

UNIVERSIDADE DO ESTADO DO AMAZONAS  
ESCOLA SUPERIOR DE CIÊNCIAS DA SAÚDE  
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOTECNOLOGIA E RECURSOS NATURAIS

MODELAGEM COMPARATIVA DA METALOPROTEASE VMBMP  
DE *Bothrops Moojeni* POR HOMOLOGIA ESTRUTURAL

KARLA NUNES DA SILVA

MANAUS – AM

2010

KARLA NUNES DA SILVA

MODELAGEM COMPARATIVA DA METALOPROTEASE VMBMP  
DE *Bothrops Moojeni* POR HOMOLOGIA ESTRUTURAL

Dissertação apresentada ao Programa de Pós-graduação em Biotecnologia e Recursos Naturais da Universidade do Estado do Amazonas – UEA, como requisito para a obtenção do grau de mestre.

Orientação: PROF. DR. JORGE LUÍS LÓPEZ LOZANO

MANAUS – AM

2010

*Ao prof. Dr. Luís Lozano,  
que foi o maior incentivador da realização deste trabalho.  
Ele me ensinou como cultivar a autonomia para buscar o conhecimento,  
E me mostrou que uma chance, uma vez concedida, não tem preço.  
Dedico este trabalho.*

## AGRADECIMENTOS

A profa. MSc. Priscila V. Capriles Golliat (National Laboratory for Scientific Computing – LNCC/MCT/RJ), pelas valiosas considerações sobre este trabalho, bem como o suporte na metodologia de validação dos modelos obtidos computacionalmente.

Ao prof. Dr. Jorge Luis López-Lozano (UEA), por toda a sua disponibilidade em orientar este trabalho, também estendendo o agradecimento à toda a equipe do Laboratório de Genômica e Proteômica da UEA.

Ao prof. Dr. Oscar Henrique Pereira Ramos (Département d'Endocrinologie, Métabolisme et Cancer – EMC Institut Cochin – France) pelo suporte ao algoritmo adaptado ao problema de modelagem comparativa de metaloproteases.

Aos amigos Ana Fabíola Peres, Rafael Siza, Maria das Dores Nogueira, André Higa, André Luís Ferreira, Bruno Medeiros, Elzafã Gomes e João Paulo Catunda, por seguirem nesta caminhada juntos na pós-graduação. Especialmente, à profa. Érica Maria Castro pelas orientações recebidas e pelo valioso aprendizado compartilhado na temática de modelagem molecular; e à Profa. MSc. Rebecca Tavares, pelos conhecimentos compartilhados acerca dos estudos sobre proteínas e toxinas de serpentes.

À minha família (pai, mãe e meu irmão Carlos Eduardo Nunes) por todo o apoio, compreensão e carinho que têm dedicado a mim, por todo esse tempo de vida acadêmica.

À Deus, que tem me dado sinais todos os dias para seguir um caminho novo e diferente, em busca do desenvolvimento da Educação, Ciência e Tecnologia no Estado do Amazonas.

Ao Programa de Pós-graduação em Biotecnologia e Recursos Naturais da Universidade do Estado do Amazonas – UEA, pelo suporte ao curso.

À CAPES/FAPEAM, pela concessão do suporte financeiro.

## RESUMO

O presente trabalho tem por objetivo a obtenção de um modelo estrutural 3D através da técnica de modelagem comparativa de uma metaloprotease VMBMP\_BOTMO, encontrada na peçonha da espécie de serpente *Bothrops moojeni*. O modelo proposto foi modelado a partir de uma metaloprotease homóloga, encontrada na peçonha de *Bothrops asper*. A escolha desta metaloprotease foi realizada com base na similaridade estrutural entre as duas macromoléculas, evidenciada por alinhamento da cadeia de resíduos de aminoácidos. De posse de um molde estrutural (*B. asper*), foi gerado um modelo tridimensional da metaloprotease de *B. moojeni*. Este modelo passou por um processo de validação, com o auxílio de softwares específicos para esta etapa, onde foi possível a análise da conformidade da estrutura tridimensional proposta para a macromolécula. Uma vez conhecido o arranjo tridimensional molecular, é possível utilizar esta biomolécula para duas aplicações importantes: (1) o design de um inibidor para outras metaloproteases, presentes em venenos de serpentes do mesmo gênero; (2) no design de fármacos como os indutores de resposta imune, contra doenças endêmicas tropicais.

Palavras-chaves: *Bothrops moojeni*, metaloprotease, modelagem molecular comparativa, design de fármacos.

