for new schistosomicide drugs. In mammals have been described eight isoforms. Isoform 1 of human NTPDase (HsNTPDase1), also known as CD39, is present in the endothelial cells of blood vessels, and it is responsible for the regulation of extracellular concentrations of ATP and ADP. In this paper, we present 3D models of Sm1 and CD39, constructed by homology modeling technique. Due to the similarity of 3D structures of both models, it was necessary to perform a comparative structural analysis of both enzymes by molecular dynamics simulations and receptor-ligand docking. Cavities prediction were made in Sm1's model and two cavities were selected, the active site and an alternative binding site proposed in this work. After, it was performed molecular dynamics simulations involving models of Sm1 and CD39, and PDB3ZX3 and PDB3CJA templates. Two simulations were performed for each model, one in the presence of ANP (ATP analog) in the active-site and other in the absence. In models' simulations, structures were coupled to a POPC membrane, and the simulation time was 32ns. Simulation time for templates was 40ns. Simulations showed that the alternative site in Sm1 without ANP presented conformational changes which

were not observed in simulations of CD39 and Sm1 with ANP. This suggests that the presence of ligand in the active site contributes in stabilizing the structure of Sm1 and CD39. Druggability and clustering analysis were performed with conformations generated through simulations. In this work, we proposed the use of GLCM methodology with K-means algorithm for clustering analysis, aiming to select conformations from trajectories that could be used in receptor-ligand docking. We analyzed the affinity of three ligands (ANP, ARL67153 and praziquantel) in the active site and in the alternative site of Sm1 and CD39 enzymes. The best results for active site were obtained with ANP and ARL67153 ligands, in all simulations. For alternative site, the best results were obtained with conformations from the simulation of Sm1 without ANP in the active site. For CD39's simulations, none of the ligands were docked within the cavity of alternative site. With regard to praziquantel, the best results were observed in the active site of Sm1 and CD39. It is possible to conclude that the alternative site of Sm1 is a region which can be used for docking of possible inhibitors of this enzyme.

Keywords: *Schistosoma mansoni*. Schistosomiasis. Ecto-NTPDases. SmATPDase1. CD39. Comparative Modeling. Molecular Dynamics. Ligand-receptor *Docking*.

SUMÁRIO

1	INTRODUÇÃO.....	17
1.1	Esquistossomose	17
1.2	As Ecto-Nucleosídeo Trifosfato Difosfohidrolases	19
1.3	ATP-difosfohidrolases de <i>S. mansoni</i>	22
1.4	Mecanismo Catalítico das NTPDases	24
1.4.1	<i>Modelo proposto por Zebisch e Sträter (2008)</i>	24
1.4.2	<i>Modelo proposto por Kozakiewicz (2008)</i>	25
1.5	Modelagem Comparativa	27
1.6	Dinâmica Molecular	30
1.7	<i>Docking receptor-ligante</i>	35
1.8	Justificativa	40
2	OBJETIVOS.....	42
2.1	Objetivo Geral	42
2.2	Objetivos Específicos	42
3	METODOLOGIA.....	43
3.1	Construção dos modelos tridimensionais	43
3.1.1	<i>Predição de peptídeo sinal, de domínio transmembranar e de estrutura secundária</i>	44
3.1.2	<i>Alinhamento global</i>	44
3.1.3	<i>Geração dos modelos</i>	45
3.1.4	<i>Otimização de alças</i>	46
3.1.5	<i>Validação dos modelos</i>	47
3.1.6	<i>Adição dos hidrogênios</i>	47
3.2	Dinâmica Molecular	47
3.2.1	<i>Preparação dos sistemas</i>	48
3.2.2	<i>Etapas da Simulação por Dinâmica molecular</i>	50
3.3	Relação área por lipídio	51
3.4	Análise do RMSD e Raio de Giro	52

3.5	Predição de cavidades	52
3.6	Análise das cavidades	53
3.7	Conjunto de conformações	54
3.8	GLCM e análise de agrupamento	55
3.9	Estudos de <i>docking</i> receptor-ligante	61
3.10	Resumo da metodologia	65
4	RESULTADOS E DISCUSSÃO	67
4.1	Filogenia das NTPDases de mamíferos e parasitas	67
4.2	Os modelos 3D da Sm1 e CD39	67
4.2.1	<i>Validação dos modelos</i>	74
4.2.2	<i>Resíduos catalíticos</i>	76
4.2.3	<i>Domínios transmembranares</i>	77
4.2.4	<i>Perfil eletrostático dos modelos</i>	80
4.2.5	<i>Ligações dissulfeto</i>	80
4.3	Dinâmica Molecular	82
4.3.1	<i>Relação área por lipídio</i>	83
4.3.2	<i>Cálculo das Energias</i>	84
4.3.2.1	<i>Energias dos moldes</i>	84
4.3.2.2	<i>Energias dos modelos</i>	87
4.3.2.3	<i>Energias membranas</i>	87
4.3.3	<i>RMSD das trajetórias</i>	92
4.3.3.1	<i>RMSD dos ECD</i>	93
4.3.3.2	<i>RMSD das TM</i>	100
4.3.4	<i>Raio de giro</i>	103
4.3.4.1	<i>Raio de giro dos ECD</i>	103
4.3.4.2	<i>Raio de giro das TM</i>	103
4.3.5	<i>Águas conservadas</i>	109
4.3.6	<i>O íon cálcio</i>	109
4.4	Análise de drogabilidade das cavidades	109
4.5	Análise de agrupamento	115
4.6	Conformações da Sm1 e o movimento de K279	118
4.7	Estudos de <i>docking</i> receptor-ligante	122

<i>4.7.1</i>	<i>Análises de docking na cavidade-1</i>	124
<i>4.7.1.1</i>	<i>Análise de redocking do ANP</i>	124
<i>4.7.1.2</i>	<i>Análise dos dockings ao longo das simulações</i>	126
<i>4.7.2</i>	<i>Análises de docking na cavidade-2</i>	129
<i>4.7.3</i>	<i>Conclusões dos estudos de docking</i>	131
5	CONCLUSÕES e PERSPECTIVAS	132
5.1	Perspectivas	134
	REFERÊNCIAS	135
6	ANEXO	144

LISTA DE ILUSTRAÇÕES

1.1	Mapa com a distribuição das espécies de <i>Schistosoma</i> no mundo.	18
1.2	Ciclo de vida do <i>Schistosoma</i>	19
1.3	Ligações dissulfeto presentes na RnNTPDase1 e RnNTPDase2.	22
1.4	Região do sítio ativo da NTPDase2 de <i>Rattus norvegicus</i>	25
1.5	Resíduos catalíticos da apirase de batata.	26
1.6	Posicionamento do íon metal e moléculas de água no sítio ativo da apirase de batata.	26
1.7	Esquema dos potenciais de um campo de força.	32
1.8	Representação das condições periódicas de contorno.	35
1.9	Esquema dos modelos de reconhecimento molecular.	37
3.1	Consenso dos preditores de SS e de TM para a Sm1.	45
3.2	Consenso dos preditores de SS e TM para a CD39.	45
3.3	Região do sítio ativo do molde PDB3CJA.	46
3.4	Estrutura final dos modelos inseridas em membrana POPC.	49
3.5	Cavidades selecionadas para os estudos de <i>docking</i>	53
3.6	Ângulos de orientação para o cálculo das co-ocorrências.	56
3.7	Exemplo de construção de uma matriz GLCM.	56
3.8	Fluxograma da metodologia.	66
4.1	Árvore filogenética das sequências de aminoácidos das NTPDases de mamíferos e parasitas.	69
4.2	Estruturas primária e secundária da Sm1.	70
4.3	Topologia e estrutura terciária da CD39.	71
4.4	Estruturas primária e secundária da CD39.	72
4.5	Topologia e estrutura terciária da CD39.	73
4.6	Perfil do DOPE para três conformações da Sm1 e os moldes.	75
4.7	Perfil do DOPE para três conformações da CD39 e os moldes.	75
4.8	Estrutura 3D e sítio ativo da Sm1 e CD39.	78
4.9	Resíduos catalíticos pelo modelo de (KOZAKIEWICZ et al., 2008).	79

4.10	Potencial de superfície eletrostático da Sm1 e CD39.	81
4.11	Potencial de superfície eletrostático da Sm1 e CD39 na região do sítio-ativo.	81
4.12	Ligações dissulfeto presentes na Sm1 e CD39.	82
4.13	Área por lipídio durante os 20ns de equilíbrio da Sm1 com e sem ANP.	84
4.14	Área por lipídio durante os 20ns de equilíbrio da CD39 com e sem ANP.	85
4.15	Energia total dos moldes durante os 40ns de simulação.	86
4.16	Energia total da Sm1 e Sm1-ANP durante 32ns de simulação.	88
4.17	Energia total da CD39 e CD39-ANP durante 32ns de simulação.	89
4.18	Energia total da membrana POPC durante 32ns de simulação da proteína Sm1.	90
4.19	Energia total da membrana POPC durante 32ns de simulação da proteína CD39.	91
4.20	Variação do RMSD dos moldes.	94
4.21	Variação do RMSD do ECD da Sm1 e Sm1-ANP.	95
4.22	Alinhamento estrutural do ECD da Sm1.	96
4.23	Alinhamento estrutural do ECD da Sm1-ANP.	97
4.24	Variação do RMSD do ECD da CD39 e CD39-ANP.	99
4.25	Variação do RMSD das TM da Sm1 e Sm1-ANP.	101
4.26	Variação do RMSD das TM da CD39 e CD39-ANP.	102
4.27	Gráfico do raio de giro dos moldes.	104
4.28	Gráfico do raio de giro do ECD da Sm1 e Sm1-ANP.	105
4.29	Gráfico do raio de giro do ECD da CD39 e CD39-ANP.	106
4.30	Gráfico do raio de giro das TM da Sm1 e Sm1-ANP.	107
4.31	Gráfico do raio de giro das TM da CD39 e CD39-ANP.	108
4.32	Variação do <i>Drugscore</i> das cavidades 1 e 2 da Sm1 ao longo da DM.	111
4.33	Variação do <i>Drugscore</i> das cavidades 1 e 2 da Sm1-ANP ao longo da DM.	112
4.34	Variação do <i>Drugscore</i> das cavidades 1 e 2 da CD39 ao longo da DM.	113
4.35	Variação do <i>Drugscore</i> das cavidades 1 e 2 da CD39-ANP ao longo da DM.	114
4.36	Alinhamento estrutural do ECD de conformações da Sm1 com e sem ligante.	119
4.37	Alinhamento estrutural do domínio ECD de três conformações da Sm1.	121
4.38	<i>Redocking</i> do ANP na cavidade-1 do <i>frame-0</i> da Sm1-ANP.	124
4.39	<i>Redocking</i> do ANP na cavidade-1 do <i>frame-80</i> da CD39-ANP.	125
4.40	<i>Docking</i> do praziquantel na cavidade-1 do <i>frame-0</i> da CD39-ANP.	127
4.41	<i>Docking</i> do ANP (verde) e ARL67 (rosa) na cavidade-1 do <i>frame-219</i> da Sm1-ANP.	128

4.42	<i>Docking</i> do praziquantel na cavidade-2 do <i>frame-219</i> da Sm1-ANP.	129
4.43	<i>Docking</i> do ANP (verde) e ARL67 (azul) na cavidade-2 do <i>frame-0</i> da Sm1.	130
6.1	Gráfico do Errat para o modelo da Sm1.	145
6.2	Gráfico do Errat para a Sm1-ANP ao final da minimização.	146
6.3	Gráfico do Errat para a Sm1-ANP ao final do aquecimento.	147
6.4	Gráfico do Errat do modelo para a CD39.	148
6.5	Gráfico do Errat para a CD39-ANP ao final da minimização.	149
6.6	Gráfico do Errat para a CD39-ANP ao final do aquecimento.	150
6.7	Gráfico do Z-score para o modelo da Sm1.	151
6.8	Gráfico do Z-score para a Sm1-ANP ao final da minimização.	151
6.9	Gráfico do Z-score para a Sm1-ANP ao final do aquecimento.	152
6.10	Gráfico do Z-score do modelo para a CD39.	152
6.11	Gráfico do Z-score para a CD39-ANP ao final da minimização.	153
6.12	Gráfico do Z-score para a CD39-ANP ao final do aquecimento.	153

LISTA DE TABELAS

3.1	Heteroátomos presentes nos cristais analisados para a etapa de construção do modelo da Sm1 e CD39.	43
3.2	Dimensões das caixas de simulação.	49
3.3	Concentrações dos sais.	50
3.4	Resíduos usados na centralização das caixas para simulação do <i>docking</i>	64
4.1	Validação das estruturas 3D da Sm1 e CD39.	76
4.2	Valores do RMSD para os moldes e modelos nas simulações.	93
4.3	Resumo dos agrupamentos obtidos com os métodos <i>g_cluster</i> e GLCM para as quatro dinâmicas moleculares testadas.	115
4.4	Estrutura média e respectivo <i>drugscore</i> dos agrupamentos da Sm1 com os métodos <i>g_cluster</i> e GLCM.	116
4.5	Estrutura média e respectivo <i>drugscore</i> dos agrupamentos da Sm1-ANP com os métodos <i>g_cluster</i> e GLCM.	116
4.6	Estrutura média e respectivo <i>drugscore</i> dos agrupamentos da CD39 com os métodos <i>g_cluster</i> e GLCM.	117
4.7	Estrutura média e respectivo <i>drugscore</i> dos agrupamentos da CD39-ANP com os métodos <i>g_cluster</i> e GLCM.	117
4.8	Estruturas médias selecionadas para os estudos de <i>docking</i> da Sm1, Sm1-ANP, CD39 e CD39-ANP.	118
4.9	Resultados do <i>XPGScore</i> nos <i>dockings</i> das quatro simulações de dinâmica molecular.	123

1 INTRODUÇÃO

1.1 Esquistossomose

A esquistossomose é uma doença parasitária causada por helmintos do gênero *Schistosoma*, sendo considerada pela Organização Mundial da Saúde uma doença negligenciada (COLLEY; SECOR, 2007). Esta doença ocorre predominantemente em países tropicais da África, Sudeste Asiático, América do Sul, Oriente Médio e Caribe (GRYSEELS et al., 2006). Estima-se que em 2009, havia nessas regiões 239 milhões de pessoas infectadas pelo parasito (KING et al., 2011), com aproximadamente 300 mil mortes anuais (FERREIRA; OLIVA; ANDRICOPULO, 2015). No Brasil, estima-se que 25 milhões de pessoas vivam em áreas endêmicas, sendo que de 4 a 6 milhões estão infectadas (LEAL et al., 2012).

As principais espécies que parasitam o homem são o *S. mansoni*, *S. haematobium*, *S. japonicum*, *S. mekongi* e *S. intercalatum*. Enquanto que as espécies *S. japonicum* e *S. mekongi* são encontradas na Ásia, as espécies *S. haematobium* e *S. intercalatum* ocorrem na África. A espécie *S. mansoni* ocorre na África e nas Américas, sendo a responsável pela esquistossomose no Brasil (GRYSEELS et al., 2006). A figura 1.1 mostra o mapa com a distribuição nos continentes das espécies de *Schistosoma* que parasitam o homem.

O ciclo de vida do parasito envolve cinco fases (ovo, miracídio, cercária, esquistossômulo e adulto) e dois hospedeiros (intermediário e definitivo) (GRYSEELS et al., 2006). O primeiro hospedeiro é um caramujo do gênero *Biomphalaria*. Existem no Brasil três espécies de caramujos que são hospedeiros naturais, o *B. glabrata*, *B. tenagophila* e *B. straminea* (BRASIL, 1995). É no caramujo que o parasito evolui da fase miracídio para cercária sendo, portanto, o hospedeiro intermediário e indispensável no ciclo de vida do parasito. A cercária é a forma infectante, vivendo livremente na água.

Quando em contato com a pele ou mucosa humana, a cercária é capaz de penetrar ativamente sendo transportada pela corrente sanguínea até os vasos do sistema porta-hepático (PEARCE; MACDONALD, 2002). No homem, o parasito evolui para a forma esquistossômulo e depois adulto. Um homem infectado (hospedeiro definitivo) passa a excretar os ovos do parasito adulto através das fezes e urina (GRYSEELS et al., 2006),

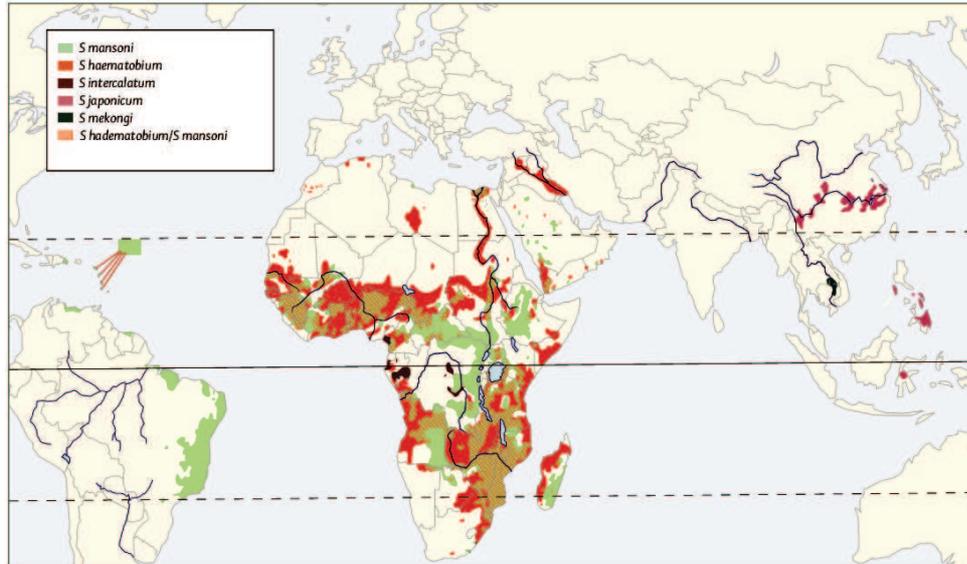


Figura 1.1: Mapa com a distribuição das espécies de *Schistosoma* no mundo. (Fonte: Gryseels et al, 2006) .

que permanecem viáveis no meio ambiente por até 7 dias. Quando em contato com a água, dos ovos são liberados os miracídeos que poderão infectar novos caramujos. A figura 1.2 mostra um resumo do ciclo de vida do *Schistosoma*.

A esquistossomose mansoni apresenta uma fase aguda e outra crônica (GRYSEELS et al., 2006). A fase aguda, também conhecida como febre de Katayama, caracteriza-se pela perda de peso, dores abdominais, diarreia e hepatoesplenomegalia (GRYSEELS et al., 2006). Já a fase crônica caracteriza-se por ser uma fase onde os parasitos adultos começam a produzir os ovos. Esses ovos são transportados pela corrente sanguínea para os pulmões, baço e fígado (GRYSEELS et al., 2006). Enzimas proteolíticas liberadas pelos ovos estimulam uma resposta imunológica onde células T, macrófagos e eosinófilos se concentram ao redor dos ovos, causando uma reação inflamatória granulomatosa (PEARCE; MACDONALD, 2002). A fase crônica pode apresentar-se de forma branda com sintomas como perda de apetite e diarreia. Também pode ser assintomática ou pode se apresentar de forma mais grave caracterizada por hepatoesplenomegalia, cirrose e hemorragias (GRYSEELS et al., 2006).

Atualmente, a principal forma de tratamento da esquistossomose é a farmacoterapia, sendo o praziquantel a droga mais aplicada no tratamento. Esse fármaco atua sobre o parasito adulto provocando um influxo de Ca^{+2} através da membrana do tegumento o que leva a uma paralisia da contração muscular, sendo os canais de cálcio do tegumento os prováveis alvos do praziquantel (DOENHOFF; CIOLI; UTZINGER, 2008). Entretanto, já

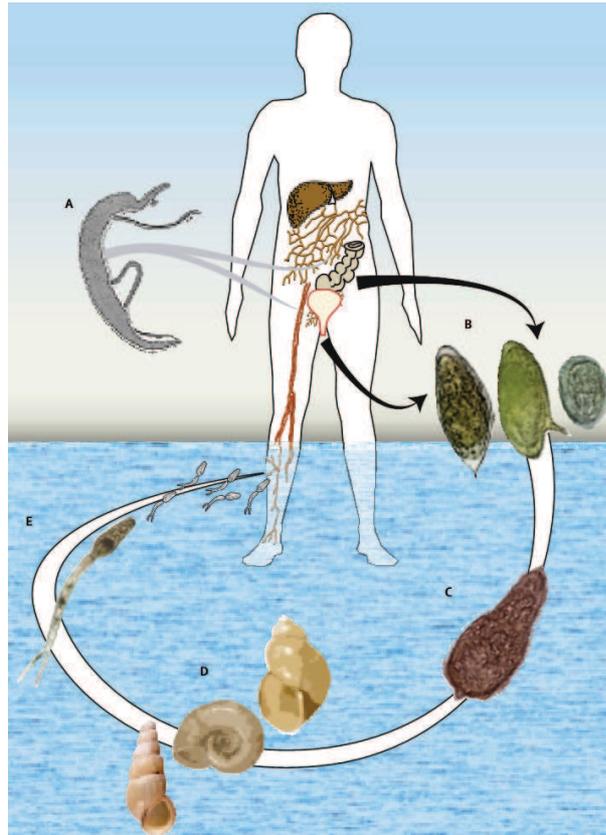


Figura 1.2: Ciclo de vida do *Schistosoma*. A- parasito adulto, B- ovos, C- miracídio, D- caramujo, E- cercária. (Fonte: Gryseels et al, 2006) .

foram relatados na literatura casos onde o tratamento com o praziquantel não apresentou efeito sobre o parasito e que esses casos são crescentes. Esses fatos seriam indícios da ocorrência de cepas resistentes ao praziquantel (DOENHOFF; CIOLI; UTZINGER, 2008).

Em 2007, Colley e Secor propuseram uma agenda de pesquisa sobre a esquistossomose conhecida como “*A Schistosomiasis Research Agenda*” (COLLEY; SECOR, 2007). Nela, os autores apresentam um conjunto de iniciativas que vão desde pesquisa básica sobre a doença e o parasito, até medidas de saúde pública. Uma dessas iniciativas seria a identificação de proteínas do parasito que possam ser alvos de novos candidatos a esquistossomicidas (COLLEY; SECOR, 2007).

1.2 As Ecto-Nucleosídeo Trifosfato Difosfohidrolases

As NTPDases (EC 3.6.1.5), também conhecidas como ATP-difosfohidrolases ou apirases, são enzimas que hidrolisam nucleosídeos difosfatados e trifosfatados em nucleosídeos monofosfatados, na presença de íons bivalentes, principalmente o cálcio e o magnésio.

Uma importante característica das NTPDases é a ocorrência de cinco regiões conservadas chamadas de ACR (*Apyrase Conserved Regions*). Todas as cinco ACR estão localizadas em torno do sítio ativo das NTPDases (ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011).

Uma das primeiras NTPDases a serem identificadas foi a apirase da batata (*Solanum tuberosum*), quando em 1956 foi proposto por Cohn e Meek um modelo para o mecanismo catalítico da enzima (COHN; MEEK, 1957). Nos mamíferos, já foram descritas oito isoformas de NTPDases (ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011). Estas oito isoformas estão divididas em três grupos: um primeiro grupo formado pelas isoformas NTDase1, NTDase2, NTDase3 e NTDase8, sendo extracelulares e localizadas na superfície externa das células; um grupo formado pelas NTDase5 e NTDase6, que são secretadas; e um grupo formado pelas isoformas NTDase4 e NTDase7, que são intracelulares (ligadas à membrana de organelas citoplasmáticas) (SANSOM, 2012; KNOWLES, 2011).

O primeiro grupo das NTPDases de mamíferos é conhecido como Ecto-NTPDases, e estão presentes em vários órgãos e tecidos, como o endotélio dos vasos sanguíneos, fígado, pâncreas, glândulas salivares e rim, atuando na regulação das concentrações dos nucleotídeos extracelulares (ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011). É o grupo que vem recebendo mais atenção devido ao seu envolvimento com o processo de sinalização purinérgico (KNOWLES, 2011; ZIMMERMANN; ZEBISCH; STRÄTER, 2012).

A sinalização purinérgica é um processo em que nucleosídeos trifosfatados e difosfatados, ATP e ADP por exemplo, atuam como moléculas sinalizadoras estimulando receptores purinérgicos conhecidos como P2 (ROBSON; SEVIGNY; ZIMMERMANN, 2006; SANSOM, 2012). Foram descritos dois grupos de receptores P2, um grupo conhecido como P2X formado por 7 subtipos e ativados principalmente por ATP, e um segundo grupo chamado de P2Y formado por 8 subtipos e estimulados por ATP mas também ADP, UTP e UDP (ROBSON; SEVIGNY; ZIMMERMANN, 2006; SANSOM, 2012; ZIMMERMANN; ZEBISCH; STRÄTER, 2012). Em vários processos fisiológicos como a coagulação sanguínea, proliferação e diferenciação celular, apoptose, inflamação, resposta imunológica, há o envolvimento da sinalização purinérgica (ROBSON; SEVIGNY; ZIMMERMANN, 2006; SANSOM, 2012).

Segundo a literatura, quando o substrato for o ATP, as isoformas NTPDase2, NTPDase3 e NTPDase8 fariam primeiro a hidrólise do ATP liberando o ADP e em seguida a hidrólise do ADP liberando o AMP, devido ao acúmulo de uma quantidade considerável de ADP antes da hidrólise para AMP (ZIMMERMANN; ZEBISCH; STRÄTER, 2012). Em contraste, a NTPDase1 faria a hidrólise do ATP direto para AMP, devido à ausência de quantidades significativas do ADP como intermediário (KACZMAREK et al., 1996; ZIMMERMANN; ZEBISCH; STRÄTER, 2012). Porém, a hidrólise do UTP pela NTPDase1 leva a um acúmulo de quantidades significativas de UDP (ZIMMERMANN; ZEBISCH; STRÄTER, 2012). A NTPDase1 de humanos é também conhecida como CD39 e está presente nas células do endotélio dos vasos sanguíneos, baço, placenta e células do sistema imunológico como as células B e dendríticas (ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011).

A estrutura das Ecto-NTPDases é caracterizada por um domínio extracelular (ECD - do inglês, *Extracellular Domain*), onde estão localizadas as ACR e o sítio ativo, e dois domínios transmembranares localizados próximos às extremidades N-terminal (Nter) e C-terminal (Cter) (GRINTHAL; GUIDOTTI, 2006; ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011). Os domínios transmembranares são formados por duas hélices, chamadas de TM1 e TM2, e atuam na ancoragem da enzima na superfície externa das células (ZEBISCH et al., 2013).

Em 2008 e 2012, foram publicadas as estruturas dos cristais da NTPDase2 e NTPDase1 de *Rattus norvegicus*, denominadas RnNTPDase2 e RnNTPDase1, respectivamente. Somente o domínio ECD foi cristalizado (ZEBISCH; STRÄTER, 2008; ZEBISCH et al., 2012). Uma diferença entre elas é que a RnNTPDase1 é capaz de hidrolisar tanto o ATP como o ADP, sem preferência (ZEBISCH et al., 2012). Ao contrário, a RnNTPDase2 apresentou uma preferência por ATP (ZEBISCH; STRÄTER, 2008).

Uma importante característica da RnNTPDase1 e RnNTPDase2 é a presença de cinco pontes dissulfeto no domínio ECD, o que garante uma estabilidade a estrutura de ambas as enzimas (ZEBISCH et al., 2013). As figuras 1.3(a) e 1.3(b) mostram a estrutura da RnNTPDase1 (PDB3ZX3) e RnNTPDase2 (PDB3CJA) e em destaque as cinco pontes dissulfeto.

A estrutura do cristal de uma Ecto-NTPDase, presente na bactéria *Legionella pneumophila* (LpNTPDase), foi publicada em 2010 (VIVIAN et al., 2010). Segundo

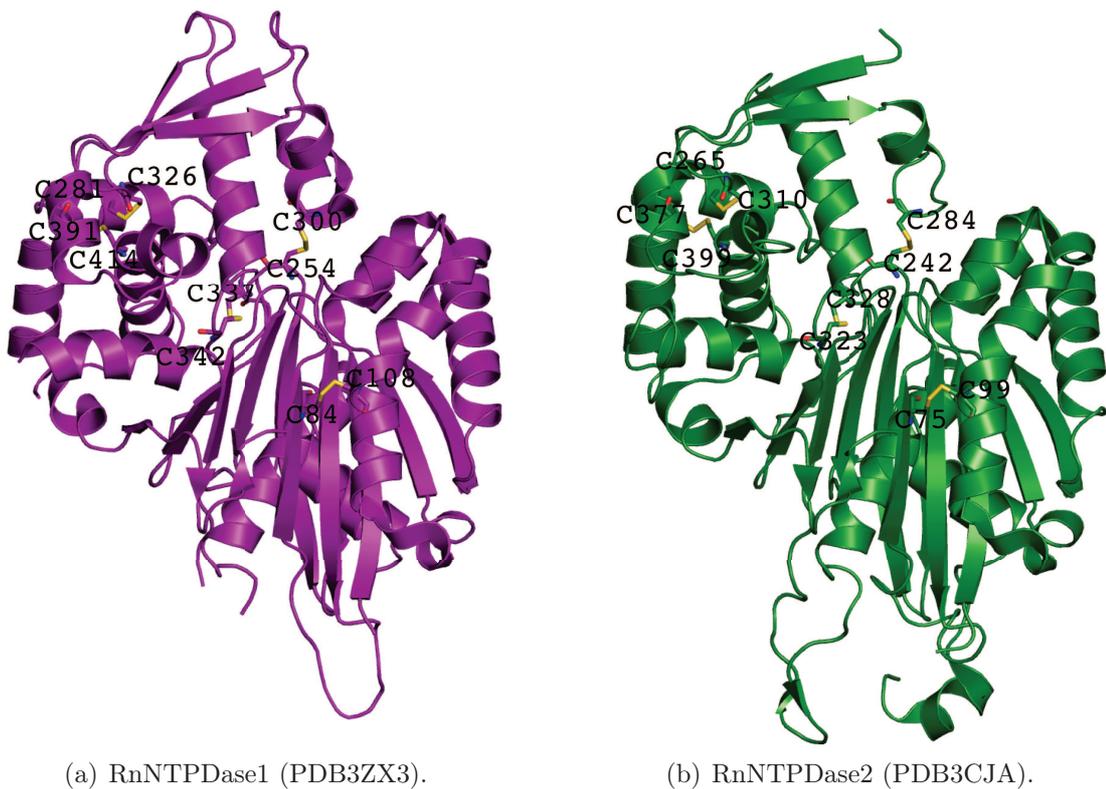


Figura 1.3: Ligações dissulfeto presentes na RnNTPDase1 e RnNTPDase2. Em destaque as cinco pontes dissulfeto (amarelo) presentes na (a) RnNTPDase1 (magenta) e (b) RnNTPDase2 (verde).

Sansom e colaboradores (2008), essa forma é similar à NTPDase1 de humanos e teria um papel importante na replicação da bactéria no interior das células do hospedeiro (SANSOM et al., 2008; VIVIAN et al., 2010).

1.3 ATP-difosfohidrolases de *S. mansoni*

Em 1993, foi evidenciada pela primeira vez a presença de atividade ATP-difosfohidrolásica no tegumento de vermes adultos de *S. mansoni* (VASCONCELOS et al., 1993). Posteriormente, foram identificadas duas isoformas, SmATPDase1 e SmATPDase2, no tegumento do parasito adulto, com massa molecular em torno de 63KDa e 55KDa, respectivamente (VASCONCELOS et al., 1996). A presença das duas isoformas em outras fases do ciclo de vida do parasito, como os ovos e miracídios, foi confirmada posteriormente (LEVANO-GARCIA et al., 2007; FARIA-PINTO et al., 2004).

Foi demonstrado que ambas as isoformas, isoladas de ovos do parasito, foram reconhecidas pelo soro de camundongos infectados e não pelo soro de camundongos

saudáveis, indicando que essas enzimas são reconhecidas como antígenos pelo sistema imunológico do hospedeiro (FARIA-PINTO, 2008). Também foi demonstrado, por técnica ELISA, que anticorpos IgG produzidos em camundongos infectados experimentalmente por *S. mansoni*, apresentaram alta reatividade contra a apirase de batata (FARIA-PINTO, 2008). Tal imunoreatividade cruzada, sugere o compartilhamento de domínios antigênicos conservados entre as isoformas do parasito e a apirase de batata (FARIA-PINTO, 2008).

Em 2003, foi publicada a caracterização do gene da SmATPDase1 (Sm1), responsável pela síntese da isoforma de 63KDa (DEMARCO et al., 2003). Segundo os autores, essa isoforma estaria localizada na superfície externa do tegumento em todas as fases de vida do parasito, exceto no ovo. A estrutura da Sm1 seria muito semelhante a da CD39 de humanos, com um domínio ECD e dois domínios transmembranares (TM1 e TM2) (DEMARCO et al., 2003).

Em 2007, Levano-Garcia e colaboradores publicaram a caracterização do gene da SmATPDase2 (Sm2), que codifica a isoforma de 55KDa. Segundo os autores, diferentemente da Sm1, a isoforma Sm2 possui uma única região transmembranar que é clivada após sua síntese. Também foi sugerido que a Sm2 seria sintetizada por células do tegumento e transportadas e excretadas pelo tegumento de parasitos adultos (LEVANO-GARCIA et al., 2007). Em 2014, Souza e colaboradores publicaram o modelo 3D completo da Sm2. O modelo proposto foi feito usando a técnica de modelagem comparativa e apresenta semelhança estrutural com a NTPDase6 de humanos (SOUZA et al., 2014).

A atividade enzimática das isoformas Sm1 e Sm2 depende de íons bivalentes. Para a Sm1 a atividade enzimática foi maior na presença de Ca^{+2} enquanto que para a Sm2 sua atividade foi maior na presença de Mg^{+2} (LEVANO-GARCIA, 2008). Um estudo de proteômica do tegumento do verme adulto de *S. mansoni*, feito por Castro-Borges e colaboradores, mostrou que a Sm1 é a segunda proteína mais expressa no tegumento (CASTRO-BORGES et al., 2011). Também foi demonstrado que a supressão, usando a técnica de RNAi, de enzimas presentes no tegumento de vermes adultos e que estão envolvidas na catálise do ATP e ADP, somente a supressão da Sm1 apresentou redução na capacidade de clivagem do ATP e ADP extracelular (DA'DARA et al., 2014).

Devido as propriedades catalíticas, localização e importância dos nucleotídeos di- e trifosfatados na ativação hemostática e de células do sistema imunológico, foi sugerido

que a Sm1 estaria envolvida na regulação das concentrações de nucleótidos em torno do parasito, o que contribuiria com a sua evasão do sistema de defesa do hospedeiro mamífero (VASCONCELOS et al., 1993; VASCONCELOS et al., 1996; DEMARCO et al., 2003; LEVANO-GARCIA et al., 2007; BHARDWAJ; SKELLY, 2009).

No presente trabalho, nós apresentamos um modelo tridimensional das proteínas Sm1 e CD39. Para a construção dos modelos, usamos a técnica de modelagem comparativa via programa Modeller, aplicando uma abordagem de restrição de estrutura secundária e restrições de distâncias entre átomos em contato nas regiões de TM1 e TM2.

1.4 Mecanismo Catalítico das NTPDases

Na literatura, foram descritos dois modelos sobre o mecanismo catalítico das NTPDases. A seguir, serão descritos esses dois modelos.

1.4.1 Modelo proposto por Zebisch e Sträter (2008)

O modelo proposto por Zebisch e Sträter (2008) é baseado na NTPDase2 de *Rattus norvegicus* (PDB3CJA). Pelo modelo, a hidrólise ocorre pelo ataque nucleofílico de uma molécula de água no grupo fosfato terminal (γ -fosfato) do ANP. A posição da água nucleofílica seria direcionada pelas interações com os resíduos A123, E165 e S206 da NTPDase2. Uma segunda molécula de água estaria interagindo com o resíduo Q208 e a água nucleofílica (ZEBISCH; STRÄTER, 2008). A Figura 1.4 mostra a molécula de água nucleofílica (água 1) posicionada a uma distância de 3,11 Å do grupo fosfato terminal, os resíduos que direcionam a posição dessa água nucleofílica e o íon cálcio acompanhado por quatro moléculas de água (ZEBISCH; STRÄTER, 2008).

Pela figura 1.4 é possível ver que o ângulo formado entre a molécula de água 1 e o grupo γ -fosfato é de 168,7° (ZEBISCH; STRÄTER, 2008). Segundo os autores do modelo, nenhuma outra molécula de água ou aminoácido estaria melhor posicionado para um ataque nucleofílico do que a molécula de água 1 (ZEBISCH; STRÄTER, 2008). A ativação da água nucleofílica seria feita pelo grupo carboxila da cadeia lateral do resíduo E165, que encontra-se a uma distância de 2,6 Å da água nucleofílica (ZEBISCH; STRÄTER, 2008).

Experimentos de mutagênese no resíduo E174 da CD39 de humanos, e que corresponde ao resíduo E165, mostraram a perda completa da atividade catalítica da enzima

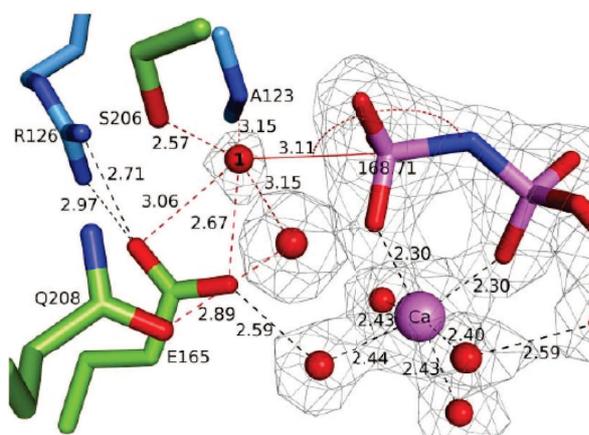


Figura 1.4: Região do sítio ativo da NTPDase2 de *Rattus norvegicus*. Destaque para a água nucleofílica (água 1 - em vermelho) posicionada a uma distância de 3,11 Å do grupo fosfato terminal do ligante ANP (em rosa), os resíduos (em azul e verde) que direcionam a posição da água nucleofílica e íon cálcio (em rosa) coordenado por quatro moléculas de água. (Fonte: Retirado de Zebisch e Sträter (2008)) .

(DROSOPOULOS et al., 2000; ZEBISCH; STRÄTER, 2008).

No caso do íon metálico, segundo Zebisch e Sträter (2008), este funcionaria como um catalizador polarizando uma das ligações P-O do grupo fosfato terminal. Durante o ataque nucleofílico, as cargas negativas dos estados de transição são estabilizadas pelo íon metálico (ZEBISCH; STRÄTER, 2008). A posição desse íon é coordenada pela presença de quatro moléculas de água (ZEBISCH et al., 2013) (ver Figura 1.4).

1.4.2 Modelo proposto por Kozakiewicz (2008)

Também em 2008, Kozakiewicz e colaboradores propuseram um modelo para a catálise na NTPDase1 baseado em um modelo da apirase de batata (*Solanum tuberosum*). O modelo proposto sugere um outro mecanismo catalítico envolvendo dois conjuntos de resíduos, onde o primeiro grupo (S54, T127 e E170) seria responsável pela hidrólise do γ -fosfato, enquanto que o segundo grupo (T55 e E78) atuaria na hidrólise do β -fosfato (KOZAKIEWICZ et al., 2008). A figura 1.5 mostra a região do sítio ativo da apirase de batata e os resíduos envolvidos na catálise.

Estudo feito com a mutação S57A na CD39 de humanos, correspondente a S54 da apirase de batata, provocou um aumento da atividade ADPase e diminuição da atividade ATPase (DROSOPOULOS et al., 2000; KOZAKIEWICZ et al., 2008). Segundo os autores do modelo, o resíduo S54 seria responsável pelo posicionamento do ATP no sítio ativo da

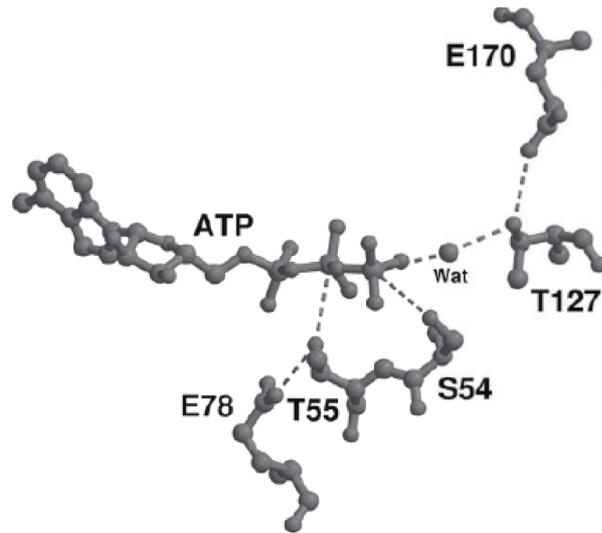


Figura 1.5: Resíduos catalíticos da apirase de batata. O sítio apresenta os principais resíduos envolvidos na interação da enzima com o substrato (ATP) e com a molécula de água nucleofílica (Wat). (Fonte: Retirado de Kozakiewicz, et al, 2008) .

enzima (KOZAKIEWICZ et al., 2008). O resíduo T127 seria o responsável pela ativação de uma molécula de água que faria o ataque nucleofílico do grupo γ -fosfato, enquanto que o resíduo E170 ajudaria no posicionamento da água nucleofílica (KOZAKIEWICZ et al., 2008).

Com relação ao segundo grupo de resíduos, que atuaria na hidrólise do β -fosfato, o resíduo T55 seria o responsável pelo ataque nucleofílico (KOZAKIEWICZ et al., 2008). Os autores do modelo não deixam claro se haveria a participação de uma molécula de água nesse segundo ataque nucleofílico, como ocorreria na hidrólise do ATP.

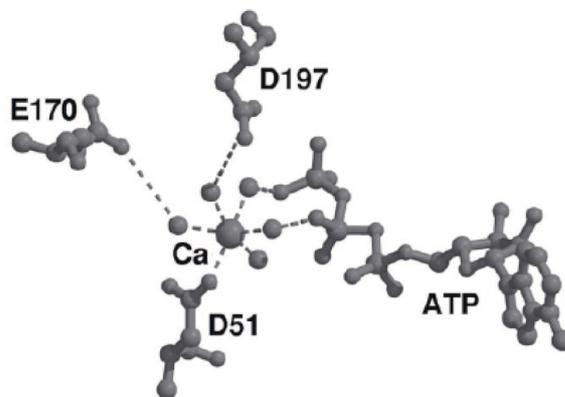


Figura 1.6: Posicionamento do íon metal e moléculas de água. Destaque para o íon cálcio, e as moléculas de água e resíduos que atuam no posicionamento do íon no sítio ativo. (Fonte: Retirado de Kozakiewicz, et al, 2008) .

No modelo catalítico de Kozakiewicz e colaboradores (2008), assim como no modelo

de Zebisch e Sträter (2008), o íon metálico não interage diretamente com os resíduos da enzima mas por meio de moléculas de água. Essas mesmas moléculas de água, juntamente com os resíduos D50, D197 e E170, seriam responsáveis pelo posicionamento do íon metálico no sítio ativo da enzima (ver figura 1.6) (KOZAKIEWICZ et al., 2008).

1.5 Modelagem Comparativa

Sabe-se que a função de uma proteína é determinada por sua sequência de aminoácidos e pela sua estrutura tridimensional (3D) (SANCHEZ et al., 2000). O advento das tecnologias de sequenciamento do ácido desoxirribonucleico (DNA, do inglês *deoxyribonucleic acid*), levou ao surgimento de muitos projetos de sequenciamento de genomas em larga escala. Isto abriu as portas para a possibilidade de se identificar todo o conjunto de proteínas de um organismo. Porém, esta possibilidade esbarrou no alto custo financeiro dos dois principais métodos experimentais de predição da estrutura tridimensional de proteínas, a cristalografia por difração de raios-X e a espectroscopia de ressonância magnética nuclear. Diante de tal situação, métodos teóricos, surgem como uma importante ferramenta na predição da estrutura 3D de uma proteína (JOHN; SALI, 2003).

As famílias de proteínas são formadas por membros que apresentam duas importantes características: estrutura tridimensional conservada e alta similaridade na estrutura primária. Esta conservação, resultado da evolução, permite que se faça a predição da estrutura 3D de todos os membros de uma família mesmo quando se conhece apenas a estrutura 3D de um único membro (SALI, 1995; SANCHEZ et al., 2000). Isto é um importante fundamento da modelagem comparativa, também conhecida como modelagem por homologia.

A técnica de modelagem comparativa se baseia no alinhamento estrutural da proteína alvo, cuja estrutura 3D não é conhecida, com uma ou mais proteínas cuja estrutura 3D já foi resolvida por métodos experimentais, e as coordenadas cartesianas de seus átomos estão depositadas em um banco de dados como o PDB (*Protein Data Bank*) (BERMAN et al., 2000). Estas últimas são conhecidas como proteínas moldes e são evolutivamente relacionadas com a proteína alvo (SALI, 1995). É importante salientar que a técnica depende da identidade entre as estruturas primárias da proteína alvo e dos moldes. Quanto

maior essa identidade, maior é a acurácia do método (SANCHEZ et al., 2000; BAKER; SALI, 2001).

A modelagem comparativa consiste em quatro etapas (VERLI, 2014):

- (i) **Busca e seleção dos moldes:** Nessa etapa, são feitos alinhamentos locais entre as estruturas primárias da proteína alvo e das proteínas depositadas no PDB. O objetivo é encontrar no banco de dados proteínas que tenham alta similaridade com a proteína alvo. Para tal, é usado um programa de alinhamento local entre sequências de proteínas, como o BLAST (*Basic Local Alignment Search Tool*) (ALTSCHUL et al., 1990).

O BLASTP faz alinhamentos pareados entre as regiões com maior número de resíduos idênticos ou similares das sequências de proteínas que estão sendo analisadas. Para cada alinhamento são computados o número de resíduos idênticos e de resíduos similares em cada posição do alinhamento.

No processo de seleção dos moldes é levado em conta o grau de similaridade entre as estruturas primárias da proteína alvo e das proteínas relacionadas pelo BLASTP. A chance de uma determinada proteína do PDB ser selecionada como molde será tanto maior quanto maior for o número de resíduos idênticos e similares encontrados no alinhamento e a cobertura desse alinhamento (BAKER; SALI, 2001; GOLIATT, 2007).

- (ii) **Alinhamento global entre as sequências:** Após a seleção dos moldes, são feitos alinhamentos globais entre a sequência de aminoácidos da proteína alvo e as sequências das proteínas moldes. O alinhamento global envolve toda a extensão das sequências que estão sendo alinhadas e busca maximizar o número de pareamentos entre resíduos idênticos ou similares. Esta etapa é de grande importância pois a construção do modelo depende fundamentalmente de uma ótima correspondência entre os resíduos da sequência alvo e os resíduos da sequência molde (SANCHEZ et al., 2000). Um exemplo de programa para alinhamento global é o Clustal (SIEVERS et al., 2011).

- (iii) **Construção do modelo:** A construção do modelo pelo método de satisfação de restrições espaciais se dá por meio do uso de diferentes tipos de informações. A seguir uma descrição sucinta do método implementado no programa Modeller (SALI, 1995;

ESWAR et al., 2006) usado, no presente trabalho, na construção dos modelos da Sm1 e CD39.

