

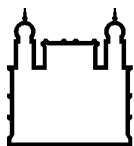
MINISTÉRIO DA SAÚDE  
FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação Biologia Computacional e Sistemas

ESTUDO DE METALOPROTEASES DE *Leishmania (Viannia) braziliensis*  
PERTENCENTES AO CROMOSSOMO 10 DA FAMÍLIA M8

AMANDA SUTTER DE OLIVEIRA HAMMES

Rio de Janeiro  
Março de 2017



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

*AMANDA SUTTER DE OLIVEIRA HAMMES*

Estudo de metaloproteases de *Leishmania (Viannia) braziliensis* pertencentes ao cromossomo 10 da família M8

Tese apresentada ao Instituto Oswaldo Cruz como parte dos requisitos para obtenção do título de Doutor em Biologia Computacional e Sistemas

**Orientador (es):** Prof. Dr. Ernesto Raúl Caffarena  
Prof. Dr. Carlos Roberto Alves

**RIO DE JANEIRO**  
**Março de 2017**

Hammes, Amanda Sutter de Oliveira.

ESTUDO DE METALOPROTEASES DE *Leishmania (Viannia) braziliensis* PERTENCENTES AO CROMOSSOMO 10 DA FAMÍLIA M8 / Amanda Sutter de Oliveira Hammes. - Rio de Janeiro, 2017.  
158 f.; il.

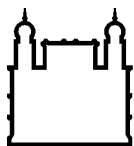
Tese (Doutorado) - Instituto Oswaldo Cruz, Pós-Graduação em Biologia Computacional e Sistemas, 2017.

Orientador: Ernesto Raul Caffarena.  
Co-orientador: Carlos Roberto Alves.

Bibliografia: f. 99-109

1. *Leishmania (Viannia) braziliensis*. 2. leishmanolisina. 3. modelagem comparativa. 4. dinâmica molecular. 5. docking. I. Título.

Elaborada pelo Sistema de Geração Automática de Ficha Catalográfica da Biblioteca de Manguinhos/ICICT com os dados fornecidos pelo(a) autor(a).



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## **INSTITUTO OSWALDO CRUZ**

**Programa de Pós-Graduação em Biologia Computacional e Sistemas**

**AUTOR: AMANDA SUTTER DE OLIVEIRA HAMMES**

**ESTUDO DE METALOPROTEASES DE *Leishmania (Viannia) braziliensis*  
PERTENCENTES AO CROMOSSOMO 10 DA FAMÍLIA M8**

**ORIENTADORES: Prof. Dr. Ernesto Raúl Caffarena  
Prof. Dr. Carlos Roberto Alves**

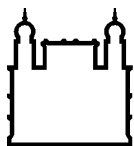
**Aprovada em: \_\_\_\_/\_\_\_\_/\_\_\_\_**

**EXAMINADORES:**

**Prof. Dra. Ana Carolina Ramos Guimarães – Presidente e revisor (Fiocruz/RJ)  
Prof. Dra. Rafaela Salgado Ferreira (UFMG/MG)  
Prof. Dra. Patricia Cuervo Escobar (Fiocruz/RJ)  
Prof. Dr. Laurent Emmanuel Dardenne – Suplente (LNCC/RJ)  
Prof. Dr. Floriano Paes Silva Junior – Suplente (Fiocruz/RJ)**

Rio de Janeiro, 07 de março de 2017





Ministério da Saúde

**FIOCRUZ**

**Fundação Oswaldo Cruz**

**Anexar a cópia da Ata que será entregue pela SEAC já assinada.**

## Agradecimentos

Primeiramente gostaria de agradecer ao Ernesto por todos esses anos que trabalhamos juntos, pela exemplar orientação, pelos conselhos, por nossa amizade. Foi muito bom tê-lo como orientador porque é um orientador presente, sabe ouvir o aluno e muito do que sou hoje, devo a ele. Muito obrigada de coração, por tudo! Tenho certeza que construímos uma amizade que não se encerra aqui.

Em segundo lugar, mas não menos importante, gostaria de agradecer ao Carlinhos, meu coorientador. Obrigada pelas sugestões, por diversas vezes sentarmos para entender os resultados e revisar o texto.

À Mariana Silva-Almeida por me ceder as sequências das metaloproteases, na qual trabalhei nesta tese e ao Franklin Silva pelas sugestões dadas durante as reuniões.

Ao meu amor, Felipe, por me aguentar nos momentos de pressão e prazos. Por ser meu melhor programador, que automatiza minhas tarefas intermináveis. Muito obrigada por estar ao meu lado neste momento especial. Te amo!

Aos meus amigos do laboratório, vulgo sala dos alunos do PROCC, obrigada pelo convívio. Em especial, gostaria de agradecer à Deborah, minha gêmea, que me ajudou muito no design das figuras, no paper da tese, no uso do Photoshop e nas análises da DM. Lu, você é “a cara”! Obrigada pelas ajudas durante a tese, pela amizade, pelos sarcasmos (Ernestina). Sem você, nosso grupo não seria o mesmo. Vanessinha, nossa Ariel, obrigada pelo apoio e dicas durante os seminários. E Rafa! Não tem nem o que dizer! Sem palavras para agradecer à quem sempre está disposto a ajudar! Obrigada por tudo galera!

A todo o corpo docente da BCS pelo compartilhamento da ciência de forma tão esclarecedora. Aos amigos pesquisadores Paulo Ricardo, Fabrício Alves e, em especial, Maurício Costa e Ana Carolina Guimarães, pelo convívio, apoio e compreensão.

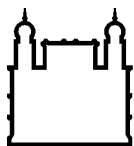
À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Instituto Oswaldo Cruz (IOC) pelo auxílio financeiro recebido ao longo desses quatro anos de trabalho.

Aos meus pais, José Luis e Luciane, por sempre me incentivarem nos meus estudos, por me darem conselhos e pelo exemplo de carinho e simplicidade. Obrigada por entenderem minhas ausências em festas e comemorações durante o doutorado.

Às minhas irmãs, por entenderem meu compromisso com a ciência e por não estar sempre disponível para ouvi-las nos seus mais diversos assuntos. Agora vou melhorar e serei uma irmã melhor! =)

*“Que a tua vida não seja estéril.  
Sê útil. Deixa rasto.”*

**São Josemaria Escrivá**



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

### ***ESTUDO DE METALOPROTEASES DE Leishmania (Viannia) braziliensis PERTENCENTES AO CROMOSSOMO 10 DA FAMÍLIA M8***

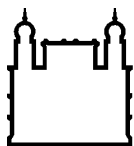
#### RESUMO

#### **TESE DE DOUTORADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS**

**Amanda Sutter de Oliveira Hammes**

O genoma de *Leishmania (Viannia) braziliensis* possui um número significativo de genes de metaloproteases. Entretanto, não se sabe por que esse parasito apresenta uma grande quantidade de genes em um determinado cromossomo. Neste estudo, apresentamos e discutimos os dados das estruturas tridimensionais de um conjunto de metaloproteases suportando que diferenças entre elas são cruciais para a adaptação do parasita no hospedeiro. Técnicas de modelagem comparativa foram utilizadas para a construção dos modelos 3D destas enzimas. Posteriormente, realizamos simulações de dinâmica molecular para examinar a estabilidade dos modelos gerados. O método de *docking* molecular foi aplicado para análise do modo de ligação de três inibidores conhecidos de metaloproteases (batimastat, marimastat e prinomastat), e dessa forma, ranquear as proteínas que obtiverem os menores valores de energia total. Nossos resultados apresentaram um comportamento dinâmico similar para todos os sistemas, embora o perfil eletrostático apresentou diferenças nos grupos definidos, principalmente para o motivo SSV. Depois do agrupamento realizado com base em dados geométricos do sítio de ligação, as estruturas de mesmo motivo se mantiveram juntas sugerindo a existência de uma microheterogeneidade no cromossomo 10 da família M8. Resultados do *docking* mostraram que as conformações ranqueadas por menor energia não têm relação com os motivos encontrados, porém algumas estruturas conseguiram bom resultado na interação com mais de um inibidor.

**Palavras Chave:** *Leishmania* spp; *Leishmania (Viannia) braziliensis*; proteases; metalo; leishmanolisina, modelagem comparativa; dinâmica molecular; *docking*.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

## INSTITUTO OSWALDO CRUZ

### INSIGHTS INTO METALLOPROTEASES FROM CHROMOSOME 10 OF *Leishmania (Viannia) braziliensis* M8 FAMILY

#### ABSTRACT

#### DOCTORAL THESIS IN COMPUTATIONAL SYSTEMS BIOLOGY

**Amanda Sutter de Oliveira Hammes**

The genome of *Leishmania (Viannia) braziliensis* has a significant number of metalloproteases genes. However, it is not known why this parasite has a large number of genes on a given chromosome. In this study, we present and discuss the data from the three-dimensional structures of a set of metalloproteases supporting that the differences among them are crucial for the adaptation of the parasite to hosts. Comparative modeling techniques were used to build 3D models of these enzymes. Afterward, molecular dynamics simulations were performed to examine the stability of the proteins. The docking method was applied for the analysis of the binding mode of the three known inhibitors of metalloproteases (batimastat, marimastat and prinomastat), and thus to rank the proteins that present the lowest total energies values. Our results showed a similar dynamic behavior for all systems, although the electrostatic profile showed differences in the defined groups, mainly for SSV motif. After clustering, based on geometrical data of the binding site, the structures were kept together suggesting the existence of a microheterogeneity in the chromosome 10 of M8 family. Docking results showed that the conformations ranked for lower energy have no relation with the defined motifs, on the other hand, some structures have managed good results in the interaction with more than one inhibitor.

**Keywords:** *Leishmania* spp; *Leishmania (Viannia) braziliensis*; proteases; metallo; leishmanolysin, comparative modeling; molecular dynamics; docking.

# ÍNDICE

<b>1 INTRODUÇÃO .....</b>	<b>1</b>
1.1 CONSIDERAÇÕES GERAIS .....	1
1.2 O PARASITO E O CICLO BIOLÓGICO.....	2
1.2.1 <i>Fatores de Virulência</i> .....	5
1.2.2 <i>Tratamento</i> .....	5
1.3 PROTEASES .....	6
1.4 METALOPROTEASES .....	11
1.4.1 <i>Metaloproteases de matriz - Leishmanolisinas</i> .....	12
1.4.2 <i>Metaloproteases de L. (V.) braziliensis</i> .....	14
1.5 PREDIÇÃO DE ESTRUTURA TRIDIMENSIONAL DE PROTEÍNAS .....	17
1.5.1 <i>Modelagem Comparativa</i> .....	18
1.5.1.1 Satisfação de Restrições Espaciais .....	20
1.5.1.2 União de Corpos Rígidos .....	21
1.5.1.3 Principais Programas de Modelagem Comparativa .....	22
1.5.2 <i>Métodos de novo</i> .....	22
1.5.2.1 Programas que usam o método <i>de novo</i> .....	22
1.5.3 <i>Threading</i> .....	23
1.5.3.1 Programas que usam método <i>Threading</i> .....	23
1.6 OTIMIZAÇÃO DA GEOMETRIA .....	24
1.7 DINÂMICA MOLECULAR.....	26
1.8 MÉTODOS DE ATRACAMENTO MOLECULAR ( <i>DOCKING</i> ) .....	29
<b>2 OBJETIVOS .....</b>	<b>32</b>
2.1 OBJETIVO GERAL.....	32
2.2 OBJETIVOS ESPECÍFICOS .....	32
<b>3 MATERIAIS E MÉTODOS .....</b>	<b>33</b>
3.1 SEQUÊNCIAS DE METALOPROTEASES DE <i>L. (V.) BRAZILIENSIS</i> .....	33
3.2 PREDIÇÃO DA LOCALIZAÇÃO CELULAR DAS METALOPROTEASES .....	33
3.3 PREDIÇÃO DA ESTRUTURA TRIDIMENSIONAL DE METALOPROTEASES .....	34
3.3.1 <i>Programas/servidores utilizados na construção dos modelos</i> .....	36
3.3.2 <i>Validação dos modelos</i> .....	38

3.3.3 Refinamento dos modelos .....	40
3.4 DINÂMICA MOLECULAR.....	41
3.4.1 Análise das Simulações de Dinâmica Molecular.....	42
3.4.1.1 Raiz do Desvio Médio Quadrático (RMSD).....	42
3.4.1.2 Flexibilidade dos resíduos do sítio catalítico.....	43
3.4.1.3 Clusterização.....	43
3.4.1.4 Perfil Eletrostático.....	43
3.4.1.5 Dendrograma estrutural.....	44
3.4.2 Identificação de hotspots.....	45
3.5 ATRACAMENTO MOLECULAR ( <i>DOCKING</i> ) .....	46
3.5.1 Preparação dos receptores para docking.....	48
3.5.2 Preparação dos ligantes para docking.....	50
3.5.3 Análise das simulações de Docking.....	51
<b>4 RESULTADOS E DISCUSSÃO.....</b>	<b>52</b>
4.1 VISÃO GERAL DAS PROTEÍNAS DO CROMOSSOMO 10 .....	52
4.1.1 Predição da Localização Celular .....	55
4.2 BUSCA DE MOLDES PARA CONSTRUÇÃO DO MODELO 3D.....	56
4.3 COMPARAÇÃO DOS PROGRAMAS DE MODELAGEM BASEADA EM MOLDE .....	58
4.3.1 Validação da estrutura 3D do conjunto teste.....	62
4.4 CONSTRUÇÃO E VALIDAÇÃO DAS DEMAIS ESTRUTURAS COM O SERVIDOR SWISS-MODEL .....	67
4.5 PROTONAÇÃO DOS RESÍDUOS QUE COORDENAM O ÍON ZINCO .....	70
4.6 AVALIAÇÃO DA ESTABILIDADE DA DINÂMICA MOLECULAR .....	72
4.7 AVALIAÇÃO DA FLUTUAÇÃO DOS RESÍDUOS DA REGIÃO DO SÍTIO CATALÍTICO .....	76
4.8 ENERGIA POTENCIAL.....	77
4.9 AVALIAÇÃO DO PERFIL ELETROSTÁTICO.....	78
4.10 DENDROGRAMA ESTRUTURAL .....	82
4.11 PROTONAÇÃO DOS LIGANTES PARA <i>DOCKING</i> .....	84
4.12 <i>DOCKING</i> MOLECULAR: ANÁLISE DAS CONFORMAÇÕES POR MENOR ENERGIA DE LIGAÇÃO .....	85
4.13 ANÁLISE DAS INTERAÇÕES ENTRE PROTEÍNA E LIGANTE .....	87
4.14 ANÁLISE DOS <i>HOTSPOTS</i> DAS METALOPROTEASES .....	91

4.15 COMPARAÇÃO ENTRE OS <i>HOTSPOTS</i> E OS RESÍDUOS DE INTERAÇÃO COM OS LIGANTES .....	93
<b>5 CONCLUSÕES.....</b>	<b>97</b>
<b>6 REFERÊNCIAS.....</b>	<b>99</b>
<b>ANEXO A.....</b>	<b>110</b>
<b>ANEXO B.....</b>	<b>114</b>
<b>ANEXO C.....</b>	<b>115</b>
<b>ANEXO D.....</b>	<b>125</b>
<b>ANEXO E.....</b>	<b>128</b>
<b>ANEXO F.....</b>	<b>129</b>
<b>ANEXO G.....</b>	<b>130</b>
<b>ANEXO H.....</b>	<b>136</b>
<b>ANEXO I.....</b>	<b>140</b>



## ÍNDICE DE FIGURAS

<b>Figura 1:</b> Distribuição de espécies do gênero <i>Leishmania</i> na transmissão da leishmaniose tegumentar americana em 2005.....	3
<b>Figura 2:</b> Ciclo Biológico de <i>Leishmania</i> spp. ....	4
<b>Figura 3:</b> Clãs e famílias de proteases de <i>Leishmania</i> (V.) <i>braziliensis</i> . Classificação das proteases anotadas no genoma de <i>Leishmania</i> spp.....	9
<b>Figura 4:</b> Quantidade de genes de proteases distribuídos nos cromossomos de <i>L.</i> (V.) <i>braziliensis</i> . ....	10
<b>Figura 5:</b> Estrutura do sítio catalítico do clã metzincina.....	12
<b>Figura 6:</b> Região do sítio catalítico de uma leishmanolisina de <i>L. major</i> . ....	13
<b>Figura 7:</b> Quantidade de resíduos de aminoácidos em cada sequência de metaloproteases do cromossomo 10 de <i>L.</i> (V.) <i>braziliensis</i> . ....	15
<b>Figura 8:</b> Acurácia e aplicação de modelos de estrutura de proteínas.....	19
<b>Figura 9:</b> Modelagem comparativa usando o método de satisfação de restrições espaciais do programa Modeller. ....	21
<b>Figura 10:</b> Superfície de energia potencial para simulação de dinâmica molecular. ...	25
<b>Figura 11:</b> Exemplo de uma equação empregada para aproximar as forças atômicas que governam o movimento molecular. ....	27
<b>Figura 12:</b> Etapas do atracamento molecular. ....	29
<b>Figura 13:</b> Desenho esquemático da construção dos modelos. ....	35
<b>Figura 14:</b> Representação do protocolo de refinamento utilizado pelo servidor KoBaMIN.....	40
<b>Figura 15:</b> Método de rolamento utilizado pelo servidor 3V. ....	45
<b>Figura 16:</b> Representação gráfica da caixa cúbica (azul) onde o grid é gerado.....	49
<b>Figura 17:</b> Estruturas 2D dos inibidores usados nas simulações de docking.....	50
<b>Figura 18:</b> Estrutura 2D dos ligantes desprotonados pelo programa Marvin. ....	50
<b>Figura 19:</b> Alinhamento da região do sítio catalítico com a identificação dos três motivos: DLX, SSV e NRI no início da região de inserção. ....	53
<b>Figura 20:</b> Dendrograma da organização dos genes do cromossomo 10 de <i>L.</i> (V.) <i>braziliensis</i> . ....	54
<b>Figura 21:</b> Predição da localização celular realizada com o servidor CELLO.....	55

<b>Figura 22:</b> Predição da localização celular da metaloprotease de <i>Leishmania major</i> .	56
<b>Figura 23:</b> Alinhamento do molde 1LML com o melhor modelo construído da sequência LbrM.10.1710, LbrM.10.1610 e LbrM.10.1550 com os programas i-tasser, Swiss-Model, Modeller, ModPipe, HHPred e Fugue.	59
<b>Figura 24:</b> Estruturas geradas pelo ANOLEA.	66
<b>Figura 25:</b> Alinhamento estrutural entre proteína molde e modelo construído pelo programa Swiss-Model e o gráfico DOPE correspondente.	68
<b>Figura 26:</b> Gráfico DOPE da estrutura PDB id 1LML, usada como molde para a construção dos modelos de <i>L. V. braziliensis</i> .	69
<b>Figura 27:</b> Alinhamento entre a sequência LbrM.10.1550 e a proteína molde 1LML.	70
<b>Figura 28:</b> Protonação dos resíduos His165, His169 e His235 que participam da coordenação do Zinco.	72
<b>Figura 29:</b> Desvio RMS dos átomos da cadeia principal (backbone) representado em função do tempo de simulação.	74
<b>Figura 30:</b> Desvios RMS dos átomos da cadeia principal (backbone) da região de inserção das proteínas representado em função do tempo de simulação.	75
<b>Figura 31:</b> RMS de Flutuação representado por resíduo da região de inserção das proteínas.	76
<b>Figura 32:</b> Estrutura de motivo DLX com destaque para as regiões de maior flutuação.	77
<b>Figura 33:</b> Energia potencial total das estruturas simuladas.	78
<b>Figura 34:</b> Perfil eletrostático da proteína PDB id 1LML.	78
<b>Figura 35:</b> Diferenças no perfil eletrostático da região de inserção das metaloproteases de <i>Leishmania (V.) braziliensis</i> em pH 7.	79
<b>Figura 36:</b> Diferenças no perfil eletrostático da região de inserção das metaloproteases de <i>Leishmania (V.) braziliensis</i> em pH 5.	81
<b>Figura 37:</b> Dendrograma estrutural das metaloproteases do cromossomo 10 de <i>L. (V.) braziliensis</i> .	83
<b>Figura 38:</b> Curva da distribuição de micro espécies por pH e estado de protonação dos ligantes usados nas simulações de docking.	85

## LISTA DE TABELAS

<b>Tabela 1:</b> Códigos de acesso das estruturas tridimensionais que contêm pelos menos o domínio catalítico dos membros pertencentes ao clã metzincina depositadas no banco de dados público, PDB – Protein Data Bank.....	16
<b>Tabela 2:</b> Subgrupos e regiões identificadas por análise sequencial das sequências de leishmanolisinas presentes no cromossomo 10 de <i>L. (V.) braziliensis</i> . ....	52
<b>Tabela 3:</b> Identificação do molde com os valores de cobertura, identidade e similaridade com a sequência alvo. ....	57
<b>Tabela 4:</b> RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1710. ....	60
<b>Tabela 5:</b> RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1610. ....	60
<b>Tabela 6:</b> RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1550. ....	61
<b>Tabela 7:</b> Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1710. ....	62
<b>Tabela 8:</b> Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1610. ....	63
<b>Tabela 9:</b> Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1550. ....	63
<b>Tabela 10.</b> Resumo dos servidores/programas que obtiveram o melhor resultado em cada um dos critérios avaliados para as sequências do conjunto teste.....	67
<b>Tabela 11:</b> Parâmetros de validação estatística dos modelos LbrM.10.0470, LbrM.10.1550 e LbrM.10.1560. ....	69
<b>Tabela 12:</b> Estado de protonação das Histidinas do sítio catalítico calculado com o servidor H++. ....	71
<b>Tabela 13.</b> Valores numéricos da região de inserção das metaloproteases.....	82
<b>Tabela 14:</b> Menor energia obtida entre cada inibidor de metaloproteases e as proteínas em estudo. ....	86
<b>Tabela 15:</b> Resíduos que interagem em mais de 50% das metaloproteases de <i>Leishmania (V.) braziliensis</i> com o ligante Batimastat. ....	88
<b>Tabela 16.</b> Resíduos que interagem em mais de 50% das metaloproteases de <i>Leishmania (V.) braziliensis</i> com o ligante Marimastat. ....	89

<b>Tabela 17.</b> Resíduos que interagem em mais de 50% das metaloproteases de <i>Leishmania (V.) braziliensis</i> com o ligante Prinomastat.....	90
<b>Tabela 18:</b> Energia de Gibbs dos hotspots definidos após a troca dos resíduos da região de inserção por Alanina.....	92
<b>Tabela 19:</b> Identificação das metaloproteases que possuem interação com hotspots nas simulações de docking para os ligantes Batimastat, Marimastat e Prinomastat...	94

## **LISTA DE SIGLAS E ABREVIATURAS**

ANOlea – Atomic Non-Local Environment Assessment  
APBS – Adaptive Poisson-Boltzmann Solver  
BLAST - Basic Local Alignment Search Tool  
CASP - Comparative Assessment of Methods for Protein Structure Prediction  
CELLO – subCELLular LOcalization predictor  
DM - Dinâmica molecular  
DOPE - Discrete Optimized Protein Energy  
ELL – Energia livre de ligação  
ENCAD - Energy Calculation and Dynamics  
I-Tasser - Iterative Threading Assembly Refinement  
iTOL – Interactive Tree of Life  
GIPL - Glicoinositolfosfolípídeos  
GPI – Glicosilfosfatidilinositol  
KoBaMIN - Knowledge Based MINimization  
LGA – Algoritmo Genético Lamarckiano, na sigla em inglês  
LOMETS - Local Meta-Threading Server  
MMPs - Metaloproteases de matriz, na sigla em inglês  
MUSTER - MUlti-Source ThreadER  
NCBI – National Center for Biotechnology Information  
PDB – Protein Data Bank  
PME - Particle-Mesh Ewald  
PMEMD – Particle-Mesh Ewald Molecular Dynamics  
PPA - Sequence profile-profile alignment  
RMN - Ressonância Magnética Nuclear  
RMSD – Raiz do desvio médio quadrático, na sigla em inglês  
RMSF – Flutuação da raiz média quadrática, na sigla em inglês

# 1 INTRODUÇÃO

## 1.1 Considerações Gerais

As leishmanioses englobam um grupo de doenças parasitárias causadas por parasitos da Ordem Kinetoplastida, Família Trypanosomatidae e gênero *Leishmania* (Ross 1903a, Ross 1903b). Os parasitos do gênero *Leishmania* são responsáveis por produzir doenças no homem e em outros mamíferos, possuem uma ampla distribuição geográfica e podem apresentar três principais tipos de leishmanioses: visceral, cutânea e muco-cutânea. Em humanos, a forma visceral, geralmente conhecida como Kala-azar, é a forma mais severa da doença tendo uma estimativa entre 200 a 400 mil novos casos a cada ano sendo que cerca de 30 mil casos levam ao óbito. Já a forma cutânea é a forma mais comum, com aproximadamente 1,2 milhões de casos novos relatados por ano. A forma muco-cutânea é derivada da forma cutânea mas produz lesões destrutivas das mucosas da nasofaringe que destroem a cartilagem do nariz e do palato provocando deformações graves, desfigurando a face (WHO 2014).

O gênero *Leishmania* se agrupa em dois gêneros: *Leishmania* e *Viannia*; que se diferenciam segundo o padrão de desenvolvimento dos parasitas no intestino do vetor invertebrado. Atualmente, mais de 20 espécies de *Leishmania* são conhecidas por infectar o homem e mais de 30 espécies de flebotomíneos foram descritos como prováveis vetores para os parasitos do gênero *Leishmania* (Desjeux 1996). No subgênero *Leishmania*, os parasitas podem se desenvolver no intestino anterior e médio do flebotomo, enquanto em *Viannia* estes podem se desenvolver no intestino posterior (Lainson, Shaw *et al.* 1987).

No Brasil, as principais espécies causadoras de *Leishmania* são: *Leishmania (Viannia) braziliensis*, *L. (V.) guyanensis* e *Leishmania (Leishmania) amazonensis* (Grimaldi Jr, Tesh *et al.* 1989). A *L. (V.) braziliensis* é o principal agente etiológico da forma muco-cutânea da doença nas Américas, e o mais frequente agente da leishmaniose tegumentar no Brasil (Rangel and Lainson 2003). Inicialmente, a doença estava limitada a áreas rurais e a pequenas localidades urbanas, mas com o passar do tempo, se expandiu para grandes centros (Brasil 2009).

Casos de leishmaniose são relatados em todo o território nacional (Figura 1). No estado do Rio de Janeiro, casos autóctones e epidemias de leishmaniose tegumentar vêm sendo registrados desde o início do século XX. Existem várias

regiões de alta incidência, principalmente no norte e nordeste fluminense, mas também existem focos em zonas urbanas e semiurbanas, como na Zona Oeste da cidade do Rio de Janeiro. Uma possível explicação para estas epidemias pode ser dada pelo processo de ocupação de áreas florestais para a formação de espaço urbano, como em Jacarepaguá em 1974 (Marzochi, Coutinho *et al.* 1979). Foram registrados cerca de 4.900 casos de leishmaniose tegumentar no Estado do Rio de Janeiro entre 1985 e 2003, sendo que desses, cerca de 1.300 foram descritos no Município do Rio de Janeiro, onde os mosquitos das espécies *Lutzomyia intermedia* e *Lutzomyia migonei* já foram identificados naturalmente infectados com *Leishmania (V.) braziliensis* (de Pita-Pereira, Alves *et al.* 2005), o que demonstra que a doença permanece em franca atividade nesta área. Outras informações mais recentes, como o número estimado de casos de morte anual por leishmaniose cutânea e visceral são disponibilizados na página<sup>1</sup> da organização mundial da saúde (WHO).

## 1.2 O parasito e o ciclo biológico

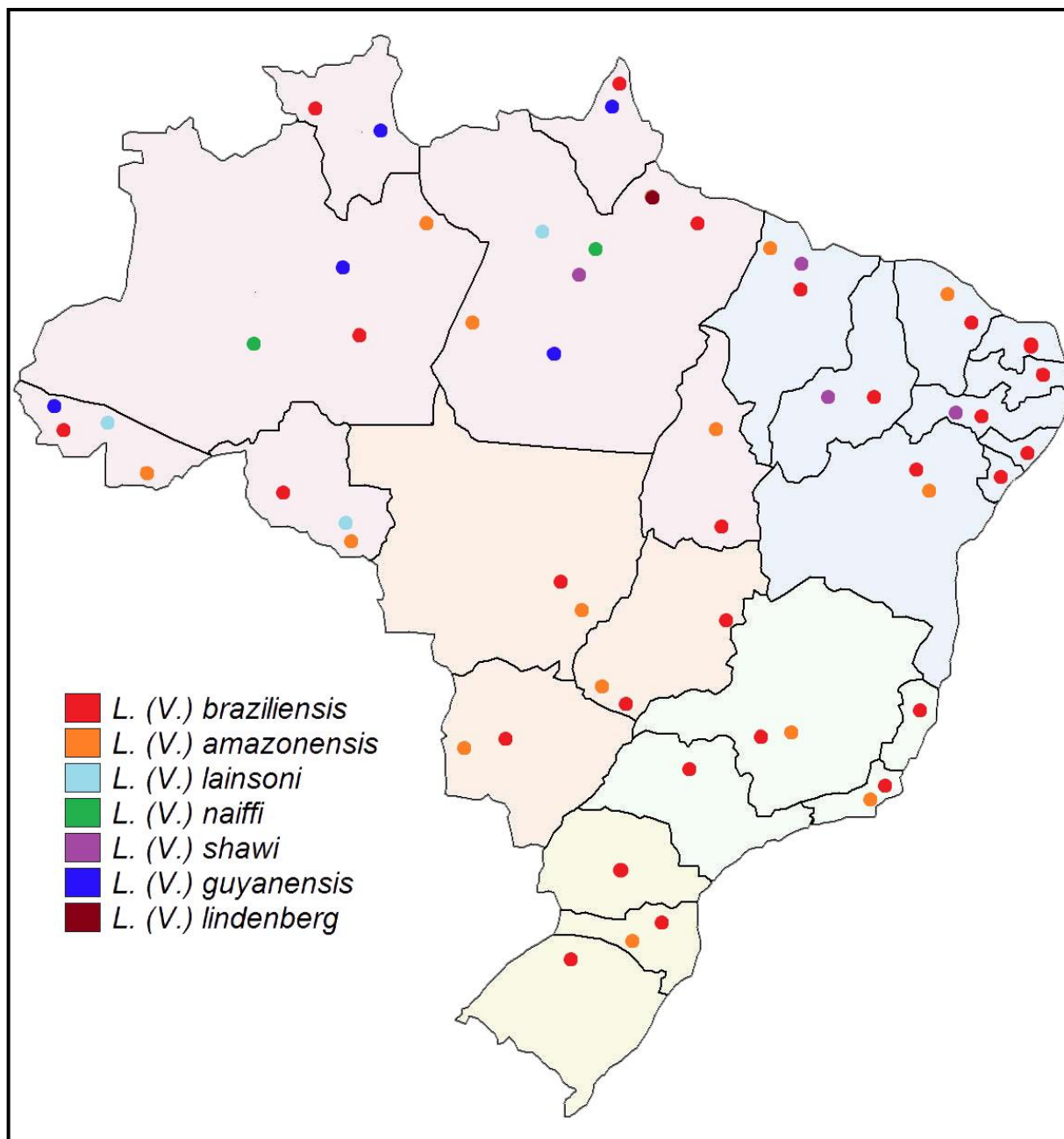
As espécies do gênero *Leishmania* apresentam seu ciclo biológico dividido em duas fases: um no hospedeiro invertebrado e outro no hospedeiro vertebrado. De acordo com a fase do ciclo de vida, o parasito pode apresentar duas formas: amastigota e promastigota. A forma infectante se dá através das promastigotas, formas extracelulares com flagelo livre, que uma vez fagocitadas pelas células de defesa do organismo do hospedeiro, ao invés de serem digeridas por estas, sofrem transformação, dando origem à forma amastigota, forma intracelular sem flagelo livre (Muskus and Marín 2002).

A forma promastigota das leishmanias entra em contato com o hospedeiro vertebrado no momento do repasto sanguíneo do inseto vetor sendo rapidamente fagocitada pelos macrófagos, fundindo-se mais tarde com o lisossoma para formar o fagolisossoma, onde ocorre a diferenciação para as formas amastigotas (Cohen-Freue, Holzer *et al.* 2007). As formas amastigotas se reproduzem, rompendo a célula infectada e são rapidamente fagocitadas, ficando na forma intracelular. Quando ingeridas pelo mosquito, transformam-se novamente em promastigotas no

---

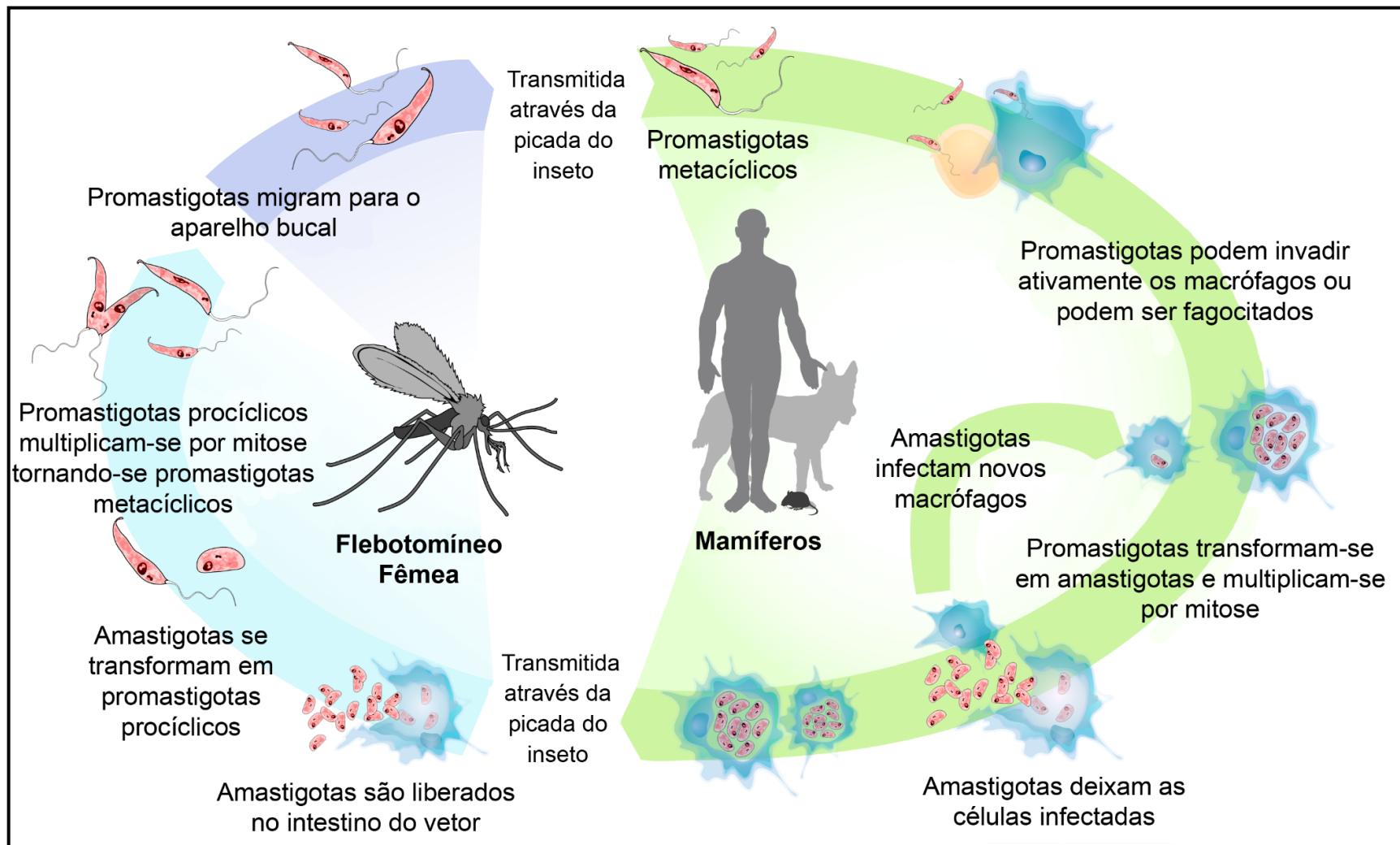
<sup>1</sup> <http://www.who.int/leishmaniasis/en/>

intestino do agente vetor, retornando ao aparelho bucal do mesmo de forma a serem transferidas para outro hospedeiro, completando o ciclo (Vannier-Santos, Martiny *et al.* 2002) (Figura 2).



**Figura 1:** Distribuição de espécies do gênero *Leishmania* na transmissão da leishmaniose tegumentar americana em 2005. Adaptado de Ministério da Saúde/Serviço de Vigilância em Saúde/2007.





**Figura 2:** Ciclo Biológico de *Leishmania* spp. Figura retirada do artigo de revisão publicado por (Harhay, Olliaro et al. 2011).

### 1.2.1 Fatores de Virulência

Parasitas do gênero *Leishmania* possuem um arsenal variado de componentes estruturais e funcionais, essenciais para sua adaptação nos hospedeiros vertebrados e invertebrados (Conceição-Silva and Alves 2014). Na revisão realizada por Silva-Almeida e colaboradores (Silva-Almeida, Pereira *et al.* 2012), as proteases são apontadas como fatores de virulência em *Leishmania* spp na infecção no hospedeiro mamífero.

### 1.2.2 Tratamento

Apesar da grande prevalência mundial e dos conhecimentos da relação entre o parasito e o seu hospedeiro, poucos avanços foram obtidos no tratamento das leishmanioses. A base do tratamento atual foi descrita por Gaspar Viana em 1912 (Vianna 1912), quando foi descoberta a ação curativa dos antimoniais sobre a doença cutânea. Somente na década de 1940 foram introduzidos os antimoniais pentavalentes, como o estibogluconato de sódio (Pentostan®) e o antimoniato de N-metil glucamina (Glucantime®), na quimioterapia.

Atualmente, esses fármacos são os recomendados no tratamento das leishmanioses, embora apresentem inconvenientes como alta cardiotoxicidade, esquema de tratamento longo e eficiência questionável (Bryceson 2001). Pacientes que apresentam lesões resistentes ao tratamento não o torna necessariamente portador de uma cepa de *Leishmania* resistente. O parasito pode apresentar mecanismos já descritos de resistência intrínseca a medicamentos e o regime de tratamento ter que ser alterado devido à resistência generalizada. Casos de insucesso na quimioterapia com os antimoniais pentavalentes podem estar relacionados com fatores advindos tanto do parasito quanto do hospedeiro.

Estudos controlados comparando a eficácia e toxicidade *in vivo* das formulações dos compostos antimoniais (N-metil-glucamina e estibogluconato de Sódio BP88) usados no Brasil para o tratamento das leishmanioses são poucos. Em estudos recentes, observou-se que quando utilizados na dosagem de 20 mg de  $Sb^{+5}/Kg/dia$  (Deps, Viana *et al.* 2000) ou 15 mg de  $Sb^{+5}/Kg/dia$  durante 20 dias consecutivos (Berman, Wiernik *et al.* 1997) ambas as formulações teriam a mesma eficácia, apesar da maior toxicidade observada no grupo tratado com estibogluconato de sódio BP88.

Os fármacos empregados como segunda escolha neste tratamento são a pentamidina e a anfotericina B que, além do alto custo possuem problemas quanto a via de administração (Saldanha, Romero *et al.* 1999). Outros compostos vêm sendo estudados quanto à sua atividade leishmanicida nas doenças tegumentar e visceral (pirimetamina, oketoconazol, metotrexato, adriamicina e miltefosina) e, mesmo assim, ainda não se tem um fármaco que apresente um bom índice terapêutico e uma baixa toxicidade (Sereno, Guilvard *et al.* 2001).

O estado da arte dos agentes quimioterapêuticos empregados atualmente no combate das leishmanioses nos mostra que os medicamentos usados hoje surgiram a partir de programas de seleção empíricos. Neste transcurso, pouco se valorizou as diferenças bioquímicas e metabólicas entre a *Leishmania* e seus hospedeiros mamíferos, o que forneceria, indiscutivelmente, uma alternativa razoável para o desenvolvimento de novos agentes quimioterapêuticos.

Com o objetivo de buscar variáveis bioquímicas que estabeleçam diferenças entre as leishmanias e seus hospedeiros, diversos estudos indicam uma variação bioquímica durante a diferenciação do parasito após a infecção dos hospedeiros vertebrados e invertebrados (Köhler 1986, Gull 2001), o que poderiam constituir na seleção de alvos para quimioterapias mais eficazes e com menor resistência (Croft 2001). Ainda que o impacto exato das variações bioquímicas nas manifestações clínicas não esteja definida, há evidências de que esses componentes bioquímicos possam modular as interações entre a *Leishmania* e células imunes do hospedeiro (Silva-Almeida, Pereira *et al.* 2012).

### **1.3 Proteases**

As proteases de *Leishmania* spp. estão envolvidas nos mais diversos mecanismos, tais como virulência e patogênese. A ação patogênica do parasito está relacionada com a destruição celular provocada pela reprodução das formas amastigotas. As funções exercidas pelas proteases vão desde a digestão de proteínas para fins nutritivos a refinados controles de funções de proteína hidrolisando ligações peptídicas em determinados substratos (Barrett 1994). Em *Leishmania* spp, essas enzimas desempenham um papel fundamental na interação entre o parasito e o hospedeiro, uma vez que têm o potencial de degradar proteínas e peptídeos que participam de uma gama de funções biológicas.

As proteases são vistas como importantes proteínas sinalizadoras que estão envolvidas no correto funcionamento da maioria das funções biológicas. Em humanos, a atividade anormal dessas enzimas pode estar relacionada direta ou indiretamente com diversas patologias cardiovasculares e inflamatórias, cânceres e desordens neurológicas (Turk 2006).

O termo proteinases se refere a uma subdivisão de peptidases que são as endopeptidases. Estas enzimas atuam nas regiões internas da cadeia polipeptídica e não nas porções externas N- ou C-terminais (denominadas aminopeptidases e carboxipeptidases, respectivamente), sendo este tipo de enzimas chamadas de exopeptidases. Assim, o termo *in* de prote<sup>i</sup>nase vem da língua inglesa que significa dentro, e se refere ao local de clivagem, no interior da molécula proteica (Barrett, Rawlings *et al.* 2001, Barrett, Tolle *et al.* 2003). Dependendo do grupo reativo envolvido na catálise e a natureza do sítio catalítico, elas podem ser subdivididas em serino- (EC 3.4.21), cisteíno- (EC 3.4.22), aspártico- (EC 3.4.23), metalo- (EC 3.4.24) e treonino-peptidases (EC 3.4.25) (Rawlings, Barrett *et al.* 2009).

Essas enzimas são extremamente abundantes no genoma, e juntamente com seus homólogos correspondem a aproximadamente 2% dos genes codificadores em alguns organismos (Rawlings and Barrett 2000, Puente, Sánchez *et al.* 2003). Estão distribuídas em uma grande variedade de fontes, como: plantas, animais e micro-organismos.

No trabalho de Silva-Almeida e colaboradores (Silva-Almeida, Souza-Silva *et al.* 2014) foi feito um levantamento bibliográfico sobre as proteases presentes no genoma de *Leishmania* spp. Pensando no uso de proteases como alvos para quimioterapia, foi verificado que essas enzimas são diferentes das enzimas correspondentes em seus hospedeiros mamíferos, como humanos e cachorros, sendo validado através da genômica comparativa dos genes de *Leishmania* spp depositados em banco de dados públicos.

O sequenciamento completo do genoma de diversos micro-organismos, levou a uma modernização do sistema de classificação das proteases, pela necessidade de abranger o diversificado repertório catalítico encontrado na natureza, e de catalogar e sistematizar a grande quantidade de informação produzida ao longo dos anos.

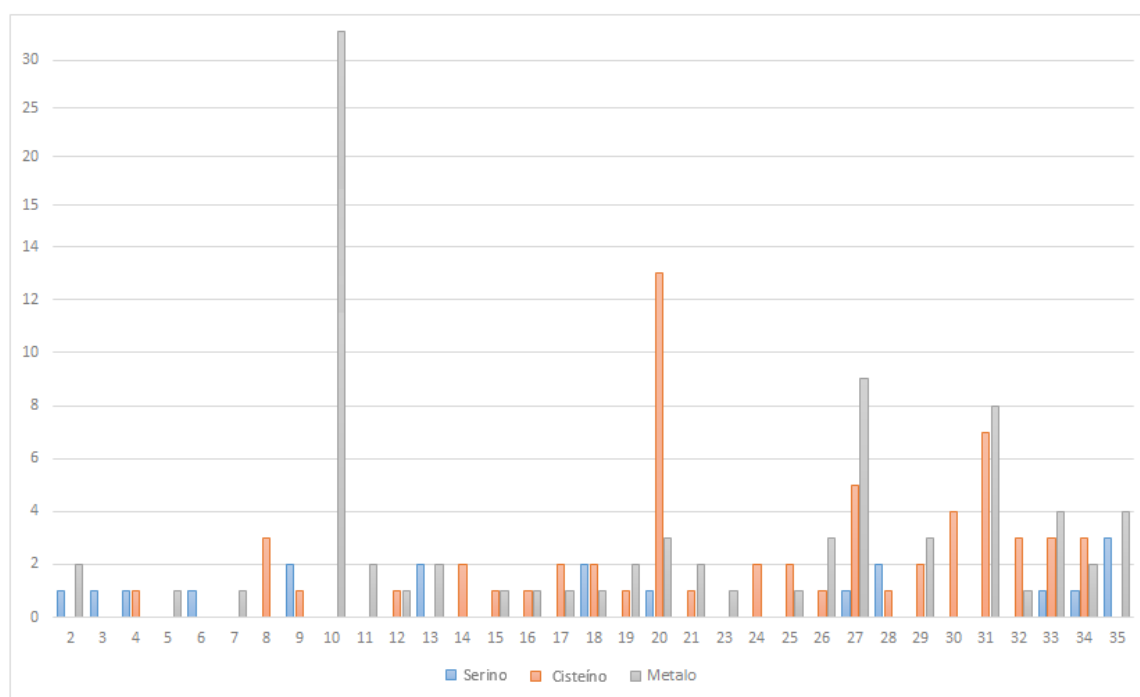
Dessa forma, foi criado o banco de dados MEROPS (<http://merops.sanger.ac.uk>), que agrupa as enzimas em famílias de acordo com a homologia baseada na identidade entre sequências de aminoácidos, onde as

famílias cujos genes provem de uma mesma origem ancestral são agrupadas em clãs. Este tipo de informação é determinado pela estrutura terciária das peptidases (Rawlings, Waller *et al.* 2014). Na Figura 3 é possível observar os clãs e famílias das quatro proteases e a quantidade de genes encontrados no genoma de *Leishmania (V.) braziliensis*.

	Clan	Familia	Qtd			
Aspártico	AA	A28	1	3		
	AD	A22	2			
Cisteino	CA	C1	5	62		
		C2	30			
		C12	2			
		C19	16			
		C48	1			
		C51	2			
		C54	2			
	CD	C65	1			
		C13	1			
	CF	C14	1			
		C15	1			
	Metalo	MA	M1		5	89
			M3		11	
M8			36			
M32			2			
M41			6			
MC		M48	1			
		M49	1			
ME		M14	3			
		M16	5			
MF		M17	7			
		M24	7			
MG	M18	1				
	M20	2				
MH	M6	4				
Serino	SB	S8	2	17		
	SC	S9	6			
		S10	2			
	SF	S26	1			
	SJ	S16	3			
	SP	S59	1			
	ST	S54	2			

**Figura 3:** Clãs e famílias de proteases de *Leishmania (V.) braziliensis*. Classificação das proteases anotadas no genoma de *Leishmania* spp. Nomenclatura baseada na base de dados MEROPS. Os números nos quadrados representam o total estimado de proteases em cada família. Adaptado de Silva-Almeida (Silva-Almeida, Souza-Silva *et al.* 2014).

O genoma de *L. (V.) braziliensis*, espécie alvo deste estudo, possui 35 cromossomos (Britto, Ravel *et al.* 1998) com um total de 171 genes referente a proteases. A distribuição gênica corresponde a um total de 3, 62, 89 e 17 genes para aspártico, cisteíno, metalo e serino proteases (Figura 3), respectivamente. A quantidade de genes presentes por cromossomo pode ser vista na Figura 4. É possível observar que os genes de serino proteases se concentram em maior quantidade no cromossomo 35. Em cisteíno proteases, o cromossomo 20 é o que possui a maior quantidade de genes. Já em metaloproteases, os genes se concentram no cromossomo 10 e é o cromossomo com a maior quantidade de genes comparado às demais proteases.



**Figura 4:** Quantidade de genes de proteases distribuídos nos cromossomos de *L. (V.) braziliensis*.

Dada a nomenclatura do gene é possível identificar o cromossomo ao qual este está relacionado assim como seu alelo. Por exemplo, para o gene LbrM.10.0500, LbrM significa que se trata de um gene de metaloprotease de *Leishmania (V.) braziliensis*, 10, significa que este gene pertence ao cromossomo 10 e 0500 é o alelo correspondente. Essa nomenclatura será usada na sessão de resultados.

## 1.4 Metaloproteases

As metaloproteases zinco dependentes são classificadas em quatro grupos distintos (Hooper 1994, Gomis-Rüth 2003): (i) zincinas, que contêm a sequência conservada HEXXH como motivo de ligação ao zinco; (ii) inverzincinas, que possuem o domínio zincina de ligação ao zinco invertido (HXXE); (iii) carboxipeptidase, que possuem o motivo de ligação ao zinco, HXXE e; (iv) DD-carboxipeptidase, que apresenta o motivo de ligação ao zinco HXH. Nos motivos apresentados, X representa um aminoácido qualquer. Para ligação ao metal, todos os grupos possuem como primeiro e segundo ligantes resíduos de histidina, que atua como uma base ou ácido durante o processo catalítico, exceto nas carboxipeptidases. Já o terceiro ligante pode ser um resíduo de histidina (exceto nas inverzincinas) ou um ácido glutâmico (zincinas e inverzincinas). O maior grupo é representado pelas zincinas, subdivididas em gluzincinas (aquelas em que o terceiro ligante do zinco é o ácido glutâmico) e metzincinas (cujo terceiro ligante do zinco é uma histidina).