## ABSTRACT

This study aims to carry out a structural model three-dimensional, through of comparative modeling of a metalloproteinase VMBMP\_BOTMO, found in the venom of snake species *Bothrops moojeni*. The proposed model was built from a homologous metalloproteinase, it found in the venom of *Bothrops asper*. The choice of this metalloproteinase was based on structural similarity between the two macromolecules, as evidenced by alignment 2D of the chain of amino acid residues. Armed with such a structural (*B. asper*), was generated a three-dimensional model of metalloproteinase *B. moojeni*. This model is being validated, with the help of specific software for this step, which will allow the analysis of agreement of 3D structural, it proposed for the biomolecule. Once known three-dimensional molecular arrangement, we can use this molecule in two important applications: (1) design of inhibitor drugs for another metalloproteinases, presents in snake venoms from the same genre *Bothrops*; (2) in design of drugs, like inductors of immune response, against tropical endemic diseases.

Keywords: *Bothrops moojeni*, metalloproteinase, comparative molecular modeling, drug design.

## LISTA DE TABELAS

Tabela 1 – Esquema das influências entre variáveis de espontaneidade, energia e entropia. ....	25
Tabela 2 – Interpretação dos resultados para a avaliação do Molprobit. ....	36
Tabela 3 – Características da metaloprotease P85314 VMBMP_BOTMO. ....	44
Tabela 4 – Amostra de seqüências com identidade acima de 49% com a metaloprotease de <i>B.moojeni</i> . ....	56
Tabela 5 – Valores médios de energia para os modelos gerados. ....	63
Tabela 6 – Resultados de qualidade geométrica e estereoquímica, obtidos por Molprobit. ....	66
Tabela 7 – Comparativo para os resultados dos cinco modelos, obtidos no servidor Harmony. ....	70
Tabela 8 – Comparativo dos <i>z-scores</i> obtidos pelo ProSA-web. ....	73

## LISTA DE FIGURAS

Fig. 1 – Estruturas dos 20 principais aminoácidos constituintes das proteínas (Adaptado NELSON & COX, 2007).	12
Fig. 2 – Etapas de enovelamento de uma proteína: estruturas primária, secundária, terciária e quaternária (Adaptado NELSON & COX, 2007).	12
Fig. 3 – Possíveis rotações livres em ligações covalentes de um polipeptídeo (adaptado NELSON & COX, 2007)...	13
Fig. 4 – Gráfico de Ramachandran, para regiões permitidas de ângulos $\Phi$ e $\Psi$ (Adaptado NELSON & COX, 2007).	14
Fig. 5 – Estrutura de arranjo de resíduos de aminoácidos para a conformação alfa-hélice (a). Estrutura alfa-hélice vista de cima (b) (Adaptado KOOLMAN & ROEHM, 2005).	15
Fig. 6 – Demonstração de $\beta$ -folhas paralelas e antiparalelas (Adaptado NELSON & COX, 2007).	16
Fig. 7 – Região de $\beta$ -turn para a união de duas estruturas secundárias de um polipeptídeo (Adaptado KHOLMANN & ROEHM, 2007).	17
Fig. 8 – Estruturas terciárias coincidentes com a estrutura biologicamente ativa da proteína. (Adaptado MURRAY <i>et al.</i> , 2003).	19
Fig. 9 – Esquema do efeito de solvatação em molécula apolar.	24
Fig. 10 – Relação das variações da entropia total do entorno ( $\Delta S$ ) e da energia de Gibbs ( $\Delta G$ ).	26
Fig. 11 – Ocorrência da espécie de serpente <i>Bothrops moojeni</i> no território brasileiro (BUTANTAN, 2000).	42
Fig. 12 – Exemplar de serpente <i>Bothrops moojeni</i> . Fonte: Bernarde, 2009.	42
Fig. 13 – Seqüência de aminoácidos da VMBMP_BOTMO (P85314), de <i>Bothrops moojeni</i> .	43
Fig. 14 – Esquema geral do procedimento de modelagem molecular estrutural.	48
Fig. 15 – Seqüência de aminoácidos da VMBP1_BOTAS (P83512), de <i>Bothrops asper</i> .	51
Fig. 16 – Estrutura tridimensional da metaloprotease VMBP1_BOTAS P83512 ( <i>B. asper</i> ).	52
Fig. 17 – Script de alinhamento das seqüências alvo e molde, a ser executado pelo aplicativo MODELLER.	53
Fig. 18 – Script para modelagem molecular da seqüência-alvo em função da seqüência-molde. O arquivo de saída oferece cinco modelos selecionados.	54
Fig. 19 – Alinhamento múltiplo para as seqüências de metaloproteases com identidade acima de 70% com a metaloprotease de <i>B. moojeni</i> (P85314). Obtido através do aplicativo ClustalW ( <a href="http://services.uniprot.org/clustalw">http://services.uniprot.org/clustalw</a> ).	59
Fig. 20 – Alinhamento da seqüência-alvo ( <i>B. moojeni</i> P85314.1) e da seqüência-molde ( <i>B. asper</i> P.83512.2). Obtido através do aplicativo ClustalW ( <a href="http://services.uniprot.org/clustalw">http://services.uniprot.org/clustalw</a> ).	60
Fig. 21 – Delimitação de regiões de <i>loops</i> para a seqüência-alvo de <i>B. moojeni</i> .	61
Fig. 22 – Predição da estrutura secundária de VMBMP_BOTMO de <i>B. moojeni</i> .	61
Fig. 23 – Gráficos de Ramachandran fornecidos pelo Molprobitry para o modelo P1.	65
Fig. 24 – Gráfico de detecção de erros grosseiros no enovelamento do modelo P1.	67
Fig. 25 – Detecção de erros locais no modelo P1 (verde), através do comparativo com a seqüência reversa de aminoácidos (vermelho).	69
Fig. 26 – Estrutura tridimensional sugerida pelo Harmony para o modelo P1.	71
Fig. 27 – Qualidade global estimada pelo ProSA-web para o modelo P1.	72
Fig. 28 – Qualidade local estimada para o modelo P1.	74
Fig. 29 – Correspondência de seqüências 3D-1D para o modelo P1.	76
Fig. 30 – Sobreposição de modelo teórico obtido de metaloprotease <i>B. moojeni</i> (VMBMP_BOTMO) e o modelo experimental (BmooMPa-I).	78
Fig. 31 – Sobreposição de modelo obtido de metaloprotease <i>B. moojeni</i> P85314 (colorido) e <i>B. asper</i> P83512 (cinza) como molde utilizado.	78
Fig. 32 – Representação esquemática do modelo experimental de BmooMPa-I, elucidando as posições de pontes dissulfeto, histidinas catalíticas e metal zinco.	79
Fig. 33 – Representação esquemática do modelo teórico de VMBMP_BOTMO: o <i>loop</i> catalítico (vermelho), as histidinas do sítio catalítico (verde).	80
Fig. 34 – Representação esquemática do sítio ativo e a participação do íon zinco nas estruturas da metaloprotease de <i>B. moojeni</i> (A), na adamalisina-II (B) e na astacina (C). Os resíduos de aminoácidos interagentes com o íon zinco são representados em esferas. As interações do zinco catalítico com as histidinas são representadas em pontilhado.	80