O ponto de partida da construção do modelo são as informações obtidas a partir do alinhamento global entre a sequência da proteína alvo e a sequência da proteína molde. Informações como comprimento das ligações e ângulos diedrais são transferidos da proteína molde para o modelo (SALI, 1995; ESWAR et al., 2006).

Em seguida, as informações do alinhamento são avaliadas por funções densidade probabilidade que foram obtidas a partir de análises estatísticas de uma base de dados formada pelo alinhamento de 416 proteínas, de diferentes famílias, com estruturas 3D resolvidas experimentalmente (SALI, 1995; ESWAR et al., 2006). Essas funções expressam relações quantitativas, como a correlação entre duas distâncias C_{α} - C_{α} equivalentes de proteínas similares ou a correlação entre ângulos diedrais equivalentes de duas proteínas similares, e são usadas como restrições espaciais (SALI, 1995; ESWAR et al., 2006).

O passo seguinte do método é a combinação das restrições espaciais com termos de energia de campos de força em uma função objetivo. O modelo é obtido pela otimização da função objetivo por meio de métodos que visam minimizar as violações das restrições espaciais (*e.g.* gradiente conjugado, máximo declive) (SALI, 1995; ESWAR et al., 2006).

Uma outra técnica usada na geração dos modelos é a dos corpos rígidos. Um exemplo de programa que usa essa técnica é o Swiss-Model (BIASINI et al., 2014). Na modelagem comparativa por corpos rígidos, a construção do modelo tridimensional da proteína é dividida em três partes: modelagem de regiões conservadas, inserção de alças e inserção de cadeias laterais.

As regiões estruturalmente conservadas da proteína são definidas por meio de preditores de estrutura secundária. Essas regiões são então alinhadas com a estrutura do molde (VERLI, 2014).

Em seguida são inseridas no modelo as regiões variáveis, que geralmente correspondem a alças que conectam as regiões conservadas. Banco de dados de estruturas onde as alças são classificadas pelo número de resíduos e pelo tipo de estrutura secundária que conectam, podem ser usados na obtenção da cadeia

principal (VERLI, 2014).

Após a inserção das alças, são feitas a inserção das cadeias laterais dos aminoácidos. Nesta etapa, bibliotecas de rotâmeros podem ser usadas (VERLI, 2014). Essas bibliotecas diminuem o tempo computacional pois facilitam a busca por conformações preferenciais de cadeias laterais que apresentam ângulos de torção favoráveis (XIONG, 2006).

- (iv) **Validação do modelo:** Nesta etapa, é avaliada a qualidade do(s) modelo(s) gerado(s). A qualidade do(s) modelo(s) depende de avaliações estereoquímica, energia (DOPE (*Discrete Optimized Protein Energy*), calculado pelo programa Modeller), conservação de domínios e ligações (VERLI, 2014). São usados na avaliação programas como o Procheck (LASKOWSKI et al., 1983), Errat (COLOVOS; YEATES, 1993) e Verify3D (EISENBERG; LÜTHY; BOWIE, 1997). Caso sejam identificados erros na estrutura 3D de um modelo, torna-se necessário refazer alguns dos passos anteriores do processo, seja desde ajustar o alinhamento global até a escolha de novos moldes.

Além de propor um modelo tridimensional para a Sm1 e CD39, o presente trabalho pretende verificar mudanças conformacionais dessas enzimas por meio de simulações computacionais de dinâmica molecular. A seguir, será apresentada uma breve descrição dos fundamentos teóricos da técnica de dinâmica molecular usada no presente trabalho.

1.6 Dinâmica Molecular

A dinâmica molecular é uma técnica de simulação computacional que fornece informações a respeito dos movimentos em um sistema de partículas em função do tempo (KARPLUS; MCCAMMON, 2002). A partir da posição inicial de cada partícula do sistema, são geradas novas posições a cada intervalo de tempo constituindo a trajetória do sistema. Essa trajetória depende do potencial de interação entre as partículas e da resolução das equações de movimento da mecânica clássica (GOLIATT, 2007).

Simulações de dinâmica molecular têm sido aplicadas a uma variedade de problemas envolvendo macromoléculas biológicas, em especial proteínas. Problemas como interação proteína-ligante, mecanismos de reação enzimática, desnaturação e enovelamento de

proteínas, mudanças na conformação da proteína, são exemplos de aplicação dessa técnica (PONDER; CASE, 2003; CASE et al., 2005).

O potencial de interação entre os átomos em um sistema é representado por uma função chamada de Campo de Força. A forma geral de um campo de força para um sistema com N átomos é dada pela equação 1.1.

$$V(r_i) = \frac{1}{2} \sum_{n=1}^{N_b} K_{bn}(b_n - b_{0n})^2 + \frac{1}{2} \sum_{n=1}^{N_\theta} K_{\theta n}(\theta_n - \theta_{0n})^2 + \frac{1}{2} \sum_{n=1}^{N_\xi} K_{\xi n}(\xi_n - \xi_{0n})^2 + \frac{1}{2} \sum_{n=1}^{N_\tau} K_{\tau n}[1 + \cos(n\tau_n - \delta_n)] + \sum_{i \leq j}^N \left(\frac{B_{ij}}{r_{ij}} \right)^{12} - \left(\frac{A_{ij}}{r_{ij}} \right)^6 + \sum_{i \leq j}^N \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad (1.1)$$

onde o primeiro termo é o potencial harmônico linear para N_b ligações químicas e seus movimentos vibracionais, k_b é a constante de Hook, b é o comprimento instantâneo da ligação e b_0 é o comprimento de equilíbrio. O segundo termo é o potencial harmônico angular para N_θ ângulos formados por três átomos ligados consecutivamente, K_θ é a constante de Hook, θ é o ângulo entre as duas ligações e θ_0 é o ângulo de equilíbrio. O terceiro termo é o potencial diedral impróprio que se refere aos N_ξ ângulos entre os planos formados por quatro átomos, K_ξ é a constante de Hook, ξ é o ângulo entre os planos e ξ_0 é o ângulo de equilíbrio. O quarto termo é o potencial diedral próprio que descreve as N_τ rotações das ligações químicas formadas, K_τ constante que define a barreira de rotação das ligações químicas, n é o número de mínimos, τ é a variação do ângulo de rotação e σ é o ângulo de diferença de fase (0° ou 180°). O quinto termo é o potencial de Lennard-Jones, onde $\left(\frac{A_{ij}}{r_{ij}}\right)^{12}$ modela a repulsão entre átomos muito próximos, $\left(\frac{A_{ij}}{r_{ij}}\right)^6$ modela as interações atrativas de van der Waals e r_{ij} é a distância entre os átomos i e j . O sexto termo é o potencial de Coulomb onde q_i e q_j são as cargas parciais dos átomos i e j , respectivamente, r_{ij} é a distância entre os átomos i e j , ϵ_r constante dielétrica do meio e ϵ_0 é a permissividade elétrica do vácuo.

Os quatro primeiros termos modelam interações entre átomos ligados e os dois últimos termos modelam interações entre átomos não-ligados. A figura 1.7 mostra um esquema dos potenciais calculados pelo campo de força.

Sabendo como os átomos de um sistema interagem entre si, é necessário, então, definir a posição e velocidade iniciais dos átomos que compõem o sistema. Para uma proteína, a posição inicial dos átomos é dada pelas coordenadas espaciais, seja em uma estrutura resolvida experimentalmente ou em um modelo 3D. Para a atribuição das velocidades iniciais, um procedimento geralmente utilizado é o uso da distribuição de Maxwell-Boltzmann dependente da temperatura do sistema (MARTÍNEZ, 2007). Como

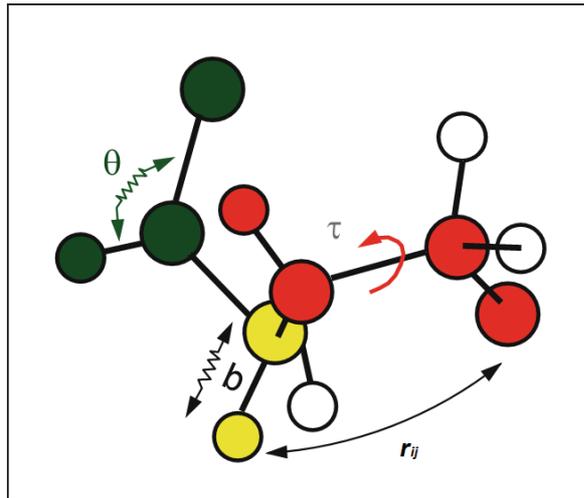


Figura 1.7: Esquema dos potenciais de um campo de força. No esquema, b representa o potencial harmônico de ligação, θ representa o potencial angular, τ representa o potencial torcional e r_{ij} a distância entre átomos não ligados. (Fonte: Schlick, 2010) .

a força sobre cada átomo do sistema é o oposto do gradiente da energia potencial, ou seja, $\vec{F} = -\nabla V(r)$, é possível calcular a aceleração de cada átomo.

Estabelecidas as condições iniciais e a aceleração \vec{a} de cada átomo do sistema, pode-se atualizar as posições e velocidades nas etapas subsequentes da simulação. Isso pode ser feito por meio da resolução das equações newtonianas de movimento. Para o instante $t + \Delta t$ as velocidades e posições dos átomos de um sistema podem ser atualizados pelas equações 1.2 e 1.3 dadas abaixo.

$$\vec{v}(t + \Delta t) = \vec{v}(t) + \vec{a}(t)\Delta t + \dots \quad (1.2)$$

$$\vec{s}(t + \Delta t) = \vec{s}(t) + \vec{v}(t)\Delta t + \frac{\vec{a}(t)\Delta t^2}{2} + \dots \quad (1.3)$$

Dadas as novas posições e velocidades no instante $t + \Delta t$, os potenciais interatômicos são recalculados, definindo novas forças e novas acelerações. Dessa forma, novas posições e velocidades serão recalculadas em um novo ciclo. Este ciclo resulta na trajetória de todos os átomos do sistema ao longo do tempo (MARTÍNEZ, 2007).

A equação 1.2 é a expansão da série de Taylor com erro associado da ordem de Δt^2 , enquanto que a equação 1.3 é a expansão da série de Taylor em torno do ponto $s(t)$ com erro associado da ordem de Δt^3 . Afim de reduzir o erro no cálculo das velocidades, outros algoritmos foram propostos como o Velocity-Verlet (MARTÍNEZ, 2007).

No algoritmo Velocity-Verlet as posições são atualizadas em $t + \Delta t$ e as velocidades em $t + \frac{\Delta t}{2}$. Desta forma a velocidade e as posições dos átomos de um sistema são atualizados pelas respectivas equações 1.4 e 1.5 descritas a seguir:

$$\vec{v}(t + \frac{\Delta t}{2}) = \vec{v}(t) + \frac{\vec{a}(t)\Delta t}{2} \quad (1.4)$$

$$\vec{s}(t + \Delta t) = \vec{s}(t) + \vec{v}(t + \frac{\Delta t}{2})\Delta t \quad (1.5)$$

Antes do início da simulação é necessário a preparação do sistema. As etapas de preparação podem variar conforme o sistema biológico a ser modelado. Em geral, são necessárias uma etapa de adição do solvente e uma de adição dos íons. Caso a simulação envolva uma proteína transmembranar, as etapas seriam as seguintes: (i) adição da membrana, (ii) solvatação, (iii) adição dos íons e (iv) definição de condições periódicas de contorno.

- (i) **Adição da membrana:** Diversas proteínas estão associadas à membranas de organelas celulares ou a membrana plasmática. Isso significa que durante a montagem de sistemas que envolvam proteínas transmembranares é necessário o uso de um modelo parametrizado de membrana.

Um modelo de membrana que é muito usado em simulações de dinâmica molecular é o modelo POPC (palmitoil-oleil-fosfatidilcolina) por ser um dos tipos de lipídeos mais frequentes na composição de membranas biológicas (NELSON; COX, 2009), além de estar parametrizado para vários campos de força.

- (ii) **Solvatação:** Uma importante característica das simulações de dinâmica molecular é a inclusão explícita da água como solvente. Um modelo de água muito usado nessas simulações é o TIP3P (JORGENSEN et al., 1983). Neste modelo de água são considerados três sítios de interação: um sítio no oxigênio e dois nos hidrogênios. O ângulo entre as ligações OH é de 104.52° , a distância OH é de $0,9572\text{\AA}$, com carga parcial de $-0,834$ para o oxigênio e $0,417$ para os hidrogênios (JORGENSEN et al., 1983).

Outros modelos de água também foram propostos como o TIP4P (JORGENSEN et al., 1983) e o TIP5P (MAHONEY; JORGENSEN, 2000). A principal diferença

entre esses modelos está no número de sítios de interação (JORGENSEN et al., 1983; MAHONEY; JORGENSEN, 2000). Quanto mais sítios de interação maior o custo computacional (RICK; STUART; BERNE, 1994). A escolha do modelo a ser usado depende do campo de força a ser usado na simulação.

- (iii) **Adição dos íons:** Os íons possuem papel fundamental na fisiologia celular. Para muitas enzimas a atividade catalítica depende da presença de íons. Nesta etapa, são adicionados íons buscando reproduzir as concentrações fisiológicas, como por exemplo as concentrações de Na^+ e K^+ no sangue.
- (iv) **Condições periódicas de contorno:** Pela figura 1.8 podemos ver que as partículas do sistema estão restritas a uma caixa tridimensional chamada de célula de simulação. Portanto, as partículas estão imersas em um sistema finito sujeitas aos efeitos de borda, ou seja, a cada passo da simulação, partículas do sistema poderiam desaparecer através da borda. Afim de evitar este problema, são usadas condições periódicas de contorno. Isto permite que no entorno da célula de simulação sejam adicionadas réplicas idênticas fazendo com que o sistema seja simulado como se fosse de tamanho infinito. Assim, toda partícula na célula de simulação tem uma cópia exata nas células vizinhas. Este arranjo permite que quando uma partícula deixe a célula de simulação seja substituída por uma outra idêntica, conservando assim o número de partículas.

A simulação de dinâmica molecular envolve quatro etapas (VERLI, 2014): (i) minimização, (ii) tratamento térmico, (iii) equilíbrio e (iv) produção. Essas etapas são descritas a seguir:

- (i) **Minimização:** Durante a preparação do sistema para simulação, podem ocorrer problemas relacionados à geometria inicial dos átomos que compõem o sistema. Por exemplo, quando o programa adiciona o solvente pode ocorrer que um átomo de hidrogênio de uma molécula de água fique muito próximo de um outro átomo de hidrogênio da cadeia lateral de um resíduo, levando a uma repulsão eletrostática. Para evitar tais problemas, é necessário ajustes na configuração inicial do sistema por meio da minimização da energia potencial. Métodos como o Gradiente Conjugado e Máximo Declive podem ser usados no processo de minimização da energia potencial do sistema (MACKAY; CROSS; HAGLER, 1989).

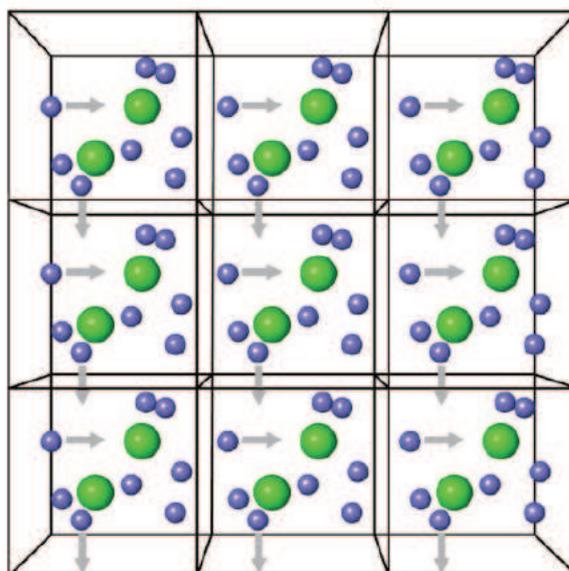


Figura 1.8: Representação das condições periódicas de contorno. No centro a caixa de simulação e em torno as réplicas da mesma. (Fonte: Verli, 2014) .

- (ii) **Tratamento térmico:** O tratamento térmico consiste na elevação gradual da temperatura do sistema de um valor pré-determinado até a temperatura desejada para a simulação. Esta etapa é importante quando a simulação envolve uma proteína cuja a estrutura 3D foi obtida em temperatura diferente da fisiológica, como na cristalografia de raios-X, ou então estruturas obtidas por modelagem comparativa.
- (iii) **Equilibração:** Nesta etapa, busca-se equilibrar a energia potencial de interação das partículas do sistema com a energia cinética, evitando variações bruscas da energia total do sistema na etapa de produção (SCHLICK, 2010). Assim como na minimização, a equilibração pode ser dividida em duas fases, uma com restrição no movimento da proteína e outra sem nenhuma restrição.
- (iv) **Produção:** A etapa de produção caracteriza-se pelo início da geração das conformações que irão compor a trajetória do sistema ao longo do tempo de simulação.

1.7 Docking receptor-ligante

Muito dos processos celulares envolve a interação entre uma proteína (receptor) e uma pequena molécula (ligante), que atua como substrato ou inibidor, ou desencadeando uma resposta biológica (agonista) ou bloqueando a resposta (antagonista). Por conta disso, o

conhecimento do modo de interação entre ligantes e receptores torna-se fundamental no planejamento racional de fármacos. Uma técnica computacional que tem sido usada na previsão do modo de interação e da afinidade de um ligante por um receptor é a técnica denominada atracamento receptor-ligante ou *docking* receptor-ligante (VERLI, 2014). Ao longo do texto iremos nos referir ao *docking* receptor-ligante simplesmente como *docking*.

As metodologias computacionais de *docking* estão baseadas no modelo chave-fechadura proposto por Emil Fisher (1894) (FISCHER, 1894). Por este modelo, o ligante é a “chave”, o receptor é a “fechadura”, e o “buraco da fechadura” seria o sítio de ligação. A interação entre o ligante e o receptor estaria relacionada com as ações de “abrir ou trancar” a porta. Entretanto, as estruturas do receptor proteico e do ligante são flexíveis, de tal maneira que durante a interação tanto o ligante como o receptor podem apresentar mudanças em suas conformações (VERLI, 2014).

Para explicar as mudanças na conformação do receptor foi proposto por Koshland o modelo do encaixe induzido, onde o ligante induz alterações conformacionais no receptor a fim de otimizar a interação receptor-ligante (MEIRELES et al., 2011). Em contraposição foi proposto o modelo da seleção conformacional. Por este modelo, um receptor apresenta um conjunto de conformações e o ligante seleciona uma determinada conformação do receptor, deslocando o equilíbrio químico de tal forma que a conformação selecionada tenha a sua proporção aumentada no conjunto das conformações (MEIRELES et al., 2011; VERLI, 2014).

Estudos experimentais mostram que os dois modelos podem coexistir, a seleção conformacional desempenhando um papel dominante na definição das mudanças conformacionais de larga escala e que são exploradas pelo ligante, especialmente na fase inicial da ligação (MEIRELES et al., 2011; VERLI, 2014). Já o encaixe induzido explicaria as alterações locais e específicas que ocorrem no receptor induzidas pelo ligante, tais como rearranjos das cadeias laterais dos resíduos próximos ao ligante (MEIRELES et al., 2011; VERLI, 2014). A figura 1.9 mostra um esquema dos modelos de reconhecimento molecular.

O reconhecimento molecular entre proteína e ligante está baseado na complementaridade de características físico-químicas e estruturais. As características físico-químicas definem o grau de afinidade e de especificidade do ligante pela proteína e estão relacionadas com interações intermoleculares, tais como ligações de hidrogênio,

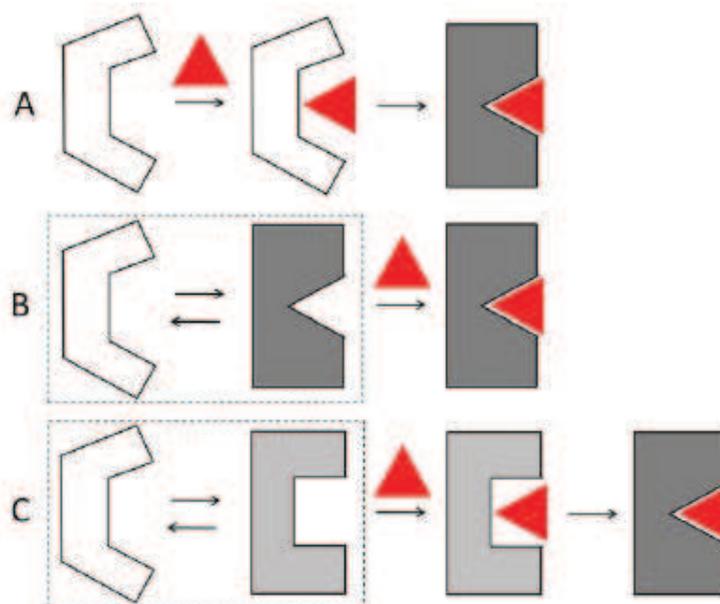


Figura 1.9: Esquema dos modelos de reconhecimento molecular. Em (A) o modelo encaixe induzido, (B) modelo seleção conformacional e em (C) seleção conformacional seguido de encaixe induzido. (Fonte: Meireles, 2011) .

interações de van der Waals, interações eletrostáticas e interações hidrofóbicas. Já as características estruturais estão associadas com as variações na orientação espacial das ligações químicas (VERLI, 2014).

O processo de reconhecimento molecular é dirigido por uma combinação de efeitos entálpicos e entrópicos que podem ser estimados através da energia livre de ligação de Gibbs e que, por sua vez, está relacionada com a constante de equilíbrio de ligação, ou seja:

$$\Delta G_{lig} = \Delta H - T\Delta S = -RT\ln K_{eq} \quad (1.6)$$

onde ΔH é a variação da entalpia, T é a temperatura absoluta, ΔS é a variação da entropia e R é a constante universal dos gases. A constante de equilíbrio de ligação é determinada experimentalmente e pode ser representada pela constante de associação:

$$K_a = \frac{[RL]}{[R][L]} \quad (1.7)$$

onde $[R]$ é a concentração do receptor, $[L]$ a concentração do ligante e $[RL]$ a concentração do complexo receptor-ligante.

No *docking* os algoritmos de busca são usados para explorar a superfície da energia livre de ligação a fim de encontrar as melhores conformações do ligante durante o *docking*. Se os

efeitos entálpicos e entrópicos forem corretamente modelados pela função de energia, então o mínimo global da superfície de energia estará associado ao modo de ligação receptor-ligante encontrado experimentalmente. Devido às aproximações introduzidas no modelo de interação molecular, nem sempre o mínimo global satisfaz este importante requisito (GUEDES; MAGALHÃES; DARDENNE, 2014).

Existem três estratégias relativas à flexibilidade do receptor e do ligante: (1) a proteína é considerada rígida e apenas os graus de liberdade translacionais e rotacionais do ligante são explorados, (2) a proteína é considerada rígida e todos os graus de liberdade do ligante são explorados e, (3) a proteína é considerada totalmente ou parcialmente flexível e todos os graus de liberdade do ligante são considerados. A maioria dos algoritmos de *docking* considera a proteína rígida e o ligante flexível, permitindo rotações das ligações do tipo sp^3 sendo que o comprimento e os ângulos das ligações permanecem fixados (GUEDES; MAGALHÃES; DARDENNE, 2014).

Os algoritmos de busca são classificados em três grupos de acordo com a metodologia usada para explorar a flexibilidade do ligante: (1) métodos de busca sistemática, (2) métodos de busca determinística e (3) e métodos de busca estocástica (BROOIJMANS; KUNTZ, 2003).

Os algoritmos de busca sistemática exploram todos os graus de liberdade durante a pesquisa, como a rotação de todos diedrais do ligante (BROOIJMANS; KUNTZ, 2003). Um exemplo de busca sistemática são os algoritmos de busca incremental, um tipo de abordagem onde o ligante é separado em fragmentos. Em um primeiro momento um fragmento âncora é encaixado no sítio do receptor e, posteriormente, todos os demais fragmentos são adicionados, sequencialmente, por meio de ligações químicas rotacionáveis ao fragmento âncora até a reconstrução total do ligante (BROOIJMANS; KUNTZ, 2003). Esta estratégia é amplamente empregada na abordagem (*de novo*) em que novos ligantes são desenvolvidos por meio do *docking* de fragmentos dentro do sítio de ligação do receptor (GUEDES; MAGALHÃES; DARDENNE, 2014).

Nos métodos determinísticos o estado inicial do sistema determina as modificações que ocorrem no estado seguinte, de tal maneira que o resultado final é altamente dependente da estrutura inicial. Para um estado inicial, mantido os mesmos parâmetros, é obtido o mesmo resultado na saída (BROOIJMANS; KUNTZ, 2003). Métodos de minimização de energia e simulação por dinâmica molecular são exemplos de busca determinística

(BROOIJMANS; KUNTZ, 2003). A minimização de energia tem como objetivo explorar a superfície de energia potencial usando a direção associada com o gradiente da energia, levando o sistema ao mínimo local mais próximo. O principal problema associado a esta estratégia é a sua incapacidade para superar as barreiras de energia para alcançar outros mínimos locais e assim explorar a superfície de energia de forma mais eficiente (GUEDES; MAGALHÃES; DARDENNE, 2014). Já a dinâmica molecular apresenta as seguintes vantagens: (1) a inclusão de solvente explícito e (2) todos graus de liberdade do ligante e da proteína podem ser consideradas durante a simulação. As desvantagens são o alto custo computacional e a tendência do sistema ficar preso em um mínimo local (GUEDES; MAGALHÃES; DARDENNE, 2014).

Os algoritmos de busca estocástica alteram aleatoriamente todos os graus de liberdade do ligante (translação, rotação e conformacional), gerando uma grande diversidade de soluções. Essas conformações do ligante são avaliadas de acordo com um critério probabilístico para decidir se cada conformação deve ou não ser rejeitada. A principal desvantagem é que não há qualquer garantia de convergência para a solução ótima, e várias execuções do algoritmo são necessários para maximizar a probabilidade de encontrar um mínimo de energia global. Métodos como Monte Carlo e Algoritmos Genéticos são exemplos de busca estocástica (GUEDES; MAGALHÃES; DARDENNE, 2014).

Algoritmos de busca estocásticos fazem mudanças aleatórias, geralmente mudando um grau de liberdade do ligante de cada vez. Essas conformações do ligante são avaliadas de acordo com um critério probabilístico para decidir se cada conformação deve ou não ser rejeitada. A fim de melhorar a convergência, são feitas várias execuções do algoritmo a fim de maximizar a probabilidade de encontrar um mínimo de energia global. Exemplos de pesquisas estocásticas são Monte Carlo (MC) métodos e algoritmos evolucionários (GUEDES; MAGALHÃES; DARDENNE, 2014; BROOIJMANS; KUNTZ, 2003).

Durante o processo do *docking* os algoritmos de busca geram muitas conformações do ligante. Funções de avaliação (*score*) são usadas para avaliar a qualidade das conformações geradas. Uma função de avaliação deve ter três objetivos: (1) ser capaz de determinar o modo de ligação, (2) prever a afinidade de ligação entre a proteína e o ligante (energia livre de ligação), e (3) classificar corretamente os ligantes em ativos e inativos (triagem virtual) (HUANG; GRINTER; ZOU, 2010).

As funções de avaliação podem ser divididas em três grupos: baseadas em campo de

força, empíricas e baseadas em conhecimento (HUANG; GRINTER; ZOU, 2010).

Funções de avaliação baseadas em campo de força são obtidas a partir de um campo de força usados em simulações de dinâmica molecular. Este tipo de função de avaliação considera as energias de interação receptor-ligante (termos do campo de força para átomos não ligados) e a energia interna do ligante (termos para átomos ligados e não ligados) (GUEDES; MAGALHÃES; DARDENNE, 2014).

As funções empíricas procuram reproduzir a afinidade de ligação obtidas de dados experimentais. A maioria das funções de pontuação empírica é baseada no modelo em que a afinidade de ligação pode ser decomposta em termos que refletem as várias contribuições para a ligação. Os coeficientes associados aos termos funcionais podem ser obtidos por meio de uma análise de regressão de dados a respeito da afinidade de ligação de estruturas obtidas experimentalmente (JR.; DIAS, 2008; GUEDES; MAGALHÃES; DARDENNE, 2014).

Funções baseadas em conhecimento são construídas a partir de análises estatísticas entre pares de átomos de complexos receptor-ligante resolvidos experimentalmente. Seus termos são derivados a partir das frequências observadas de interações específicas pré-definidas entre os pares de átomos de cada complexo, o que permite capturar efeitos de interações mais específicas. Uma desvantagem das funções baseadas em conhecimentos é que dependem de um conjunto de treinamento bastante amplo para a parametrização (VERLI, 2014).

1.8 Justificativa

A esquistosomose é uma doença negligenciada causada por helmintos do gênero *Schistosoma*. No Brasil a espécie responsável pela doença é o *S. mansoni*, com uma estimativa de 4 a 6 milhões de infectados. O fármaco mais usado no tratamento da doença é o praziquantel. Porém, há relatos na literatura de casos de indivíduos infectados que não responderam ao praziquantel. Tal situação levanta a necessidade de desenvolvimento de novos fármacos para o tratamento da doença, bem como a seleção de proteínas do parasita que possam ser alvos de esquistossomicidas.

Um grupo de proteínas apresentadas como potenciais alvos moleculares de esquistossomicidas é o das NTPDases. Estas enzimas hidrolisam nucleotídeos di e

trifosfatados (e.g. ADP e ATP) em presença de cátions bivalentes, e estão descritas em diferentes organismos.

Nos mamíferos foram descritas oito isoformas de NTPDases, quatro estariam presentes na superfície externa das células, sendo, por isso, chamadas de Ecto-NTPDases. A isoforma 1 de humanos (HsNTPDase1), mais conhecida como CD39, está presente nas células do endotélio dos vasos sanguíneos e seriam responsáveis pela regulação das concentrações extracelulares de ATP e ADP. A estrutura das Ecto-NTPDases é caracterizada por um domínio extracelular (ECD - Extracellular Domain), onde estão localizadas as ACR e o sítio ativo, e dois domínios transmembranares, hélices TM1 e TM2, localizados próximos as extremidades N-terminal (Nter) e C-terminal (Cter). Os domínios transmembranares ancoram a enzima na superfície externa das células (GRINTHAL; GUIDOTTI, 2006; ROBSON; SEVIGNY; ZIMMERMANN, 2006; KNOWLES, 2011).

No parasita *S. mansoni* foi descrita uma NTPDase ancorada na membrana plasmática das células do tegumento do verme adulto e conhecida como SmATPDase1 ou Sm1. É a segunda proteína mais expressa no tegumento dos vermes adultos (CASTRO-BORGES et al., 2011). Devido a localização e a importância dos nucleotídeos di e trifosfatos na ativação hemostática e de células do sistema imunológico, foi sugerido que a Sm1 estaria envolvida na regulação das concentrações de nucleotídeos em torno do parasita, o que contribuiria com a sua evasão do sistema imunológico do hospedeiro mamífero (VASCONCELOS et al., 1993; VASCONCELOS et al., 1996; DEMARCO et al., 2003; LEVANO-GARCIA et al., 2007; BHARDWAJ; SKELLY, 2009). A estrutura da Sm1 seria semelhante as Ecto-NTPDases de mamíferos, inclusive a CD39, com um domínio ECD, onde estão localizadas as ACR e o sítio ativo, e duas hélices transmembranares (TM1 e TM2) (DEMARCO et al., 2003).

Baseado nos pressupostos acima, tem sido proposto o uso da Sm1 como possível alvo molecular de novos esquistosomicidas. Porém, devido à semelhança estrutural entre a Sm1 e a CD39 se faz necessário um estudo comparativo da estrutura tridimensional das duas enzimas. No presente trabalho propomos uma análise comparativa dos modelos 3D da Sm1 e CD39 por meio de simulações de dinâmica molecular e *docking* receptor-ligante.

2 OBJETIVOS

2.1 Objetivo Geral

Realizar uma análise comparativa entre a NTPDase1 do parasita *Schistosoma mansoni* (SmATPDase1 ou Sm1) e a NTPDase1 de humanos (CD39) em busca de diferenças conformacionais em cada uma das proteínas e entre elas que possam ser relevantes em estudos de *docking* receptor-ligante.

2.2 Objetivos Específicos

- Construção dos modelos tridimensionais completos da Sm1 e CD39.
- Identificar nos modelos cavidades com potencial drogável.
- Acompanhar ao longo de trajetórias de dinâmica molecular as modificações estruturais em toda a proteína e nas cavidades selecionadas, em ambos os modelos.
- Identificar conformações preferenciais em cada simulação por meio de análise de agrupamento.
- Selecionar dos grupos preferenciais as conformações cujas cavidades tenham maior potencial drogável para estudos de *docking* receptor-ligante.
- Avaliar por meio de estudos de *docking* receptor-ligante a afinidade dos ligantes ANP, ARL67 e praziquantel às cavidades selecionadas.

3 METODOLOGIA

3.1 Construção dos modelos tridimensionais

O primeiro passo na construção dos modelos 3D da Sm1 (UniProtID: Q7YTA4) e CD39 (UniProtID: P49961) foi a busca por moldes. Foram feitos alinhamentos da estrutura primária de ambas as enzimas contra o banco de dados do PDB usando a ferramenta BLASTp (<http://www.rcsb.org/pdb/search/advSearch.do>). Os candidatos foram avaliados com base nos valores de identidade, similaridade, cobertura dos alinhamentos, qualidade das estruturas (resolução na cristalografia e se havia fragmentos faltando na estrutura 3D), e presença ou não de ligantes no sítio ativo.

Os melhores resultados obtidos foram os cristais da NTPDase1 (PDB3ZX3)(ZEBISCH et al., 2012) e NTPDase2 (PDB3CJA, PDB3CJ1, PDB3CJ7 e PDB3CJ9)(ZEBISCH; STRÄTER, 2008), todos de *Rattus norvegicus*. O cristal PDB3ZX3 apresentou 37% de identidade e 53% de similaridade com a Sm1, e 74% de identidade e 85% de similaridade com a CD39. Para os demais cristais a identidade foi de 33% e a similaridade de 49% com a Sm1, enquanto que para a CD39 a identidade foi de 46% e similaridade de 62%. A tabela 3.1 apresenta os heteroátomos presentes em cada uma dos cristais analisados.

Tabela 3.1: Heteroátomos presentes nos cristais analisados para a etapa de construção do modelo da Sm1 e CD39.

Heteroátomo	3CJA	3CJ1	3CJ7	3CJ9	3ZX3
Ca ⁺²	x	-	-	x	-
Água	x	x	x	x	x
PO ₄	-	-	-	x	-
AMP	-	-	x	x	-
ANP	x	-	-	-	-

As células marcadas com x indica presença do heteroátomo no molde. As estruturas das colunas 2-4 foram obtidas de (ZEBISCH; STRÄTER, 2008) e a estrutura da coluna 5 foi obtida de (ZEBISCH et al., 2012)

Todos os cristais recuperados no PDB correspondem somente ao domínio extracelular (ECD, do inglês *Extracellular Domain*). Como não foram encontrados moldes para os domínios transmembranares (TM1 e TM2), não foi possível usar programas de modelagem de proteínas membranares, como o Memoir (<http://opig.stats.ox.ac.uk/webapps/memoir/php/index.php>) e o Medeller(<http://opig.stats.ox.ac.uk/webapps/medeller/>)

>). O passo seguinte foi necessário para a modelagem dos domínios transmembranares.

3.1.1 Predição de peptídeo sinal, de domínio transmembranar e de estrutura secundária

Para a predição da estrutura secundária da Sm1 e CD39 foram usados os programas PSIPRED (<http://bioinf.cs.ucl.ac.uk/psipred/>) (MCGUFFIN; BRYSON; JONES, 2000), Jpred3 (<http://www.compbio.dundee.ac.uk/www-jpred/>) (COLE; BARBER; BARTON, 2008) e JUFO (http://www.meilerlab.org/index.php/servers/show?s_id=5) (MEILER; BAKER, 2000). A predição de peptídeo sinal foi feita por meio dos programas SignalP4.0 (<http://www.cbs.dtu.dk/services/SignalP/>) (PETERSEN et al., 2011) e Phobius (<http://phobius.sbc.su.se/>) (KÄLL; KROGH; SONNHAMMER, 2007). Ambos os programas não detectaram regiões de peptídeo sinal para as sequências da Sm1 e CD39.

Para os domínios transmembranares, foram usados os preditores Phobius, TMHMM Server v.2.0 (<http://www.cbs.dtu.dk/services/TMHMM/>) (KROGH et al., 2001) e TMpred (http://www.ch.embnet.org/software/TMPRED_form.html) (HOFMANN; STOFFEL, 1993). A definição de quais resíduos das hélices TM1 e TM2 estariam em contato foi feita com base em dois preditores de contato, específicos para predição de contatos em hélices transmembranares: (i) TMhit web server (<http://bio-cluster.iis.sinica.edu.tw/TMhit/index.php>) (LO et al., 2009); (ii) TMhhcp (<http://protein.cau.edu.cn/tmhhcp/>) (WANG et al., 2011). Todos os contatos aplicados nos modelos da Sm1 e CD39 foram restringidos a uma distância de 8Å entre os carbonos- α , com desvio padrão de 0,5Å. A seleção dos contatos foi feita com base no *score* maior que 0,5 para o preditor TMhhcp, e probabilidade maior que 0,8 para o preditor TMhit.

3.1.2 Alinhamento global

O alinhamento global de cada uma das sequências de aminoácidos da Sm1 e CD39 com os moldes selecionados, PDB3CJA e PDB3ZX3, foi feita no sítio eletrônico do programa Clustal Ω (<http://www.ebi.ac.uk/Tools/msa/clustalo/>) (SIEVERS et al., 2011). Com base nos resultados dos preditores de domínios transmembranares, de estrutura secundária e de contato, e no alinhamento global entre as sequências, foram definidos os intervalos

que compõem cada um dos domínios transmembranares da Sm1 e CD39. As figuras 3.1 e 3.2 mostram o consenso dos preditores de estrutura secundária e de domínios transmembranares para os trechos correspondentes às TM1 e TM2 na Sm1 e CD39.

		TM1																													
Pos_Sm1_ali	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	
Sm1	N	L	H	K	I	L	T	I	L	A	L	I	I	L	V	I	S	F	I	G	L	L	I	I	I	V	L	Y	I	L	
Jpred	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C
Confidence	4	3	7	8	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	7	5	0	6	7	7	
Psipred	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	
Confidence	4	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	9	8	9	8	7	6	6	3	0	3		
Jufo	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	
Confidence	0,51	0,59	0,65	0,67	0,66	0,66	0,67	0,65	0,66	0,63	0,62	0,65	0,64	0,67	0,65	0,62	0,63	0,63	0,65	0,63	0,62	0,57	0,61	0,56	0,55	0,46	0,47	0,47	0,45	0,45	

		TM2																									
Pos_Sm1_ali	508	509	510	511	512	513	514	515	516	517	518	519	520	521	522	523	524	525	526	527	528	529	530	531	532	533	
Sm1	L	F	V	C	L	F	I	L	L	L	L	L	I	I	G	S	I	I	G	L	L	L	T	R	N	F	
Jpred	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	
Confidence	6	0	6	8	9	9	9	9	9	9	9	9	9	9	9	9	8	7	4	4	4	5	5	5	5	4	
Psipred	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	
Confidence	5	7	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	8	7	6	6	4	0	
Jufo	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	
Confidence	0,47	0,48	0,56	0,63	0,64	0,65	0,63	0,64	0,65	0,65	0,64	0,63	0,64	0,65	0,61	0,59	0,61	0,6	0,6	0,6	0,58	0,55	0,5	0,46	0,44	0,4	

Figura 3.1: Resultado dos preditores de SS e TM para a Sm1. Resultado dos preditores de estrutura secundária (em vermelho) e o consenso dos preditores de domínios transmembranares (em verde) para as regiões correspondentes às TM1 e TM2 na Sm1.

		TM1																															
Pos_Seq	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39								
CD39	K	N	I	L	A	I	L	G	F	S	S	I	I	A	V	I	A	L	L	A	V	G	L	T	Q								
Psipred	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H								
confidence	0	3	6	8	8	8	8	7	9	9	9	9	9	9	9	9	9	8	7	7	5	3	0	2	5	7							
Jpred	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C								
confidence	3	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	9	8	6	3	0	3	0	0	5								
JUFO	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C	H	H	H	H	C								
confidence	0,62	0,67	0,65	0,62	0,66	0,64	0,65	0,64	0,60	0,62	0,64	0,65	0,64	0,61	0,60	0,65	0,62	0,57	0,40	0,41	0,41	0,45	0,48	0,42	0,51								

		TM2																															
Pos_Seq	476	477	478	479	480	481	482	483	484	485	486	487	488	489	490	491	492	493	494	495	496	497	498	499	500								
CD39	T	Y	V	F	L	M	V	L	F	S	L	V	L	F	T	V	A	I	I	G	L	L	I	F	H								
Psipred	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C								
confidence	3	2	6	7	8	9	9	9	9	9	9	9	9	9	9	9	9	9	8	7	7	6	3	7	6								
Jpred	C	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C								
confidence	4	4	6	7	8	8	9	8	7	4	7	9	9	9	9	9	9	9	8	6	1	0	0	0	0								
JUFO	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	H	C								
confidence	0,38	0,37	0,47	0,45	0,53	0,52	0,59	0,58	0,64	0,63	0,64	0,63	0,64	0,62	0,63	0,61	0,63	0,60	0,57	0,50	0,52	0,47	0,41	0,34	0,67								

Figura 3.2: Resultado dos preditores de SS e TM para a CD39. Resultado dos preditores de estrutura secundária (em vermelho) e o consenso dos preditores de domínios transmembranares (em verde) para as regiões correspondentes às TM1 e TM2 na CD39.

3.1.3 Geração dos modelos

Para a construção dos modelos das estruturas 3D da Sm1 e CD39, foi usado a versão 9.10 do programa Modeller (ESWAR et al., 2006). Durante a geração dos modelos, foram aplicadas restrições de estrutura secundária e de contato nos trechos que correspondem

as hélices TM1 e TM2 (figuras 3.1 e 3.2). Essas restrições foram configuradas no *script* do Modeller responsável pela geração dos modelos.

Foram gerados 100 modelos com a presença de ANP (phosphoaminophosphonic acid-adenylate ester, análogo do ATP) e Ca^{+2} no sítio ativo. Além disso, foram consideradas seis moléculas de água que, segundo a literatura, são necessárias na estabilização do íon metal e no ataque nucleofílico (ZEBISCH et al., 2013). As seis moléculas de água (O503, O504, O505, O506, O507, O508) foram selecionadas a partir do cristal PDB3CJA (figura 3.3).

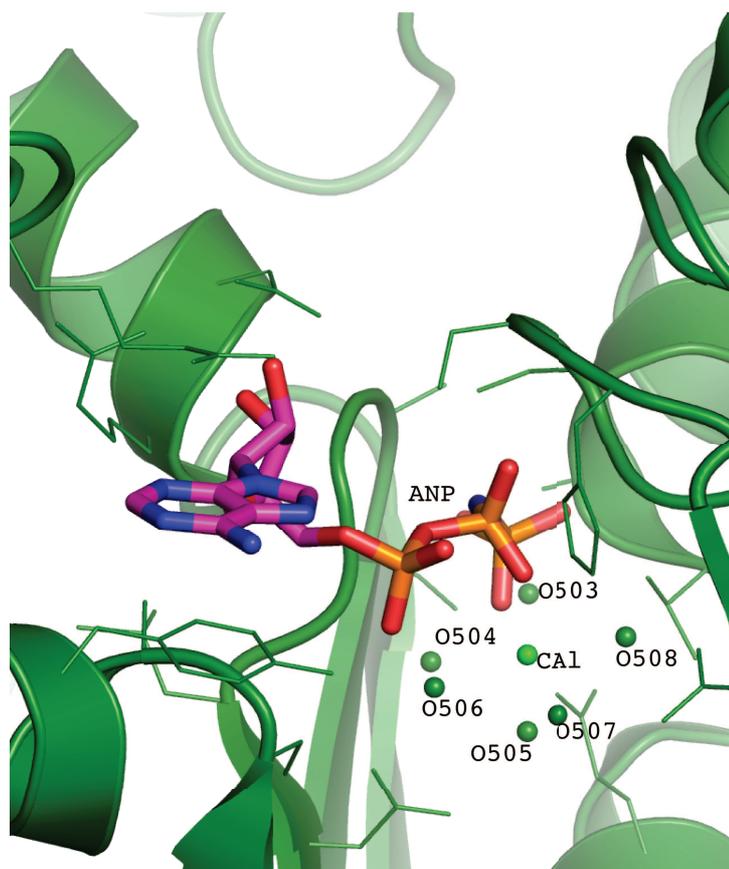


Figura 3.3: Região do sítio ativo do molde PDB3CJA. Em destaque seis moléculas de água (verde), o íon cálcio (verde claro) e o ligante ANP (rosa).

3.1.4 Otimização de alças

Após a geração dos modelos da estrutura 3D da Sm1 e CD39, foi feita otimização das alças que conectam as TM ao domínio ECD. No caso da Sm1 os dois trechos otimizados foram S66-Y73 e G493-S499. No modelo da CD39 os trechos otimizados foram V39-N46 e M461-S475.

3.1.5 Validação dos modelos

Os modelos gerados na etapa anterior, foram submetidos a um processo de validação a fim de avaliar a qualidade das estruturas e selecionar o melhor modelo. Nessa etapa, foram usados os programas Procheck (<http://www.ebi.ac.uk/thornton-srv/software/PROCHECK/>) (LASKOWSKI et al., 1983) e MolProbity (<http://molprobity.biochem.duke.edu/>) (DAVIS et al., 2007) para a avaliação da qualidade estereoquímica. Também foram avaliados os valores do DOPE normalizado e molpdf calculados pelo programa Modeller para cada modelo gerado. Análises adicionais foram feitas com os programas ProSa-web (<https://prosa.services.came.sbg.ac.at/prosa.php>) e SAVes (<http://services.mbi.ucla.edu/SAVES/>).

3.1.6 Adição dos hidrogênios

Após a seleção dos melhores modelos da Sm1 e CD39, iniciou-se o processo de adição de hidrogênios usando a ferramenta *Prep Wizard* do programa Maestro9.3.5 (Schrödinger, LLC, New York, NY). Nesse módulo do Maestro, também foram feitas a otimização e minimização dos hidrogênios em pH 7,4 (valor de simulação para o ambiente fisiológico da proteína). A adição dos hidrogênios nos modelos, foi feita seguindo a verificação do pKa dos resíduos usando a ferramenta *web* PROPKA 3.1 (<http://propka.ki.ku.dk/>) (OLSSON et al., 2011).

Após a adição dos hidrogênios, as estruturas foram novamente avaliadas com relação a qualidade energética e estereoquímica. Apenas uma estrutura da Sm1 e CD39 (com melhores resultados de validação) foi selecionada para as simulações de dinâmica molecular.