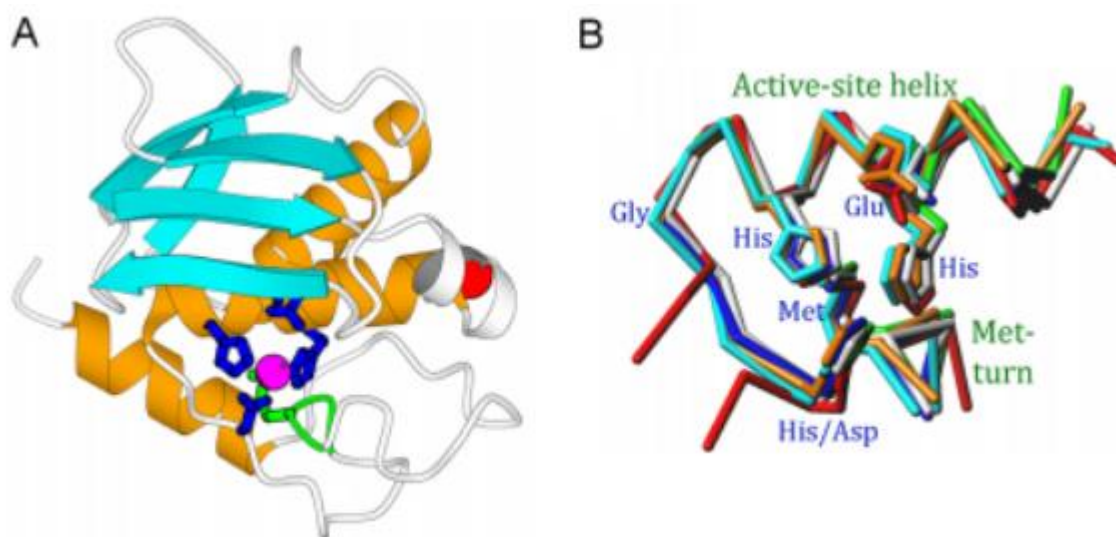
Metzincinas contêm proteínas com um pró-domínio N-terminal envolvida na manutenção de latência, um domínio catalítico de protease e domínios envolvidos em interações proteína-proteína e célula-célula além de outras funções reguladoras. O domínio da protease é caracterizado por um C-terminal estendido do motivo de ligação do zinco, HEXXHXXGXX(H/D), com a marca de uma glicina e uma terceira histidina ou aspartato na ligação ao zinco. Além disso, há uma metionina presente em um loop conservado chamado de Met-turn (Bode, Gomis-Rüth *et al.* 1993, Stöcker, Grams *et al.* 1995, Gomis-Rüth 2003). Este resíduo de metionina parece estabilizar os resíduos de histidina, favorecendo a coordenação do metal.

Metzincinas se dividem em famílias, sete das quais foram caracterizadas a nível estrutural por pelo menos um de seus membros: astacinas, ADAMs/adamalisinas/reprolisinas, serralisinas, metaloproteases de matriz, snapalisinas, leishmanolisinas e papalisinas. Atualmente existem mais de 200 estruturas de metzincinas depositadas no banco de dados de proteínas (PDB – Protein Data Bank (Berman, Westbrook *et al.* 2000)). As estruturas revelam que as



metzincinas compartilham um sítio ativo comum mas cada família tem elementos estruturais distintos (Gomis-Rüth 2009).

A estrutura do sítio catalítico desta classe é conservada. Na Figura 5 ilustra-se a estrutura tridimensional de uma snapalissina de *S. caespitosus* (PDB ID 1c7k) onde é possível observar os resíduos His83, Glu84, His87, His/Asp93 que participam da coordenação do íon.



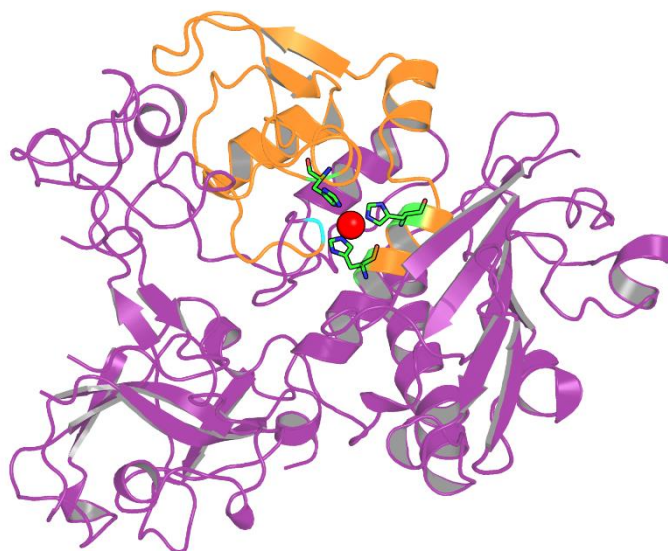
**Figura 5:** Estrutura do sítio catalítico do clã metzincina. Os elementos de estrutura secundária são comuns a todas as metzincinas. A: em laranja as  $\alpha$ -hélices, em ciano, as folhas- $\beta$  e o Met-turn é mostrado em verde. B: sete protótipos de metzincinas estruturalmente caracterizadas foram sobrepostas ilustrando a  $\alpha$ -hélice do sítio ativo incluindo as cadeias laterais das histidinas e aspartato, coordenadores do íon zinco e também o glutamato e o Met-turn. O íon do sítio catalítico foi omitido para uma melhor visualização. As cores utilizadas com o respectivo clã são: azul – astacina; ciano – adamalissina II; vermelho – leishmanolisina; verde – MMP-8; branco – aeruginolisina; amarelo – snapalissina e laranja – ulilissina. Esta figura foi retirada do artigo publicado por (Tallant, García-Castellanos *et al.* 2010).

#### 1.4.1 Metaloproteases de matriz - Leishmanolisinas

Dentre as famílias de metzincina, se destacam as metaloproteases de matriz (MMPs) e as leishmanolisinas. A família das MMPs foram as primeiras descritas em vertebrados, mas também têm sido encontradas em invertebrados e plantas. Até o momento existem pelo menos 18 distintas metaloproteases de matriz em vertebrados, e, como exemplo, pode-se citar a MMP-1, que é uma Colagenase 1, MMP-6, que é uma metaloprotease ácida; a MMP-14, que é uma membrana de metaloprotease de matriz tipo 1, dentre outras. Este grupo foi originalmente definido a partir da dependência de zinco para a atividade catalítica e a presença de grandes

peptídeos responsáveis pela latência. O alinhamento entre as estruturas de MMPs mostra a conservação do sítio catalítico das mesmas (Woessner 2002).

A leishmanolisina ou glicoproteína de 63 kDa (gp63), é uma metaloprotease dependente de zinco com grande similaridade estrutural no domínio de ligação a esse íon na comparação com diferentes leishmanolisinas. Essa enzima é um dos principais componentes de superfície presente em todas as espécies do gênero *Leishmania*, especialmente nas formas promastigotas, representando mais de 1% do total de proteínas do parasito (Jaffe and Dwyer 2003). A única estrutura tridimensional disponível no PDB resolvida por métodos experimentais é a de um membro da espécie de *L. major* (PDB ID: 1lml) (Schlagenhauf, Etges *et al.* 1998). Essa proteína foi sintetizada como um precursor inativo de 602 resíduos no retículo endoplasmático com um sinal e um pró peptídeo de 100 resíduos de aminoácidos no qual incluía um resíduo de cisteína altamente conservado que atua potencialmente como ativador deste processo. A ativação libera uma metaloprotease madura de ~280 resíduos seguido de ~200 resíduos no domínio C-terminal. Entre a glicina e a terceira histidina do motivo de consenso longo, são inseridos 60 resíduos (Figura 6).



1LML **H**EMA**H**ALGF**S**GPFFEDARIVANVPNVRGKNFDVPVINS**S** 39  
TAVAKAREQYGC**D**TLEYLEVEDQGGAGSAG**S****H**IKMRNAQDE**L****M** 82

**Figura 6:** Região do sítio catalítico de uma leishmanolisina de *L. major* (PDB ID: 1lml). A região em laranja representa a região de inserção, as histidinas que coordenam o zinco estão representadas em verde, em ciano, está o Met-turn e o zinco está representado por uma esfera vermelha.

Esta glicoproteína, ligada à membrana plasmática por uma âncora de glicosilfosfatidilinositol (GPI), é expressa em grande quantidade na superfície de promastigotas metacíclicos infectantes ao hospedeiro vertebrado e tem um baixo nível de expressão na forma amastigota intracelular. Essa diferença na expressão de gp63 pode estar relacionada com sua função no processo de entrada do parasita na célula hospedeira, na sobrevivência dos amastigotas, e na modulação da resposta imune do hospedeiro (Olivier, Gregory *et al.* 2005).

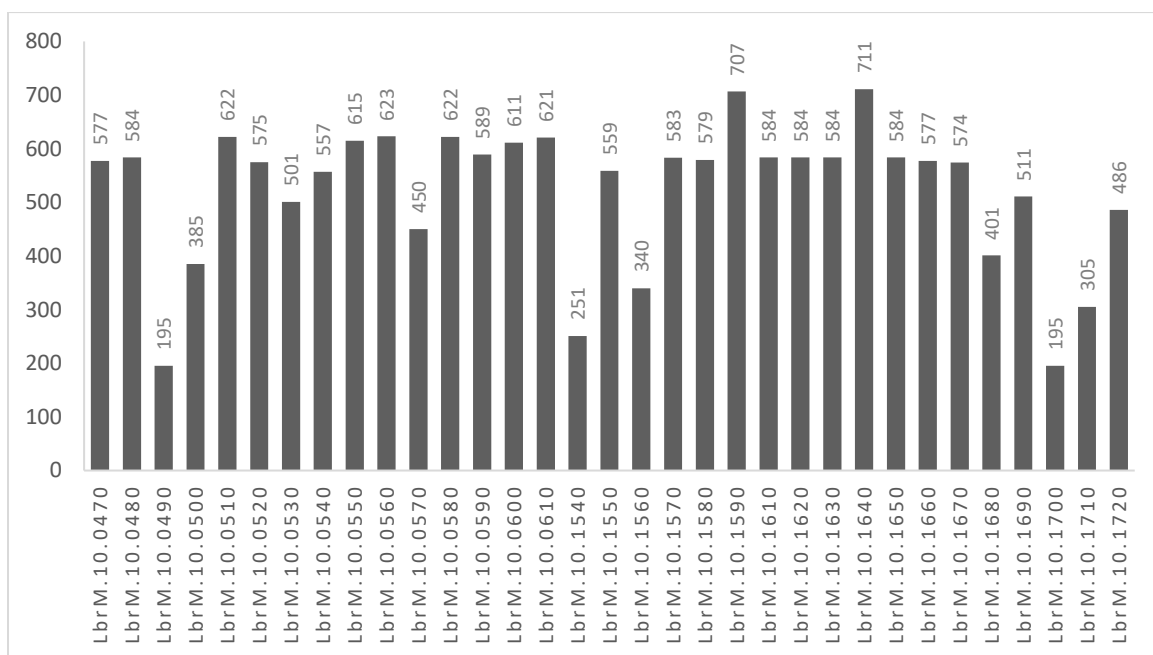
Diversas funções de gp63 vêm sendo descritas relacionadas ao sucesso da infecção das *Leishmanias* no hospedeiro vertebrado, como: (i) resistência à lise mediada pela clivagem do componente C3b do Sistema Complemento em C3bi, facilitando a fixação da *Leishmania* ao macrófago; (ii) proteção da degradação dentro do fagolisossomo de macrófagos (Hsiao, Yao *et al.* 2008); (iii) inibição da quimiotaxia de monócitos e neutrófilos; (iv) ligação às células *natural killer* inibindo sua proliferação e facilitando assim a infecção parasitária (Lieke, Nylen *et al.* 2008); e (v) evasão da morte apoptótica mediada por peptídeos antimicrobianos (Kulkarni, McMaster *et al.* 2006).

Como revisado por Silva-López (da Silva-López 2010), inibidores de metaloproteases foram utilizados em culturas de promastigotas de *L. major*, mas tais compostos não demonstraram o efeito citotóxico esperado. Além disso, a gp63 tem sido empregada em vários estudos de imunização experimental contra diferentes espécies de *Leishmania* e tem conferido proteção efetiva em camundongos suscetíveis. Entretanto, outros animais também susceptíveis à infecção por *Leishmania*, como algumas espécies de cães e macacos, não desenvolveram resposta imune protetora quando estimulados por diferentes formulações contendo leishmanolisina. Estudos continuam sendo conduzidos para estudar a melhor formulação e administração da gp63 que confira maior imunoproteção, e sua possível utilização em vacinas contra leishmanioses.

#### **1.4.2 Metaloproteases de *L. (V.) braziliensis***

As metaloproteases são as que apresentam a maior quantidade de genes, 89. Ela está dividida em 7 clãs (MA, MC, ME, MF, MG, MH e MP) e possui 14 famílias (M1, M3, M6, M8, M14, M16, M17, M18, M20, M24, M32, M41, M48 e M49). O cromossomo 10 é o que apresenta uma maior quantidade de alelos, 33 (Figura

4) e este pertence à família M8, clã Ma. No cromossomo 10 da família M8 de *Leishmania (V.) braziliensis*, o número de aminoácidos pode variar entre 195 e 711 resíduos (Figura 7). Tendo em vista a grande quantidade de sequências de metaloproteases presentes no cromossomo 10 deste parasita, o objetivo deste trabalho é encontrar as principais diferenças entre essas proteínas.



**Figura 7:** Quantidade de resíduos de aminoácidos em cada sequência de metaloproteases do cromossomo 10 de *L. (V.) braziliensis*.

Até o momento, não existem estruturas tridimensionais resolvidas por métodos experimentais, como cristalografia ou ressonância magnética nuclear, de metaloproteases de *Leishmania (V.) braziliensis*. No banco de dados público de proteínas (PDB) (Berman, Westbrook *et al.* 2000) há mais de 200 estruturas que contêm pelos menos o domínio catalítico dos membros pertencentes ao clã metzincina (Tabela 1), porém, somente 1 estrutura foi resolvida experimentalmente para a espécie *Leishmania* (PDB ID 1lml).

Para *L. (V.) braziliensis*, a busca em banco de dados de estrutura de proteínas retorna 6 proteínas, das quais 1 Dihydroorotate dehydrogenase, de código 4WZH (Reis, Lorenzato *et al.* 2015), 1 Nucleoside diphosphate kinase b, de código 4KPC (Vieira, de Giuseppe *et al.* 2015), 2 Peroxidoxin, de códigos 4KB3 (dado ainda não publicado) e 4KCE (dado ainda não publicado), 1 6-phosphogluconolactonase, de código 3CH7 (dado ainda não publicado) e 1 eukaryotic initiation factor 5a, de código 1X6O (dado ainda não publicado).

Portanto, para o estudo do comportamento do sítio catalítico das metaloproteases de *L. (V.) braziliensis* é necessário a utilização de métodos computacionais que podem prever a estrutura dessas proteínas baseada somente em sua sequência de aminoácidos.

**Tabela 1:** Códigos de acesso das estruturas tridimensionais que contêm pelo menos o domínio catalítico dos membros pertencentes ao clã metzincina depositadas no banco de dados público, PDB – Protein Data Bank. (Gomis-Rüth 2009)

Proteína	Código PDB
<b>ASTACINAS</b>	
Astacina ( <i>Astacus astacus</i> )	1ast 1iaa 1iab 1iac 1iad 1iae 1qji 1qjj
Tolloid-like protease 1 ( <i>Homo sapiens</i> )	3edi
Bone morphogenetic protein 1 ( <i>H. sapiens</i> )	3edh 3edg
<b>SERRALISINAS</b>	
Aeruginolisina ( <i>Pseudomonas aeruginosa</i> )	1kap 1jiw 1akl
Alkaline protease ( <i>Pseudomonas</i> sp. TACII 18)	1g9k 1h71 1omj 1om6 1om7 1om8 1o0t 1o0q
Serralisina ( <i>Serratia marcescens</i> )	1af0 1sat 1srp 1smp
PrtC ( <i>Erwinia chrysanthemi</i> )	1go7 1go8 1k7g 1k7i 1k7q
<b>ADAMALISINA/ADAMs</b>	
Adamalisina II ( <i>Crotalus adamanteus</i> )	1iag 2aig 3aig 4aig
Atrolisina-C (Ht-d/c) ( <i>Crotalus atrox</i> )	1atl 1dth 1htd
Bothropasin ( <i>Bothrops jararaca</i> )	3dsl
Taiwan habu MP ( <i>Trimeresurus mucrosquamatus</i> )	1kuf 1kug 1kui 1kuk
H2-proteinase ( <i>Trimeresurus flavoviridis</i> )	1wni
Acutolisina-C ( <i>Agkistrodon acutus</i> )	1qua
Acutolisina-A ( <i>A. acutus</i> )	1bsw 1bud
BaPI ( <i>Bothrops asper</i> )	1nd1
Vascular apoptosis-inducing protein1 VAP1 ( <i>C. atrox</i> )	2ero 2erp 2erq
Fibrin(ogen)olytic MP ( <i>A. acutus</i> )	1yp1
Catrocolastatina/VAP2B ( <i>C. atrox</i> )	2dw0 2dw1 2dw2
Russel's viper venom MP ( <i>Daboia russelli siamensis</i> )	2e3x
TNF- $\alpha$ -converting enzyme (ADAM-17) ( <i>H. sapiens</i> )	1bkc 3edz 3b92 3cki 2oi0 2i47 2fv9 2fv5 2ddf 2a8h 1zxc
ADAM-33 ( <i>H. sapiens</i> )	1r54 1r55
ADAMT-1 ( <i>H. sapiens</i> )	2jih 2v4b
ADAMT-4 ( <i>H. sapiens</i> )	2rjp 3b2z
ADAMT-5 ( <i>H. sapiens</i> )	3b8z 2rjp 2rjq
<b>MATRIXINAS/MMPs</b>	
Fibroblast collagenase (MMP-1) ( <i>H. sapiens</i> ; <i>Sur scrofa</i> )	966c 1ayk 1cge 1cgf 1cgl 1fbl 1hfc 1mnc 2ayk 2tcl 3ayk 4ayk 1su3 2j0t 2clt
Gelatinase A (MMP-2) ( <i>H. sapiens</i> )	1ck7 1hov 1qib 1gxd 1eak

Proteína	Código PDB
MATRIXINAS/MMPs (cont.)	
Stromelisina-1 (MMP-3) ( <i>H. sapiens</i> )	1b3d 1b8y 1biw 1bm6 1bqo 1c3i 1caq 1ciz 1cqr 3usn 1uea 2d1o 1c8t 2jt6 2jt5 2jnp 1d5j 1d7x 1d8f 1d8m 1g05 1g49 1g4k 1hfs 1hy7 1sln 1slm 1qia 1qic 1ums 1umt 1usn 2srt 2usn
Matrilisina (MMP-7) ( <i>H. sapiens</i> )	1mmp 1mmq 1mmr 2ddy
Neutrophil collagenase (MMP-8) ( <i>H. sapiens</i> )	1jan 1a85 1a86 1bzs 1i73 1i76 1jao 1jap 1jaq 1jh1 1jj9 1kbc 1mmb 1mnc 1zvx 2oy4 1zs0 1zp5 2oy2 1jh1
Gelatinase B (MMP-9) ( <i>H. sapiens</i> )	1gkc 1gkd 1l6j 2ow0 2ow1 2ow2 2ovz 2ovx
Stromelisina-2 (MMP-10) ( <i>H. sapiens</i> )	1q3a
Stromelisina-3 (MMP-11) ( <i>Mus musculus</i> )	1hv5
Macrophage elastase (MMP-12) ( <i>H. sapiens</i> )	1jiz 1jk3 2z2d 2k2g 3ba0 2oxu 2oxz 2oxw 2hu6 1z3j 1ycm 1y93 1rmz 1os9 1os2 2poj 1utz 1utt 1ros
Collagenase-3 (MMP-13) ( <i>H. sapiens</i> ; <i>Mus musculus</i> )	456c 830c 1cxv 1eub 1fls 1fm1 2pjt 2ozr 2ow9 2e2d 2d1n 1ztq 1you 1xur 1xud 1xuc
MT1-MMP (MMP-14) ( <i>H. sapiens</i> )	1bqq 1buu
MT3-MMP (MMP-16) ( <i>H. sapiens</i> )	1rm8
Enamelisina (MMP-20) ( <i>H. sapiens</i> )	2jsd
SNAPALISINAS	
<i>Streptomyces caespitosus</i> neutral protease	1c7k 1kuh
LEISHMANOLISINAS	
<i>Leishmania major</i> leishmanolisina	1lml
PAPALISINAS	
<i>Methanosarcina acetivorans</i> ulilisina	2cki 2j83

## 1.5 Predição de estrutura tridimensional de proteínas

A predição da estrutura tridimensional de uma proteína e seu padrão de enovelamento fornecem informações relevantes sobre sua provável função biológica e, conseqüentemente, na identificação desta como um potencial alvo farmacológico. Já é bem estabelecido que a evolução tende a conservar funções que dependem mais diretamente das estruturas tridimensionais do que das próprias similaridades entre as sequências de aminoácidos das proteínas (Sánchez, Pieper *et al.* 2000, Cherkasov and Jones 2004). O enovelamento proteico permite que aminoácidos sequencialmente distantes interajam em uma mesma região espacial da proteína, formando distintos padrões conformacionais, dependendo das condições físico-químicas do ambiente.

A determinação de estruturas por métodos experimentais é um trabalho laborioso, custoso e com algumas limitações, já que nem sempre é possível ser realizado para todas as proteínas. Por outro lado, métodos de modelagem comparativa têm tido êxito na predição de estruturas 3D. Porém, a aplicação destes métodos só é possível para proteínas que possuem estruturas similares às aquelas já conhecidas por métodos experimentais. Por esta razão, o número de sequências conhecidas é bem maior que o número de estruturas tridimensionais.

Para verificar e avaliar a qualidade e eficiência dos métodos de predição de estruturas existentes, uma reunião bianual CASP (*Comparative Assessment of Methods for Protein Structure Prediction*) é realizada desde 1994 (Moult, Fidelis *et al.* 2009). Sequências proteicas são escolhidas como alvo e são organizadas para serem modeladas por dois principais métodos: baseado em estrutura molde, que incluem modelagem comparativa e *threading* e; métodos independentes de estrutura molde, chamados de primeiros princípios, que englobam os métodos *de novo* e *ab initio*. Ao final é divulgado o desempenho dos programas utilizados.

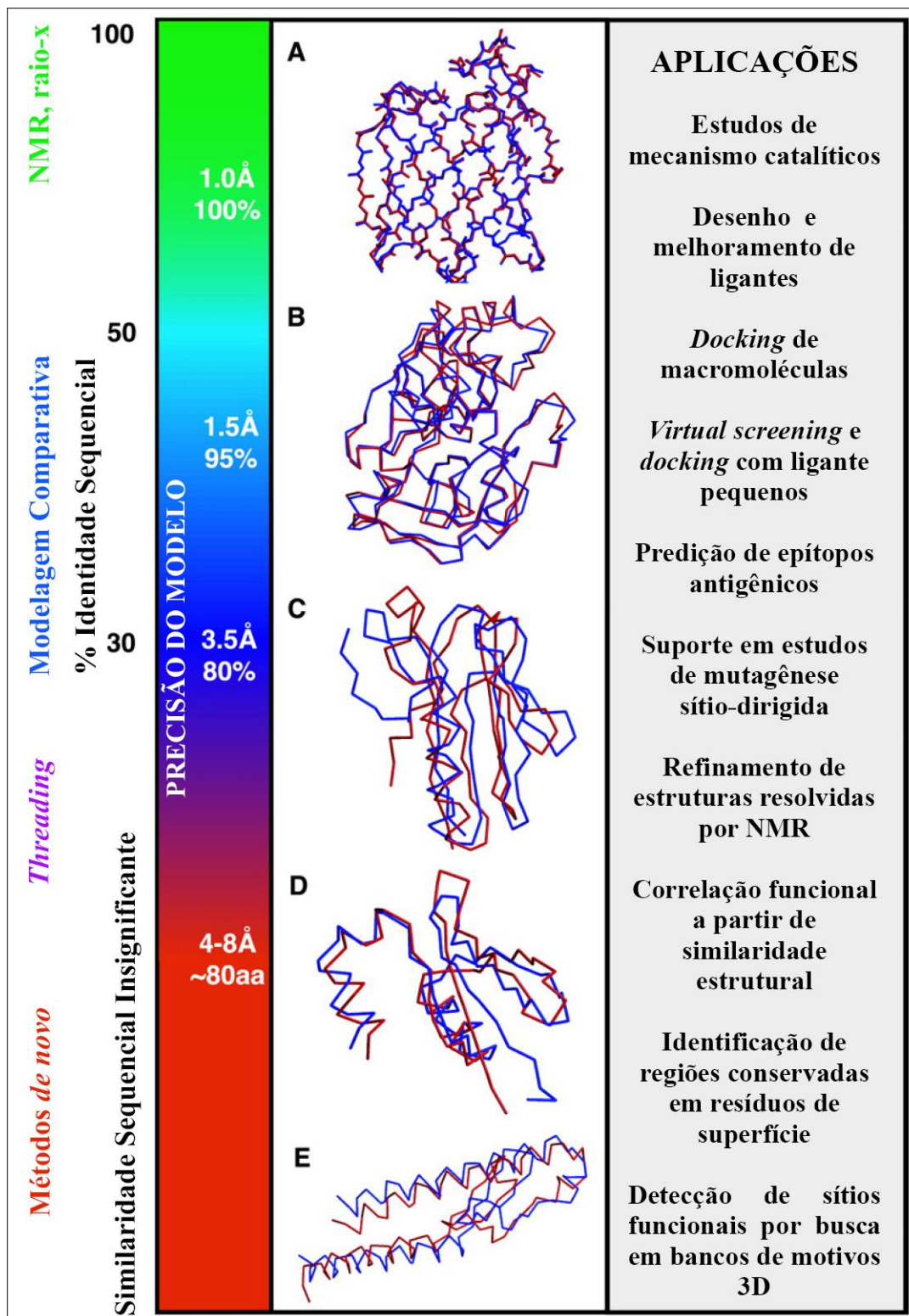
Para escolher o método a ser aplicado na construção de um modelo, é necessário verificar se há estruturas resolvidas experimentalmente que possam servir como molde para a sequência alvo a ser modelada.

### **1.5.1 Modelagem Comparativa**

A modelagem comparativa é um método empírico que utiliza o alinhamento construído entre a sequência da proteína a ser modelada (sequência alvo) e a sequência de uma proteína relacionada estruturalmente (molde) (Sánchez and Šali 1999). O seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. Geralmente, consideram-se como limites mínimos de aplicabilidade do método o valor de 30% de identidade entre a sequência da proteína de interesse e o molde, para proteínas com um mínimo de 100 aminoácidos na sua composição (Figura 8).

O processo de obtenção de um modelo virtual de proteína através de modelagem comparativa envolve quatro etapas, dentre elas: (i) identificação dos moldes, (ii) alinhamento das sequências, (iii) construção do modelo e (iv) validação dos mesmos (Sternberg 1996). Dentre os mais utilizados métodos de modelagem comparativa para a construção do modelo, estão o método de modelagem pela satisfação de restrições espaciais e o método de modelagem pela união de corpos rígidos.





**Figura 8:** Acurácia e aplicação de modelos de estrutura de proteínas. São mostrados os diferentes intervalos de aplicabilidade da modelagem comparativa, threading, e métodos de novo; a acurácia correspondente dos modelos estruturais de proteínas; e exemplos de aplicações. (A a C) Exemplos de modelos comparativos baseados em cerca de 60% (A), 40% (B), e 30% (C) de identidade sequencial com a sua estrutura molde. (D e E) Exemplos de predições do método de novo com software Rosetta para o experimento CASP4. Predições de estruturas estão em vermelho, e estruturas reais em azul. Adaptado de Baker e Sali, 2001.

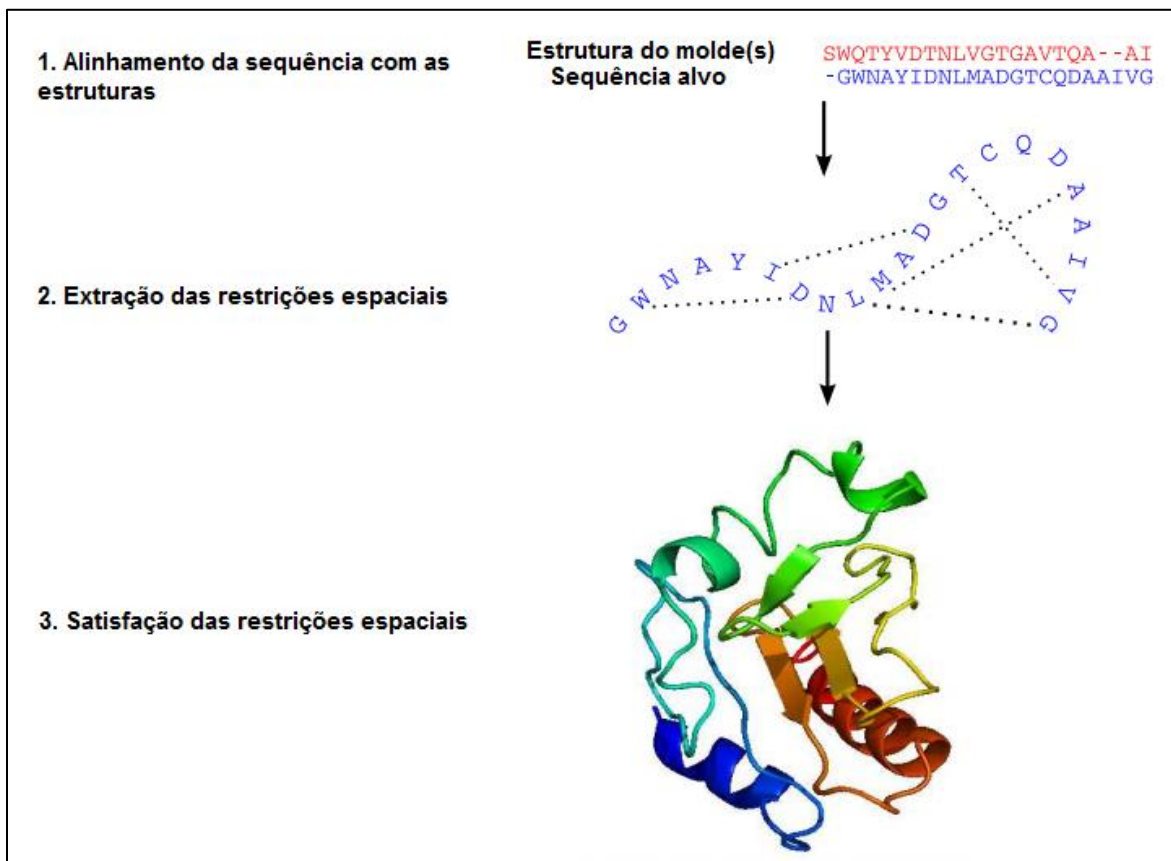


### 1.5.1.1 Satisfação de Restrições Espaciais

Este método é realizado mediante a geração de um conjunto de restrições espaciais que são aplicadas à sequência alvo (Figura 9). Dentre as restrições utilizadas pelos programas, estão as aplicadas nas ligações entre três átomos (ângulos planos) e entre quatro átomos (ângulos diedrais). Além destas restrições, um campo de força controla as propriedades estereoquímicas entre os átomos, aplicando restrições físico-químicas (Rössle 2004). Todas as restrições físico-químicas e espaciais aplicadas ao modelo são combinadas em uma função, que é minimizada durante o processo de construção do modelo (Fiser and Sali 2003).

As restrições espaciais podem vir de diferentes tipos de fontes, como: estruturas homólogas, refinamento por RMN (Ressonância Magnética Nuclear), regras de empacotamento de estruturas secundárias, reconstrução por imagem em microscopia eletrônica, potenciais de campos de força de átomo-átomo e resíduo-resíduo, dentre outros.

Programas que utilizam esse método fazem o alinhamento entre as sequências alvo e de referência. A partir do resultado do alinhamento, as restrições espaciais, como distâncias e ângulos, são extraídas do molde e são transferidas para o modelo. O tamanho das ligações bem como seus ângulos preferenciais são obtidos de campos de força e os dados para interações entre os átomos não ligados são obtidos por análise estatística de um grupo representativo de estruturas conhecidas.



**Figura 9:** Modelagem comparativa usando o método de satisfação de restrições espaciais do programa Modeller (Šali and Blundell 1993). Adaptado do manual do programa.

### 1.5.1.2 União de Corpos Rígidos

Nesta metodologia, a construção da proteína alvo ocorre por meio das transferências de coordenadas atômicas dos moldes utilizados. Após o alinhamento entre as sequências alvo e molde, os resíduos são sobrepostos e trocados, mas as coordenadas espaciais são mantidas (Goldsmith-Fischman and Honig 2003).

Essa técnica baseia-se na divisão das estruturas das proteínas em três conjuntos: cadeias laterais, região de voltas (loop) e as regiões conservadas. Dessa forma, cada região da molécula participa separadamente na construção do modelo, no entanto a geometria molecular é preservada considerando-se um conjunto de ligações ou distâncias entre os sítios da molécula, cujos cumprimentos são fixos (Morgon and Coutinho 2007).

Programas que utilizam esse método, como o Swiss-Model, fazem a construção do modelo baseados no fato de que proteínas homólogas possuem regiões estruturalmente conservadas como  $\alpha$ -hélices e folhas- $\beta$ . As regiões conservadas são definidas através da predição de estruturas secundárias. Após a

definição, o alinhamento entre as estruturas é realizado considerando a média das posições dos C $\alpha$  das sequências de aminoácidos das regiões conservadas nas sequências de referência. As regiões que não satisfazem as exigências são chamadas de regiões variáveis que geralmente são regiões de voltas que conectam as regiões conservadas.

### **1.5.1.3 Principais Programas de Modelagem Comparativa**

Atualmente existem diversos servidores online/programas, que utilizam a metodologia de modelagem comparativa. Entre eles, podem-se citar Fugue (Shi, Blundell *et al.* 2001), HHpred (Söding, Biegert *et al.* 2005), I-Tasser (Zhang 2008, Roy, Kucukural *et al.* 2010), Modeller (Šali and Blundell 1993) e, sua versão automatizada, ModPipe e Swiss-Model (Arnold, Bordoli *et al.* 2006).

### **1.5.2 Métodos *de novo***

De forma a manter a independência de moldes de proteínas homólogas, foram desenvolvidos os chamados métodos *de novo*, que usam informações provenientes de bancos de estruturas de proteínas determinadas empiricamente, sem a necessidade de haver identidade significativa com a sequência alvo.

A estratégia utilizada neste método é baseada em dois pressupostos: (i) todas as informações necessárias sobre a estrutura de uma proteína estão contidas em sua sequência de aminoácidos; e (ii) acredita-se que as proteínas se enovelam para um estado nativo, ou um conjunto de estados nativos, que se encontra no (ou próximo ao) mínimo global de energia livre.

#### **1.5.2.1 Programas que usam o método *de novo***

Como exemplo, pode-se citar alguns dos servidores que usam o método *de novo* como QUARK (Xu and Zhang 2012) e ROBETTA (Chivian, Kim *et al.* 2005). O primeiro realiza a predição da estrutura tridimensional de proteínas sem a utilização de moldes. A criação do modelo é dividida em três passos: (i) obtenção do perfil da sequência por meio da predição de estrutura secundária, valores de ângulos de torção, acessibilidade ao solvente e posições de  $\beta$ -turn seguido da geração de fragmentos da sequência de consulta entre 1-20 resíduos de

comprimento; (ii) simulações para a construção da estrutura utilizando o método “replica-exchange Monte Carlo” guiada pelo campo de força QUARK baseado em conhecimentos de nível atômico e (iii) agrupamento dos fragmentos e refinamento *full-atom*.

O servidor ROBETTA funciona de forma semelhante ao QUARK. O protocolo por ele empregado utiliza fragmentos de estruturas existentes no PDB a fim de orientar um conjunto de fragmentos com funções de energia. Esse protocolo chamado de “reconstrói e refina” é composto por uma etapa que emprega um modelo simplificado, chamado modelagem de baixa resolução, seguida de uma etapa que utiliza um modelo *full-atom*, chamado modelagem de alta resolução (Bonneau, Tsai *et al.* 2001).

### **1.5.3 Threading**

O método de *threading* ou *folding recognition*, como também é conhecido, é um método de predição que considera que o número de dobramentos possíveis é finito e ao invés de procurar por todo o espaço conformacional, se limita a buscar pelo melhor dobramento possível, dentre as proteínas que possuem estrutura conhecida (Leach 2001). Para tanto, os métodos de *threading* alinham uma sequência proteica de busca diretamente nas estruturas tridimensionais conhecidas, buscando assim encontrar o dobramento da proteína de interesse.

As limitações deste método surgem de dois pontos principais: (i) assim como na modelagem comparativa, se a identidade entre as sequências alvo e molde for muito baixa, é possível que o enovelamento daquela sequência simplesmente não corresponda à estrutura nativa da sequência de interesse, podendo dar lugar a um modelo completamente errado; (ii) os modelos gerados com este método apresentam uma resolução relativamente baixa, dificultando seu uso em estudos que exigem posicionamento preciso dos átomos como no caso do *docking* molecular de ligantes no sítio da proteína (Gibas and Jambeck 2001).

#### **1.5.3.1 Programas que usam método *Threading***

Dentre os programas que utilizam o método *Threading*, pode-se citar I-Tasser (Iterative Threading Assembly Refinement) (Roy, Kucukural *et al.* 2010),

MUSTER (MUlti-Source ThreadER) (Wu and Zhang 2008), LOMETS (Local Meta-Threading Server) (Wu and Zhang 2007) e SPARKS-X (Yang, Faraggi *et al.* 2011).

O servidor I-Tasser combina os métodos *threading*, *ab initio* e refinamento estrutural para construção de estruturas tridimensionais. O MUSTER é um servidor que usa um algoritmo *threading* empregando uma extensão de “sequence profile-profile alignment” (PPA), que combina uma variedade de sequências e recursos estruturais gerados por meio de várias ferramentas como acessibilidade ao solvente, ângulos de torção, dentre outros. O LOMETS utiliza a técnica *meta-threading*, ou seja, gera as predições de estruturas tridimensionais a partir de modelos consenso de uma variedade de servidores individuais e o SPARKS-X é uma otimização do método *threading* SPARKS que altera a função de pontuação no alinhamento e incorpora técnicas que realizam melhores predições da estrutura secundária, ângulos de torção e superfície acessível ao solvente.

## 1.6 Otimização da Geometria

Proteínas obtidas por modelagem comparativa necessitam de refinamento (Holtje, Folkers *et al.* 1998), pois, durante a construção dos modelos, as conformações da cadeia lateral e os loops podem ser criados aleatoriamente. Com isso, pode haver necessidade de ajustes para corrigir comprimentos de ligação e interações atômicas desfavoráveis produzidas entre átomos muito próximos. O processo para se estabelecer uma condição energeticamente favorável consiste em adequar o modelo ao campo de força que o descreve mediante otimização da geometria.

A otimização da geometria é uma técnica que visa encontrar um conjunto de coordenadas que minimizam a energia potencial do sistema de interesse. Devido ao elevado número de graus de liberdade das macromoléculas biológicas, uma exploração completa da superfície multidimensional de energia é praticamente impossível. Dois algoritmos são os mais usados para minimização de energia: máximo declive (*steepest-descent*) e gradiente conjugado (Lipkowitz and Boyd 1997).

Em sistemas macromoleculares, o algoritmo *steepest-descent* se mostra mais efetivo e rápido nos passos iniciais quando se está muito longe do mínimo e utiliza somente a informação do gradiente na coordenada atualizada. O segundo algoritmo utilizado na técnica de otimização da geometria é o método dos

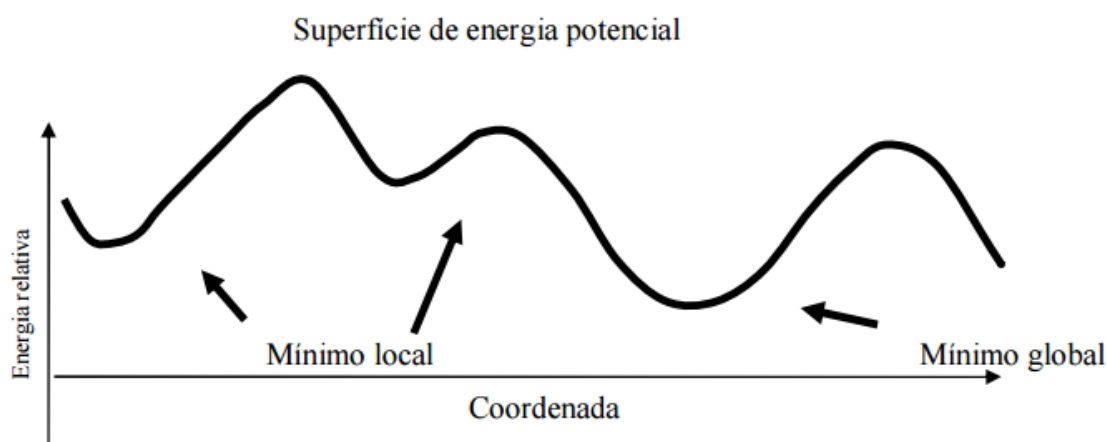
gradientes conjugados. Este método permite, em geral, uma exploração da superfície mais completa do que o método *steepest-descent*.

O método do gradiente conjugado utiliza para determinação do passo seguinte, além do valor do gradiente no ponto atual, o valor do gradiente obtido no passo anterior. Com isso, o algoritmo evita que se tome várias vezes uma mesma direção, tornando o método mais preciso. Por isso, é comum utilizarmos os dois métodos para otimização da geometria antes da dinâmica molecular (Lipkowitz and Boyd 1997).

Quando uma estrutura é gerada a partir das técnicas de predição da estrutura tridimensional, uma técnica mais apurada que a minimização da geometria deve ser realizada neste novo modelo. Isso se deve ao fato que a geometria obtida pelo processo de otimização pode representar uma provável adaptação encontrada pelo algoritmo na região de mínimo local.

Portanto, para encontrar a estrutura 3D mais favorável energeticamente é necessário investigar o comportamento conformacional através da superfície de energia potencial para que se possa aproximar o modelo o máximo possível da região de mínimo global (Figura 10).

Previamente à execução das simulações por dinâmica molecular é necessário realizar a minimização dos sistemas moleculares para eliminar deformações na geometria da molécula. Caso um sistema seja submetido sem antes ser realizada a minimização de energia, este poderá apresentar temperaturas irreais nos primeiros passos da dinâmica e o sistema poderá visitar configurações indesejáveis do espaço de fase (Pascutti 2002).



**Figura 10:** Superfície de energia potencial para simulação de dinâmica molecular. Retirado de (Pacheco 2009).

## 1.7 Dinâmica Molecular

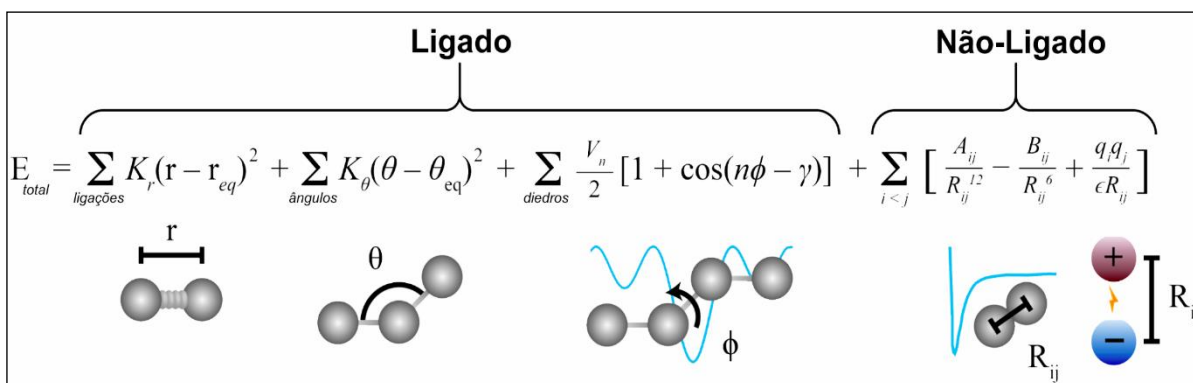
A Dinâmica molecular (DM) é um método de simulação computacional que utiliza um campo de forças clássico para descrever as interações moleculares a partir da resolução numérica das equações clássicas de movimento. A força que atua sobre cada átomo é obtida a partir das informações presentes nos campos de força clássicos.

Genericamente atribui-se a denominação de campo de força ao conjunto de equações que determinam como a energia potencial de um sistema modifica com a posição de seus átomos. A função energia potencial de uma molécula é calculada por campos de força que podem considerar todos os átomos da proteína em seus cálculos (*full-atom*), como por exemplo, GROMOS (Scott, Hünenberger *et al.* 1999), AMBER (Pearlman, Case *et al.* 1995), OPLS (Jorgensen and Tirado-Rives 1988) e CHARMM (Brooks, Brucoleri *et al.* 1983), ou campos de força que não consideram todos os átomos (*coarse-grained*), como é o caso do MARTINI (Marrink, Risselada *et al.* 2007).

Cada campo de força estabelece um conjunto de equações matemáticas dedicados a reproduzir aspectos do comportamento molecular, como o estiramento de ligações químicas, a deformação de um ângulo de ligação ou a torção de um diedro. Estas equações, por sua vez, são parametrizadas para reproduzir o comportamento dos compostos de interesse (Verli 2014). Os campos de força clássicos possuem duas classes de funções potenciais (Figura 11):

(i) interações ligadas são representadas por ligações covalentes (potencial de estiramento), dobramento angular, diedros próprios e impróprios.

(ii) interações não-ligadas são computadas baseando-se em uma lista de átomos vizinhos não ligados, dentro de um certo raio de corte. Essas ligações consistem no potencial de *Lennard-Jones* e interação de *Coulomb*.



**Figura 11:** Exemplo de uma equação empregada para aproximar as forças atômicas que governam o movimento molecular. As interações atômicas que governam o movimento molecular podem ser divididas entre aquelas causadas por átomos que estão quimicamente ligados uns aos outros, e as causadas por átomos que não estão ligados. As ligações químicas e ângulos atômicos são modelados usando molas simples, e ângulos diedros (ou seja, as rotações sobre uma ligação) são modelados usando uma função sinusoidal que aproxima as diferenças de energia entre conformações eclipsadas e escalonados. Interações não-ligadas surgem devido a forças de Van der Waals que são modeladas utilizando o potencial de Lennard-Jones; e interações eletrostáticas são modeladas utilizando a lei de Coulomb. Adaptado de Durrant e McCammon, 2011.

Para cada sistema é aconselhável analisar qual campo de força é o mais apropriado porque cada um contém um conjunto de potenciais empíricos específicos para cada sistema (Leach 2001).

Campos de força AMBER utilizam o potencial de *Lennard-Jones* mais cargas pontuais para átomos soltos ou distantes; para átomos diferentes é utilizado ainda regras de mistura de *Lorentz-Berthelot*. Interações intramoleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedrais do tipo impróprio. Átomos conectados por três ligações apresentam tratamento misto: são regidos por *Lennard-Jones* e potencial de *Coulomb* (Lindorff-Larsen, Piana *et al.* 2010).

O campo de força CHARMM utiliza as mesmas regras de potenciais não ligados que o AMBER. As diferenças são que as interações intramoleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedros do tipo impróprio, além de termo não escalados do tipo *Coulomb*. Torções angulares do tipo impróprias são utilizadas em casos especiais com fins de restrição de planaridade de resíduos aromáticos (MackKerell, Feig *et al.* 2004).

Já o campo de força GROMOS é um campo de força baseado na unificação dos átomos. Neste caso, os átomos de hidrogênio apolares, ou seja, aqueles ligados a átomos de carbono, são unidos a este elemento, dando origem a um



pseudo-átomo representando as propriedades de grupos CH, CH<sub>2</sub> ou CH<sub>3</sub>. Exceção se dá para o grupo CH de anéis aromáticos, que tem os átomos de hidrogênio descritos explicitamente. É parametrizado para usar as mesmas formas funcionais que AMBER, sem a regra geral de mistura de *Lennard-Jones*. As relações intramoleculares são diferenciadas. Flexibilidade das ligações é produzida pelo potencial de estiramento de quarta potência, enquanto ângulos se retorcem com torções do tipo próprias e os diedrais são tratados pela série de cossenos, *Lennard-Jones* e termos de *Colulomb* não escalados (Oostenbrink, Villa *et al.* 2004).

A solução numérica das equações de movimento a partir de coordenadas e velocidades iniciais, para o movimento de cada átomo submetido ao campo de força, é obtida através de um algoritmo computacional (Allen and Wilson 1989, Frenkel and Smit 1996). Todos eles baseiam-se na resolução das equações de Newton a partir de uma expansão em série de potenciais de Taylor da posição de cada partícula e em cada incremento no tempo. Um algoritmo utilizado é o de Verlet que resolve as equações de movimento de Newton para cada átomo e em cada incremento no tempo onde as novas posições dos átomos são calculadas levando em conta as posições dos átomos no instante de tempo anterior e o valor da força no tempo atual.