Fig. 35 – Representação esquemática da sobreposição das cadeias principais ( <i>backbones</i> ) das metaloproteases e das adamalinas, revelando a divergência estrutural no <i>loop</i> catalítico quanto a classificação em hemorrágica e não-hemorrágica. ....	82
Fig. 36 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P1. ....	90
Fig. 37 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P2. ....	91
Fig. 38 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P3. ....	92
Fig. 39 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P4. ....	93
Fig. 40 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P5. ....	94
Fig. 41 – Estrutura tridimensional sugerida pelo Harmony para o modelo P2. ....	95
Fig. 42 – Estrutura tridimensional sugerida pelo Harmony para o modelo P3. ....	95
Fig. 43 – Estrutura tridimensional sugerida pelo Harmony para o modelo P4. ....	96
Fig. 44 – Estrutura tridimensional sugerida pelo Harmony para o modelo P5. ....	96
Fig. 45 – Qualidade global estimada pelo ProSA-web para o modelo P2. ....	97
Fig. 46 – Qualidade global estimada pelo ProSA-web para o modelo P3. ....	97
Fig. 47 – Qualidade global estimada pelo ProSA-web para o modelo P4. ....	98
Fig. 48 – Qualidade global estimada pelo ProSA-web para o modelo P5. ....	98
Fig. 49 – Qualidade local estimada para o modelo P2. ....	99
Fig. 50 – Qualidade local estimada para o modelo P3. ....	99
Fig. 51 – Qualidade local estimada para o modelo P4. ....	100
Fig. 52 – Qualidade local estimada para o modelo P5. ....	100
Fig. 53 – Correspondência de seqüências 3D-1D para o modelo P2. ....	101
Fig. 54 – Correspondência de seqüências 3D-1D para o modelo P3. ....	101
Fig. 55 – Correspondência de seqüências 3D-1D para o modelo P4. ....	101
Fig. 56 – Correspondência de seqüências 3D-1D para o modelo P5. ....	102
Fig. 57 – Correspondência de seqüências 3D-1D para o molde ( <i>B.asper</i> ). ....	102