3.2 Dinâmica Molecular

Para as simulações de dinâmica molecular (DM) foram preparados quatro sistemas. Dois sistemas envolvendo o modelo da Sm1, um onde o modelo apresentava o Ca^{+2} e ANP no sítio ativo, e um segundo sistema sem o ANP no sítio ativo. Para o modelo da CD39 também foram preparados dois sistemas, um com Ca^{+2} e ANP no sítio ativo da proteína, e outro sem o ANP no sítio ativo. Isso foi feito com o objetivo de observar mudanças conformacionais nas proteínas na presença e ausência do ligante no sítio ativo.

Também foram realizadas simulações de dinâmica molecular para cada um dos moldes PDB3CJA e PDB3ZX3. Por não apresentarem os domínios TM na sua estrutura 3D, foram simulados apenas os domínios ECD.

Todas as simulações envolvendo os moldes e os modelos, foram executadas no programa NAMD versão 2.9 (<<http://www.ks.uiuc.edu/Research/namd/>>) (PHILLIPS et al., 2005), usando o campo de força CHARMM27 (MACKERELL et al., 1998). A seguir serão descritos os passos de preparação dos sistemas e configuração das etapas das simulações de dinâmica molecular realizadas no presente trabalho.

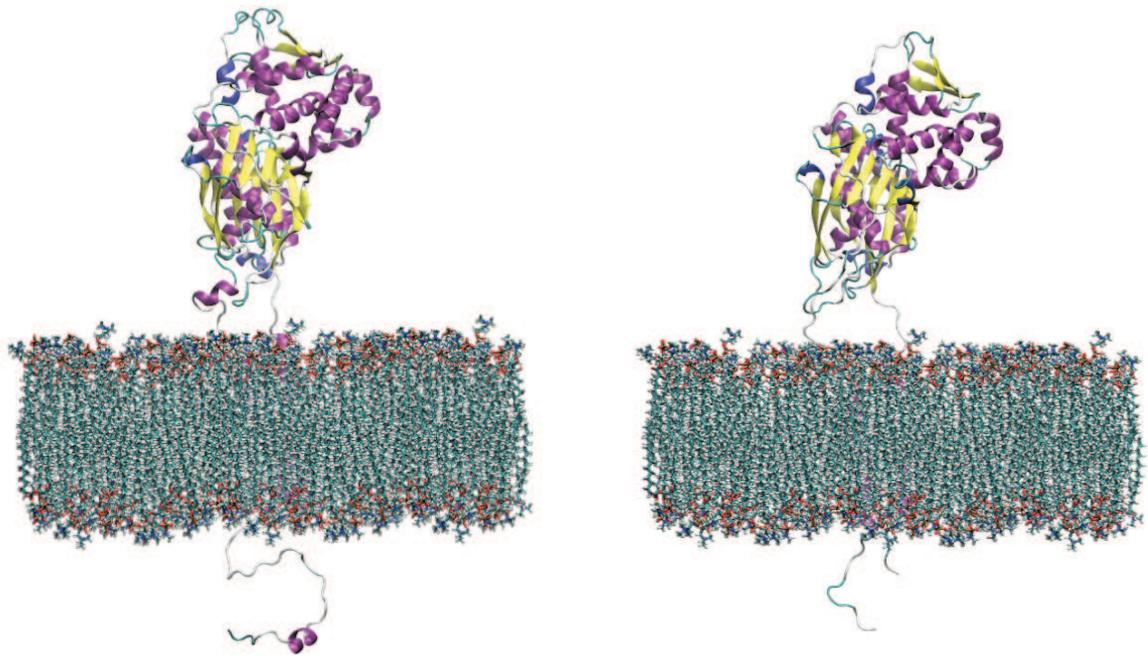
3.2.1 *Preparação dos sistemas*

A preparação de cada um dos quatro sistemas para a simulação de dinâmica molecular seguiu as seguintes etapas:

- (i) **Combinação dos modelos com a membrana:** Após a adição dos hidrogênios, foi feita a combinação do modelo da Sm1 e CD39 com uma membrana do tipo POPC. A escolha de tal modelo deve-se ao fato de a fosfatidilcolina ser o principal componente da membrana plasmática do tegumento do *S. mansoni* (ALLAN; PAYARES; EVANS, 1983; KUSEL; AL-ADHAMI; DOENHOFF, 2007). A dimensão da membrana no plano cartesiano foi de 120x120Å. Cada uma das camadas foi formada por 203 moléculas de POPC. A ferramenta usada na construção do modelo da membrana foi o *Membrane Builder* do programa VMD (*Visual Molecular Dynamics*) (HUMPHREY; DALKE; SCHULTEN, 1996). Nas simulações que envolviam o ligante ANP, a parametrização do ligante com o campo de força do CHARMM27 foi feita usando o portal do *SwissParam* (ZOETE et al., 2011).

Em seguida, a membrana e o modelo foram combinados usando o *script* `combine.tcl` disponível no endereço *web* <http://users.mccammon.ucsd.edu/rlaw/ctbp_workshop_rlaw.htm>. Durante a combinação do modelo com a membrana, foi feito um poro na membrana com um raio de 0.8Å em torno das TM1 e TM2 para ancoragem das mesmas. As figuras 3.4(a) e 3.4(b) mostram a Sm1 e CD39 integradas com a membrana, respectivamente.

- (ii) **Solvatação:** As caixas de simulação envolvendo os moldes e os modelos foram solvatadas usando o modelo de água TIP3P, usando a ferramenta *Add Solvation*



(a) Estrutura da Sm1 acoplada à membrana.

(b) Estrutura da CD39 acoplada à membrana.

Figura 3.4: Estrutura final dos modelos inseridas em membrana POPC.

Box do programa VMD.

Para a solvatação dos sistemas envolvendo os moldes foram adicionados 24.343 moléculas de água. Nos dois sistemas envolvendo a Sm1 foram adicionados 69.212 moléculas de água, enquanto que nos dois sistemas envolvendo a CD39 foram adicionados 55832 moléculas de água.

Todas as caixas de simulação são ortorrômbicas. Na tabela 3.2 são apresentadas as dimensões das caixas.

Tabela 3.2: Dimensões das caixas de simulação.

Caixa	x	y	z
Sm1/Sm1+ANP	118	118	210
CD39/CD39+ANP	118	118	180
3ZX3/3CJA	83	88	104

Dimensões em Ångströms.

- (iii) **Adição dos íons:** O último passo da montagem dos sistemas foi a adição de íons. Os íons foram adicionados na forma de sais de cloreto e as concentrações encontram-se na tabela 3.3. Tais concentrações mimetizam o ambiente sanguíneo do hospedeiro humano e são necessárias na manutenção da atividade enzimática. A adição dos íons

foi feita por meio da ferramenta *Add Ions* do programa VMD, para todos os sistemas envolvendo os moldes e modelos.

Tabela 3.3: Concentrações dos sais.

Sais	Concentrações (mM)
NaCl	150
KCl	5
MgCl ₂	2
CaCl ₂	5

3.2.2 *Etapas da Simulação por Dinâmica molecular*

Para todas as etapas das simulações de dinâmica molecular adotou-se o programa NAMD versão 2.9 (<<http://www.ks.uiuc.edu/Research/namd/>>) (PHILLIPS et al., 2005), usando o campo de força CHARMM27.

O *ensemble* adotado foi o NPT, com aplicação de condições periódicas de contorno. Para integração das interações eletrostáticas de longa distância foi usado o algoritmo PME, com raio de corte de 11Å. Para o controle da temperatura foi adotado o algoritmo *Langevin Piston*, e para o controle da pressão foi adotado o método *Nosé-Hoover Langevin Piston*, ambos implementados no programa NAMD. Para as ligações envolvendo hidrogênios foi usado o algoritmo SHAKE (LEIMKUHLER; SKELL, 1994).

A seguir será descrita a configuração das etapas para cada uma das simulações de dinâmica molecular realizadas no presente trabalho:

- (i) **Minimização:** O método de minimização adotado foi o Gradiente Conjugado. A minimização foi dividida em duas etapas:
- Primeira etapa: foram realizados 150.000 passos com a proteína rígida.
 - Segunda etapa: foram realizados 250.000 passos adicionais com a proteína flexível (*i.e.*, todo o sistema sem restrição).

(ii) **Aquecimento:**

Durante a etapa de aquecimento a temperatura do sistema foi elevada de 277K, temperatura da estrutura no cristal de referência para a construção do modelo 3D, para 310K (temperatura média do meio sanguíneo no hospedeiro humano). O tempo de simulação desta etapa foi de 700ps, com passo de integração de 2fs.

(iii) **Equilíbrio:** Durante a equibração, a temperatura foi mantida constante em 310K e a pressão em 1 atm. Os sistemas foram equibrados nas seguintes condições:

- Modelos: como os sistemas envolvem membranas, a equibração foi dividida em duas etapas. Primeiro com as proteínas rígidas, sendo feita uma equibração somente da membrana durante 10ns, sendo a cada nanossegundo avaliada a relação área por lipídeo através do *plugin Membrane Analysis* (GUIXAGONZALEZ et al., 2014) do programa VMD. Em seguida, foram simulados mais 10ns, sendo 3ns com as cadeias laterais flexíveis e mais 7ns com as proteínas flexíveis.
- Moldes: as simulações foram feitas com a proteína fixa durante 1ns.

(iv) **Produção:** Após a equibração, cada sistema foi simulado com temperatura constante a 310K e 1 atm de pressão. Nesta etapa, os modelos foram simulados durante 25ns enquanto os moldes foram simulados por 40ns.

Após a etapa de produção da dinâmica molecular foram selecionadas conformações das trajetórias a cada 100ps, totalizando 250 conformações mais a última conformação da equibração, que passou a ser considerada a conformação de referência nas etapas de análise de RMSD e de cavidades. Essa última conformação da equibração, no próximo capítulo será indentificada como conformação-0 (ou *frame-0*).

3.3 Relação área por lipídeo

Para simulações de dinâmica molecular envolvendo modelos de membrana, como o POPC, a relação entre a área da membrana e o número de moléculas de lipídeo que a forma é chamada de relação área por lipídeo (Equação 3.1) (EDHOLM; NAGLE, 2005). Os valores da relação área por lipídeo obtidos durante a simulação de dinâmica molecular, devem ser comparados com valores experimentais a fim de validar os resultados obtidos durante a simulação (BENZ et al., 2005; KUCERKA et al., 2015).

$$a(x) = \frac{A(x)}{N} \quad (3.1)$$

onde $A(x)$ é a área da membrana e N é o número de moléculas de lipídeo que formam a membrana (EDHOLM; NAGLE, 2005).

No presente trabalho a análise da relação área por lipídio foi feita usando a ferramenta *Membrane Analysis Tool* do programa VMD.

3.4 Análise do RMSD e Raio de Giro

Para avaliar o comportamento estrutural das proteínas ao longo das simulações, foram calculados o RMSD (do inglês, *Root Mean Square Deviation*) e o raio de giro. As análises do RMSD (Equação 3.2) foram feitas usando o plugin *RMSD Trajectory Tool* do programa VMD, durante toda a trajetória. Para as análises do raio de giro foi usado o programa *g_gyrate* do pacote Gromacs (BERENDSEN; SPOEL; DRUNEN, 1995) versão 4.6.5.

$$RMSD = \sqrt{\sum_i^N \frac{d_i^2}{N}} \quad (3.2)$$

onde N é o número de átomos pareados no alinhamento e d_i é a distância, em Ångström, entre os átomos no pareamento i do alinhamento.

3.5 Predição de cavidades

A predição de cavidades nas estruturas 3D dos modelos foi feita usando o programa DoGSiteScorer versão 2.0.3 para plataforma Linux64 (VOLKAMER et al., 2012a), sendo executado localmente. Para a predição das cavidades, o programa faz uso de uma malha tridimensional cuja a aresta pode ser ajustada entre 0,2 Å e 1,0 Å, juntamente com um filtro gaussiano que é usado para identificar cavidades na superfície da proteína que sejam adequadas para acomodar átomos de ligantes (VOLKAMER et al., 2010). Foi esse filtro que deu origem ao nome do programa (*DoG - Difference of Gaussian*).

O DoGSiteScorer também faz predição da drogabilidade para cada cavidade predita. A drogabilidade seria a capacidade de uma proteína de se ligar com alta afinidade e especificidade a pequenas moléculas com potencial a serem fármacos (HAJDUK; HUTH; TSE, 2005; KRASOWSKI et al., 2011). Para cada cavidade predita em uma proteína pelo DoGSiteScorer, é atribuído uma pontuação que seria a confiança da drogabilidade da cavidade. Essa pontuação é chamada de *drugscore* (VOLKAMER et al., 2012b). Para a predição do valor do *drugscore* o programa faz uso de máquina de vetor suporte (SVM, do inglês, *Support Vector Machine*) (VOLKAMER et al., 2012b), onde são usados os seguintes

descritores: volume, proporção de resíduos apolares e profundidade (VOLKAMER et al., 2012a).

Para a calibração do melhor tamanho de malha para os moldes e modelos, foi avaliado o valor do *drugscore* para diferentes tamanhos de malha entre [0, 2, 1, 0] com intervalos de 0,1. Após os testes, observamos uma convergência do valor do *drugscore* para uma malha de 0,7 Å.

A partir dessas análises foram definidas duas cavidades, sendo uma delas o sítio-ativo (ao longo do texto será chamada de cavidade-1), e a segunda cavidade, que será chamada ao longo do texto de cavidade-2, situada na porção superior do domínio ECD logo acima do sítio ativo (Figura 3.5). É importante destacar que a escolha da cavidade-2 para todas as estruturas foi feita com base na predição de cavidades feita no modelo da Sm1, tomando como referência o alinhamento estrutural entre elas.

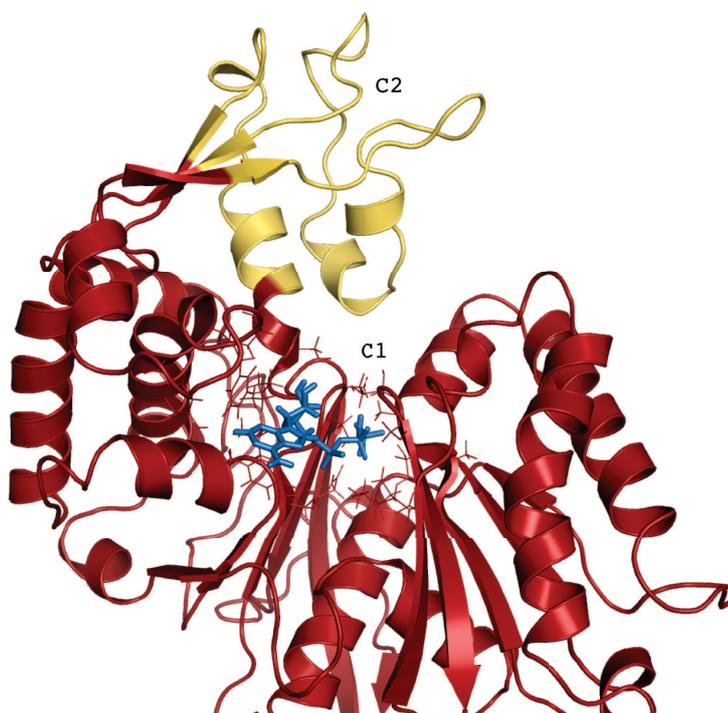


Figura 3.5: Cavidades selecionadas para os estudos de *docking*. Cavidade-1 (C1) abrange a região do sítio ativo com o ANP em azul. Em amarelo a cavidade-2 (C2).

3.6 Análise das cavidades

Nesta etapa, as cavidades foram analisadas ao longo das 250 conformações selecionadas da trajetória da simulação, mais a conformação-0 (conformação-referência). Para tal,

foram gerados arquivos no formato “.pdb” para cada uma das cavidades selecionadas e em cada uma das 250 conformações. Para a análise de drogabilidade, foi usado o parâmetro *drugscore*, calculado pelo programa DoGSiteScorer.

3.7 Conjunto de conformações

Proteínas são estruturas dinâmicas em seu ambiente fisiológico, o que é muitas vezes crucial para as diversas funções biológicas dessas macromoléculas (CHENG et al., 2012). O ligante ao interagir com o sítio de ligação da proteína muitas vezes induz mudanças na conformação, como movimentos das cadeias laterais dos resíduos que formam o sítio de ligação (CHENG et al., 2012).

A inclusão da flexibilidade da proteína continua a ser um desafio na área de *docking* receptor-ligante (GUEDES; MAGALHÃES; DARDENNE, 2014). Hoje é amplamente aceito que as proteínas existem como um conjunto de estados conformacionais preexistentes, em vez de um único estado nativo (CHANGEUX; EDELSTEIN, 2011).

Uma estratégia que tem sido usada para tratar a flexibilidade da proteína em estudos de *docking* receptor-ligante, é o uso de um conjunto de conformações (do inglês, *ensemble docking*) representativas da flexibilidade da proteína (GUEDES; MAGALHÃES; DARDENNE, 2014). Estas conformações podem ser obtidas por métodos experimentais, tais como difração por raios-X (PARK; KUFAREVA; ABAGYAN, 2010) ou RMN (Ressonância Magnética Nuclear) (DAMM; CARLSON, 2007), ou ainda por meio de métodos computacionais como modelagem comparativa (NOVOA et al., 2010), modos normais (SPERANDIO et al., 2010) e dinâmica molecular (NICHOLS et al., 2011).

Um problema relacionado a essa estratégia é o tempo computacional empregado na realização das simulações de *docking* receptor-ligante (VERLI, 2014). O tempo computacional aumenta proporcionalmente com o aumento do número de conformações (KORB et al., 2012; VERLI, 2014).

O emprego de algoritmos de agrupamento tem sido indicado como uma metodologia de pré-seleção das conformações, reduzindo significativamente o seu número sem grande perda da informação sobre a flexibilidade do receptor (VERLI, 2014).

A análise de agrupamento tem sido uma metodologia muito empregada, por exemplo, na análise de trajetórias de simulações por dinâmica molecular. O agrupamento das

diferentes conformações da proteína ao longo da trajetória, possibilita uma compreensão das mudanças significativas que ocorrem na estrutura da proteína (LINDORFF-LARSEN; FERKINGHOFF-BORG, 2009).

Um programa bastante usado em análise de agrupamento de trajetórias é o *g_cluster* do pacote GROMACS. Ele possui diferentes algoritmos de agrupamento e usa como medida de similaridade o RMSD. No presente trabalho, foram feitos testes com diferentes valores de *cutoff* para o RMSD, usando o método *gromos* de agrupamento. A seguir apresentaremos uma proposta de análise de agrupamento baseado na metodologia do GLCM.

3.8 GLCM e análise de agrupamento

Neste trabalho, apresentamos uma proposta de análise de agrupamento das trajetórias baseado na metodologia de processamento de imagem GLCM (do inglês *Gray Level Co-occurrence Matrix*). A intenção era selecionar conformações da Sm1 e CD39, ao longo das trajetórias, para estudos de *docking* receptor-ligante.

Em 1979, Haralick propôs o método GLCM para analisar a textura de imagens baseada na extração de descritores de uma matriz GLCM dos diferentes níveis de cinza de uma imagem (NANNI et al., 2013). Na abordagem do GLCM a textura é escolhida como a mais importante característica para representar imagens. A textura contém informações importantes sobre o arranjo estrutural e as relações entre os *pixels* (NASCIMENTO, 2003).

A metodologia do GLCM faz uso de estatísticas de segunda ordem, pois analisa as co-ocorrências existentes entre um *pixel* e seus vizinhos, ou seja, a análise leva em conta um conjunto de *pixels* que estão relacionados através de algum padrão (SILVA, 2013). Em uma matriz GLCM são computadas as co-ocorrências entre os níveis de cinza presentes em uma imagem, a uma dada distância pps (*pixel pair spacing*) e um ângulo Θ de orientação, que pode ser de 0° , 45° , 90° e 135° (ver figura 3.6) (BARALDI; PARMIGGIANI, 1995; SILVA, 2013).

A Figura 3.7 mostra um exemplo de montagem de uma matriz GLCM. À esquerda uma imagem com oito diferentes níveis (tonalidades) de cinza e à direita a matriz GLCM. A célula da linha 1 e coluna 1 da matriz GLCM possui valor 1, o que indica que na imagem há apenas uma situação onde dois *pixels* com tonalidades igual a 1 são adjacentes

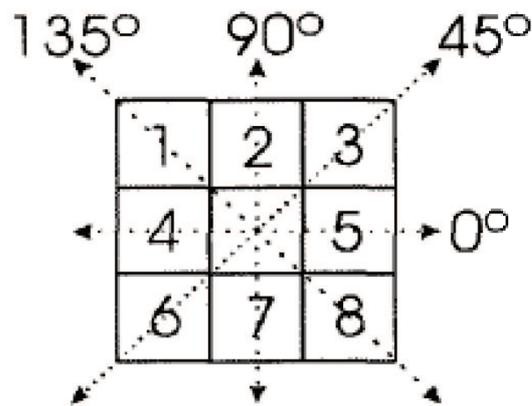


Figura 3.6: Ângulos de orientação para o cálculo das co-ocorrências. (Fonte: Fonte: Retirado de Nascimento, 2003) .

na horizontal. Já na linha 1 e coluna 2 da matriz GLCM, o valor 2 indica que a imagem possui duas situações onde um *pixel* com tonalidade 1 de cinza é adjacente na horizontal a um *pixel* com tonalidade 2 de cinza.

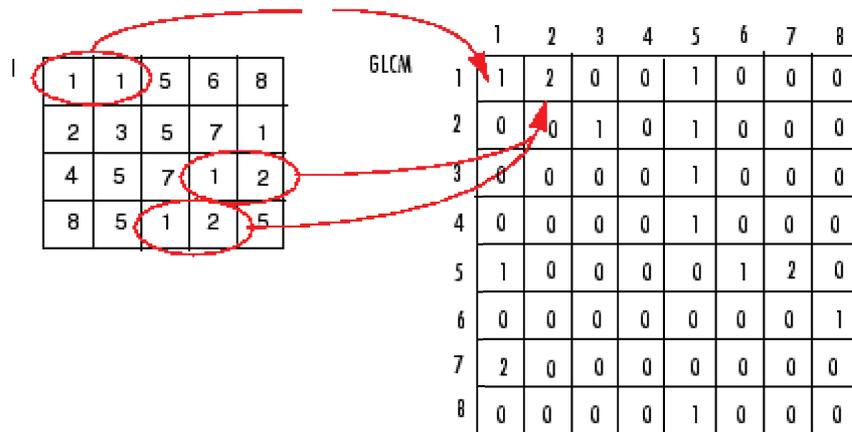


Figura 3.7: Exemplo de construção de uma matriz GLCM. (Fonte: Fonte: Retirado de <http://www.mathworks.com/help/images/ref/graycomatrix.html?requestedDomain=www.mathworks.com>) .

A partir da matriz GLCM podem ser extraídas várias medidas (descritores) que caracterizam a textura de uma imagem. Haralick propôs 14 descritores, entretanto seis descritores foram considerados mais relevantes: contraste, correlação, energia, entropia, homogeneidade e variância (BARALDI; PARMIGGIANI, 1995). A dissimilaridade também tem sido incluída na lista (NASCIMENTO, 2003). Nem sempre será necessário a utilização de todos esses descritores, o que na verdade, pode piorar o desempenho do método ao invés de melhorar (SILVA, 2013).

Essa técnica tem sido empregada em diferentes problemas de classificação, como processamento de imagem de satélites (LU; BATISTELLA, 2005), no reconhecimento facial (ELEYAN; DEMIREL, 2011), análise morfológica de células biológicas (CHEN et al., 2012) e na identificação de tumores cancerígenos (NOBIS et al., 2013).

Matrizes de co-ocorrências também tem sido usada juntamente com algoritmos de agrupamento em problemas como, busca e segmentação em base de dados de imagens médicas (RAMAMURTHY; CHANDRAN, 2012) e de dados obtidos por espectrometria de massa de proteínas (NAEGLE et al., 2012).

A seguir serão apresentadas as etapas das análises de agrupamento das trajetórias das quatro simulações de dinâmica molecular, usando o método GLCM.

- (i) **Cálculo das distâncias entre os átomos do *backbone*:** A primeira etapa consistiu na geração de matrizes de distância entre os átomos do *backbone* para cada uma das duas cavidades analisadas e em cada uma das 251 conformações das quatro simulações, ou seja, foram geradas 502 matrizes para cada simulação. A distância calculada entre os átomos do *backbone* foi a euclidiana. Para essa etapa foi implementado um *script* em Perl para fazer a leitura das coordenadas dos átomos e calcular a distância. Para as etapas seguintes foi implementado um *script* em Python que faz uso de bibliotecas de processamento de imagens e de análise de agrupamento.
- (ii) **Geração e análise das matrizes de co-ocorrência:** A etapa seguinte consistiu na normalização de cada uma das matrizes de distância (M), dada por

$$M_{ij} = \left(\frac{d_{ij}}{d_{max}} \right)^2 \quad (3.3)$$

onde d_{ij} é a distância entre os átomos i e j da estrutura, e d_{max} é a distância máxima computada na matriz.

As distâncias normalizadas são analisadas como se fossem níveis de cinza por meio de matrizes de co-ocorrência. As propriedades de textura usadas na análise das matrizes de co-ocorrência foram o contraste, homogeneidade, correlação e dissimilaridade. A seguir, é apresentada uma breve descrição de cada uma das propriedades.

- **Contraste:** O contraste é uma estimativa das variações locais dos níveis de cinza entre pares de pixels e é dada pela seguinte formulação:

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j)(i - j)^2 \quad (3.4)$$

onde o valor $p(i, j)$ é o elemento da matriz de co-ocorrência cuja linha é i e a coluna é j , e L é a quantidade de níveis de cinza.

- **Correlação:**

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j) \left[\frac{(i - \mu_i)(j - \mu_j)}{(S_i)(S_j)} \right] \quad (3.5)$$

onde μ_i e μ_j são as médias, e S_i e S_j são os desvios-padrão de i e j , respectivamente.

- **Homogeneidade:**

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} \frac{p(i, j)}{1 + (i - j)^2} \quad (3.6)$$

- **Dissimilaridade:**

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p(i, j)|i - j| \quad (3.7)$$

A dissimilaridade e o contraste assumem valores maiores para imagens compostas principalmente por diferentes níveis de cinza, enquanto que a homogeneidade resulta em valores maiores para níveis de cinza similares (NASCIMENTO, 2003).

- (iii) **Agrupamento:** O método de agrupamento usado foi o *K-Means* (HARTIGAN; WONG, 1979). No *K-Means* o número de grupos (K) é definido pelo usuário. Em seguida, cada elemento i a ser agrupado é atribuído ao conjunto cujo centróide estiver mais próximo. O centróide de cada grupo é então atualizado levando-se em conta os elementos que compõem o grupo. Os passos anteriores de atribuição e atualização são repetidos até que não haja mais mudança dos elementos entre os grupos, ou seja, até que não haja mais atualização dos centróides.

A escolha dos centróides iniciais é comumente feita de forma aleatória. Já a atribuição de um elemento a um determinado grupo é feita com base na distância em relação ao centróide. Uma medida de distância frequentemente usada é a distância euclidiana. Seja $(x_1, x_2, x_3, \dots, x_n)$ um conjunto de elementos, onde cada elemento é

um vetor de d dimensões, e $(C_1, C_2, C_3, \dots, C_k)$ K conjuntos. O *K-Means* irá separar os n elementos em K conjuntos buscando minimizar a soma dos quadrados das distâncias dentro do conjunto, isto é:

$$\sum_{i=1}^K \sum_{x \in C_i} (x_i - c_i)^2 \quad (3.8)$$

onde c_i é o centróide do conjunto C_i .

É importante destacar que, no presente trabalho, queremos selecionar conformações dos modelos ao longo das trajetórias por meio de agrupamentos, e que essas conformações serão as que estiverem mais próxima dos centroídes dos grupos e serão chamadas de estruturas médias. Como medida de similaridade entre os grupos foi empregada a silhueta (ROUSSEEUW, 1987). Seja $x_i \in C_i$, onde x é um elemento e C um conjunto, $a(x_i)$ é a dissimilaridade média do elemento x_i em relação a todos os demais elementos do mesmo conjunto C_i , e $b(x_i)$ a dissimilaridade média do elemento x_i em relação aos elementos que pertencem aos demais C_k conjuntos. Assim, a silueta é dada pela formulação 3.9:

$$S(i) = \frac{b(x_i) - a(x_i)}{\max\{a(x_i), b(x_i)\}} \quad (3.9)$$

Dessa forma, as médias $S(i)$ de todos os elementos de um conjunto é uma medida do quão bem estão agrupados os elementos.

É importante destacar que neste trabalho o valores de K não foram definidos durante a realização dos agrupamentos. O valor foi obtido usando o método de Otimização por Enxame de Partículas (*PSO - Particle Swarm Optimization*). O número máximo de grupos foi definido como sendo 20 agrupamentos, e tendo como critério de parada do algoritmo PSO o número máximo de interações igual a 5.

O método PSO é uma técnica estocástica de otimização desenvolvida por Kennedy e Eberhart (1995), e inspirada no comportamento social como os de bando de pássaros e cardume de peixes (KENNEDY; EBERHART, 1995). O processo de otimização no PSO ocorre por meio do aprendizado em grupo (cooperação) e do aprendizado individual (competição). A cooperação e competição são as principais características que guiam o algoritmo do PSO na exploração do espaço de busca

(PESSIN; OSORIO, 2009).

O funcionamento do PSO é dado pela evolução de um enxame de partículas, onde cada partícula possui uma posição x e uma velocidade v . A posição é a representação da solução do problema, e a velocidade é o peso do direcionamento para o deslocamento das partículas. A cada iteração, a posição de cada partícula é atualizada seguindo dois melhores valores: a melhor posição dentro da população e a melhor posição encontrada pela partícula (PESSIN; OSORIO, 2009). A posição da partícula é atualizada até que seja satisfeita a condição de parada.

A posição de cada partícula é atualizada pela seguinte equação 3.10:

$$x_i(t + 1) = x_i(t) + v_i(t) \quad (3.10)$$

onde $x_i(t)$ é a posição da partícula i no instante t , e $v_i(t)$ é a velocidade da partícula no instante t (PESSIN; OSORIO, 2009).

Por sua vez, a velocidade de cada partícula é atualizada pela equação 3.11:

$$v_i(t + 1) = wv_i(t) + c_1r_1(pbest - x_i(t)) + c_2r_2(gbest - x_i(t)) \quad (3.11)$$

onde $v_i(t)$ é a velocidade da partícula no instante t , w é o parâmetro que representa a inércia da partícula, $pbest$ (*particle best*) é a melhor posição encontrada pela partícula, $gbest$ (*group best*) é a melhor posição encontrada pelo grupo, r_1 e r_2 são números aleatórios entre 0 e 1, c_1 e c_2 são chamados de confiança (PESSIN; OSORIO, 2009).

Pela formulação 3.11 é possível ver que a velocidade é um parâmetro que consiste de uma soma de três termos: (i) a velocidade no instante t ; (ii) o componente cognitivo que direciona o indivíduo para a melhor posição encontrada pela partícula até o instante t ; (iii) o componente social, que direciona o indivíduo para a melhor posição encontrada pelo grupo até o instante t (PESSIN; OSORIO, 2009).

Os parâmetros c_1 e c_2 são usados para definir a importância da tendência individual ou social. O PSO padrão, também conhecido como Modelo Completo, trabalha com valores de confiança iguais ($c_1 = c_2$), atribuindo igual importância à experiência do grupo e à experiência individual (PESSIN; OSORIO, 2009). Se o PSO usa apenas

informação do grupo ($c_2 > 0$), não considerando a experiência individual ($c_1 = 0$), o modelo é chamado de Modelo Social. Por fim, quando o PSO usa apenas informação da própria partícula ($c_1 > 0$), não considerando a experiência do grupo ($c_2 = 0$), o modelo é chamado de Modelo Cognitivo (PESSIN; OSORIO, 2009).

Para evitar uma convergência rápida do algoritmo impedindo, assim, uma melhor exploração do espaço de busca, foram introduzidos os parâmetros r_1 e r_2 no ajuste da velocidade (ENGELBRECHT, 2005; PESSIN; OSORIO, 2009). O parâmetro de inércia (w) permite o ajuste no que se refere à busca local ou global. Quanto mais perto de 1 for este valor, mais global é a busca, enquanto valores mais próximos de 0 possibilitam uma busca local (PESSIN; OSORIO, 2009).

Os resultados dos agrupamentos baseados na metodologia do GLCM, foram comparados com análises de agrupamento realizadas no programa *g_cluster* do pacote GROMACS. Para os testes de *docking* foram selecionadas conformações que foram centróides e que apresentaram para as duas cavidades selecionadas os melhores valores do *drugscore* calculado pelo programa DoGSiteScorer.

3.9 Estudos de *docking* receptor-ligante

Para os testes de atracamento molecular (*docking* receptor-ligante) foram usadas conformações da Sm1 e CD39, selecionadas na etapa da análise de agrupamento, contra dois ligantes, o ARL67156 e o praziquantel.

O ARL67156 (6-N,N-dietil-D- β - γ -dibromometileno adenosina trifosfato) é um análogo do ATP (SANSOM et al., 2007). Dados experimentais mostram que este composto é um inibidor que compete com o ATP pelo sítio ativo da CD39, porém, é considerado um inibidor fraco (LÉVESQUE et al., 2007). Já o praziquantel foi selecionado por ser atualmente o principal fármaco usado no tratamento da esquistossomose (DOENHOFF; CIOLI; UTZINGER, 2008).

Para os estudos de *docking* molecular também foi usado o ANP com a finalidade de comparar as energias de interação do ARL67156 e do praziquantel, bem como pelo fato do ANP já estar presente nos modelos das proteínas. Os arquivos com as coordenadas dos átomos de cada um dos ligantes foram baixados do banco de dados *ZINC* <<http://zinc.docking.org/>> (IRWIN; SHOICHET, 2005). A fim de facilitar a diferenciação

entre o ANP do modelo e o ANP baixado do ZINC, ao longo do texto este último será nomeado de ANP-Zinc.

A seguir serão descritas as etapas de preparação das conformações da Sm1 e CD39 que foram selecionadas previamente, a preparação dos ligantes, a preparação das caixas (*grids*), e a configuração das análises de *docking* molecular:

(i) **Preparação das conformações:** Cada uma das conformações selecionadas foram preparadas na ferramenta *PrepWizard* do pacote Maestro9.3.5 (Schrödinger, LLC, New York, NY). Para as conformações decorrentes das simulações onde o ANP estava presente no sítio ativo das proteínas, foram feitas três preparações:

- Preparação 1: proteína, ANP, moléculas de água e cálcio.
- Preparação 2: proteína, moléculas de água e cálcio.
- Preparação 3: somente a proteína.

A primeira preparação foi usada no *redocking* do ANP, a preparação 2 no *docking* do ANP-Zinc e ARL67156, e a preparação 3 no *docking* do ANP-Zinc, ARL67156 e praziquantel.

Para as conformações que vieram das simulações de DM onde o ANP não estava presente no sítio ativo das proteínas, foram feitas duas preparações:

- Preparação 1: proteína, moléculas de água e cálcio.
- Preparação 2: somente proteína.

A primeira a preparação foi usada no *docking* do ANP-Zinc e ARL67156, e a segunda preparação no *docking* do ANP-Zinc, ARL67156 e praziquantel.

As moléculas de água foram selecionadas com base na posição em relação ao íon cálcio e ao resíduo catalítico (E201 na Sm1 e E174 na CD39), e no alinhamento com o molde PDB3CJA.

(ii) **Preparação dos ligantes:** Os ligantes foram preparados usando a ferramenta *LigPrep4.0.5* (Schrödinger, LLC, New York, NY). Para o preparo foi usado o campo de força OPLS2005 (JORGENSEN; MAXWELL; TIRADO-RIVES, 1996), com pH igual a 7,4 ($\pm 0,2$).

(iii) **Preparação das caixas:** As caixas para simulação do *docking* foram preparadas no programa *Glide5.8.5* (Schrödinger, LLC, New York, NY), usando o campo de força OPLS2005. Para as conformações que vieram das simulações onde o ANP estava presente no sítio ativo das proteínas, foram preparadas quatro caixas:

- Caixa 1: cavidade-1 com ANP, moléculas de água e cálcio.
- Caixa 2: cavidade-1 com moléculas de água e cálcio.
- Caixa 3: cavidade-1, somente a proteína.
- Caixa 4: cavidade-2, somente a proteína.

A caixa 1 foi usada para o *redocking* do ANP. As caixas 2 e 3 no *docking* do ARL67156, praziquantel e ANP-Zinc. E a caixa 4 foi usada no *docking* do ARL67156, praziquantel e ANP-Zinc.

Para as conformações que vieram das simulações com ausência do ANP no sítio ativo das proteínas, foram preparadas três caixas:

- Caixa 1: cavidade-1 com moléculas de água e cálcio.
- Caixa 2: cavidade-1, somente a proteína.
- Caixa 3: cavidade-2, somente a proteína.

As caixas 1, 2 e 3 no *docking* do ARL67156, praziquantel e ANP-Zinc.

O programa *Glide* permite que as caixas sejam centralizadas a partir de um conjunto de resíduos, calculando o centro de massa do conjunto de resíduos. Todas as caixas para simulação de *docking* receptor-ligante usadas no presente trabalho foram centralizadas usando resíduos da cavidade-1 ou cavidade-2. A tabela 3.4 apresenta a relação dos resíduos usados na centralização das caixas.

(iv) **Simulações de *Docking*:**

As simulações de *docking* receptor-ligante foram realizadas com o programa *Glide5.8.5* (Schrödinger, LLC, New York, NY). Nas simulações, todos os ligantes foram testados usando a mesma caixa para uma mesma conformação e cavidade, de tal maneira que diferenças na energia de interação e na pontuação (*score*) da conformação do ligante não sejam atribuídas a diferenças entre as caixas.

Tabela 3.4: Resíduos usados na centralização das caixas para simulação do *docking*.

Proteína	Cavidade-1	Cavidade-2
Sm1	D78, T154, A155, E201, D232, S237, Q239, W483, Y397 e F441	K279-K299 e D312-S348
CD39	D54, T131, A132, E174, D213, S218, Q220, W450, F365 e Y408	K258-I277 e V290-F314

O Glide usa o método de busca sistemática para a busca de conformações do ligante (FRIESNER et al., 2004). Este programa usa uma série de filtros hierárquicos para procurar as possíveis localizações do ligante na região do sítio ativo do receptor, que resultam em um conjunto inicial de conformações do ligante. Estas conformações são selecionadas a partir de uma enumeração exaustiva dos mínimos de energia no espaço de ângulos torcionais do ligante. Dadas estas conformações do ligante, são realizadas triagens sobre todo o espaço de busca disponível (conformação, orientação e posição) com o objetivo de localizar conformações promissoras do ligante (FRIESNER et al., 2004).

Após a triagem inicial, as conformações do ligante são minimizadas no domínio do receptor usando uma função de energia baseada em mecânica molecular clássica (neste caso tirada do campo de força OPLS-AA - (FRIESNER et al., 2004)) conjuntamente com um modelo dielétrico dependente da distância. Por fim, são selecionadas de três a seis conformações de mais baixa energia para serem submetidas a uma simulação de Monte Carlo que examina os mínimos torcionais nas proximidades. Em alguns casos este procedimento é necessário para orientar adequadamente grupos periféricos e, ocasionalmente, alterar os ângulos de torção (FRIESNER et al., 2004).

Uma função de pontuação, *GlideScore*, é usada para predizer a afinidade de ligação e ordenar as conformações do ligante (FRIESNER et al., 2004). Existem duas versões do *GlideScore*, a versão SP (*Standart-Precision*) e a versão XP (*Extra-Precision*) (FRIESNER et al., 2004). A versão SP do *GlideScore* é uma função mais indulgente e adequada para identificar ligantes que possuem uma razoável propensão para se ligar na região de busca. Esta versão procura minimizar falsos negativos e é adequada

para rastreio (*virtual screening*) em banco de dados de ligantes (FRIESNER et al., 2004).

A versão XP do *GlideScore* é uma função que impõe maiores penalidades para as conformações que violam princípios físico-químicos, como grupos fortemente polares estarem adequadamente expostos ao solvente (FRIESNER et al., 2004; FRIESNER et al., 2006). Esta versão do *GlideScore* procura minimizar os falsos positivos, e pode ser útil em estudos nos quais apenas um número limitado de compostos serão considerados experimentalmente e as conformações de cada composto testado precisa ter a melhor qualidade possível (FRIESNER et al., 2004; FRIESNER et al., 2006). A versão do *GlideScore* adotada no presente trabalho foi a XP, com o ligante flexível e minimização das conformações geradas durante o *docking*.

3.10 Resumo da metodologia

O fluxograma mostrado na figura 3.8 resume as etapas da metodologia proposta no presente trabalho.

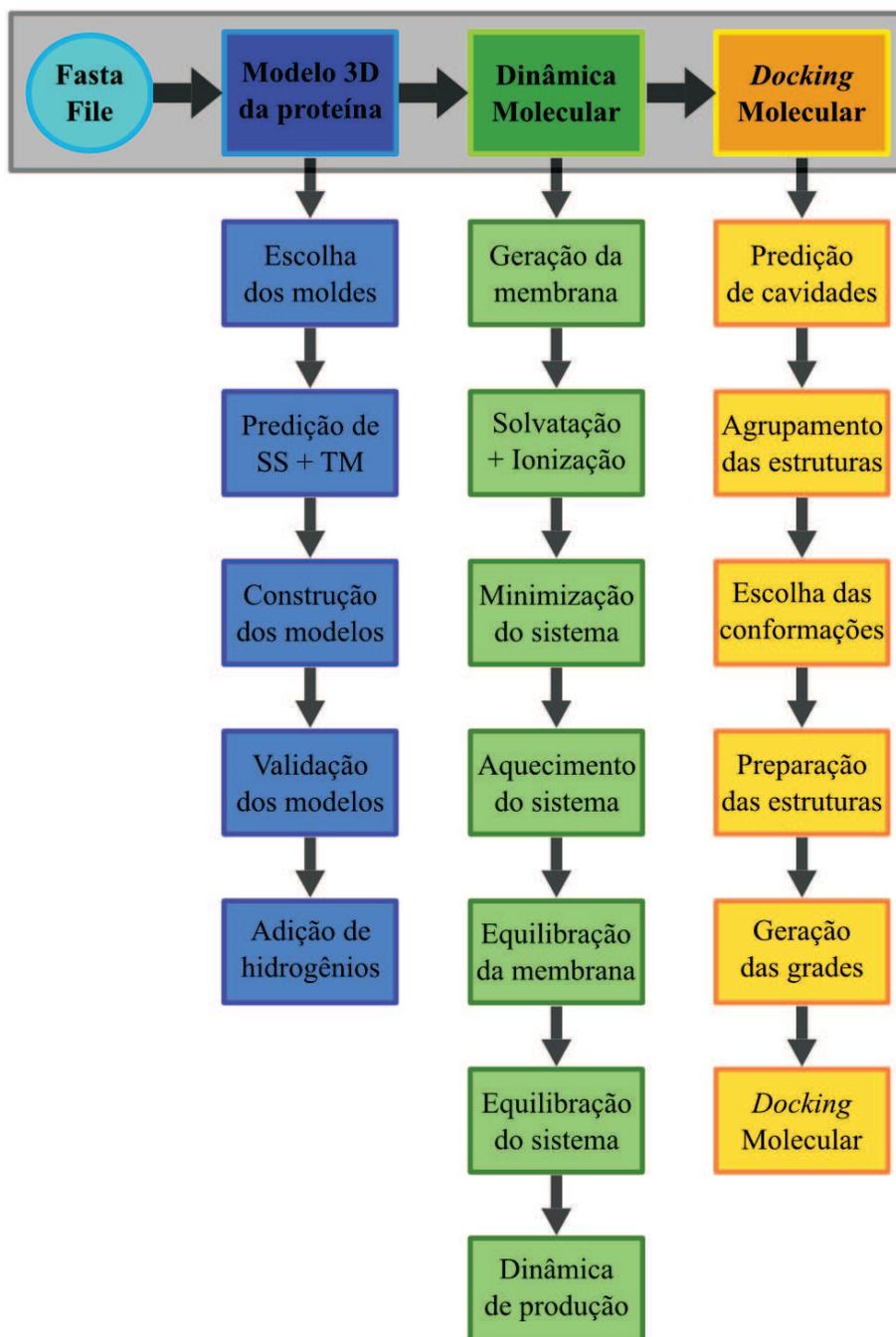


Figura 3.8: Fluxograma da metodologia.

4 RESULTADOS E DISCUSSÃO

4.1 Filogenia das NTPDases de mamíferos e parasitas

A Figura 4.1 apresenta as relações filogenéticas entre as NTPDases de mamíferos e de parasitas. A filogenia foi feita no programa ClustalΩ usando o método de distância *neighbor-joining*, e editada no programa FigTree (<<http://tree.bio.ed.ac.uk/software/figtree/>>). O teste de *Bootstrap* foi calculado no programa MEGA v6.0 (TAMURA et al., 2013). Tais relacionamentos filogenéticos podem fornecer pistas sobre a função das NTPDases de parasitas que ainda não foram demonstradas (SANSOM, 2012). Também podem explicar a semelhança estrutural entre a Sm1 e CD39.

Pela figura 4.1 é possível ver que as Ecto-NTPDases de mamíferos e dos parasitas estão em clados separados. Entretanto, é importante destacar que a Sm1 e a NTPDase de *Plasmodium falciparum* (PfNTPDase) estão filogeneticamente mais relacionadas com as Ecto-NTPDases de mamíferos do que com as demais NTPDases de parasitas. Isto torna-se ainda mais interessante tendo em conta os seguintes aspectos: (i) a presença de duas TM na Sm1 e nas Ecto-NTPDases de mamíferos, em especial na CD39; (ii) são extracelulares, estão ancoradas na superfície externa da membrana plasmática por meio das TM; (iii) devido a sua localização, o provável envolvimento da Sm1 na inibição da ativação plaquetária (VASCONCELOS et al., 1993).

No que diz respeito ao *P. falciparum*, foi demonstrado que as plaquetas são capazes de matar o parasita dentro do eritrócito, sendo que este efeito foi abolido na presença de aspirina (MCMORRAN et al., 2009). A presença de um gene de NTPDase no genoma de *P. falciparum*, e a predição de duas TM, sugerem que esta enzima esteja ancorada na superfície do parasita de maneira similar à CD39 (SANSOM, 2012).

4.2 Os modelos 3D da Sm1 e CD39

As figuras 4.2 e 4.4 mostram o alinhamento global entre os moldes PDB3ZX3 e PDB3CJA e as sequências da Sm1 e CD39, respectivamente. Nas figuras aparecem em destaque as cinco regiões conservadas (ACR), com os resíduos catalíticos marcados em cinza. É possível observar que esses resíduos são conservados entre as sequências, indicando que eles são necessários para a atividade catalítica das NTPDases (ZEBISCH; STRÄTER, 2008).

Experimentos de mutagênese no resíduo E174 da CD39 de humanos, e que corresponde ao resíduo E165 da NTPDase2 de *Rattus norvegicus*, mostraram a perda completa da atividade catalítica da enzima (DROSOPOULOS et al., 2000).

Os modelos tridimensionais da Sm1 (figura 4.3) e CD39 (figura 4.5) propostos neste trabalho contemplam toda a estrutura, ou seja, apresentam um domínio ECD e dois domínios transmembranares TM1 (Nter) e TM2 (Cter). O domínio ECD, por sua vez, é dividido em dois subdomínios ECD-I e ECD-II, com o sítio-ativo localizado em uma cavidade entre esses dois subdomínios. A estrutura tridimensional dos modelos da Sm1 e CD39 é semelhante com a das demais Ecto-NTPDases de mamíferos, como por exemplo a NTPDase2 de *Rattus norvegicus* (RnNTPDase2) (ZIMMERMANN; ZEBISCH; STRÄTER, 2012).