A avaliação das forças para obtenção das acelerações é o processo que mais consome tempo computacional na dinâmica molecular. O tempo gasto no cálculo das forças depende da complexidade da função de energia potencial. Um outro algoritmo também utilizado em dinâmica molecular é o “*summed Verlet*” ou “*leapfrog*”. O algoritmo *leapfrog* apresenta menor probabilidade a erros numéricos, sendo um dos algoritmos mais estáveis, simples e eficientes. Além disso, pode ser aplicado em sistemas constituídos de fluidos simples a biopolímeros. A força que atua sobre cada átomo é obtida a partir das informações presentes nos campos de força clássicos (Kukol 2008) e das velocidades nos tempos intermediários.

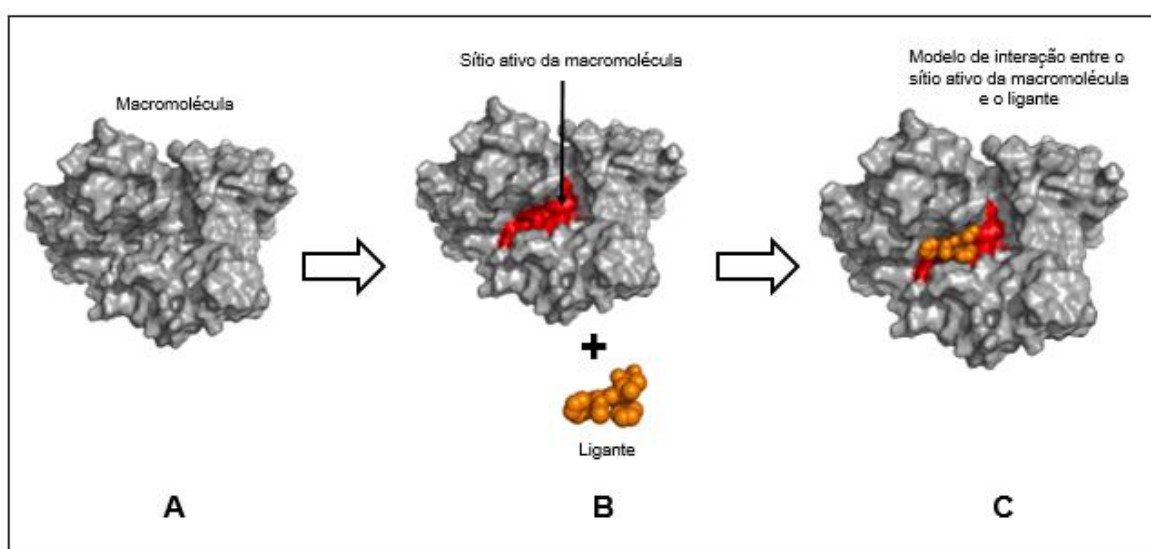
A simulação por Dinâmica Molecular abre a possibilidade de um estudo mais detalhado dos sistemas de macromoléculas com análise na escala microscópica e em tempos da ordem de nanossegundos. Com esta metodologia simula-se o comportamento dinâmico dos sistemas sendo possível o acompanhamento da evolução temporal dos mesmos como também a determinação de propriedades dinâmicas, estruturais e termodinâmicas.

## 1.8 Métodos de Atracamento Molecular (*Docking*)

Os métodos de atracamento molecular podem proporcionar uma melhor compreensão da interação entre proteínas e outras moléculas. Dentre as possíveis aplicabilidades do método, se destaca a predição de se ligantes ainda não testados experimentalmente podem interagir favoravelmente com a macromolécula. Obter uma descrição acurada e automatizada do processo de reconhecimento molecular, utilizando metodologias computacionais, pode permitir a redução do tempo e dos altos custos envolvidos no desenvolvimento de novos medicamentos (Morgon and Coutinho 2007, Jorgensen 2009).

O sucesso da técnica pode ser medido comparando os resultados preditos com os modos de ligação já conhecidos determinados por métodos experimentais como cristalografia de raios-x ou ressonância magnética nuclear, e medidas de afinidade determinadas em ensaios *in vitro*. Para a realização do atracamento molecular, três etapas são necessárias (Figura 12): definição da estrutura da molécula alvo, localização do sítio de ligação e predição do modo e da afinidade de ligação de um ligante utilizando algoritmos específicos (Morgon and Coutinho 2007).

A predição do modo e a afinidade de ligação é realizada através de algoritmos de busca e funções de avaliação, dois aspectos principais que diferenciam os programas de atracamento molecular.



**Figura 12:** Etapas do atracamento molecular. Em A, observa-se a macromolécula, em B, o sítio ativo da macromolécula e o ligante e em C, o modelo de interação entre o sítio da macromolécula e o ligante.

Os métodos de busca são usados para amostrar as possíveis orientações do ligante ligado ao alvo proteico, considerando os graus de liberdade translacionais, rotacionais e conformacionais. Eles são divididos em 3 categorias: busca sistemática, busca determinística e busca estocástica. A busca sistemática consiste em explorar combinatoriamente todos os graus de liberdade da molécula. Um exemplo desta busca são os algoritmos conhecidos como construção incremental ou baseados em fragmentos. A busca determinística é caracterizada por conseguir reproduzir sempre a mesma conformação de saída dados os mesmos parâmetros e coordenadas de entrada; esta busca pode ser exemplificada pelos métodos clássicos de simulação por dinâmica molecular e minimização de energia. Já a busca estocástica, como o próprio nome sugere, é caracterizada por utilizar variáveis aleatórias para produzir diferentes conformações de saída para um mesmo conjunto de parâmetros de entrada. Isso ocorre, pois, os algoritmos utilizam “sementes” diferentes para a geração de números pseudo-aleatórios. Essa busca pode ser exemplificada pelos métodos de Monte Carlo, *Simulated Annealing* e algoritmos evolucionários (Morgon and Coutinho 2007).

As funções de avaliação também podem ser divididas em três classes: funções baseadas em campo de força, funções empíricas e funções baseadas em conhecimento (Kitchen, Decornez *et al.* 2004).

As funções baseadas em campos de força utilizam métodos de mecânica molecular para quantificar a energia proveniente das interações entre o ligante e receptor, assim como a energia interna do ligante. As funções G-score e D-score, implementadas no pacote de programas SYBYL (Meng, Shoichet *et al.* 1992, Jones, Willett *et al.* 1997) e a função GOLDScore, implementada no programa GOLD (Verdonk, Cole *et al.* 2003), são exemplos de funções baseadas em campo de força.

As funções empíricas são baseadas no ajuste de equações provenientes de campos de força, a partir de dados experimentais, ou na contagem de interações, visando calcular com maior exatidão a energia livre de interação do ligante com seu alvo macromolecular. F-score, ChemScore e X-score são alguns exemplos de funções empíricas.

As funções baseadas no conhecimento utilizam dados estatísticos dos potenciais de pares atômicos de interação, os quais são derivados de conjuntos de dados proveniente de complexos cristalográficos entre proteínas e ligantes para descrever as interações entre ligante e proteína.

Apesar das simplificações empregadas, as funções de avaliação implementadas nos programas de *docking* são úteis nos processos de descoberta de novos compostos com atividade biológica e na elucidação dos prováveis modos de interação entre compostos já existentes e macromoléculas. Programas mais recentes utilizam a combinação dessas funções para produzirem funções consenso de forma a reduzir os erros associados a elas e com isso aumentar a qualidade dos resultados, melhorando uma pose no sítio de ligação ou permitindo uma maior interação desse ligante com a macromolécula (Morgon and Coutinho 2007).

A flexibilidade das moléculas é uma questão importante a ser levada em consideração. Os primeiros algoritmos que faziam simulações de *docking*, tratavam tanto o receptor quanto o ligante como moléculas rígidas considerando apenas os graus de liberdade translacionais e rotacionais da molécula ligante. Atualmente, existem programas que consideram a flexibilidade total do ligante, flexibilidade parcial do receptor e alguns casos apresenta alternativas de flexibilidade do receptor, como simulações curtas de dinâmica molecular.

Nas últimas duas décadas, diversos programas de *docking* têm sido propostos tanto para uso comercial quanto para uso acadêmico. Dentre os programas existentes, podem-se citar os mais conhecidos: Autodock (Goodsell, Morris *et al.* 1996), Autodock Vina (Trott and Olson 2010), Glide (Friesner, Banks *et al.* 2004), GOLD (Verdonk, Cole *et al.* 2003), FlexX (Rarey, Kramer *et al.* 1996), Surflex (Spitzer and Jain 2012) e DockThor (de Magalhães, Barbosa *et al.* 2012).

## 2 OBJETIVOS

### 2.1 Objetivo Geral

Caracterização estrutural de metaloproteases de *Leishmania (V.) braziliensis*, com destaque para o comportamento dinâmico de seus sítios catalíticos.

### 2.2 Objetivos Específicos

- Investigar as diferenças entre metaloproteases referente ao cromossomo 10 da família M8 de *Leishmania (V.) braziliensis* mediante utilização de metodologias computacionais;
- Analisar a estabilidade dos modelos mediante dinâmica molecular na procura de diferenças substanciais no arranjo de resíduos no sítio ativo;
- Identificar o modo de ligação das metaloproteases modeladas com inibidores conhecidos;
- Eleger os inibidores que obtiverem melhor resultado na interação com as metaloproteases a partir da energia de interação visando conseguir fármacos mais potentes e com especificidade.

## 3 MATERIAIS E MÉTODOS

### 3.1 Sequências de metaloproteases de *L. (V.) braziliensis*

As sequências de metaloproteases foram obtidas do banco de dados GeneDB<sup>2</sup> (Logan-Klumpler, De Silva *et al.* 2011) onde está depositado o genoma de *L. (V.) braziliensis*. No total, 89 sequências foram extraídas (Figura 3), dentre estas, 36 estão presentes na família M8 e 33 dessas, pertencem ao cromossomo 10 (Silva-Almeida, Souza-Silva *et al.* 2014).

As 33 sequências do cromossomo 10 que serão estudadas neste trabalho estão presentes nos alelos: 0470, 0480, 0490, 0500, 0510, 0520, 0530, 0540, 0550, 0560, 0570, 0580, 0590, 0600, 0610, 1540, 1550, 1560, 1570, 1580, 1590, 1610, 1620, 1630, 1640, 1650, 1660, 1670, 1680, 1690, 1700, 1710 e 1720.

As sequências foram alinhadas utilizando o servidor T-coffee Expresso (Armougom, Moretti *et al.* 2006) e verificamos que os alelos 0490, 1540, 1700 e 1720 não possuem o motivo conservado identificador das proteínas dessa classe e, portanto, foram excluídas do conjunto a ser analisado. Dessa forma, reduzimos o estudo para 29 proteínas.

### 3.2 Predição da Localização Celular das metaloproteases

A predição da localização celular das metaloproteases foi realizada com os servidores, TopPred v. 1.1 (Von Heijne 1992, TopPred 1994) e CELLO (subCELLular LOcalization predictor) (Yu, Lin *et al.* 2004, Yu, Chen *et al.* 2006). O servidor TopPred indica as regiões extracelulares, transmembranares, citoplasmática e loops de uma dada proteína. Este servidor gera uma ou mais estruturas com as respectivas probabilidades de acerto na predição. Dentre os valores de confiança resultantes da predição, usamos àquele que apresentou maior confiança.

Já o servidor CELLO, indica a possível localização da proteína na célula a partir de um valor de confiabilidade da predição. Ambos os servidores possuem um banco de dados com informações de proteínas de localização celular conhecidas e utiliza-as para a predição das que não conhece.

---

<sup>2</sup> [www.genedb.org](http://www.genedb.org)

### 3.3 Predição da estrutura tridimensional de metaloproteases

A construção dos modelos tridimensionais das metaloproteases de *L. (V.) braziliensis* foi realizada empregando o método de modelagem baseada em molde tanto na forma manual quanto automatizada. As 29 sequências pertencentes ao cromossomo 10, foram submetidas ao servidor online BLAST, que faz uma busca contra o banco de dados de proteínas (PDB), a fim de encontrar proteínas similares que possam ser utilizadas como molde na modelagem comparativa. O desenho esquemático dessa etapa está demonstrado na Figura 13.

Uma vez definido o molde, as 29 sequências foram ranqueadas seguindo três critérios: (i) as sequências alvo que possuem maior porcentagem de cobertura; (ii) maior porcentagem de identidade e; (iii) maior porcentagem de similaridade com a proteína molde. A fim de avaliar qual dos programas/servidores de modelagem teriam um melhor resultado sob o conjunto de proteínas, 10% das sequências com melhores classificações a partir da filtragem foram modeladas usando seis programas/servidores: Fugue, HHpred, i-tasser, Modeller, ModPipe e Swiss-Model. Após validação da modelagem, refinamento das estruturas seguido da validação, o servidor Swiss-Model apresentou um melhor resultado e, portanto, foi o selecionado para a modelagem das sequências restantes do cromossomo 10.





### 3.3.1 Programas/servidores utilizados na construção dos modelos

O Modeller é o programa de modelagem comparativa que emprega o método de satisfação de restrição espacial. A construção da estrutura tridimensional de uma sequência de aminoácidos é feita a partir do alinhamento desta com uma proteína homóloga de estrutura conhecida. Além da construção do modelo 3D, o programa disponibiliza funções adicionais que incluem a modelagem de regiões de volta na estrutura da proteína através do método *de novo*, otimização de vários modelos de estrutura de proteína em relação a flexibilidade definida na função objetivo, alinhamento múltiplo de sequências de proteína e/ou estruturas, clusterização, busca em base de dados de sequências, comparação de estrutura de proteínas, entre outras. Esse programa também permite criar o gráfico de avaliação de energia através do método DOPE (*Discrete Optimized Protein Energy*), que é um potencial estatístico otimizado para avaliar o modelo construído em relação a seu molde. Alternativamente, DOPE pode gerar um perfil de energia resíduo por resíduo para o modelo inicial tornando possível ao usuário identificar as regiões problemáticas na estrutura modelo.

ModPipe é um software totalmente automatizado que constrói modelos de estrutura de proteína para um grande número de sequências e que só recebe como entrada o identificador da sequência alvo e sua sequência fasta. O pipeline utilizado tem como base o programa Modeller, que pertence ao grupo Andrej Sali, criador de ambos programas. O programa automatiza as três principais etapas do processo de modelagem comparativa, que é o alinhamento entre sequência alvo e a estrutura molde, construção de modelos e avaliação dos mesmos. Além do protocolo padrão, ModPipe permite que o usuário execute o programa informando qual estrutura molde deseja usar para construir o modelo de determinada sequência. A principal vantagem de utilização deste programa é a automatização quando se trabalha com diversas sequências.

O programa Fugue faz o reconhecimento de homólogos distantes por comparação entre a sequência alvo e a estrutura molde. Para reconhecer homólogos distantes, este utiliza HOMSTRAD (Mizuguchi, Deane *et al.* 1998), que é uma base de dados de alinhamentos de estruturas de proteínas para famílias homólogas. Além disso, Fugue usa penalidades de *gap*, dependentes de estrutura, que calculam a penalidade para os resíduos de acordo com sua posição em relação aos elementos de estrutura secundária e da acessibilidade ao solvente, resultando

em alinhamentos mais precisos. Em suma, o programa varre um banco de dados de perfis estruturais, calcula a pontuação de compatibilidade da sequência-estrutura e produz uma lista de potenciais homólogos e alinhamentos. No caso das metaloproteases, o molde encontrado foi uma leishmanolisina de *L. major* de código PDB id 1LML (Schlagenhauf, Etges *et al.* 1998). Tendo em vista que esse servidor é online, o modelo foi criado de forma automatizada.

O servidor HHpred foi o primeiro servidor a implementar a comparação par a par do perfil de modelos ocultos de Markov (*Hidden Markov Model*). O primeiro passo da execução consiste na construção do alinhamento entre homólogos da sequência alvo através da busca com PSI-BLAST contra a base de dados não redundante do NCBI (*National Center for Biotechnology Information*). Este programa permite a busca em uma variedade de base de dados como PDB (Bourne, Address *et al.* 2004), SCOP (Murzin, Brenner *et al.* 1995), Pfam (Sonnhammer, Eddy *et al.* 1998), SMART (Schultz, Milpetz *et al.* 1998), COG (Tatusov, Fedorova *et al.* 2003) e CDD (Marchler-Bauer, Panchenko *et al.* 2002), tendo como entrada a sequência de aminoácidos da sequência alvo ou um alinhamento múltiplo da sequência alvo com os moldes encontrados. A partir do alinhamento, o modelo estrutural é calculado a partir do programa Modeller, de forma automática no servidor.

O servidor i-tasser (*Iterative Threading Assembly Refinement*) utiliza métodos *threading*, *ab initio* e refinamento estrutural para a construção de estruturas 3D. A predição é constituída basicamente por três etapas: (i) Alinhamento multi-fragmentos, a partir da sequência de aminoácidos. O servidor primeiramente tenta recuperar proteínas moldes com dobras similares (superestruturas secundárias) a partir da biblioteca PDB do LOMETS (Wu and Zhang 2007), uma abordagem *meta-threading* instalada localmente. (ii) Agrupamento, fragmentos contínuos são extraídos das estruturas moldes e reagrupados em modelos completos por meio de simulações de Monte Carlo do tipo troca entre réplicas (*replica-exchange Monte Carlo*) utilizando *threading*, regiões não alinhadas (principalmente os loops) são construídas por método *ab initio*. (iii) Refinamento, a montagem de conformações por fragmentos é executada novamente utilizando as estruturas-modelos (centroides) da etapa anterior. Essa segunda interação é realizada para refinar os centroides eliminando possíveis interferências estéricas. *Clusters* (agrupamentos de estruturas) são gerados

novamente e as configurações de menor energia são selecionadas. Por fim, realiza-se uma otimização da energia das ligações de hidrogênio.

Por fim, o servidor Swiss-Model é o único dos servidores citados que utiliza a metodologia de união de corpos rígidos. Ele utiliza um servidor automatizado que oferece três níveis de interação com o usuário. O primeiro modo permite a submissão de uma sequência de aminoácidos de uma proteína de interesse para construção do modelo 3D. A seleção do modelo, seu alinhamento e construção são feitos automaticamente pelo servidor desde que o molde disponível tenha mais de 25% de identidade sequencial com a sequência desejada. No modo alinhamento, o processo de modelagem é iniciado com base na submissão do arquivo que contém o alinhamento entre a sequência alvo e o molde escolhido pelo usuário. O último modo permite ao usuário submeter um modelo otimizado ao servidor. Para esse modo, é utilizado o DeepView, disponível no visualizador Swiss-Pdb Viewer. Ele contém as estruturas de molde sobrepostas e o alinhamento entre a sequência alvo e os moldes. Este modo permite que o usuário controle diversos parâmetros como, por exemplo, a seleção do molde ou a inserção de gaps no alinhamento. Com este servidor utilizamos o primeiro modo oferecido.

Todos os programas/servidores utilizaram como molde a proteína leishmanolisina de *Leishmania major* disponível no banco de dados público de proteínas (PDB) sob código PDB id 1LML (Schlagenhauf, Etges *et al.* 1998). Todos os parâmetros dos programas/servidores foram mantidos em seus valores padrão para a construção dos modelos.

### **3.3.2 Validação dos modelos**

A validação dos modelos é uma etapa essencial a ser realizada após a construção dos mesmos. Na validação, a qualidade do enovelamento global da proteína, possíveis erros estruturais em regiões localizadas e os parâmetros estereoquímicos são avaliados (Johnson, Srinivasan *et al.* 1994). Como a qualidade do modelo depende da estrutura da proteína-molde utilizada e do alinhamento calculado, é importante verificar se existem diferenças conformacionais entre os elementos de estrutura secundária (regiões conservadas) das estrutura-molde e da estrutura-modelada.

A qualidade estereoquímica do modelo foi realizada com o programa PROCHECK (Laskowski, MacArthur *et al.* 1993), que avalia os comprimentos de

ligação, os ângulos planos, a planaridade dos anéis de cadeias laterais, a quiralidade, as conformações das cadeias laterais, a planaridade das ligações peptídicas, os ângulos torcionais da cadeia principal e das cadeias laterais, o impedimento estérico entre pares de átomos não-ligados e a qualidade do gráfico de Ramachandran (Ramachandran and Sasisekharan 1968).

O gráfico de Ramachandran indica a distribuição dos ângulos de torção  $\phi$  (Phi) e  $\psi$  (Psi) dos resíduos da estrutura. Uma estrutura de boa qualidade deve possuir 90% ou mais dos seus resíduos nas regiões permitidas ou generosamente permitidas do gráfico.

Auxiliado pelo programa PROCHECK, o fator-G fornece uma média dos desvios de uma dada propriedade estereoquímica (distância de ligação covalente, ângulos da cadeia principal, ângulos diedrais, entre outros) baseada nas distribuições observadas destes parâmetros estereoquímicos em sua base de dados. Um valor baixo do fator-G representa uma baixa probabilidade de ocorrência do parâmetro que se observa para um dado resíduo. Resíduos que se encontram em regiões não permitidas do diagrama de Ramachandran terão um fator-G baixo (ou muito negativo), indicando inconsistências na geometria total da estrutura avaliada. O fator-G global abrange os valores de fator-G para toda a estrutura em um único número.

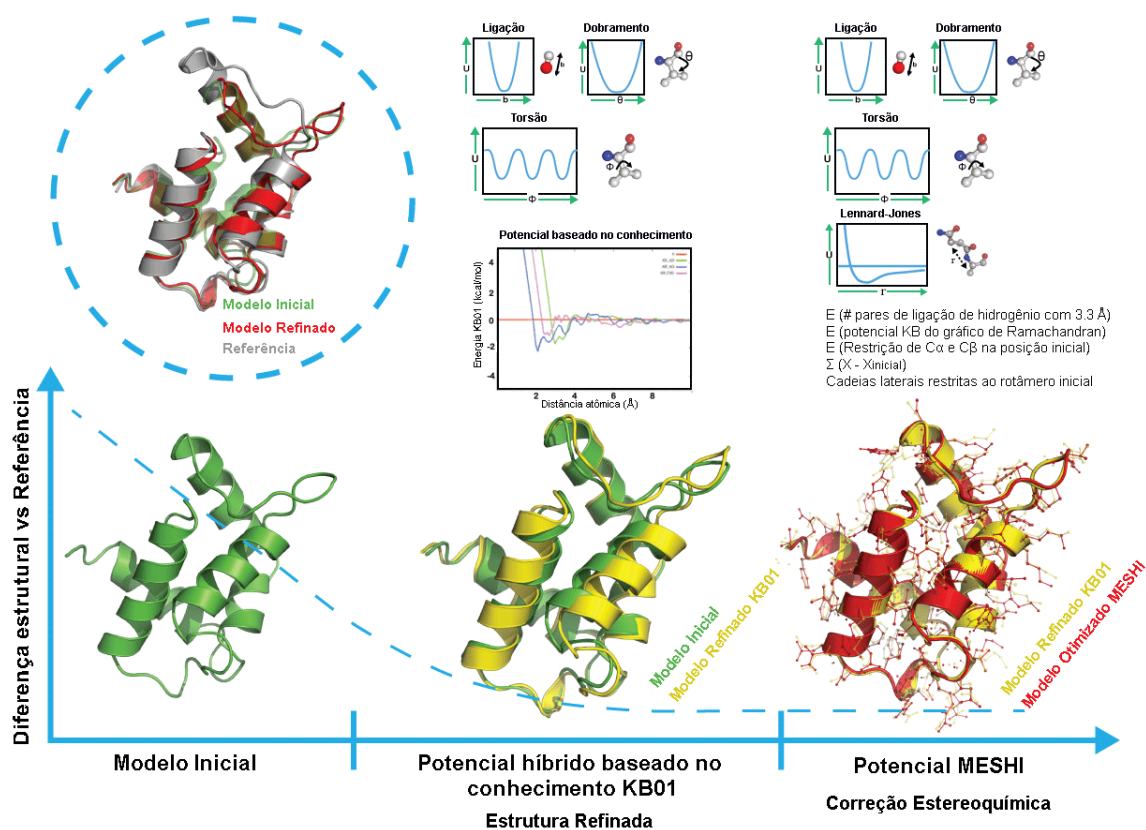
Os programas de validação ANOLEA (*Atomic Non-Local Environment Assessment*) (Melo and Feytmans 1998) e QMEAN (Benkert, Tosatto *et al.* 2008), disponíveis pelo workspace do Swiss-Model<sup>3</sup> também foram considerados. O servidor ANOLEA avalia o ambiente dos átomos pesados por meio de cálculos de energia, sendo possível determinar zonas de alta energia na proteína, as quais estão relacionadas a erros pontuais ou regiões de interação. Este cálculo gera uma pontuação indicando quantos aminoácidos estão em zonas de maior energia e quais são eles. Já o QMEAN é uma função de pontuação capaz de derivar tanto globalmente (toda a estrutura) quanto realizar estimativas locais (resíduos) na determinação de erros na base de um único modelo. Ele mostra no modelo gerado as regiões que precisam ser melhor modeladas afim de obter resultados mais confiáveis. A avaliação do QMEAN fornece uma pontuação que varia de 0 a 1, sendo os valores mais próximos de 1 correlacionados com a melhor qualidade do modelo.

---

<sup>3</sup> [http://swissmodel.expasy.org/workspace/?func=tools\\_structureassessment1](http://swissmodel.expasy.org/workspace/?func=tools_structureassessment1)

### 3.3.3 Refinamento dos modelos

Com o objetivo de refinar os modelos construídos com os programas de modelagem baseada em molde, foi utilizado o servidor KoBaMIN (Rodrigues, Levitt *et al.* 2012). O processo de refinamento deste servidor é realizado em dois passos (Figura 14): (i) utilizando o servidor ENCAD (Levitt, Hirshberg *et al.* 1995) que usa um algoritmo de minimização de energia (KB01) altamente convergente que inclui o efeito do solvente em todos os átomos (Summa and Levitt 2007, Chopra, Summa *et al.* 2008) e; (ii) outra minimização de energia usando o programa MESHI (Kalisman, Levi *et al.* 2005), que corrige as posições das cadeias laterais e outros problemas encontrados na estereoquímica da proteína.



**Figura 14:** Representação do protocolo de refinamento utilizado pelo servidor KoBaMIN. Adaptado de Rodrigues, Levitt *et al.* 2012.

Após o refinamento, o programa retorna um arquivo com as coordenadas da proteína refinada e novamente essas estruturas são submetidas a avaliação pelos programas citados na sessão de validação dos modelos.

### 3.4 Dinâmica Molecular

As estruturas refinadas pelo KoBaMIN foram as utilizadas nas simulações de dinâmica molecular. O estado de protonação dos resíduos envolvidos na coordenação do zinco foram determinados usando o servidor H++ (Anandakrishnan, Aguilar *et al.* 2012), que faz a predição do estado de protonação baseando-se em ligações de hidrogênio favoráveis, energias locais e valores de pKa. Nas histidinas His165, His169 e His235, o Nitrogênio  $\epsilon$  foi desprotonado a fim de manter a coordenação do íon de zinco.

As simulações foram realizadas com o programa AMBER v14 (Case, Babin *et al.* 2014), que utiliza o PMEMD (*Particle-Mesh Ewalds Molecular Dynamics*). O campo de força utilizado nas simulações de dinâmica molecular foi o ff14SB (Maier, Martinez *et al.* 2015).

A preparação dos arquivos de entrada para as simulações de DM foi realizada utilizando a ferramenta LEaP do programa AMBER. As ligações dissulfeto presentes nos modelos/molde foram relacionadas no script de preparação desses arquivos. A neutralização do sistema foi dada adicionando uma quantidade apropriada de íons Na<sup>+</sup>.

Para solvatação do sistema, foi utilizado uma caixa de moléculas de água retangular do tipo TIP3P (Jorgensen, Chandrasekhar *et al.* 1983), com 9 Å a partir dos átomos das extremidades da proteína nas três direções ortogonais até a borda da caixa gerando um tamanho aproximado de  $80 \pm 4,5$ ,  $76 \pm 2,0$  e  $83 \pm 2,7$  Å. TIP3P é provavelmente o modelo de água mais utilizado devido a boas propriedades dinâmicas e termodinâmicas que proporciona para a maioria dos campos de força (de Castro 2012) e é compatível com o campo de força utilizado.

Os sistemas foram simulados com condições periódicas de contorno em uma caixa de água retangular e o método PME (*Particle Mesh Ewald*) (Essmann, Perera *et al.* 1995) foi utilizado para tratar as interações eletrostáticas de longo alcance com um cut-off de 10 Å. Duas minimizações foram realizadas: (1) com restrição de posição imposta aos átomos pesados do sistema com aplicação de potencial de

Hooke com constante de força no valor de  $5,0 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$ , sendo os primeiros 1000 passos realizados com o algoritmo *steepest-descent* e outros 1000 passos utilizando gradiente conjugado; (2) sem restrição de posição com 2500 passos usando o algoritmo *steepest-descent* e mais 2500 passos usando gradiente conjugado. A minimização será interrompida se a raiz quadrada média dos elementos cartesianos do gradiente for inferior ao valor de  $1.0 \times 10^{-4} \text{ kcal/mol.\AA}$ . As coordenadas da trajetória foram salvas a cada 100 passos.

Após a minimização de energia, o sistema foi gradualmente aquecido de 20 até 310K por 1ns mantendo restrição de posição da proteína. Em seguida, foi realizado a equilibração do sistema, com pressão constante de 1 bar, reduzindo progressivamente a restrição de posição imposta aos átomos pesados. O valor de restrição inicial foi  $10,0 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  sendo simulado por 100ps. Para cada um dos demais valores de restrição: 5,0, 2,5, 1,0, 0,5, 0,25, 0,1 e  $0,05 \text{ kcal mol}^{-1} \text{ \AA}^{-1}$  e mais uma simulação sem nenhuma restrição também foi realizado em 100ps cada. Após equilibração, foi realizada a fase de produção em temperatura constante de 310K por 50ns (Anexo A). Todas as simulações foram realizadas no ensemble estatístico NPT.

Os servidores utilizados para as simulações dos sistemas possuem processador Intel Xeon com 6 núcleos, placa gráfica Nvidia GTX 780ti, 64 gigabytes de memória RAM e sistema operacional Centos 6.5.

### **3.4.1 Análise das Simulações de Dinâmica Molecular**

#### **3.4.1.1 Raiz do Desvio Médio Quadrático (RMSD)**

O RMSD é a medida da média da distância entre os átomos de duas estruturas. O valor final mostra quanto o conjunto de átomos de uma proteína se desvia do conjunto de átomos de referência. Em DM, mesmo após a etapa prévia de equilíbrio, o relaxamento das estruturas ainda continua por tempos que vão além de centenas de picossegundos. Para avaliar o comportamento global das estruturas foi calculado o RMSD durante a trajetória de simulação para avaliar se as estruturas geradas já se encontravam em equilíbrio. Valores uniformes indicam pouca oscilação, enquanto que oscilações bruscas na média indicam mudanças na conformação. Para fazer o cálculo de RMSD, utilizamos a ferramenta *ptraj*, incluída no pacote AmberTools14.

Sua definição matemática é dada pela fórmula:

$$RMSD = \sqrt{\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^n (x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2} \quad (1)$$

onde  $n$  é o número de átomos do sistema;  $x, y$  e  $z$  são as coordenadas espaciais do sistema; e,  $i, j$  são dois átomos respectivos.

#### 3.4.1.2 Flexibilidade dos resíduos do sítio catalítico

Para verificar o comportamento do zinco e dos resíduos do sítio catalítico usamos a flutuação da raiz média quadrática (*Root Mean Square Fluctuations*, RMSF). O RMSF é calculado comparando-se as estruturas instantâneas obtidas em cada passo da simulação com a estrutura inicial antes da DM. As simulações foram convertidas para o formato do programa GROMACS (Pronk, Páll *et al.* 2013) para então utilizarmos a ferramenta *gmx rmsf*. O objetivo desta análise foi verificar a flexibilidade dos resíduos do sítio dos modelos ao longo das simulações.

#### 3.4.1.3 Clusterização

A clusterização permite agrupar estruturas similares de uma simulação. A estrutura representativa do maior cluster da simulação foi selecionada utilizando a ferramenta *gmx cluster* do pacote GROMACS (Pronk, Páll *et al.* 2013), utilizando o algoritmo GROMOS (Daura, Gademann *et al.* 1999) para determinar a associação de um cluster com raio de corte de 0,1nm no *backbone* para cada grupo de cluster.

#### 3.4.1.4 Perfil Eletrostático

Para se obter o perfil eletrostático, utilizamos somente a região do sítio catalítico da estrutura representativa do maior cluster da dinâmica molecular. Submetemos o mesmo ao servidor pdb2pqr (Dolinsky, Nielsen *et al.* 2004), que automatiza muitas das tarefas comuns de preparação de estruturas para cálculos eletrostáticos. Isto inclui a adição de um número limitado de átomos pesados que faltam em estruturas biomoleculares, otimização da proteína para ligações de hidrogênio favoráveis, determinação do pKa de cadeias laterais, atribuição de



cargas e parâmetros de raio a partir de uma variedade de campos de força, obtendo-se um arquivo convertido em formato PQR. O campo de força utilizado no servidor foi o Amber e os valores de pH foram 5,0 e 7,0.

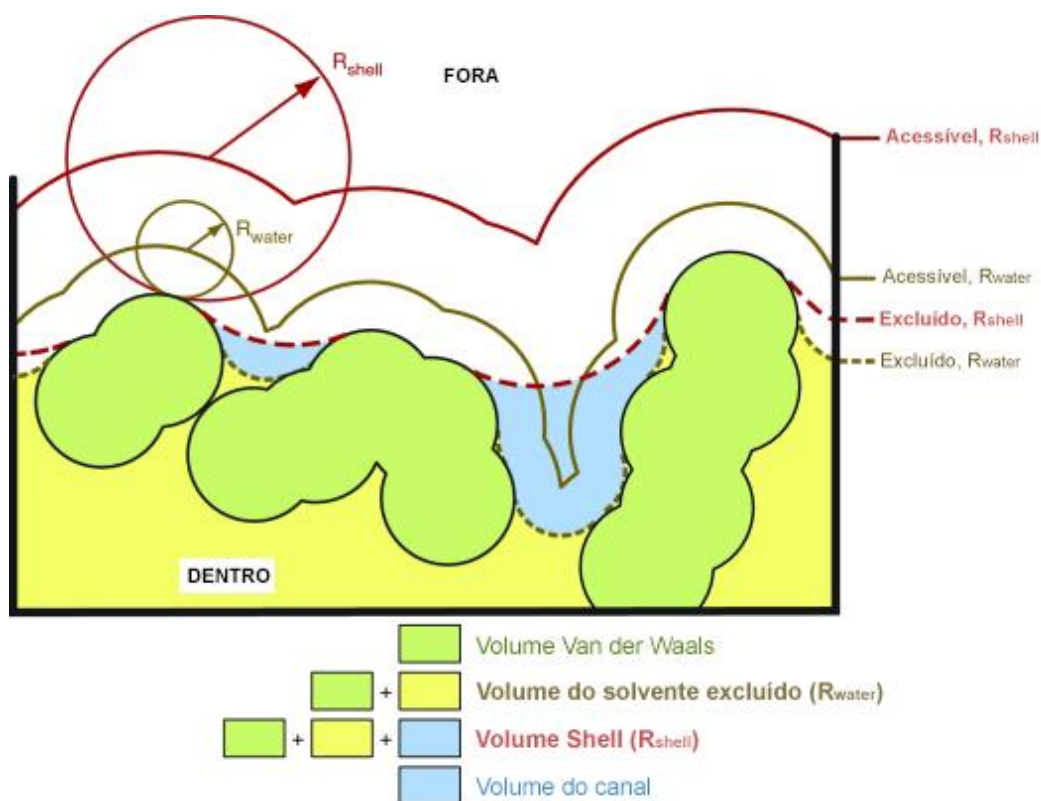
O resultado do mapa do potencial eletrostático é visualizado pelo programa UCSF Chimera (Pettersen, Goddard *et al.* 2004), que utiliza a ferramenta APBS (Adaptive Poisson-Boltzmann Solver) (Baker, Sept *et al.* 2001) e possui uma ferramenta de coloração da superfície eletrostática que é aplicada à superfície molecular. Os valores mínimo e máximo utilizados para a coloração do perfil eletrostático foram -3 e 3, respectivamente.

#### **3.4.1.5 Dendrograma estrutural**

Utilizamos o servidor 3V (Voss and Gerstein 2010) que calcula, dentre outros parâmetros, volume, área, esfericidade e raio efetivo de uma proteína dada. Este servidor usa o método de rolamento (Richards 1977). Este método funciona essencialmente colocando uma sonda virtual ou uma bola de um determinado raio em torno da superfície de van der Waals da macromolécula (Figura 15).

Como resultado, uma representação volumétrica da proteína é dada, além de valores numéricos. Submetemos a região de inserção das proteínas do cromossomo 10 e usamos os valores padrão (raio de rolamento = 3 Å) do servidor para a geração dos resultados. O volume é calculado em Å<sup>3</sup> e a área, em Å<sup>2</sup>.

A esfericidade é uma medida de quanto o volume se assemelha a uma esfera dado que uma esfera perfeita tem a esfericidade de 1,0 e um cubo tem esfericidade de 0,81. O valor da esfericidade é calculado a partir do volume sobre a área da proteína. Já o raio efetivo é o raio de uma esfera que tem a mesma área de superfície que a proporção do volume da proteína.



**Figura 15:** Método de rolamento utilizado pelo servidor 3V. Adaptado de (Voss and Gerstein 2010).

### 3.4.2 Identificação de *hotspots*

Para identificar os resíduos importantes, foi realizada uma análise com o programa foldX (Schymkowitz, Borg *et al.* 2005). Este programa possui um campo de força empírico que foi desenvolvido para a rápida avaliação do efeito das mutações sobre a estabilidade, a dobragem e a dinâmica das proteínas e ácidos nucleicos. O comando AlaScan faz uma varredura por Alanina na proteína e avalia o efeito direto que cada resíduo mutado tem na estabilidade global da proteína.

Cada posição da proteína (exceto Glicina e Alanina) é trocada por Alanina e a posição dos resíduos vizinhos é otimizada. O resultado dado é a diferença em energia ( $\Delta\Delta G$ , em kcal/mol) entre o resíduo mutado e a estrutura nativa, decompostas nos termos de energia usados no programa:

$$\Delta\Delta G(\text{troca}) = \Delta G(\text{mutado}) - \Delta G(\text{nativo}) \quad (2)$$

A regra utilizada neste programa para identificar a estabilidade é:

- $\Delta\Delta G(\text{troca}) > 0$ : a mutação é desestabilizada
- $\Delta\Delta G(\text{troca}) < 0$ : a mutação é estabilizada

A margem de erro do foldX é de cerca de 0,5 kcal/mol. Alterações neste intervalo são insignificantes. Alguns trabalhos definem o valor adotado para a definição de uma mutação. Ofran e Rost (Ofran and Rost 2007) definem que se um resíduo na troca muda a energia de ligação da proteína substancialmente ( $\Delta\Delta G > 2,5$  kcal/mol), então este resíduo é considerado *hotspot*. Nesta tese, usamos este mesmo critério.

### 3.5 Atracamento Molecular (*Docking*)

As simulações de *docking* foram realizadas com o programa Autodock4 (Morris, Goodsell *et al.* 1998). Este programa utiliza um algoritmo genético lamarckiano (LGA) para encontrar a melhor conformação do ligante no sítio ativo do receptor. O algoritmo genético lamarckiano é baseado na teoria da evolução de Lamarck, onde aos descendentes são transmitidas as mudanças adquiridas durante a vida de seus antecessores.

Este algoritmo utiliza, para avaliação das conformações, uma função empírica que determina a interação através da energia livre de ligação receptor-ligante. A equação 3 descreve a função implementada neste programa.

$$\Delta G = \Delta G_{vdw} + \Delta G_{hbond} + \Delta G_{ele} + \Delta G_{conf} + \Delta G_{solv} \quad (3)$$

onde  $\Delta G_{vdw}$  é o termo que representa a contribuição de van der Waals,  $\Delta G_{hbond}$  é o termo que descreve a contribuição das ligações hidrogênio,  $\Delta G_{ele}$ , representa as contribuições do potencial eletrostático,  $\Delta G_{conf}$ , descreve a contribuição entre os ângulos dos átomos (termo torcional) e  $\Delta G_{solv}$ , a contribuição de solvatação.

Utilizamos este programa, pois, recentemente, o grupo desenvolvedor do Autodock implementou um campo de força que inclui um potencial especializado na descrição de interações de ligantes coordenados pelo zinco. A esse novo campo de força foi dado o nome de Autodock4<sub>Zn</sub> (Santos-Martins, Forli *et al.* 2014).

O novo campo de força descreve tanto componentes energéticos quanto geométricos da interação. Foi calibrado com um conjunto de dados de 292 complexos cristalográficos contendo zinco. Simulações de *redocking* mostraram que com a utilização deste campo de força há melhorias significativas no desempenho tanto na estimativa da energia livre de ligação quanto no RMSD (*Root Mean Square Deviation*) calculado com a estrutura da pose cristalográfica. Esse novo campo de força foi implementado no programa Autodock sem modificações no seu código-fonte.

Autodock4<sub>Zn</sub> foi desenvolvido pois o campo de força padrão implementado no Autodock considerava a distância de equilíbrio de van der Waals para átomos envolvidos na coordenação do zinco valores significativamente maiores do que as distâncias de coordenação (por exemplo, para o nitrogênio, a distância é usada com valor de 2,49 Å comparada com a distância de coordenação que é de 2,0 Å). Em segundo lugar, devido à falta de termos especializados para a coordenação do metal, a direcionalidade não era levada em consideração. Finalmente, enquanto o termo eletrostático é muito eficaz na descrição das interações envolvendo cargas parciais, ele faz com que a função de energia seja altamente sensível a grupos fortemente carregados, tais como metais com cargas formais.

Nos modelos que usam cargas Gasteiger (Huey, Morris *et al.* 2007) no Autodock, cargas mais negativas são atribuídas a átomos de oxigênio do que nitrogênio e enxofre, resultando assim em candidatos preferenciais por quelar metal carregado positivamente. Enquanto esta aproximação é acurada suficientemente para interações com magnésio, não é suficiente para descrever com propriedade as preferências de coordenação do zinco.

A coordenação do zinco requer um tratamento especializado. A função da energia livre de ligação (ELL) do campo de forças padrão do Autodock inclui os seguintes termos (eq 4):

$$ELL = W_{vdw}(vdW) + W_{hb}(Hbond) + W_{elec}(elec) + W_{sol}(desolv) + W_{tor}(torsDoF)$$

onde a energia livre de ligação é calculada como uma soma de van der Waals (*vdW*), ligação hidrogênio (*Hbond*), Coulomb (*elec*), desolvatação (*desolv*) e a entropia torcional do ligante (*torsDoF*); cada termo possui um peso *W* que é estimado usando um modelo de regressão linear.

Neste campo de forças novo, o potencial eletrostático para o zinco foi desabilitado e configurado para ter carga parcial zero. As interações par a par de cada tipo de átomo envolvido na coordenação do zinco foi definido como um novo termo no potencial de energia. Para os tipos de átomos N, OA e SA, potenciais esféricos como  $V_{Zn,N}$ ,  $V_{Zn,OA}$ ,  $V_{Zn,SA}$  foram definidos para refletir as distâncias de coordenação conhecidas através da adaptação do potencial de van der Waals no campo de força do Autodock (eq 5):

$$V_{ij} = \varepsilon_{ij} \left[ \left( \frac{r_{ij}}{r} \right)^{12} - 2 \left( \frac{r_{ij}}{r} \right)^6 \right] \quad (5)$$

A distância de equilíbrio dos pares  $r_{ij}$  entre zinco e os tipos de átomos N, OA, e SA foram ajustados para 2,0, 2,1 e 2,25 Å, respectivamente, e os valores de  $\varepsilon$  foram estimados. Para o tipo NA, uma nova direção tetraédrica  $V_{TZ,NA}$  foi definida e a interação com zinco foi separada em dois componentes. O componente repulsivo é mediado pelo átomo de zinco enquanto que o componente atrativo é mediado por um pseudoátomo TZ que foi adicionado à tabela padrão do campo de força. O pseudoátomo é adicionado à estrutura do receptor em todos os complexos onde a geometria de coordenação tetraédrica é presente.

Com isso, o seguinte potencial é adicionado à eq. 4:

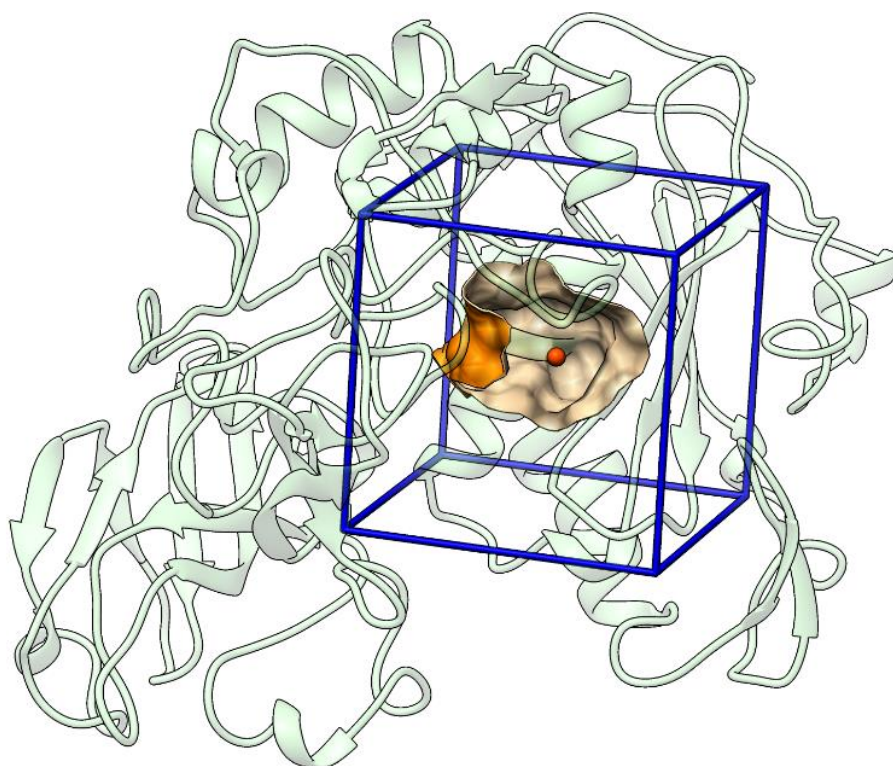
$$ZincCoord = V_{TZ,NA} + V_{Zn,N} + V_{Zn,OA} + V_{Zn,SA} \quad (6)$$

Dessa forma, a energia livre de ligação torna-se uma combinação linear dos cinco termos padrão do Autodock mais o novo potencial de coordenação par a par do zinco.

### 3.5.1 Preparação dos receptores para *docking*

Cada uma das proteínas foi preparada automaticamente através do script *prepare\_receptor4.py* disponibilizado pelo programa Autodock. A dimensão da grade de energia foi de 22,5 Å em cada direção (x, y e z) com discretização de 0,375 Å, estando a grade centralizada no zinco (Figura 16). Esse tamanho foi definido após a sobreposição de todos os ligantes para que pudéssemos encontrar

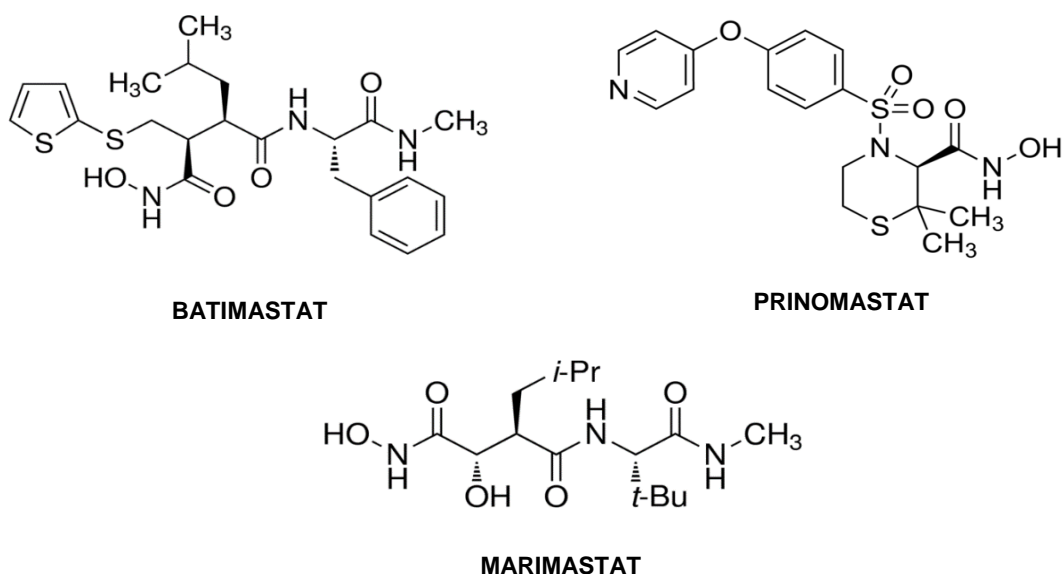
o tamanho que acomodasse o maior ligante no sítio da proteína. Cargas Gasteiger (Gasteiger and Marsili 1978) foram atribuídas à proteína.



**Figura 16:** Representação gráfica da caixa cúbica (azul) onde o grid é gerado. A esfera laranja escuro indica a posição do zinco e em laranja mais claro, encontra-se a superfície acessível ao solvente com até 4 Å de distância do íon.