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
1.1	IMPORTÂNCIA BIOLÓGICA DAS PROTEÍNAS	10
1.2	ESTRUTURA DAS PROTEÍNAS E ARRANJO TRIDIMENSIONAL	11
1.3	TERMODINÂMICA DE ENOVELAMENTO DAS PROTEÍNAS	19
1.3.1	Formação da estrutura terciária: enovelamento ( <i>folding</i> )	19
1.3.2	Hipótese termodinâmica para o enovelamento das proteínas	20
1.3.3	Hipótese cinética para o enovelamento das proteínas	21
1.3.4	Termodinâmica vs cinética	23
1.3.5	A interação hidrofóbica	23
1.3.6	Entropia, entalpia e energia livre de Gibbs	24
1.3.7	Espontaneidade e energia de Gibbs	26
1.4	MÉTODOS DE PREDIÇÃO DA ESTRUTURA TRIDIMENSIONAL	27
1.5	ASPECTOS DA MODELAGEM MOLECULAR POR HOMOLOGIA	28
1.5.1	Breve histórico da modelagem molecular	28
1.5.2	Predição por modelagem comparativa	31
1.5.3	Banco de dados: comparação de seqüências de aminoácidos	31
1.5.4	Similaridade entre seqüências	32
1.5.5	Ferramentas para modelagem da estrutura 3D por homologia estrutural comparativa	33
1.5.6	Ferramentas para validação de modelos	34
1.6	GÊNERO <i>BOTHROPS</i>	39
1.6.1	Ocorrência	39
1.6.2	Domínios e metaloprotease	39
1.6.3	Mecanismo de catálise das metaloproteases	40
1.6.4	A espécie <i>Bothrops moojeni</i>	41
<b>2</b>	<b>JUSTIFICATIVA</b>	<b>45</b>
<b>3</b>	<b>OBJETIVOS</b>	<b>47</b>
3.1	OBJETIVOS GERAIS	47
3.2	OBJETIVOS ESPECÍFICOS	47
<b>4</b>	<b>MATERIAIS E MÉTODOS</b>	<b>48</b>
4.1	OBTENÇÃO DE ESTRUTURAS SIMILARES (SEQÜÊNCIAS-MOLDES)	49
4.2	ALINHAMENTO DE SEQÜÊNCIA-ALVO VERSUS MOLDE	50
4.3	OBTENÇÃO DO MODELO PRIMÁRIO	52
4.4	VALIDAÇÃO DE MODELOS	54
<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>55</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS</b>	<b>77</b>
<b>7</b>	<b>REFERÊNCIAS</b>	<b>84</b>

# 1 INTRODUÇÃO

## 1.1 *Importância biológica das proteínas*

As proteínas são importantes substâncias catalisadoras para reações metabólicas, sinalizadoras de processos celulares e constituintes estruturais de cabelos, ossos, tendões e dentes. Há uma imensa variedade e sofisticação das proteínas no organismo humano, refletindo a especificidade que cada uma delas possui para determinados processos biológicos. O enovelamento de um novo polipeptídeo sintetizado, se tornando uma proteína biologicamente funcional resulta em um arranjo tridimensional específico (conformação). Durante o enovelamento, algumas modificações pós-traducionais podem adicionar novos agrupamentos químicos, ou remover segmentos peptídicos que são desnecessários para o bom funcionamento da proteína (RAMOS, 2007).

Exemplos de funcionalidade das proteínas são algumas deficiências nutricionais ou genéticas, e até mesmo de origem desconhecida, que venham a impedir o enovelamento protéico, tornando-se assim muito prejudiciais à saúde. São os casos de Doença de Creutzfeldt-Jakob, do Tremor epizoótico, do Mal de Alzheimer, da encefalopatia espongiiforme bovina (doença da vaca-louca), dentre outras (BAKER & SALI, 2001).

Outro exemplo de funcionalidade das proteínas seria no processo da ligação de um substrato (reagente) ao sítio catalítico de uma enzima (proteína), as duas moléculas interagentes precisam ser estruturalmente complementares, para que ocorra um acoplamento efetivo e que a catálise ocorra com eficiência. Como também na ligação de um hormônio e seu receptor, essa condição de complementaridade também é requerida, e na interação entre uma molécula de antígeno com uma molécula de anticorpo específico (STOCKER & BODE, 1995).

Portanto, a conformação correta para uma funcionalidade biológica eficiente é fator importante para a análise da proteína, não bastando apenas a seqüência primária de aminoácidos, mas sim toda a disposição tridimensional dos seus átomos constituintes.

Por sua vez, os termos “configuração” e “conformação” são frequentemente confundidos. A configuração se refere quanto ao relacionamento geométrico de átomos, como aquela que permite a distinção entre um aminoácido levógiro ou dextrógiro. Uma interconversão de alternativas configuracionais requer a quebra de ligações covalentes. Já a conformação se refere ao relacionamento espacial de cada um dos átomos de uma molécula. A interconversão entre as conformações ocorre sem quebra de ligações covalentes, com a manutenção de uma configuração, apenas pela rotação de átomos em torno de ligações simples (KACZANOWSKI & ZIELENKIEWICZ, 2008).