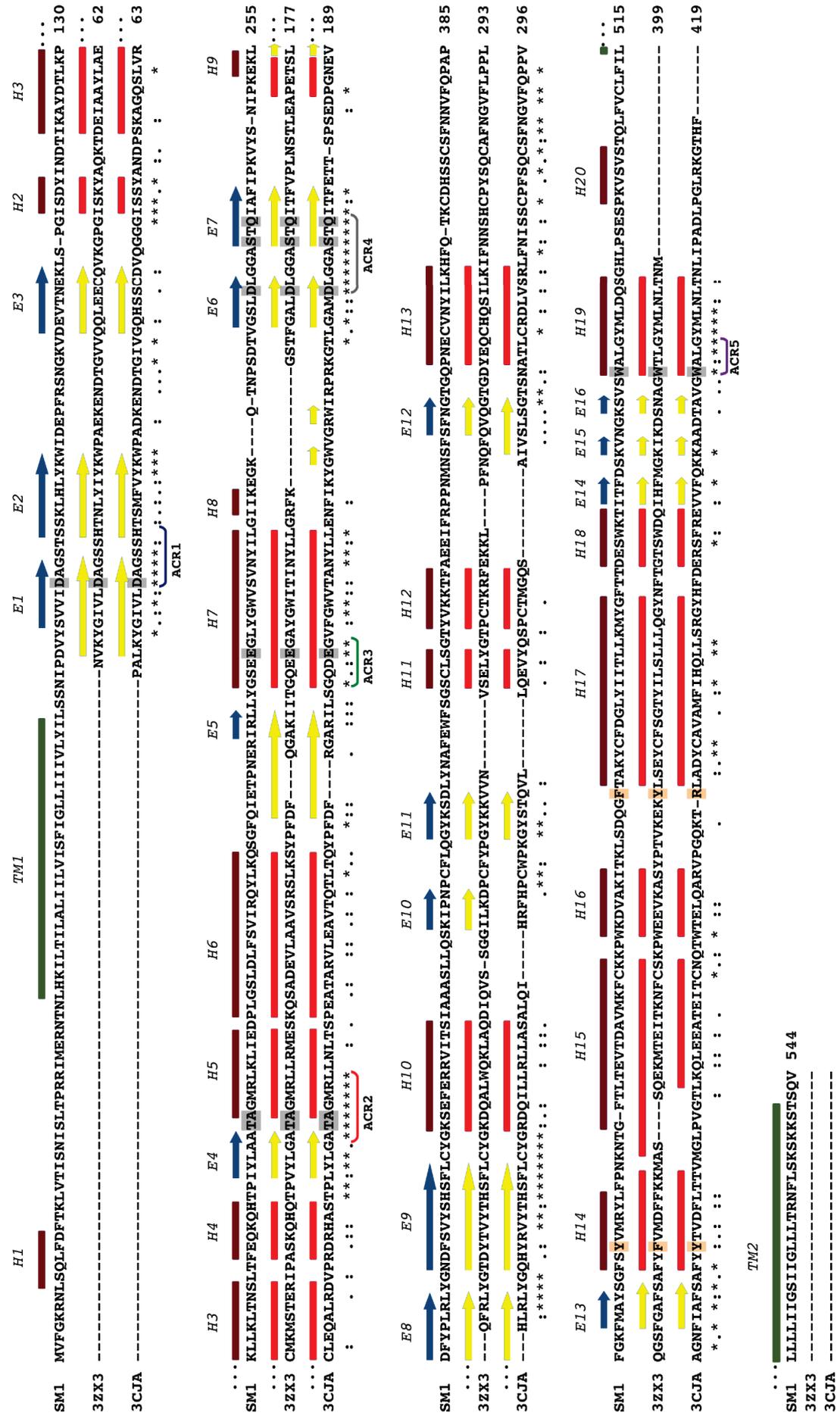


Figura 4.2: Estruturas primária e secundária da Sm1.

Alinhamento da sequência de amniácidos da Sm1 e dos moldes 3ZX3 e 3CJA. Os resíduos que participariam da catálise e presentes nas ACR estão marcados em cinza. As fitas (E) estão representadas por setas e as hélices (H) por retângulos. E e H estão numeradas da extremidade Nter para Cter. Os subdomínios do ECD são: ECD-I (E1-E5, H2-H6, E15-E16 e H19) e ECD-II (E6-E14 e H7-H18). A composição dos subdomínios do ECD é baseada no trabalho de Zebisch e Sträter (2008).

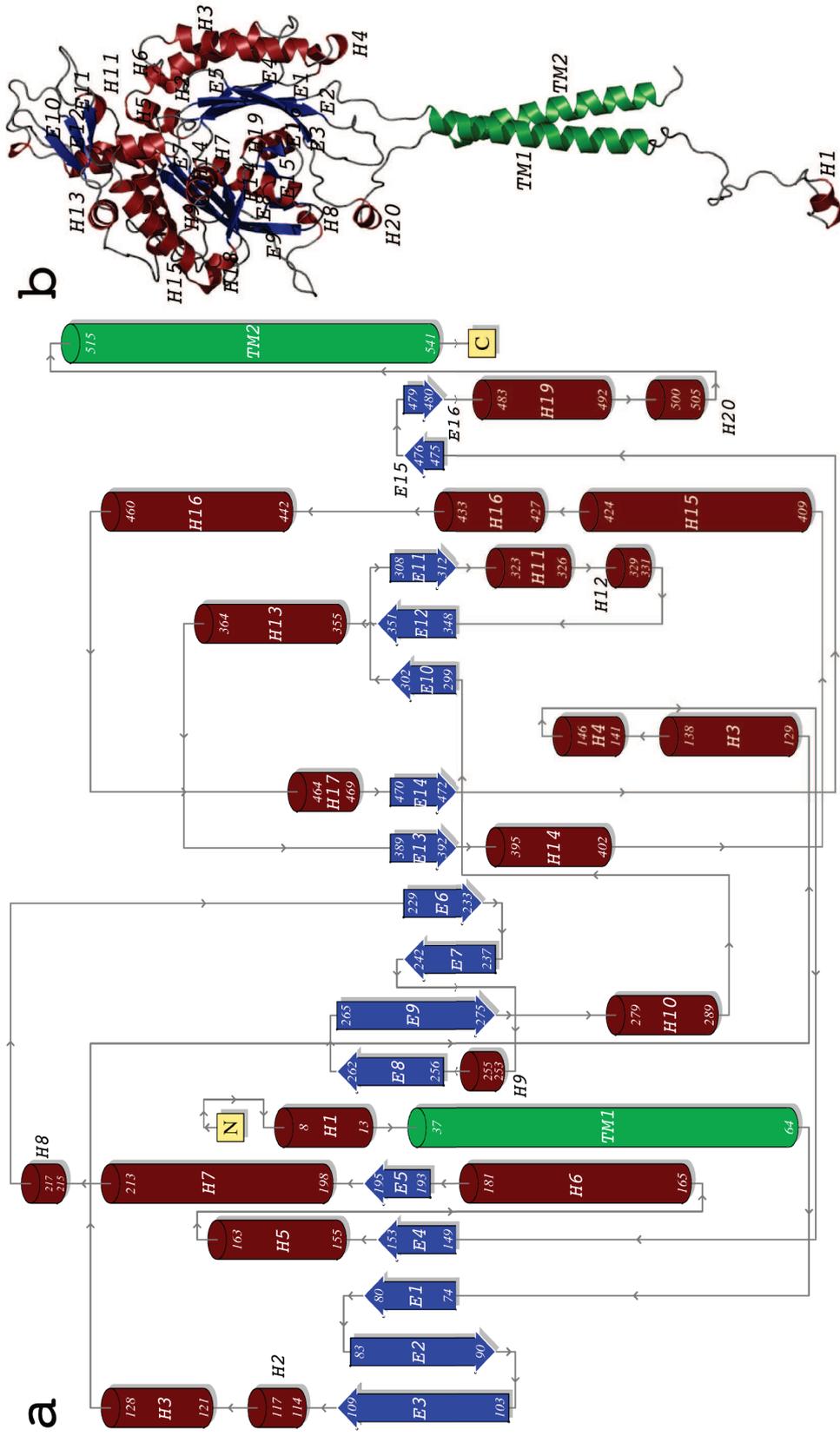


Figura 4.3: Topologia e estrutura terciária da CD39.

(a) Diagrama de topologia do modelo da Sm1. As fitas (E) e hélices (H) estão numeradas da extremidade Nter para Cter. O diagrama foi gerado usando o programa PDBsum (LASKOWSKI; CHISTYAKOV; THORNTON, 2005). (b) Modelo 3D da Sm1. Os subdomínios do ECD são: ECD-I (E1-E5, H2-H6, E15-E16 e H19) e ECD-II (E6-E14 e H7-H18). A composição dos subdomínios do ECD é baseada no trabalho de Zebisch e Sträter (2008).

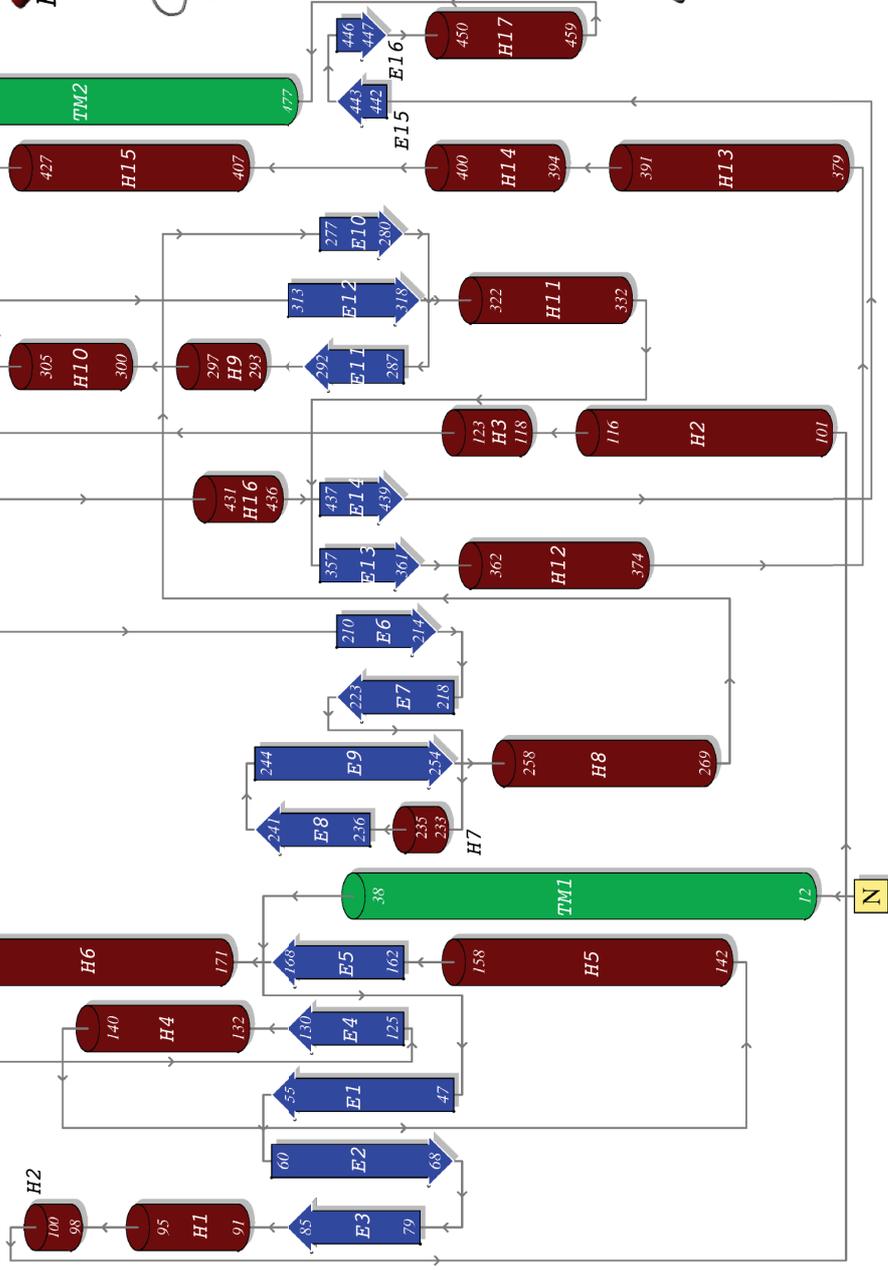
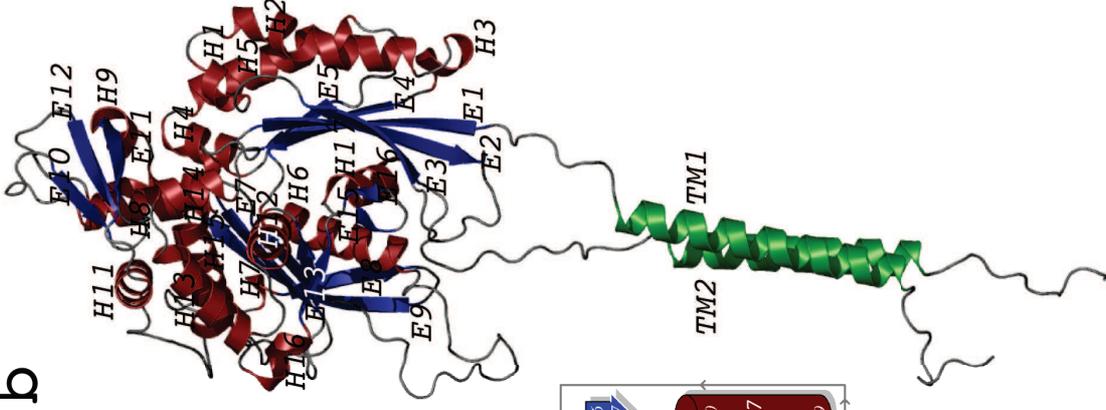
a**b**

Figura 4.5: Topologia e estrutura terciária da CD39.

(a) Diagrama de topologia do modelo da CD39. As fitas (E) e hélices (H) estão numeradas da extremidade Nter para Cter. O diagrama foi gerado usando o programa PDBsum (LASKOWSKI; CHISTYAKOV; THORNTON, 2005). (b) Modelo 3D da CD39. Os subdomínios do ECD são: ECD-I (E1-E5, H1-H5, H17 e E15-E16) e ECD-II (E6-E14 e H6-H16). A composição dos subdomínios do ECD é baseada no trabalho de Zebisch e Sträter (2008).

4.2.1 Validação dos modelos

Para a validação dos modelos 3D da Sm1 e CD39 foram avaliadas três conformações de ambas as proteínas: (conf-1) conformação inicial obtida a partir da modelagem comparativa após a adição de hidrogênios e que foi submetida à simulação de dinâmica molecular; (conf-2) conformação obtida ao final da etapa de minimização da dinâmica molecular; (conf-3) conformação obtida ao final da etapa de aquecimento da dinâmica molecular.

A seguir serão mostrados os gráficos dos perfis do DOPE calculados pelo programa Modeller para as três conformações referentes aos modelos da Sm1, CD39 e moldes. O gráfico mostra o valor do DOPE para cada um dos resíduos dos modelos e dos moldes em função da posição do resíduo no alinhamento múltiplo.

A figura 4.6 mostra o gráfico dos perfis do DOPE para as três conformações da Sm1 e dos moldes PDB3ZX3 e PDB3CJA. Pelo gráfico é possível observar que os perfis das três conformações da Sm1 apresentam-se oscilando entre os perfis dos dois moldes. Também é possível observar uma acomodação da estrutura 3D da Sm1 após as etapas da minimização e do aquecimento. Os dois trechos de resíduos inicial e final e que não são cobertos pelos moldes, correspondem aos domínios transmembranares.

A figura 4.7 mostra o gráfico dos perfis do DOPE para as três conformações da CD39 e dos moldes PDB3ZX3 e PDB3CJA. Pode-se observar que os perfis da CD39 são mais semelhantes aos perfis dos moldes do que os da Sm1, provavelmente em função da maior identidade e similaridade entre as sequências de resíduos da CD39 e dos moldes (figura 4.4). Também é possível verificar que o perfil do DOPE do modelo da CD39 acompanha mais o perfil do molde PDB3ZX3, provavelmente devido a uma maior identidade e a similaridade entre as sequências estas duas proteínas.

Na CD39 também é possível observar uma maior acomodação da estrutura 3D após as etapas da minimização e do aquecimento. Os dois trechos de resíduos inicial e final e que não são cobertos pelos moldes, correspondem aos domínios transmembranares.

Para o processo de validação dos modelos, também foram avaliados os resultados dos programas Procheck (LASKOWSKI et al., 1983), Molprobity (DAVIS et al., 2007), ProSA-Web (WIEDERSTEIN; SIPPL, 2007) e Errat (COLOVOS; YEATES, 1993). A tabela 4.1 apresenta um resumo dos principais parâmetros avaliados no processo de validação dos modelos 3D da Sm1 e CD39. Pela tabela é possível ver que em ambas

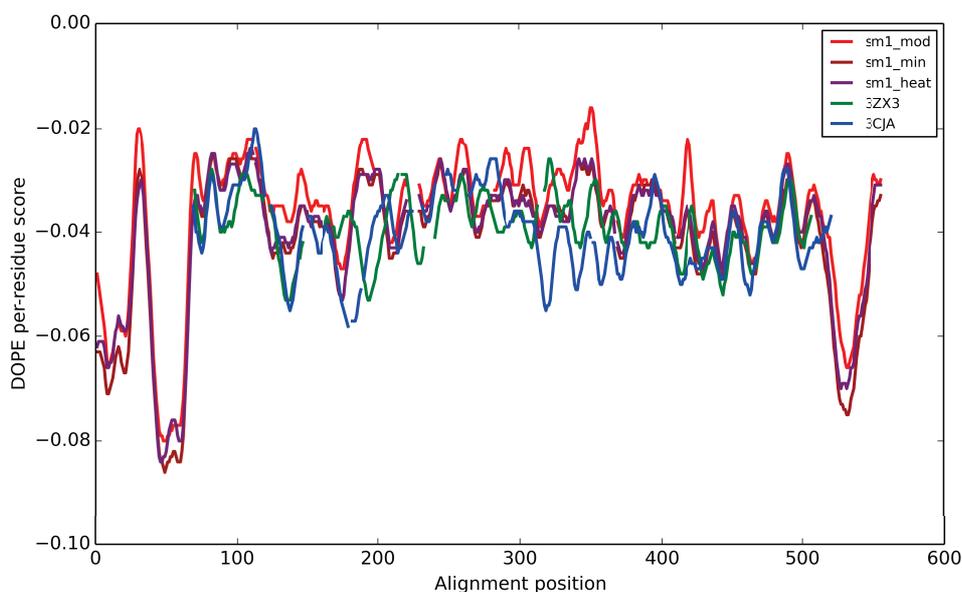


Figura 4.6: Perfil do DOPE para três conformações da Sm1 e os moldes. “sm1_mod” é a conformação inicial obtida a partir da modelagem comparativa após a adição de hidrogênios (conf-1), “sm1_min” é a conformação da Sm1 ao final da minimização da dinâmica molecular (conf-2), e “sm1_heat” é a conformação da Sm1 ao final da etapa de aquecimento da dinâmica molecular (conf-3).

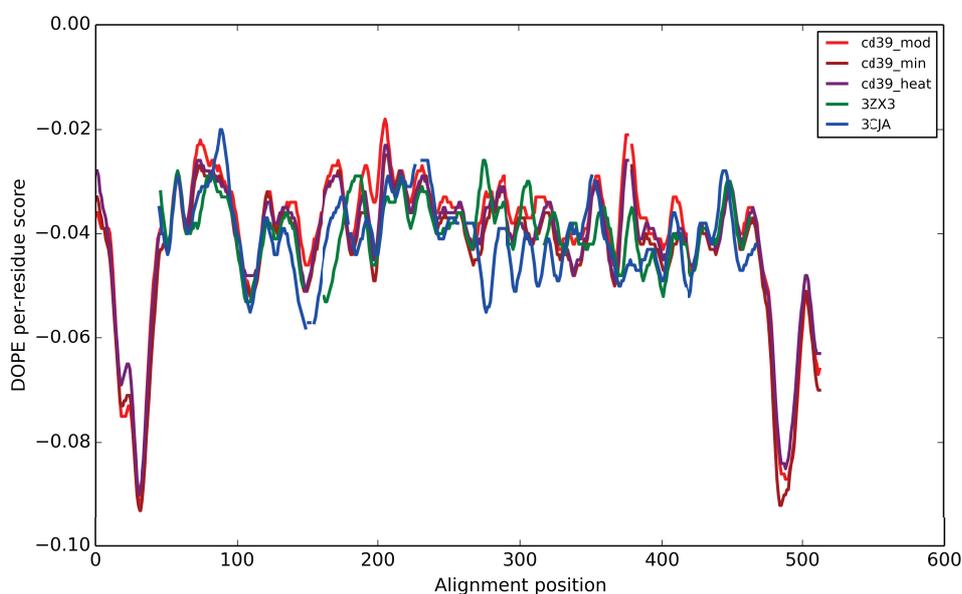


Figura 4.7: Perfil do DOPE para três conformações da CD39 e os moldes. “cd39_mod” é a conformação inicial obtida a partir da modelagem comparativa após a adição de hidrogênios (conf-1), “cd39_min” é a conformação da Sm1 ao final da minimização da dinâmica molecular (conf-2), e “cd39_heat” é a conformação da Sm1 ao final da etapa de aquecimento da dinâmica molecular (conf-3).

as estruturas há uma melhora nos valores do Z -score (calculado pelo ProSA-web) e do Errat (gráficos no Anexo).

Com relação aos resultados do Procheck e do Molprobit, os valores apresentados na tabela apontam uma tendência de piora na qualidade das estruturas 3D nas etapas de minimização e aquecimento, provavelmente pelo fato das estruturas ainda estarem se acomodando ao ambiente de simulação. No caso da Sm1, o resíduo K406, localizado em uma região de volta e próximo a um resíduo de prolina (P404), é o único que permanece em região desfavorável nas três conformações. Na CD39, assim como na Sm1, apenas um único resíduo, T430 e localizado em uma região de volta, permanece em região desfavorável após as etapas de minimização e aquecimento da dinâmica molecular.

Tabela 4.1: Validação das estruturas 3D da Sm1 e CD39.

Estrutura	Z -score	Errat	Out1	F1	Out2	F2
Sm1-conf1	-5,66	48,577	3	99,4%	12	97,8%
Sm1-conf2	-5,42	80,594	11	97,7%	14	97,4%
Sm1-conf3	-6,09	82,218	7	98,6%	17	96,9%
CD39-conf1	-9,41	71,657	6	98,7%	5	99,0%
CD39-conf2	-9,32	93,737	7	98,4%	9	98,2%
CD39-conf3	-9,08	92,340	8	98,4%	10	98,0%

Out1 são os resíduos em região desfavorável (*outliers*) e F1 são os resíduos em região favorável, obtidos com o programa Procheck. Out2 são os resíduos em região desfavorável (*outliers*) e F2 são os resíduos em região favorável obtidos com o programa Molprobit.

Os resultados da validação sugerem que houve uma melhora na qualidade dos modelos após as etapas de minimização e aquecimento da dinâmica molecular. O aumento do número de resíduos em regiões desfavoráveis do diagrama de Ramachandran, pode estar associado com o fato das estruturas de ambas as proteínas estarem acomodando-se. Portanto, decidimos pela continuidade das simulações de dinâmica molecular envolvendo os modelos da Sm1 e CD39.

4.2.2 Resíduos catalíticos

Baseado no modelo catalítico proposto por Zebisch e Sträter (2008), foi possível identificar nas estruturas 3D da Sm1 e CD39 os prováveis resíduos do sítio ativo que atuariam na catálise. Isso foi possível através dos alinhamentos entre as estruturas 3D da Sm1 e CD39 com os moldes PDB3ZX3 e PDB3CJA.

Na Sm1, os resíduos seriam D78 (na ACR1), T154 e A155 (ambos na ACR2), E201 (na

ACR3), D232, S237 e Q239 (na ACR4) e W483 (na ACR5) (Figura 4.8). Enquanto que na CD39 os resíduos seriam D54 (na ACR1), T131 e A132 (na ACR2), E174 (na ACR3), D213, S218 e Q220 (na ACR4) e W450 (na ACR5) (Figura 4.8). Além deles há dois resíduos que fazem interações do tipo *stacking* com a base nitrogenada do nucleotídeo, são eles: Y397 e F441, na Sm1, e F365 e Y408, na CD39. É possível observar que os resíduos da Sm1 e CD39 que participam da catálise estão todos distribuídos entre as cinco regiões conservadas (ACR).

Também a partir do alinhamento estrutural, foi possível identificar na estruturas 3D da Sm1 e CD39 os resíduos catalíticos descritos no modelo catalítico proposto por Kozakiewicz e colaboradores (2008). Os resíduos da Sm1 que estariam envolvidos no ataque nucleofílico do γ -fosfato seriam S81 e T154, e no β -fosfato seria T82 (Figura 4.9(a)). Já os resíduos da CD39 que fariam o ataque nucleofílico do γ -fosfato seriam S57 e T131, e no β -fosfato seria S58 (Figura 4.9(b)).

4.2.3 Domínios transmembranares

As hélices transmembranares podem ser divididas em duas regiões: (i) o segmento hidrofóbico que interage com a porção hidrofóbica da bicamada lipídica, (ii) e o segmento formado por resíduos que estão próximos da interface água-lipídio da membrana (NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). O segmento hidrofóbico das hélices transmembranares é rico em resíduos hidrofóbicos como valina, leucina, alanina, isoleucina e fenilalanina (ULMSCHNEIDER; SANSOM, 2001; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Resíduos de glicina e prolina também aparecem preferencialmente no segmento hidrofóbico das hélices transmembranares (ULMSCHNEIDER; SANSOM, 2001; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006).

As regiões de interface água-lipídio são mais ricas em resíduos como a histidina e o triptofano (ULMSCHNEIDER; SANSOM, 2001; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Resíduos carregados como a lisina, arginina, aspartato e glutamato também são encontrados frequentemente nas regiões de interface (ULMSCHNEIDER; SANSOM, 2001; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Resíduos carregados positivamente, são mais encontrados na face *cis* das membranas biológicas, enquanto os carregados negativamente são mais encontrados na face *trans* (ULMSCHNEIDER; SANSOM, 2001; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Os resíduos carregados positivamente

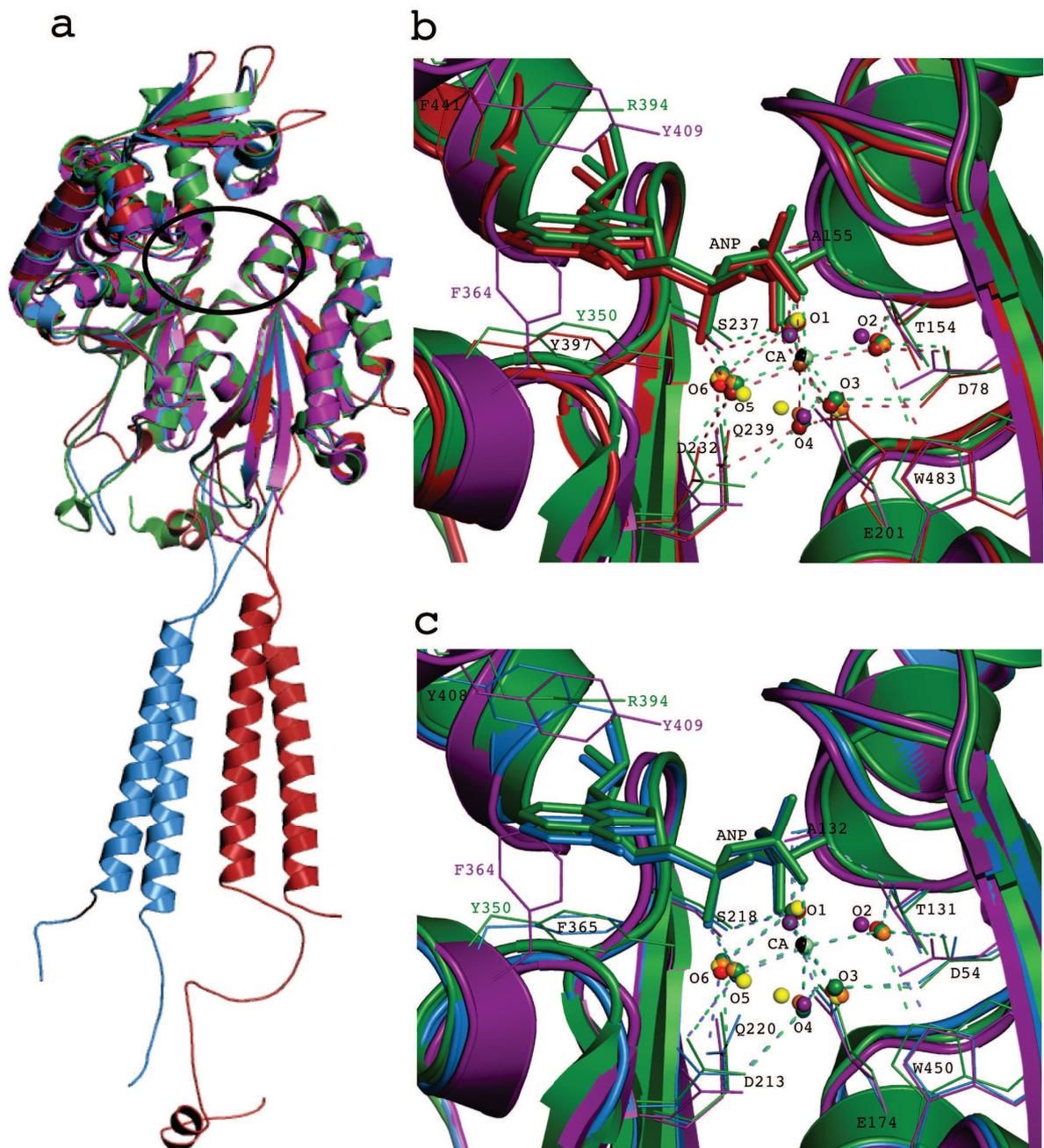


Figura 4.8: Estrutura 3D e sítio ativo da Sm1 e CD39.

(a) Alinhamento estrutural do modelo da Sm1 (vermelho) e CD39 (azul) com os moldes PDB3ZX3 (roxo) e PDB3CJA (verde). (b) Alinhamento estrutural da Sm1 com os moldes na região do sítio catalítico. (c) Alinhamento estrutural da CD39 com os moldes na região do sítio catalítico. Em destaque o ligante AMPPNP (*stick*), resíduos catalíticos (preto), e águas conservadas (O1, O2, O3, O4, O5 e O6) entre as estruturas PDB3CJ1 (vermelho claro), PDB3CJA7 (amarelo), PDB3CJ9 (laranja), PDB3CJA (verde) e PDB3ZX3 (roxo), e o íon cálcio presente nos cristais PDB3CJ9 (preto), PDB3CJA (verde claro), SmATPDase1 (marrom) e CD39 (atrás do cálcio presente no cristal PDB3CJA). As linhas tracejadas indicam os contatos entre os resíduos catalíticos, as moléculas de água conservadas e o íon cálcio. Também é possível observar os resíduos do *stacking* na Sm1 (Y397 e F441) e CD39 (F365 e Y408).

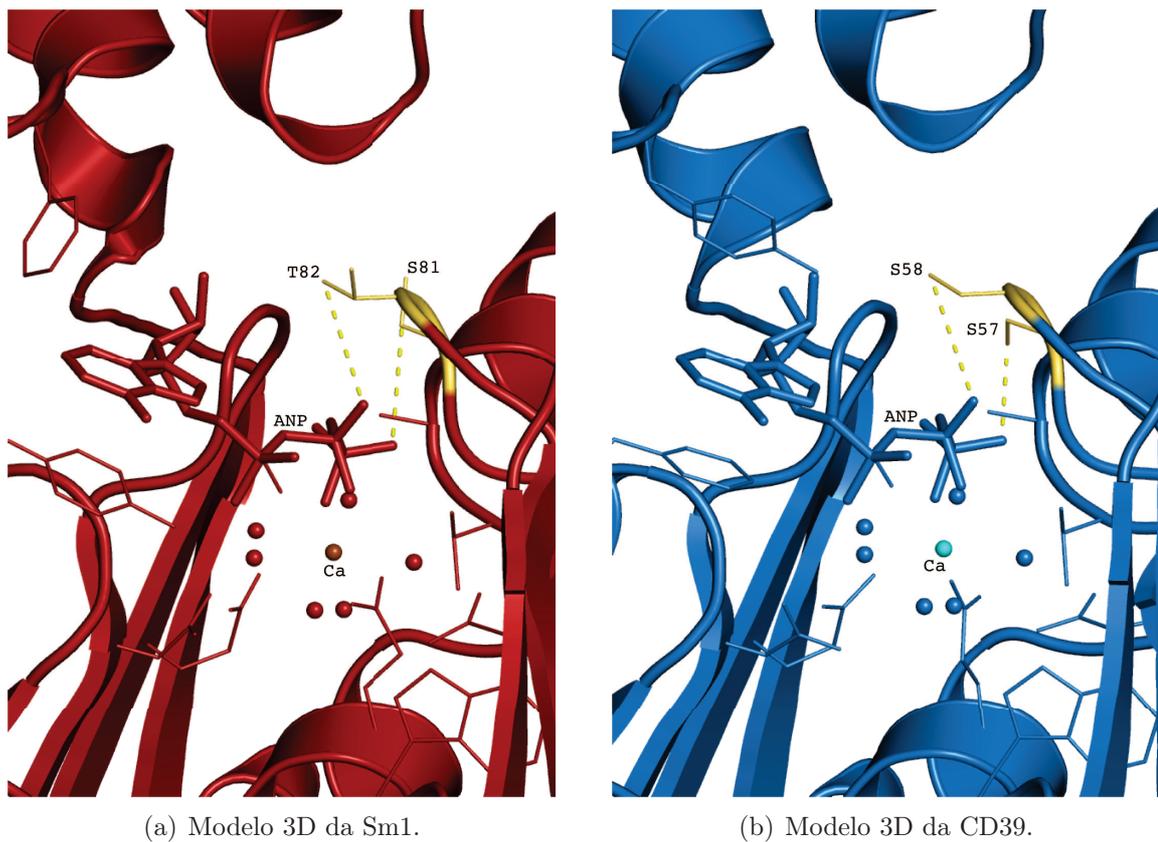


Figura 4.9: Resíduos catalíticos pelo modelo de (KOZAKIEWICZ et al., 2008). Em *stick* amarelo estão destacados os resíduos catalíticos em interação com o ligante ANP.

e com cadeia lateral mais longa, como a lisina e arginina, podem projetar a porção alifática para a região mais hidrofóbica da membrana, enquanto que o agrupamento amino é projetado para a região de interface água-lipídio da membrana (STRANDBERG; KILLIAN, 2003; NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Resíduos como a serina e a treonina possuem um importante papel na interação hélice-hélice, e podem ser encontrados nas duas regiões das hélices transmembranares (NYHOLM; ÖZDIREKCAN; KILLIAN, 2006).

Os domínios transmembranares (TM) de ambos os modelos da Sm1 e CD39 são formados por duas hélices. Foi sugerido na literatura que a mobilidade das TM regula a atividade catalítica no sítio-ativo da CD39 (GRINTHAL; GUIDOTTI, 2006). A remoção de uma ou ambas as TM reduziria em até 90% a atividade enzimática, indicando que a atividade enzimática depende de ambas TM (GRINTHAL; GUIDOTTI, 2004), e que ambas as hélices apresentam um alto grau de movimento rotacional (GRINTHAL; GUIDOTTI, 2004; GRINTHAL; GUIDOTTI, 2006).

Da'dara e colaboradores (2014) demonstraram que a remoção de ambas TM

compromete as atividades ATPásica e ADPásica da Sm1. Eles concluíram que os domínios transmembranares ajudam a proteína a se manter em uma conformação favorável a atividade enzimática (DA'DARA et al., 2014).

Veremos na sessão *Dinâmica molecular*, que nas quatro simulações de dinâmica molecular envolvendo os modelos, não foram observadas mudanças significativas nas TM que tivesse relação com mudanças no domínio ECD.

4.2.4 Perfil eletrostático dos modelos

Neste trabalho, também foram avaliados os perfis eletrostáticos dos modelos da Sm1 e CD39 usando os programas pdb2pqr v1.9 (DOLINSKY et al., 2004) e APBS v1.4 (BAKER et al., 2001). A figura 4.10 mostra o perfil eletrostático dos modelos da Sm1 e CD39. O perfil eletrostático da Sm1 (figuras 4.10(a) e (b)) apresenta-se mais carregado positivamente, inclusive na região dos domínios transmembranares. O perfil mais positivo nos domínios transmembranares da Sm1, pode ser explicado pela presença de resíduos carregados positivamente, como H38 e K39 na TM1, e R531 na TM2. Esses três resíduos carregados positivamente estão localizados próximo a face intracelular (*cis*) da TM1 e TM2, o que vai de encontro com a literatura (NYHOLM; ÖZDIREKCAN; KILLIAN, 2006). Diferentemente da Sm1, a CD39 apresenta um perfil eletrostático mais negativo com domínios transmembranares mais neutros (figuras 4.10(c) e (d)).

A figura 4.11 mostra o perfil eletrostático na região do sítio-ativo dos modelos da Sm1 e CD39. O perfil eletrostático do sítio ativo da Sm1 é mais carregado positivamente, porém na região onde localiza-se o resíduo D438 o perfil está mais negativo. Já o perfil eletrostático do sítio ativo da CD39 é mais carregado negativamente, com uma região mais positiva devido a presença de dois resíduos de lisina (K407 e K405). Essa diferença no perfil eletrostático pode interferir na interação dos nucleotídeos di- e trifosfatados na região do sítio ativo, em função das cargas negativas presentes nos grupos fosfatos.

4.2.5 Ligações dissulfeto

Assim como nos seus moldes, os modelos da Sm1 e CD39 também apresentam ligações dissulfeto no domínio ECD. A diferença é que a Sm1 possui três ligações, enquanto que CD39 apresenta cinco ligações, como nos moldes PDB3ZX3 e PDB3CJA. As figuras 4.12(a) e 4.12(b) mostram os domínios ECD da Sm1 e CD39 e em destaque as ligações

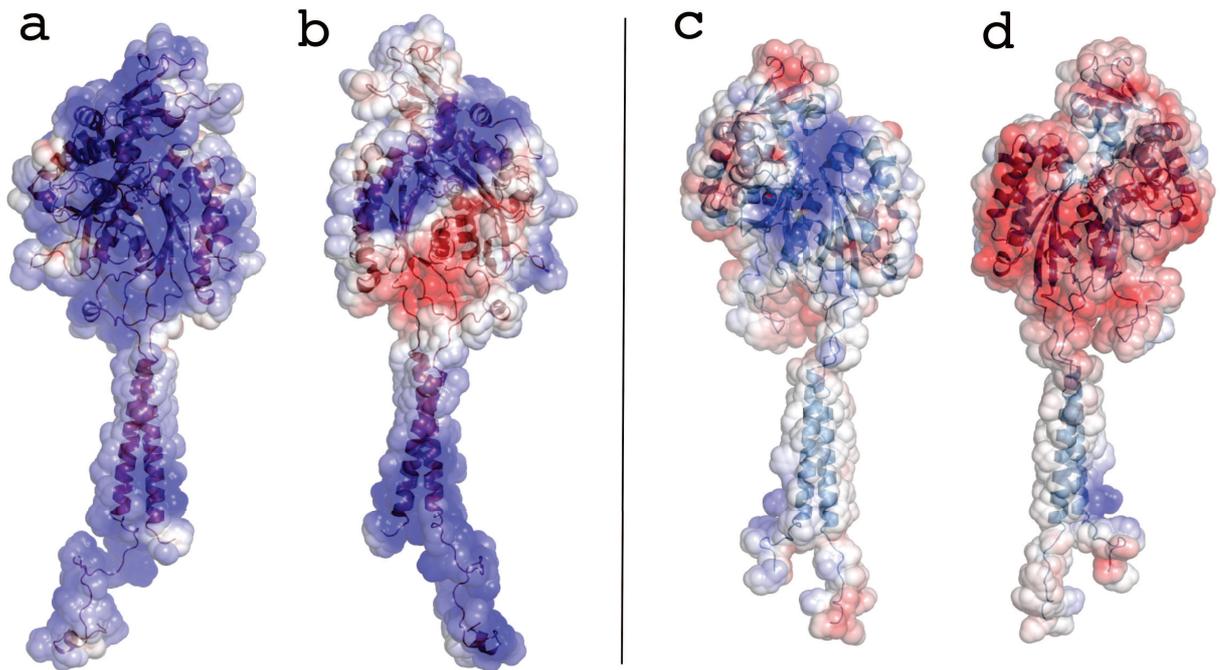


Figura 4.10: Potencial de superfície eletrostático da Sm1 e CD39.

(a) Visão frontal do perfil eletrostático do modelo da Sm1. (b) Visão de trás do perfil eletrostático do modelo da Sm1. (c) Visão frontal do perfil eletrostático do modelo da CD39. (d) Visão atrás do perfil eletrostático do modelo da CD39. Potencial eletrostático variando de -5 (vermelho) a 5 (azul) kT/e.

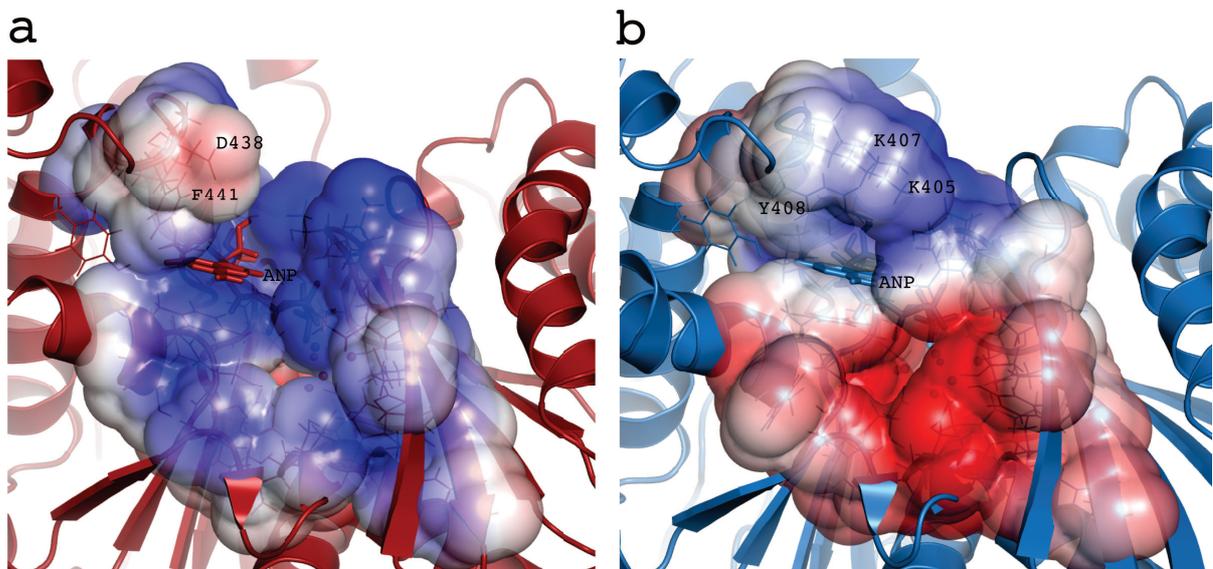


Figura 4.11: Potencial de superfície eletrostático da Sm1 e CD39 na região do sítio-ativo. (a) Perfil eletrostático do sítio-ativo no modelo da Sm1. (b) Perfil eletrostático so sítio-ativo no modelo da CD39. As estruturas 3D são mostradas em *Cartoon*, o ligante (ANP) em *Licorice*, o Ca^{2+} e molécula de água em esfera. Potencial eletrostático variando de -3 (vermelho) a 3 (azul) kT/e.

dissulfeto. Veremos na seção seguinte que as duas ligações dissulfeto a menos na Sm1 (que seriam equivalentes aos pares C255-C301 e C84-C108 na CD39) parecem contribuir para a diferença na dinâmica dessa proteína em relação aos moldes e a CD39.

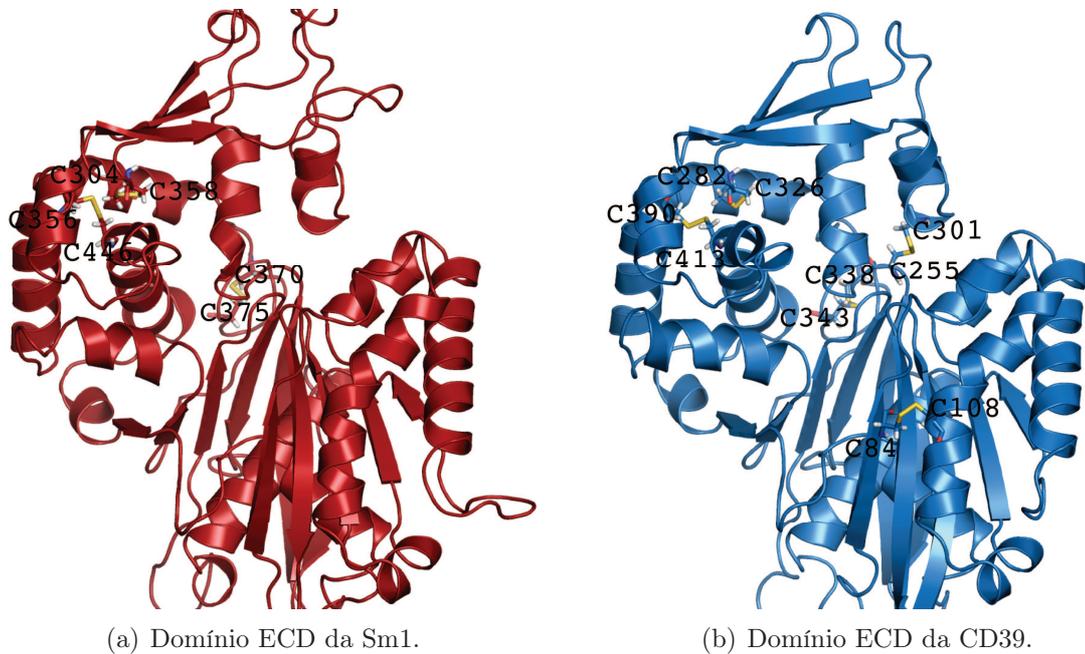


Figura 4.12: Em destaque as pontes dissulfeto (amarelo) presentes na Sm1 e CD39.

4.3 Dinâmica Molecular

Nesta seção será discutido os resultados referentes às simulações de dinâmica molecular realizadas no presente trabalho. No total foram simulados 6 sistemas: (i) modelo da Sm1 sem o ligante ANP presente no sítio ativo e na presença do íon cálcio (Sm1); (ii) modelo da Sm1 com o ligante ANP e o íon cálcio presentes no sítio ativo (Sm1-ANP); (iii) modelo da CD39 sem o ligante ANP no sítio ativo e na presença do íon cálcio (CD39); (iv) modelo da CD39 com o ligante ANP e o íon cálcio presentes no sítio ativo (CD39-ANP); (v) molde PDB3CJA com ANP e íon cálcio presentes no sítio ativo; (vi) molde PDB3ZX3 sem o ligante e sem heteroátomos no sítio ativo. Os cristais dos moldes apresentam apenas o domínio ECD. O tempo de simulação para os modelos foi de 32ns e para os moldes foi de 40ns.