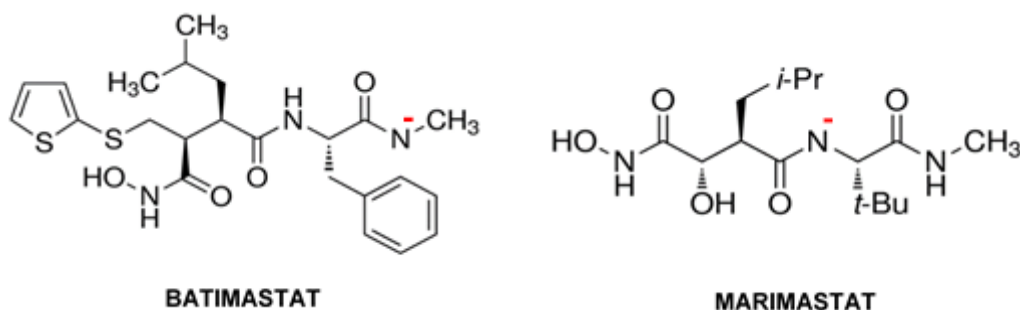
### 3.5.2 Preparação dos ligantes para *docking*

Os ligantes usados nas simulações de *docking* foram: Batimastat (BAT), Prinomastat (PRI) e Marimastat (MAR), três inibidores conhecidos de metaloproteases (Figura 17). As estruturas foram obtidas no site da Sigma-Aldrich<sup>4</sup> sob os números SML0041 (Batimastat), PZ0198 (Prinomastat) e M2699 (Marimastat).



**Figura 17:** Estruturas 2D dos inibidores usados nas simulações de *docking*.

Para o cálculo da protonação, utilizamos o programa Marvin Sketch<sup>5</sup> v. 15.6.8, que gera possíveis conformações do ligante de acordo com o pH. O intervalo de pKa para a geração dos resultados foi de 4 a 8. Tendo em vista que esses ligantes vão atuar em um ambiente de pH 7.0, os ligantes batimastat e marimastat apresentaram nitrogênio desprotonado (Figura 18).



**Figura 18:** Estrutura 2D dos ligantes desprotonados pelo programa Marvin.

<sup>4</sup> [www.sigmaaldrich.com/brazil.html](http://www.sigmaaldrich.com/brazil.html)

<sup>5</sup> [www.chemaxon.com](http://www.chemaxon.com)

A preparação do arquivo para as simulações de *docking* consistiu na adição de todos os hidrogênios. A criação do arquivo.pdbqt foi feita através do script *prepare\_ligand4.py*. Cargas Gasteiger também foram usadas nos ligantes. Os scripts utilizados na etapa do *docking* estão disponíveis no Anexo B.

### 3.5.3 Análise das simulações de *Docking*

O arquivo de saída do programa Autodock4<sub>zn</sub> lista as conformações por ordem de menor energia e mostra em qual execução da simulação as estruturas foram obtidas. Após esta verificação, a conformação a ser analisada é separada em um novo arquivo acrescido das coordenadas da proteína.

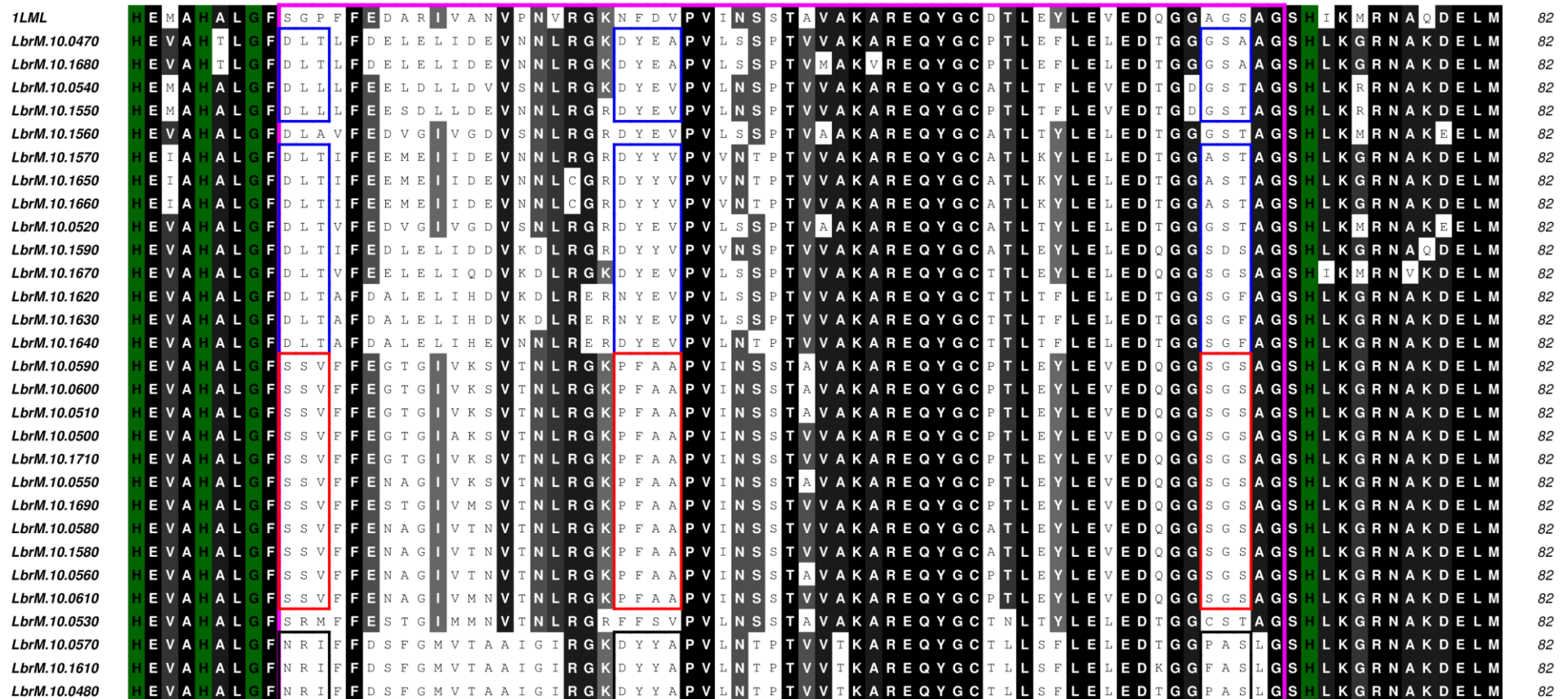
O programa Maestro<sup>6</sup>, da empresa Schrödinger, possui uma versão gratuita para visualização das moléculas. Este programa foi utilizado para mapear as interações entre proteína-ligante. Para visualização também foi utilizado o programa PyMOL (DeLano 2002).

---

<sup>6</sup> <https://www.schrodinger.com/maestro>

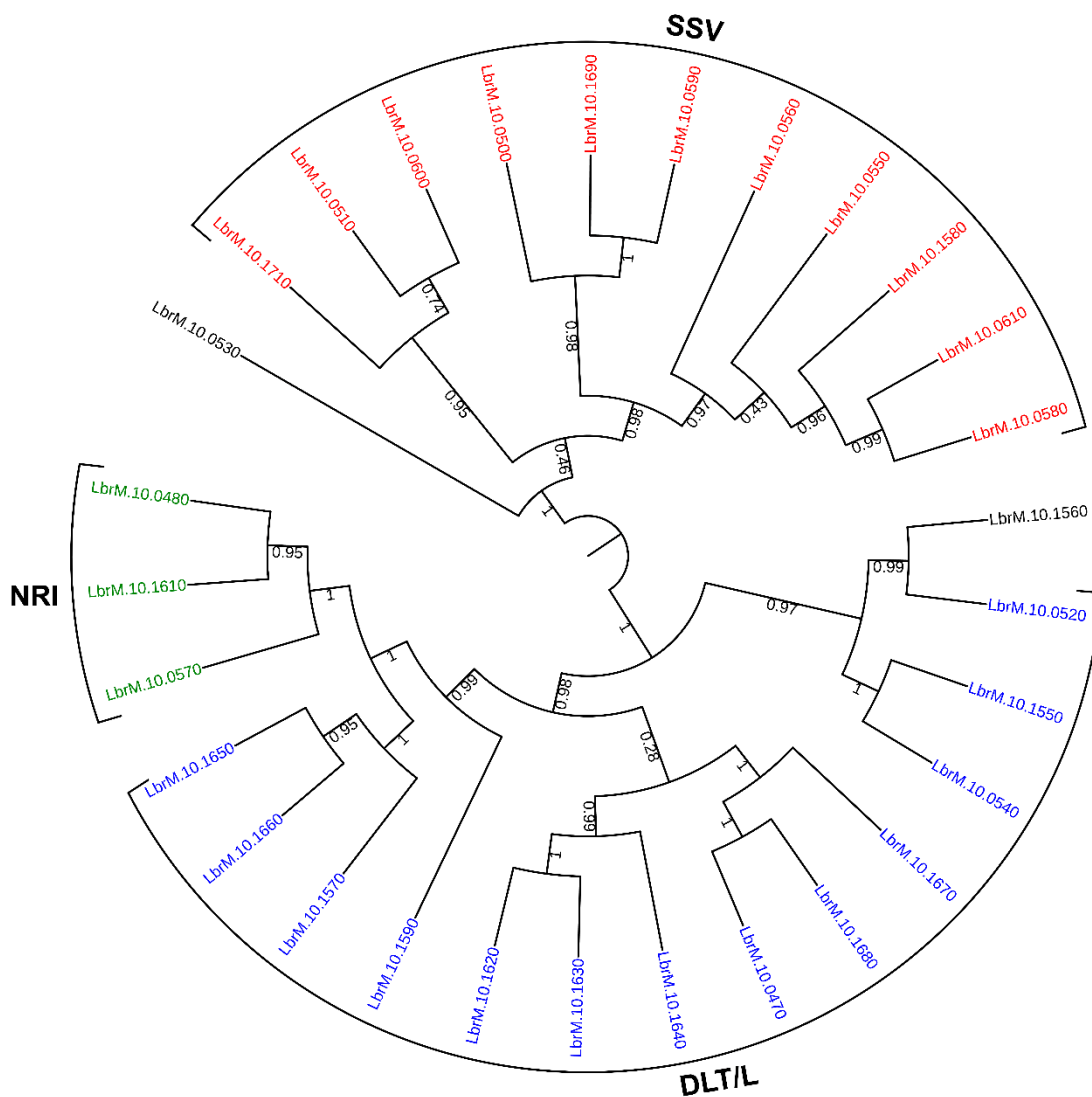






**Figura 19:** Alinhamento da região do sítio catalítico com a identificação dos três motivos: DLX, SSV e NRI no início da região de inserção (caixa magenta). Caixas azul, vermelha e preta indicam as regiões que se repetem nessas sequências para DLX, SSV e NRI, respectivamente, onde X pode ser substituído por Thr ou Leu. Regiões na cor preta indicam a identidade entre os aminoácidos das sequências, regiões em cinza, indicam a similaridade, e, regiões em branco mostram os resíduos que diferem entre as mesmas. As histidinas de coordenação do Zinco e a Glicina conservada estão destacadas em verde. A sequência 1LML no topo corresponde a leishmanosina na qual a estrutura cristalográfica é conhecida. Figura gerada com o programa ALINE (Bond and Schuettelkopf 2009).

Essas regiões representam a principal diferença entre as sequências deste cromossomo. Com o objetivo de agrupar os genes deste cromossomo, submetemos o resultado do alinhamento global das sequências ao servidor iTOL (Letunic and Bork 2016), para gerar um dendrograma da organização desses genes (Figura 20).



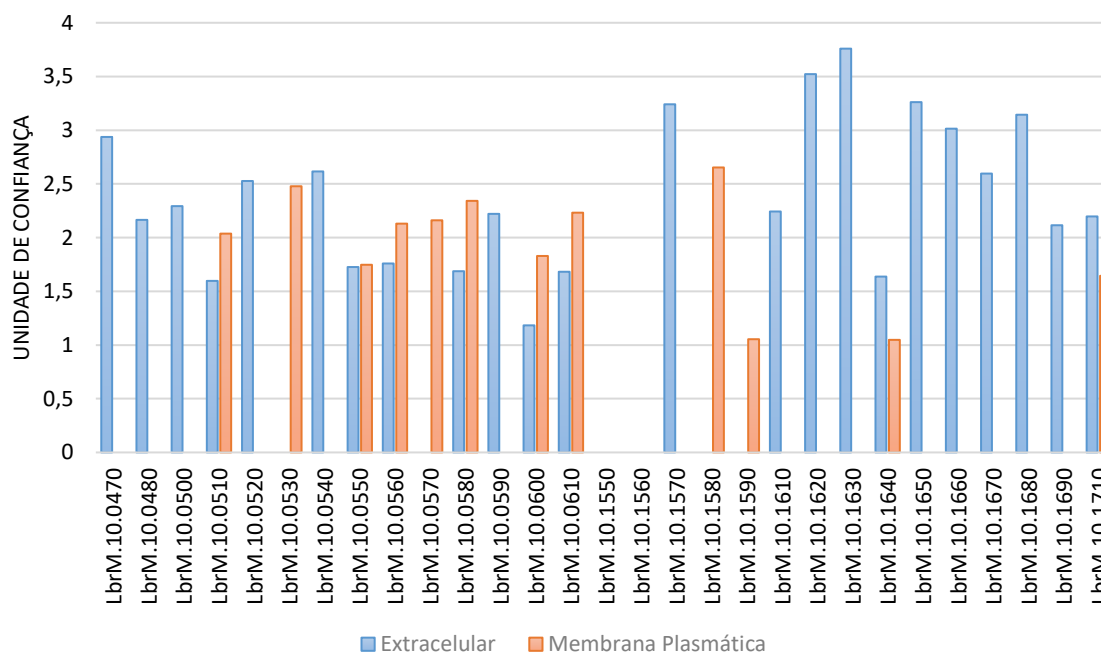
**Figura 20:** Dendrograma da organização dos genes do cromossomo 10 de *L. (V.) braziliensis*. As tríades próximas aos nomes dos genes indicam os aminoácidos característicos da sequência. Os números nos ramos representam a distância evolutiva entre diferentes proteínas homólogas. As sequências LbrM.10.0530 e LbrM.10.1560 são as de motivo único, SRM e DLA, respectivamente. Figura gerada com o servidor iTOL.

É possível observar que as sequências com motivos DLX, onde X pode ser substituído por uma Leucina ou Treonina, ficaram agrupadas no dendrograma juntamente com o motivo NRI, enquanto que as sequências com motivo SSV se

agruparam separadamente. Este fato demonstra que apenas com a estrutura primária dessas proteínas é possível observar uma separação em dois principais grupos.

#### 4.1.1 Predição da Localização Celular

Os servidores CELLO e TopPred indicaram que a maior parte das metaloproteases de *Leishmania (V.) braziliensis* são extracelulares. Este resultado está de acordo já que as proteínas deste cromossomo apresentaram alta identidade com uma metaloprotease de *Leishmania major*, que também é expressa no ambiente extracelular (Schlagenhauf, Etges *et al.* 1998). A Figura 21 mostra as unidades de confiança que o servidor CELLO obteve para cada sequência do cromossomo 10 de *L. braziliensis*.

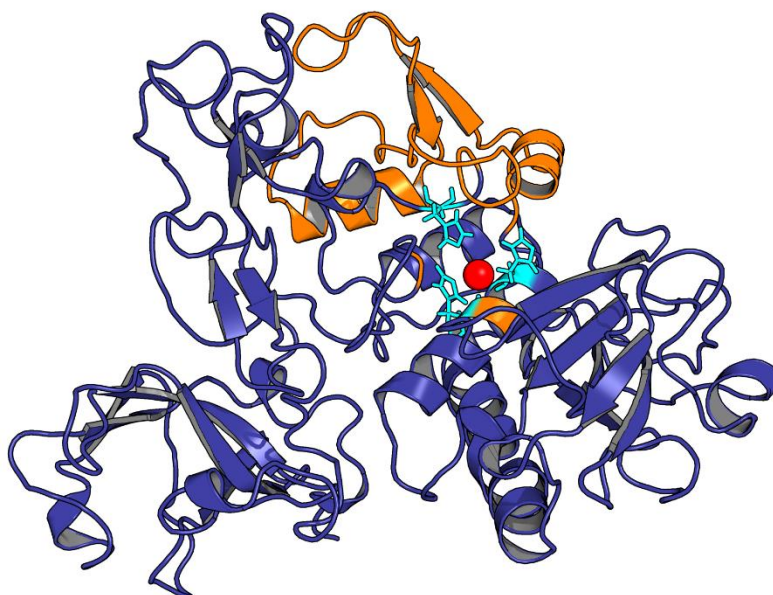


**Figura 21:** Predição da localização celular realizada com o servidor CELLO. O eixo Y indica as unidades de confiança para as regiões extracelulares e membrana plasmática.

Os resultados do CELLO indicam 51% das sequências com predição de serem expressas no ambiente extracelular. Para algumas sequências, o servidor fez a predição em duas regiões com valores de confiança para cada uma. Com predição de localização em regiões de membrana plasmática e extracelular, 27% das proteínas apresentaram essa combinação. Já a predição somente em

membrana plasmática, LbrM.10.0530 (motivo SRM) e LbrM.10.1580 (motivo SSV) foram os resultados obtidos. É curioso que as sequências LbrM.10.1550 e LbrM.10.1560 não apresentaram sua predição em regiões extracelulares ou em membrana plasmática. Este resultado pode estar relacionado com o tamanho dessas sequências, que são menores que as demais. Da predição somente em região extracelular, 81% das sequências deste resultado são compostas por sequências de motivo DLT.

O servidor TopPred indica quais partes de cada proteína pertence a região extracelular, transmembranar, citoplasmática e loop. Todas as sequências foram submetidas ao servidor e 75% dessas apresentaram localização na porção extracelular. A Figura 22 indica a predição realizada para a estrutura cristalográfica de uma metaloprotease de *Leishmania major*, PDB id 1LML.



**Figura 22:** Predição da localização celular da metaloprotease de *Leishmania major* (PDB id 1LML). A região extracelular está indicada em azul e a região de inserção está indicada em laranja. O íon Zinco está representado por uma esfera vermelha e as histidinas que coordenam o metal são representadas em sticks na cor ciano.

## 4.2 Busca de Moldes para Construção do modelo 3D

O primeiro passo para a modelagem comparativa é a verificação da identidade sequencial com alguma proteína de estrutura tridimensional já conhecida em bancos de dados públicos. Para isso, usamos a ferramenta de busca

por alinhamento local, BLASTp<sup>8</sup>, do NCBI (*National Center for Biotechnology Information*), contra o PDB.

O resultado do BLAST indicou o mesmo molde para todas as sequências. A protease de superfície de código PDB id 1LML (Schlagenhauf, Etges *et al.* 1998) é uma leishmanolisina de *Leishmania major* resolvida experimentalmente pelo método de difração de raio-x com resolução de 1,86 Å.

Ao resultado foi aplicado um critério de classificação que leva em consideração a maior porcentagem de cobertura, maior porcentagem de identidade e maior porcentagem de similaridade com a proteína molde. O objetivo dessa classificação é identificar as sequências a serem usadas como alvo em programas/servidores de modelagem baseada em molde para teste. Os valores são apresentados na Tabela 3.

**Tabela 3:** Identificação do molde com os valores de cobertura, identidade e similaridade com a sequência alvo. As sequências destacadas em cinza foram as escolhidas para teste com diferentes programas/servidores de modelagem baseada em molde.

Gene	PDB id	Cobertura	Identidade	Similaridade
LbrM.10.1710	1LML	97%	71%	80%
LbrM.10.0570	1LML	97%	61%	72%
LbrM.10.1680	1LML	97%	60%	74%
LbrM.10.0500	1LML	96%	69%	79%
LbrM.10.1690	1LML	90%	71%	79%
LbrM.10.0530	1LML	88%	66%	76%
LbrM.10.0540	1LML	85%	62%	73%
LbrM.10.1580	1LML	82%	70%	80%
LbrM.10.1660	1LML	82%	61%	75%
LbrM.10.0470	1LML	82%	60%	74%
LbrM.10.0520	1LML	82%	60%	73%
LbrM.10.1570	1LML	81%	62%	75%
LbrM.10.1650	1LML	81%	61%	73%
LbrM.10.0480	1LML	81%	60%	71%
LbrM.10.1610	1LML	81%	60%	71%
LbrM.10.1620	1LML	81%	59%	73%
LbrM.10.1630	1LML	81%	59%	72%
LbrM.10.0590	1LML	80%	72%	81%
LbrM.10.1670	1LML	78%	63%	76%
LbrM.10.0550	1LML	77%	70%	80%
LbrM.10.0600	1LML	77%	68%	78%
LbrM.10.0560	1LML	76%	70%	80%

<sup>8</sup><http://blast.ncbi.nlm.nih.gov/Blast.cgi>

(continuação)

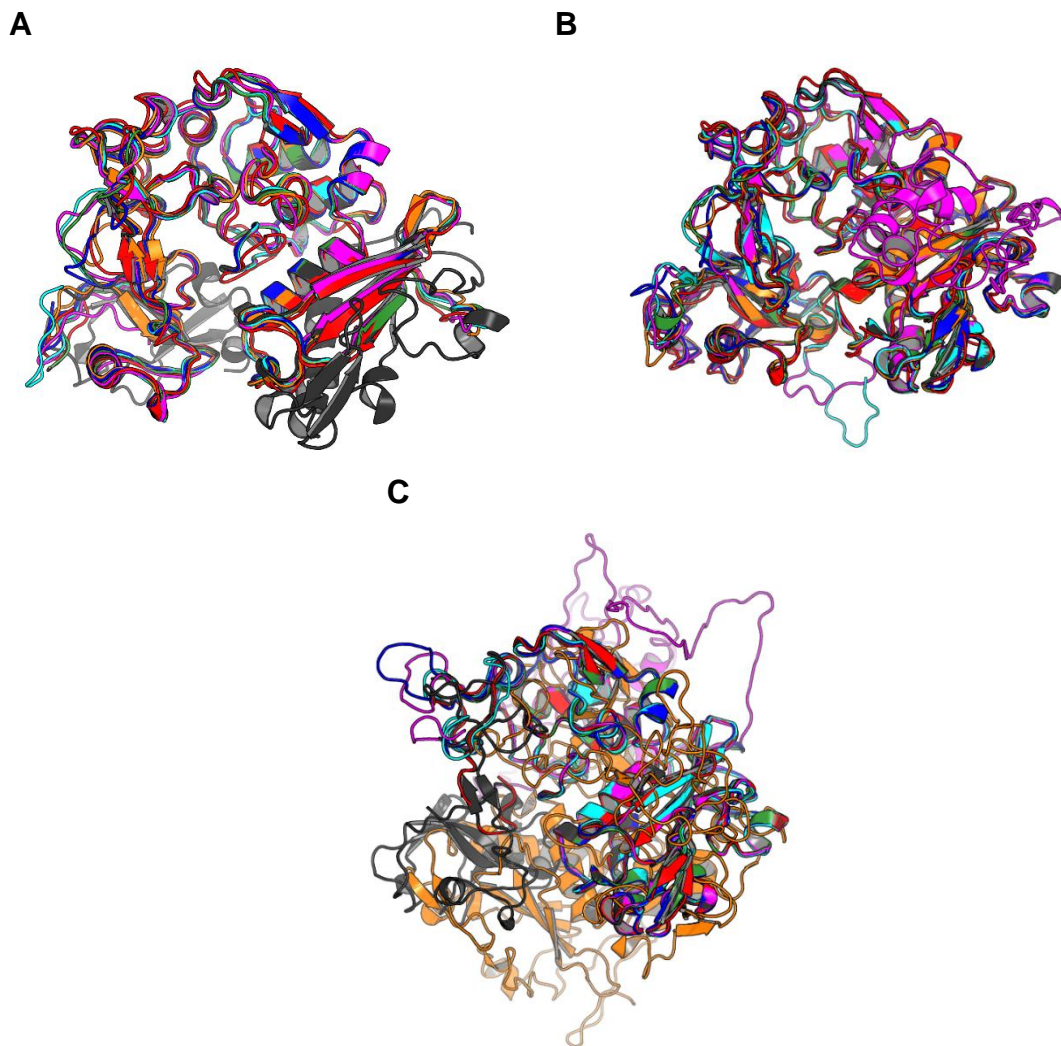
Gene	PDB id	Cobertura	Identidade	Similaridade
LbrM.10.0580	1LML	76%	70%	79%
LbrM.10.0610	1LML	76%	70%	79%
LbrM.10.0510	1LML	76%	69%	77%
LbrM.10.1560	1LML	74%	65%	77%
LbrM.10.1590	1LML	67%	63%	74%
LbrM.10.1640	1LML	66%	60%	73%
LbrM.10.1550	1LML	50%	65%	78%

Das sequências marcadas em cinza, a sequência LbrM.10.1710 é a que apresentou uma maior cobertura com a proteína molde enquanto que a sequência LbrM.10.1550 encontrou apenas 50% de cobertura. Além dessas duas, a sequência LbrM.10.1610 apresenta valor mediano na cobertura com a sequência molde. Usar essas três sequências em diferentes programas de modelagem é uma boa avaliação desses programas/servidores tendo em vista a diferença de cobertura e identidade entre proteína alvo e molde.

### 4.3 Comparação dos programas de modelagem baseada em molde

As sequências LbrM.10.1710, LbrM.10.1550 e LbrM.10.1610 foram submetidas aos programas/servidores Fugue, HHpred, i-tasser, Modeller, ModPipe e Swiss-Model e um modelo de cada servidor foi selecionado. No alinhamento estrutural entre a proteína usada como molde e os modelos construídos por esses programas, observamos que as estruturas secundárias se sobrepuseram (Figura 23). Para os métodos baseados no uso de estruturas moldes resolvidas experimentalmente, um modelo pode ser considerado de boa qualidade se o valor de RMSD (*Root Mean Square Deviation*) obtido pela sobreposição entre a cadeia peptídica de regiões conservadas do modelo gerado e a da estrutura molde for menor que 2 Å (Verli 2014).





**Figura 23:** Alinhamento do molde 1LML (cinza) com o melhor modelo construído da sequência (A) LbrM.10.1710, (B) LbrM.10.1610 e (C) LbrM.10.1550 com os programas i-tasser (magenta), Swiss-Model (verde), Modeller (laranja), ModPipe (azul), HHPred (ciano) e Fugue (vermelho).

É possível notar que na Figura 23c, as estruturas são diferentes e não se alinham por completo. Este fato é devido a porcentagem de cobertura entre a proteína molde com a sequência da proteína modelada ser de somente 50%. Quanto menor a cobertura, mais difícil será para os programas de modelagem comparativa conseguirem construir modelos que se aproximem do molde. Cada programa/servidor vai utilizar seus algoritmos para construir a proteína na região que não há informação de estrutura. Os modelos para LbrM.10.1710 e LbrM.10.1610 foram parecidos entre si e com o molde. Nas Tabelas 4, 5 e 6, as matrizes indicam os valores de RMSD entre a proteína molde e as estruturas obtidas para LbrM.10.1710, LbrM.10.1610 e LbrM.10.1550 por cada um dos



programas. Os valores marcados em cinza são os resultados que mais divergem dos demais.

**Tabela 4:** RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1710.

	Molde 1LML	Fugue	HHpred	i- tasser	Modeller	ModPipe	Swiss- Model
Molde 1LML	0	1,499	0,489	0,752	0,516	0,595	0,400
Fugue		0	1,496	1,442	1,539	1,521	1,409
HHpred			0	0,721	0,465	0,459	0,388
i-tasser				0	0,892	0,834	0,721
Modeller					0	0,478	0,440
ModPipe						0	0,457
Swiss- Model							0

RMSD é calculado em Ångstrom (Å)

**Tabela 5:** RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1610.

	Molde 1LML	Fugue	HHpred	i- tasser	Modeller	ModPipe	Swiss- Model
Molde 1LML	0	1,377	0,429	0,469	0,568	0,469	0,370
Fugue		0	1,349	1,509	1,531	1,489	1,367
HHpred			0	0,456	0,499	0,443	0,347
i-tasser				0	0,614	0,559	0,450
Modeller					0	0,474	0,529
ModPipe						0	0,452
Swiss- Model							0

RMSD é calculado em Ångstrom (Å)

**Tabela 6:** RMSD entre as estruturas construídas por cada programa/servidor da sequência LbrM.10.1550.

	Molde 1LML	Fugue	HHpred	i- tasser	Modeller	ModPipe	Swiss- Model
Molde 1LML	0	0,298	0,434	0,238	20,387	0,416	0,374
Fugue		0	0,614	0,518	21,268	0,584	0,495
HHpred			0	0,507	21,565	0,379	0,413
i-tasser				0	25,252	0,496	0,409
Modeller					0	21,257	20,480
ModPipe						0	0,352
Swiss- Model							0

RMSD é calculado em Ångstrom (Å)

Conforme pode ser observado na Figura 23a, as estruturas obtidas para LbrM.10.1710 ficaram muito próximas. Na Tabela 4 são apresentados os valores de RMSD entre a estrutura modelada com cada programa/servidor utilizado. Pode-se observar que o modelo construído pelo servidor Fugue foi o que mais apresentou diferença com relação aos demais, tendo valor de RMSD maior que 1Å. Em segundo lugar, o servidor i-tasser apresentou valores de RMSD entre 0,721 Å e 0,892 Å, indicando que o modelo construído por ele também apresenta diferenças estruturais com relação aos demais. As principais diferenças entre as estruturas se dá nos loops da proteína e não nas estruturas de  $\alpha$ -hélice e folha- $\beta$ .

A Tabela 5 mostra resultados de RMSD maiores que 1 Å somente para o modelo construído pelo servidor Fugue. Os demais modelos apresentaram valores de RMSD abaixo de 0,614 Å. A Figura 23b mostra o alinhamento entre essas estruturas onde é possível ver a sobreposição das mesmas.

Assim como já citado anteriormente sobre a porcentagem de cobertura da proteína LbrM.10.1550, as estruturas não se sobrepuseram completamente. Dentre os programas utilizados para a modelagem dessa sequência, o programa Modeller apresentou os piores resultados (Tabela 6), indicando RMSD maior que 20 Å entre os modelos construídos pelos demais servidores. Ao analisarmos os valores de RMSD dos modelos construídos com relação ao molde, observamos que o modelo que mais se aproximou do molde utilizado foi o modelo construído pelo servidor i-tasser, com RMSD de 0,238 Å. Este servidor obtêm o melhor resultado com

proteínas de baixa cobertura devido à metodologia utilizada na construção do modelo.

#### 4.3.1 Validação da estrutura 3D do conjunto teste

Para avaliar a estereoquímica dos modelos, foi gerado o gráfico de Ramachandran para as estruturas construídas pelos seis programas/servidores. Além dessa análise, os valores do fator G, pontuação QMEAN6, Z-Score também foram avaliados. O resultado para as estruturas LbrM.10.1710, LbrM.10.1610 e LbrM.10.1550 estão apresentados nas Tabelas 7, 8 e 9.

**Tabela 7:** Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1710.

		Fugue	HHpred	i-tasser	Modeller	ModPipe	Swiss-Model
Ramachandran (%)	Regiões favoráveis	81,3	92,3	86,2	91,6	91,4	91,3
	Regiões permitidas	11,5	6,9	9,2	6,9	6,6	7,5
	Generosamente permitidas	3,6	0,4	1,5	0,8	1,2	0,4
	Regiões não permitidas	3,6	0,4	3,1	0,8	0,8	0,8
Fator-G	Diedros	-0,49	-0,04	-0,24	-0,11	-0,10	-0,12
	Covalente	0,59	0,64	0,63	0,64	0,65	0,64
	Global	-0,05	<b>0,23</b>	0,11	0,19	0,20	0,19
	QMEAN6	0,586	0,623	0,579	0,59	<b>0,641</b>	0,628
	Z-Score	-2,037	-1,719	-2,214	-2,087	<b>-1,463</b>	-1,581

Os valores apresentados são do melhor modelo de modelagem da estrutura LbrM.10.1710 em cada um dos programas/servidores usados.

**Tabela 8:** Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1610.

		Fugue	HHpred	i-tasser	Modeller	ModPipe	Swiss-Model
Ramachandran (%)	Regiões favoráveis	85,0	91,6	81,3	88,7	90,1	92,5
	Regiões permitidas	8,6	7,0	11,9	9,0	8,2	6,2
	Generosamente permitidas	3,4	0,6	4,1	1,0	0,5	0,5
	Regiões não permitidas	2,9	0,8	2,7	1,4	1,2	0,7
Fator-G	Diedros	-0,47	-0,11	-0,24	-0,21	-0,13	-0,12
	Covalente	0,60	0,65	0,61	0,63	0,63	0,64
	Global	-0,04	<b>0,19</b>	0,10	0,13	0,17	<b>0,19</b>
	QMEAN6	0,591	0,537	0,535	0,559	<b>0,653</b>	0,648
	Z-Score	-2,104	-2,555	-2,579	-2,308	<b>-1,343</b>	-1,402

Os valores apresentados são do melhor modelo de modelagem da estrutura LbrM.10.1610 em cada um dos programas/servidores usados.

**Tabela 9:** Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para a estrutura LbrM.10.1550.

		Fugue	HHpred	i-tasser	Modeller	ModPipe	Swiss-Model
Ramachandran (%)	Regiões favoráveis	83,2	90,5	69,7	76,0	92,4	91,9
	Regiões permitidas	15,4	8,2	24,4	17,5	5,8	6,5
	Generosamente permitidas	0,4	0,4	3,6	4,4	1,1	0,0
	Regiões não permitidas	1,1	0,8	2,3	2,1	0,7	1,5
Fator-G	Diedros	-0,46	-0,14	-0,46	-0,39	-0,07	-0,09
	Covalente	-5,90	0,65	-1,49	0,61	0,64	0,64
	Global	-2,90	0,18	-0,96	0,02	<b>0,22</b>	0,20
	QMEAN6	0,661	0,497	0,47	0,178	0,689	<b>0,732</b>
	Z-Score	-1,314	-2,994	-3,296	-6,643	-0,997	<b>-0,442</b>

Os valores apresentados são do melhor modelo de modelagem da estrutura LbrM.10.1550 em cada um dos programas/servidores usados.

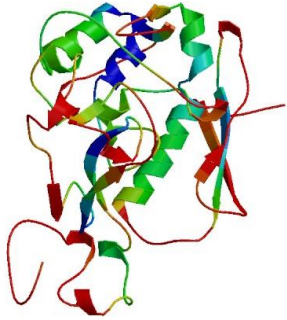
De acordo com os resultados do gráfico de Ramachandran, os servidores HHpred e Swiss-Model foram os que apresentaram as maiores porcentagens de aminoácidos em regiões favoráveis somados às regiões permitidas, com valor médio de 98,8% e 98,6%, respectivamente. Já o resultado do Fator-G, que indica com valores negativos inconsistências na geometria total da estrutura, valores mais positivos indicam os melhores resultados desse parâmetro. Este parâmetro não teve definição sobre qual programa apresentou melhor resultado visto que para a estrutura LbrM.10.1710, o servidor HHpred apresentou o resultado mais alto enquanto que para LbrM.10.1610, os servidores HHpred e Swiss-Model apresentaram o mesmo valor e para LbrM.10.1550, o programa ModPipe é o de maior valor no parâmetro.

A análise do QMEAN6 indica o programa ModPipe como o melhor para a construção dos modelos visto que os valores são mais próximos de 1, indicando maior probabilidade de estarem corretos. Porém, para a estrutura LbrM.10.1550, o servidor Swiss-Model é o que apresentou maior valor e para as outras sequências mostrou valor bem próximo do resultado para o programa ModPipe.

O último parâmetro avaliado nestas Tabelas é o Z-score. Este parâmetro é uma medida absoluta da qualidade do modelo. O valor é comparado com as pontuações de estruturas referência de alta resolução resolvidas por cristalografia de raio-x. Os valores maiores indicam que a estrutura do modelo é mais provável de ser correta. Com isso, temos o programa ModPipe e servidor Swiss-Model com os melhores valores.

Com o objetivo de avaliar os erros na estrutura 3D, a análise do gráfico gerado pelo ANOLEA indica regiões com potencial de confiança do modelo. Para representar este resultado, a Figura 24 ilustra as estruturas 3D onde a coloração é dada de acordo com um gradiente de cor variando do azul (regiões mais confiáveis) ao vermelho (regiões não confiáveis).

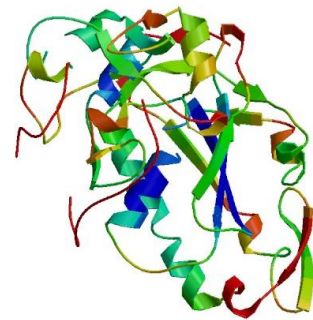
Fugue: LbrM.10.1710



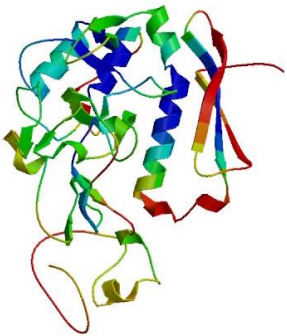
Fugue: LbrM.10.1610



Fugue: LbrM.10.1550



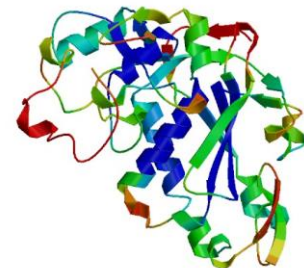
HHpred: LbrM.10.1710



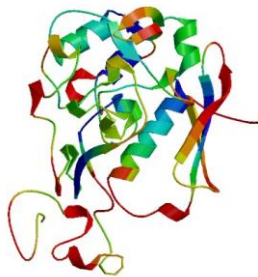
HHpred: LbrM.10.1610



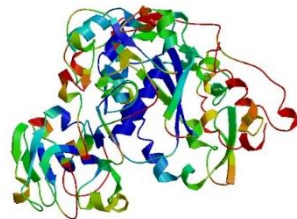
HHpred: LbrM.10.1550



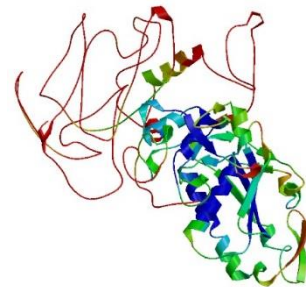
i-tasser: LbrM.10.1710



i-tasser: LbrM.10.1610



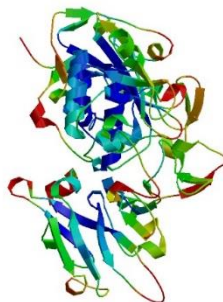
i-tasser: LbrM.10.1550



Modeller: LbrM.10.1710



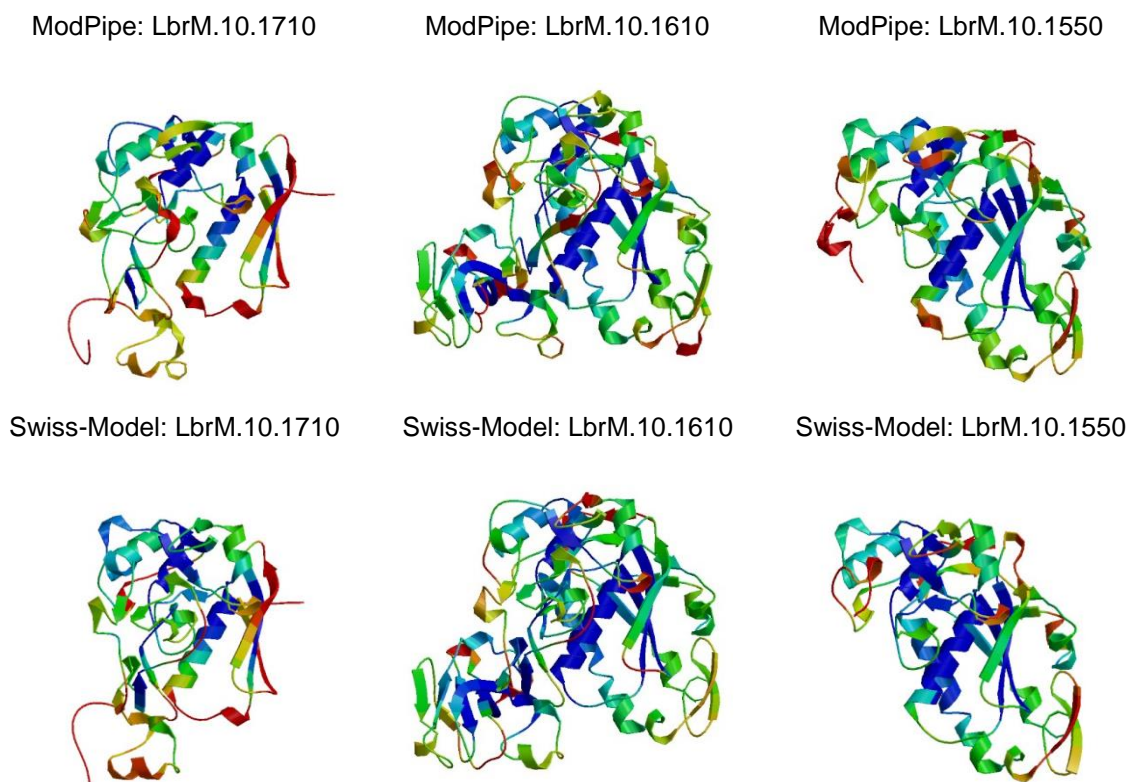
Modeller: LbrM.10.1610



Modeller: LbrM.10.1550



(continuação)



**Figura 24:** Estruturas geradas pelo ANOLEA. O gradiente de cor define regiões mais confiáveis (azul) e regiões não confiáveis (vermelho).

A Figura 24 mostra as estruturas obtidas com cada um dos programas/servidores para as três sequências escolhidas. A estrutura LbrM.10.1710, que possui maior cobertura com a proteína molde, apresentou menos regiões em vermelho na estrutura modelada com o programa HHpred. Para as demais, LbrM.10.1610 e LbrM.10.1550, o programa Swiss-Model foi o melhor. Além disso, vale destacar que os únicos programas/servidores que consideraram o íon Zn na estrutura do modelo foram os programas Modeller e Swiss-Model.

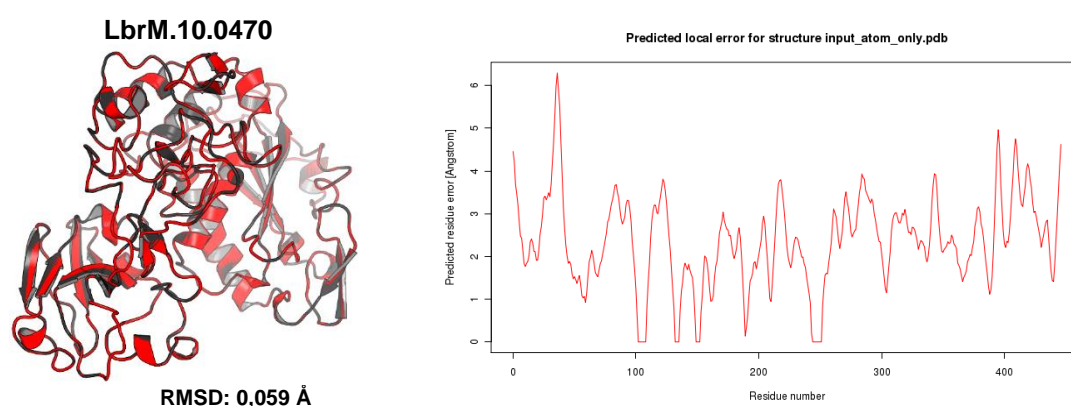
Levando em consideração o melhor resultado de cada programa/servidor para as três estruturas avaliadas, os resultados se dividem principalmente entre os servidores ModPipe e Swiss-Model. A tabela 10 mostra o programa/servidor que obteve o melhor resultado considerando cada um dos critérios avaliados. Além disso, o servidor Swiss-Model foi capaz de construir os modelos levando em consideração o íon presente na estrutura molde e por este motivo, este servidor foi o escolhido para a modelagem das demais sequências do conjunto de dados.

**Tabela 10.** Resumo dos servidores/programas que obtiveram o melhor resultado em cada um dos critérios avaliados para as sequências do conjunto teste.

Critérios	Estruturas		
	LbrM.10.1710	LbrM.10.1610	LbrM.10.1550
<b>RMSD</b>	Swiss-Model	Swiss-Model	i-tasser
<b>Ramachandran</b>	HHpred	Swiss-Model	HHpred
<b>Fator-G</b>	HHpred	Swiss-Model/HHpred	ModPipe
<b>QMEAN6</b>	ModPipe	ModPipe	Swiss-Model
<b>Z-score</b>	ModPipe	ModPipe	Swiss-Model
<b>ANOLEA</b>	HHpred	Swiss-Model	Swiss-Model

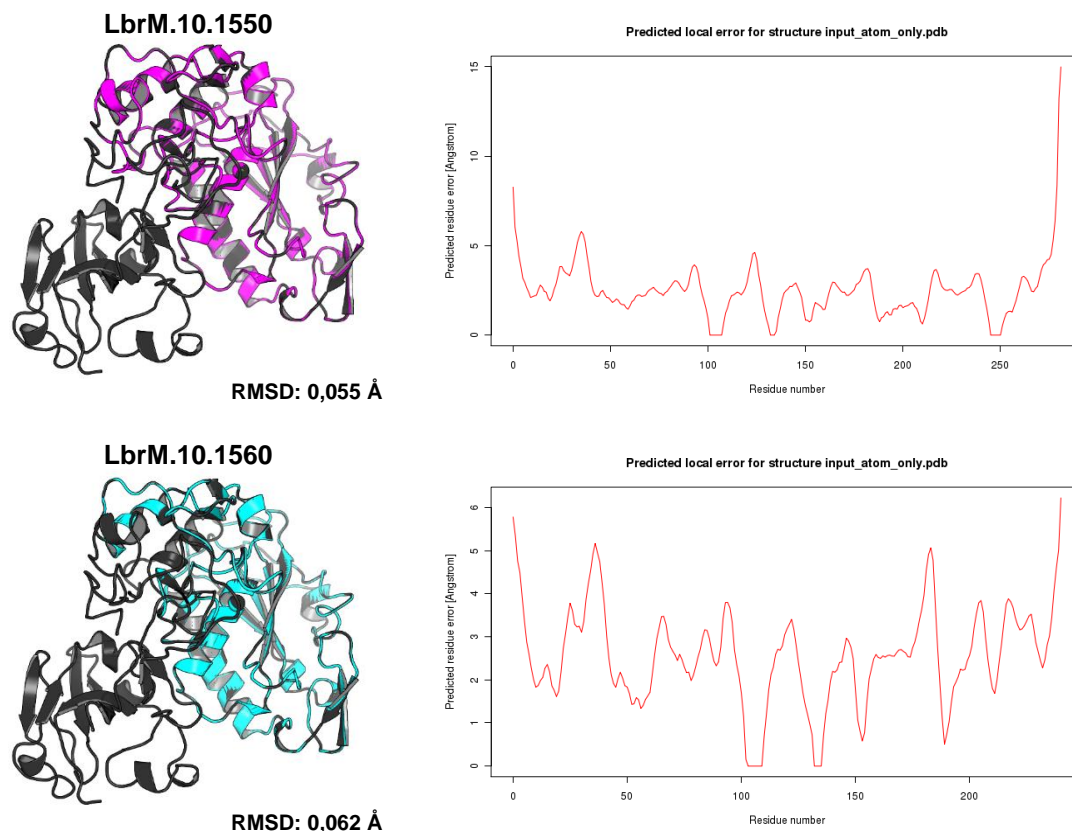
#### 4.4 Construção e Validação das demais estruturas com o servidor Swiss-Model

As vinte e seis sequências restantes foram submetidas ao servidor Swiss-Model. O servidor retorna três modelos para cada sequência e a estrutura de maior pontuação QMEAN6 foi a escolhida para validação. O alinhamento estrutural entre o molde e o alvo demonstra que há identidade estrutural entre elas. A Figura 25 mostra este resultado para a estrutura LbrM.10.0470, LbrM.10.1550 e LbrM.10.1560 com o gráfico de energia DOPE. As demais estruturas estão disponíveis no Anexo D.



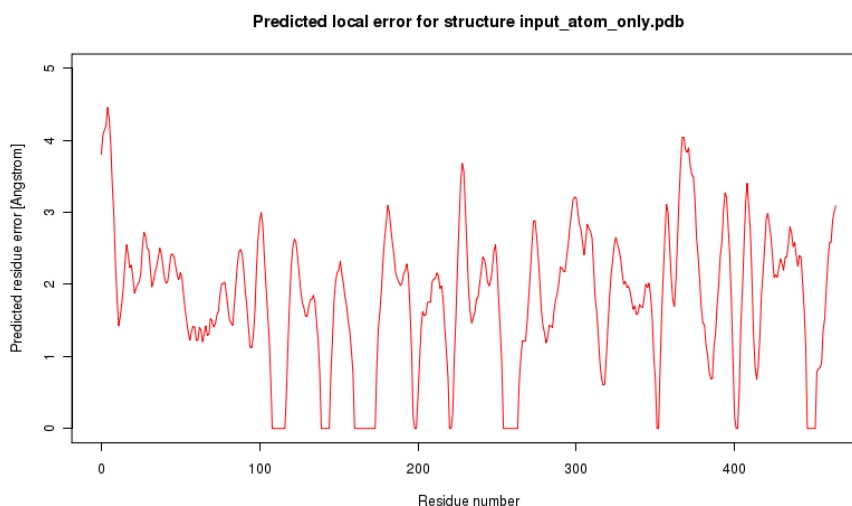


(continuação)



**Figura 25:** Alinhamento estrutural entre proteína molde (preto) e modelo construído pelo programa Swiss-Model e o gráfico DOPE correspondente.

Com o gráfico DOPE pode-se observar que para a estrutura LbrM.10.1550 há um pouco mais que 250 resíduos e para LbrM.10.1560, há um pouco mais que 200 resíduos enquanto que para LbrM.10.0470 há mais de 400 resíduos, se sobrepondo assim a estrutura da proteína molde. Este gráfico mostra as regiões na proteína que apresentam valores de erro por resíduo e podem ser comparados com o gráfico DOPE da proteína molde (Figura 26).



**Figura 26:** Gráfico DOPE da estrutura PDB id 1LML, usada como molde para a construção dos modelos de *L. V. braziliensis*.