## **1.2 Estrutura das proteínas e arranjo tridimensional**

As proteínas são as moléculas mais abundantes nas células vivas, formadas por peptídeos, que por sua vez são constituídos por aminoácidos. Segundo Lesk (2002), a seqüência de aminoácidos é fator determinante para a predição da estrutura tridimensional de um polipeptídeo, assim como sua função biológica. O estudo de comparativo entre seqüências de aminoácidos permite então a determinação do grau de identidade e de similaridade entre estas.

Existem 20 tipos de aminoácidos principais, ilustrados na Fig.1 As moléculas de aminoácidos unem-se por ligações covalentes chamadas ligações peptídicas, a qual por uma reação de desidratação, interliga um grupamento  $\alpha$ -carboxila de um resíduo de aminoácido e o grupo  $\alpha$ -amino de outro resíduo de aminoácido.

A síntese e enovelamento de uma proteína segue uma ordem de etapas de formação das cadeias polipeptídicas (Fig.2). O enovelamento das cadeias é realizado segundo as variações de energia livre de Gibbs ( $\Delta G$ ), assim como a variação de entalpia ( $\Delta H$ ) e a entropia ( $\Delta S$ ) no processo de enovelamento (*fold*ing), formando assim as estruturas dos polipeptídeos.

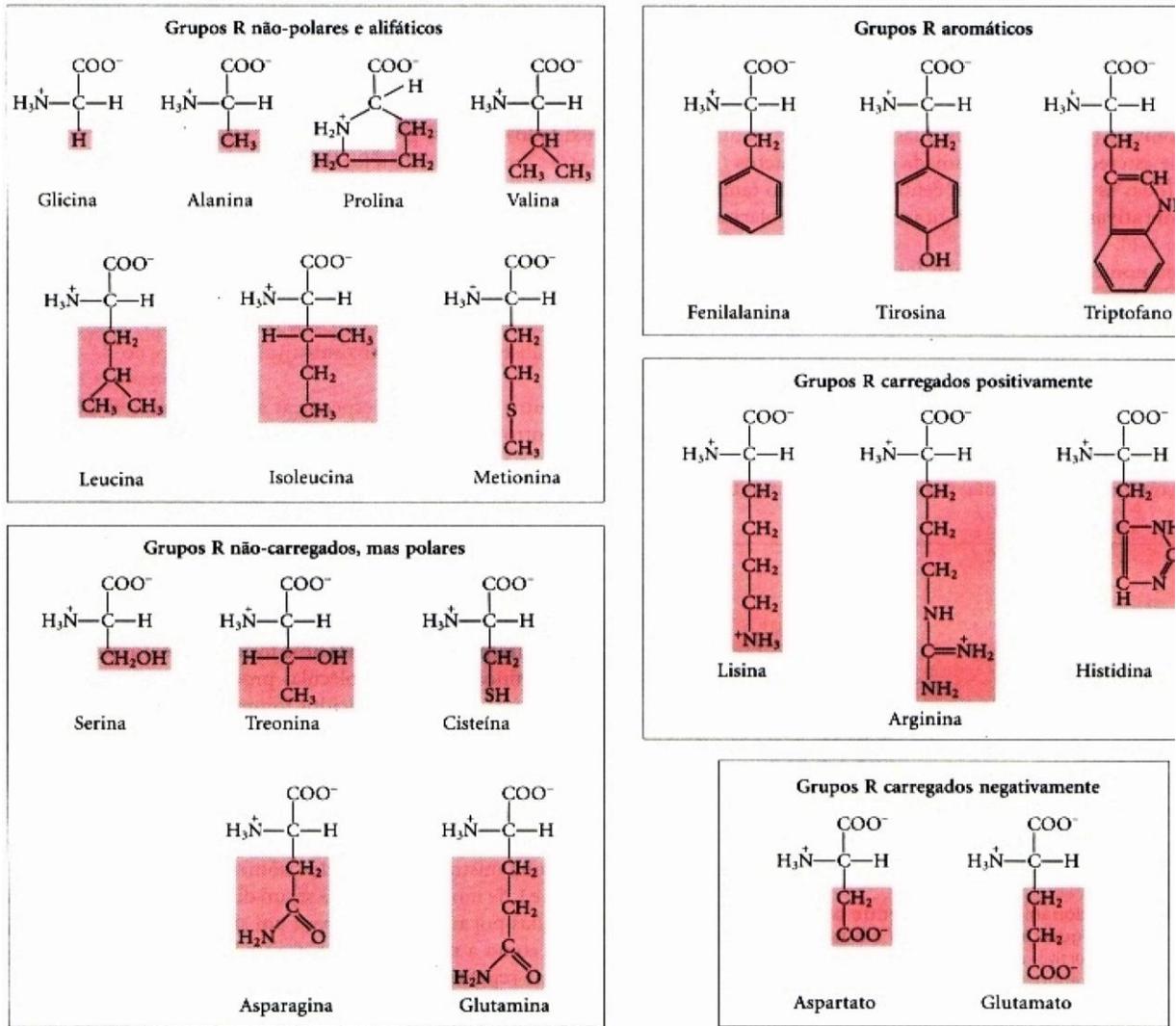


Fig. 1 – Estruturas dos 20 principais aminoácidos constituintes das proteínas (Adaptado NELSON & COX, 2007).