4.3.1 Relação área por lipídio

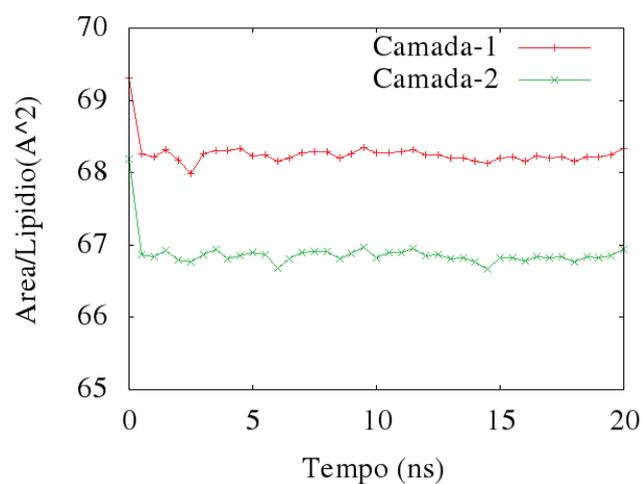
seção *Para* simulações de dinâmica molecular envolvendo modelos de membrana, a relação entre a área da membrana e o número de moléculas de lipídio que a formam é chamada de relação área por lipídio (EDHOLM; NAGLE, 2005). Os valores da relação área por lipídio obtidos durante a simulação de dinâmica molecular, devem ser comparados com valores experimentais a fim de validar os resultados obtidos durante a simulação (BENZ et al., 2005; KUCERKA et al., 2015). A seguir serão apresentados os gráficos referentes a variação da relação área por lipídeo observadas nas quatro simulações realizadas no presente trabalho.

As figuras 4.13(a) e 4.13(b) mostram os gráficos da relação área por lipídeo durante os 20ns de equilíbrio das simulações envolvendo a Sm1. A camada-1 corresponde a camada extracelular da bicamada lipídica, enquanto a camada-2 corresponde a camada intracelular.

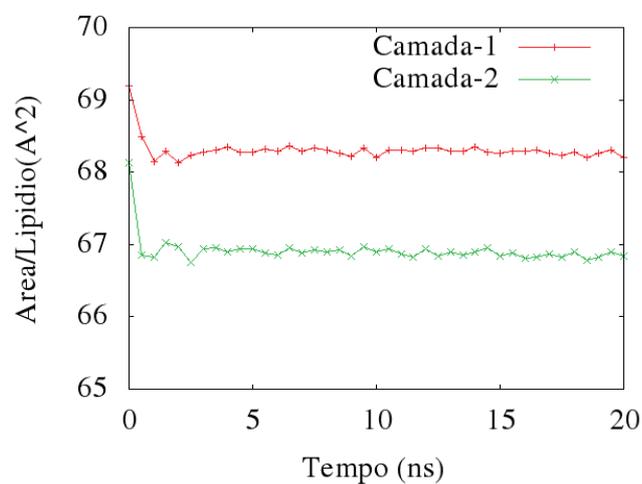
A média da área por lipídio para a simulação envolvendo a Sm1 sem ANP foi de $68,26\text{\AA}^2$, para a camada extracelular, e $66,87\text{\AA}^2$, para a camada intracelular. Para a simulação envolvendo a Sm1 com ANP (Sm1-ANP), a média da área por lipídio foi $68,31\text{\AA}^2$, para a camada extracelular, e $66,91\text{\AA}^2$, para a camada intracelular. Os valores da área por lipídeo observados podem ser corroborados por estimativas experimentais que apontam, para membrana POPC, uma relação de $68,3\pm 1,5 \text{\AA}^2/\text{molécula POPC}$ (KUCERKA; TRISTRAM-NAGLE; NAGLE, 2006).

As figuras 4.14 e 4.14(b) mostram os gráficos da relação área por lipídeo durante os 20ns de equilíbrio das simulações envolvendo a CD39. Na simulação envolvendo a CD39 sem ANP, a média da área por lipídio foi de $67,87\text{\AA}^2$, para a camada extracelular, e $66,48\text{\AA}^2$, para a camada intracelular. Para a simulação envolvendo a CD39 com ANP (CD39-ANP), a média da área por lipídio foi $67,83\text{\AA}^2$, para a camada extracelular, e $66,45\text{\AA}^2$, para a camada intracelular. Os valores obtidos também correspondem aos apresentados por estimativas experimentais (KUCERKA; TRISTRAM-NAGLE; NAGLE, 2006).

A menor relação área por lipídeo na camada intracelular em relação a camada extracelular deve-se ao fato de que durante o acoplamento das proteínas na membrana houve uma maior remoção de moléculas de lipídeo na camada extracelular. Nas simulações envolvendo a Sm1 a camada extracelular ficou com 197 moléculas e a camada intracelular com 193 moléculas. Nas simulações envolvendo a CD39 a camada extracelular ficou com



(a) Simulação da Sm1.



(b) Simulação da Sm1-ANP.

Figura 4.13: Área por lipídio durante os 20ns de equilíbrio da Sm1 com e sem ANP.
Camada-1: extracelular. Camada-2: intracelular

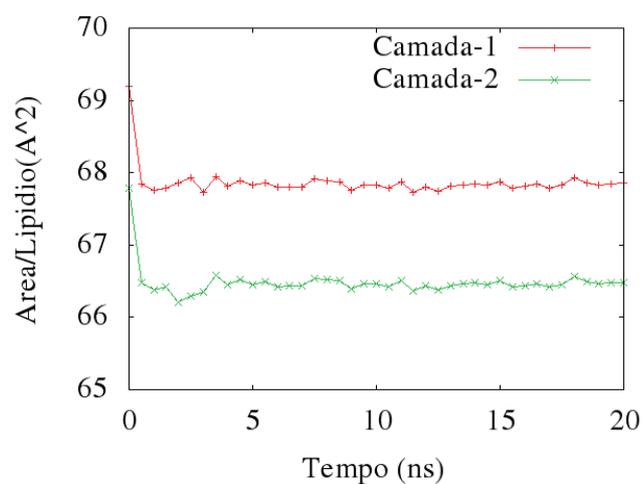
198 moléculas e a camada intracelular com 194 moléculas.

4.3.2 Cálculo das Energias

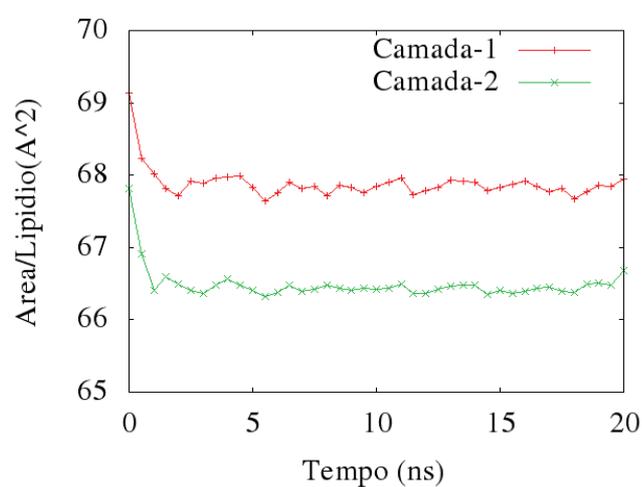
A seguir serão apresentados os gráficos de energias das proteínas e membranas, referentes às simulações de dinâmica molecular envolvendo os moldes PDB3CJA e PDB3ZX3, e os modelos da Sm1 e CD39.

4.3.2.1 Energias dos moldes

As figuras 4.15(a) e 4.15(b) mostram os gráficos de energias durante os 40ns de simulação de dinâmica molecular envolvendo os moldes PDB3CJA (RnNTPDase2) e PDB3ZX3



(a) Simulação da CD39.

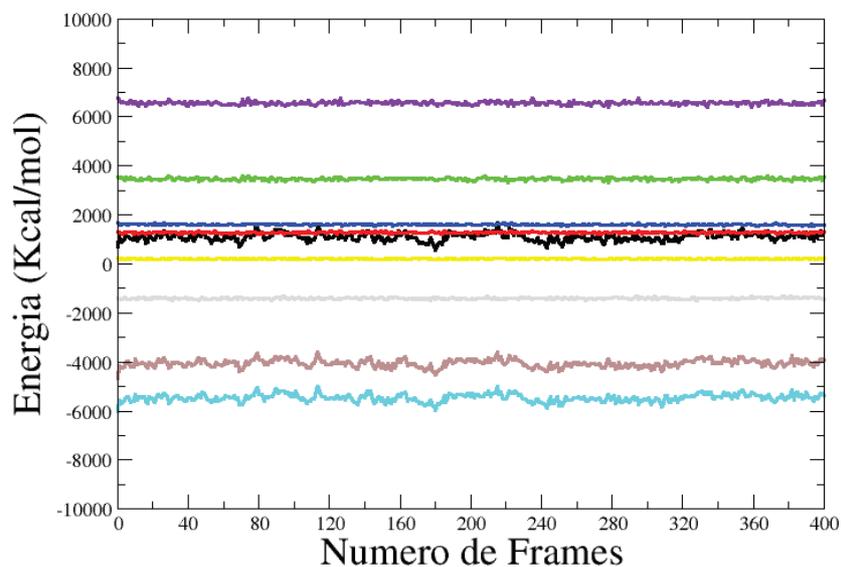


(b) Simulação da CD39-ANP.

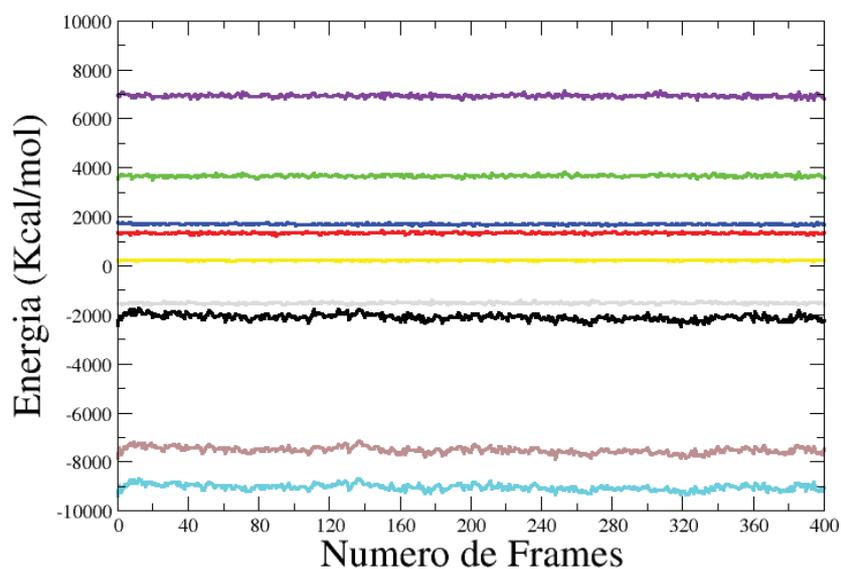
Figura 4.14: Área por lipídio durante os 20ns de equilíbrio da CD39 com e sem ANP. Camada-1: extracelular. Camada-2: intracelular

(RnNTPDase1). São apresentados os gráficos de energia total e das demais energias calculadas pelo campo de força. É possível observar que os gráficos referentes às energias de átomos não ligados, eletrostática e Van der Waals apresentam valores negativos, enquanto que as energias de átomos ligados, angular, diedral e impróprio apresentam valores positivos.

É possível observar que a energia total da PDB3ZX3 apresenta oscilação até aproximadamente 24ns e seu valor é positivo durante toda a simulação. Por outro lado, a energia total da PDB3CJA apresenta uma menor oscilação e apresenta valor negativo. Isso pode ser em função da presença do ligante ANP, já que a literatura indica que as NTPDases apresentam uma conformação mais relaxada na presença de nucleotídeos



(a) Energia total da proteína PDB3ZX3 (RnNTPDase1).



(b) Energia total da proteína PDB3CJA (RnNTPDase2).

Figura 4.15: Energia total dos moldes durante os 40ns de simulação. Energias: Total (preto), Ligação (vermelho), Angular (verde), Dedral (azul), Impróprio (amarelo), Eletrostática (marrom), Van der Waals (cinza), Conformacional (violeta), Não-ligados (azul claro). O intervalo de tempo entre os *frames* é de 100ps.

trifosfatados no sítio ativo (ZEBISCH et al., 2013).

4.3.2.2 Energias dos modelos

As figuras 4.16(a) e 4.16(b) mostram os gráficos de energias durante os 32ns de simulação de dinâmica molecular envolvendo o modelo da Sm1. Os gráficos mostram que a energia total da Sm1 na ausência ou presença do ANP está oscilando. Isso indica que em ambos os sistemas a proteína provavelmente ainda não equilibrou.

As figuras 4.17(a) e 4.17(b) mostram os gráficos de energias durante os 32ns de simulação de dinâmica molecular envolvendo o modelo da CD39. Assim como na Sm1, os gráficos referentes a CD39 na ausência ou presença do ANP também mostram uma oscilação na energia total. O que indica que em ambos os sistemas a proteína ainda não equilibrou.

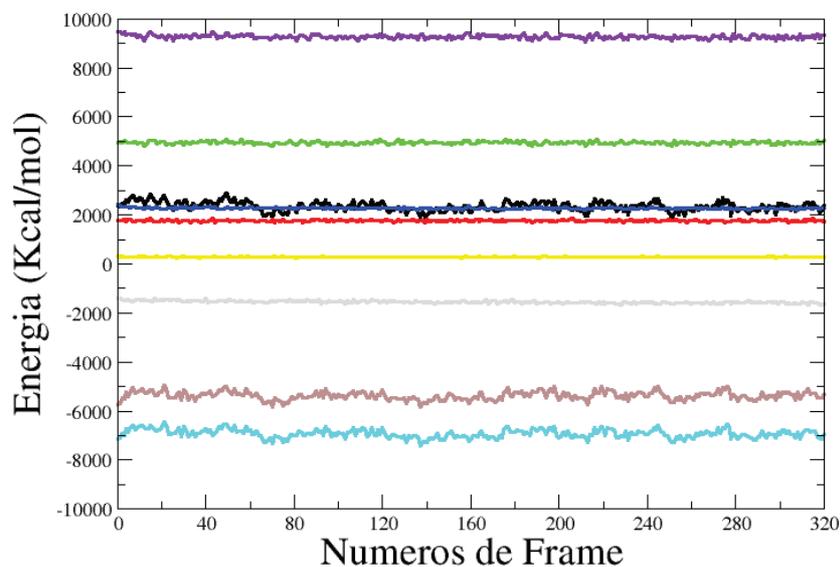
Nos dois modelos, podemos observar que a energia total apresentam valores positivos seguindo o mesmo comportamento apresentado pelo molde PDB3ZX3. Para as simulações da Sm1-ANP e CD39-ANP a presença do ligante no sítio ativo não levou a uma energia total negativa como no molde PDB3CJA, provavelmente por esses modelos terem sido construídos considerando também o molde PDB3ZX3.

4.3.2.3 Energias membranas

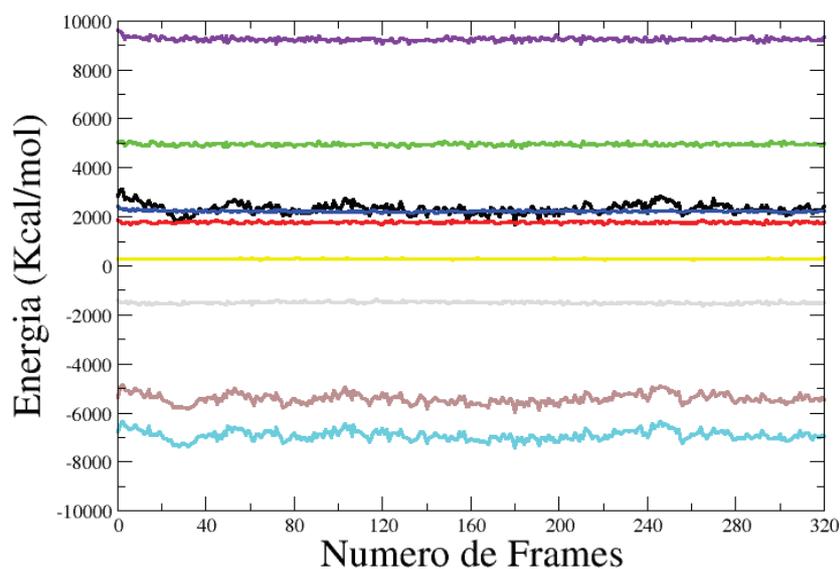
As figuras 4.18(a) e 4.18(b) mostram os gráficos da energia total da membrana POPC durante os 32ns de simulação de dinâmica molecular envolvendo o modelo da Sm1. Aqui também são mostrados os gráficos das demais energias calculadas pelo campo de força. Observa-se que os gráficos referentes às energias de átomos não ligados, eletrostática e Van der Waals apresentam valores negativos, enquanto que as energias de átomos ligados, angular, diedral e impróprio apresentam valores positivos.

Pelos gráficos é possível observar uma estabilização da energia total da membrana POPC nos sistemas envolvendo a Sm1 na ausência ou presença do ANP. Porém, observa-se que a energia total de ambos os sistemas apresenta valores positivos. Embora a relação área por lipídeo tenha estabilizado, como mostrado na seção *Relação área por lipídeo*, é provável que a membrana, em ambos os sistemas Sm1 e Sm1-ANP, ainda não tenha equilibrado.

Os gráficos 4.19(a) e 4.19(b) mostram a variação da energia total da membrana POPC durante os 32ns de simulação de dinâmica molecular envolvendo o modelo da CD39. Os gráficos mostram uma estabilização da energia total da membrana POPC nos sistemas



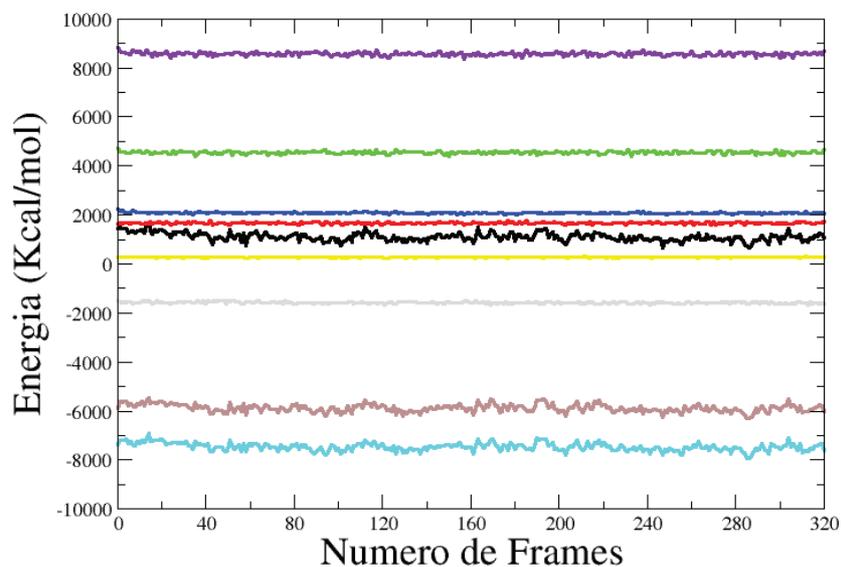
(a) Simulação da Sm1.



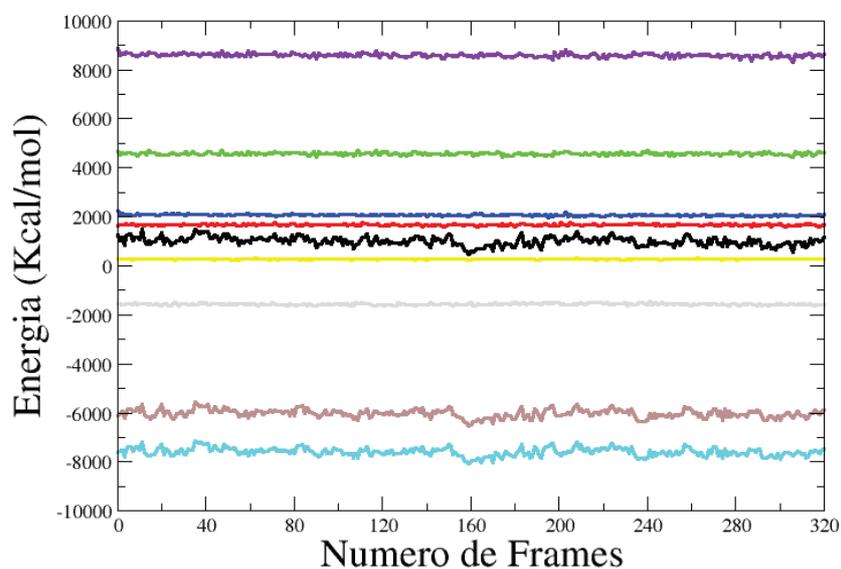
(b) Simulação da Sm1-ANP.

Figura 4.16: Energia total da Sm1 e Sm1-ANP durante 32ns de simulação. Energias: Total (preto), Ligação (vermelho), Angular (verde), Diedral (azul), Impróprio (amarelo), Eletrostática (marrom), Van der Waals (cinza), Conformacional (violeta), Não-ligados (azul claro). O intervalo de tempo entre os *frames* é de 100ps.

envolvendo a CD39 na ausência ou presença do ANP. Mas também é possível observar que a energia total também apresenta valores positivos. Embora a área por lipídeo em



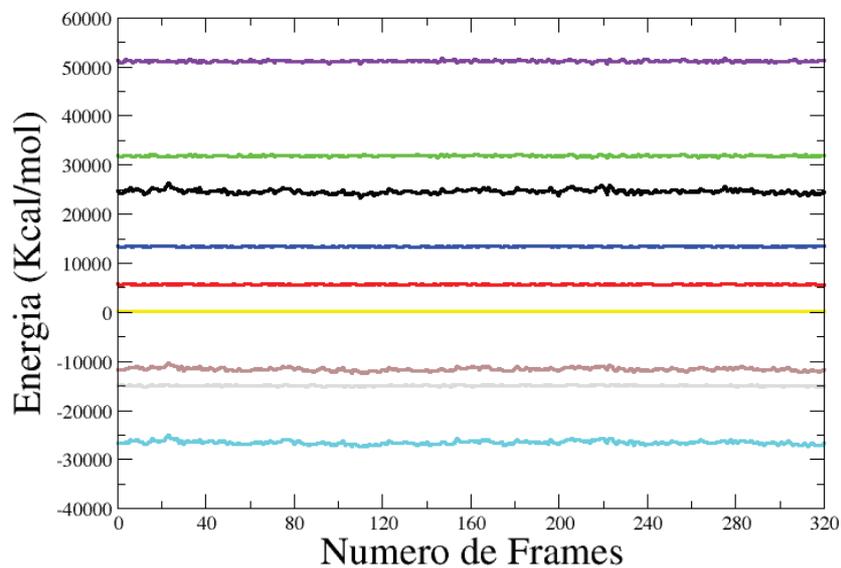
(a) Simulação da CD39.



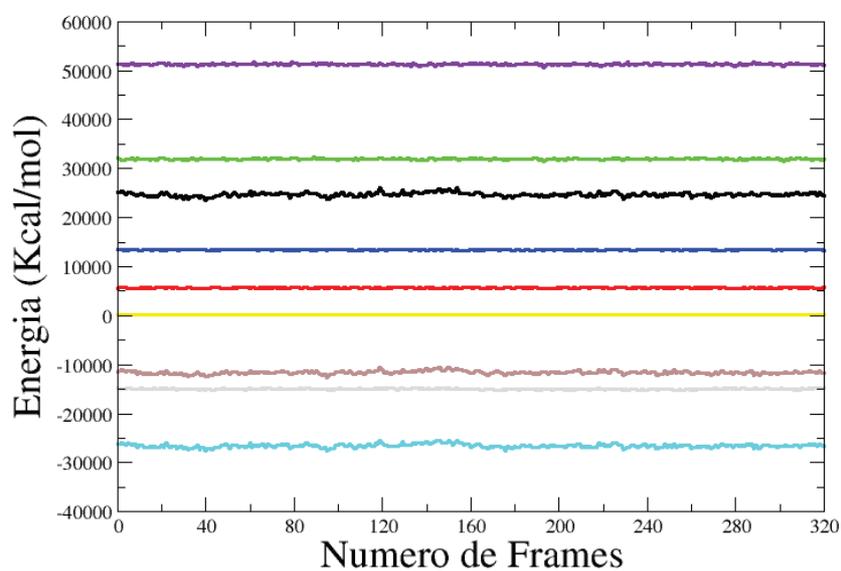
(b) Simulação da CD39-ANP.

Figura 4.17: Energia total da CD39 e CD39-ANP durante 32ns de simulação. Energias: Total (preto), Ligação (vermelho), Angular (verde), Diedral (azul), Impróprio (amarelo), Eletrostática (marrom), Van der Waals (cinza), Conformacional (violeta), Não-ligados (azul claro). O intervalo de tempo entre os *frames* é de 100ps.

ambos os sistemas CD39 e CD39-ANP tenha estabilizado, como mostrado na seção seção *Relação área por lipídio*, é provável que a membrana ainda não tenha equilibrado.

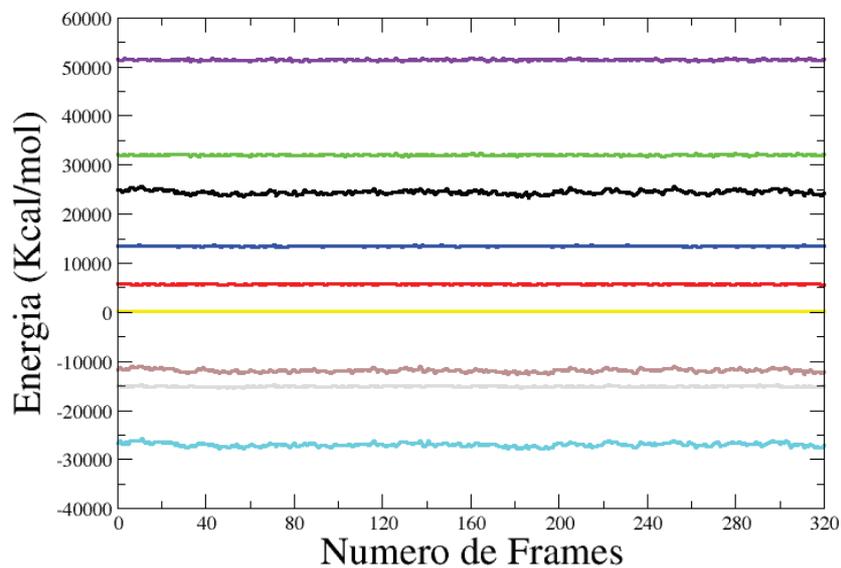


(a) Simulação da Sm1.

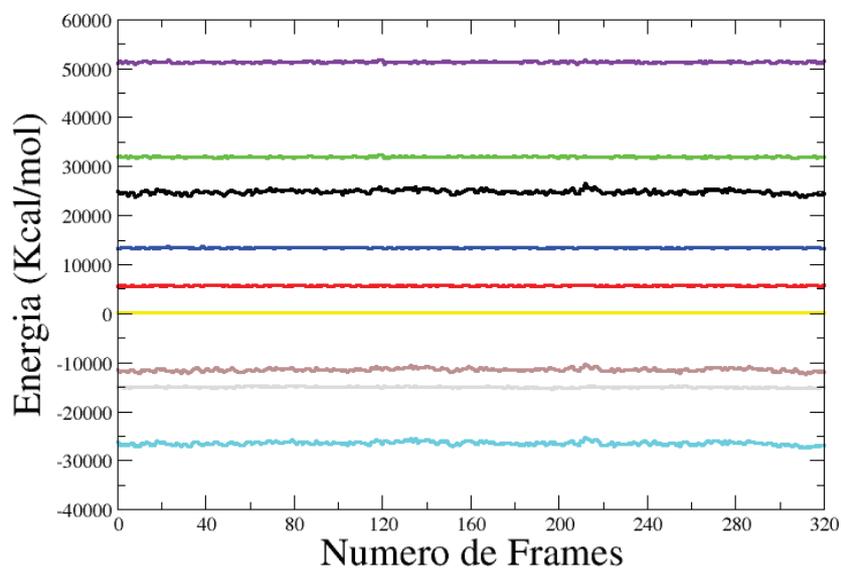


(b) Simulação da Sm1-ANP.

Figura 4.18: Energia total da membrana POPC durante 32ns de simulação. Energias: Total (preto), Ligação (vermelho), Angular (verde), Diedral (azul), Impróprio (amarelo), Eletrostática (marrom), Van der Waals (cinza), Conformacional (violeta), Não-ligados (azul claro). O intervalo de tempo entre os *frames* é de 100ps.



(a) Simulação da CD39.



(b) Simulação da CD39-ANP.

Figura 4.19: Energia total da membrana POPC durante 32ns de simulação. Energias: Total (preto), Ligação (vermelho), Angular (verde), Diedral (azul), Impróprio (amarelo), Eletrostática (marrom), Van der Waals (cinza), Conformacional (violeta), Não-ligados (azul claro). O intervalo de tempo entre os *frames* é de 100ps.

4.3.3 RMSD das trajetórias

A tabela 4.2 apresenta os valores do RMSD para os moldes e modelos na simulações de dinâmica molecular. Para os modelos da Sm1 e CD39, os valores do RMSD são apresentados por subdomínios ECD1 e ECD2 e domínios transmembranares TM1 e TM2. Para os moldes foi considerado somente os subdomínios do ECD devido à ausência dos domínios transmembranares nos cristais.

A definição dos subdomínios ECD1 e ECD2 dos modelos foi feita com base no alinhamento estrutural destes com os moldes e na literatura (ZEBISCH; STRÄTER, 2008). O subdomínio ECD1 da Sm1 é formado pelos resíduos que estão nos intervalos Y73-G198 e G478-G493, enquanto que o subdomínio ECD2 é formado pelos resíduos S199-K477. Na CD39 o subdomínio ECD1 é formado pelos resíduos dos intervalos N46-G171 e G445-N460, enquanto que o subdomínio ECD2 é formado pelos resíduos Q172-Q444.

Observando a tabela 4.2 é possível ver que os valores mínimo, máximo, média e desvio padrão do RMSD em cada uma das simulações estão próximos.

Quando se compara os valores do RMSD referentes ao subdomínio ECD1 dos modelos com os dos moldes é possível ver que eles são similares, indicando que as mudanças conformacionais do ECD1 entre os moldes e modelos é similar. Por outro lado, quando observamos os valores do RMSD no ECD2 podemos ver que eles são maiores nas simulações onde o ligante ANP está presente no sítio ativo. Isso pode estar relacionado com a interação (*stacking*) da adenina com resíduos do ECD2, como Y397 e F441 (Sm1), e F365 e Y408 (CD39). A exceção é o ECD2 da Sm1 sem o ANP no sítio ativo (cavidade-1), que é o que apresenta os maiores valores de RMSD.

Com relação aos valores do RMSD dos domínios transmembranares, podemos observar que para a TM1 os valores são ligeiramente maiores para as simulações da Sm1 e Sm1-ANP. O mesmo podemos observar para a TM2. Esses resultados sugerem um pouco mais de movimento nas TM da Sm1 e Sm1-ANP, do que na CD39 e CD39-ANP. Porém, quando se compara as simulações com ligante e sem ligante, para um mesmo modelo, não foram observadas diferenças. Isso sugere que o ligante não fez diferença na movimentação dos domínios transmembranares.

A seguir serão apresentados os gráficos do RMSD para os domínios ECD e transmembranares em função do tempo para todas as simulações de dinâmica molecular envolvendo os moldes e modelos.

Tabela 4.2: Valores do RMSD para os moldes e modelos nas simulações.

Domínio	Estrutura	Min	Max	Média	DP
ECD-I	3ZX3	1,094	3,742	2,720	0,346
	3CJA	1,001	3,045	1,238	0,350
	Sm1	1,577	3,115	2,530	0,214
	Sm1-ANP	1,313	2,664	2,098	0,191
	CD39	0,898	1,978	1,587	0,138
	CD39-ANP	1,021	2,038	1,695	0,172
ECD-II	3ZX3	1,105	3,797	2,202	0,245
	3CJA	0,996	2,579	1,745	0,251
	Sm1	1,562	6,878	4,782	1,153
	Sm1-ANP	1,535	4,600	3,675	0,459
	CD39	1,249	2,330	1,898	0,182
	CD39-ANP	1,284	4,295	3,309	0,556
TM1	Sm1	0,724	2,367	1,537	0,322
	Sm1-ANP	0,779	2,692	1,896	0,335
	CD39	0,628	1,434	0,989	0,138
	CD39-ANP	0,597	1,885	1,276	0,281
TM2	Sm1	0,808	2,739	1,929	0,297
	Sm1-ANP	1,329	2,554	1,876	0,204
	CD39	0,472	1,162	0,756	0,142
	CD39-ANP	0,573	1,449	0,969	0,160

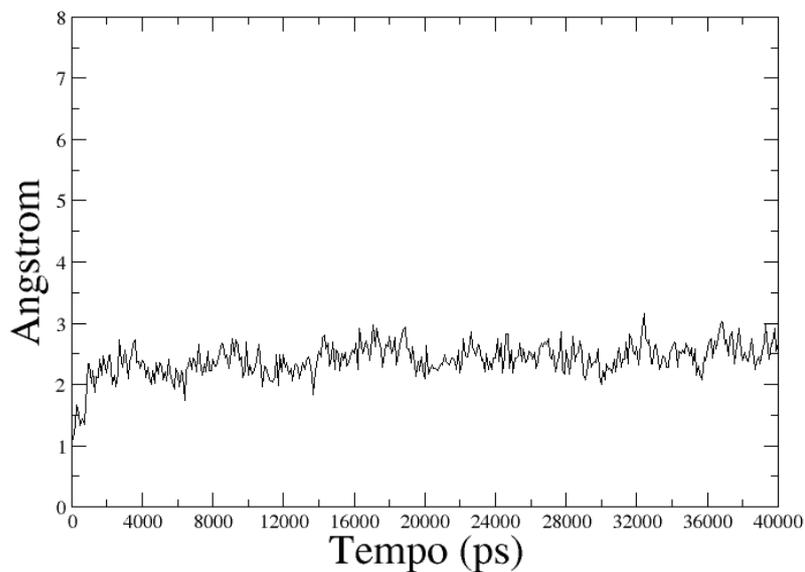
Min é o valor mínimo do RMSD, Max é o valor máximo do RMSD e DP é o desvio padrão.

4.3.3.1 RMSD dos ECD

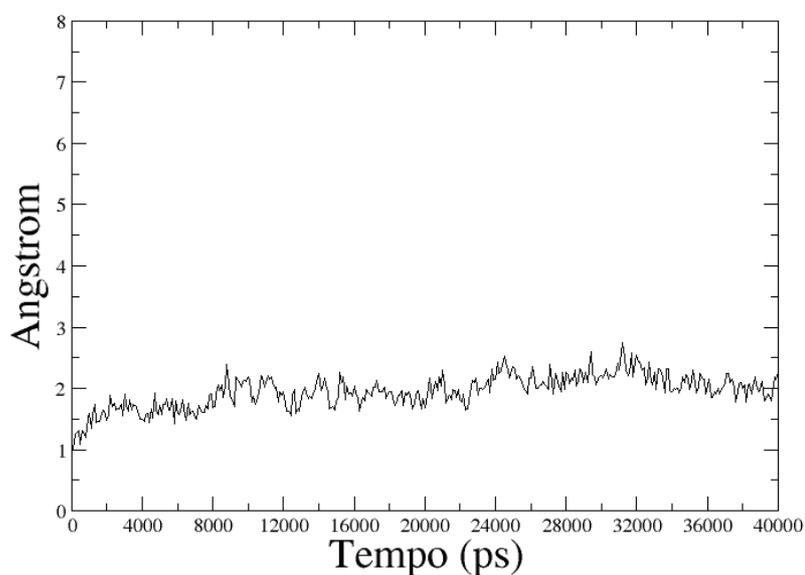
As figuras 4.20(a) e 4.20(b) mostram a variação do RMSD do domínio ECD (total) nas simulações de dinâmica molecular envolvendo os moldes PDB3ZX3 e PDB3CJA. Pelos gráficos é possível observar uma estabilização na variação do RMSD para os domínios ECD dos moldes, o que indicaria uma estabilização nas mudanças conformacionais dos domínios ECD de ambas as proteínas.

Os gráficos apresentados em 4.21 mostram a variação do RMSD do domínio ECD da Sm1 e Sm1-ANP. Na figura 4.21(a) é possível observar um aumento do RMSD ao longo da trajetória, o que sugere mudanças conformacionais no domínio ECD. Essa maior variação do ECD estaria provavelmente associada a tais fatores: (i) a estrutura ainda não equilibrou e busca por um mínimo de energia; (ii) a ausência de ligações dissulfeto na Sm1 (em relação a CD39 e aos moldes), que seriam responsáveis por manter a estabilidade do domínio ECD-II em relação ao sítio ativo da proteína; (iii) ausência do ligante no sítio ativo.

A ausência de duas ligações dissulfeto no ECD da Sm1 pode estar contribuindo para



(a) RMSD da proteína PDB3ZX3 (RnNTPDase1).

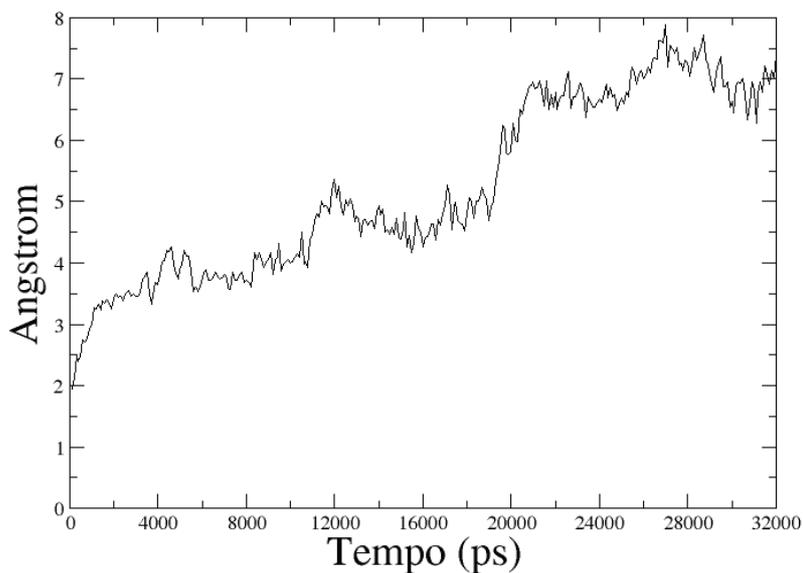


(b) RMSD da proteína PDB3CJA (RnNTPDase2).

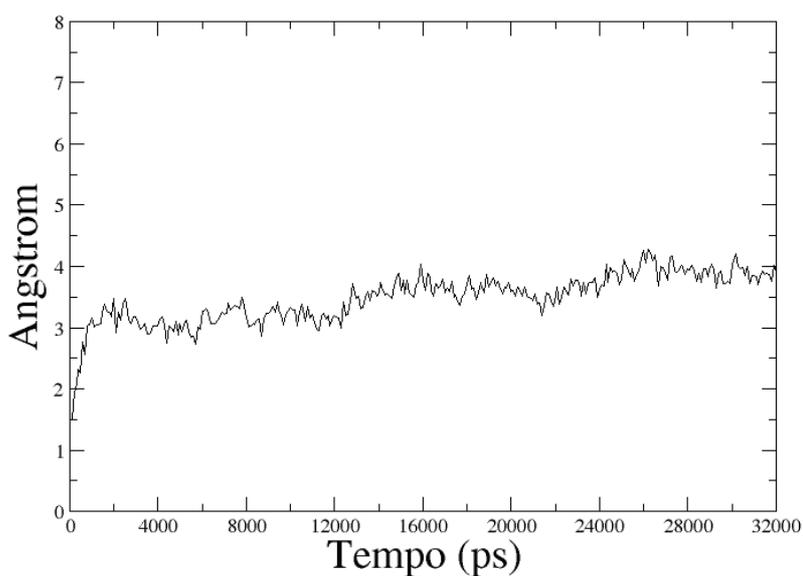
Figura 4.20: Variação do RMSD nas simulações com os moldes. RMSD calculado usando $C\alpha$.

uma maior movimentação desse domínio. É importante destacar que as duas ligações dissulfeto ausentes na Sm1, correspondem àquelas formadas pelos pares C84-C108 e C255-C301 presentes na CD39 (ver figura 4.12(b)). Na região da cavidade-2 da Sm1 não há uma

ligação dissulfeto, correspondente a ligação formada pelo C255-C301 da CD39 (ligação entre a região do sítio ativo e a cavidade-2), o que provavelmente estaria permitindo uma maior movimentação dessa região na Sm1.



(a) RMSD da Sm1.



(b) RMSD da Sm1-ANP.

Figura 4.21: Variação do RMSD referente ao domínio ECD nas simulações envolvendo a Sm1 e Sm1-ANP. RMSD calculado usando o $C\alpha$.

A figura 4.22 mostra o alinhamento do domínio ECD da Sm1 em três conformações diferentes. Essas três conformações são o *frame-0* (conformação-referência) e os *frames* 55 e 171 provenientes da análise de agrupamento da trajetória da Sm1 (seção *Análise de Agrupamento*). Nessa figura é possível observar uma movimentação da região ECD-II onde está localizada a cavidade-2 (região em destaque). Na figura é mostrado em destaque o resíduo K279 cuja movimentação no sítio ativo parece estar associada com a movimentação da região da cavidade-2. Esse ponto será melhor abordado na seção *Conformações da Sm1 e o movimento de K279*.

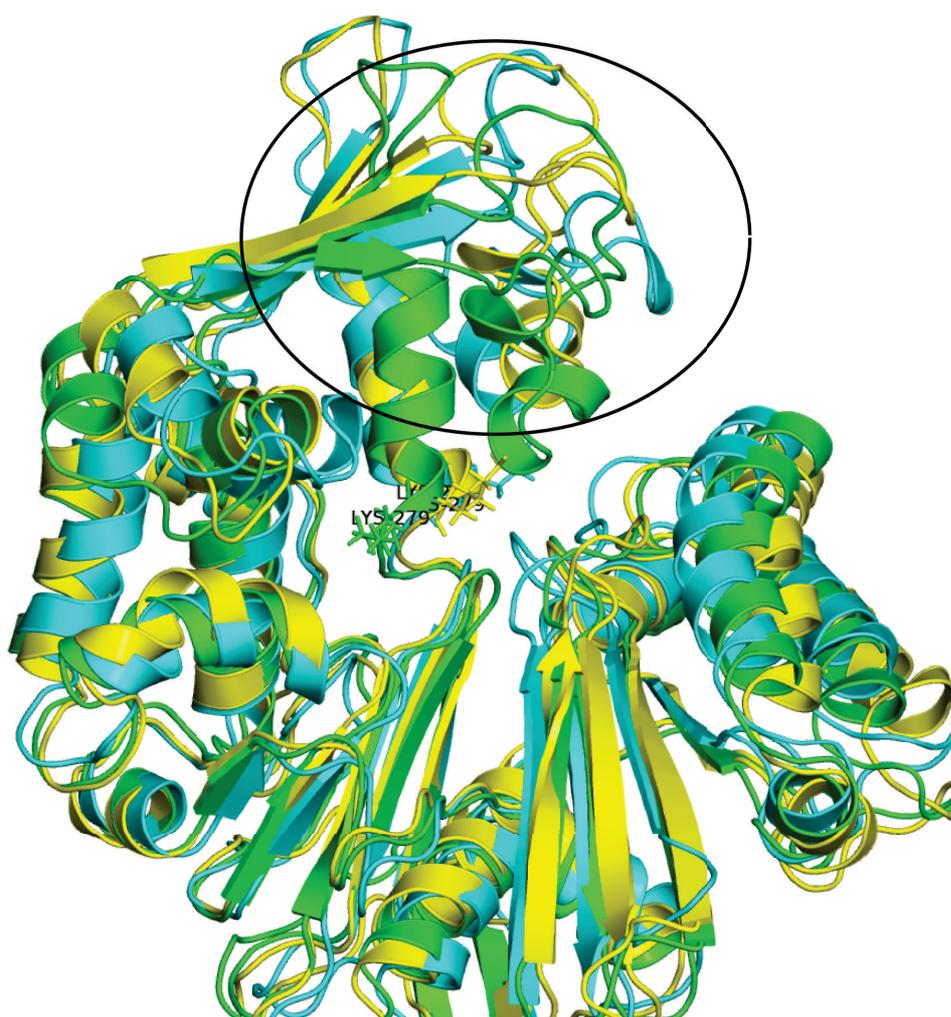


Figura 4.22: Alinhamento estrutural do ECD da Sm1. Em azul o *frame-0* (conformação-referência), em amarelo o *frame-55* e em verde o *frame-171*.

Sobre o ligante, a literatura indica que a presença do ligante no sítio ativo contribui para a estabilidade da proteína (ZEBISCH et al., 2013). No gráfico da simulação da Sm1-ANP (figura 4.21(b)), é possível observar que a presença do ligante parece estabilizar a estrutura de forma a apresentar menor variação do RMSD no domínio ECD, semelhante

à variação do ECD dos moldes (Figura 4.20). Portanto, a ausência do ligante também poderia estar contribuindo para uma maior movimentação do domínio ECD na Sm1 simulada sem a presença do ligante.

A figura 4.23 mostra o domínio ECD da Sm1-ANP em três conformações. Essas três conformações são o *frame-0* (conformação-referência) e os *frames* 149 e 219 provenientes da análise de agrupamento da trajetória da Sm1-ANP (seção *Análise de Agrupamento*). Pela figura é possível ver que a região da cavidade-2 apresenta-se mais estável em comparação com a Sm1 simulada sem a presença do ligante no sítio ativo (figura 4.22).

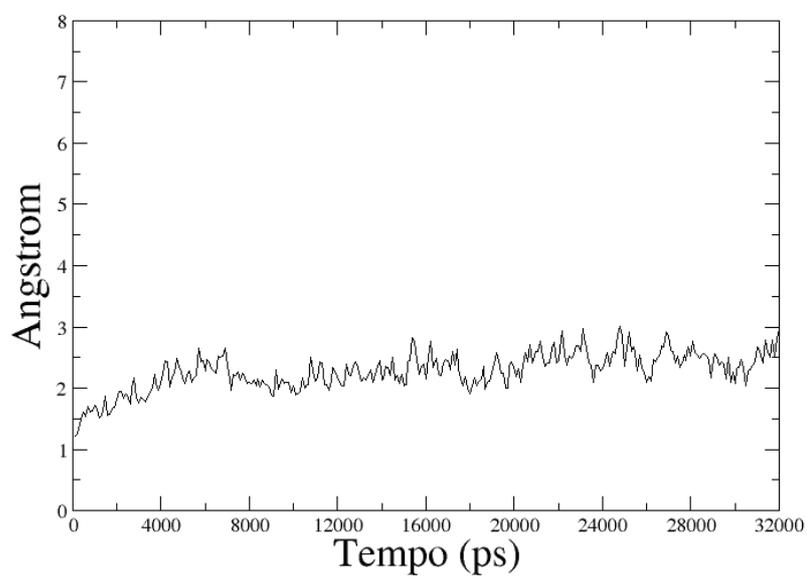


Figura 4.23: Alinhamento estrutural do ECD da Sm1-ANP. Em azul o *frame-0* (conformação-referência), em amarelo o *frame-149* e em verde o *frame-219*.