Os parâmetros usados para validação destes modelos estão apresentados na Tabela 11. Todas as estruturas mostraram que mais de 85% dos aminoácidos se encontram em regiões favoráveis, e, se somado com as regiões permitidas, este número aumenta para 95%. A qualidade global dada pelo fator-G variou de 0,04 a 0,37 entre os modelos. Valores entre 0 e 0,5 são aceitáveis. Os valores da pontuação do QMEAN6 estão acima de 0,6, sendo que valores mais próximos de 1 indicam um modelo de melhor qualidade. Modelos de baixa qualidade têm pontuação Z fortemente negativa. Quanto mais positivo este valor, melhor o modelo. Após validação dos modelos, estes foram utilizados como estrutura inicial para as simulações de dinâmica molecular.

**Tabela 11:** Parâmetros de validação estatística dos modelos LbrM.10.0470, LbrM.101550 e LbrM.10.1560.

		LbrM.10.0470	LbrM.10.1550	LbrM.10.1560
<b>Ramachandran (%)</b>	Regiões favoráveis	85,4	87,7	87,7
	Regiões permitidas	12,9	10,8	11,4
	Generosamente permitidas	1,2	0,4	0,5
	Regiões não permitidas	0,5	1,2	0,5
<b>Fator-G</b>	Diedros	-0,09	-0,06	-0,08
	Covalente	0,24	0,35	0,35
	Global	0,05	0,10	0,10
	QMEAN6	0,625	0,746	0,705
	Z-Score	-1,671	-0,292	-0,680

É importante destacar que o modelo LbrM.10.1560 apresenta uma quantidade menor de aminoácidos (340 resíduos) com relação à proteína molde (478 resíduos), porém a região do sítio catalítico está presente na estrutura modelada. Já o modelo LbrM.10.1550, apesar de possuir em sua estrutura primária 559 resíduos, o servidor Swiss-Model construiu o modelo com 297 aminoácidos com base no alinhamento realizado (Figura 27). Contudo, a região do sítio catalítico foi preservada e os 263 aminoácidos restantes (contidos no início de sua estrutura primária) não foram levados em consideração. As demais estruturas se sobrepuseram no alinhamento entre o modelo e a proteína molde (Anexo D).

```

Model_01 MDVSRGRRRARAPWPAVVRFL LATSQRVTLSPAQHASAGRPRRLRHSPSSDTRPIRGWKQQPE 65
1lml.1.A -----

Model_01 GVGPRLLLVAAARPWGAREAAACGAPRPCFPLGFSAPHSFSA LPPRYLVHAARTTTVH LPLSLR 130
1lml.1.A -----

Model_01 SAPPPLHTLARTRSTTGTASMSRDRSSTPRRRSDAARLMRLAAAGLVMAGAAA VWAQAGHHCI 195
1lml.1.A -----

Model_01 HDKLQARVLQSV AQQRPPGSSVSALGLPYVSADWALADSTSPSV AHSADWGT LRI LLSVEDLNDP 260
1lml.1.A ----->DQVNWGLR L IAVS EDDLDF 22

Model_01 DYYCSYVGQLVNNHAGALDMCKAED ILTDAKRNTLVGSIIPQALQLHTERL KVRQVQGSWKVTGM 325
1lml.1.A RYHRCARVGGV HAGA C AED ILT EKRDI LVS IPQA QLHT ER LK QVQGV KVTDM 87

Model_01 TGDVCGTFKVP EAHVT VGVSNADPFVLYVASVPSKPGVM AWAVMCQAF FDDRPAVGVINIPAA YIR 390
1lml.1.A VGD CGI FKVP CAHT ECG SNTD FVMYY SVPS E EGVLA WATTCQ FSDGH PAVGV IIPAA IA 152

Model_01 SAYDQIMLRSVTHEMAHALGFDLLLFEE S DLLDEVNNLRGRDYEVPV LNSPTVVAKAREQYGCPT 455
1lml.1.A SRYDQLVTRVVTHEMAHALGFS GEEFEDARI VAVN PNRGKNFI VPV NS STAVAKAREQYGCPT 217

Model_01 LTFLEVEDTGDGSTAGSHLKR RN AKDELMAPVMGAGYYTALT MATFEDFGFYRVDFTKAEVMPWG 520
1lml.1.A LEV LEVED QGGAGSAGSH KMRNAQDEL MAPA AAGYYTALT MATFED FG FYRVDFTKAEVMPWG 282

Model_01 RALQLRLHHQEVHGG QHHAVAGDVLQLHREFVSVPHPPP 559
1lml.1.A QNACCAFLT NKC ----- 297

```

**Figura 27:** Alinhamento entre a sequência LbrM.10.1550 e a proteína molde 1LML. Este alinhamento foi realizado pelo servidor Swiss-Model para a construção do modelo 3D desta proteína.

#### 4.5 Protonação dos resíduos que coordenam o íon Zinco

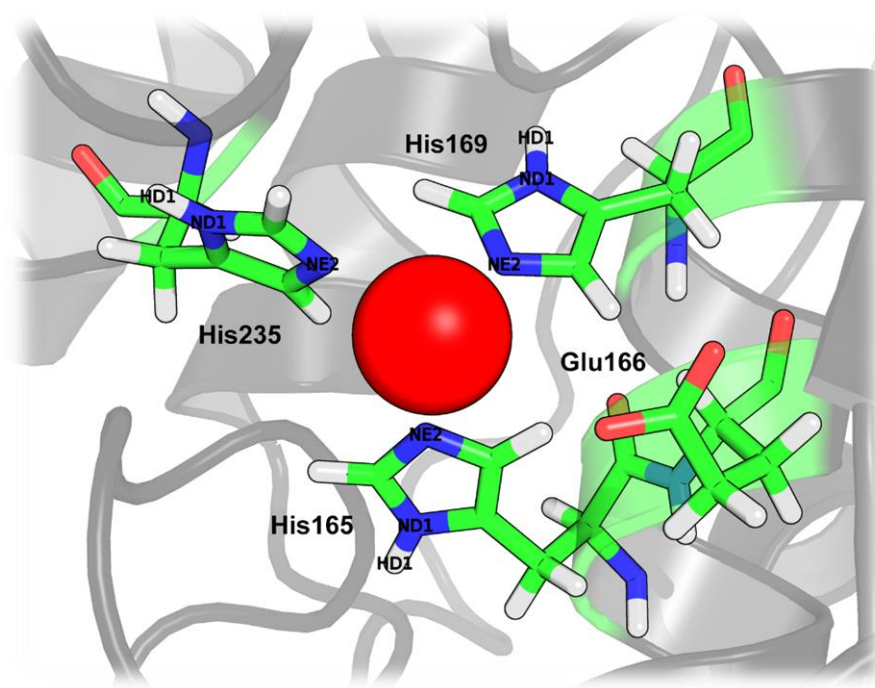
A estrutura de código PDB id 1LML, usada como molde na modelagem comparativa, foi submetida ao servidor H++ para a verificação da protonação dos resíduos de coordenação do metal (Tabela 12). Os parâmetros utilizados foram: salinidade = 0,15 M,  $\epsilon_{int} = 10$ ,  $\epsilon_{ext} = 80$  e pH = 7,0. O ponto isoelétrico foi de 5,12.

Segundo os resultados obtidos, os resíduos His165, His169 e His235, foram protonados no N $\delta$  (Figura 28), de acordo com o papel desses resíduos de coordenarem o átomo Zn. O valor estimado para Glu166 indica que este resíduo se encontra desprotonado.

**Tabela 12:** Estado de protonação das Histidinas do sítio catalítico calculado com o servidor H++.

Resíduos	$\epsilon_{int} = 10$	Resíduos	$\epsilon_{int} = 10$	Resíduos	$\epsilon_{int} = 10$	Resíduos	$\epsilon_{int} = 10$
Val1	7,5	Asp86	3,6	Tyr213	> 12,0	Asp350	< 0,0
Arg3	> 12,0	Asp90	4,0	Asp216	3,5	Tyr351	10,9
Asp4	3,8	Asp94	1,9	Glu219	4,6	Tyr358	11,3
Arg11	> 12,0	Lys96	> 12,0	Tyr220	11,8	Asp360	3,4
Glu17	2,8	His101	5,7	Glu222	< 0,0	Arg366	> 12,0
Asp18	< 0,0	Glu104	2,9	Glu224	< 0,0	Glu369	3,7
Asp21	1,4	Asp110	< 0,0	Asp225	< 0,0	His371	6,0
Tyr24	11,4	Tyr114	> 12,0	<b>His235</b>	<b>6,1</b>	Asp382	2,9
His25	7,0	Glu121	4,8	Lys237	> 12,0	Arg385	> 12,0
Arg28	> 12,0	Glu122	5,1	Arg239	> 12,0	Asp388	1,7
His32	6,8	Asp136	0,8	Asp243	0,9	Arg392	> 12,0
Lys34	> 12,0	His138	7,6	Glu244	< 0,0	Lys394	11,6
Asp35	2,7	Arg154	> 12,0	Tyr254	> 12,0	Tyr397	10,4
His36	11,3	Tyr155	10,5	Tyr255	> 12,0	Asp407	3,1
Glu46	4,9	Asp156	1,6	Asp265	2,2	Arg411	> 12,0
Asp47	2,0	Arg161	> 12,0	Tyr269	> 12,0	Tyr413	> 12,0
Glu52	3,3	<b>His165</b>	<b>10,2</b>	Asp272	1,1	His418	6,2
Lys53	> 12,0	<b>Glu166</b>	<b>0,2</b>	Lys275	> 12,0	Asp422	3,8
Arg54	> 12,0	<b>His169</b>	<b>7,7</b>	Glu277	< 0,0	Tyr423	> 12,0
Asp55	1,0	Glu179	4,6	Lys293	> 12,0	Arg431	> 12,0
Lys59	11,1	Asp180	4,1	Glu296	3,8	Glu433	4,0
His60	4,4	Arg182	> 12,0	Arg311	> 12,0	Glu442	3,0
His69	2,6	Arg191	> 12,0	Arg316	> 12,0	Tyr446	10,3
Glu71	2,0	Lys193	> 12,0	Arg326	> 12,0	Tyr452	11,6
Arg72	> 12,0	Asp196	4,1	His327	6,2	Glu454	3,3
Lys74	10,7	Lys208	11,7	Tyr333	> 12,0	Lys464	> 12,0
Lys81	11,4	Arg210	> 12,0	Tyr336	> 12,0	Asp465	6,0
Lys83	11,3	Glu211	3,9	Asp339	2,7		

Os resíduos que coordenam o metal estão indicados em negrito.



**Figura 28:** Protonação dos resíduos His165, His169 e His235 que participam da coordenação do Zinco. O íon Zn está representado pela esfera vermelha.

#### 4.6 Avaliação da estabilidade da Dinâmica Molecular

Todos os sistemas foram simulados por 50 nanossegundos. Na Figura 29 é possível observar o valor de RMSD das estruturas simuladas representado em função do tempo.

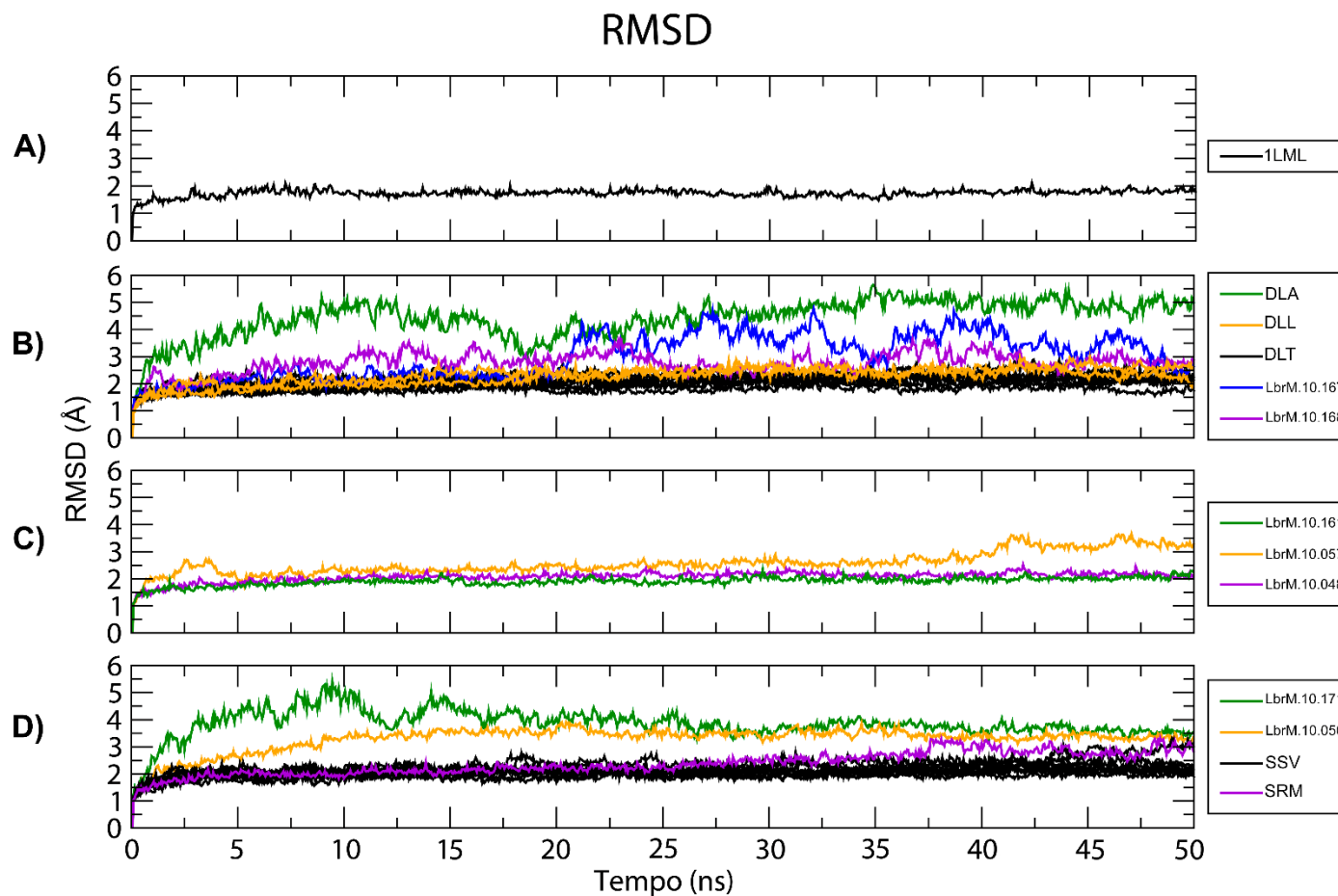
Na Figura 29A é possível ver que após 10ns a proteína, usada como molde na modelagem comparativa - PDB id 1LML, se estabiliza. Já para as estruturas com motivo DLX, observa-se um comportamento diferente entre as mesmas. A estrutura LbrM.10.1560, de motivo DLA, apresentou variação entre 3 Å e 5,5 Å ao longo da simulação, se diferenciando das demais. Este fato pode ser explicado pelo tamanho desta proteína que possui menos aminoácidos que as demais. Já as estruturas de motivo DLL, LbrM.10.0540 e LbrM.101550, apresentaram perfis similares se mantendo estabilizado em torno de 2,5 Å. Vale destacar que a estrutura LbrM.10.1550 também possui tamanho reduzido (<300 aa) e se estabilizou a partir de 10 ns variando entre 2 e 2,5 Å. As estruturas de motivo DLT mantiveram um perfil ao longo da simulação sendo consideradas estáveis após 10ns. Porém, duas estruturas deste motivo apresentaram comportamento diferenciado. A estrutura LbrM.10.1670 se manteve estável até 20 ns de simulação e depois teve uma

variação entre 2,5 e 4,5 Å. Já a estrutura LbrM.10.1680 não se estabilizou no tempo simulado tendo variação entre 2 e 3,5 Å.

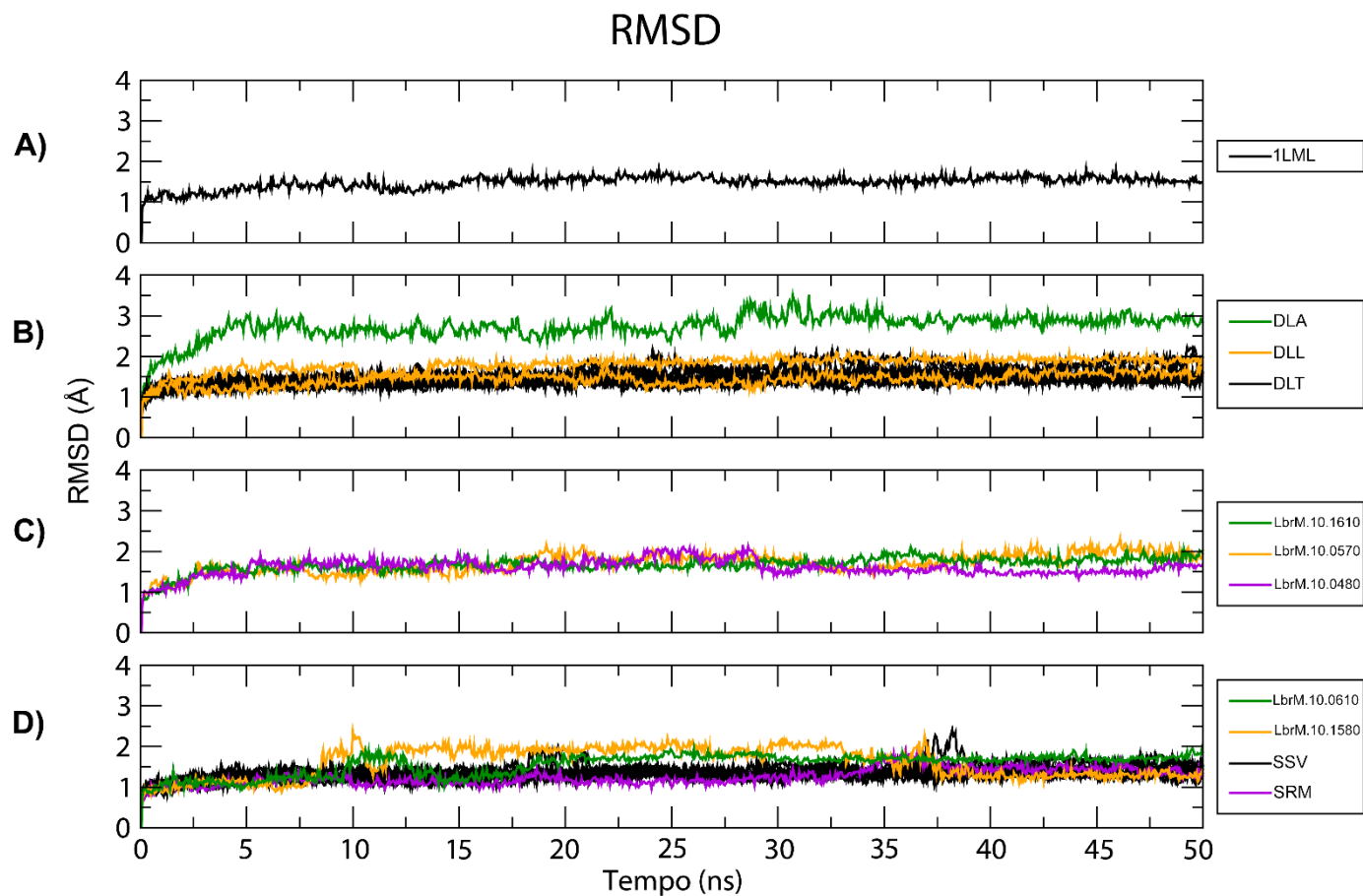
A Figura 29C mostra como as estruturas de motivo NRI se comportaram ao longo da simulação. Duas delas, LbrM.10.0480 e LbrM.10.1610 se estabilizaram após 10 ns em torno de 2 Å. Já a estrutura LbrM.10.0570 se manteve estabilizada entre 10 e 40 ns mas após esse tempo, teve uma variação na estrutura entre 3 e 3,5 Å. Em 29D, as estruturas de motivo SXX são apresentadas. A estrutura LbrM.10.0530, motivo SRM, se manteve estável até 30ns com RMSD em torno de 2 Å mas após este tempo, teve mudança conformacional variando entre 2,5 e 3 Å. As estruturas de motivo SSV se estabilizaram após 10 ns com RMSD em torno de 2 Å. Porém, duas estruturas deste motivo apresentaram perfis diferentes. A estrutura LbrM.10.0500 e LbrM.10.1710 começaram a simulação com grandes diferenças estruturais com relação a estrutura inicial, porém após 30 ns ambas se estabilizaram em torno de 3,5 Å.

Com o objetivo de analisar somente o comportamento do sítio catalítico, gráficos mostrando a variação da estrutura ao longo da simulação foram gerados somente com a região de inserção dessas proteínas (Figura 30). É possível observar que praticamente todas as estruturas tiveram a região de inserção estabilizada em torno de 1,5 Å. A estrutura de motivo DLA, se manteve estável, porém com RMSD entre 2,5 e 3 Å e as estruturas de motivo DLL se estabilizaram após 15 ns. Vale destacar as duas estruturas, LbrM.10.0610 e LbrM.10.1580, que tiveram a região de inserção estabilizada somente após 40 ns.

Após DM, a clusterização foi realizada e a estrutura representativa do maior cluster foi selecionada para as análises posteriores.



**Figura 29:** Desvio RMS dos átomos da cadeia principal (*backbone*) representado em função do tempo de simulação. A: proteína de código PDB id 1LML; B: estruturas com motivo DLX. A estrutura do motivo DLA está representada em verde, as estruturas do motivo DLL estão representadas em laranja e as de motivo DLT, estão em preto; C: estruturas de motivo NRI; D: estruturas de motivo SXX. A estrutura do motivo SRM está representada em violeta, e as estruturas de motivo SSV, são mostradas em preto.

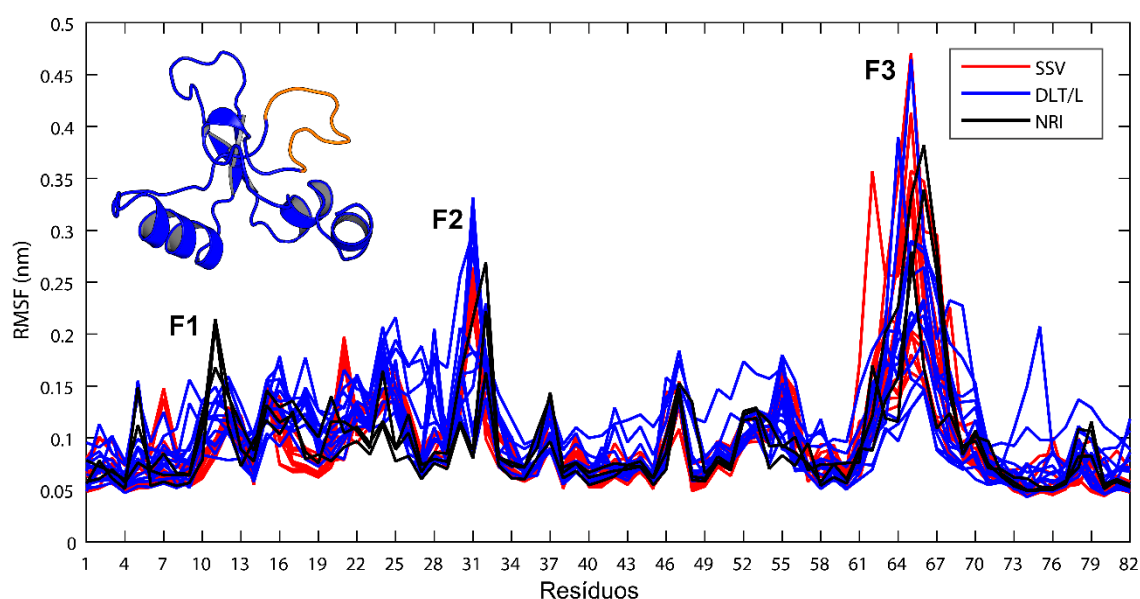


**Figura 30:** Desvios RMS dos átomos da cadeia principal (*backbone*) da região de inserção das proteínas representado em função do tempo de simulação. A: proteína de código PDB id 1LML; B: estruturas com motivo DLX. A estrutura do motivo DLA está representada em verde, as estruturas do motivo DLL estão representadas em laranja e as de motivo DLT, estão em preto; C: estruturas de motivo NRI; D: estruturas de motivo SXX. A estrutura do motivo SRM está representada em violeta, e as estruturas de motivo SSV, são mostradas em preto.



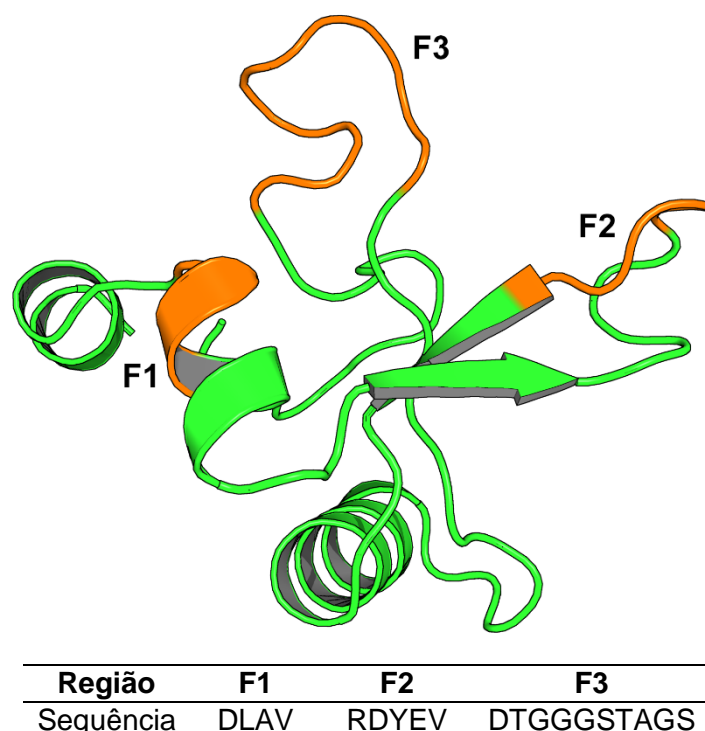
## 4.7 Avaliação da flutuação dos resíduos da região do sítio catalítico

Para mais detalhes sobre a região de inserção, a Figura 31 mostra os valores de RMSF para os modelos simulados durante a dinâmica molecular. As flutuações da maioria das estruturas apresentam variações nos mesmos resíduos comparados com a estrutura usada como molde, PDB id 1LML. Os resíduos presentes na região de inserção entre as posições 61 e 70 são os que apresentam maior flutuação por estar localizado em uma região de *coil*, conforme mostra a estrutura no topo esquerdo.



**Figura 31:** RMS de Flutuação representado por resíduo da região de inserção das proteínas. A estrutura no topo à esquerda apresenta em laranja, a região de maior flutuação (F3) correspondente aos resíduos 61 a 70.

As regiões de maior flutuação são mostradas na Figura 32. Três regiões foram selecionadas e nomeadas de F1 a F3. A flutuação F1 tem uma parte de estrutura em alfa-hélice e a flutuação F2 e F3 são compostas por regiões de *coil*.

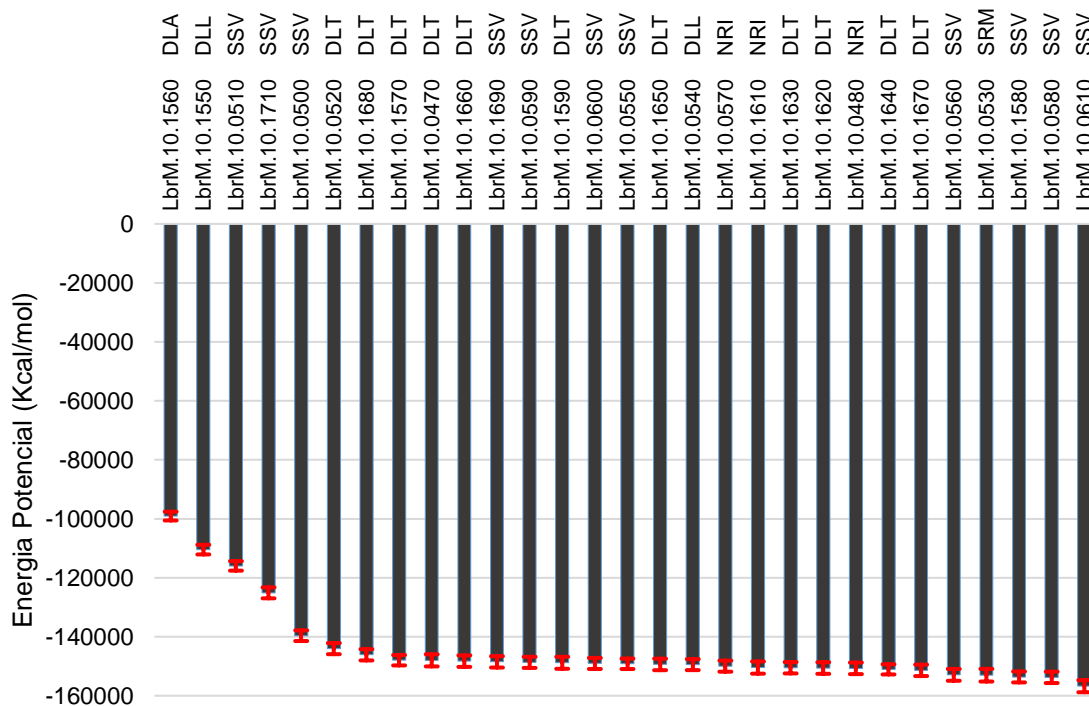


**Figura 32:** Estrutura de motivo DLX com destaque para as regiões de maior flutuação, representadas em laranja.

## 4.8 Energia Potencial

A energia potencial das proteínas não apresentou diferenças significativas em seus valores que pudessem ser relacionados com os motivos identificados. Na Figura 33, um gráfico da energia potencial total é mostrado. Pode-se observar que a estrutura que teve a maior energia de todas foi a LbrM.10.1560, de motivo DLA, seguida da LbrM.10.1550, de motivo DLL, e duas sequências de motivo SSV, LbrM.10.0510 e LbrM.10.1710. As demais sequências de motivo SSV, SRM, DLT, DLL e NRI apresentaram valores de energia entre -140000 kcal/mol e -150000 kcal/mol, o que não permite classificá-las em grupos distintos.

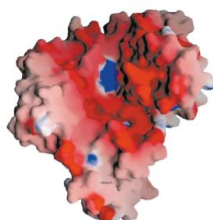
Vale destacar que a estrutura LbrM.10.1710, que obteve a quarta maior energia potencial, mostrou um comportamento diferenciado das demais estruturas de motivo SSV conforme apresentado no gráfico de RMSD da Figura 29.



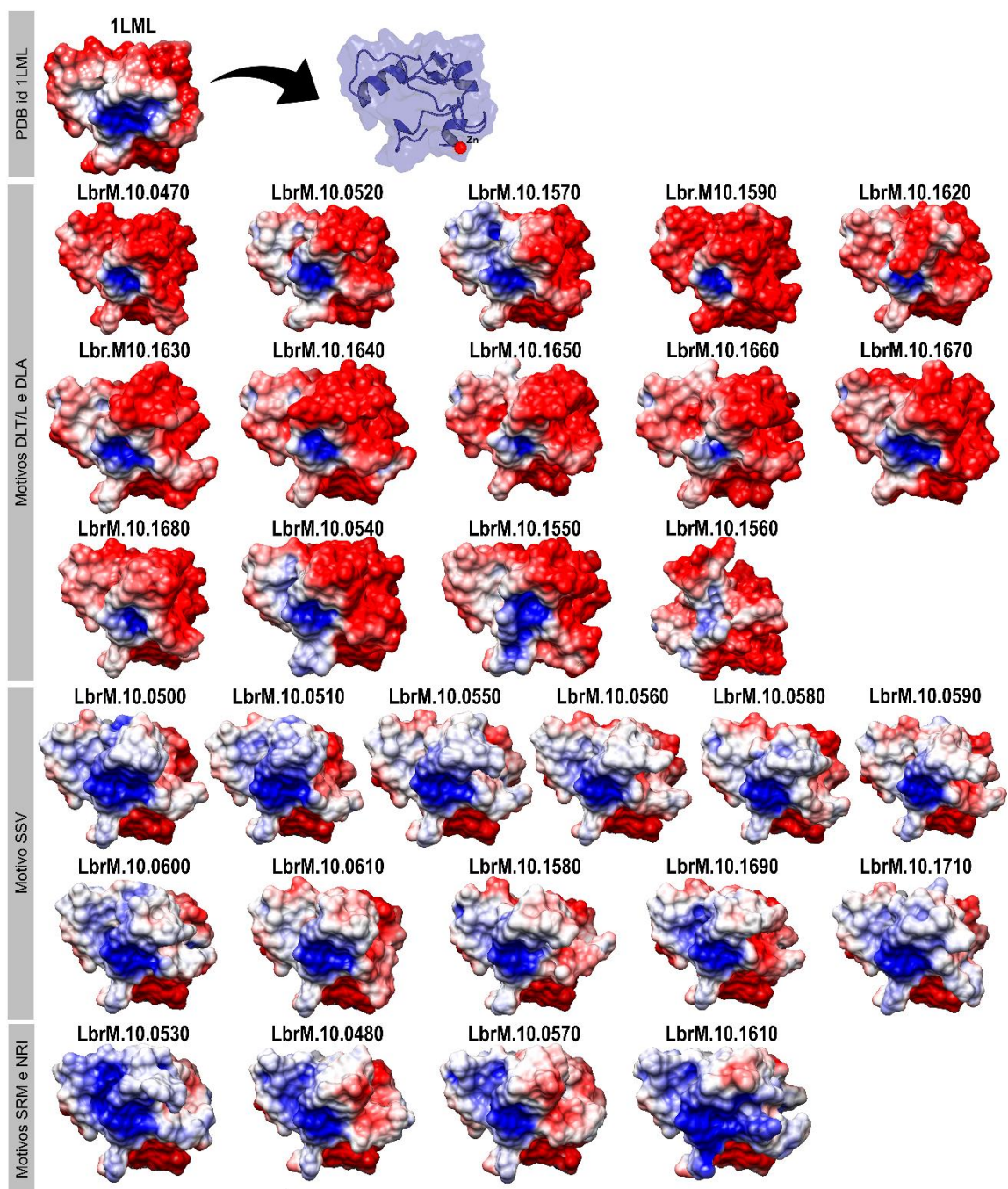
**Figura 33:** Energia potencial total das estruturas simuladas. No topo do gráfico, o motivo de cada estrutura é mostrado ao lado de seu nome identificador. A barra vermelha indica o desvio padrão.

## 4.9 Avaliação do perfil eletrostático

Com os resultados obtidos da DM, observamos que não houve diferença que pudesse classificar as estruturas em grupos distintos. Com esse objetivo, foi gerado o perfil eletrostático da região de inserção dessas enzimas. Primeiramente, fizemos uma comparação entre os perfis da estrutura PDB id 1LML, estrutura usada como molde na modelagem comparativa, e também esta estrutura simulada por DM com os mesmos parâmetros usados nos modelos. Para avaliar se o perfil obtido equivale ao observado pelo grupo que realizou a cristalografia, comparamos o perfil da figura publicada no artigo da proteína PDB id 1LML (Schlagenhauf, Etges *et al.* 1998) (Figura 34) com o perfil eletrostático das proteínas do cromossomo 10 (Figura 35). Segundo Silva-López (da Silva-López 2010), a faixa de pH ótimo em que as metaloproteases atuam varia de 7 a 9 e então decidimos usar pH 7 nesta análise.



**Figura 34:** Perfil eletrostático da proteína PDB id 1LML. Retirado de (Schlagenhauf, Etges *et al.* 1998), Figura 1b.



**Figura 35:** Diferenças no perfil eletrostático da região de inserção das metaloproteases de *Leishmania (V.) braziliensis* em pH 7. A localização do íon Zn está indicada na estrutura PDB id 1LML. A superfície molecular é colorida de acordo com o potencial eletrostático usando o software Chimera, onde vermelho, branco e azul correspondem potenciais ácidos, neutros e básicos, respectivamente. As estruturas estão divididas pelos motivos identificados na caixa cinza.

O perfil eletrostático da proteína PDB id 1LML, usada na simulação de DM, apresentou uma maior contribuição negativa na região de inserção. Este fato

comprova suas características após a DM já que o perfil eletrostático desta proteína apresentada no artigo que a descreve, é similar ao obtido.

Os perfis dos modelos simulados indicam, claramente, uma diferenciação entre as proteínas deste cromossomo. Proteínas com motivo DLX apresentaram um perfil com contribuição negativa, parecidas com a proteína utilizada como molde.

Há um segundo grupo, dos motivos, SXX e NRI que apresentam um perfil com maior contribuição positiva. Vale ressaltar que todas as proteínas foram estruturalmente alinhadas e foi obtida a mesma posição para a geração dos perfis, não sendo possível que ela estivesse em posição diferente para que a coloração do perfil mudasse.

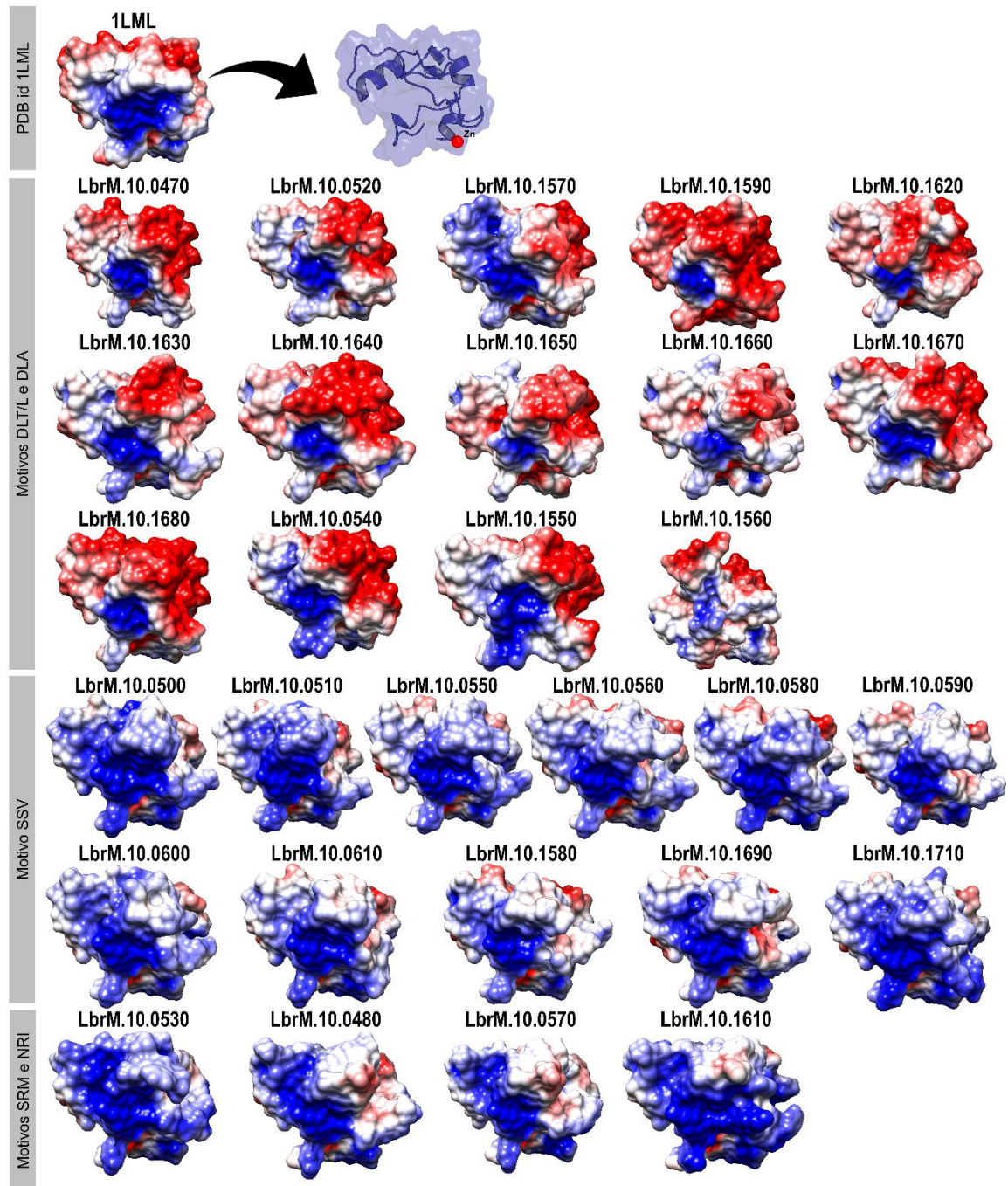
Um detalhe importante a ser citado é que todos os perfis foram gerados sem considerar o íon Zn a fim de comparação com a proteína PDB id 1LML, que não teve o íon inserido em seu perfil eletrostático. Por este motivo, a região central das proteínas não apresenta uma região mais positiva conforme se esperaria.

Pouco se sabe sobre as mudanças globais que acontecem a nível transcricional em ambos parasita e hospedeiro durante a infecção. Depledge *et al* (Depledge, Evans *et al.* 2009), estudaram 60 genes reguladores de *Leishmania (V.) braziliensis* e encontraram que 5 são preferencialmente expressos em amastigotas, 35 preferencialmente expressos em metacíclicos e, 11 em procíclicos. Nenhum dos genes estudados pertence ao cromossomo 10 e não há, na literatura, informação sobre o transcriptoma destes genes.

Em *Leishmania spp.*, a forma encontrada nos hospedeiros é amastigota. Levando em consideração que a multiplicação ocorre em um vacúolo resultante da fusão do fagossomo e lisossomo, e que o pH do fagolisossomo é ácido, em torno de 5, geramos um novo perfil eletrostático considerando este valor (Figura 36).

Esse novo perfil mostra que mesmo dentro de um ambiente de pH ácido, existirá os mesmos grupos encontrados no perfil eletrostático com pH 7. A diferença é que em pH 5, as proteínas possuem uma contribuição mais positiva na região de inserção.





**Figura 36:** Diferenças no perfil eletrostático da região de inserção das metaloproteases de *Leishmania (V.) braziliensis* em pH 5. A localização do íon Zn está indicada na estrutura PDB id 1LML. A superfície molecular é colorida de acordo com o potencial eletrostático usando o software Chimera, onde vermelho, branco e azul correspondem potenciais ácidos, neutros e básicos, respectivamente. As estruturas estão divididas pelos motivos identificados na caixa cinza.

## 4.10 Dendrograma estrutural

Os dados de estrutura obtidos pelo servidor 3V são mostrados na Tabela 13. Esses valores são usados como dados de entrada na análise da distância euclidiana realizada no programa estatístico R. O script criado para esta análise está detalhado no Anexo F.

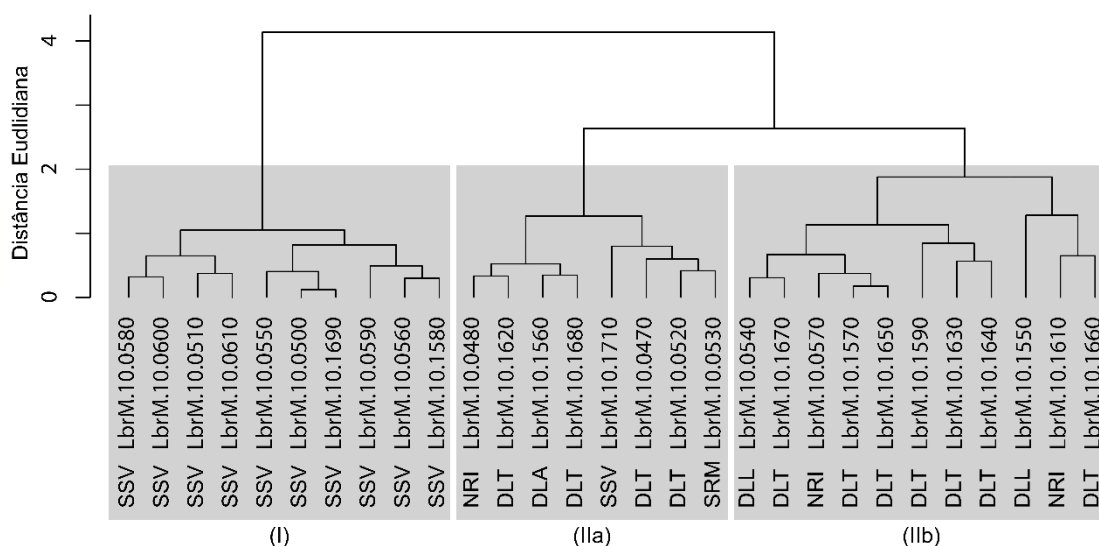
**Tabela 13.** Valores numéricos da região de inserção das metaloproteases.

Estrutura	Volume	Área de Superfície	Esfericidade	Raio Efetivo
LbrM.10.0470	14023	4101	0,69	10,26
LbrM.10.0480	14104	4168	0,68	10,15
LbrM.10.0500	13712	3983	0,70	10,33
LbrM.10.0510	13633	3931	0,70	10,40
LbrM.10.0520	13933	4060	0,69	10,29
LbrM.10.0530	13952	4116	0,68	10,17
LbrM.10.0540	14423	4119	0,70	10,50
LbrM.10.0550	13725	4052	0,68	10,16
LbrM.10.0560	13655	4070	0,68	10,06
LbrM.10.0570	14352	4126	0,69	10,43
LbrM.10.0580	13548	3937	0,70	10,32
LbrM.10.0590	13598	4090	0,67	9,97
LbrM.10.0600	13573	3968	0,69	10,26
LbrM.10.0610	13679	3951	0,70	10,39
LbrM.10.1550	14715	4302	0,67	10,26
LbrM.10.1560	14132	4125	0,69	10,28
LbrM.10.1570	14329	4156	0,69	10,34
LbrM.10.1580	13692	4086	0,68	10,05
LbrM.10.1590	14278	4353	0,65	9,84
LbrM.10.1610	14550	4243	0,68	10,29
LbrM.10.1620	14123	4210	0,67	10,06
LbrM.10.1630	14284	4254	0,67	10,07
LbrM.10.1640	14380	4245	0,67	10,16
LbrM.10.1650	14301	4160	0,68	10,31
LbrM.10.1660	14591	4162	0,69	10,52
LbrM.10.1670	14428	4171	0,69	10,38
LbrM.10.1680	14069	4125	0,68	10,23
LbrM.10.1690	13725	3991	0,69	10,32
LbrM.10.1710	13845	4097	0,68	10,14

O volume é calculado em  $\text{Å}^3$  e a área de superfície em  $\text{Å}^2$ .

O dendrograma gerado (Figura 37) divide as sequências em dois grupos, sendo o segundo subdividido em outros dois. O primeiro grupo é composto pelas estruturas de motivo SSV. O segundo grupo possui 19 elementos, sendo 8 pertencentes ao grupo IIa e 11, ao grupo IIb. O grupo IIa contém 4 estruturas de motivo DLT, a estrutura de motivo DLA, 1 estrutura de motivo NRI, e 1 estrutura de motivo SSV. Já o grupo IIb reuniu 7 estruturas de motivo DLT, 2 estruturas de motivo DLL e 2 estruturas de motivo NRI.

É notável que a maioria das estruturas de motivo SSV pertencem ao grupo I, reforçando a particularidade deste grupo e sugerindo a existência de uma micro-heterogeneidade no cromossomo 10 de *L. braziliensis*.



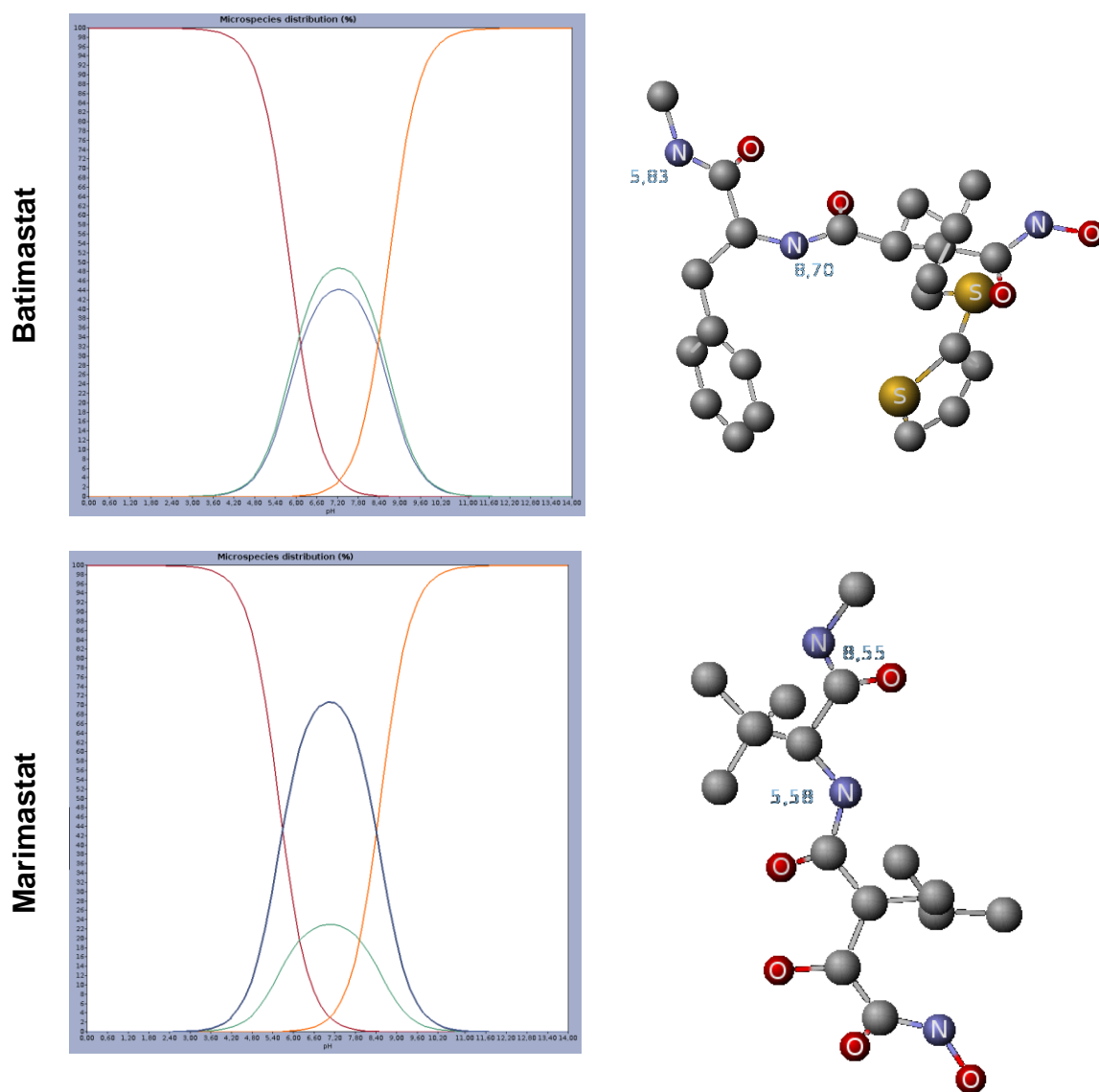
**Figura 37:** Dendrograma estrutural das metaloproteases do cromossomo 10 de *L. (V.) braziliensis*. A escala no eixo Y representa o peso baseado na distância euclidiana das variáveis normalizadas. Os blocos em cinza destacam a formação dos grupos I, IIa e IIb. O motivo que identifica a estrutura está próximo do nome da mesma.