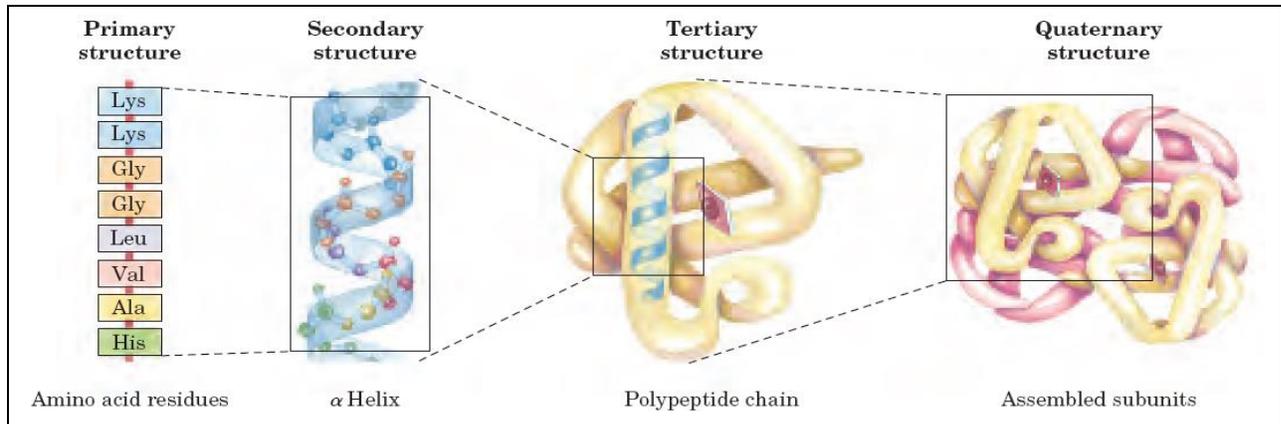


Fig. 2 – Etapas de enovelamento de uma proteína: estruturas primária, secundária, terciária e quaternária (Adaptado NELSON & COX, 2007).

A conformação da estrutura secundária é permitida através da liberdade rotacional entre as ligações entre os carbonos- $\alpha$  do eixo seqüencial de aminoácidos. Tomando como exemplo a estrutura do aminoácido glicina, as rotações livres são possíveis somente em três casos de ligações covalentes do polipeptídeo: N –  $C_{\alpha}$  (amino terminal – carbono alfa),  $C_{\alpha}$  – C (carbono alfa – carbono carbonílico) e, com uma liberdade rotacional bem menor, a ligação entre C – N (carbono carbonílico – amino terminal), segundo mostra a Fig.3 (adaptado NELSON & COX, 2007).

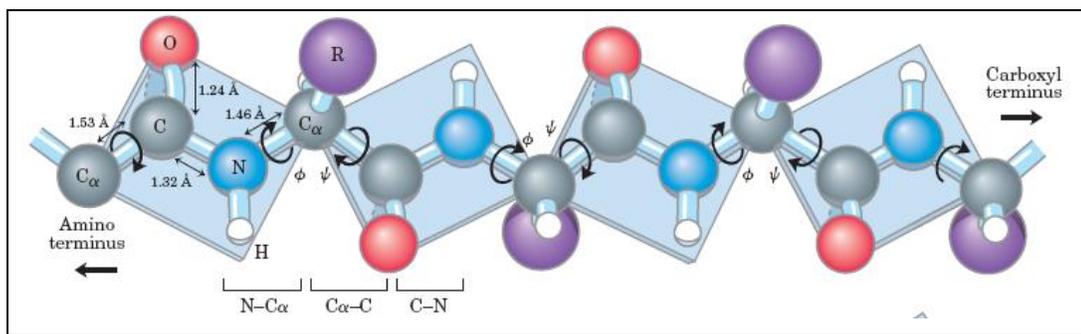


Fig. 3 – Possíveis rotações livres em ligações covalentes de um polipeptídeo (adaptado NELSON & COX, 2007).