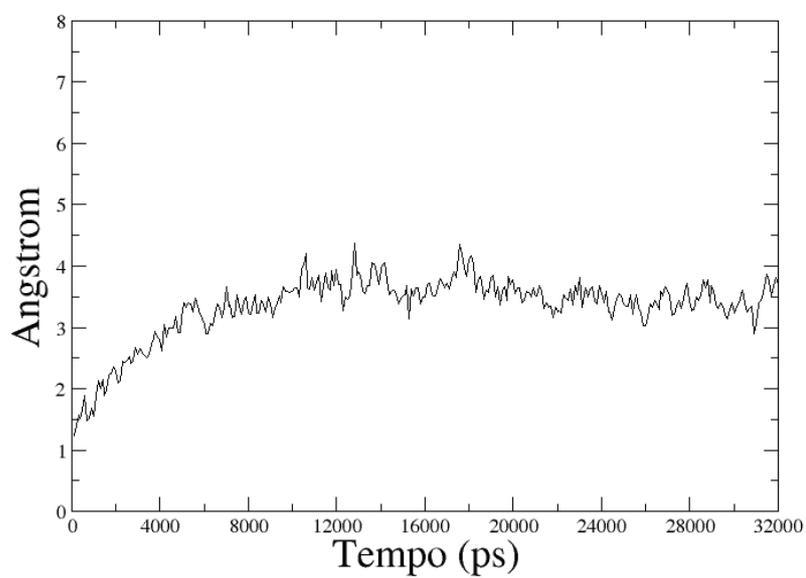
A figura 4.24 mostra a variação do RMSD do domínio ECD da CD39 com e sem o ANP no sítio ativo. No gráfico da CD39 (figura 4.24(a)) é possível observar uma estabilização na variação do RMSD, o que indica uma estabilização conformacional do domínio ECD durante a trajetória da dinâmica molecular. A variação do RMSD da CD39 é muito semelhante à variação do RMSD do molde PDB3ZX3 (figura 4.20(a)). É interessante lembrar que a CD39, bem como a RnNTPDase1 e RnNTPDase2 possuem cinco ligações dissulfeto, e essas proporcionam uma maior estabilidade do domínio ECD (ZEBISCH; STRÄTER, 2008; ZEBISCH et al., 2013).

Com relação ao gráfico 4.24(b) ele mostra uma estabilização na variação do RMSD para o domínio ECD da CD39-ANP. Quando comparado com o gráfico do ECD da CD39 (4.24), podemos ver que os valores do RMSD na CD39-ANP é um pouco maior do que na CD39. Isso provavelmente se deve à uma maior movimentação do ECD2 na CD39-ANP (tabela 4.2). Além da presença das cinco ligações dissulfeto, o ligante também está contribuindo com a estabilidade do ECD.

Os resultados da análise do RMSD para o domínio ECD nos moldes e modelos sugerem: (i) que o ligante contribui para estabilidade do domínio ECD, o que pode ser visto ao comparar os gráficos do RMSD da Sm1 na ausência (4.21(a)) e presença do ligante (4.21(b)) no sítio ativo; (ii) que o maior número de pontes dissulfeto presentes no ECD da CD39, RnNTPDase1 e RnNTPDase2 possibilitam uma maior estabilidade do domínio ECD em relação ao domínio ECD da Sm1, o que pode ser visto ao comparar os gráficos 4.21(a), 4.20(a) e 4.24(a).



(a) RMSD da CD39.



(b) RMSD da CD39-ANP.

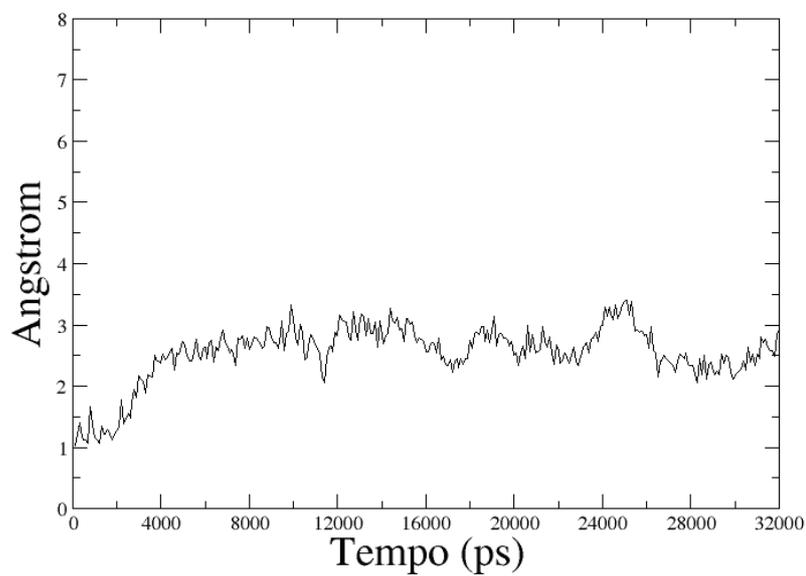
Figura 4.24: Variação do RMSD referente ao domínio ECD nas simulações envolvendo a CD39 e CD39-ANP. RMSD calculado usando o $C\alpha$.

4.3.3.2 RMSD das TM

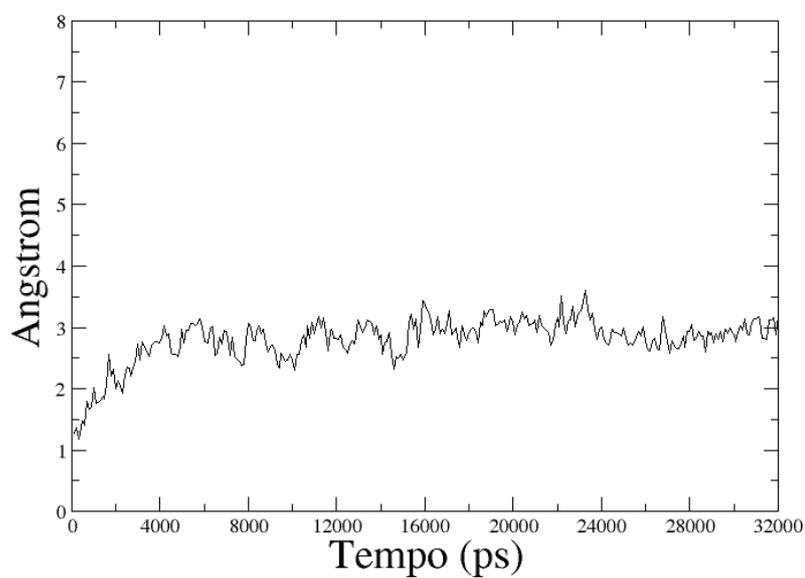
Nos gráficos da figura 4.25 pode-se observar que houve uma estabilização no RMSD referente aos domínios transmembranares da Sm1 nas simulações com (4.25(b)) e sem ligante (4.25(a)) no sítio ativo da proteína. Também é possível observar uma maior flutuação no gráfico do RMSD da Sm1 sem o ligante ANP (figura 4.21(a)), o que sugere uma maior estabilidade da proteína quando na presença do ligante.

Nos gráficos da figura 4.26 é possível observar uma estabilização no RMSD referente aos domínios transmembranares da CD39 e CD39-ANP. Isso sugere que os domínios transmembranares não apresentaram mudanças significativas na conformação ao longo do tempo simulado.

Podemos concluir pelos gráficos do RMSD referentes aos domínios transmembranares de ambas estruturas modeladas, que tais domínios não apresentaram mudanças significativas ao longo das simulações. Também não foi possível observar no tempo de simulação realizado, uma relação entre as mudanças conformacionais das TM e a região do sítio ativo de ambas as enzimas, conforme proposto por Grinthal e Guidotti (2006).

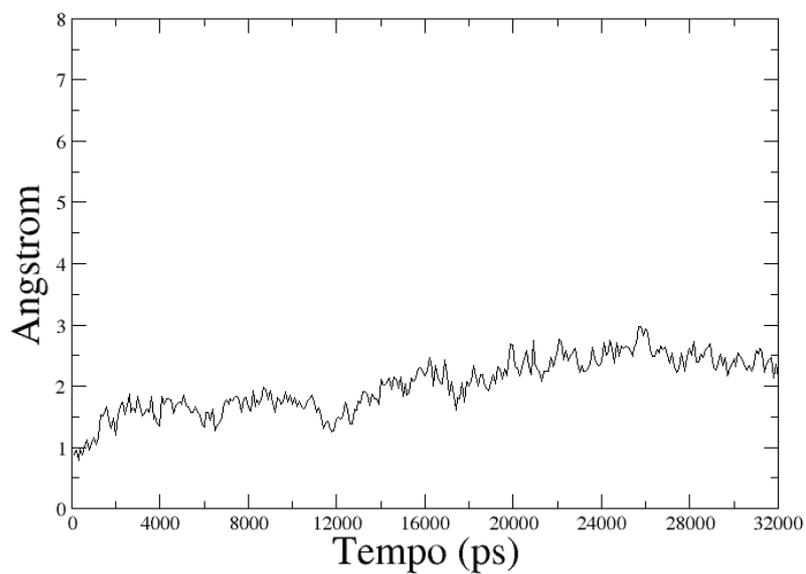


(a) RMSD da Sm1.

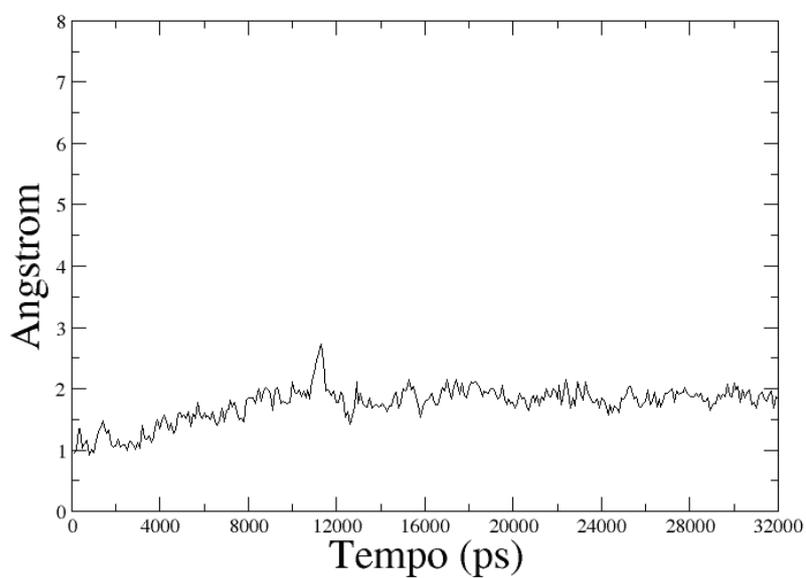


(b) RMSD da Sm1-ANP.

Figura 4.25: Variação do RMSD referente às TM nas simulações envolvendo a Sm1 e Sm1-ANP. RMSD calculado usando o $C\alpha$.



(a) RMSD da CD39.



(b) RMSD da CD39-ANP.

Figura 4.26: Variação do RMSD referente às TM nas simulações envolvendo a CD39 e CD39-ANP. RMSD calculado usando $C\alpha$.

4.3.4 Raio de giro

A seguir serão apresentados os gráficos do raio de giro para os domínios ECD e transmembranares em função do tempo para todas as simulações de dinâmica molecular envolvendo os moldes e modelos.

4.3.4.1 Raio de giro dos ECD

A figura 4.27 apresenta os gráficos do raio de giro dos domínios ECD dos moldes PDB3ZX3 (4.27(a)) e PDB3CJA (4.27(b)). É possível observar uma estabilidade em ambas as proteínas, como apontaram as análises do RMSD.

Pelo gráfico 4.28(a) observa-se que a dinâmica do domínio ECD da Sm1 na ausência do ligante difere da dinâmica observada nos domínios ECD dos moldes e da CD39. O mesmo não ocorre com a Sm1-ANP, ou seja, o gráfico do raio de giro da Sm1-ANP (figura 4.28(b)) apresenta uma estabilidade semelhante ao observado nos gráficos do raio de giro da CD39 e moldes.

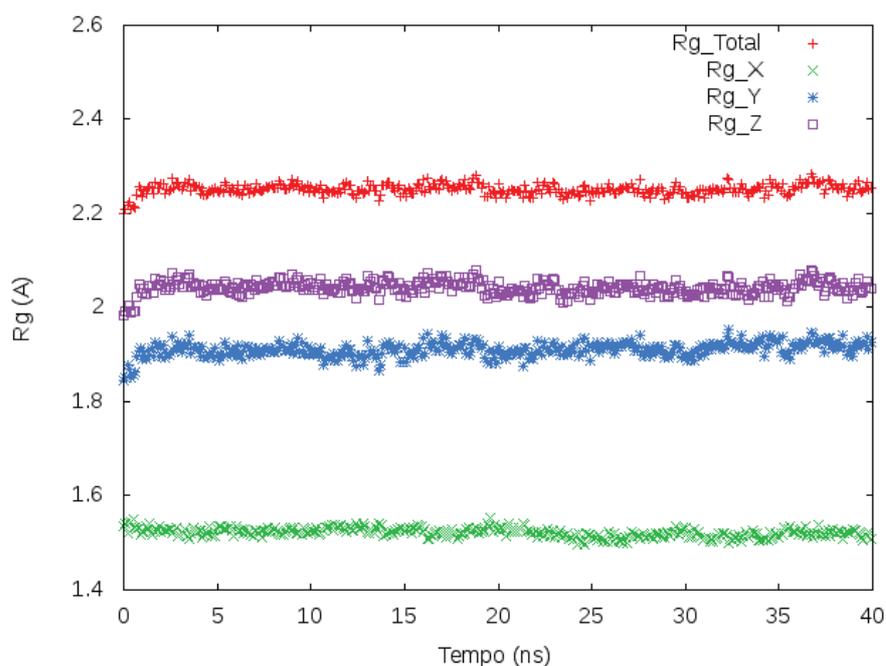
Os gráficos do raio de giro referentes às simulações envolvendo a CD39 com (4.29(b)) e sem ligante (4.29(a)) indicam uma estabilidade do domínio ECD em ambas as simulações. A estabilidade observada mesmo na simulação na ausência do ligante, reforça a importância das ligações dissulfeto, como já foi observado nas análises do RMSD.

A análise do raio de giro reforça a observação de que as mudanças conformacionais que ocorrem na Sm1, em destaque as mudanças na região da cavidade-2, são observadas somente na ausência do ligante, e não ocorrem na CD39 e nos moldes. Portanto, a análise do raio de giro, assim como a do RMSD, apontam para a importância das ligações dissulfeto e a presença do ligante no sítio ativo na estabilidade do domínio ECD.

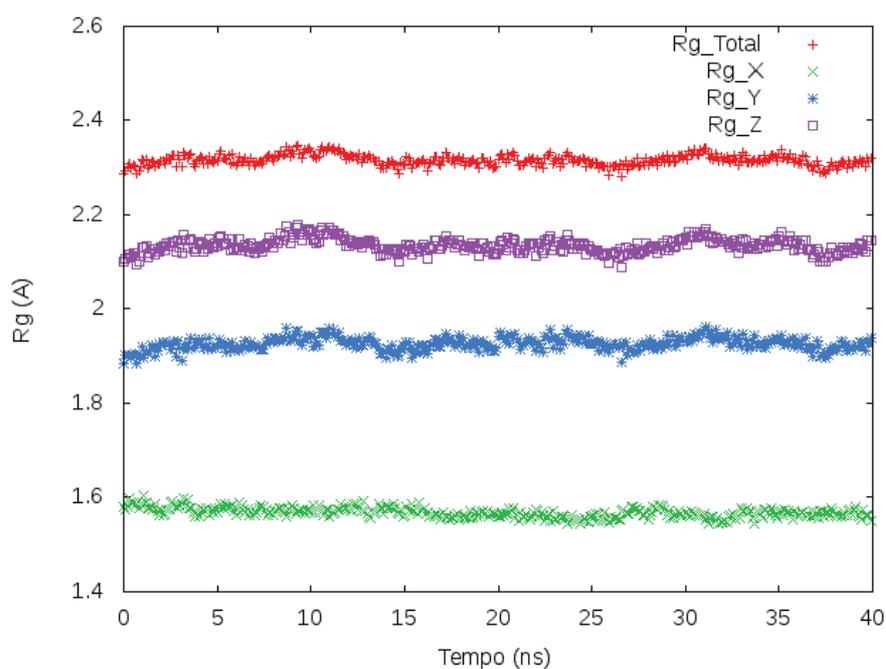
4.3.4.2 Raio de giro das TM

As figuras 4.30(a) e 4.30(b) mostram os gráficos do raio de giro referente às TM nas simulações de dinâmica molecular envolvendo o modelo da Sm1 na presença e ausência do ligante ANP. Nesses gráficos é possível observar que não houve nenhuma variação significativa no raio de giro da Sm1 e Sm1-ANP. Isso indica uma estabilidade dos domínios transmembranares durante as simulações de dinâmica molecular envolvendo essa enzima, o que também pôde ser observado pela análise do RMSD.

As figuras 4.31(a) e 4.31(b) mostram os gráfico do raio de giro referente às TM nas



(a) Raio de giro da proteína PDB3ZX3 (RnNTPDase1).

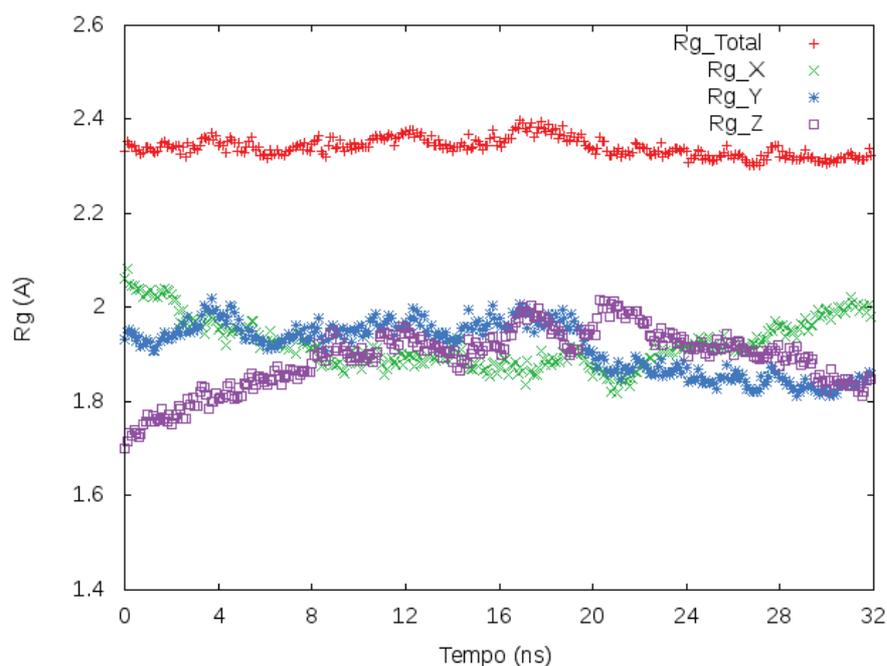


(b) Raio de giro da proteína PDB3CJA (RnNTPDase2).

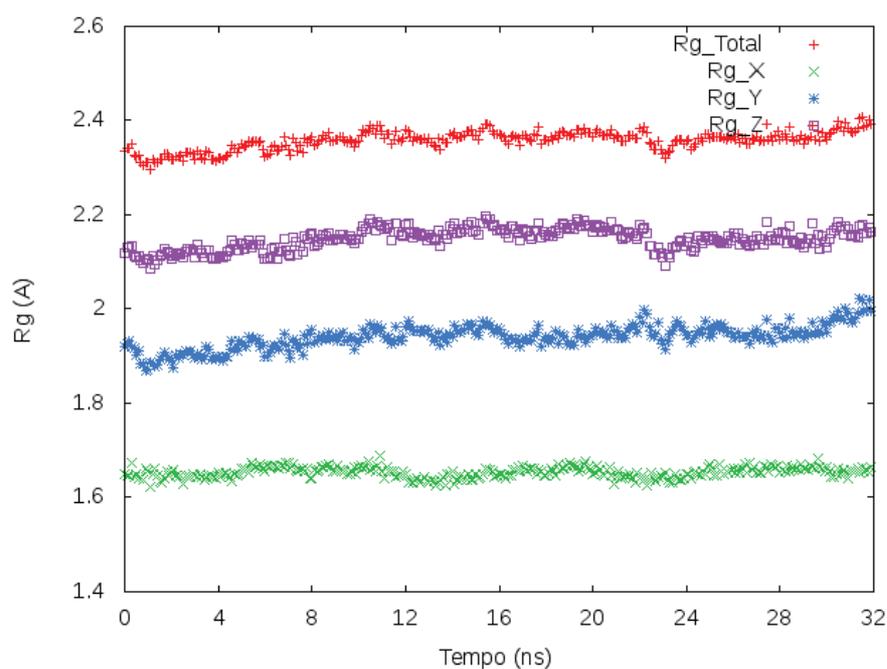
Figura 4.27: Variação do raio de giro nas simulações com os moldes.

simulações de dinâmica molecular da CD39 e CD39-ANP. Pelos gráficos observa-se uma estabilização do raio de giro referente aos domínios transmembranares da CD39 e CD39-ANP ao longo das simulações, o que também foi observado na análise do RMSD dessa enzima.

Os gráficos do raio de giro referentes aos domínios transmembranares da Sm1 e



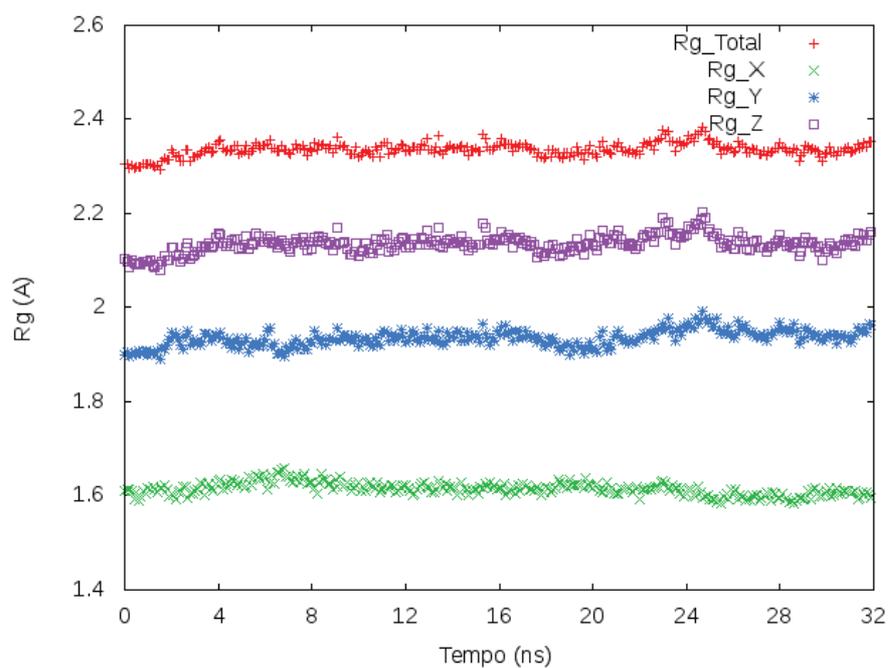
(a) Raio de giro da Sm1.



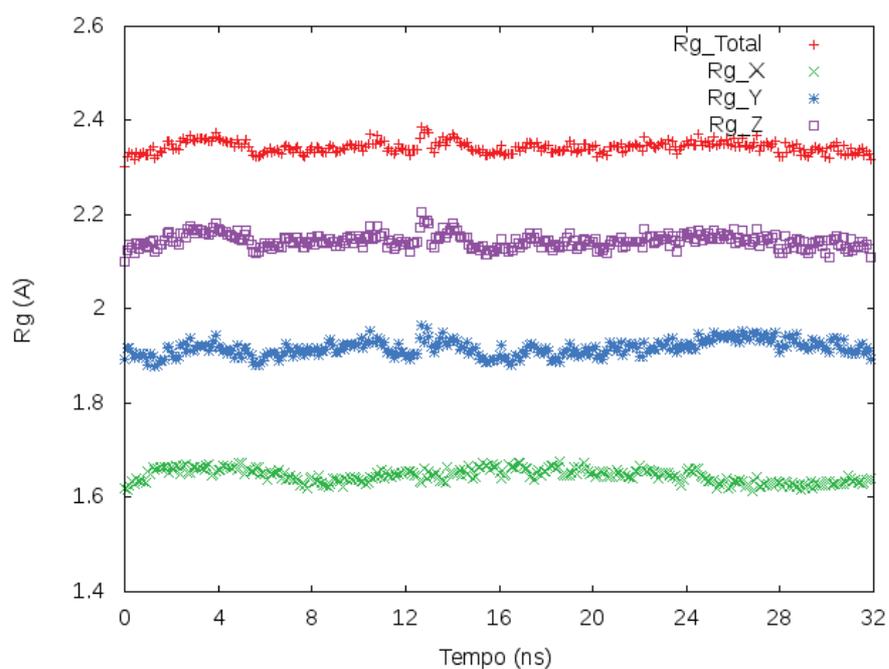
(b) Raio de giro da Sm1-ANP.

Figura 4.28: Gráfico do raio de giro referente ao domínio ECD nas simulações envolvendo a Sm1 e Sm1-ANP. RMSD calculado usando $C\alpha$.

CD39, mostram que tais domínios não apresentaram mudanças significativas ao longo das simulações. Assim como nas análises do RMSD, as análises do raio de giro das TM também não apontam uma relação entre as mudanças conformacionais das TM e a região do sítio ativo de ambas as enzimas (proposto por Grinthal e Guidotti (2006)). Talvez seja



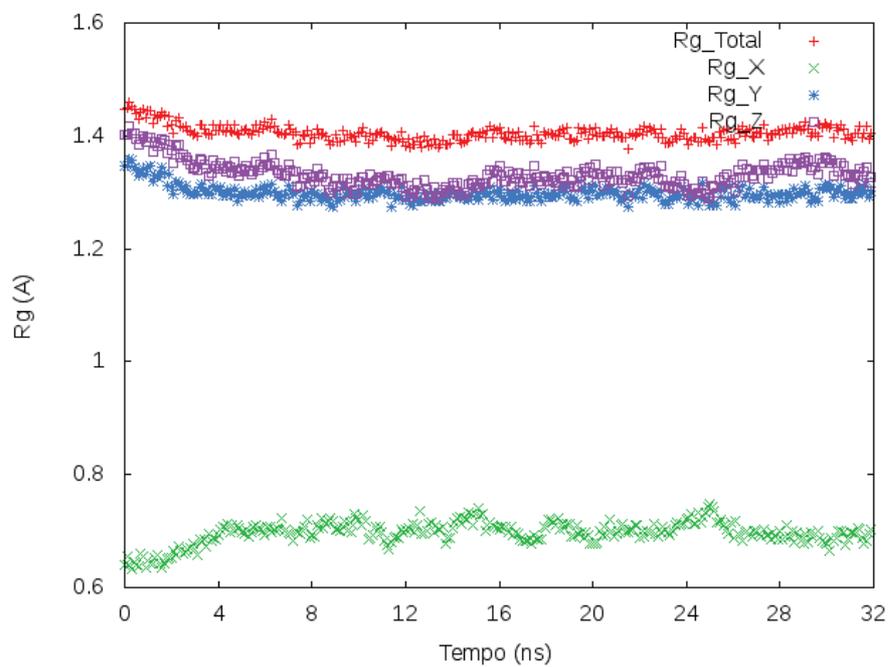
(a) Raio de giro da CD39.



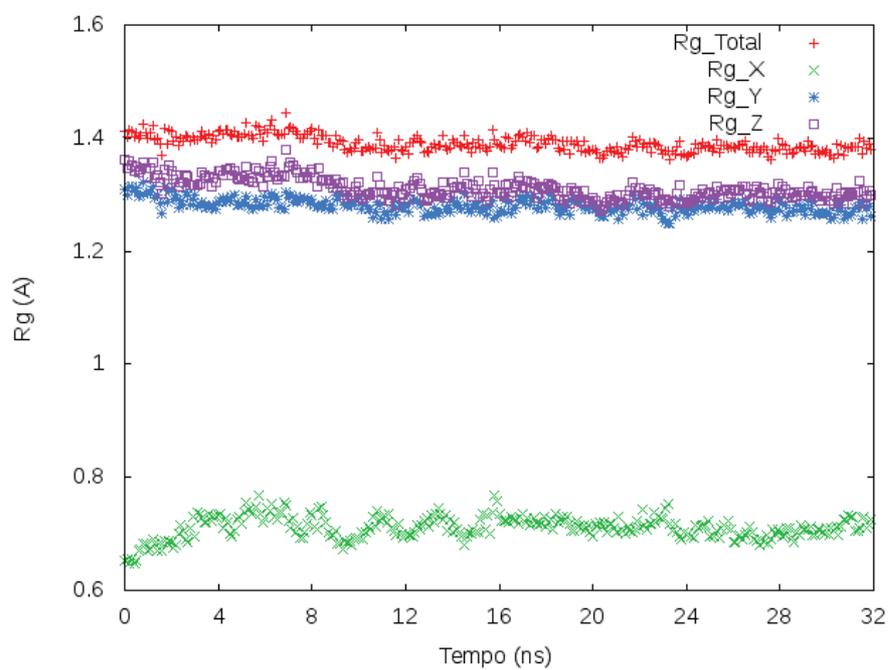
(b) Raio de giro da CD39-ANP.

Figura 4.29: Gráfico do raio de giro referente ao domínio ECD nas simulações envolvendo a CD39 e CD39-ANP.

necessário um maior tempo de simulação para ser possível observar essa relação.

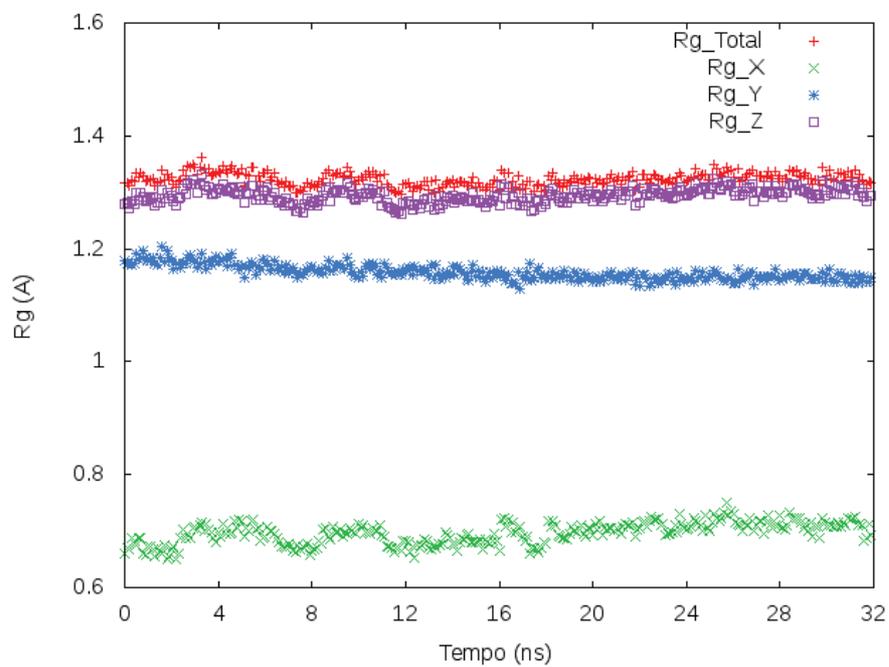


(a) Raio de giro da Sm1.

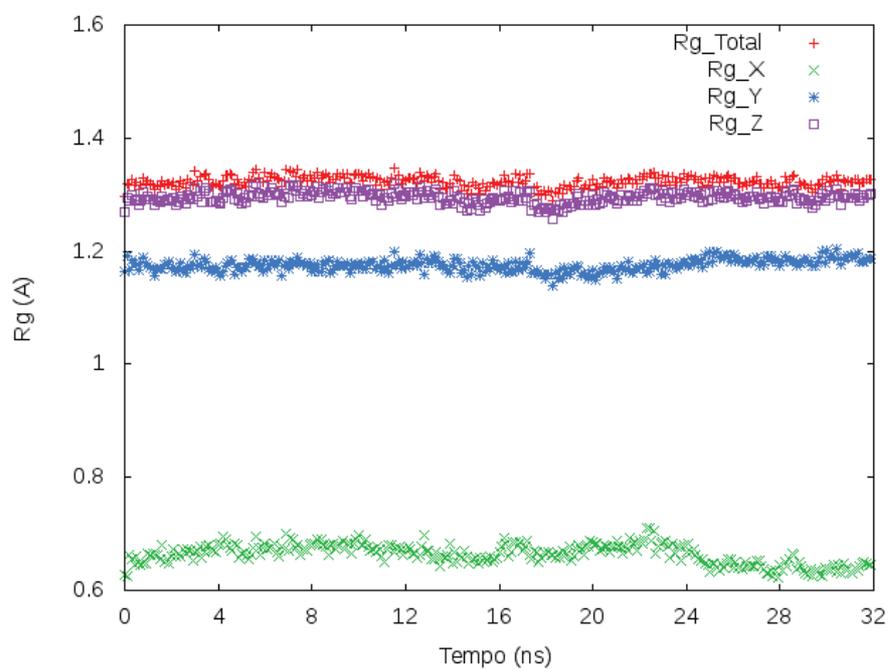


(b) Raio de giro da Sm1-ANP.

Figura 4.30: Gráfico do raio de giro referente às TM nas simulações envolvendo a Sm1 e Sm1-ANP.



(a) Raio de giro da CD39.



(b) Raio de giro da CD39-ANP.

Figura 4.31: Gráfico do raio de giro referente às TM nas simulações envolvendo a CD39 e CD39-ANP.

4.3.5 *Águas conservadas*

A respeito das seis moléculas de água consideradas como conservadas por Zebisch e colaboradores (2013), ao final de cada nanosegundo de simulação, foi verificada a conservação dessas moléculas. Em todas as simulações com o ligante ANP presente no sítio ativo (cavidade-1) foi possível verificar que todas as seis moléculas de água inicialmente no sistema foram substituídas por outras moléculas também de água, sendo conservadas suas posições. Já nas simulações sem o ligante ANP no sítio ativo, não foi possível observar a conservação de posição das moléculas de água ao final de cada nanosegundo de simulação.

4.3.6 *O íon cálcio*

Nas simulações de dinâmica molecular também foi verificada a permanência e interações do íon cálcio no sítio ativo (cavidade-1), independente da presença ou não do ligante. Para todas as simulações com a presença do ligante ANP no sítio ativo, foi verificado que o íon cálcio interage com os grupos γ -fosfato e β -fosfato do ANP, e com até cinco moléculas de água. Para as simulações na ausência do ligante ANP no sítio ativo, foi observado que o íon cálcio permanece no sítio ativo durante toda a trajetória, porém, interagindo com o resíduo D78 na Sm1 e D54 na CD39, e até cinco moléculas de água.

4.4 *Análise de drogabilidade das cavidades*

As análises de drogabilidade envolveram as conformações geradas na etapa de produção (25ns) das quatro simulações de dinâmica molecular com modelos da Sm1 e CD39. Essas análises foram necessárias para auxiliar na seleção de conformações a serem usadas nos estudos de *docking*. A seguir serão apresentados os gráficos da variação do *drugscore* para as duas cavidades ao longo das trajetórias das quatro simulações.

Nos gráficos 4.32(a) e 4.32(b) são mostradas a variação do *drugscore* na cavidade-1 e cavidade-2 da Sm1 sem o ligante ANP no sítio ativo. Pelo gráfico 4.32(a), observa-se que os valores do *drugscore* referentes a cavidade-1 da Sm1 mantêm-se altos (próximos de 0,8) ao longo da trajetória, com poucas conformações apresentando valores mais baixos (abaixo de 0,5). Até o final dos 25ns de simulação da dinâmica de produção há um ligeiro aumento do *drugscore*.

Para a cavidade-2 (figura 4.32(b)) os valores do *drugscore* são mais altos para as

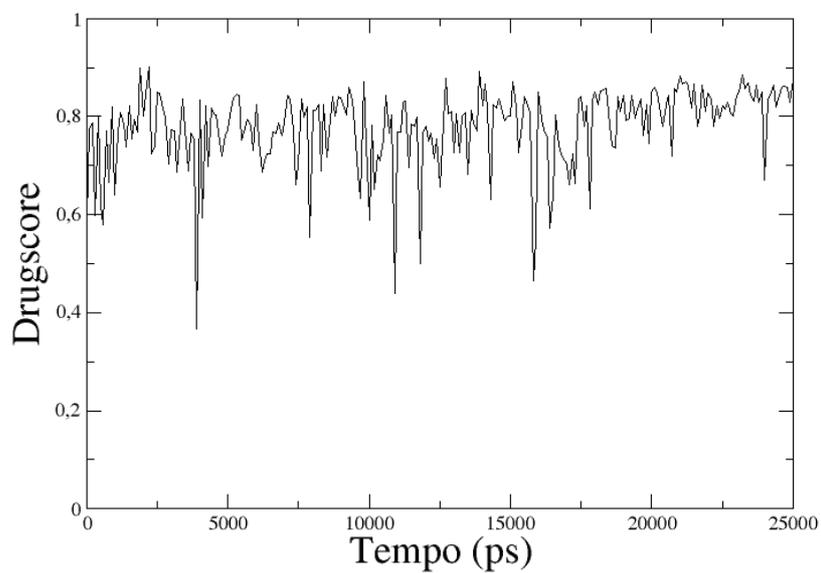
conformações da proteína que estão no início dos 25ns da etapa de produção da DM. Porém, ao longo da trajetória os valores do *drugscore* diminuem (abaixo de 0,5), sendo essa diminuição mais proeminente próximo aos 22ns de simulação seguida de novo aumento do valor, atingindo o mesmo patamar próximo aos 25ns. Essa diminuição do *drugscore* pode estar associada à movimentação da região da Sm1 onde está localizada a cavidade-2 (domínio ECD-II). A manutenção desses valores só poderia ser verificada com um maior tempo de simulação.

Com relação a cavidade-1 na simulação da Sm1-ANP, pode-se observar pelo gráfico 4.33(a) uma oscilação do valor do *drugscore* em intervalos de tempo mais longos, apresentando mais conformações com valores abaixo de 0,5. Para a cavidade-2, é possível observar pelo gráfico 4.33(b) que os valores do *drugscore* mantêm-se estáveis ao longo da simulação.

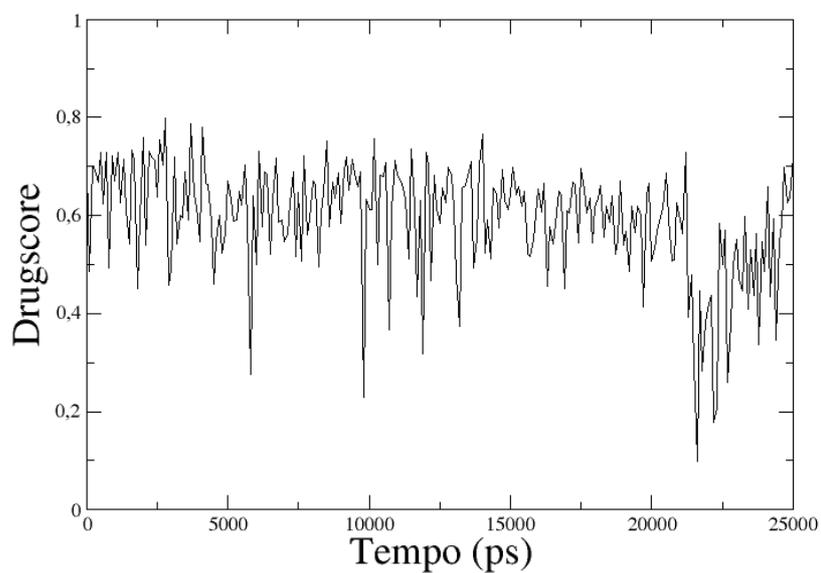
Aqui é importante lembrar dos dados já apresentados nesse trabalho sobre a maior movimentação da região da cavidade-2 que ocorre na simulação da Sm1 sem ANP, e da maior estabilidade do domínio ECD na simulação da Sm1 com ANP na cavidade-1. Tais fatores poderiam estar contribuindo para o comportamento observado para o *drugscore*.

A seguir, são apresentados os gráficos referentes aos valores do *drugscore* na simulação da CD39 para as cavidades 1 (4.34(a)) e 2 (4.34(b)). Ambos os gráficos do *drugscore* para a CD39 apresentam uma acentuada oscilação, principalmente para a cavidade-2. Porém, os valores do *drugscore* para a cavidade-1 foram melhores (mais próximos de 0,8) do que para a cavidade-2.

O gráfico 4.35(a) mostra a variação do *drugscore* na cavidade-1 da CD39. Pode-se observar uma diminuição do *drugscore* até 15.000ps de simulação, e em seguida um aumento do *drugscore*. Por outro lado, no gráfico que mostra a variação do *drugscore* na cavidade-2 da CD39 (figura 4.35(b)), pode-se observar que os valores são baixos e uma grande oscilação ao longo da trajetória.

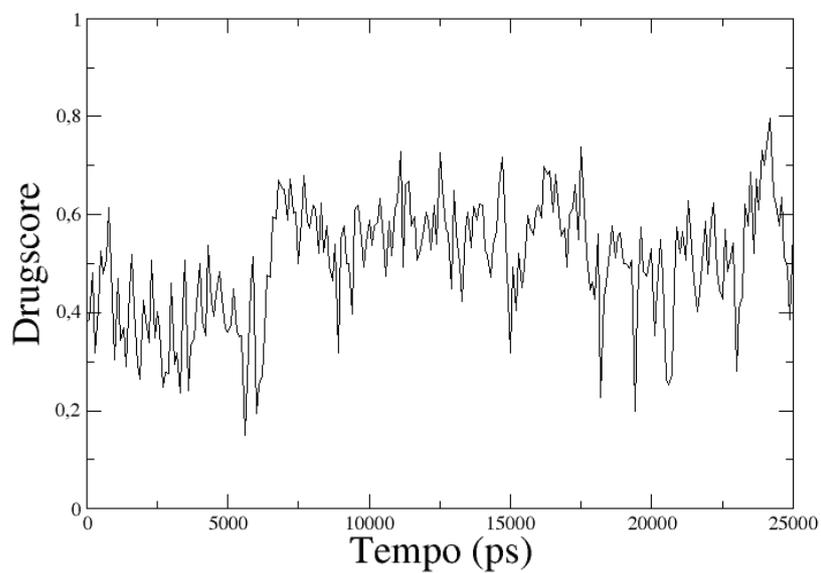


(a) Cavidade-1.

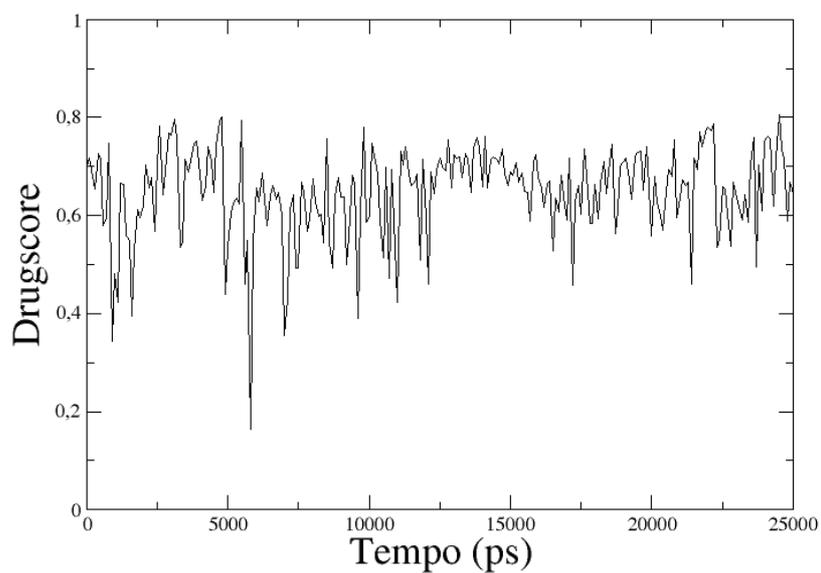


(b) Cavidade-2.

Figura 4.32: Variação do *Drugscore* das cavidades 1 e 2 da Sm1 ao longo da DM.

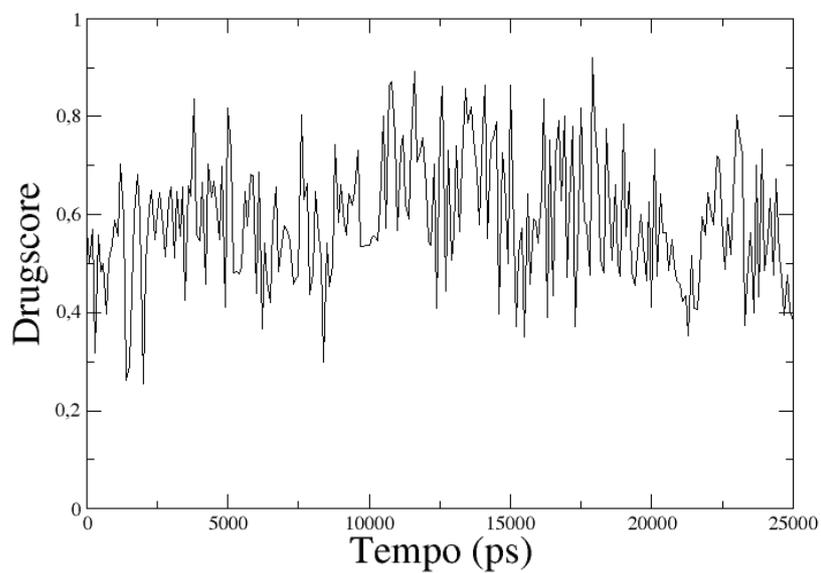


(a) Cavidade-1.

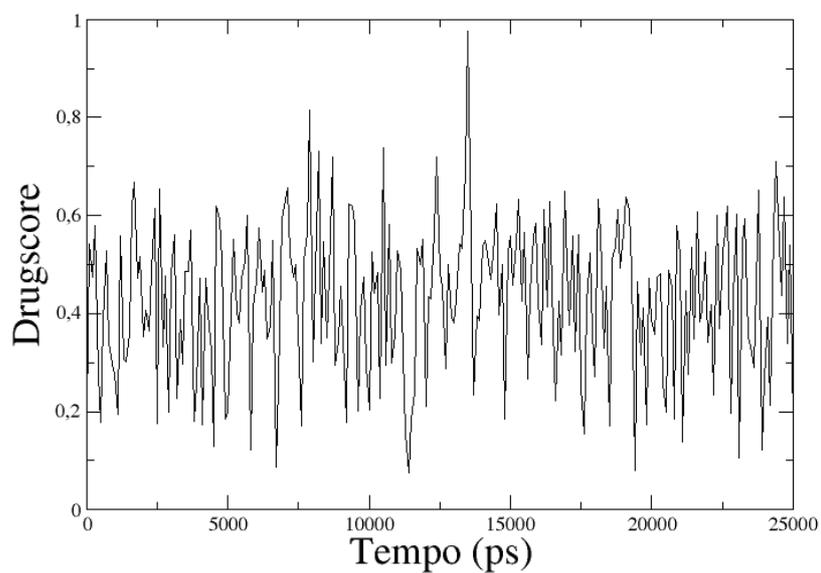


(b) Cavidade-2.