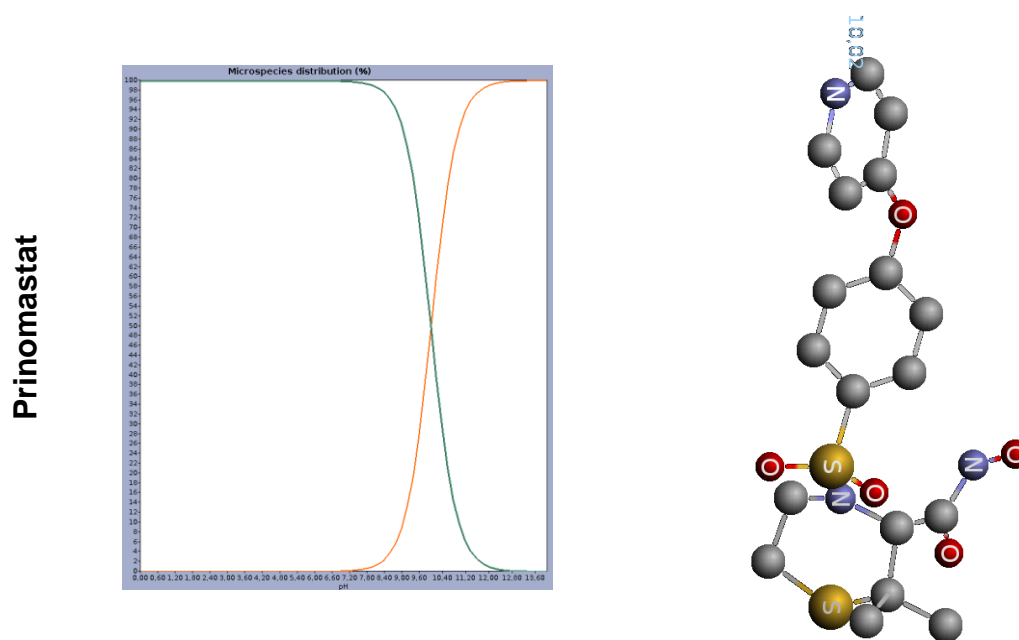


## 4.11 Protonação dos ligantes para *docking*

A análise de protonação dos ligantes foi realizada com o *plugin* pKa do programa Marvin. Este *plugin* calcula os valores de pKa de todos os ganhos e perdas de próton dos átomos com base na distribuição das cargas parciais. O gráfico gerado mostra a curva da distribuição de micro espécies por pH, o que permite a determinação da forma de maior probabilidade de protonação no pH especificado. A protonação dos ligantes foi realizada considerando a atuação em pH 7.

O resultado da protonação para as moléculas Batimastat, Marimastat e Prinomastat em pH 7 é mostrada na Figura 38 juntamente com os respectivos gráficos de distribuição de microespécies.





**Figura 38:** Curva da distribuição de micro espécies por pH e estado de protonação dos ligantes usados nas simulações de *docking*.

É possível observar que para as moléculas Batimastat e Marimastat, quatro estados de protonação foram obtidos. A curva de maior porcentagem em pH 7 foi a escolhida para utilizar a molécula correspondente no *docking*. A molécula Prinomastat apresentou uma distribuição diferente das demais, onde não consta nenhuma curva em pH 7. Com isso, conclui-se que a molécula Prinomastat não necessita de protonação para atuar neste pH.

#### 4.12 *Docking* Molecular: Análise das conformações por menor energia de ligação

O *docking* foi realizado com o programa Autodock4<sub>Zn</sub> para cada um dos ligantes com cada uma das 29 proteínas do cromossomo 10 de *Leishmania (V.) braziliensis*. Este programa gera 10 conformações para cada molécula. A análise do modo de ligação dos ligantes com as proteínas foi realizada com a conformação de menor energia. Não há informação de como é o modo de ligação real desses inibidores com metaloproteases de *Leishmania (V.) braziliensis* visto que não há nenhuma estrutura tridimensional conhecida e depositada em bancos de dados público. Portanto, esta análise tem como objetivo verificar o valor da energia de

ligação entre os ligantes com as proteínas e saber se há alguma relação com os motivos definidos. A Tabela 14 apresenta os valores de cada simulação realizada.

**Tabela 14:** Menor energia obtida entre cada inibidor de metaloproteases e as proteínas em estudo.

Motivo	Proteína	Inibidores		
		Batimastat	Marimastat	Prinomastat
SSV	LbrM.10.0600	<b>-8,79</b>	-6,05	<b>-8,83</b>
DLT	LbrM.10.1640	-6,95	-6,18	-8,05
DLT	LbrM.10.1670	-6,75	-5,43	-7,48
DLT	LbrM.10.1620	-6,73	-4,74	-6,61
DLT	LbrM.10.1650	-6,69	-5,88	-7,86
SSV	LbrM.10.0550	-6,52	-5,15	-7,14
DLT	LbrM.10.0520	-6,40	-4,32	-6,92
NRI	LbrM.10.1610	-6,36	-5,13	-8,38
SSV	LbrM.10.0610	-6,21	-5,49	-7,92
DLL	LbrM.10.1550	-6,19	<b>-7,20</b>	<b>-8,83</b>
DLL	LbrM.10.0540	-6,17	-4,56	-7,10
SRM	LbrM.10.0530	-6,13	-6,81	-7,98
DLT	LbrM.10.1630	-6,04	-4,80	-6,74
DLT	LbrM.10.1590	-6,02	-6,24	-7,87
DLT	LbrM.10.1570	-5,95	-5,07	-7,11
SSV	LbrM.10.0580	-5,85	-4,96	-7,64
SSV	LbrM.10.1580	-5,74	-4,54	-7,28
SSV	LbrM.10.1690	-5,70	-4,64	-7,02
SGP	1LML	-5,65	-4,61	-7,05
SSV	LbrM.10.0510	-5,64	-3,71	-7,49
DLT	LbrM.10.1660	-5,63	-5,19	-6,68
NRI	LbrM.10.0480	-5,63	-4,95	-6,57
SSV	LbrM.10.0560	-5,59	-4,63	-6,53
DLT	LbrM.10.0470	-5,55	-5,35	-7,34
SSV	LbrM.10.0590	-5,53	-4,36	-7,26
DLT	LbrM.10.1680	-5,44	-5,99	-7,12
NRI	LbrM.10.0570	-5,23	-4,01	-6,74
SSV	LbrM.10.0500	-5,08	-4,20	-7,83
DLA	LbrM.10.1560	-4,42	-3,96	-6,27
SSV	LbrM.10.1710	-4,28	-4,13	-5,83

Os valores de energia são calculados em kcal/mol.  
As linhas em cinza indicam as proteínas que tiveram a menor energia de ligação.  
Os motivos foram coloridos para melhor identificação.

Considerando as conformações de menor energia para cada um dos ligantes, temos que a estrutura LbrM.10.0600 (motivo SSV) apresentou menor valor de energia tanto para Batimastat quanto para Prinomastat. Já a proteína LbrM.10.1550 (motivo DLL), apresentou menor valor de energia para Marimastat e Prinomastat.

Esta tabela está classificada por ordem de menor energia obtida pelo ligante Batimastat. Pode-se observar que o topo da tabela é composto majoritariamente por proteínas de motivo DLX. Ao classificarmos esta tabela por ordem de menor energia levando em consideração os valores obtidos pelo ligante Marimastat, tivemos igual observação. Já para o ligante Prinomastat, as proteínas ranqueadas são majoritariamente estruturas de motivo SXX.

Este fato pode indicar que proteínas de motivo SXX podem ter maior especificidade em se ligar no inibidor Prinomastat enquanto que proteínas de motivos DLX e NRI teriam maior probabilidade de se ligarem em inibidores como Batimastat e Marimastat.

A maior energia de ligação foi obtida para a proteína LbrM.10.1710, tanto para Batimastat quanto para Prinomastat e para a proteína LbrM.10.0510 para o ligante Marimastat. A segunda maior energia foi obtida para a proteína LbrM.10.1560 para os três ligantes usados no *docking*, o que indica que essa proteína provavelmente teria pouca chance de se ligar a esses inibidores.

#### **4.13 Análise das interações entre proteína e ligante**

A conformação de menor energia dos ligantes nas simulações de *docking* foram analisadas com o objetivo de identificar os resíduos da proteína envolvidos em ligações hidrogênio (*backbone* ou cadeia lateral), eletrostática, coordenação do metal, dentre outras, e comparar os resíduos envolvidos nas interações com cada um dos ligantes. Essa análise foi realizada com o programa Maestro.

Uma tabela foi obtida a partir do mapeamento de todas as interações entre cada proteína com os ligantes Batimastat, Marimastat e Prinomastat. Nas Tabelas 15, 16 e 17, somente os resíduos que interagiram com mais de 50% das proteínas são apresentados com relação aos ligantes Batimastat, Marimastat e Prinomastat, respectivamente. As figuras referentes a interação de cada proteína com estes ligantes são mostradas nos Anexos G, H e I.

**Tabela 15:** Resíduos que interagem em mais de 50% das metaloproteases de *Leishmania (V.) braziliensis* com o ligante Batimastat.

	Gly123	Val124	Leu125	Ala126	Trp127	Ala128	Thr162	Hid165	Glu166	Hid169	Hid235	Ala247	Pro248	Ala249	Ser250	Zn
1LML																
100470																
100480																
100500																
100510																
100520																
100530																
100540																
100550																
100560																
100570																
100580																
100590																
100600																
100610																
101550																
101560																
101570																
101580																
101590																
101610																
101620																
101630																
101640																
101650																
101660																
101670																
101680																
101690																
101710																
%	56,66	90	66,66	90	73,33	50	76,66	90	100	63,33	83,33	53,33	76,66	63,33	50	100

Os espaços preenchidos em cinza claro indicam interações  $\pi$ - $\pi$  ou eletrostáticas e os espaços preenchidos em cinza escuro indica ligação hidrogênio com a cadeia lateral ou *backbone*.

**Tabela 16.** Resíduos que interagem em mais de 50% das metaloproteases de *Leishmania (V.) braziliensis* com o ligante Marimastat.

	Gly123	Val124	Leu125	Ala126	Trp127	Thr162	Hid165	Glu166	Hid235	Ala247	Pro248	Ala249	Zn
1LML													
100470													
100480													
100500													
100510													
100520													
100530													
100540													
100550													
100560													
100570													
100580													
100590													
100600													
100610													
101550													
101560													
101570													
101580													
101590													
101610													
101620													
101630													
101640													
101650													
101660													
101670													
101680													
101690													
101710													
%	63,33	86,66	66,66	96,66	63,33	80	96,66	100	83,33	50	80	63,33	100

Os espaços preenchidos em cinza claro indicam interações  $\pi$ - $\pi$  ou eletrostáticas e os espaços preenchidos em cinza escuro indica ligação hidrogênio com a cadeia lateral ou *backbone*.

**Tabela 17.** Resíduos que interagem em mais de 50% das metaloproteases de *Leishmania (V.) braziliensis* com o ligante Prinomastat.

	Val124	Leu125	Ala126	Trp127	Thr162	Hid165	Glu166	Hid235	Pro248	Ala249	Ser250	Zn
1LML												
100470												
100480												
100500												
100510												
100520												
100530												
100540												
100550												
100560												
100570												
100580												
100590												
100600												
100610												
101550												
101560												
101570												
101580												
101590												
101610												
101620												
101630												
101640												
101650												
101660												
101670												
101680												
101690												
101710												
%	76,66	53,33	83,33	63,33	70	86,66	100	90	80	66,66	53,33	100

Os espaços preenchidos em cinza claro indicam interações  $\pi$ - $\pi$  ou eletrostáticas e os espaços preenchidos em cinza escuro indica ligação hidrogênio com a cadeia lateral ou *backbone*.

Três análises foram realizadas para cada ligante levando-se em consideração a proteína completa. A primeira análise contém apenas resíduos que interagiram de 50 a 70% das proteínas. A segunda análise leva em consideração os resíduos com 70 a 90% e por fim, os resíduos que interagiram em mais de 90% das proteínas.

Para o ligante Batimastat, a primeira análise resultou a interação de Gly123, Leu125, Ala128, His169, Ala247, Ala249 e Ser250. Na segunda análise tivemos Trp127, Thr162, His235 e Pro248. Em mais de 90% a interação ocorreu com Val224, Ala126, His165 e Glu166, além de todas as proteínas interagirem com o Zinco. Com Marimastat, a primeira análise teve os resíduos Gly123, Leu125, Trp127, Ala247 e Ala249 envolvidos em interações. Na segunda análise, Val224, Thr162, His235 e Pro248 e os que mais interagem com as proteínas são os resíduos Ala126, His165 e Glu166, além do Zinco. Por fim, Prinomastat interagiu com Leu125, Trp127, Ala249 e Ser250 na primeira análise, os resíduos Val124, Ala126, Thr162, His165 e Pro248 são os resultados da segunda na análise e com mais de 90%, Glu166, His235 e Zinco.

É importante notar que os resíduos de coordenação do Zinco, His165, His169 e His235, estão presentes nas interações com todos os ligantes avaliados envolvidos tanto em ligações de hidrogênio com o ligante quanto participando da coordenação do íon. O resíduo Glu166, que compõe o motivo identificador de metaloproteases, também interage com 100% das proteínas. Val124, Leu125, Ala126, Trp127, Thr162, Pro248 e Ala249 são os resíduos comuns que interagem com todos os ligantes.

#### **4.14 Análise dos *hotspots* das metaloproteases**

Usamos o programa foldX para identificar os *hotspots* das metaloproteases. Para a proteína PDB id 1LML, o programa retornou 102 resíduos com valor de  $\Delta\Delta G$  maior que 2,5 kcal/mol na troca por Alanina. Considerando somente a região de inserção, a quantidade de resíduos diminuiu para 18, sendo eles: Met167, Leu171, Gly172, Phe173, Phe178, Ile183, Gly192, Pro198, Ile200, Tyr213, Gly214, Cys215, Leu221, Gly233, Ile236, Lys237, Leu245 e Met246. A Tabela 18 mostra os valores de energia de Gibbs para estes resíduos e para os resíduos encontrados nas demais proteínas.



**Tabela 18:** Energia de Gibbs dos *hotspots* definidos após a troca dos resíduos da região de inserção por Alanina.

	<b>1LML</b>	<b>DLA</b>	<b>DLL</b>	<b>DLT</b>	<b>NRI</b>	<b>SRM</b>	<b>SSV</b>
Met167	3,78	-	3,08	-	-	-	-
Val167	-	2,53	-	-	-	-	-
Leu171	2,67	3,10	3,43	-	2,81	-	3,71
Gly172	2,71	3,07	5,64	3,42	4,68	3,00	5,59
Phe173	3,57	3,49	4,14	3,79	3,97	4,51	4,15
Phe177	-	-	-	-	3,77	2,95	-
Phe178	3,14	-	-	3,45	3,11	3,25	3,32
Gly182	-	2,57	-	-	-	-	2,94
Ile183	2,73						
Leu183	-	-	-	2,51	-	-	-
Met183	-	-	-	-	3,20	-	-
Gly189	-	-	-	-	2,78	-	-
Leu190	-	-	3,50	-	-	-	-
Gly192	3,66	-	-	3,15	3,39	2,53	2,89
Tyr195	-	-	-	2,72	3,28	-	-
Pro198	2,84						
Val199	-	-	2,71	-	-	2,50	2,72
Ile200	4,42	-	-	-	-	-	4,29
Leu200	-	3,54	3,11	4,18	2,85	3,75	-
Tyr213	3,63	-	-	-	3,17	4,47	-
Gly214	5,43	-	-	2,76	-	3,50	-
Cys215	4,53	-	5,64	5,36	5,21	5,09	5,24
Phe220	-	-	2,62	-	-	-	-
Tyr220	-	-	-	-	-	2,63	-
Leu221	3,41	-	3,49	3,18	-	3,19	2,75
Glu222	-	-	4,04	-	-	3,78	-
Leu223	-	2,64	-	2,66	-	3,70	-
Val223	-	-	2,69	-	-	-	-
Asp225	-	-	-	2,56	-	-	-
Gly227	-	-	2,86	-	4,42	4,87	-
Gly229	-	-	3,40	-	-	-	-
Gly233	2,83	-	-	-	3,73	3,85	-
Ile236	2,80						
Leu236	-	-	-	-	-	2,90	-
Lys237	2,78	-	2,50	-	-	-	2,60
Gly238	-	-	-	3,33	-	-	-
Glu244	-	-	3,34	4,85	-	-	-
Leu245	3,04	-	2,78	3,25	-	2,63	3,07
Met246	3,23	-	2,66	-	-	-	2,73

Os valores indicam a energia de Gibbs e estão calculados em kcal/mol.

Uma proteína de cada motivo foi usada para a análise de *hotspots* visto que a identidade é alta entre proteínas de mesmo motivo. Dessa forma, para o motivo

DLT, foi escolhida a estrutura LbrM.10.0470; para o motivo DLL, LbrM.10.0540; para NRI, LbrM.10.0570 e para SSV, LbrM.10.0590. Os motivos DLA e SRM só possuem uma estrutura cada, LbrM10.1560 e LbrM.10.0530, respectivamente.

Os resíduos Gly172 e Phe173 foram os únicos que se mantiveram como *hotspot* em todas as proteínas. Val167 é um *hotspot* específico da proteína de motivo DLA. Leu190, Phe220, Val223 e Gly229 são *hotspots* específicos de proteínas de motivo DLL. Leu183, Asp225, Gly238 são *hotspots* específico de proteínas de motivo DLT. Tyr220 e Leu236 são específicos para proteínas de motivo SRM e as proteínas de motivo SSV não apresentam nenhum *hotspot* específico para as mesmas.

#### **4.15 Comparação entre os *hotspots* e os resíduos de interação com os ligantes**

Esta análise foi realizada com objetivo de verificar as proteínas que tiveram resíduos identificados como *hotspots* e que também tivessem interação com os ligantes utilizados nas simulações de *docking*. Esta análise foi feita somente com os resíduos pertencentes a região de inserção das metaloproteases. Vale ressaltar, que nenhum dos *hotspots* identificados teve interação com mais de 50% das proteínas, conforme análise anterior (seção 4.14). A Tabela 19 mostra quais proteínas interagiram com cada um dos *hotspots* e as identifica a partir de seus motivos.

**Tabela 19:** Identificação das metaloproteases que possuem interação com *hotspots* nas simulações de *docking* para os ligantes Batimastat, Marimastat e Prinomastat.

	Gly172	Phe173	Phe177	Gly227	Gly229	Gly233	Glu244	Leu245	Met246	
Batimastat		LbrM.10.0470		LbrM.10.0500		LbrM.10.0470		LbrM.10.0500		<div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">SSV</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">DLT</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">SRM</div> <div style="border: 1px solid black; padding: 2px; margin-bottom: 2px;">DLL</div>
							LbrM.10.0510	LbrM.10.0510		
								LbrM.10.0520	LbrM.10.0520	
								LbrM.10.0530		
							LbrM.10.0550	LbrM.10.0550		
				LbrM.10.0560			LbrM.10.0600	LbrM.10.0600		
		LbrM.10.0610								
		LbrM.10.1550			LbrM.10.1550					
								LbrM.10.1570		
								LbrM.10.1620	LbrM.10.1620	
							LbrM.10.1640	LbrM.10.1640		
				LbrM.10.1650						
				LbrM.10.1660						
							LbrM.10.1680	LbrM.10.1680		
	Marimastat					LbrM.10.0470	LbrM.10.0470			
								LbrM.10.0510		
							LbrM.10.0530	LbrM.10.0530	LbrM.10.0530	
				LbrM.10.0550			LbrM.10.0550	LbrM.10.0550		
								LbrM.10.0560		
		LbrM.10.0590						LbrM.10.0610		
						LbrM.10.1550				
								LbrM.10.1570		
							LbrM.10.1640	LbrM.10.1640		
							LbrM.10.1680	LbrM.10.1680	LbrM.10.1680	

(continuação)

	Gly172	Phe173	Phe177	Gly227	Gly229	Gly233	Glu244	Leu245	Met246
Prinomastat				LbrM.10.0480	LbrM.10.0470			LbrM.10.0470	
							LbrM.10.0500	LbrM.10.0500	
							LbrM.10.0510	LbrM.10.0510	
								LbrM.10.0530	
				LbrM.10.0550				LbrM.10.0550	
							LbrM.10.0560	LbrM.10.0560	
			LbrM.10.0570						
		LbrM.10.0590	LbrM.10.0590	LbrM.10.0590					
								LbrM.10.0600	LbrM.10.0600
								LbrM.10.0610	
								LbrM.10.1550	
					LbrM.10.1580				
							LbrM.10.1640	LbrM.10.1640	LbrM.10.1640
							LbrM.10.1680	LbrM.10.1680	LbrM.10.1680
						LbrM.10.1690	LbrM.10.1690		
		LbrM.10.1710							

SSV
DLT
NRI
SRM
DLL

A legenda com os motivos está apresentada ao lado de cada tabela.

Levando em consideração os resíduos da região de inserção que foram identificados como *hotspots*, relacionamos as proteínas envolvidas em qualquer interação para cada ligante. Os resíduos são Gly172, Phe173, Phe177, Gly227, Gly229, Gly233, Glu244, Leu245 e Met246. Esses resíduos interagem com menos de 50% das proteínas, havendo casos de interagir com somente uma proteína específica, como é o caso do resíduo Gly172 e Phe177 que só interagem com a proteína LbrM.10.0590 na simulação com o ligante Prinomastat e Gly229 que só interagem com a proteína LbrM.10.0470 nas simulações com os ligantes Marimastat e Prinomastat.

As proteínas que interagem com os três ligantes considerando o resíduo Leu125 são LbrM.10.0510, LbrM.10.0530, LbrM.10.0550, LbrM.10.1640 e LbrM.10.1680, e, LbrM.10.1640 e LbrM.10.1680, considerando o resíduo Glu244. Nesta análise, duas proteínas são as de motivo SSV, uma de motivo SRM, e duas de motivo DLT. Pode-se concluir que a interação do ligante na proteína não leva em consideração o motivo que a proteína apresenta já que com diferentes inibidores, a interação se mantém.

Outro fato importante é que mais proteínas de motivo DLT interagiram com o ligante Batimastat enquanto que uma quantidade maior de proteínas de motivo SSV interagiram mais com Prinomastat. O ligante Marimastat não apresentou preferência por nenhum dos motivos, já que 5 proteínas envolvidas nas interações com os resíduos de *hotspots* são pertencentes ao motivo SSV e 4 delas, ao motivo DLT.

## 5 CONCLUSÕES

As principais conclusões desta tese podem se condensar nos principais tópicos:

- A partir do alinhamento entre as sequências foi possível distinguir três subgrupos que apresentaram regiões específicas de resíduos ao longo da região de inserção: DLT/L, SSV e NRI.
- Com o alinhamento global das sequências foi possível observar uma separação em dois principais grupos. O primeiro contendo as sequências de motivos DLT/L e NRI e o segundo com as sequências de motivos SSV;
- A análise de localização celular mostrou que 75% das proteínas apresentam o sítio ativo na porção extracelular, confirmando sua característica já que a proteína usada como molde também é extracelular;
- Dos seis programas de modelagem baseada em molde, o programa Swiss-Model foi o que apresentou melhores resultados na validação das estruturas construídas;
- Sequências com baixa cobertura no alinhamento com a proteína molde não apresentaram resultados satisfatórios com o programa de modelagem comparativa, Modeller;
- A abordagem de dinâmica molecular clássica aplicada neste trabalho foi pertinente para validação da estabilidade estrutural dos modelos tridimensionais propostos;
- A análise da flutuação dos resíduos contidos na região de inserção das proteínas mostrou um comportamento similar para as proteínas de mesmo subgrupo. As regiões de maior flutuação são regiões coil;
- Os valores obtidos de energia potencial das estruturas simuladas não possuem relação com os motivos identificados;
- Três padrões distintos de densidade de cargas foram identificados com a análise do perfil eletrostático, tanto para a análise em pH 7 quanto em pH 5, tendo uma alta correspondência com os subgrupos definidos previamente;
- A análise dos valores geométricos da região de inserção das proteínas permitiu o reagrupamento resultando na separação do subgrupo SSV das demais estruturas;

- Proteínas dos subgrupos NRI e DLX são evolutivamente mais próximas entre si do que as proteínas do subgrupo SSV. Entretanto, na análise estrutural a densidade de cargas entre esses dois subgrupos se mostrou o oposto. A identificação de geometrias similares em seus sítios ativo revelam sua habilidade de acomodar ligantes de tamanhos similares de forma apropriada;
- A diversidade estrutural provavelmente confere ao parasita a habilidade de interagir com diversos tipos de peptídeos, favorecendo assim o processo de infecção;
- Tanto as características estruturais específicas quanto os parâmetros físico-químicos são relevantes para o entendimento da habilidade dessas proteínas na interação com ligantes e potenciais inibidores;
- Enquanto a geometria conservada do sítio ativo pode favorecer interações moleculares com ligantes em volumes mais estreitos, diferentes distribuições de carga atômica podem resultar em interações de um amplo espectro de ligantes com propriedades físico-químicas variáveis;
- Simulações de *docking* realizadas com o programa Autodock4<sub>Zn</sub> foram satisfatórias visto que este programa tem um campo de força específico para a identificação do Zinco das metaloproteases;
- Proteínas do subgrupo SSV podem ter maior especificidade em se ligar ao inibidor Prinomastat por apresentar menores valores de energia de ligação. Já proteínas dos subgrupos DLX e NRI teriam maior possibilidade de se ligarem em inibidores como Batimastat e Marimastat;
- A análise da interação entre a proteína e o ligante mostrou que em todas as simulações as histidinas de coordenação do Zinco estão envolvidas, ora em ligações hidrogênio, ora participando da coordenação do íon;
- A interação do ligante com resíduos da proteína identificados como *hotspots* não leva em consideração o motivo que a proteína apresenta já que com diferentes inibidores, a interação se mantém.

## 6 REFERÊNCIAS

Maestro, Schrödinger. New York, NY, LLC.

Allen, M. P. and M. R. Wilson (1989). "Computer simulation of liquid crystals." Journal of computer-aided molecular design **3**(4): 335-353.

Anandakrishnan, R., B. Aguilar and A. V. Onufriev (2012). "H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations." Nucleic acids research **40**(W1): W537-W541.

Armougom, F., S. Moretti, O. Poirot, S. Audic, P. Dumas, B. Schaeli, V. Keduas and C. Notredame (2006). "Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee." Nucleic acids research **34**(suppl 2): W604-W608.

Arnold, K., L. Bordoli, J. Kopp and T. Schwede (2006). "The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling." Bioinformatics **22**(2): 195-201.

Baker, N. A., D. Sept, S. Joseph, M. J. Holst and J. A. McCammon (2001). "Electrostatics of nanosystems: application to microtubules and the ribosome." Proceedings of the National Academy of Sciences **98**(18): 10037-10041.

Barrett, A. J. (1994). "Classification of peptidases." Methods in enzymology **244**: 1.

Barrett, A. J., N. D. Rawlings and E. A. O'Brien (2001). "The MEROPS database as a protease information system." Journal of structural biology **134**(2): 95-102.

Barrett, A. J., D. P. Tolle and N. D. Rawlings (2003). "Managing peptidases in the genomic era." Biological chemistry **384**(6): 873-882.

Benkert, P., S. C. Tosatto and D. Schomburg (2008). "QMEAN: A comprehensive scoring function for model quality assessment." Proteins: Structure, Function, and Bioinformatics **71**(1): 261-277.

Berman, E., P. Wiernik, R. Vogler, E. Vélez-García, A. Bartolucci and F. S. Whaley (1997). "Long-term follow-up of three randomized trials comparing idarubicin and daunorubicin as induction therapies for patients with untreated acute myeloid leukemia." Cancer **80**(S11): 2181-2185.

Berman, H. M., J. Westbrook, Z. Feng, G. Gilliland, T. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne (2000). "The protein data bank." Nucleic acids research **28**(1): 235-242.

Bode, W., F.-X. Gomis-Rüth and W. Stöckler (1993). "Astacins, serralyisins, snake venom and matrix metalloproteinases exhibit identical zinc-binding environments (HEXXHXXGXXH and Met-turn) and topologies and should be grouped into a common family, the 'metzincins'." FEBS letters **331**(1-2): 134-140.



Bond, C. S. and A. W. Schuettelkopf (2009). "ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments." Acta Crystallographica Section D: Biological Crystallography **65**(5): 510-512.

Bonneau, R., J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C. E. Strauss and D. Baker (2001). "Rosetta in CASP4: progress in ab initio protein structure prediction." Proteins: Structure, Function, and Bioinformatics **45**(S5): 119-126.

Bourne, P. E., K. J. Address, W. F. Bluhm, L. Chen, N. Deshpande, Z. Feng, W. Fleri, R. Green, J. C. Merino-Ott and W. Townsend-Merino (2004). "The distribution and query systems of the RCSB Protein Data Bank." Nucleic acids research **32**(suppl 1): D223-D225.

Brasil (2009). Guia de vigilância epidemiológica. M. d. S. S. d. V. e. S. D. d. V. Epidemiológica. Brasília. **7**.

Britto, C., C. Ravel, P. Bastien, C. Blaineau, M. Pagès, J.-P. Dedet and P. Wincker (1998). "Conserved linkage groups associated with large-scale chromosomal rearrangements between Old World and New World Leishmania genomes." Gene **222**(1): 107-117.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." Journal of computational chemistry **4**(2): 187-217.

Bryceson, A. (2001). "A policy for leishmaniasis with respect to the prevention and control of drug resistance." Tropical Medicine & International Health **6**(11): 928-934.

Case, D., V. Babin, J. Berryman, R. Betz, Q. Cai, D. Cerutti, T. Cheatham Iii, T. Darden, R. Duke and H. Gohlke (2014). "Amber 14."

Cherkasov, A. and S. J. Jones (2004). "Structural characterization of genomes by large scale sequence-structure threading." BMC bioinformatics **5**(1): 1.

Chivian, D., D. E. Kim, L. Malmström, J. Schonbrun, C. A. Rohl and D. Baker (2005). "Prediction of CASP6 structures using automated Robetta protocols." Proteins: Structure, Function, and Bioinformatics **61**(S7): 157-166.

Chopra, G., C. M. Summa and M. Levitt (2008). "Solvent dramatically affects protein structure refinement." Proceedings of the National Academy of Sciences **105**(51): 20239-20244.

Cohen-Freue, G., T. R. Holzer, J. D. Forney and W. R. McMaster (2007). "Global gene expression in Leishmania." International journal for parasitology **37**(10): 1077-1086.

Conceição-Silva, F. and C. R. Alves (2014). Leishmanioses do continente americano. Leishmanioses do continente americano, Editora Fiocruz.

Croft, S. L. (2001). "Monitoring drug resistance in leishmaniasis." Tropical Medicine & International Health **6**(11): 899-905.

da Silva-López, R. E. (2010). "Proteases de Leishmania: novos alvos para o desenvolvimento racional de fármacos." Quim. Nova **33**(7): 1541-1548.

Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark (1999). "Peptide folding: when simulation meets experiment." Angewandte Chemie International Edition **38**(1-2): 236-240.

de Castro, R. M. (2012). "Determinação computacional dos efeitos da mutagênese em interfaces proteína-DNA."

de Magalhães, C. S., C. H. d. S. Barbosa, D. M. Almeida and L. E. Dardenne (2012). Improving differential evolution accuracy for flexible ligand docking using a multi-solution strategy. Intelligent Data Engineering and Automated Learning-IDEAL 2012, Springer: 688-698.

de Pita-Pereira, D., C. R. Alves, M. B. Souza, R. P. Brazil, Á. L. Bertho, A. de Figueiredo Barbosa and C. C. Britto (2005). "Identification of naturally infected *Lutzomyia intermedia* and *Lutzomyia migonei* with *Leishmania (Viannia) braziliensis* in Rio de Janeiro (Brazil) revealed by a PCR multiplex non-isotopic hybridisation assay." Transactions of the Royal Society of Tropical Medicine and Hygiene **99**(12): 905-913.

DeLano, W. L. (2002). "The PyMOL molecular graphics system."

Depledge, D. P., K. J. Evans, A. C. Ivens, N. Aziz, A. Maroof, P. M. Kaye and D. F. Smith (2009). "Comparative expression profiling of *Leishmania*: modulation in gene expression between species and in different host genetic backgrounds." PLoS Negl Trop Dis **3**(7): e476.

Deps, P. D., M. C. Viana, A. Falqueto and R. Dietze (2000). "Evaluation of the efficacy and toxicity of N-methyl-glucamine vs BP88® Sodium Stibogluconate in the treatment of localized cutaneous leishmaniasis." Revista da Sociedade Brasileira de Medicina Tropical **33**(6): 535-543.

Desjeux, P. (1996). "Leishmaniasis: public health aspects and control." Clinics in dermatology **14**(5): 417-423.

Di Tommaso, P., S. Moretti, I. Xenarios, M. Orobity, A. Montanyola, J.-M. Chang, J.-F. Taly and C. Notredame (2011). "T-Coffee: a web server for the multiple sequence alignment of protein and RNA sequences using structural information and homology extension." Nucleic acids research **39**(suppl 2): W13-W17.

Dolinsky, T. J., J. E. Nielsen, J. A. McCammon and N. A. Baker (2004). "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations." Nucleic acids research **32**(suppl 2): W665-W667.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen (1995). "A smooth particle mesh Ewald method." The Journal of chemical physics **103**(19): 8577-8593.

Fiser, A. and A. Sali (2003). "Modeller: generation and refinement of homology-based protein structure models." Methods in enzymology **374**: 461-491.

Frenkel, D. and B. Smit (1996). Understanding molecular simulations: from algorithms to applications, Academic Press.

Friesner, R. A., J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley and J. K. Perry (2004). "Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy." Journal of medicinal chemistry **47**(7): 1739-1749.

Gasteiger, J. and M. Marsili (1978). "A new model for calculating atomic charges in molecules." Tetrahedron Letters **19**(34): 3181-3184.

Gibas, C. and P. Jambeck (2001). Developing bioinformatics computer skills, " O'Reilly Media, Inc."

Goldsmith-Fischman, S. and B. Honig (2003). "Structural genomics: computational methods for structure analysis." Protein Science **12**(9): 1813-1821.

Gomis-Rüth, F. X. (2003). "Structural aspects of the metzincin clan of metalloendopeptidases." Molecular biotechnology **24**(2): 157-202.

Gomis-Rüth, F. X. (2009). "Catalytic domain architecture of metzincin metalloproteases." Journal of biological chemistry **284**(23): 15353-15357.

Goodsell, D. S., G. M. Morris and A. J. Olson (1996). "Automated docking of flexible ligands: applications of AutoDock." Journal of Molecular Recognition **9**(1): 1-5.

Grimaldi Jr, G., R. B. Tesh and D. McMahon-Pratt (1989). "A review of the geographic distribution and epidemiology of leishmaniasis in the New World." The American journal of tropical medicine and hygiene **41**(6): 687-725.

Gull, K. (2001). "The biology of kinetoplastid parasites: insights and challenges from genomics and post-genomics." International journal for parasitology **31**(5): 443-452.

Harhay, M. O., P. L. Olliaro, D. L. Costa and C. H. N. Costa (2011). "Urban parasitology: visceral leishmaniasis in Brazil." Trends in parasitology **27**(9): 403-409.

Holtje, H., G. Folkers and A. Luzar (1998). "Molecular Modeling, Basic Principles and Applications." Computers in Physics **12**(1): 41-41.

Hooper, N. M. (1994). "Families of zinc metalloproteases." FEBS letters **354**(1): 1-6.

Hsiao, C.-H. C., C. Yao, P. Storlie, J. E. Donelson and M. E. Wilson (2008). "The major surface protease (MSP or GP63) in the intracellular amastigote stage of *Leishmania chagasi*." Molecular and biochemical parasitology **157**(2): 148-159.

Huey, R., G. M. Morris, A. J. Olson and D. S. Goodsell (2007). "A semiempirical free energy force field with charge-based desolvation." Journal of computational chemistry **28**(6): 1145-1152.

Jaffe, C. L. and D. M. Dwyer (2003). "Extracellular release of the surface metalloprotease, gp63, from *Leishmania* and insect trypanosomatids." Parasitology research **91**(3): 229-237.

Johnson, M. S., N. Srinivasan, R. Sowdhamini and T. L. Blundell (1994). "Knowledge-based protein modeling." Critical Reviews in Biochemistry and Molecular Biology **29**(1): 1-68.

Jones, G., P. Willett, R. C. Glen, A. R. Leach and R. Taylor (1997). "Development and validation of a genetic algorithm for flexible docking." Journal of molecular biology **267**(3): 727-748.

Jorgensen, W. L. (2009). "Efficient drug lead discovery and optimization." Accounts of chemical research **42**(6): 724-733.

Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein (1983). "Comparison of simple potential functions for simulating liquid water." The Journal of chemical physics **79**(2): 926-935.

Jorgensen, W. L. and J. Tirado-Rives (1988). "The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin." Journal of the American Chemical Society **110**(6): 1657-1666.

Kalisman, N., A. Levi, T. Maximova, D. Reshef, S. Zafriri-Lynn, Y. Gleyzer and C. Keasar (2005). "MESHI: a new library of Java classes for molecular modeling." Bioinformatics **21**(20): 3931-3932.

Kitchen, D. B., H. Decornez, J. R. Furr and J. Bajorath (2004). "Docking and scoring in virtual screening for drug discovery: methods and applications." Nature reviews Drug discovery **3**(11): 935-949.

Köhler, P. (1986). "Progress in molecular parasitology." Experientia **42**(4): 377-386.

Kukol, A. (2008). Molecular modeling of proteins, Springer.

Kulkarni, M. M., W. R. McMaster, E. Kamysz, W. Kamysz, D. M. Engman and B. S. McGwire (2006). "The major surface-metalloprotease of the parasitic protozoan, *Leishmania*, protects against antimicrobial peptide-induced apoptotic killing." Molecular microbiology **62**(5): 1484-1497.

Lainson, R., J. J. Shaw, W. Peters and R. Killick-Kendrick (1987). Evolution, classification and geographical distribution, Academic Press.

Laskowski, R. A., M. W. MacArthur, D. S. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." Journal of applied crystallography **26**(2): 283-291.

Leach, A. R. (2001). Molecular modelling: principles and applications, Pearson education.

Letunic, I. and P. Bork (2016). "Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees." Nucleic acids research: gkw290.

Levitt, M., M. Hirshberg, R. Sharon and V. Daggett (1995). "Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution." Computer physics communications **91**(1): 215-231.

Lieke, T., S. Nylén, L. Eidsmo, W. McMaster, A. Mohammadi, A. Khamesipour, L. Berg and H. Akuffo (2008). "Leishmania surface protein gp63 binds directly to human natural killer cells and inhibits proliferation." Clinical & Experimental Immunology **153**(2): 221-230.

Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw (2010). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." Proteins: Structure, Function, and Bioinformatics **78**(8): 1950-1958.

Lipkowitz, K. B. and D. B. Boyd (1997). Reviews in Computational Chemistry:\* Volume 13\*, Wiley Online Library.

Logan-Klumpler, F. J., N. De Silva, U. Boehme, M. B. Rogers, G. Velarde, J. A. McQuillan, T. Carver, M. Aslett, C. Olsen and S. Subramanian (2011). "GeneDB— an annotation database for pathogens." Nucleic acids research **40**(D1): D98-D108.

MacKerell, A. D., M. Feig and C. L. Brooks (2004). "Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations." Journal of computational chemistry **25**(11): 1400-1415.

Maier, J. A., C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling (2015). "ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB." Journal of chemical theory and computation **11**(8): 3696-3713.

Marchler-Bauer, A., A. R. Panchenko, B. A. Shoemaker, P. A. Thiessen, L. Y. Geer and S. H. Bryant (2002). "CDD: a database of conserved domain alignments with links to domain three-dimensional structure." Nucleic acids research **30**(1): 281-283.

Marrink, S. J., H. J. Risselada, S. Yefimov, D. P. Tieleman and A. H. De Vries (2007). "The MARTINI force field: coarse grained model for biomolecular simulations." The Journal of Physical Chemistry B **111**(27): 7812-7824.

Marzochi, M., S. Coutinho, P. Sabroza and W. de Souza (1979). "[Indirect immunofluorescence reaction and intradermoreaction for American cutaneous leishmaniasis in residents of the Jacarepagua region (Rio de Janeiro). Comparative study of results observed in 1974 and 1978]." Revista do Instituto de Medicina Tropical de São Paulo **22**(3): 149-155.

Melo, F. and E. Feytmans (1998). "Assessing protein structures with a non-local atomic interaction energy." Journal of molecular biology **277**(5): 1141-1152.

Meng, E. C., B. K. Shoichet and I. D. Kuntz (1992). "Automated docking with grid-based energy evaluation." Journal of computational chemistry **13**(4): 505-524.

Mizuguchi, K., C. M. Deane, T. L. Blundell and J. P. Overington (1998). "HOMSTRAD: a database of protein structure alignments for homologous families." Protein science **7**(11): 2469-2471.

Morgon, N. H. and K. Coutinho (2007). Métodos de química teórica e modelagem molecular, Editora Livraria da Física.

Morris, G. M., D. S. Goodsell, R. S. Halliday, R. Huey, W. E. Hart, R. K. Belew and A. J. Olson (1998). "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." Journal of computational chemistry **19**(14): 1639-1662.

Moult, J., K. Fidelis, A. Kryshtafovych, B. Rost and A. Tramontano (2009). "Critical assessment of methods of protein structure prediction—Round VIII." Proteins: Structure, Function, and Bioinformatics **77**(S9): 1-4.

Murzin, A. G., S. E. Brenner, T. Hubbard and C. Chothia (1995). "SCOP: a structural classification of proteins database for the investigation of sequences and structures." Journal of molecular biology **247**(4): 536-540.

Muskus, C. E. and V. M. Marín (2002). "[Metacyclogenesis: a basic process in the biology of Leishmania]." Biomedica: Revista del Instituto Nacional de Salud **22**(2): 167-177.

Ofran, Y. and B. Rost (2007). "Protein–protein interaction hotspots carved into sequences." PLoS Comput Biol **3**(7): e119.

Olivier, M., D. J. Gregory and G. Forget (2005). "Subversion mechanisms by which Leishmania parasites can escape the host immune response: a signaling point of view." Clinical microbiology reviews **18**(2): 293-305.

Oostenbrink, C., A. Villa, A. E. Mark and W. F. Van Gunsteren (2004). "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6." Journal of computational chemistry **25**(13): 1656-1676.

Pacheco, A. (2009). MODELAGEM MOLECULAR COMPARATIVA E ESTUDOS DE ACOPLAMENTO MOLECULAR DA ENZIMA LANOSTEROL 14 $\alpha$ -DESMETILASE DO *Moniliophthora perniciosa*. Master, Universidade Estadual de Feira de Santana.

Pascutti, P. (2002). "Introdução à Modelagem e Dinâmica Molecular." Introdução à Modelagem e Dinâmica Molecular **1**: 1-38.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. DeBolt, D. Ferguson, G. Seibel and P. Kollman (1995). "AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules." Computer Physics Communications **91**(1): 1-41.

Pettersen, E. F., T. D. Goddard, C. C. Huang, G. S. Couch, D. M. Greenblatt, E. C. Meng and T. E. Ferrin (2004). "UCSF Chimera—a visualization system for exploratory research and analysis." Journal of computational chemistry **25**(13): 1605-1612.

Pronk, S., S. Páll, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson and D. van der Spoel (2013). "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit." Bioinformatics **29**(7): 845-854.

Puente, X. S., L. M. Sánchez, C. M. Overall and C. López-Otín (2003). "Human and mouse proteases: a comparative genomic approach." Nature Reviews Genetics **4**(7): 544-558.

Ramachandran, G. t. and V. Sasisekharan (1968). "Conformation of polypeptides and proteins." Advances in protein chemistry **23**: 283-437.

Rangel, E. F. and R. Lainson (2003). Flebotomíneos no Brasil, Fiocruz.

Rarey, M., B. Kramer, T. Lengauer and G. Klebe (1996). "A fast flexible docking method using an incremental construction algorithm." Journal of molecular biology **261**(3): 470-489.

Rawlings, N. D. and A. J. Barrett (2000). "MEROPS: the peptidase database." Nucleic acids research **28**(1): 323-325.

Rawlings, N. D., A. J. Barrett and A. Bateman (2009). "MEROPS: the peptidase database." Nucleic acids research: gkp971.

Rawlings, N. D., M. Waller, A. J. Barrett and A. Bateman (2014). "MEROPS: the database of proteolytic enzymes, their substrates and inhibitors." Nucleic acids research **42**(Database-Issue): 503-509.

Reis, R. A. G., E. Lorenzato, V. C. Silva and M. C. Nonato (2015). "Recombinant production, crystallization and crystal structure determination of dihydroorotate dehydrogenase from *Leishmania (Viannia) braziliensis*." Acta Crystallographica Section F: Structural Biology Communications **71**(5): 547-552.

Richards, F. M. (1977). "Areas, volumes, packing, and protein structure." Annual review of biophysics and bioengineering **6**(1): 151-176.

Rodrigues, J. P., M. Levitt and G. Chopra (2012). "KoBaMIN: a knowledge-based minimization web server for protein structure refinement." Nucleic acids research: gks376.

Ross, R. (1903a). "Further notes on Leishman's bodies." British medical journal **2**(2239): 1401.

Ross, R. (1903b). "Note on the bodies recently described by Leishman and Donovan." British medical journal **2**(2237): 1261.

Rössle, S. (2004). Desenvolvimento de um sistema computacional para modelagem comparativa em genômica estrutural: Análise de seqüências do genoma da *Gluconacetobacter diazotrophicus*, Tese de Doutorado, IBCCF/UFRJ, Brasil.

Roy, A., A. Kucukural and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nature protocols **5**(4): 725-738.

Saldanha, A. C. R., G. A. S. Romero, E. Merchan-Hamann, A. V. Magalhães and V. d. O. Macedo (1999). "Comparative study between sodium stibogluconate BP 88R and meglumine antimoniate for cutaneous leishmaniasis treatment: I. Efficacy and safety." Revista da Sociedade Brasileira de Medicina Tropical **32**(4): 383-387.

- Šali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." Journal of molecular biology **234**(3): 779-815.
- Sánchez, R., U. Pieper, F. Melo, N. Eswar, M. A. Martí-Renom, M. Madhusudhan, N. Mirković and A. Šali (2000). "Protein structure modeling for structural genomics." Nature Structural & Molecular Biology **7**: 986-990.
- Sánchez, R. and A. Šali (1999). "Comparative protein structure modeling in genomics." Journal of Computational Physics **151**(1): 388-401.
- Santos-Martins, D., S. Forli, M. J. Ramos and A. J. Olson (2014). "AutoDock4Zn: an improved AutoDock force field for small-molecule docking to zinc metalloproteins." Journal of chemical information and modeling **54**(8): 2371-2379.
- Schlagenhauf, E., R. Etges and P. Metcalf (1998). "The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63)." Structure **6**(8): 1035-1046.
- Schultz, J., F. Milpetz, P. Bork and C. P. Ponting (1998). "SMART, a simple modular architecture research tool: identification of signaling domains." Proceedings of the National Academy of Sciences **95**(11): 5857-5864.
- Schymkowitz, J., J. Borg, F. Stricher, R. Nys, F. Rousseau and L. Serrano (2005). "The FoldX web server: an online force field." Nucleic acids research **33**(suppl 2): W382-W388.
- Scott, W. R., P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger and W. F. van Gunsteren (1999). "The GROMOS biomolecular simulation program package." The Journal of Physical Chemistry A **103**(19): 3596-3607.
- Sereno, D., E. Guilvard, S. Maquaire, M. Cavaleyra, P. Holzmüller, A. Ouaiissi and J.-L. Lemesre (2001). "Experimental studies on the evolution of antimony-resistant phenotype during the in vitro life cycle of Leishmania infantum: implications for the spread of chemoresistance in endemic areas." Acta tropica **80**(3): 195-205.
- Shi, J., T. L. Blundell and K. Mizuguchi (2001). "FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties." Journal of molecular biology **310**(1): 243-257.
- Silva-Almeida, M., B. Pereira, M. L. Ribeiro-Guimarães and C. R. Alves (2012). "Proteinases as virulence factors in Leishmania spp. infection in mammals." Parasit Vectors **5**: 160.
- Silva-Almeida, M., F. Souza-Silva, B. Pereira, M. L. Ribeiro-Guimarães and C. R. Alves (2014). "Overview of the organization of protease genes in the genome of Leishmania spp." Parasit Vectors **7**(1): 387.
- Söding, J., A. Biegert and A. N. Lupas (2005). "The HHpred interactive server for protein homology detection and structure prediction." Nucleic acids research **33**(suppl 2): W244-W248.