As conformações de prolina são ainda mais restritas, devido à pouca liberdade de rotação da ligação N -  $C_{\alpha}$ . As permissões para os ângulos de rotação ( $\phi$  e  $\psi$ ) da ligação são determinadas em qualidade por um gráfico de Ramachandran (RAMACHANDRAN & SASISEKHARAN, 1968), demonstrando assim as estruturas espaciais de alfa-hélice ( $\alpha$ -hélice) e beta-folha ( $\beta$ -folha), (Fig.4). Já as regiões de *loop* são muito importantes para a estrutura tridimensional de um polipeptídeo, sendo ricas em resíduos de prolina e glicina, e configurando assim áreas de dobramento para a constituição da estrutura tridimensional. Essas regiões também podem ser ordenadas quando ocorre o acoplamento de um ligante, afetando também a funcionalidade biológica da proteína (SALI & BLUNDEL, 1993).

As proteínas constituem um grupo de compostos complexos, do ponto de vista da interação e posicionamento específico dos seus grupos químicos, num preciso arranjo tridimensional. Cada aminoácido apresenta uma maior probabilidade em assumir determinada posição para a formação de conformações estruturais terciárias. A cadeia principal de carbonos do polipeptídeo (*backbone*), contendo os grupamentos químicos, devem assumir uma

conformação que seja tanto biologicamente ativa, quanto estável fisicamente. Um desafio bastante pertinente é ainda o estado conformacional biologicamente ativo de um polipeptídeo, uma vez que estas macromoléculas são compostas de dezenas de milhares de átomos. O número de conformações que uma cadeia polipeptídica pode assumir é da ordem de  $10^{50}$ , com os dobramentos em pontos adequados, porém, para uma funcionalidade biológica eficaz da proteína madura, poucas são as conformações efetivas (LAZARIDIS, T. & KARPLUS, M., 1999).

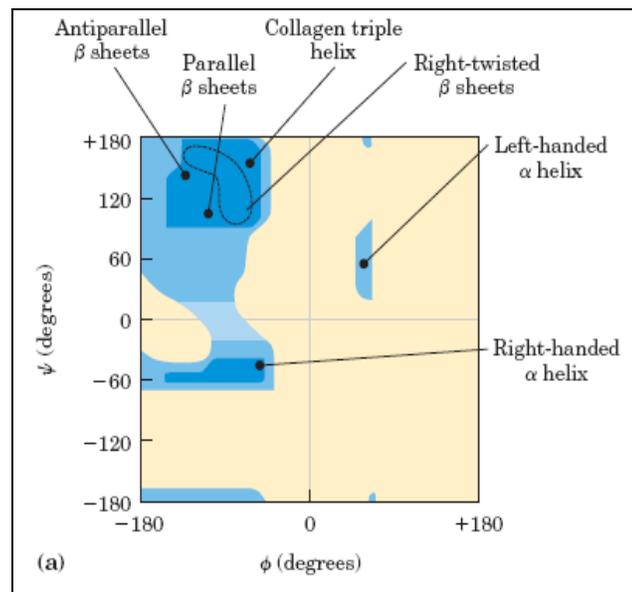


Fig. 4 – Gráfico de Ramachandran, para regiões permitidas de ângulos  $\Phi$  e  $\Psi$  (Adaptado NELSON & COX, 2007).

### **Arranjo alfa-hélice ( $\alpha$ -hélice)**

Segundo Ramachandran & Sasisekharan (1968), o *backbone* de um polipeptídeo alfa-hélice é torcido de modo uniforme, sobre cada carbono- $\alpha$ , com um ângulo  $\Phi$  de cerca de  $-57^\circ$  graus e de  $\Psi$  de aproximadamente  $-47^\circ$  graus (Fig.5). A vista de cima de um polipeptídeo em alfa-hélice se assemelha a um túnel.

Esta estrutura helicoidal é estabilizada por ligações de hidrogênio entre o grupo N-H e o grupo C=O de cada aminoácido. Desta forma, toda a cadeia principal do polipeptídeo apresenta os grupos N-H e C=O ligados, com exceção dos resíduos iniciais e terminais. A ligação de hidrogênio ocorre interligando o N-H ao C=O, evidenciada pela linha escura desenhada na

Fig.5a. Esta interação permite a formação de um longo eixo helicoidal, o qual apresenta 3,6 aminoácidos (em média) por cada dobramento das hélices. Apesar desta liberdade, os aminoácidos já possuem uma conformação bem estabelecida, e encontrando-se expostos para o exterior da hélice (Fig.5b).