Figura 4.33: Variação do *Drugscore* das cavidades 1 e 2 da Sm1-ANP ao longo da DM.

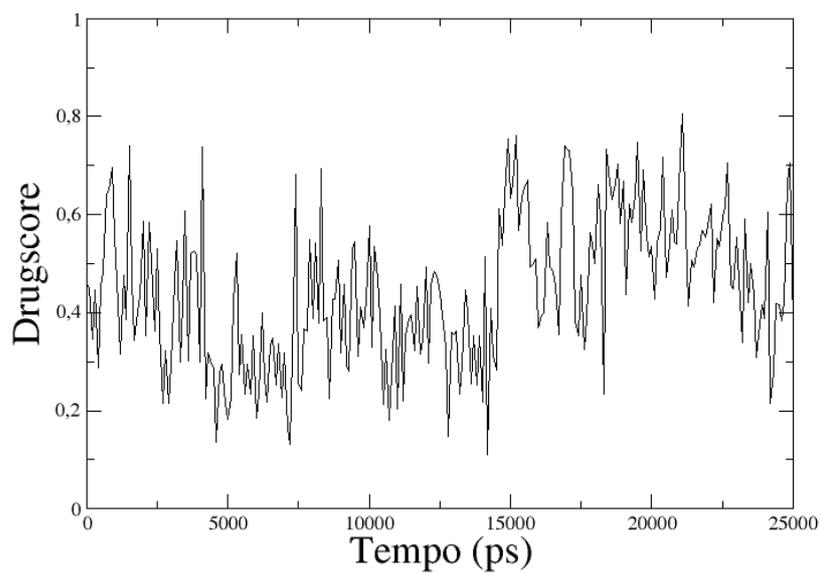


(a) Cavidade-1.

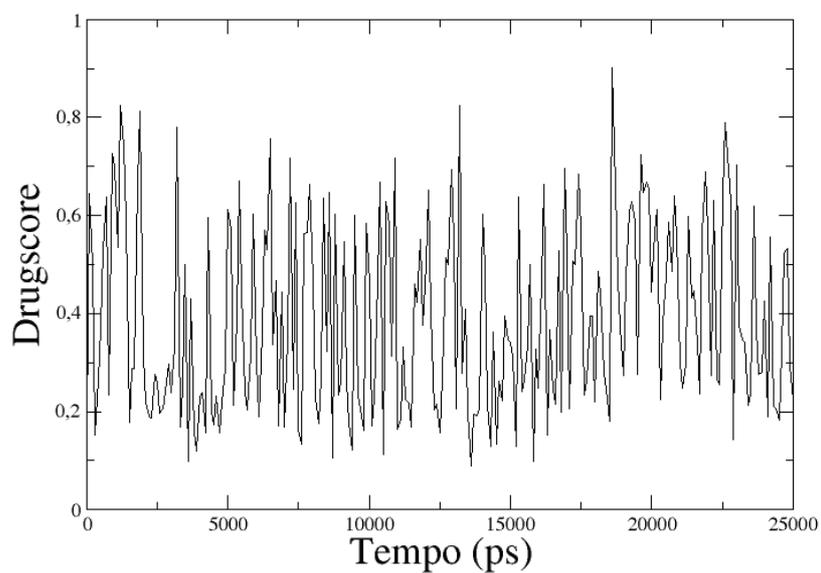


(b) Cavidade-2.

Figura 4.34: Variação do *Drugscore* das cavidades 1 e 2 da CD39 ao longo da DM.



(a) Cavidade-1.



(b) Cavidade-2.

Figura 4.35: Variação do *Drugscore* das cavidades 1 e 2 da CD39-ANP ao longo da DM.

4.5 Análise de agrupamento

Neste trabalho foram testados dois tipos de agrupamento: (i) baseado no programa *g_cluster* usando o método *gromos* de agrupamento; (ii) baseado no método GLCM usando como método de agrupamento o *K-Means*. Os resultados dos agrupamentos foram analisados em função do valor do *drugscore*.

Para o programa *g_cluster* foram feitos testes com diferentes valores de *cutoff* para o RMSD. Esses testes mostram que o número de agrupamentos formados por esse programa é dependente do valor do *cutoff* definido pelo usuário, o que torna o processo um tanto subjetivo. Na abordagem usando o GLCM com o *K-Means*, o valor de *K* também deve ser definido, entretanto, a fim de evitar essa dependência, propomos nesse trabalho o uso do algoritmo PSO para otimizar o *K* número de grupos no *K-Means*, tornando o GLCM um método mais automático para este tipo de problema.

Na Tabela 4.3 é apresentado um comparativo dos agrupamentos obtidos com o GLCM e com o *g_cluster*. Os resultados obtidos para os diferentes valores de *cutoff* também são apresentados. As tabelas 4.4, 4.5, 4.6 e 4.7 resumem os resultados encontrados para o GLCM e para o *g_cluster* com valores de *cutoff* de 1,5 para a Sm1 e de 1,0 para a CD39.

Tabela 4.3: Resumo dos agrupamentos obtidos com os métodos *g_cluster* e GLCM para as quatro dinâmicas moleculares testadas.

Região	Proteína	<i>g_cluster</i> ^a								GLCM
		1,0Å	1,1Å	1,2Å	1,3Å	1,4Å	1,5Å	1,6Å	1,7Å	<i>K</i> ^b
Cavidade-1	Sm1	8	7	5	3	2	3	2	1	3
	Sm1-ANP	10	7	6	5	4	3	3	2	2
	CD39	3	2	1	1	1	1	1	1	2
	CD39-ANP	7	4	5	3	2	2	2	1	3
Cavidade-2	Sm1	38	25	16	11	9	7	5	3	9
	Sm1-ANP	25	14	10	6	5	4	4	3	10
	CD39	1	1	1	1	1	1	1	1	6
	CD39-ANP	1	1	1	1	1	1	1	1	3

^aDiferentes valores de *cutoff* testados no *g_cluster*. ^bMelhor valor de *K* encontrado pelo PSO.

Tabela 4.4: Estrutura média e respectivo *drugscore* dos agrupamentos da Sm1 com os métodos *g_cluster* e GLCM.

Região	Método	EM ^a	<i>Drugscore</i>
Cavidade-1	<i>g_cluster</i> ^b	5	0,616
		158	0,464
		194	0,797
	GLCM	27	0,816
		55	0,752
		108	0,804
Cavidade-2	<i>g_cluster</i> ^b	2	0,701
		20	0,759
		33	0,600
		79	0,594
		154	0,619
		191	0,567
	GLCM	37	0,788
		61	0,732
		119	0,319
		122	0,468
		126	0,657
		167	0,650
		171	0,605
		223	0,209
226	0,571		

^aO valor da estrutura média (EM) corresponde ao índice do *frame* na dinâmica molecular. ^b*Cutoff* de 1,5Å.

Tabela 4.5: Estrutura média e respectivo *drugscore* dos agrupamentos da Sm1-ANP com os métodos *g_cluster* e GLCM.

Região	Método	EM ^a	<i>Drugscore</i>
Cavidade-1	<i>g_cluster</i> ^b	86	0,492
		160	0,618
		248	0,493
	GLCM	149	0,427
181		0,560	
Cavidade-2	<i>g_cluster</i> ^b	6	0,580
		63	0,632
		181	0,593
		207	0,679
	GLCM	65	0,639
		149	0,662
		162	0,616
		176	0,736
		188	0,612
		197	0,652
		214	0,460
		219	0,771
		233	0,642
		239	0,609

^aO valor da estrutura média (EM) corresponde ao índice do *frame* na dinâmica molecular. ^b*Cutoff* de 1,5Å.

Tabela 4.6: Estrutura média e respectivo *drugscore* dos agrupamentos da CD39 com os métodos *g_cluster* e GLCM.

Região	Método	EM ^a	<i>Drugscore</i>
Cavidade-1	<i>g_cluster</i> ^b	15	0,293
		179	0,921
		232	0,721
	GLCM	58	0,681
		219	0,659
Cavidade-2	<i>g_cluster</i> ^b	112	0,319
	GLCM	11	0,194
		150	0,568
		160	0,420
		187	0,534
		226	0,556
		237	0,427

^aO valor da estrutura média (EM) corresponde ao índice do *frame* na dinâmica molecular. ^b*Cutoff* de 1,0Å.

Tabela 4.7: Estrutura média e respectivo *drugscore* dos agrupamentos da CD39-ANP com os métodos *g_cluster* e GLCM.

Região	Método	EM ^a	<i>Drugscore</i>
Cavidade-1	<i>g_cluster</i> ^b	39	0,519
		72	0,131
		81	0,441
		87	0,224
		184	0,733
		213	0,413
		249	0,705
	GLCM	80	0,488
		173	0,381
		230	0,583
Cavidade-2	<i>g_cluster</i> ^b	119	0,377
	GLCM	64	0,563
		80	0,468
		245	0,183

^aO valor da estrutura média (EM) corresponde ao índice do *frame* na dinâmica molecular. ^b*Cutoff* de 1,0Å.

Como afirmado na seção anterior, a seleção das estruturas para os estudos de *docking* foi feita com base nas conformações (*frames* da dinâmica molecular) que foram estruturas média dos agrupamentos e no valor do *drugscore*. Na Tabela 4.8 estão relacionadas as conformações selecionadas e os respectivos valores do *drugscore* para as duas cavidades estudadas neste trabalho.

Tabela 4.8: Estruturas médias selecionadas para os estudos de *docking* da Sm1, Sm1-ANP, CD39 e CD39-ANP.

Proteína	EM ^a	<i>Drugscore</i> Cavidade-1	<i>Drugscore</i> Cavidade-2
Sm1	0 ^b	0,588	0,699
	37	0,689	0,788
	55	0,752	0,621
	108	0,804	0,613
	171	0,605	0,610
Sm1-ANP	0 ^b	0,382	0,699
	149	0,427	0,662
	219	0,777	0,771
CD39	0 ^b	0,569	0,188
	187	0,662	0,534
CD39-ANP	0 ^b	0,459	0,152
	80	0,488	0,468

^aO valor da estrutura média (EM) corresponde ao índice do *frame* na dinâmica molecular. ^bEsse valor corresponde ao *frame-0* (última conformação da etapa de equilibração da dinâmica molecular), considerado como estrutura de referência para análises de mudanças conformacionais.

4.6 Conformações da Sm1 e o movimento de K279

Os resultados aqui apresentados sugerem maiores mudanças conformacionais na Sm1 sem ANP, em relação ao *frame-0*, quando comparada com as demais estruturas simuladas. Nas figuras seguintes são apresentadas as conformações que foram selecionadas, com base nas análises de agrupamento e de drogabilidade, para os estudos de *docking* usando a Sm1.

A figura 4.6 mostra o alinhamento estrutural a partir do domínio ECD das conformações da Sm1 provenientes das simulações de dinâmica molecular com a presença e ausência do ligante ANP no sítio ativo, e que foram usadas nos estudos de *docking*.

Analisando as mudanças conformacionais da Sm1 sem o ligante no sítio-ativo ao logo da simulação, foi possível observar uma movimentação do resíduo K279 localizado próximo ao sítio-ativo da proteína. Essa movimentação de K279 também foi detectada analisando as estruturas médias dos agrupamentos formados pelo *K-Means*. As figuras 4.37(a) e

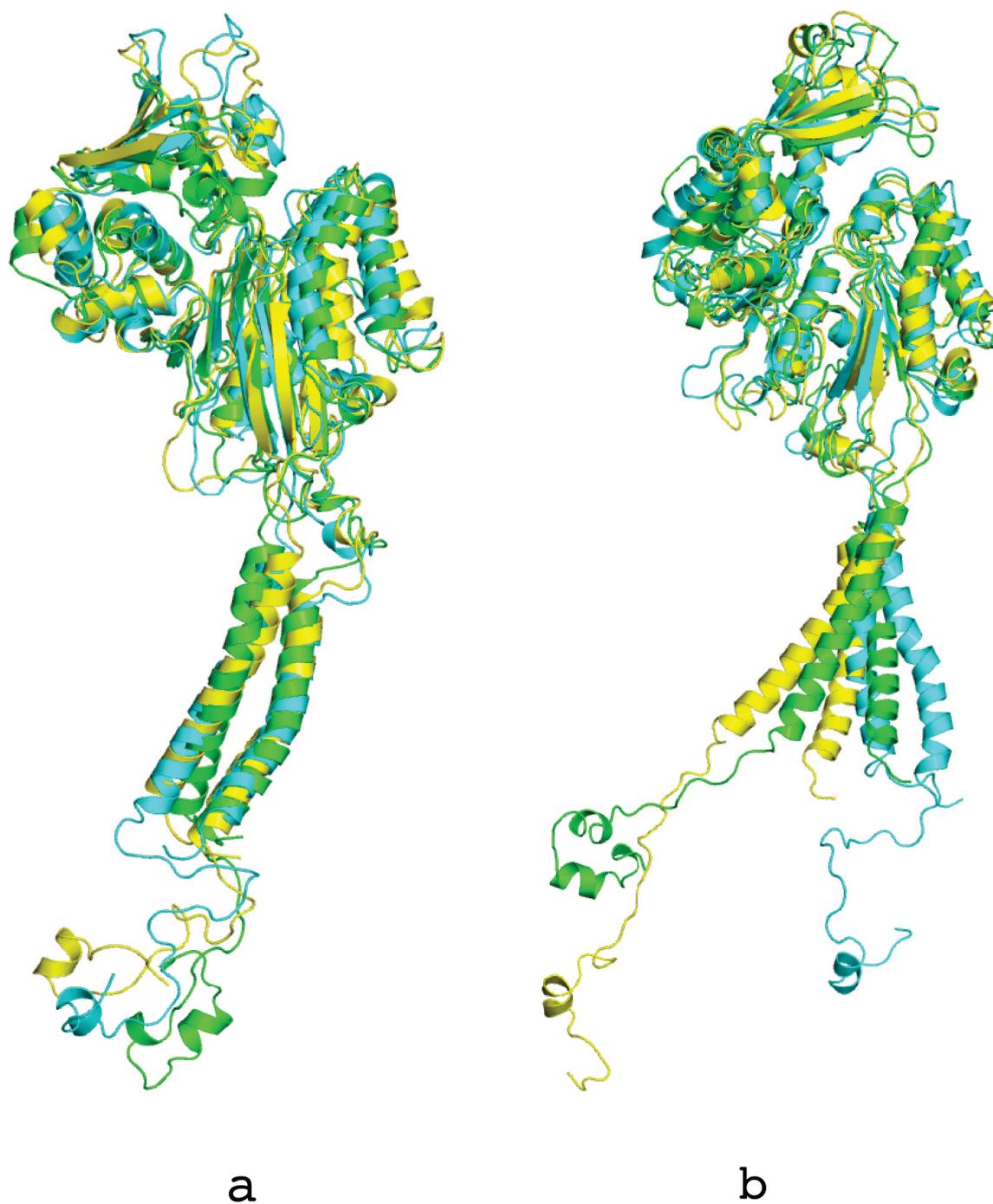


Figura 4.36: Alinhamento estrutural do ECD de conformações da Sm1 com e sem ligante.

(a) Conformações da Sm1. Em azul o *frame-0*, em amarelo o *frame-55* e em verde o *frame-171*. (b) Conformações da Sm1-ANP. Em azul o *frame-0*, em amarelo o *frame-149* e em verde o *frame-219*.

4.37(b) mostram a movimentação do domínio ECD a partir do alinhamento estrutural de três das conformações que são estruturas média, e a movimentação do resíduo K279.

Quando o ligante está presente no sítio ativo, a cadeia lateral da lisina K279 posiciona-se para dentro do sítio ativo, como pode ser visto no modelo da Sm1 (figura 4.37(b)). Porém, ao longo da trajetória a cadeia lateral de K279 foi sendo projetado para fora do sítio ativo, como pode ser observado quando comparamos os diferentes *frames*. É possível ver pela figura 4.37(b) que a lisina K279 nos *frames*-0 (em azul) e -55 (em amarelo) encontra-se em uma posição mais afastada da sua posição no modelo da Sm1-ANP (em cinza) antes da simulação DM. Entretanto, no *frame*-171 (em verde), que corresponde a pouco mais de 17ns de dinâmica de produção, o resíduo encontra-se novamente na posição observada no modelo da Sm1-ANP antes da DM, quando o sítio ativo está ocupado pelo ligante. Este movimento coincide com um aumento do *drugscore* do sítio ativo (Figura 4.32(a)). Vale destacar que essa movimentação de K279 não foi verificada na simulação da Sm1 com ANP presente no sítio ativo.



(a) Mudanças no domínio ECD.



(b) Movimentação de K279.

Figura 4.37: Alinhamento estrutural do domínio ECD de três conformações da Sm1. (b) Em azul o *frame-0*, em amarelo o *frame-55*, em verde o *frame-171*, e em cinza o modelo da Sm1 antes da simulação DM.

4.7 Estudos de *docking* receptor-ligante

Para as análises de *docking* receptor-ligante foram consideradas a pontuação (*XPGScore*), o RMSD e a posição do ligante nas cavidades. Foram realizados testes considerando a presença e a ausência do Ca^{+2} e águas conservadas no sítio ativo. A seguir serão apresentados os resultados obtidos das quatro simulações de dinâmica molecular realizadas neste trabalho com a Sm1 e CD39. Em testes preliminares, foi possível observar que para os ligantes ANP e ARL67 a melhor configuração do sítio ativo para os testes de *docking* foi considerando a presença dos heteroátomos, enquanto para o praziquantel a melhor conformação do sítio ativo foi na ausência dos heteroátomos.

A seguir, serão apresentados os melhores resultados de *docking* obtidos por ligante e por cavidade analisados neste trabalho. Os resultados do *docking* serão confrontados com os dois modelos propostos para a atividade catalítica e apresentados na subseção 4.2.3. A tabela 4.9 resume os valores das poses de menor pontuação (*XPGScore*) de cada uma das simulações. Nessa tabela podemos observar que todos os ligantes tiveram melhor pontuação na cavidade-1, com exceção ao praziquantel que não apresentou diferença observável entre as duas cavidades estudadas.

Tabela 4.9: Resultados do *XPGScore* nos *dockings* das quatro simulações de dinâmica molecular.

Ligante	Sm1			Sm1-ANP			CD39			CD39-ANP			
	frame-0	frame-37	frame-55	frame-108	frame-171	frame-0	frame-149	frame-219	frame-0	frame-187	frame-219	frame-0	frame-80
ANP (<i>Redocking</i>) ^a	-	-	-	-	-	-12,03	-13,48	-16,96	-	-	-	-10,78	-10,64
ANP-Zinc ^a	-12,39	-12,39	-12,59	-12,02	-12,61	-12,54	-14,30	-18,01	-10,71	-10,67	-10,08	-11,79	-11,79
ANP-Zinc ^c	-8,13	-7,61	-7,50	-6,86	-9,12	-4,94	-4,94	-6,59	-4,67	-8,38	-6,45	-5,64	-6,48
ARL67 ^a	-11,81	-13,20	-13,77	-12,63	-12,47	-12,34	-13,82	-18,03	-10,41	-10,65	-11,29	-13,16	-12,58
ARL67 ^c	-8,12	-7,64	-7,13	-7,58	-9,51	-8,26	-5,44	-4,98	-4,02	-6,17	-4,37	-4,83	-6,16
Praziquantel ^b	-5,00	-5,15	-3,08	-4,27	-3,81	-4,11	-4,95	-7,03	-3,92	-4,39	-9,41	-7,00	-6,00
Praziquantel ^c	-5,64	-3,92	-3,25	-3,07	-4,96	-3,04	-3,79	-7,96	-2,55	-3,49	-2,38	-2,29	-3,88

^aSítio na presença de Ca^{+2} e H_2O . ^bSítio sem Ca^{+2} e H_2O . ^cCavidade-2. O *frame-0* refere-se a última conformação obtida ao final de 10ns da última equibração. Este *frame* é considerado como conformação de referência para análise de mudança conformacional.

4.7.1 Análises de docking na cavidade-1

4.7.1.1 Análise de *redocking* do ANP

Para a validação das conformações ao longo da dinâmica no sítio ativo, foram realizados estudos de *redocking* do ligante ANP. Em cada *frame* analisado das simulações Sm1-ANP e CD39-ANP, o ligante foi removido do sítio ativo e redocado através do programa Glide. Em seguida, foi calculado o RMSD entre a conformação original do ANP no sítio ativo da proteína no *frame* em questão contra as conformações do ANP redocado neste mesmo *frame*.

Em todos os *frames* analisados, os valores do *XPGScore* para o ANP apresentaram alta afinidade no sítio ativo das proteínas Sm1 e CD39 (tabela 4.9). Na figura 4.38 é apresentada a conformação do ANP redocado no sítio da Sm1 no *frame-0* e na figura 4.39 é mostrada a conformação do ANP redocado no sítio da CD39 no *frame-80*, ambas simuladas com o ANP no sítio ativo.

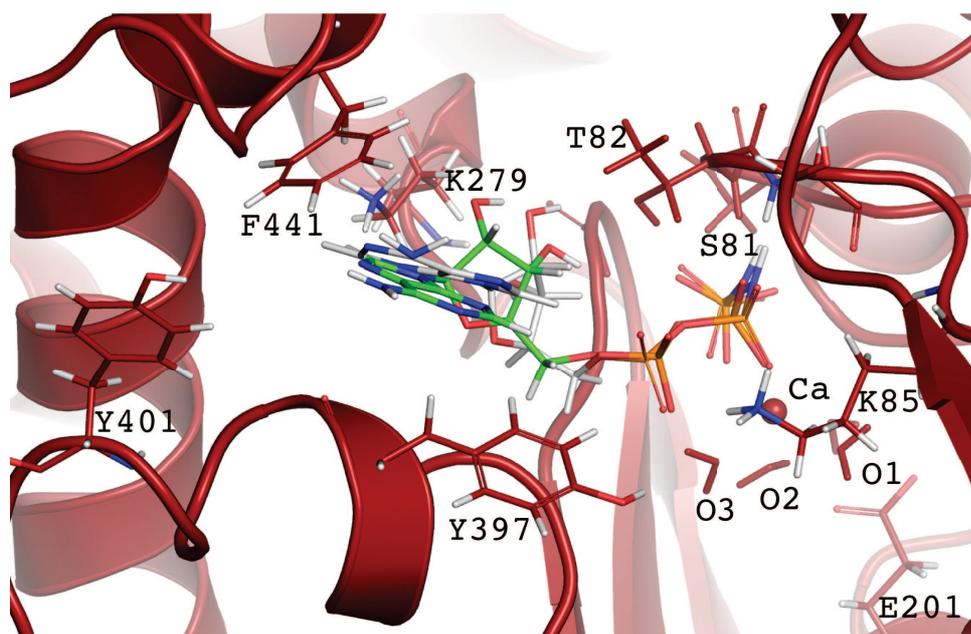


Figura 4.38: *Redocking* do ANP na cavidade-1 do *frame-0* da Sm1-ANP.

Em cor verde está representada a estrutura de melhor RMSD (2,12Å) e em cinza o ligante original.

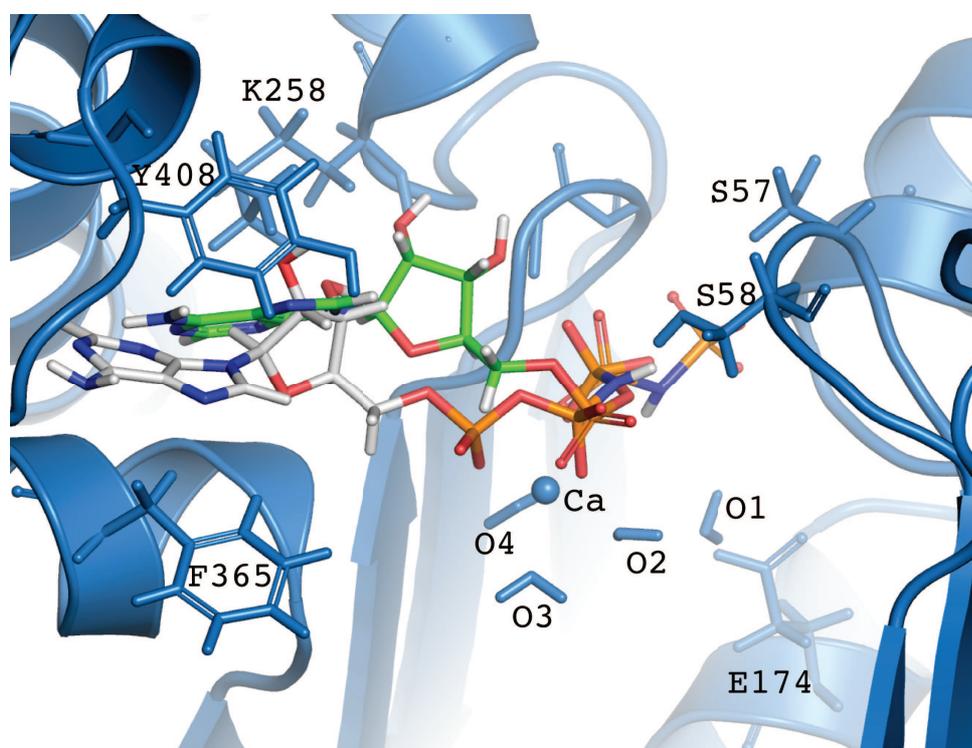


Figura 4.39: *Redocking* do ANP na cavidade-1 do *frame-80* da CD39-ANP. Em cor verde está representada a estrutura de melhor RMSD (3,69Å) e em cinza o ligante original.

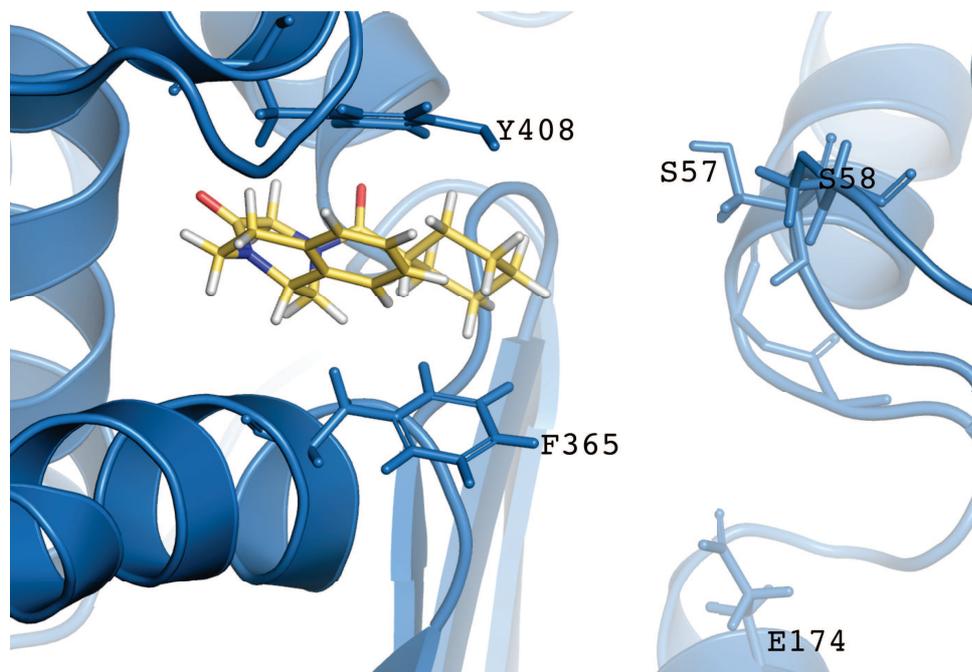
4.7.1.2 Análise dos *dockings* ao longo das simulações

Na tabela 4.9 é possível verificar que para o *frame-0*, os resultados de menor pontuação para os ligantes ARL67 e ANP são observados nos *dockings* provenientes das simulações onde o ligante ANP estava presente no sítio ativo da proteína. No *frame-0* o *docking* do ARL67 apresentou melhor pontuação ($XPGScore = -13,16$) na simulação da CD39-ANP. Nos estudos envolvendo o ARL67, os melhores resultados foram obtidos na presença de Ca^{+2} e H_2O no sítio-ativo, o que pode ser corroborado pelo trabalho de Vivian e colaboradores (2010), onde descreve o cristal da estrutura de uma NTPDase de *Legionella pneumophila* com a presença do ARL67 no sítio ativo da proteína. Segundo Vivian, as interações do ARL67 ocorrem mais mediadas por moléculas de água do que diretamente com os resíduos, como ocorre com o ANP (VIVIAN et al., 2010).

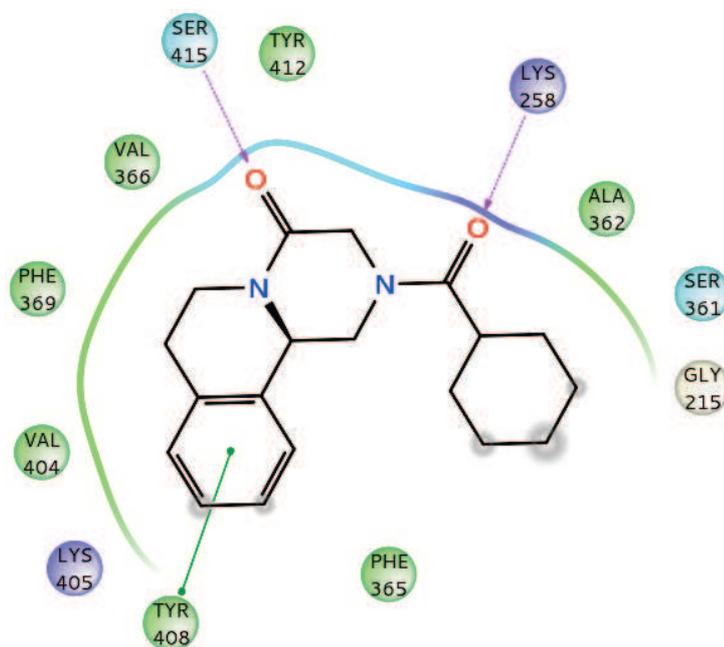
Com relação aos *dockings* do praziquantel no sítio ativo (cavidade-1), os melhores resultados foram obtidos na CD39-ANP no *frame-0* ($XPGScore = -7,00$) (figura 4.40) e na Sm1-ANP no *frame-219* ($XPGScore = -7,03$). Recentemente, Zebisch e colaboradores (2014) publicaram o cristal da estrutura do domínio ECD da NTPDase2 de *Rattus norvegicus* com a presença de sulfo-antraquinona na região do sítio ativo, onde estão localizados os resíduos que fazem *stacking* com a adenina. Podemos observar que a região de *docking* do praziquantel apresentada na Figura 4.40 é equivalente (F365 e Y408) a que Zebisch descreve para sulfo-antraquinona (Y350 e R394) (ZEBISCH et al., 2014). Assim como na CD39, o praziquantel na Sm1-ANP docou próximo a região dos resíduos Y397 e F441.

A figura 4.41 apresenta os *dockings* do ANP e ARL67 na cavidade-1 do *frame-219* da Sm1-ANP. Os valores da pontuação ($XPGScore$) foram os melhores depois dos observados no *frame-171*, ou seja, ANP com $XPGScore = -18,01$ e ARL67 $XPGScore = -18,03$, respectivamente.

Pelas figuras 4.41(b) e 4.41(c) pode-se ver que nos dois *dockings* o resíduo T82 aparece interagindo com o β -fosfato, como proposto pelo modelo de (KOZAKIEWICZ et al., 2008). Também é possível observar o resíduo S83 interagindo com o grupo α -fosfato do ANP e ARL67.

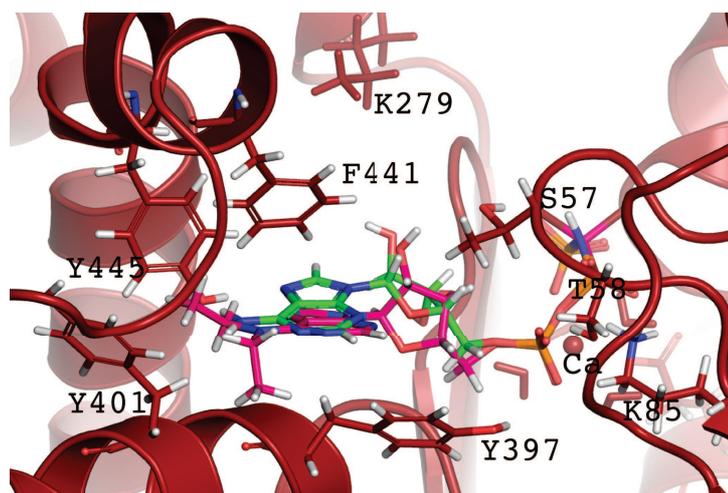


(a) Visão 3D do praziquantel na região da cavidade-1.

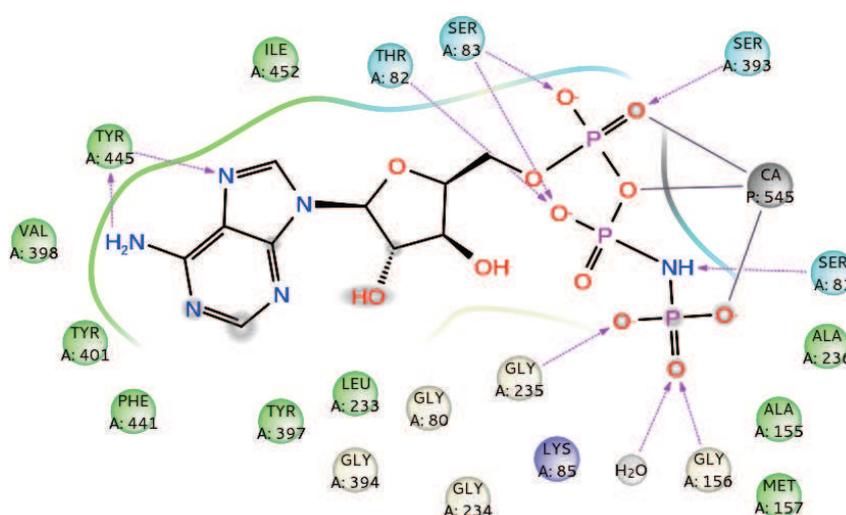


(b) Esquema 2D de interação proteína-ligante.

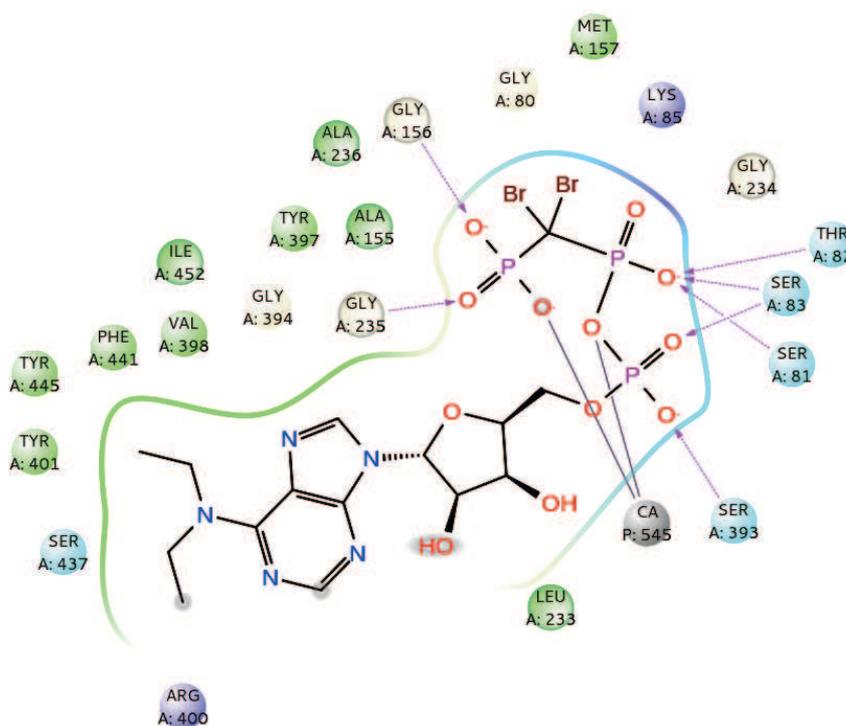
Figura 4.40: *Docking* do praziquantel na cavidade-1 do *frame-0* da CD39-ANP.



(a) Visão 3D do ANP (verde) e ARL67 (rosa) na cavidade-1.



(b) Esquema 2D de interação proteína-ANP.

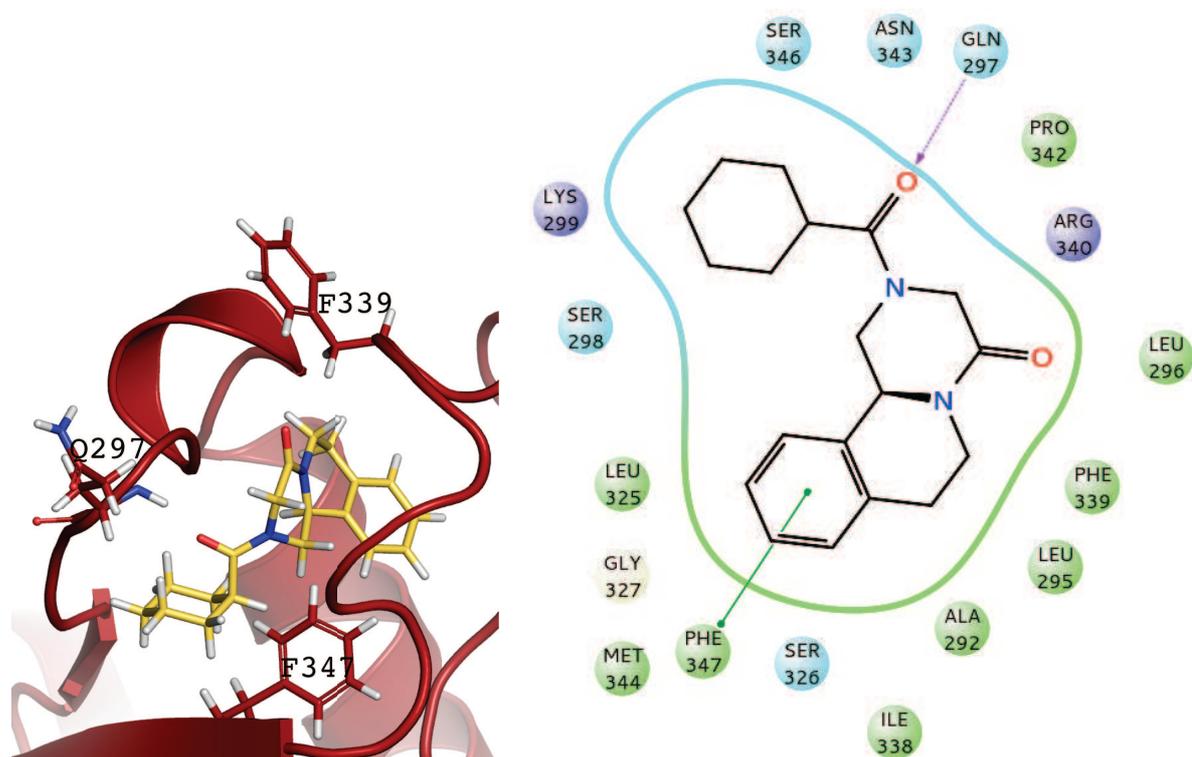


(c) Esquema 2D de interação proteína-ARL67.

Figura 4.41: Docking do ANP (verde) e ARL67 (rosa) na cavidade-1 do *frame-219* da Sm1-ANP.

4.7.2 Análises de docking na cavidade-2

Pela tabela 4.9 é possível ver que para o praziquantel o *docking* de melhor pontuação na cavidade-2 ocorreu no *frame-219* da Sm1-ANP. Na figura 4.42 pode-se observar que o *docking* do praziquantel ocorreu dentro da cavidade, por meio de interações com os resíduos Q297 e F347.



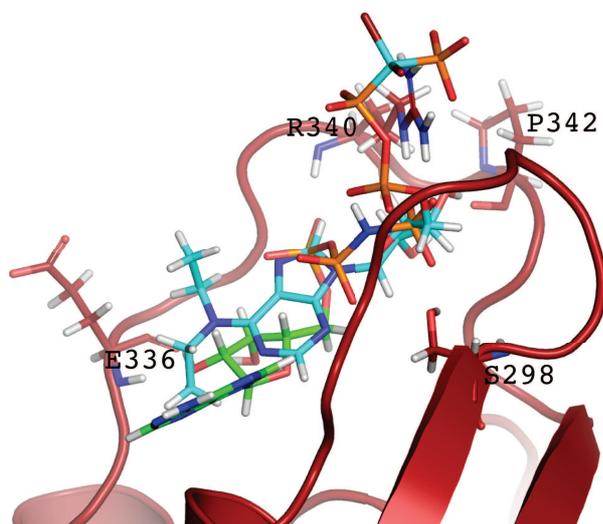
(a) Visão 3D do ANP e ARL67 na cavidade-2.

(b) Esquema 2D de interação proteína-ligante.

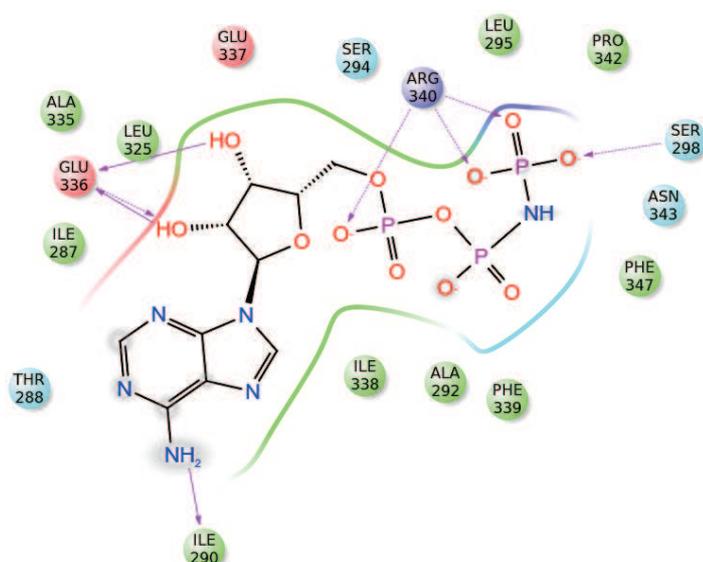
Figura 4.42: *Docking* do praziquantel na cavidade-2 do *frame-219* da Sm1-ANP.

Com relação aos ligantes ANP e ARL67, os *dockings* com as melhores pontuações foram obtidos no *frame-171* da Sm1. Porém, todos os *dockings* ocorreram fora da cavidade. Isso pode estar relacionado com a movimentação da região da cavidade durante a trajetória, como já discutido na seção RMSD dos ECD.

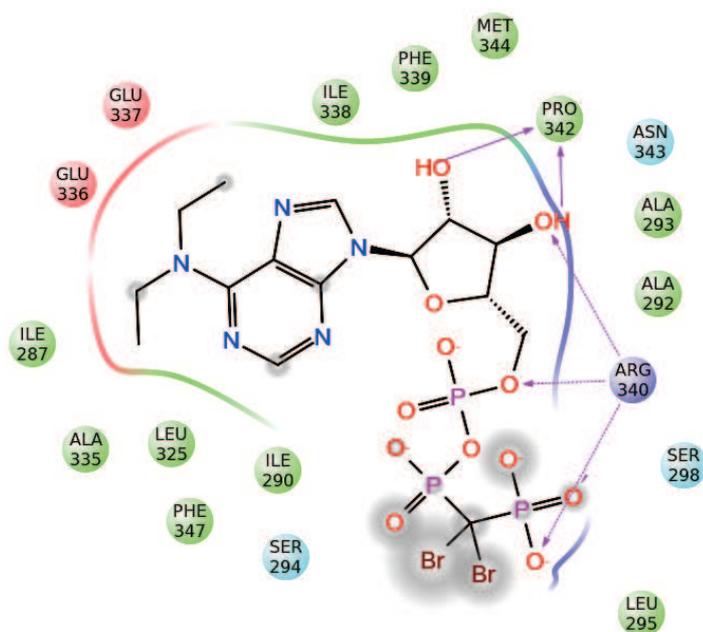
Na figura 4.43 são mostrados os *dockings* do ANP e ARL67 na cavidade-2 do *frame-0* da Sm1. Os valores da pontuação (*XPGScore*) foram os melhores depois dos observados no *frame-171*, ou seja, ANP com *XPGScore* = -8,13 e ARL67 *XPGScore* = -8,12, respectivamente.



(a) Visão 3D do ANP e ARL67 na cavidade-2.



(b) Esquema 2D de interação proteína-ANP.



(c) Esquema 2D de interação proteína-ARL67.

Figura 4.43: Docking do ANP (verde) e ARL67 (azul) na cavidade-2 do *frame-0* da Sm1.

Pela figura 4.43, pode-se observar que os grupos α e γ fosfatos de ambos os ligantes interagem com o resíduo R340. A posição da cadeia lateral de R340 posicionada mais para fora da cavidade, pode estar facilitando a interação com os grupos α e γ fosfatos. Em ambos os ligantes observa-se que a cauda de fosfatos apresentam-se expostas ao solvente, principalmente no ligante ARL67. No ligante ANP o anel da ribose está interagindo com o resíduo E336, enquanto que no ligante ARL67 o grupo ribose interage com P342 e R340.

4.7.3 Conclusões dos estudos de docking

Dos resultados dos estudos de *docking* é possível propor algumas conclusões referentes às cavidades 1 e 2. A respeito da cavidade-1, nesta foram obtidos os melhores valores de pontuação (*XPGScore*) para todos os ligantes, principalmente o ANP e ARL67. Como pode ser visto na tabela 4.9, na CD39 em todos os *dockings* o ANP foi o que apresentou os melhores valores de *XPGScore*, embora o maior valor de *XPGScore* encontrado tenha sido para o ARL67 no *frame-219* da Sm1-ANP. Porém, é possível ver que os valores do *XPGScore* para os *dockings* envolvendo o ANP e ARL67 foram próximos. Isso pode estar indicando que o ARL67 seja um ligante competitivo pelo sítio ativo para as enzimas Sm1 e CD39. Para o praziquantel, as melhores pontuações foram encontradas nos *dockings* onde o ligante encontra-se interagindo com os resíduos F365 e Y408, da CD39, ou Y397 e F441, da Sm1.

Sobre a cavidade-2, em nenhum dos *dockings* realizados com o ANP e ARL67 nesta cavidade o *XPGScore* foi melhor do que os obtidos na cavidade-1 para os mesmos ligantes. Somente o praziquantel em alguns *frames* da Sm1 (0, 55 e 171) e Sm1-ANP (219) os valores de *XPGScore* foram melhores do que os obtidos na cavidade-1. A melhor pontuação foi obtida com o ARL67 no *frame-171* da Sm1. Nenhum dos *dockings* realizados na cavidade-2 da CD39 e CD39-ANP, o ligante é encontrado dentro da cavidade. Esses resultados podem sugerir a importância da cavidade-2 na Sm1 como um sítio alternativo para o desenvolvimento de drogas mais seletivas.