Sonnhammer, E. L., S. R. Eddy, E. Birney, A. Bateman and R. Durbin (1998). "Pfam: multiple sequence alignments and HMM-profiles of protein domains." Nucleic acids research **26**(1): 320-322.

Spitzer, R. and A. N. Jain (2012). "Surflex-Dock: Docking benchmarks and real-world application." Journal of computer-aided molecular design **26**(6): 687-699.

Sternberg, M. J. (1996). Protein Structure Prediction: A Practical Approach: A Practical Approach, Oxford University Press, USA.

Stöcker, W., F. Grams, P. Reinemer, W. Bode, U. Baumann, F. X. Gomis-Rüth and D. B. McKay (1995). "The metzincins—Topological and sequential relations between the astacins, adamalysins, serralysins, and matrixins (collagenases) define a super family of zinc-peptidases." Protein Science **4**(5): 823-840.

Summa, C. M. and M. Levitt (2007). "Near-native structure refinement using in vacuo energy minimization." Proceedings of the National Academy of Sciences **104**(9): 3177-3182.

Tallant, C., R. García-Castellanos, U. Baumann and F. X. Gomis-Rüth (2010). "On the relevance of the Met-turn methionine in metzincins." Journal of Biological Chemistry **285**(18): 13951-13957.

Tatusov, R. L., N. D. Fedorova, J. D. Jackson, A. R. Jacobs, B. Kiryutin, E. V. Koonin, D. M. Krylov, R. Mazumder, S. L. Mekhedov and A. N. Nikolskaya (2003). "The COG database: an updated version includes eukaryotes." BMC bioinformatics **4**(1): 1.

TopPred, I. (1994). "an improved software for membrane protein structure predictions Claros, Manuel G.; von Heijne, Gunnar CABIOS." Computer Applications in the Biosciences **10**(6): 685-686.

Trott, O. and A. J. Olson (2010). "AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading." Journal of computational chemistry **31**(2): 455-461.

Turk, B. (2006). "Targeting proteases: successes, failures and future prospects." Nature reviews Drug discovery **5**(9): 785-799.

Vannier-Santos, M., A. Martiny and W. d. Souza (2002). "Cell biology of Leishmania spp.: invading and evading." Current pharmaceutical design **8**(4): 297-318.

Verdonk, M. L., J. C. Cole, M. J. Hartshorn, C. W. Murray and R. D. Taylor (2003). "Improved protein–ligand docking using GOLD." Proteins: Structure, Function, and Bioinformatics **52**(4): 609-623.

Verli, H. (2014). "Bioinformática: Da Biologia à Flexibilidade Molecular." São Paulo: SBBq.

Vianna, G. (1912). Resumos do 7o Congresso Brasileiro de Medicina Tropical de São Paulo, Brasil.

- Vieira, P. S., P. O. de Giuseppe, M. T. Murakami and A. H. C. de Oliveira (2015). "Crystal structure and biophysical characterization of the nucleoside diphosphate kinase from *Leishmania braziliensis*." BMC structural biology **15**(1): 1.
- Von Heijne, G. (1992). "Membrane protein structure prediction: hydrophobicity analysis and the positive-inside rule." Journal of molecular biology **225**(2): 487-494.
- Voss, N. R. and M. Gerstein (2010). "3V: cavity, channel and cleft volume calculator and extractor." Nucleic acids research: gkq395.
- WHO (2014). Leishmaniasis. Fact sheet. **375**.
- Woessner, J. F. (2002). "MMPs and TIMPs—an historical perspective." Molecular biotechnology **22**(1): 33-49.
- Wu, S. and Y. Zhang (2007). "LOMETS: a local meta-threading-server for protein structure prediction." Nucleic acids research **35**(10): 3375-3382.
- Wu, S. and Y. Zhang (2008). "MUSTER: improving protein sequence profile–profile alignments by using multiple sources of structure information." Proteins: Structure, Function, and Bioinformatics **72**(2): 547-556.
- Xu, D. and Y. Zhang (2012). "Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field." Proteins: Structure, Function, and Bioinformatics **80**(7): 1715-1735.
- Yang, Y., E. Faraggi, H. Zhao and Y. Zhou (2011). "Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates." Bioinformatics **27**(15): 2076-2082.
- Yu, C. S., Y. C. Chen, C. H. Lu and J. K. Hwang (2006). "Prediction of protein subcellular localization." Proteins: Structure, Function, and Bioinformatics **64**(3): 643-651.
- Yu, C. S., C. J. Lin and J. K. Hwang (2004). "Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions." Protein Science **13**(5): 1402-1406.
- Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC bioinformatics **9**(1): 40.

## ANEXO A

### Minimização com restrição de 5 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
  imin=1, maxcyc=2000, ncyc=1000, ntmin=1, cut=10.0, ntb=1,
  ntc=1, ntf=1, ntp=100, ioutfm = 1, ntr=1,
  restraintmask=':1-466 & !@H=', restraint_wt=5.0
/
```

### Minimização com proteína livre, sem restrição

```
&cntrl
  imin=1, maxcyc=5000, ncyc=2500, ntmin=1, cut=10.0, ntb=1,
  ntc=1, ntf=1, ntp=100, ntr=0,
/
```

### Aquecimento

```
&cntrl
  imin = 0, irest = 0, ntx = 1, nstlim = 500000, dt=0.002,
  ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0,
  cut = 10.0, nsnb = 25, ntp = 1000, ntwx = 1000,
  ntb = 2, ntp = 1, pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 0.067,
  tempi = 20.0, temp0 = 310.0, ig = -1, iwrap = 1, ioutfm = 1, ntr = 1,
  restraintmask=':1-466 & !@H', restraint_wt = 10.0, nmropt = 1
/
&wt type='TEMP0', istep1=0, istep2=50000, value1=20.0, value2=100.0 /
&wt type='TEMP0', istep1=50001, istep2=100000, value1=100.0, value2=200.0 /
&wt type='TEMP0', istep1=100001, istep2=250000, value1=200.0, value2=310.0 /
&wt type='TEMP0', istep1=250001, istep2=500000, value1=310.0, value2=310.0 /
&wt type='END' /
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfric = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 10 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
  imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
  ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
  ntp = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
  pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
  iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 10.0,
  nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfric = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 5 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 5.0,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrnc = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 2,5 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 2.5,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrnc = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 1 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 1.0,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrnc = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 0,5 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 0.5,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 0,25 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 0.25,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

### Equilibração com restrição de 0,1 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irest = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 0.1,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

## Equilibração com restrição de 0,05 Kcal/mol/Å<sup>2</sup> dos átomos da proteína

```
&cntrl
imin = 0, irect = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1, ntr = 1, restraintmask=':1-466 & !@H', restraint_wt = 0.05,
nmropt = 0
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

## Equilibração com proteína livre, sem restrição

```
&cntrl
imin = 0, irect = 1, ntx = 5, nstlim = 50000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 1000, ntwx = 1000, ntwr = 5000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1,
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

## Produção de 50ns

```
&cntrl
imin = 0, irect = 1, ntx = 5, nstlim = 25000000, dt=0.002,
ntc = 2, ntf = 2, tol = 1.0e-8, jfastw = 0, cut = 10.0, nsnb = 25,
ntpr = 25000, ntwx = 25000, ntwr = 25000, ntb = 2, ntp = 1,
pres0 = 1.01325, taup = 1.0, ntt = 3, gamma_ln = 5.0, temp0 = 310.0,
iwrap = 1, ioutfm = 1,
/
&ewald
skinnb = 2.0, order = 4, vdwmeth = 1, netfrc = 1, dsum_tol = 1.0e-6
/
```

## Anexo B

- Preparação da proteína:

```
pythonsh prepare_receptor4.py -r ../protein.pdb -o protein.pdbqt
```

- Preparação do ligante:

```
pythonsh prepare_ligand4.py -l ligante.mol2 -o ligante.pdbqt -A hydrogens
```

- Preparação do pseudoátomo que vai representar o íon Zn:

```
python zinc_pseudo.py -r protein.pdbqt -o protein_zn.pdbqt
```

- Preparação da caixa de simulação:

```
pythonsh prepare_gpf4zn.py -l ligante.pdbqt -r protein_zn.pdbqt -o protein_zn.gpf -p  
npts=60,60,60 -p gridcenter=x,y,z
```

- Comando para gerar o arquivo de configuração da caixa de simulação:

```
./autogrid4.2.5.x.20131125 -p protein_zn.gpf
```

- Comando para gerar o arquivo de configuração do *docking*:

```
pythonsh prepare_dp42.py -l ligante.pdbqt -r protein_zn.pdbqt -o ligante_protein_zn.dpf
```

- Comando para executar a simulação de docking:

```
./autodock4 -p ligante_protein_zn.dpf
```

## Anexo C

Alinhamento das sequências completas de todas as metaloproteases do cromossomo 10 presentes na família M8 de *Leishmania (Viannia) braziliensis*. A caixa vermelha indica a região de inserção.

LbrM.10.0470	MSRDRS . . . VT . ARRLM . RLAAA . . . . .	17
LbrM.10.0480	MSRDRS . . . ST . HRRR . SVAARLM . RLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.0490	MPLDSS . . . ST . HRRR . SVAARLV . RLAAA . . . GVAAALAVG . . . TAAAWAHAAATPHRCI	49
LbrM.10.0500		
LbrM.10.0510	MPLDSS . . . ST . PRRR . SVAARLM . RLAAA . . . GLVMAVGAA . . . . . AVWQAAGHHC I	45
LbrM.10.0520	MSRDRS . . . ST . HRRR . SDAARLM . RLAAA . . . AGLVMAVGAA . . . . . AVWQAAGHHC I	45
LbrM.10.0530		
LbrM.10.0540	MRLAAA . . . GL . VMAVGAA . VWAQAAGHHC IHDKLQTR . . . . .	34
LbrM.10.0550	MPLDSS . . . ST . HRRR . SDAARLM . RLAAA . . . GLVMAVGAA . . . . . AWAQAAGHHC I	45
LbrM.10.0560	MPLDSS . . . ST . HRRR . SDAARLM . RLAAA . . . GLVMAVGAA . . . . . AVWQAAGHHC I	45
LbrM.10.0570		
LbrM.10.0580	MPRDSS . . . ST . PRRR . SVAARLM . RLAAA . . . GLVMAVGAA . . . . . AVWQAAGHHC I	45
LbrM.10.0590	MPLDSS . . . ST . HRRR . SVAARLV . RLAAA . . . GVAAALAVG . . . TAAAWAHAAATPHRCI	49
LbrM.10.0600	MPLDSS . . . ST . HRRR . SVAARLV . RLAAA . . . GVAAALAVG . . . TAAAWAHAAATPHRCI	49
LbrM.10.0610	MPLDSS . . . ST . HRRR . SDAARLM . RLAAA . . . GLVMAVGAA . . . . . AWAQAAGHHC I	45
LbrM.10.1540		
LbrM.10.1550	MDVSR . . . GG . RRRAR . APA PWPVAVRFL . . . . . ATSQRVTLS . . . . . AQHASAGRPRLR	47
LbrM.10.1560	MSRDRS . . . ST . HRRR . SDAARLM . RLAAA . . . AGLVMAVGAA . . . . . AVWQAAGHHC I	45
LbrM.10.1570	MSRDRS . . . ST . HRRR . SVAARLM . RLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.1580	RVAQQR . . . SP . AVSV . SALGLPYVSAGTIS . . . SAHTVDWAL . . . . . ADSTSPS . . . . .	42
LbrM.10.1590	MRQQA . . . LA . YGTR . RPPTRAL . FGAEK . . . . . GSQRAWGRGVRGTS PPLPSRALRPSLFL	52
LbrM.10.1610	MSRDRS . . . ST . HRRR . SVAARLM . RLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.1620	MSRDRS . . . ST . HRRR . SVAARLM . RLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.1630	MSRDRS . . . ST . HRRR . SVAARLM . RLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.1640	MRQQA . . . LA . YGTR . RPPTRAL . FGAEK . . . . . GSQRAWGRGVRGTS PPLPSRALRPSLFL	52
LbrM.10.1650	MSRDRS . . . ST . HRRR . SVAARLM . WLAAA . . . GLVMAVGAAAVVWQAAGHHC IHDRLQAR	52
LbrM.10.1660	MSRDRS . . . VT . ARRLM . RLAAA . . . . . LVMVGAAAVWQAAGHHC IHDKLQAR	45
LbrM.10.1670	MHTHTA . . . AH . KPSS . SLPLP . . . . .	17
LbrM.10.1680		
LbrM.10.1690		3
LbrM.10.1700	MPLDSS . . . ST . HRRR . SVAARLV . RLAAA . . . GVAAALAVG . . . TAAAWAHAAATPHRCI	49
LbrM.10.1710		
LbrM.10.1720	MDVSR . . . GGRRAR . APA PWPVAVRFVVA . . . . . TSQRVTLS . . . . . AQHAS . . . . .	40
LbrM.10.0470		17
LbrM.10.0480	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.0490		56
LbrM.10.0500		
LbrM.10.0510	HDKLQ . TR . . . . .	52
LbrM.10.0520	HDNLQ . ARVLQ . . . . . SVA . . . . .	58
LbrM.10.0530		
LbrM.10.0540	. . . VLQSV . . . . . A . . . . .	40
LbrM.10.0550	DKL . Q . TRVLQSV . . . . . PPGSVSA . LGLPYVSAGTIS SAHTVDWALADSTSP	95
LbrM.10.0560	HDKLQ . ARVLQSV . AQQR . . . . . SPAVSVSALGLPYVSAGTIS SAHTVDWALADSTSP	96
LbrM.10.0570		
LbrM.10.0580	HDKLQ . TRVLQSV . AQQR . . . . . SPAVSVSALGLPYVSAGTIS SAHTVDWALADSTSP	96
LbrM.10.0590		56
LbrM.10.0600		56
LbrM.10.0610	DKL . Q . ARVLQSV . AQQR . . . . . SPAVSVSALGLPYVSAGTIS SAHTVDWALADSTSP	95
LbrM.10.1540		
LbrM.10.1550	HSPSSDTRPIR . . . . . GWKQPPEGVGRLLL . . . . .	73
LbrM.10.1560	HDKLQ . ARVLQ . . . . . SVA . . . . .	58
LbrM.10.1570	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.1580		42
LbrM.10.1590		52
LbrM.10.1610	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.1620	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.1630	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.1640	R . . . . . LP . . . . . PSP . . . . .	58
LbrM.10.1650	V . . . . . LQ . . . . . SVA . . . . .	58
LbrM.10.1660	V . . . . . LQ . . . . . SVA . . . . .	51
LbrM.10.1670		17
LbrM.10.1680		
LbrM.10.1690		3
LbrM.10.1700		56
LbrM.10.1710		
LbrM.10.1720		40





<i>LbrM.10.0470</i>	AQ . . . . . AAGHHC IHDRLQARVL . . . . . QSVAQRRPPGSVS . . . . .	61
<i>LbrM.10.0480</i>	. . . . .	104
<i>LbrM.10.0490</i>	. . . . .	81
<i>LbrM.10.0500</i>	. . . . .	
<i>LbrM.10.0510</i>	. . . . .	104
<i>LbrM.10.0520</i>	. . . . .	93
<i>LbrM.10.0530</i>	. . . . .	
<i>LbrM.10.0540</i>	. . . . .	75
<i>LbrM.10.0550</i>	. . . . .	103
<i>LbrM.10.0560</i>	. . . . .	104
<i>LbrM.10.0570</i>	. . . . .	
<i>LbrM.10.0580</i>	. . . . .	104
<i>LbrM.10.0590</i>	. . . . .	81
<i>LbrM.10.0600</i>	. . . . .	81
<i>LbrM.10.0610</i>	. . . . .	103
<i>LbrM.10.1540</i>	. . . . .	
<i>LbrM.10.1550</i>	SMS . . . . . RDRSSTPRR . . . . . RSDAA . . . . .	166
<i>LbrM.10.1560</i>	. . . . .	93
<i>LbrM.10.1570</i>	. . . . .	104
<i>LbrM.10.1580</i>	. . . . .	49
<i>LbrM.10.1590</i>	ILTPPSTLPPA . HVL . . . . . AGVTGPLARTRGTASMS . . . . . RDRSSTHR	133
<i>LbrM.10.1610</i>	. . . . .	104
<i>LbrM.10.1620</i>	. . . . .	86
<i>LbrM.10.1630</i>	. . . . .	86
<i>LbrM.10.1640</i>	ILTPP . PTLPPAHVL . . . . . AGVTGPLARTRSTAGTASMSRDRSSTHR	137
<i>LbrM.10.1650</i>	. . . . .	104
<i>LbrM.10.1660</i>	. . . . .	97
<i>LbrM.10.1670</i>	VMA . . . . . VVAAAVWAQAAGHHC IHDRLQARVLQSVAQRRPPGSVS . . . . .	88
<i>LbrM.10.1680</i>	. . . . .	
<i>LbrM.10.1690</i>	. . . . .	10
<i>LbrM.10.1700</i>	. . . . .	81
<i>LbrM.10.1710</i>	. . . . .	
<i>LbrM.10.1720</i>	. . . . .	41
<i>LbrM.10.0470</i>	. ALGLPYVSADP I S SAHA . . . . .	78
<i>LbrM.10.0480</i>	. . . . .	104
<i>LbrM.10.0490</i>	. . . . . DAADTVARAAD . . . . .	92
<i>LbrM.10.0500</i>	. . . . .	
<i>LbrM.10.0510</i>	. . . . .	104
<i>LbrM.10.0520</i>	. . . . .	93
<i>LbrM.10.0530</i>	. . . . .	
<i>LbrM.10.0540</i>	. . . . .	75
<i>LbrM.10.0550</i>	. . . . .	103
<i>LbrM.10.0560</i>	. . . . .	104
<i>LbrM.10.0570</i>	. . . . .	
<i>LbrM.10.0580</i>	. . . . .	104
<i>LbrM.10.0590</i>	. . . . . DAADTVARAAD . . . . .	92
<i>LbrM.10.0600</i>	. . . . . DAADTVARAAD . . . . .	92
<i>LbrM.10.0610</i>	. . . . .	103
<i>LbrM.10.1540</i>	. . . . .	
<i>LbrM.10.1550</i>	. . . . . RLMRLAAAGLVMAVGAAA VWAQAAGHHC IHDKLQARVLQSVAQHRPPGSVSALGL	222
<i>LbrM.10.1560</i>	. . . . .	93
<i>LbrM.10.1570</i>	. . . . .	104
<i>LbrM.10.1580</i>	. . . . .	49
<i>LbrM.10.1590</i>	RRSVAARLMRLAAAGLVMAVGAAA VWAQAAGHHC IHDKLQARVLQSVAQHRPPGSVSALGL	195
<i>LbrM.10.1610</i>	. . . . .	104
<i>LbrM.10.1620</i>	. . . . .	86
<i>LbrM.10.1630</i>	. . . . .	86
<i>LbrM.10.1640</i>	RRSVAARLMRLAAAGLVMAVGAAAVVWQAAGHHC VHDRLQARVLQSVAQHRPPGSVSALGL	199
<i>LbrM.10.1650</i>	. . . . .	104
<i>LbrM.10.1660</i>	. . . . .	97
<i>LbrM.10.1670</i>	. ALGLPYVS I DP I S SAHA . . . . .	105
<i>LbrM.10.1680</i>	. . . . .	
<i>LbrM.10.1690</i>	. . . . .	10
<i>LbrM.10.1700</i>	. . . . . DAADTVARAAD . . . . .	92
<i>LbrM.10.1710</i>	. . . . .	
<i>LbrM.10.1720</i>	. . . . .	41





<i>LbrM.10.0470</i>	RPAVGVINIPAA <del>N</del> IRS . PYDQLMV <del>R</del> TVT	HEVAH . TLGFDLTLFDE . . . . . LELIDEVNNLRGKDY	285
<i>LbrM.10.0480</i>	RPAVGVINIPAA <del>Y</del> IRS . PYDQIMV <del>R</del> TVT	HEVAH . ALGFNRIFFDS . FGMVTAAI . GIRGKDY	292
<i>LbrM.10.0490</i>			187
<i>LbrM.10.0500</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEG . . . . . TGIKSVTNLRGKPF	93
<i>LbrM.10.0510</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEG . . . . . TGIKSVTNLRGKPF	292
<i>LbrM.10.0520</i>	RPAVGVINIPAA <del>Y</del> IRS . AYDQIML <del>R</del> TVT	HEVAH . ALGFDLTVFED . . . . . VGIVGDVSNLRGRDY	283
<i>LbrM.10.0530</i>	QPAVGVINIPAA <del>N</del> IVS . RDDQGLT <del>R</del> AVT	HEVAH . ALGFSRMFFES . . . . . TGIMMNVNLRGRFF	171
<i>LbrM.10.0540</i>	RPAVGVINIPAA <del>Y</del> IQS . AYDQFML <del>R</del> SVT	HEMAH . ALGFDLLLFEE . . . . . LDLLDVS <del>N</del> LRGKDY	265
<i>LbrM.10.0550</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEN . . . . . AGIVKSVTNLRGKPF	291
<i>LbrM.10.0560</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEN . . . . . AGIVTNV <del>N</del> LRGKPF	292
<i>LbrM.10.0570</i>	RPAVGVINIPAA <del>H</del> IRS . PYDQLMI <del>R</del> TVT	HEVAH . ALGFNRIFFDS . FGMVTAAI . GIRGKDY	158
<i>LbrM.10.0580</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEN . . . . . AGIVTNV <del>N</del> LRGKPF	292
<i>LbrM.10.0590</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEG . . . . . TGIKSVTNLRGKPF	281
<i>LbrM.10.0600</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEG . . . . . TGIKSVTNLRGKPF	281
<i>LbrM.10.0610</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEN . . . . . AGIVMNV <del>N</del> LRGKPF	291
<i>LbrM.10.1540</i>		. . . . . MAE . VL . . . . . RM . . . . . GRN .	10
<i>LbrM.10.1550</i>	RPAVGVINIPAA <del>Y</del> IRS . AYDQIML <del>R</del> SVT	HEMAH . ALGFDLLLFEE . . . . . SDLLDEVNNLRGRDY	433
<i>LbrM.10.1560</i>	RPAVGVINIPAA <del>Y</del> IRS . AYDQIML <del>R</del> TVT	HEVAH . ALGFDLAVFED . . . . . VGIVGDVSNLRGRDY	283
<i>LbrM.10.1570</i>	RPAVGVINIPAA <del>H</del> IRS . PYDQLMI <del>R</del> TVT	HEVAH . ALGFDLTVFEE . . . . . MEIIDEVNNLRGRDY	292
<i>LbrM.10.1580</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEN . . . . . AGIVTNV <del>N</del> LRGKPF	237
<i>LbrM.10.1590</i>	RPAVGVINIPAA <del>T</del> IQS . AYDQIMV <del>R</del> TVT	HEVAH . ALGFDLTVFED . . . . . LELIDDV <del>K</del> DLRGRDY	415
<i>LbrM.10.1610</i>	RPAVGVINIPAA <del>N</del> IVS . PYDQLMI <del>R</del> TVT	HEVAH . ALGFNRIFFDS . FGMVTAAI . GIRGKDY	292
<i>LbrM.10.1620</i>	RPAVGVNIP <del>T</del> AYIQS . AYDQIMV <del>R</del> TVT	HEVAH . ALGFDLTVFEE . . . . . DAL . ELIHDV <del>K</del> DLRERNY	292
<i>LbrM.10.1630</i>	RPAVGVNIP <del>T</del> AYIQS . AYDQIMV <del>R</del> TVT	HEVAH . ALGFDLTVFEE . . . . . DAL . ELIHDV <del>K</del> DLRERNY	292
<i>LbrM.10.1640</i>	RPAVGVNIP <del>T</del> AYIQS . AYDQIMV <del>R</del> TVT	HEVAH . ALGFDLTVFEE . . . . . DAL . ELIHDV <del>K</del> DLRGRDY	419
<i>LbrM.10.1650</i>	RPAVGVINIPAA <del>H</del> IRS . PYDQLMI <del>R</del> TVT	HEIAH . ALGFDLTVFEE . . . . . MEIIDEVNNLCGRDY	292
<i>LbrM.10.1660</i>	RPAVGVINIPAA <del>H</del> IRS . PYDQLMI <del>R</del> TVT	HEIAH . ALGFDLTVFEE . . . . . MEIIDEVNNLCGRDY	285
<i>LbrM.10.1670</i>	RPAVGVINIPAA <del>T</del> IQS . AYDQTMV <del>R</del> TVT	HEVAH . ALGFDLTVFEE . . . . . LELIQDV <del>K</del> DLRGRDY	312
<i>LbrM.10.1680</i>	RPAVGVINIPAA <del>N</del> IVS . PYDQLMV <del>R</del> TVT	HEVAH . TLGFDLTLFDE . . . . . LELIDEVNNLRGKDY	109
<i>LbrM.10.1690</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFES . . . . . TGIVMSV <del>T</del> NRGKPF	201
<i>LbrM.10.1700</i>			187
<i>LbrM.10.1710</i>	HPAVGVMNIPAA <del>N</del> IVS . RYDQGTT <del>R</del> TVT	HEVAH . ALGFSSVFFEG . . . . . TGIKSVTNLRGKPF	93
<i>LbrM.10.1720</i>	RAAVGGAAAQASRQLGPR <del>S</del> ALRVRRHH	QORS <del>H</del> RRRIGA . . . . .	243
<i>LbrM.10.0470</i>	EAPVLSSTVVA <del>K</del> AREQYGCPTLEFLEL . . . . .		313
<i>LbrM.10.0480</i>	YAPVLTPTVVT <del>K</del> AREQYGCPTLLSFLEL . . . . .		320
<i>LbrM.10.0490</i>	. . . . . VAK <del>C</del> HRPV . . . . .		195
<i>LbrM.10.0500</i>	AAPVINSSTVVA <del>K</del> AREQYGCPTLEYLEV . . . . .		121
<i>LbrM.10.0510</i>	AAPVINSSTAVAKAREQYGCPTLEYLEV . . . . .		320
<i>LbrM.10.0520</i>	EVPVLSSTVVA <del>K</del> AREQYGCATLTYLEL . . . . .		311
<i>LbrM.10.0530</i>	SVPVLSSTAVAKAREQYGCNTLTYLEL . . . . .		199
<i>LbrM.10.0540</i>	EVPVLSSTVVA <del>K</del> AREQYGCATLTFLEV . . . . .		293
<i>LbrM.10.0550</i>	AAPVINSSTAVAKAREQYGCPTLEYLEV . . . . .		319
<i>LbrM.10.0560</i>	AAPVINSSTAVAKAREQYGCPTLEYLEV . . . . .		320
<i>LbrM.10.0570</i>	YAPVLTPTVVT <del>K</del> AREQYGCPTLLSFLEL . . . . .		186
<i>LbrM.10.0580</i>	AAPVINSSTVVA <del>K</del> AREQYGCATLEYLEV . . . . .		320
<i>LbrM.10.0590</i>	AAPVINSSTAVAKAREQYGCPTLEYLEV . . . . .		309
<i>LbrM.10.0600</i>	AAPVINSSTAVAKAREQYGCPTLEYLEV . . . . .		309
<i>LbrM.10.0610</i>	AAPVINSSTVVA <del>K</del> AREQYGCPTLEYLEV . . . . .		319
<i>LbrM.10.1540</i>			10
<i>LbrM.10.1550</i>	EVPVLSSTVVA <del>K</del> AREQYGCPTLTFLEV . . . . .		461
<i>LbrM.10.1560</i>	EVPVLSSTVVA <del>K</del> AREQYGCATLTYLEL . . . . .		311
<i>LbrM.10.1570</i>	YVPVNTPTVVA <del>K</del> AREQYGCATLKYLEL . . . . .		320
<i>LbrM.10.1580</i>	AAPVINSSTVVA <del>K</del> AREQYGCATLEYLEV . . . . .		265
<i>LbrM.10.1590</i>	YVPVNSPTVVA <del>K</del> AREQYGCATLEYLEV . . . . .		443
<i>LbrM.10.1610</i>	YAPVLTPTVVT <del>K</del> AREQYGCPTLLSFLEL . . . . .		320
<i>LbrM.10.1620</i>	EVPVLSSTVVA <del>K</del> AREQYGCPTLTFLEV . . . . .		320
<i>LbrM.10.1630</i>	EVPVLSSTVVA <del>K</del> AREQYGCPTLTFLEV . . . . .		320
<i>LbrM.10.1640</i>	EVPVLTPTVVA <del>K</del> AREQYGCPTLTFLEV . . . . .		447
<i>LbrM.10.1650</i>	YVPVNTPTVVA <del>K</del> AREQYGCATLKYLEL . . . . .		320
<i>LbrM.10.1660</i>	YVPVNTPTVVA <del>K</del> AREQYGCATLKYLEL . . . . .		313
<i>LbrM.10.1670</i>	EVPVLSSTVVA <del>K</del> AREQYGCPTLEYLEV . . . . .		340
<i>LbrM.10.1680</i>	EAPVLSSTVMA <del>K</del> AREQYGCPTLEFLEL . . . . .		137
<i>LbrM.10.1690</i>	AAPVINSSTVVA <del>K</del> AREQYGCPTLEYLEV . . . . .		229
<i>LbrM.10.1700</i>	. . . . . VAK <del>C</del> HRPV . . . . .		195
<i>LbrM.10.1710</i>	AAPVINSSTVVA <del>K</del> AREQYGCPTLEYLEV . . . . .		121
<i>LbrM.10.1720</i>	. . . . . GRQ <del>H</del> VAEC <del>R</del> TRHGLGHA <del>A</del> HRRLHRRPHG <del>P</del> RLPLHS <del>R</del> RRAACEQP <del>H</del> WR <del>H</del> RLHCR <del>G</del> HP		299

<i>LbrM.10.0470</i>	EDTGGGSAAGSHLKGRNAKDELMAPVSAAGYYTALTM	350
<i>LbrM.10.0480</i>	EDTGGPASLGSHLKGRNAKDELMSSVVKGGYYTALTM	357
<i>LbrM.10.0490</i>		195
<i>LbrM.10.0500</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTNLTM	158
<i>LbrM.10.0510</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTNLTM	357
<i>LbrM.10.0520</i>	EDTGGGSTAGSHLKMRNAKDELMAPAVTAGYYTALTM	348
<i>LbrM.10.0530</i>	EDTGGCSTAGSHLKGRNAKDELMAPASAAGYYTNLTM	236
<i>LbrM.10.0540</i>	EDTGDGSTAGSHLKRRNAKDELMAPVMGAGYYTALTM	330
<i>LbrM.10.0550</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	356
<i>LbrM.10.0560</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	357
<i>LbrM.10.0570</i>	EDTGGPASLGSHLKGRNAKDELMSSVVKGGYYTALTM	223
<i>LbrM.10.0580</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	357
<i>LbrM.10.0590</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTNLTM	346
<i>LbrM.10.0600</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTNLTM	346
<i>LbrM.10.0610</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	356
<i>LbrM.10.1540</i>	ASV	13
<i>LbrM.10.1550</i>	EDTGDGSTAGSHLKRRNAKDELMAPVMGAGYYTALTM	498
<i>LbrM.10.1560</i>	EDTGGGSTAGSHLKMRNAKDELMAPAVTAL	340
<i>LbrM.10.1570</i>	EDTGGASTAGSHLKGRNAKDELMAPAVTAGYYTALTM	357
<i>LbrM.10.1580</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	302
<i>LbrM.10.1590</i>	EDQGGSDSAGSHLKGRNAQDELMAPDMGAGYYTALTM	480
<i>LbrM.10.1610</i>	EDKGGFASLGSHLKGRNAKDELMSSVVKGGYYTALTM	357
<i>LbrM.10.1620</i>	EDTGGSGFAGSHLKGRNAKDELMAPASDAGYYTNLTM	357
<i>LbrM.10.1630</i>	EDTGGSGFAGSHLKGRNAKDELMAPASDAGYYTNLTM	357
<i>LbrM.10.1640</i>	EDTGGASTAGSHLKGRNAKDELMAPASDAGYYTNLTM	484
<i>LbrM.10.1650</i>	EDTGGASTAGSHLKGRNAKDELMAPAVTAGYYTALTM	357
<i>LbrM.10.1660</i>	EDTGGASTAGSHLKGRNAKDELMAPASAAGYYTALTM	350
<i>LbrM.10.1670</i>	EDQGGSGSAGSHLKMRNVKDELMAPASGAGYYTALTM	377
<i>LbrM.10.1680</i>	EDTGGGSAAGSHLKGRNAKDELMAPVSAAGYYTALTM	174
<i>LbrM.10.1690</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTNLTM	266
<i>LbrM.10.1700</i>		195
<i>LbrM.10.1710</i>	EDQGGSGSAGSHLKGRNAKDELMAPASAAGYYTALTM	158
<i>LbrM.10.1720</i>	HGGEARHPRQLPHPAGAAAA SGAAEGEAGAGQLEGDRHDGPNLRRLLQPPGSPYRWRQRRRL	361
<i>LbrM.10.0470</i>	AVFQDLGFYKADFTKAEVMPWANLASCDFLTNKKCMERNITQWPGMFCN	403
<i>LbrM.10.0480</i>	AVFQDLGFYKADFSMAEVMPWAYLATCDFLTNKKCMERNITQWPGMFCN	409
<i>LbrM.10.0490</i>		195
<i>LbrM.10.0500</i>	AVFEDLGFYKADFTKAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	213
<i>LbrM.10.0510</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	409
<i>LbrM.10.0520</i>	AVFQDLGFYKADFTKAEVMPWAYLASCDFLTNKKCMEDNITQWPGMFCN	399
<i>LbrM.10.0530</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	286
<i>LbrM.10.0540</i>	ATFEDFGFYKADFTKAEVMPWAEHASCDFLTNKKCMEDNITQWPEMFCN	381
<i>LbrM.10.0550</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	411
<i>LbrM.10.0560</i>	AVFEDLGFYKADFTKAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	412
<i>LbrM.10.0570</i>	AVFQDLGFYKADFSMAEVMPWAYLASCDFLTNKKCMERNITQWPGMFCN	275
<i>LbrM.10.0580</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	409
<i>LbrM.10.0590</i>	AVFEDLGFYKADFTKAEVMPWGRNASCDFLTNKKCMERNITQWPEMFCN	400
<i>LbrM.10.0600</i>	AVFEDLGFYKADFTKAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	398
<i>LbrM.10.0610</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	408
<i>LbrM.10.1540</i>	DF	38
<i>LbrM.10.1550</i>	ATFEDFGFYKADFTKAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	523
<i>LbrM.10.1560</i>		340
<i>LbrM.10.1570</i>	GILQDLGFYKADFTKAEVMPWAEYATCDFLTNKKCKEKNITQWPMFCN	409
<i>LbrM.10.1580</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMERNITQWPEMFCN	354
<i>LbrM.10.1590</i>	AIFQDLGFYKADFSMAEVMPWAYLATCDFLTNKKCMERNITQWPEMFCN	532
<i>LbrM.10.1610</i>	AVFQDLGFYKADFSMAEVMPWAYLATCDFLTNKKCMERNITQWPGMFCN	409
<i>LbrM.10.1620</i>	AVFQDFGFYKADFTKAEVMPWANLASCDFLTNKKCMENNITQWPEMFCN	409
<i>LbrM.10.1630</i>	AVFQDFGFYKADFTKAEVMPWANLASCDFLTNKKCMENNITQWPEMFCN	409
<i>LbrM.10.1640</i>	AVFHDHGFYKADFTKAEVMPWAYLASCDFLTNKKCMENNITQWPEMFCN	536
<i>LbrM.10.1650</i>	GILQDLGFYKADFSMAEVMPWAEYATCDFLTNKKCKEKNITQWPMFCN	409
<i>LbrM.10.1660</i>	GILQDLGFYKADFSMAEVMPWAEYATCDFLTNKKCKEKNITQWPMFCN	402
<i>LbrM.10.1670</i>	AVFEDLGFYKADFTKAEVMPWAKLASCDFLTNKKCKEKNITQWPEMFCN	426
<i>LbrM.10.1680</i>	AVFQDLGFYKADFTKAEVMPWANLASCDFLTNKKCMERNITQWPGMFCN	227
<i>LbrM.10.1690</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	321
<i>LbrM.10.1700</i>		195
<i>LbrM.10.1710</i>	AVFEDLGFYKADFSMAEVMPWGRNASCDFLTNKKCMEDNITQWPEMFCN	210
<i>LbrM.10.1720</i>	RAVRRLLGAERA	372





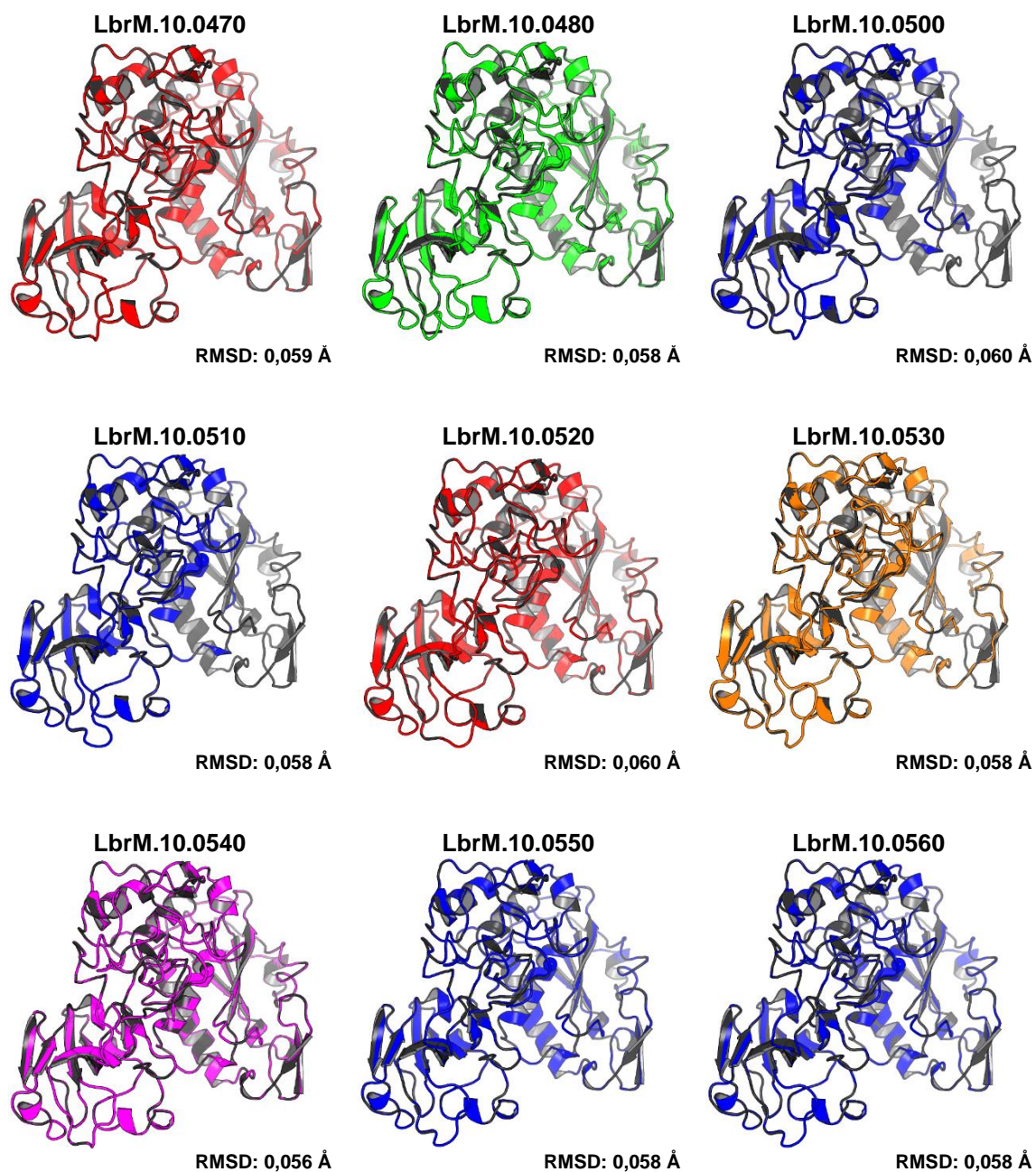




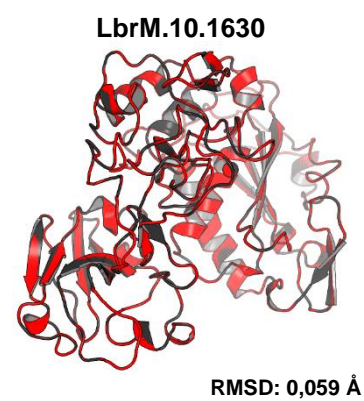
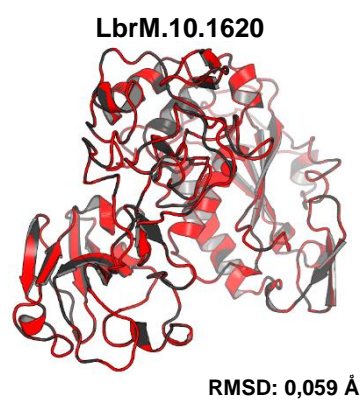
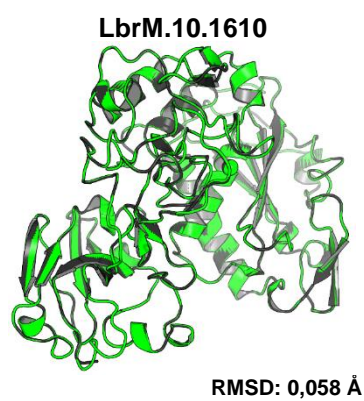
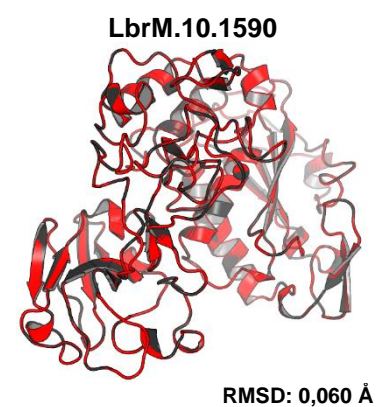
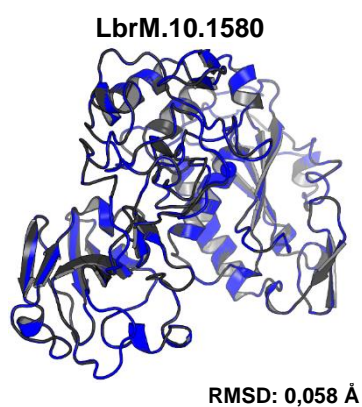
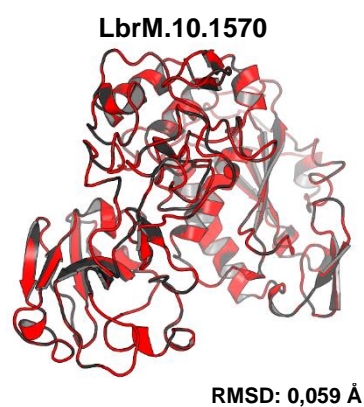
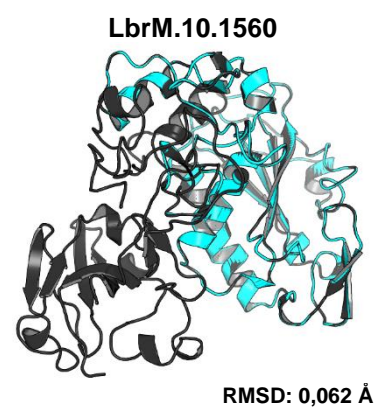
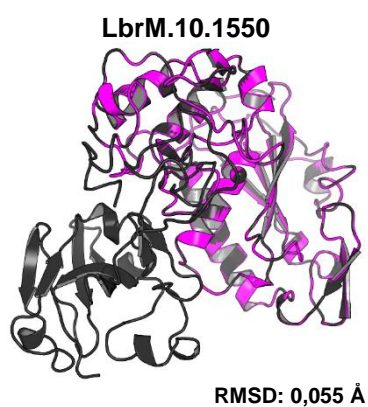
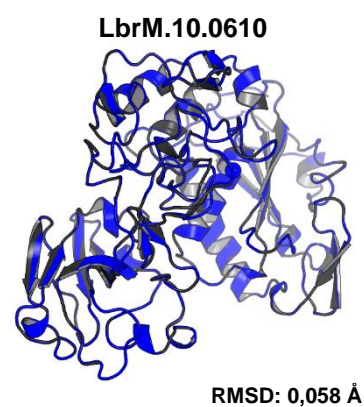
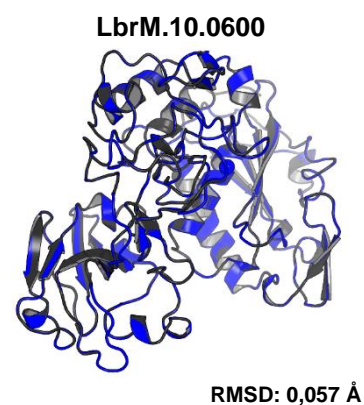
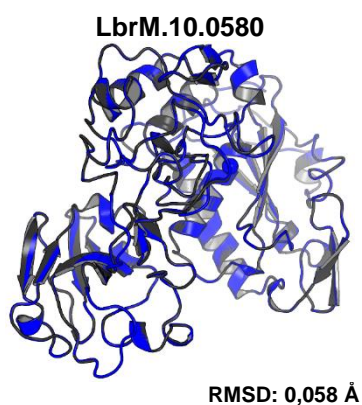
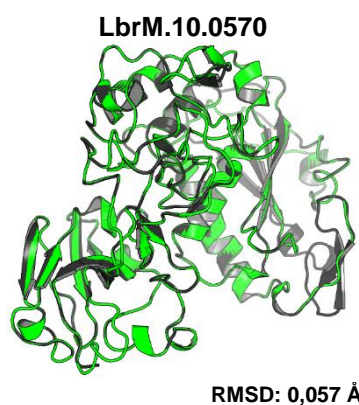
<i>LbrM.10.0470</i>	.....	577
<i>LbrM.10.0480</i>	.....	584
<i>LbrM.10.0490</i>	.....	195
<i>LbrM.10.0500</i>	.....	385
<i>LbrM.10.0510</i>	... ERMTALA . T . . . . . VTAALLGIVLAAMAGLVVGLLVISLS . . . . .	622
<i>LbrM.10.0520</i>	.....	575
<i>LbrM.10.0530</i>	... ERMTALA . T . . . . . VTAALLGIVLAAMAGLAVWLLLISLP . . . . .	501
<i>LbrM.10.0540</i>	.....	557
<i>LbrM.10.0550</i>	... ERMVLA . T . . . . . VTAVLLGIVLAAMAGLVSV . . . . .	615
<i>LbrM.10.0560</i>	... SERMTAL . A . . . . . TVTAVLLGIVLAAMAGLVVGLLVISLS . . . . .	623
<i>LbrM.10.0570</i>	.....	450
<i>LbrM.10.0580</i>	... ERMVLA . T . . . . . VTAVLLGIVLAAMAGLVVGLLVISLP . . . . .	622
<i>LbrM.10.0590</i>	GPRAAATALLV . A . . . . . ALLAIACA . . . . .	589
<i>LbrM.10.0600</i>	... ERMTALA . TVTAALLGIVLAAMAGLAVWLLLITIP . . . . .	611
<i>LbrM.10.0610</i>	... RMYVLATV . . . . . TAVLLGIVLAAMAGLVVGLLVISL . S . . . . .	621
<i>LbrM.10.1540</i>	... ERMVLA . T . . . . . VTAVLLGIVLAAMAGLVVGLLVISLS . . . . .	251
<i>LbrM.10.1550</i>	.....	559
<i>LbrM.10.1560</i>	.....	340
<i>LbrM.10.1570</i>	.....	583
<i>LbrM.10.1580</i>	... ERMVLAT . . . . . VTAVLLGIVLAAMAGLVVGLLVISLSRKLSGSRCG . . . . .	577
<i>LbrM.10.1590</i>	.....	707
<i>LbrM.10.1610</i>	.....	584
<i>LbrM.10.1620</i>	.....	584
<i>LbrM.10.1630</i>	.....	584
<i>LbrM.10.1640</i>	.....	711
<i>LbrM.10.1650</i>	.....	584
<i>LbrM.10.1660</i>	.....	577
<i>LbrM.10.1670</i>	.....	574
<i>LbrM.10.1680</i>	.....	401
<i>LbrM.10.1690</i>	... AATALLV . A . . . . . A . . . . . LLA IACA . . . . .	511
<i>LbrM.10.1700</i>	.....	195
<i>LbrM.10.1710</i>	.....	305
<i>LbrM.10.1720</i>	.....	486

## Anexo D

Alinhamento estrutural entre a proteína utilizada como molde na modelagem comparativa, PDB id 1LML, e o modelo construído pelo servidor Swiss-Model. O valor de RMSD entre as estruturas está mostrado para cada modelo.

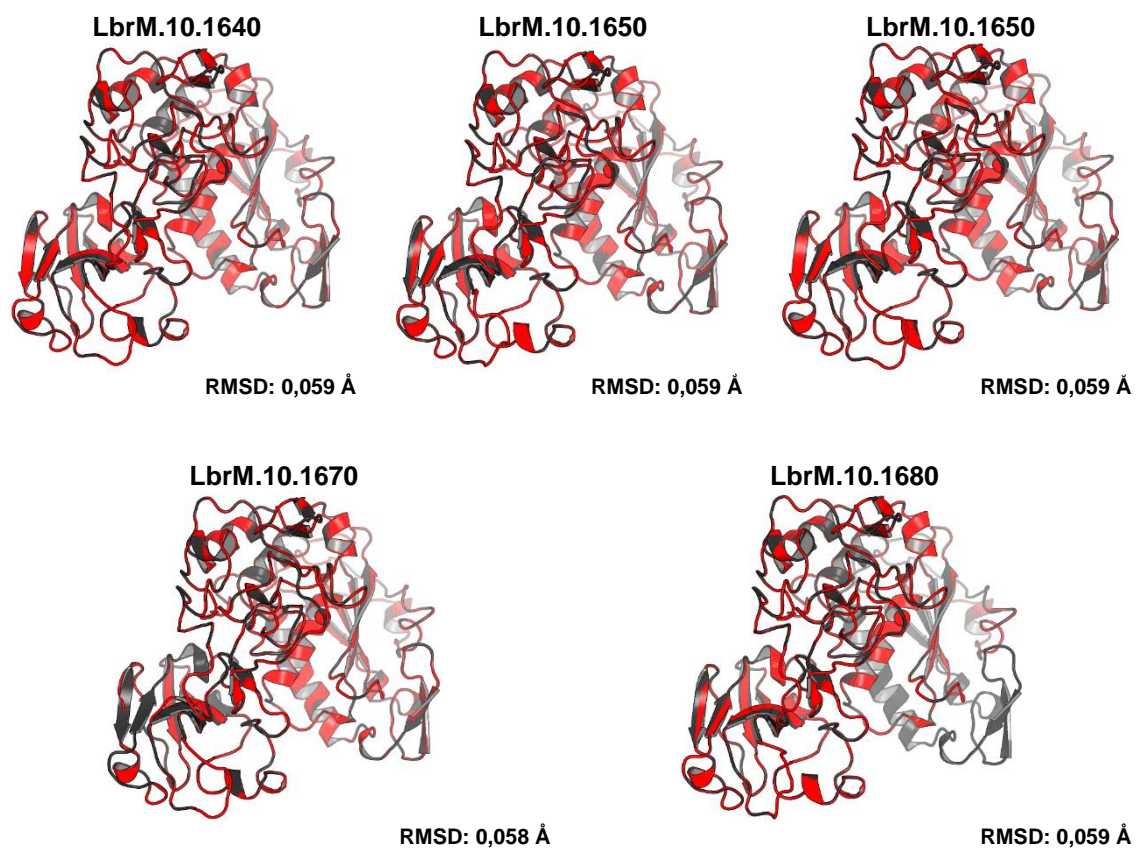


(continuação)





(continuação)



**Figura 1.** Alinhamento entre o molde PDB id 1LML (cinza) com o modelo construído pelo servidor Swiss-Model. As estruturas estão coloridas de acordo com o motivo que apresenta na região de inserção do sítio catalítico: DLT (vermelho), DLL (magenta), DLA (ciano), NRI (verde),SRM (laranja) e SSV (azul). Os valores de RMSD são dados em Angstrom (Å).

## Anexo E

**Tabela 1.** Resultado estatístico da análise do gráfico de Ramachandran, Fator-G, QMEAN6 e Z-Score para as estruturas construídas com o servidor Swiss-Model.

Proteína	Ramachandran (%)				Fator-G			QMEAN6	Z-Score
	Regiões favoráveis	Regiões permitidas	Generosamente permitidas	Regiões não permitidas	Diedros	Covalente	Global		
LbrM.10.0480	87,0	11,8	0,7	0,5	-0,13	0,28	0,04	0,625	-1,672
LbrM.10.0500	87,3	11,4	0,9	0,3	-0,09	0,34	0,09	0,595	-2,002
LbrM.10.0510	86,7	12,0	1,0	0,3	-0,10	0,32	0,07	0,594	-2,084
LbrM.10.0520	86,2	12,6	0,2	1,0	-0,08	0,33	0,09	0,668	-1,164
LbrM.10.0530	85,3	13,6	0,5	0,5	-0,09	0,31	0,08	0,664	-1,227
LbrM.10.0540	86,7	11,6	0,5	1,2	-0,09	0,30	0,07	0,651	-1,365
LbrM.10.0550	86,5	12,3	0,5	0,7	-0,08	0,33	0,09	0,661	-1,242
LbrM.10.0560	87,0	11,8	0,5	0,7	-0,08	0,30	0,07	0,658	-1,276
LbrM.10.0570	86,9	11,5	1,0	0,5	-0,11	0,32	0,07	0,639	-1,509
LbrM.10.0580	86,8	12,0	0,7	0,5	-0,09	0,34	0,08	0,658	-1,286
LbrM.10.0600	86,0	12,5	1,0	0,5	-0,08	0,32	0,08	0,717	-0,605
LbrM.10.0610	86,8	12,0	0,7	0,5	-0,08	0,32	0,08	0,643	-1,46
LbrM.10.1570	86,0	13,0	0,5	0,5	-0,10	0,31	0,07	0,645	-1,437
LbrM.10.1580	86,5	12,3	0,7	0,5	-0,08	0,32	0,08	0,644	-1,451
LbrM.10.1590	86,0	13,3	0,2	0,5	-0,10	0,30	0,07	0,665	-1,205
LbrM.10.1610	87,0	11,8	0,7	0,5	-0,12	0,29	0,05	0,644	-1,452
LbrM.10.1620	85,1	13,9	0,5	0,5	-0,12	0,31	0,06	0,656	-1,31
LbrM.10.1630	85,1	13,7	0,7	0,5	-0,11	0,30	0,06	0,638	-1,512
LbrM.10.1640	85,3	13,7	0,5	0,5	-0,11	0,32	0,07	0,658	-1,28
LbrM.10.1650	86,7	12,1	0,5	0,7	-0,10	0,27	0,05	0,634	-1,564
LbrM.10.1660	85,9	12,9	0,7	0,5	-0,09	0,34	0,09	0,637	-1,527
LbrM.10.1670	86,4	12,6	0,5	0,5	0,42	0,26	0,37	0,654	-1,354
LbrM.10.1680	84,9	13,0	1,5	0,6	-0,08	0,21	0,04	0,604	-1,927

## Anexo F

O script usado no programa R para geração do dendrograma estrutural segue abaixo.

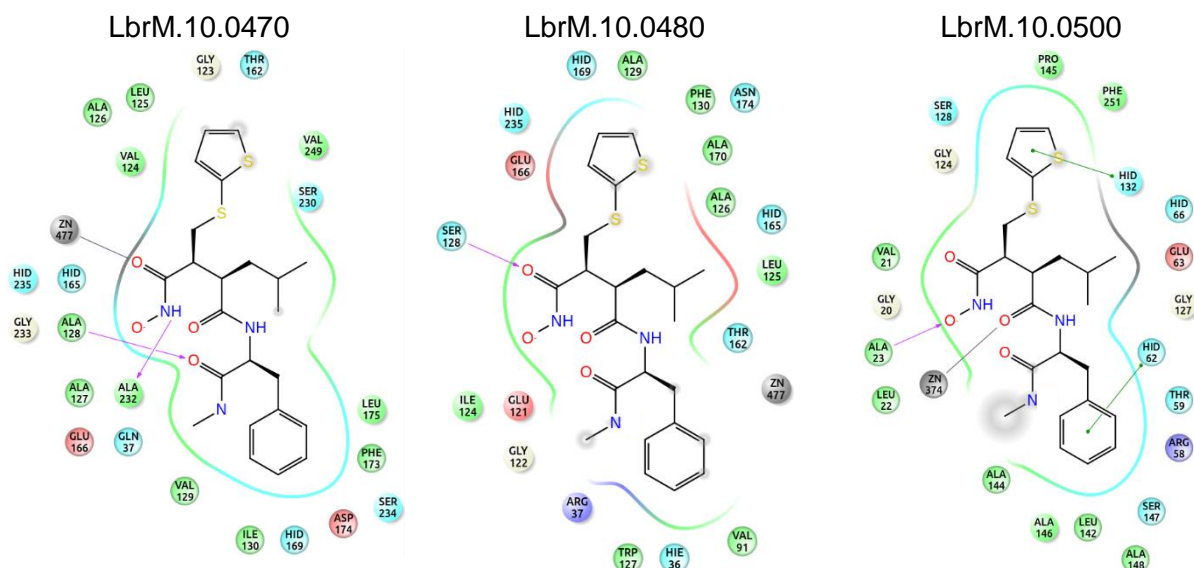
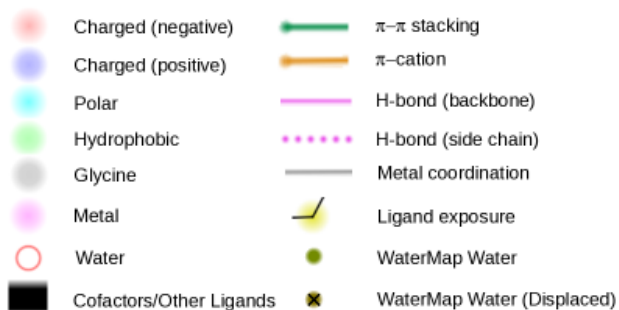
```
dados <- read.csv2(file = "Plot_dendrograma.csv")
normalizar <- function(vetor){
  (vetor - mean(vetor)) / sd(vetor)
}
dados.norm <- dados
for(j in 1:ncol(dados))
  dados.norm[,j] <- normalizar(dados[,j])

dados.final <- dados.norm[,c(2,3,4,5)]
library(vegan)
clust<-vegdist (dados.final, method = "euclidean")

plot(hclust(clust, method="average"), hang=-1,
labels=c("LbrM.10.0470", "LbrM.10.0480", "LbrM.10.0500", "LbrM.10.0510", "LbrM.10.0520", "LbrM.10.0530", "LbrM.10.0540", "LbrM.10.0550", "LbrM.10.0560", "LbrM.10.0570", "LbrM.10.0580", "LbrM.10.0590", "LbrM.10.0600", "LbrM.10.0610", "LbrM.10.1550", "LbrM.10.1560", "LbrM.10.1570", "LbrM.10.1580", "LbrM.10.1590", "LbrM.10.1610", "LbrM.10.1620", "LbrM.10.1630", "LbrM.10.1640", "LbrM.10.1650", "LbrM.10.1660", "LbrM.10.1670", "LbrM.10.1680", "LbrM.10.1690", "LbrM.10.1710"))
```

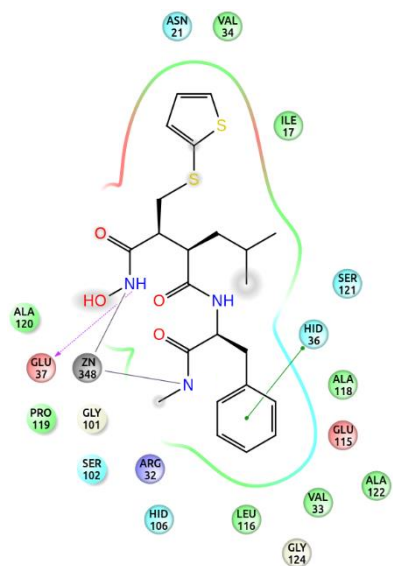
## Anexo G

Mapeamento das interações entre proteína e ligante Batimastat. A Figura abaixo representa a legenda de cores para identificação das interações proteína-ligante geradas pelo programa Maestro.

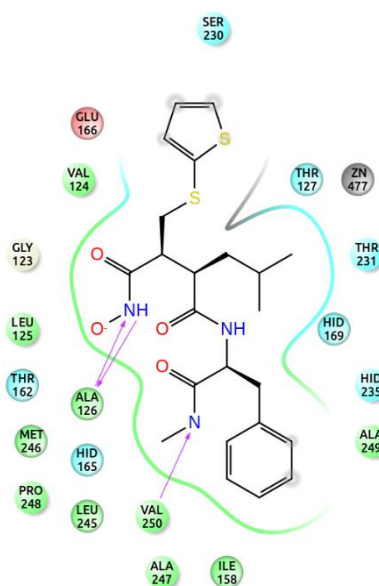


(continuação)

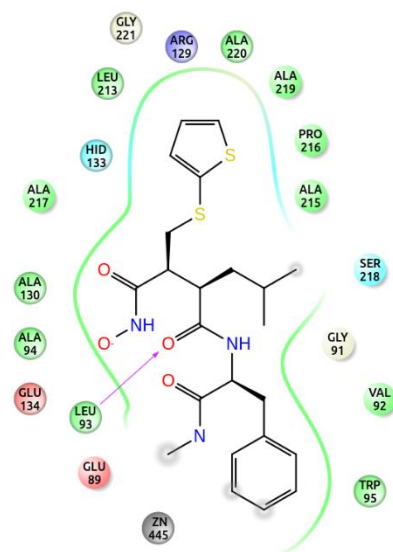
LbrM.10.0510



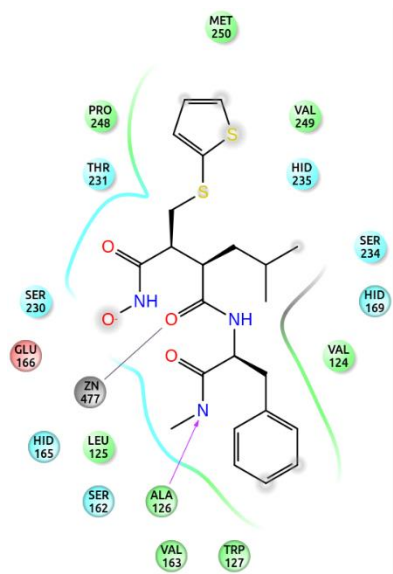
LbrM.10.0520



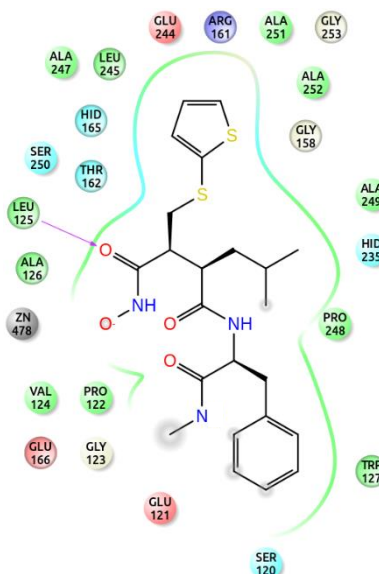
LbrM.10.0530



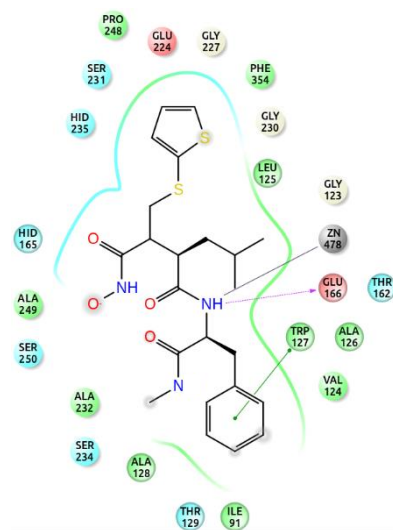
LbrM.10.0540



LbrM.10.0550



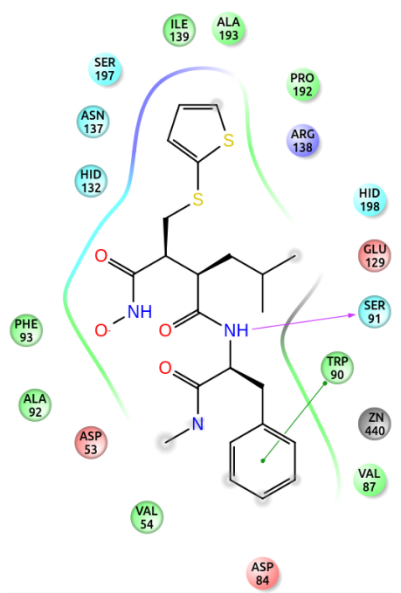
LbrM.10.0560



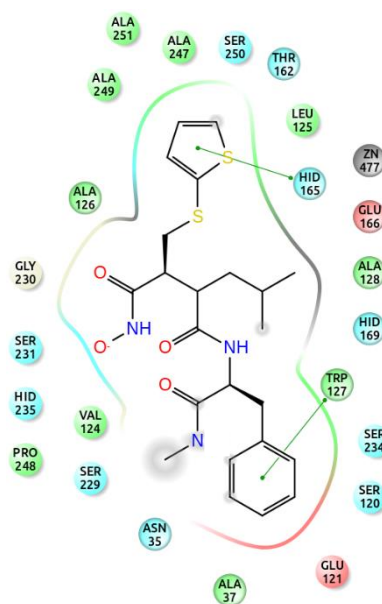


(continuação)

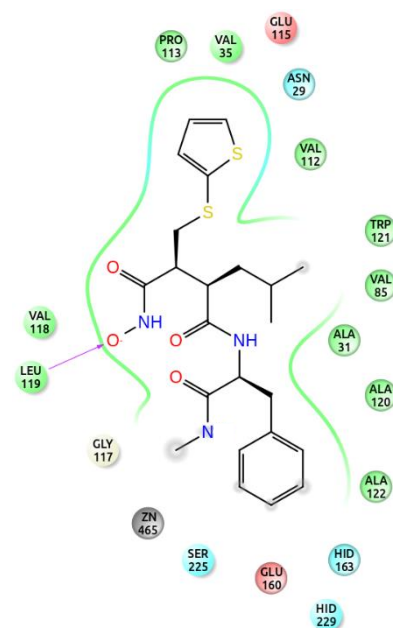
LbrM.10.0570



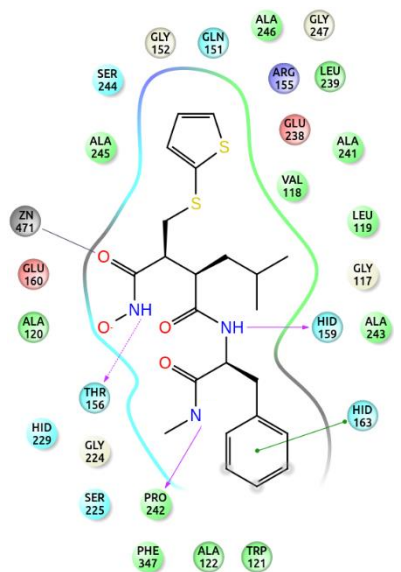
LbrM.10.0580



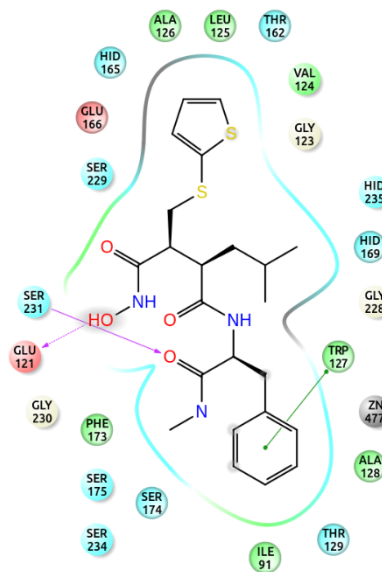
LbrM.10.0590



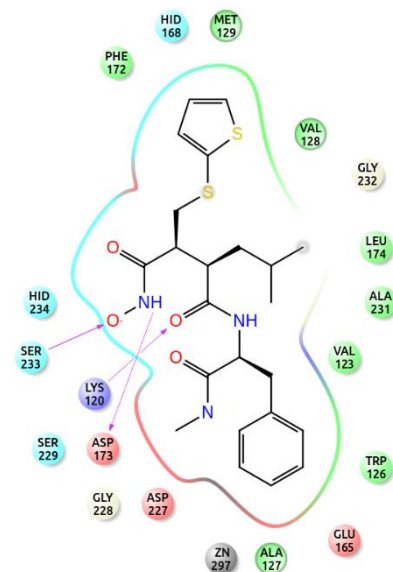
LbrM.10.0600



LbrM.10.0610

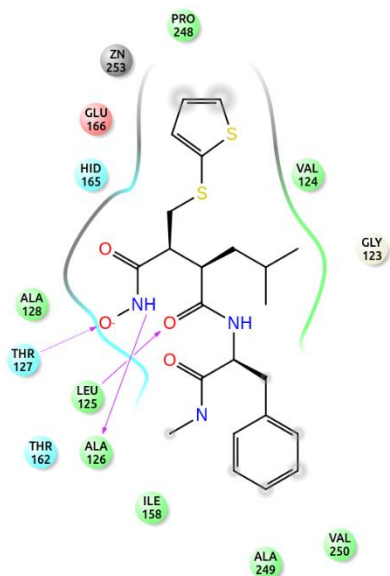


LbrM.10.1550

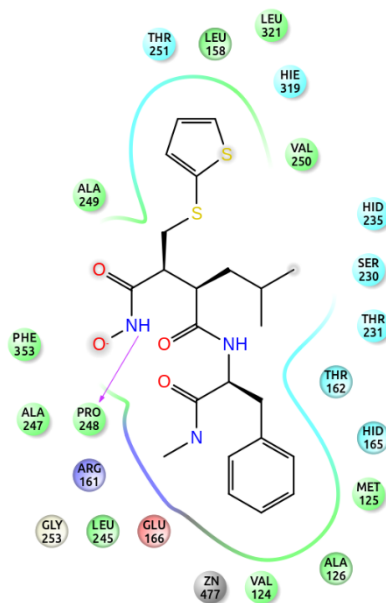


(continuação)

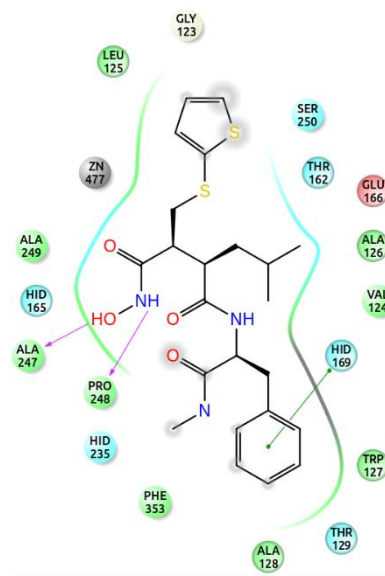
LbrM.10.1560



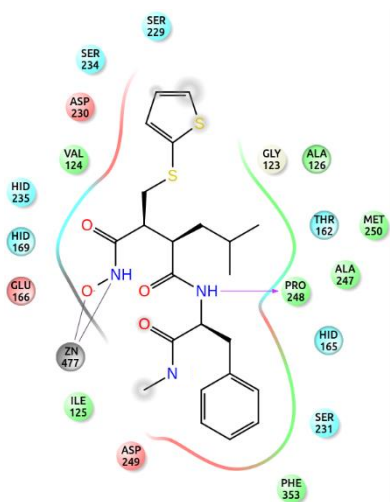
LbrM.10.1570



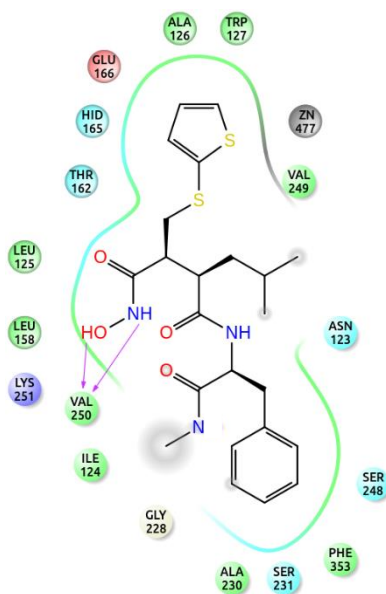
LbrM.10.1580



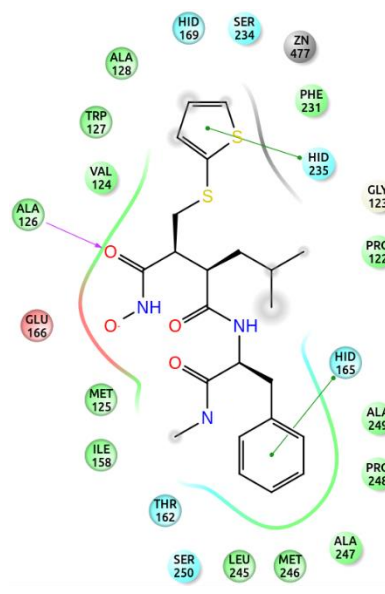
LbrM.10.1590



LbrM.10.1610

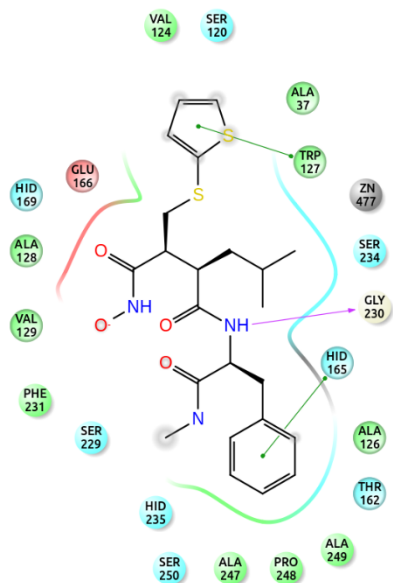


LbrM.10.1620

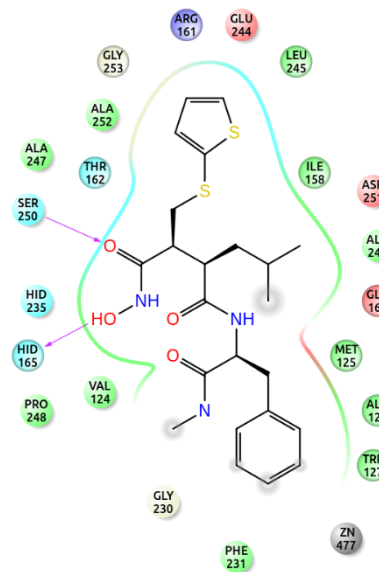


(continuação)

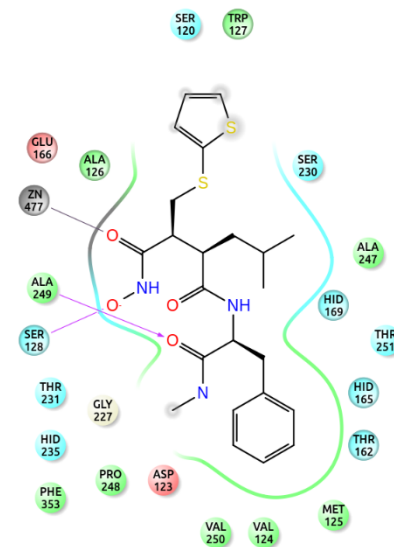
LbrM.10.1630



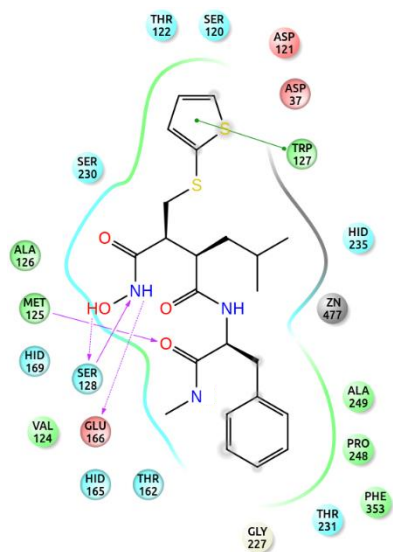
LbrM.10.1640



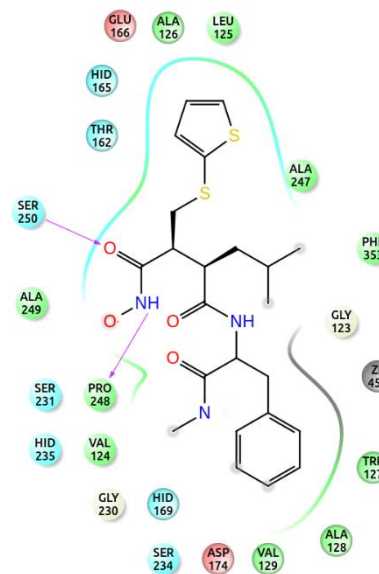
LbrM.10.1650



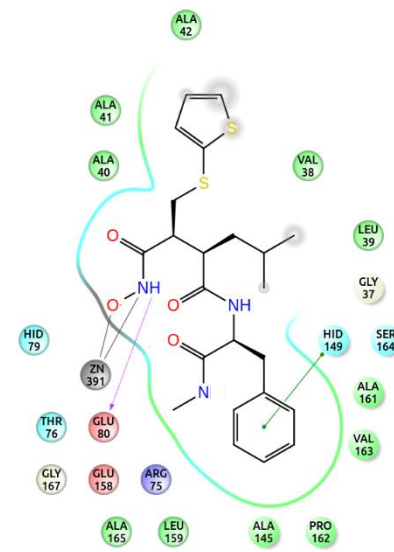
LbrM.10.1660



LbrM.10.1670



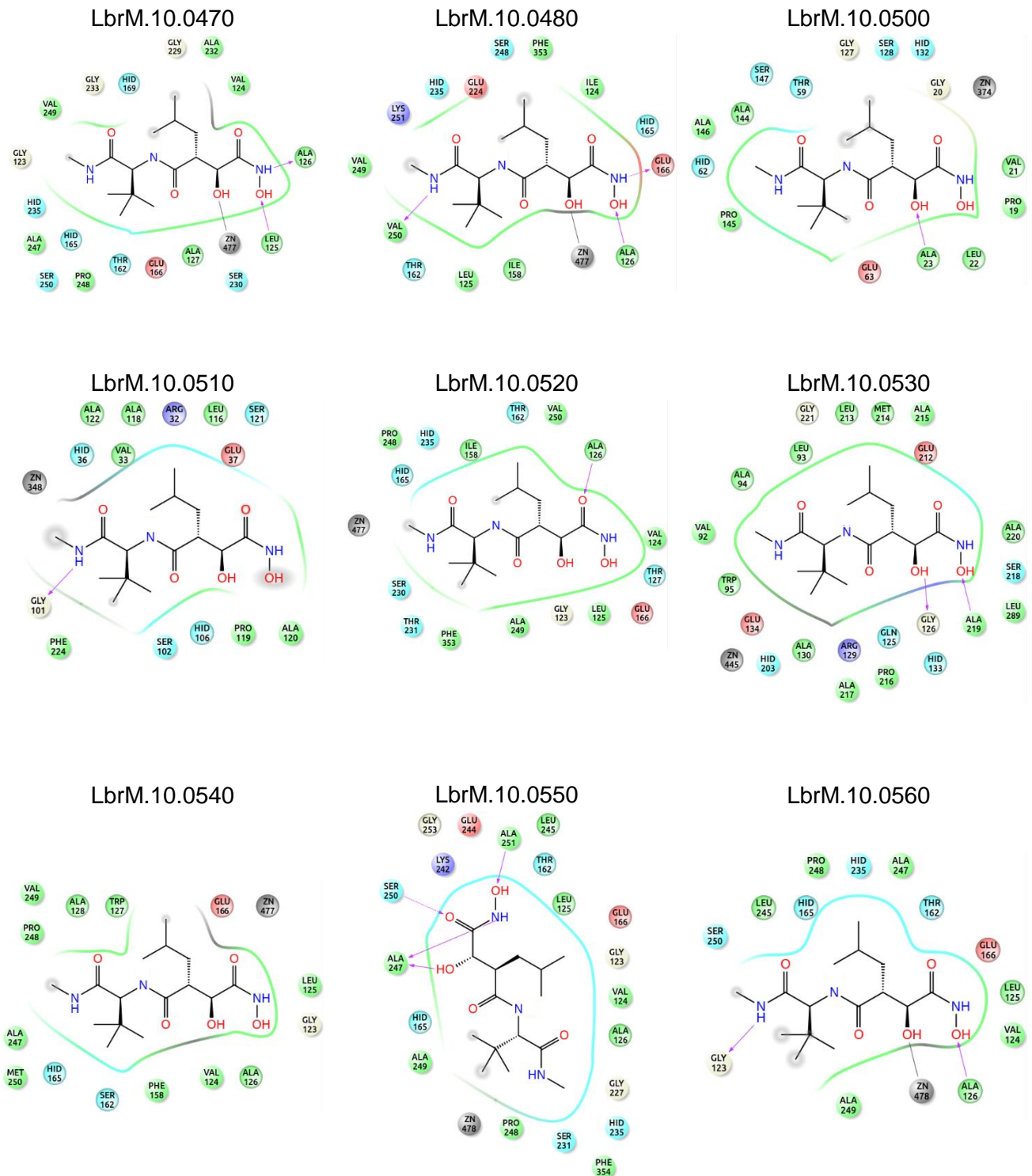
LbrM.10.1680





## Anexo H

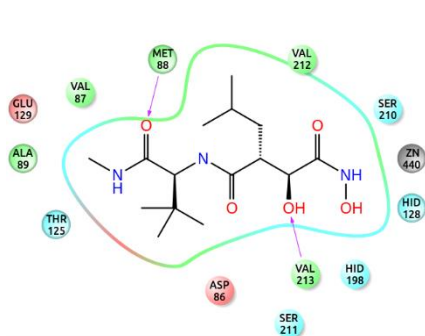
Maapeamento das interações entre proteína e ligante Marimastat.



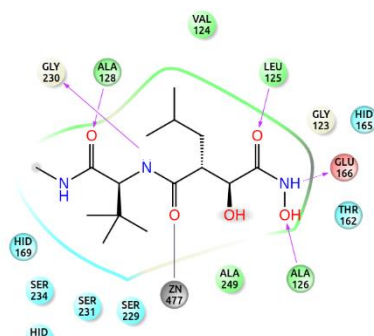


(continuação)

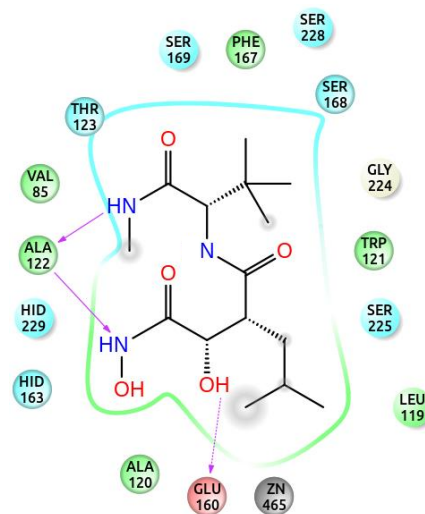
LbrM.10.0570



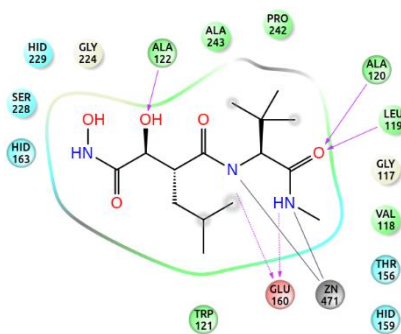
LbrM.10.0580



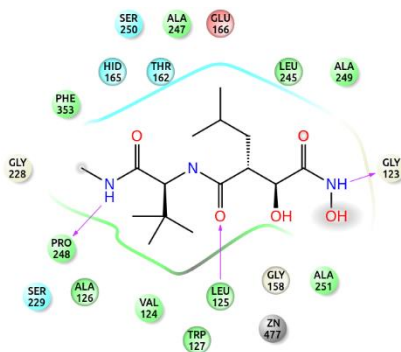
LbrM.10.0590



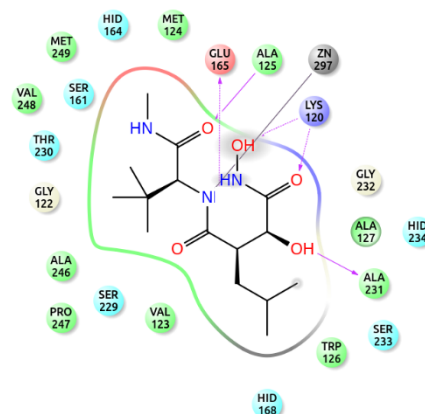
LbrM.10.0600



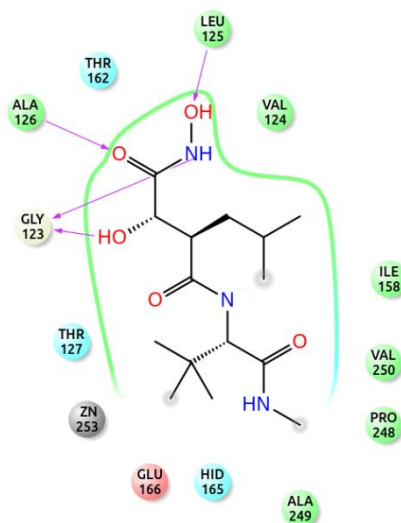
LbrM.10.0610



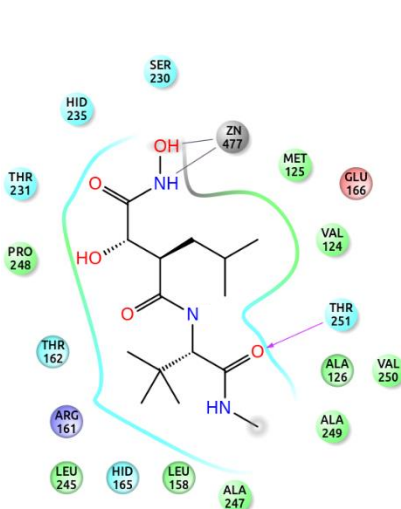
LbrM.10.1550



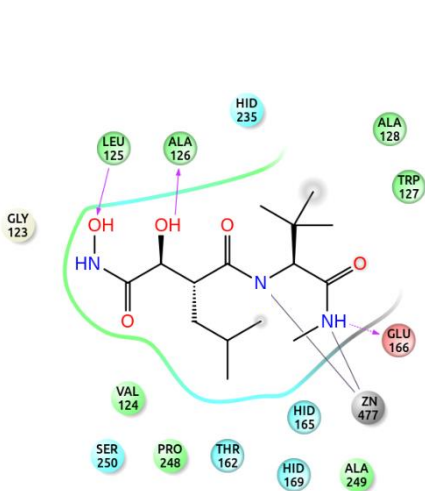
LbrM.10.1560



LbrM.10.1570

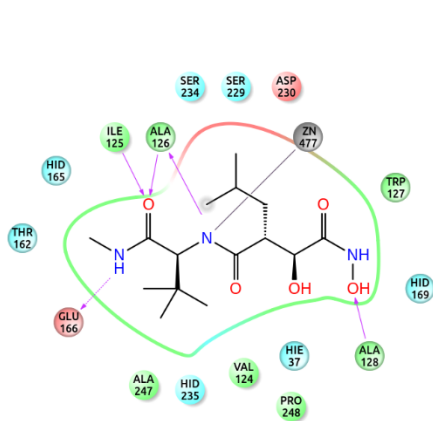


LbrM.10.1580

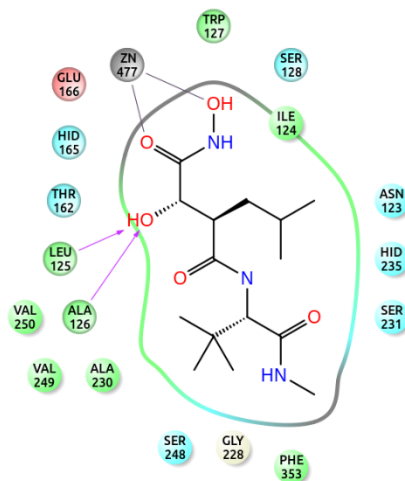


(continuação)

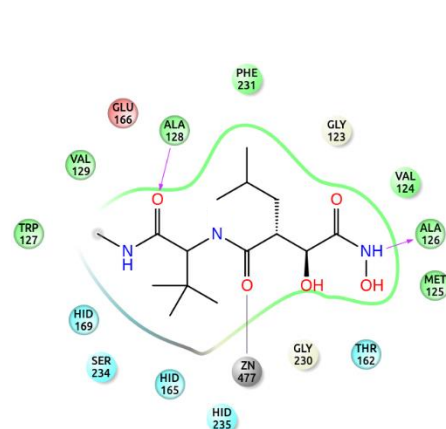
LbrM.10.1590



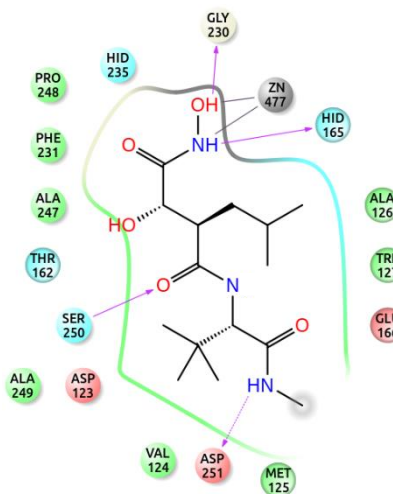
LbrM.10.1610



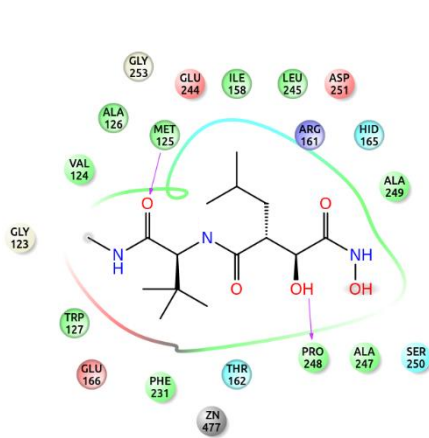
LbrM.10.1620



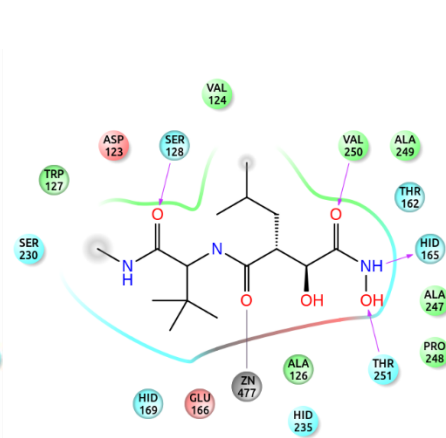
LbrM.10.1630



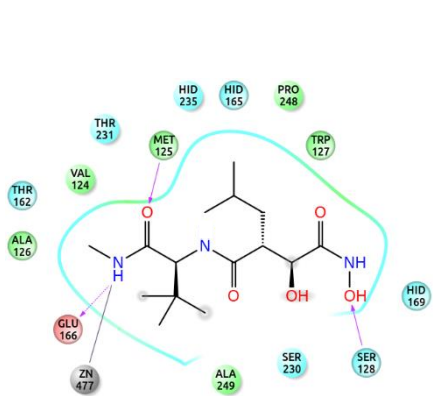
LbrM.10.1640



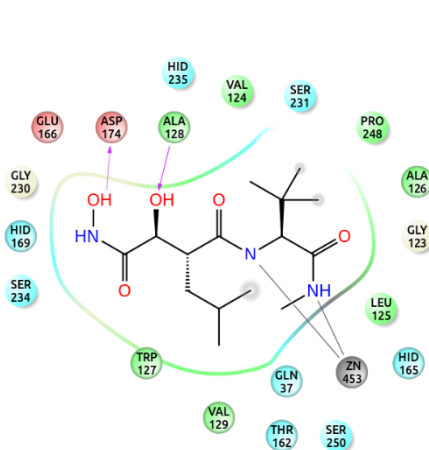
LbrM.10.1650



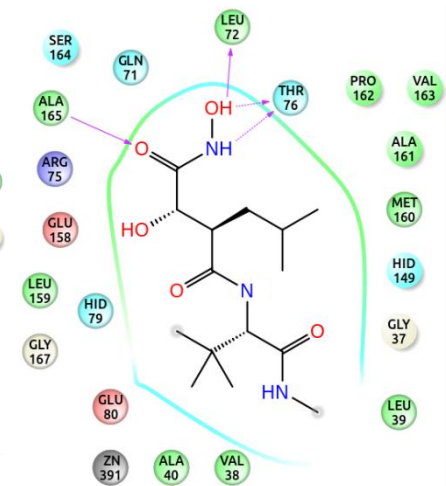
LbrM.10.1660



LbrM.10.1670

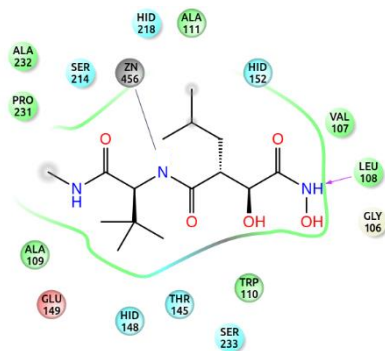


LbrM.10.1680

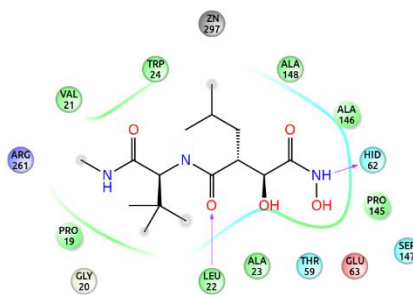


(continuação)

LbrM.10.1690



LbrM.10.1710

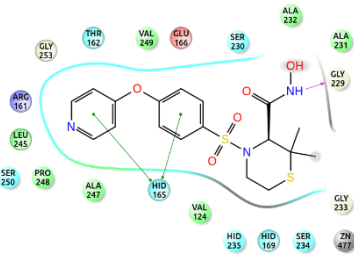




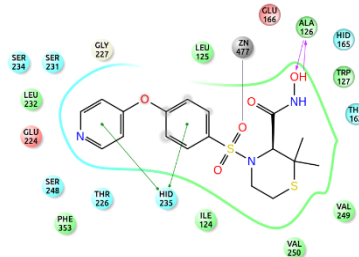
# Anexo I

Mapeamento das interações entre proteína e ligante Prinomastat.

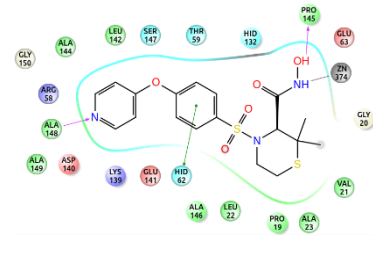
LbrM.10.0470



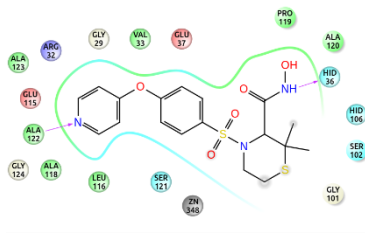
LbrM.10.0480



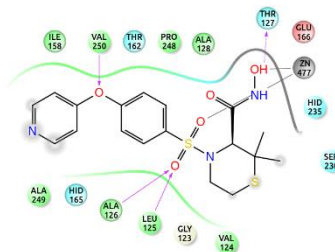
LbrM.10.0500



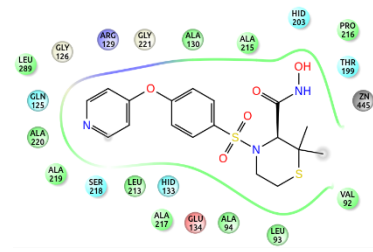
LbrM.10.0510



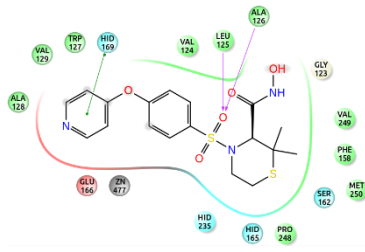
LbrM.10.0520



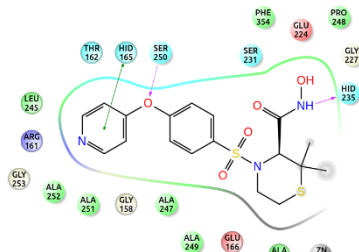
LbrM.10.0530



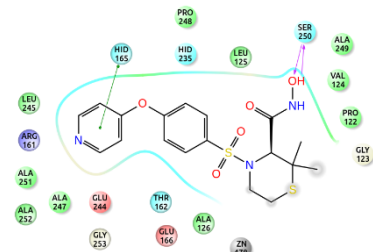
LbrM.10.0540



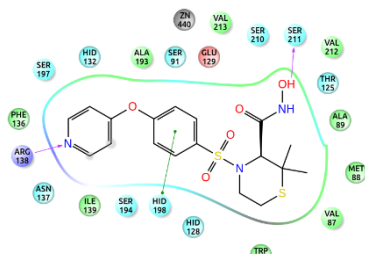
LbrM.10.0550



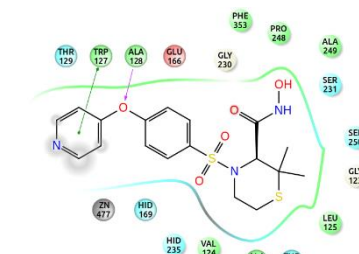
LbrM.10.0560



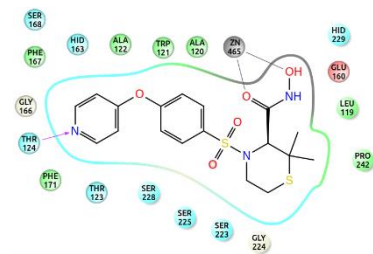
LbrM.10.0570



LbrM.10.0580



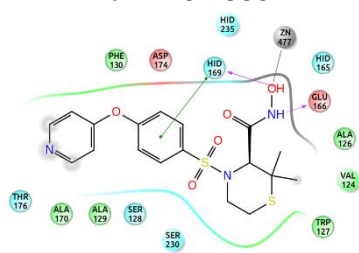
LbrM.10.0590



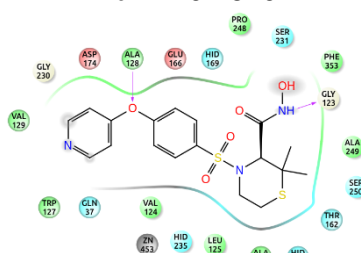


(continuação)

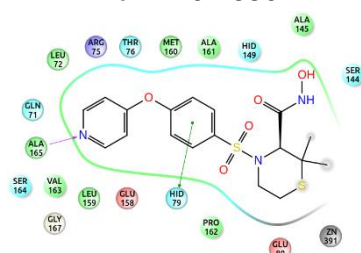
LbrM.10.1660



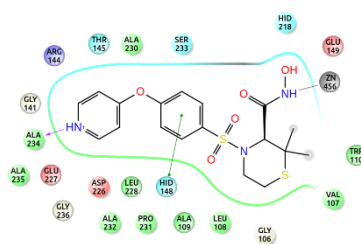
LbrM.10.1670



LbrM.10.1680



LbrM.10.1690



LbrM.10.1710