5 CONCLUSÕES e PERSPECTIVAS

Esta tese teve como objetivo realizar um estudo comparativo por meio de modelagem comparativa, simulação de dinâmica molecular e *docking* receptor-ligante de duas NTPDase1, uma presente no parasita *Schistosoma mansoni* (Sm1), e outra presente em humanos (CD39). Também foi proposta do presente trabalho, apresentar um método de agrupamento das conformações provenientes de simulações de dinâmica molecular que não dependesse de um valor de corte definido pelo usuário para o RMSD.

Neste trabalho foi possível propor os modelos tridimensionais (3D) completos das enzimas Sm1 e CD39 por meio da técnica de modelagem comparativa usando um protocolo diferenciado para as porções transmembranares sem moldes. Para tal foram usados preditores de estrutura secundária, de domínios transmembranares e de contato entre hélices transmembranares. Os resultados retornados por esses preditores foram cruzados a fim de obter um consenso a respeito dos intervalos de resíduos que formam as duas TM. Essa informação foi fundamental na definição das restrições de estrutura secundária e de contato entre resíduos usadas no programa Modeller durante a geração dos modelos. Após a construção dos modelos, estes foram eficientemente validados e em seguida preparados para simulações de dinâmica molecular e estudos de *docking*.

Na predição de cavidades, para o modelo da Sm1 objetivou-se localizar regiões que além do sítio ativo também pudessem ser usadas em estudos de *docking* receptor-ligante contra possíveis inibidores de NTPDases. Nesta etapa do trabalho foi possível definir duas cavidades a serem estudadas: a cavidade-1, que corresponde a região do sítio ativo; e a cavidade-2, um sítio alternativo com boa taxa de drogabilidade. Estas mesmas cavidades foram mapeadas no modelo 3D da CD39 para fins de comparação.

Para as simulações de dinâmica molecular, foram feitas duas preparações para cada uma das enzimas. Uma preparação onde a enzima apresentava em seu sítio ativo o ligante ANP (análogo do ATP), e uma preparação sem o ligante. A intenção era verificar se a presença e ausência do ligante no sítio ativo resultaria na produção de diferentes conformações ao longo das duas trajetórias. Analisando as simulações envolvendo os modelos da Sm1 e CD39, bem como os moldes PDB3ZX3 e PDB3CJA, foi possível observar que a presença de cinco pontes dissulfeto no ECD da CD39 e dos moldes

possibilita uma maior estabilidade do domínio, em relação à Sm1 que apresenta três pontes de dissulfeto. Por outro lado, a presença do ligante ANP na cavidade-1 (sítio ativo) também proporciona uma maior estabilidade do domínio ECD, o que foi observado ao comparar a simulação da Sm1 sem ANP com as demais. Para a região da cavidade-2 na Sm1, foi possível observar uma maior mudança conformacional na ausência do ligante no sítio ativo, quando comparada à sua simulação na presença do ligante. Entretanto, o mesmo não foi observado na região da CD39 que corresponde a cavidade-2.

Com relação a análise de agrupamento das trajetórias, o presente trabalho abordou o problema fazendo uso da metodologia GLCM em conjunto com o algoritmo *K-means*. Algoritmos de agrupamento voltados para análise de trajetórias produzidas por simulações de dinâmica molecular, como o *g_cluster*, fazem uso de um *cutoff* para o RMSD e que é definido pelo usuário. Diferentes valores de *cutoff* resultam em diferentes números de agrupamento. Para contornar essa dependência de um *cutoff*, a metodologia proposta nesta tese fez uso do algoritmo PSO para otimizar o número de agrupamentos a serem formados pelo *K-Means*. O método proposto apresentou resultados satisfatórios, sendo possível identificar de forma automática e em apenas uma execução um conjunto de grupos conformacionais representativos para os estudos seguintes.

Sobre a análise de drogabilidade, foi possível observar que a cavidade-1 foi a que apresentou melhores resultados tanto nas simulações envolvendo a Sm1 quanto a CD39. Com relação à cavidade-2, os melhores resultados foram observados nas simulações envolvendo a Sm1 com e sem ligante ANP. Essa mesma cavidade não se mostrou promissora nas conformações da CD39 provenientes das simulações com e sem o ligante ANP.

Sobre os estudos de atracamento molecular, foram avaliadas a pontuação (*XPGScore*), o RMSD e a posição dos ligantes nas cavidades. Na cavidade-1, a posição dos ligantes foi analisada de acordo com modelos de atividade catalítica propostos na literatura para NTPDases. Os melhores resultados foram obtidos para o ARL67 e ANP na cavidade-1. Na cavidade-2, os resultados mais promissores foram obtidos na Sm1 o que, juntamente com os resultados de análise da dinâmica molecular, sugerem essa como sendo uma região alternativa para atracamento de possíveis inibidores da Sm1.

5.1 Perspectivas

Desta tese também foi possível propor linhas de trabalho que poderão ser desenvolvidas futuramente e que venham a estender os resultados obtidos até o fechamento do presente trabalho. Algumas dessas propostas seriam:

- Aumentar o tempo das simulações de dinâmica molecular envolvendo tanto os moldes PDB3ZX3 e PDB3CJA, bem como os modelos da Sm1 e CD39 com e sem ligante no sítio ativo, a fim de se verificar uma maior estabilização do sistema e mudanças conformacionais mais tardias.
- Aprofundar as análises sobre a seleção de conformações de dinâmica molecular usando a metodologia GLCM.
- Dar continuidade às simulações de *docking* receptor-ligante, aprofundando as análises já feitas e realizando novas simulações com novos ligantes.

REFERÊNCIAS

- ALLAN, D.; PAYARES, G.; EVANS, W. The phospholipid and fatty acid composition of schistosoma mansoni and of its purified tegumental membranes. *Molecular and Biochemical Parasitology*, v. 23, p. 123–128, 1983.
- ALTSCHUL, S. F. et al. Basic local alignment search tool. *Journal of Molecular Biology*, v. 215, p. 403–410, 1990.
- BAKER, D.; SALI, A. Protein structure prediction and structural genomics. *Science*, v. 294, p. 93–96, 2001.
- BAKER, N. A. et al. The adaptive multilevel finite element solution of the poisson - boltzmann equation on massively parallel computers. *IBM Journal of Research and Development*, v. 45, p. 427–438, 2001.
- BARALDI, A.; PARMIGGIANI, F. An investigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters. *IEEE Transactions on Geoscience and Remote Sensing*, v. 33, p. 293–304, 1995.
- BENZ, R. W. et al. Experimental validation of molecular dynamics simulations of lipid bilayers: A new approach. *Biophysical Journal*, v. 88, p. 805–817, 2005.
- BERENDSEN, H.; SPOEL, D. van der; DRUNEN, R. van. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, v. 91, p. 43–56, 1995.
- BERMAN, H. M. et al. The protein data bank. *Nucleic Acids Research*, v. 28, p. 235–242, 2000.
- BHARDWAJ, R.; SKELLY, P. J. Purinergic signaling and immune modulation at the schistosome surface? *Trends in parasitology*, v. 25, p. 256–260, 2009.
- BIASINI, M. et al. Swiss-model: modelling protein tertiary and quaternary structure using evolutionary information. *Nucleic Acids Research*, p. 1–6, 2014.
- BRASIL, M. M. da Saúde do. *Vigilância e Controle de Moluscos de Importância Epidemiológica*. 2ª edição. ed. Brasília - DF: Editora MS, 1995.
- BROOIJMANS, N.; KUNTZ, I. D. Molecular recognition and docking algorithms. *Annual Review of Biophysics and Biomolecular Structure*, v. 32, p. 335–373, 2003.
- CASE, D. A. et al. The amber biomolecular simulation programs. *Journal of Computational Chemistry*, v. 26, p. 1668–1688, 2005.
- CASTRO-BORGES, W. et al. Abundance of tegument surface proteins in the human blood fluke schistosoma mansoni determined by qconcat proteomics. *Journal of Proteomics*, v. 74, n. 9, p. 1519–1533, 2011.
- CHANGEUX, J.-P.; EDELSTEIN, S. Conformational selection or induced-fit? 50 years of debate resolved. *F1000 Biology Reports*, v. 19, p. 1–15, 2011.

- CHEN, S. et al. Recent advances in morphological cell image analysis. *Computational and Mathematical Methods in Medicine*, v. 2012, 2012.
- CHENG, T. et al. Structure-based virtual screening for drug discovery: a problem-centric review. *IEEE International Conference on Neural Networks*, v. 14, p. 133–141, 2012.
- COHN, M.; MEEK, G. A. The mechanism of hydrolysis of adenosine di- and tri-phosphate catalysed by potato apyrase. *Biochemical Journal*, v. 66, p. 128–130, 1957.
- COLE, C.; BARBER, J. D.; BARTON, G. J. The jpred 3 secondary structure prediction server. *Nucleic Acids Research*, v. 36, p. W197–W201, 2008.
- COLLEY, D. G.; SECOR, W. E. A schistosomiasis research agenda. *PLoS Neglected Tropical Diseases*, v. 1, p. e32, 2007.
- COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. *Protein Science*, v. 2, n. 9, p. 1511–1519, 1993.
- DA'DARA, A. A. et al. Schistosome tegumental ecto-apyrase (smatpdase1) degrades exogenous pro-inflammatory and pro-thrombotic nucleotides. *PeerJ*, v. 2, p. e316, 2014.
- DAMM, K. L.; CARLSON, H. A. Exploring experimental sources of multiple protein conformations in structure-based drug design. *Journal of the American Chemical Society*, v. 129, p. 8225–8235, 2007.
- DAVIS, I. W. et al. Molprobity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Research*, v. 35, p. W375–W383, 2007.
- DEMARCO, R. et al. Molecular characterization and immunolocalization of schistosoma mansoni atp-diphosphohydrolase. *Biochemical and Biophysical Research Communications*, v. 307, p. 831–838, 2003.
- DOENHOFF, M. J.; CIOLI, D.; UTZINGER, J. Praziquantel: mechanisms of action, resistance and new derivatives for schistosomiasis. *Current Opinion in Infectious Diseases*, v. 21, p. 659–667, 2008.
- DOLINSKY, T. J. et al. Pdb2pqr: an automated pipeline for the setup of poisson-boltzmann electrostatics calculations. *Nucleic Acids Research*, v. 32, p. 665–667, 2004.
- DROSOPOULOS, J. H. F. et al. Site-directed mutagenesis of human endothelial cell ecto-adpase soluble cd39: requirement of glutamate 174 and serine 218 for enzyme activity and inhibition of platelet recruitment. *Biochemistry*, v. 39, p. 6936–6943, 2000.
- EDHOLM, O.; NAGLE, J. F. Areas of molecules in membranes consisting of mixtures. *Biophysical Journal*, v. 89, p. 1827–1832, 2005.
- EISENBERG, D.; LÜTHY, R.; BOWIE, J. U. Verify3d: Assessment of protein models with three-dimensional profiles. *Methods in Enzymology*, v. 277, p. 396–404, 1997.
- ELEYAN, A.; DEMIREL, H. Co-occurrence matrix and its statistical features as a new approach for face recognition. *TUBITAK*, v. 19, p. 97–107, 2011.

- ENGELBRECHT, A. *Fundamentals of computational swarm intelligence*. Hoboken: John Wiley and Sons, 2005.
- ESWAR, N. et al. Comparative protein structure modeling using modeller. In: *Current Protocols in Bioinformatics*. [S.l.]: John Wiley and Sons, Inc., 2006. cap. 5.6, p. 1–30.
- FARIA-PINTO, P. *Mapeamento dos domínios antigênicos compartilhados entre a apirase de batata e as isoformas de ATP difosfohidrolase de Schistosoma mansoni*. Tese (Doutorado) — Centro de Pesquisas René Rachou/Fundação Oswaldo Cruz, 2008.
- FARIA-PINTO, P. et al. Atp diphosphohydrolase from schistosoma mansoni egg: characterization and immunocytochemical localization of a new antigen. *Parasitology*, v. 129, p. 51–57, 2004.
- FERREIRA, L. G.; OLIVA, G.; ANDRICOPULO, A. D. Target-based molecular modeling strategies for schistosomiasis drug discovery. *Future medicinal chemistry*, v. 7, n. 6, p. 753–764, 2015.
- FISCHER, E. Einfluss der configuration auf die wirkung der enzyme. *Berichte der Deutschen Chemischen Gesellschaft*, v. 27, p. 2985–2993, 1894.
- FRIESNER, R. et al. Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy. *Journal of Medicinal Chemistry*, v. 47, p. 1739–1749, 2004.
- FRIESNER, R. A. et al. Extra precision glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *Journal of Medicinal Chemistry*, v. 49, p. 6177–6196, 2006.
- GOLIATT, P. V. Z. C. *Técnicas de Bioinformática e Modelagem Computacional Aplicadas ao Estudo do Genoma de Trypanosoma cruzi e de Enzimas Consideradas de Interesse no Tratamento da Doença de Chagas: Estudo Particular das Cruzipainas 1 e 2*. Tese (Doutorado) — Laboratório Nacional de Computação Científica, 2007.
- GRINTHAL, A.; GUIDOTTI, G. Dynamic motions of cd39 transmembrane domains regulate and are regulated by the enzymatic active site. *Biochemistry*, v. 43, p. 13849–13858, 2004.
- GRINTHAL, A.; GUIDOTTI, G. Cd39, ntpdase 1, is attached to the plasma membrane by two transmembrane domains. why? *Purinergic Signalling*, v. 2, p. 391–398, 2006.
- GRYSEELS, B. et al. Human schistosomiasis. *Lancet*, v. 368, p. 1106–1118, 2006.
- GUEDES, I. A.; MAGALHÃES, C. S. de; DARDENNE, L. E. Receptor-ligand molecular docking. *Biophysical Reviews*, v. 6, p. 75–87, 2014.
- GUIXA-GONZALEZ, R. et al. Membplugin: studying membrane complexity in vmd. *Bioinformatics*, v. 30, p. 1478–1480, 2014.
- HAJDUK, P. J.; HUTH, J. R.; TSE, C. Predicting protein druggability. *Drug Discovery Today*, v. 10, p. 1675–1682, 2005.
- HARTIGAN, J. A.; WONG, M. A. Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society*, v. 28, p. 100–108, 1979.

- HOFMANN, K.; STOFFEL, W. Tmbase-a database of membrane-spanning protein segments. *Biological chemistry Hoppe-Seyler*, v. 374, 6, 1993.
- HUANG, S.-Y.; GRINTER, S. Z.; ZOU, X. Scoring functions and their evaluation methods for protein-ligand docking: recent advances and future directions. *Physical Chemistry Chemical Physics*, v. 12, p. 12899–12908, 2010.
- HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Vmd: Visual molecular dynamics. *Journal of Molecular Graphics*, v. 14, p. 33–38, 1996.
- IRWIN, J. J.; SHOICHET, B. K. Zinc - a free database of commercially available compounds for virtual screening. *Journal of Chemical Information and Modeling*, v. 45, p. 177–182, 2005.
- JOHN, B.; SALI, A. Comparative protein structure modeling by iterative alignment, model building and model assessment. *Nucleic Acids Research*, v. 31, p. 3982–3992, 2003.
- JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics*, v. 79, p. 926–935, 1983.
- JORGENSEN, W. L.; MAXWELL, D. S.; TIRADO-RIVES, J. Development and testing of the opls all-atom force field on conformational energetics and properties of organic liquids. *Journal of the American Chemical Society*, v. 118, p. 11225–11236, 1996.
- JR., W. F. de A.; DIAS, R. Computational methods for calculation of ligand-binding affinity. *Current Drug Targets*, v. 9, p. 1031–1039, 2008.
- KACZMAREK, E. et al. Identification and characterization of cd39 vascular atp diphosphohydrolase. *The Journal of Biological Chemistry*, v. 271, n. 51, p. 33116–33122, 1996.
- KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, v. 9, p. 646–652, 2002.
- KENNEDY, J.; EBERHART, R. Particle swarm optimization. *IEEE International Conference on Neural Networks*, v. 20, p. 1942–1948, 1995.
- KING, C. H. et al. Utility of repeated praziquantel dosing in the treatment of schistosomiasis in high-risk communities in africa: A systematic review. *PLoS Neglected Tropical Diseases*, v. 5, n. 9, p. e1321, 2011.
- KNOWLES, A. F. The gda1-cd39 superfamily: Ntpdases with diverse functions. *Purinergic Signalling*, v. 7, n. 1, p. 21–45, 2011.
- KORB, O. et al. Potential and limitations of ensemble docking. *Journal of Chemical Information and Modeling*, v. 52, p. 1262–1274, 2012.
- KOZAKIEWICZ, A. et al. Modeling studies of potato nucleoside triphosphate diphosphohydrolase ntpdase1: an insight into the catalytic mechanism. *Acta Biochimica Polonica*, v. 55, p. 141–150, 2008.
- KRASOWSKI, A. et al. Drugpred: a structure-based approach to predict protein druggability developed using an extensive nonredundant data set. *Journal of Chemical Information and Modeling*, v. 11, p. 2829–2842, 2011.

- KROGH, A. et al. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes. *Journal of Molecular Biology*, v. 305, p. 567–580, 2001.
- KUCERKA, N. et al. Structural significance of lipid diversity as studied by small angle neutron and x-ray scattering. *Membranes*, v. 5, p. 545–472, 2015.
- KUCERKA, N.; TRISTRAM-NAGLE, S.; NAGLE, J. F. Structure of fully hydrated fluid phase lipid bilayers with monounsaturated chains. *Journal of Membrane Biology*, v. 208, p. 193–202, 2006.
- KUSEL, J. R.; AL-ADHAMI, B. H.; DOENHOFF, M. J. The schistosome in the mammalian host : understanding the mechanisms of adaptation. *Parasitology*, v. 134, p. 1477–1526, 2007.
- KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. Advantages of combined transmembrane topology and signal peptide prediction-the phobius web server. *Nucleic Acids Research*, v. 35, p. W429–W432, 2007.
- LASKOWSKI, R. A.; CHISTYAKOV, V. V.; THORNTON, J. M. Pdbsum more: new summaries and analyses of the known 3d structures of proteins and nucleic acids. *Nucleic Acids Research*, v. 33, p. 266–268, 2005.
- LASKOWSKI, R. A. et al. Procheck: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*, v. 26, p. 283–291, 1983.
- LEAL, D. A. G. et al. Acute schistosomiasis in brazilian traveler: The importance of tourism in the epidemiology of neglected parasitic diseases. *Case Reports in Infectious Diseases*, v. 2012, p. 650929, 2012.
- LEIMKUHNER, B.; SKELL, R. symplectic numerical integrators in constrained hamiltonian systems. *Journal of Computational Physics*, v. 112, p. 117–125, 1994.
- LEVANO-GARCIA, J. et al. Characterization of schistosoma mansoni atpdase2 gene, a novel apyrase family member. *Biochemical and Biophysical Research Communications*, v. 352, p. 384–389, 2007.
- LEVANO-GARCIA, J. C. *Caracterização bioquímica e imunológica das enzimas recombinantes ATP-difosfolidrolases 1 e 2 do parasita Schistosoma mansoni*. Tese (Doutorado) — Instituto de Química/Universidade de São Paulo, 2008.
- LINDORFF-LARSEN, K.; FERKINGHOFF-BORG, J. Similarity measures for protein ensembles. *PLoSOne*, v. 4, p. e4203, 2009.
- LO, A. et al. Predicting helix-helix interactions from residue contacts in membrane proteins. *Bioinformatics*, v. 25, p. 996–1003, 2009.
- LU, D.; BATISTELLA, M. Exploring tm image texture and its relationships with biomass estimation in rondônia, brazilian amazon. *Acta Amazonica*, v. 35, p. 249–257, 2005.
- LÉVESQUE, S. A. et al. Specificity of the ecto-atpase inhibitor arl 67156 on human and mouse ectonucleotidases. *British Journal of Pharmacology*, v. 152, p. 141–150, 2007.

- MACKAY, D. H. J.; CROSS, A. J.; HAGLER, A. T. The role of energy minimization in simulation strategies of biomolecular systems. In: FASMAN, G. D. (Ed.). *Prediction of Protein Structure and the Principles of Protein Conformation*. New York, USA: Plenum Press, 1989. p. 317–358.
- MACKERELL, J. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry*, v. 102, p. 3586–3616, 1998.
- MAHONEY, M. W.; JORGENSEN, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *Journal of Chemical Physics*, v. 112, p. 8910–8922, 2000.
- MARTÍNEZ, L. *Simulações de dinâmica molecular dos receptores do hormônio tireoideano*. Tese (Doutorado) — Instituto de Química/Universidade Estadual de Campinas, 2007.
- MCGUFFIN, L. J.; BRYSON, K.; JONES, D. T. The psipred protein structure prediction server. *Bioinformatics*, v. 16, p. 404–405, 2000.
- MCMORRAN, B. J. et al. Platelets kill intraerythrocytic malarial parasites and mediate survival to infection. *Science*, v. 323, p. 797–800, 2009.
- MEILER, J.; BAKER, D. Coupled prediction of protein secondary and tertiary structure. *Proceedings of the National Academy of Sciences*, v. 100, p. 12105–12110, 2000.
- MEIRELES, L. et al. Pre-existing soft modes of motion uniquely defined by native contact topology facilitate ligand binding to proteins. *Protein Science*, v. 20, p. 1645–1658, 2011.
- NAEGLE, K. M. et al. Robust co-regulation of tyrosine phosphorylation sites on proteins reveals novel protein interactions. *Molecular BioSystems*, v. 8, p. 2771–2782, 2012.
- NANNI, L. et al. Different approaches for extracting information from the co-occurrence matrix. *PLoSOne*, v. 8, p. e83554, 2013.
- NASCIMENTO, J. P. R. do. *Análise e Classificação de Imagens Baseadas em Características de Textura Utilizando Matrizes de Co-Ocorrência*. Tese (Doutorado) — Universidade Federal do Paraná, 2003.
- NELSON, D. L.; COX, M. M. *Lehninger: principles of biochemistry*. 5. ed. New York - USA: W. H. Freeman and Company, 2009.
- NICHOLS, S. et al. Predictive power of molecular dynamics receptor structures in virtual screening. *Journal of Chemical Information and Modeling*, v. 51, p. 1439–1446, 2011.
- NOBIS, M. et al. Advertisement intravital flim-fret imaging reveals dasatinib-induced spatial control of src in pancreatic cancer. *Computational and Mathematical Methods in Medicine*, v. 73, 2013.
- NOVOA, E. M. et al. Ensemble docking from homology models. *Journal of Chemical Theory and Computation*, v. 6, p. 2547–2557, 2010.
- NYHOLM, T. K. M.; ÖZDIREKCAN, S.; KILLIAN, J. A. How protein transmembrane segments sense the lipid environment. *Biochemistry*, v. 46, p. 1457–1465, 2006.

- OLSSON, M. H. et al. Propka3: Consistent treatment of internal and surface residues in empirical pka prediction. *Journal of Chemical Theory and Computation*, v. 7, p. 525–537, 2011.
- PARK, P. S.; KUFAREVA, I.; ABAGYAN, R. Improved docking, screening and selectivity prediction for small molecule nuclear receptor modulators using conformational ensembles. *Journal of Computer-Aided Molecular Design*, v. 24, p. 459–471, 2010.
- PEARCE, E. J.; MACDONALD, A. S. The immunobiology of schistosomiasis. *Nature Reviews Immunology*, v. 2, p. 499–511, 2002.
- PESSIN, G.; OSORIO, F. Otimização por enxame de partículas aplicado à formação e atuação de grupos robóticos. *Interdisciplinary Studies in Computer Science*, v. 20, p. 94–106, 2009.
- PETERSEN, T. N. et al. Signalp 4.0: discriminating signal peptides from transmembrane regions. *Nature Methods*, v. 8, p. 785–786, 2011.
- PHILLIPS, J. C. et al. Scalable molecular dynamics with namd. *Journal of Computational Chemistry*, v. 26, p. 1817–1802, 2005.
- PONDER, J. W.; CASE, D. A. Force fields for protein simulations. *Advances in Protein Chemistry*, v. 66, p. 27–85, 2003.
- RAMAMURTHY, B.; CHANDRAN, K. Content based medical image retrieval with texture content using gray level co-occurrence matrix and k-means clustering algorithms. *Journal of Computer Science*, v. 8, p. 1070–1076, 2012.
- RICK, S. W.; STUART, S. J.; BERNE, B. Dynamical fluctuating charge force fields: Application to liquid water. *Journal of Chemical Physics*, v. 101, p. 6141–6156, 1994.
- ROBSON, S. C.; SEVIGNY, J.; ZIMMERMANN, H. The e-ntpdase family of ectonucleotidases: Structure function relationships and pathophysiological significance. *Purinergic Signalling*, v. 2, p. 409–430, 2006.
- ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, p. 53–65, 1987.
- SALI, A. Comparative protein modeling by satisfaction of spatial restraints. *Molecular Medicine Today*, v. 1, p. 270–277, 1995.
- SANCHEZ, R. et al. Protein structure modeling for structural genomics. *Nature Structural Biology*, v. 7, p. 986–990, 2000.
- SANSOM, F. M. The role of the ntpdase enzyme family in parasites: what do we know, and where to from here? *Parasitology*, v. 193, n. 8, p. 963–980, 2012.
- SANSOM, F. M. et al. A bacterial ecto-triphosphate diphosphohydrolase similar to human cd39 is essential for intracellular multiplication of legionella pneumophila. *Cellular Microbiology*, v. 9, n. 8, p. 1922–1935, 2007.

- SANSOM, F. M. et al. Enzymatic properties of an ecto-nucleoside triphosphate diphosphohydrolase from legionella pneumophila. *Journal of Biological Chemistry*, v. 283, p. 12909–12918, 2008.
- SCHLICK, T. *Molecular Modeling and Simulation*. 2. ed. New York - USA: Springer, 2010.
- SIEVERS, F. et al. Fast, scalable generation of high-quality protein multiple sequence alignments using clustal omega. *Molecular Systems Biology*, v. 7, p. 539, 2011.
- SILVA, M. A. T. B. *Uso de matrizes de co-ocorrências para classificação automática de imagens de cana-de-açúcar*. Tese (Doutorado) — Departamento de Ciência da Computação/Universidade Federal de Lavras, 2013.
- SOUZA, V. C. de et al. Structural comparative analysis of secreted ntpdase models of schistosoma mansoni and homo sapiens. *Lecture Notes in Computer Science*, v. 8826, p. 91–98, 2014.
- SPERANDIO, O. et al. How to choose relevant multiple receptor conformations for virtual screening: a test case of cdk2 and normal mode analysis. *European Biophysics Journal*, v. 39, p. 1365–1372, 2010.
- STRANDBERG, E.; KILLIAN, J. A. Snorkeling of lysine side chains in transmembrane helices: how easy can it get? *FEBS Letters*, v. 544, p. 69–73, 2003.
- TAMURA, K. et al. Mega6: Molecular evolutionary genetics analysis version 6.0. *Molecular Biology and Evolution*, v. 30, p. 2725–2729, 2013.
- ULMSCHNEIDER, M. B.; SANSOM, M. S. Amino acid distributions in integral membrane protein structures. *Biochimica et Biophysica Acta - Biomembranes*, v. 1512, p. 1–14, 2001.
- VASCONCELOS, E. G. et al. Partial purification and immunohistochemical localization of atp diphosphohydrolase from schistosoma mansoni. *Journal of Biological Chemistry*, v. 271, p. 22139–22145, 1996.
- VASCONCELOS, E. G. et al. Characterization and localization of an atp-diphosphohydrolase on the external surface of the tegument of schistosoma mansoni. *Molecular and Biochemical Parasitology*, v. 58, p. 205–214, 1993.
- VERLI, H. *Bioinformática da Biologia à Flexibilidade Molecular*. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular - SBBq, 2014.
- VIVIAN, J. P. et al. Crystal structure of a legionella pneumophila ecto-triphosphate diphosphohydrolase, a structural and functional homolog of the eukaryotic ntpdases. *Structure*, v. 18, p. 228–238, 2010.
- VOLKAMER, A. et al. Analyzing the topology of active sites: On the prediction of pockets and subpockets. *Journal of Chemical Information and Modeling*, v. 50, p. 2041–2052, 2010.
- VOLKAMER, A. et al. Combining global and local measures for structure-based druggability predictions. *Journal of Chemical Information and Modeling*, v. 52, p. 360–372, 2012.

VOLKAMER, A. et al. Dogsitescorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics*, v. 28, p. 2074–2075, 2012.

WANG, X.-F. et al. Predicting residue-residue contacts and helix-helix interactions in transmembrane proteins using an integrative feature-based random forest approach. *PloS One*, v. 6, p. e26767, 2011.

WIEDERSTEIN, M.; SIPPL, M. J. Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Research*, v. 35, p. W407–W410, 2007.

XIONG, J. *Essential Bioinformatics*. 1. ed. Cambridge, UK: Cambridge University Press, 2006.

ZEBISCH, M. et al. Crystal structure of ntpdase2 in complex with the sulfoanthraquinone inhibitor psb-071. *Journal of Structural Biology*, v. 185, p. 336–341, 2014.

ZEBISCH, M. et al. Crystallographic evidence for a domain motion in rat nucleoside triphosphate diphosphohydrolase (ntpdase) 1. *Journal of Molecular Biology*, v. 415, p. 288–306, 2012.

ZEBISCH, M. et al. Crystallographic snapshots along the reaction pathway of nucleoside triphosphate diphosphohydrolases. *Structure*, v. 21, n. 8, p. 1460–1475, 2013.

ZEBISCH, M.; STRÄTER, N. Structural insight into signal conversion and inactivation by ntpdase2 in purinergic signaling. *Proceedings of the National Academy of Sciences of the United States of America*, v. 105, p. 6882–6887, 2008.

ZIMMERMANN, H.; ZEBISCH, M.; STRÄTER, N. Cellular function and molecular structure of ecto-nucleotidases. *Purinergic Signalling*, v. 8, n. 3, p. 437–502, 2012.

ZOETE, V. et al. Swissparam: A fast force field generation tool for small organic molecules. *Journal of Computational Chemistry*, v. 32, p. 2359–2368, 2011.

6 ANEXO

Gráficos do Errat (COLOVOS; YEATES, 1993), e do Z-Score gerados pelo programa ProSA-Web (WIEDERSTEIN; SIPPL, 2007).

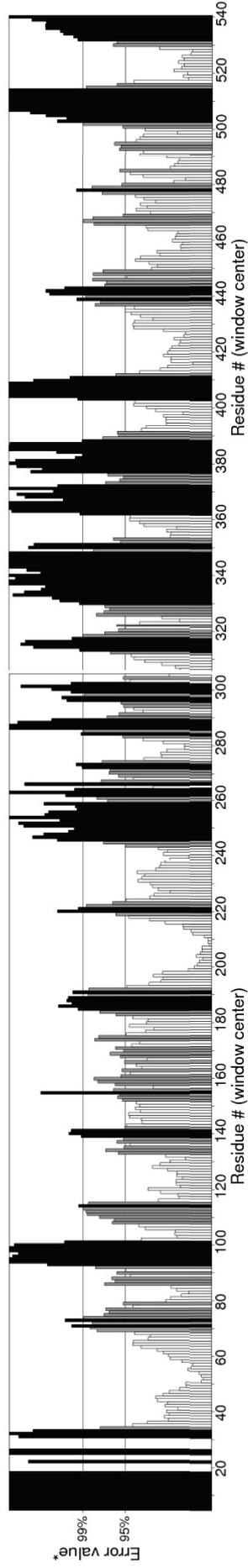


Figura 6.1: Gráfico do Errat para o modelo da Sm1.

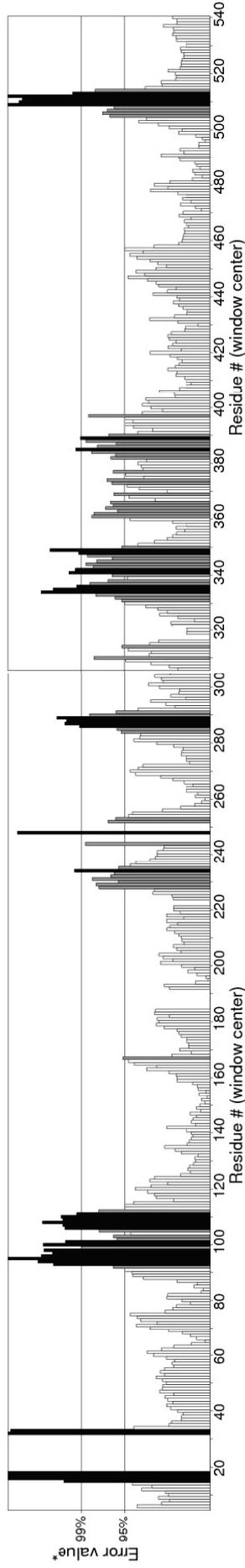


Figura 6.2: Gráfico do Errat para a Sm1-ANP ao final da minimização.

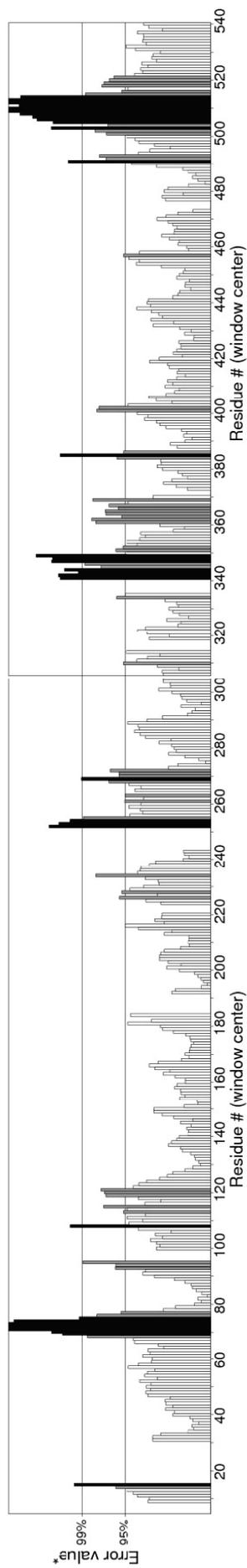


Figura 6.3: Gráfico do Errat para a Sm1-ANP ao final do aquecimento.

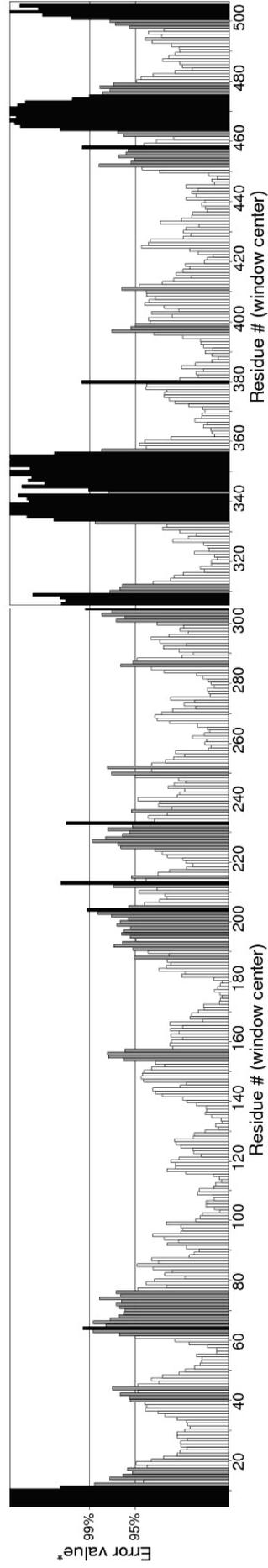


Figura 6.4: Gráfico do Errat do modelo para a CD39.

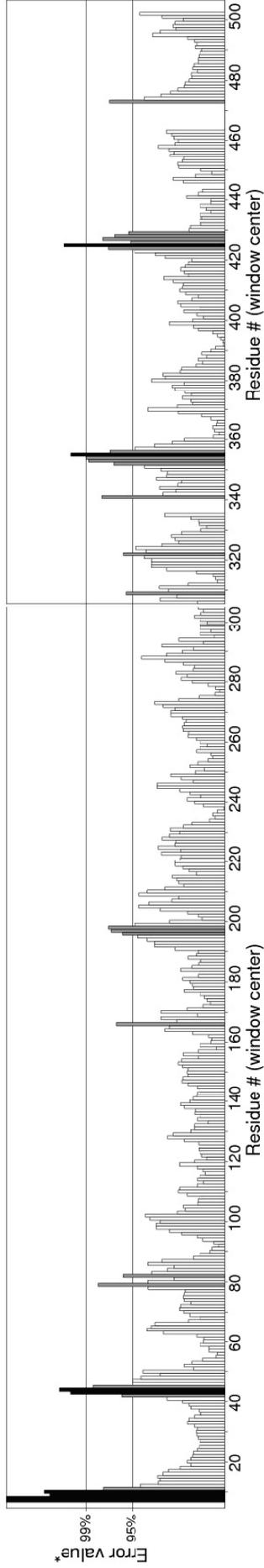


Figura 6.5: Gráfico do Errat para a CD39-ANP ao final da minimização.

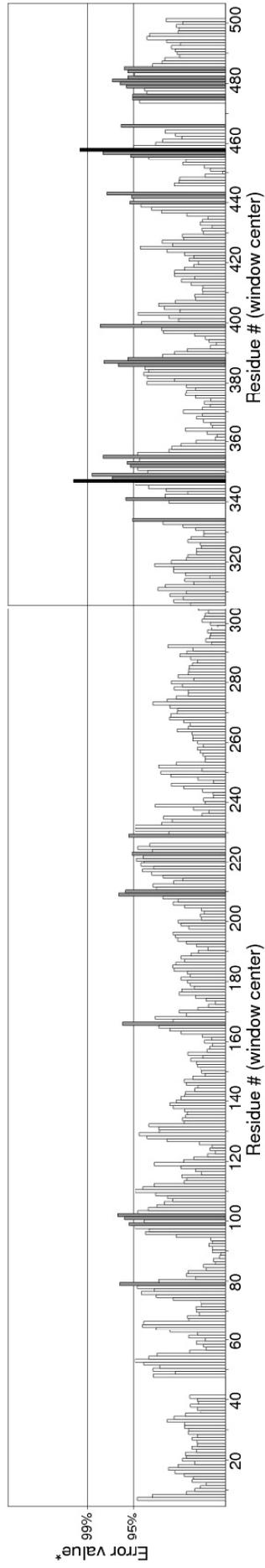


Figura 6.6: Gráfico do Errat para a CD39-ANP ao final do aquecimento.

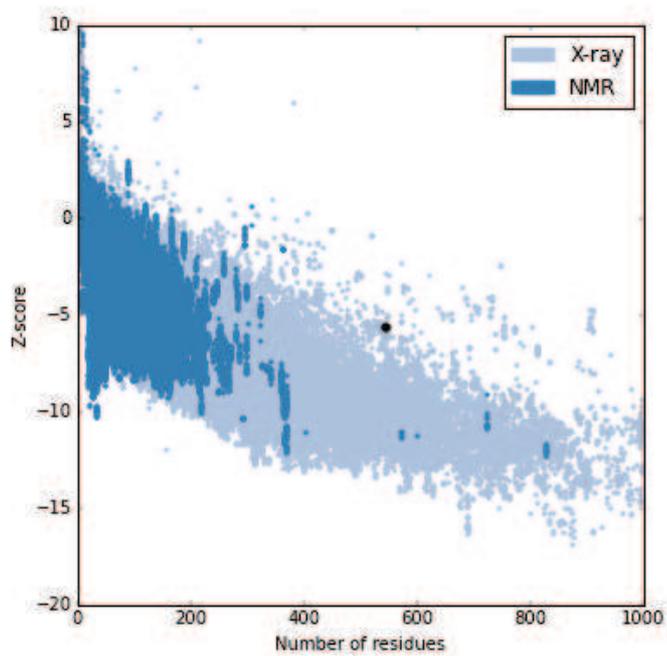


Figura 6.7: Gráfico do Z-score para o modelo da Sm1.

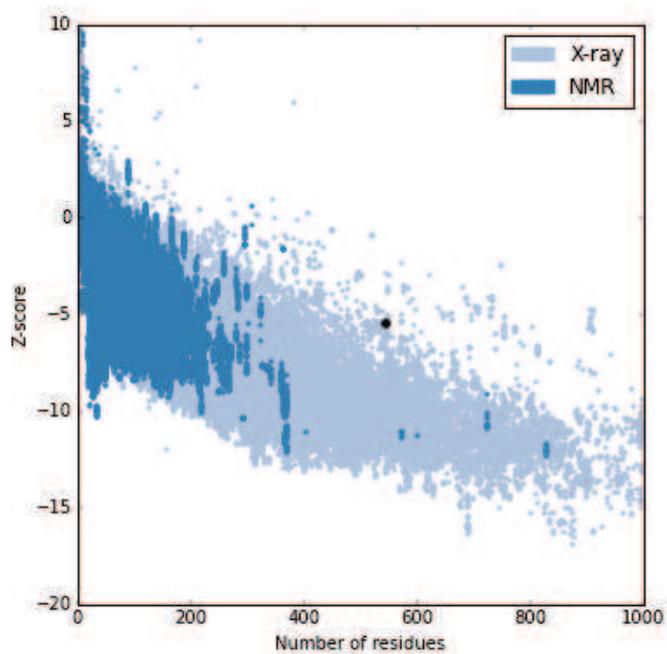


Figura 6.8: Gráfico do Z-score para a Sm1-ANP ao final da minimização.

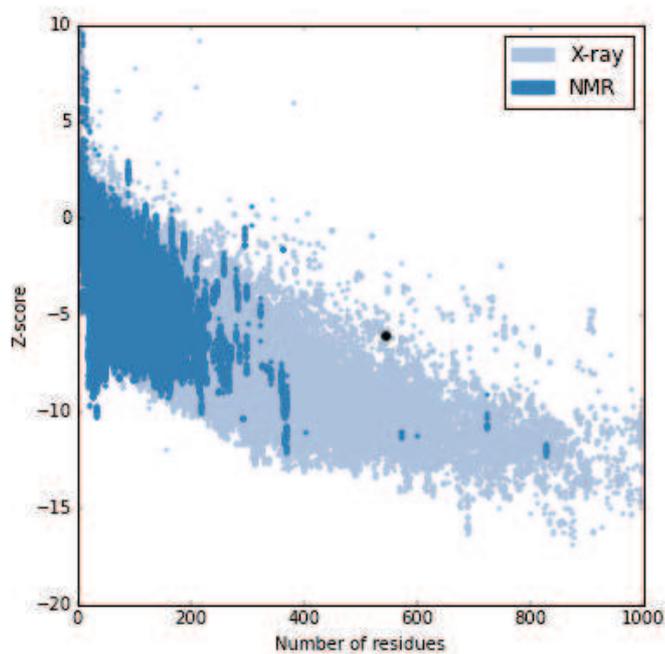


Figura 6.9: Gráfico do Z-score para a Sm1-ANP ao final do aquecimento.

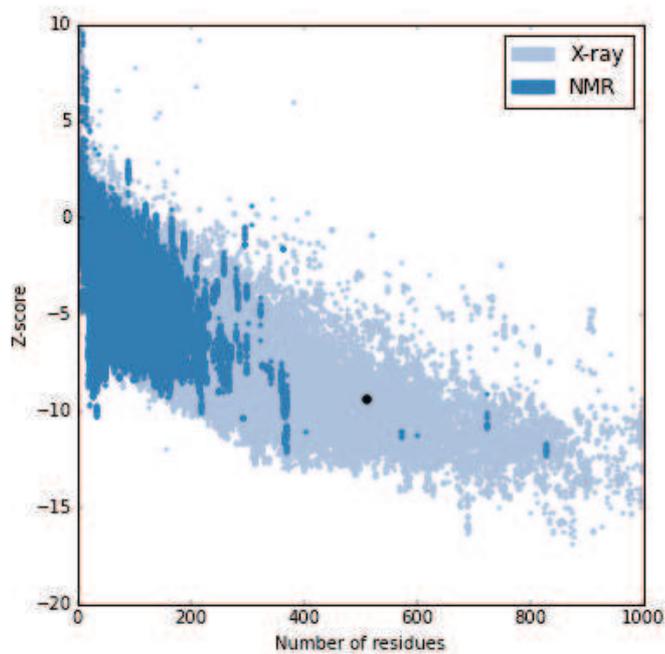


Figura 6.10: Gráfico do Z-score do modelo para a CD39.

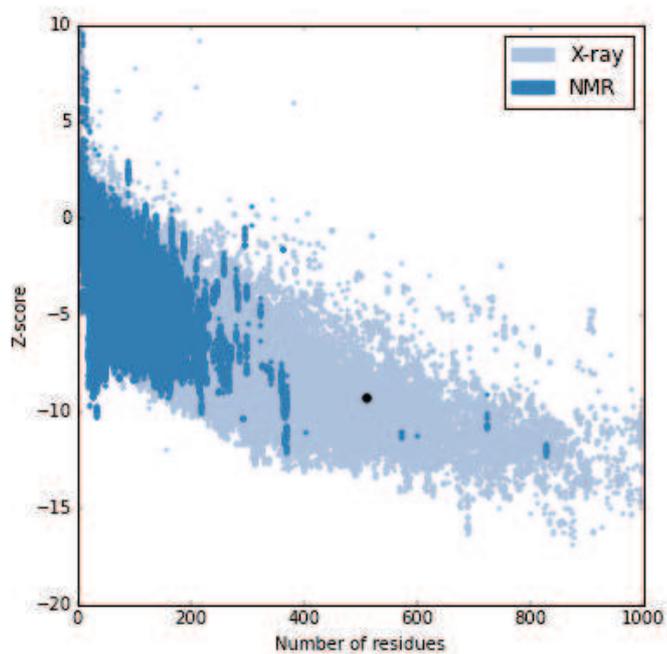


Figura 6.11: Gráfico do Z-score para a CD39-ANP ao final da minimização.

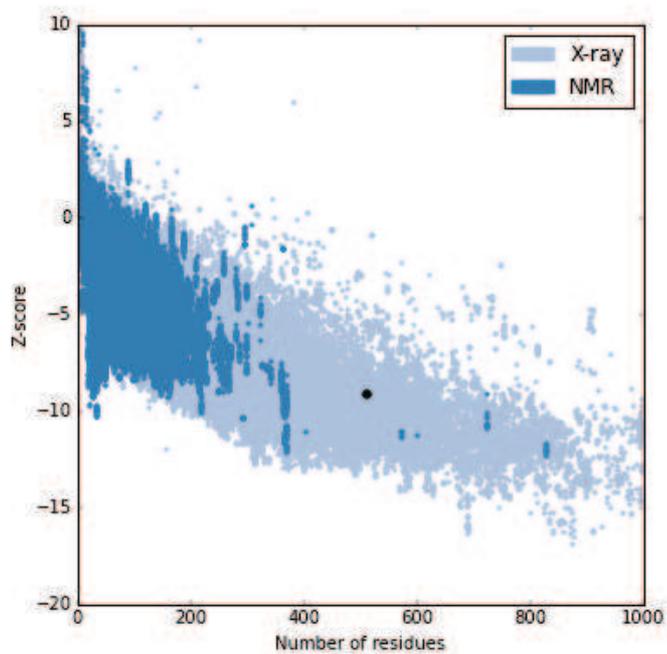


Figura 6.12: Gráfico do Z-score para a CD39-ANP ao final do aquecimento.