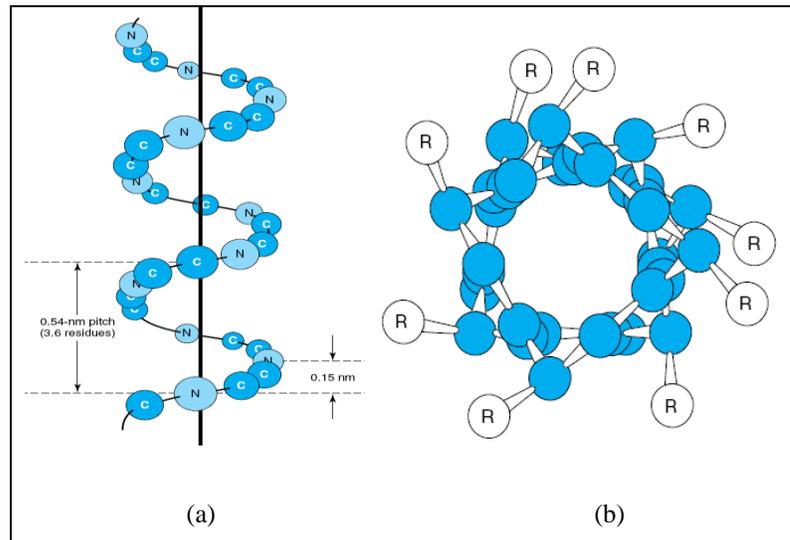


Fig. 5 – Estrutura de arranjo de resíduos de aminoácidos para a conformação alfa-hélice (a). Estrutura alfa-hélice vista de cima (b) (Adaptado KOOLMAN & ROEHM, 2005).

### *Arranjo beta-folha ( $\beta$ -folha)*

A estrutura secundária conhecida como beta-folha configura-se num arranjo onde os aminoácidos, quando vistos de ponta a ponta, formam um zigue-zague uniforme e em direções opostas (ou paralelas). Ao contrário do arranjo compacto da alfa-hélice, o peptídeo em arranjo beta-folha é bastante prolongado, onde sua maior estabilidade é devido às ligações de hidrogênio, estabelecidas entre o oxigênio da carbonila e o hidrogênio do grupo amino, pela região das ligações peptídicas da cadeia principal paralelo ou antiparalelo, segundo Ramachandran & Sasisekharan (1968).

Em ambas as conformações, observa-se um grande número de ligações de hidrogênio, de vínculos entre os segmentos ou dobras da “folha”. Com essas interações, a maior parte das estruturas em  $\beta$ -folhas não são perfeitamente planas, tendendo a uma leve torção, porém não tão

acentuada como numa estrutura  $\alpha$ -hélice (Fig.6). Pode ser evidenciado também que as cadeias laterais dispõem-se em lados opostos umas às outras.

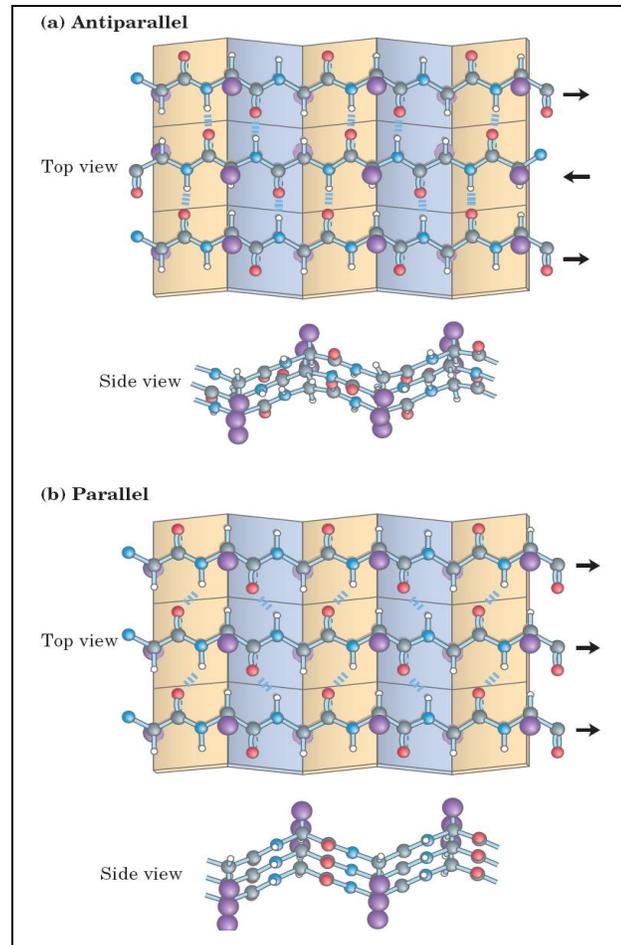


Fig. 6 – Demonstração de  $\beta$ -folhas paralelas e antiparalelas (Adaptado NELSON & COX, 2007).

### ***Loops (alças) e turns (curvas)***

Cerca de metade dos resíduos que compõem um arranjo para proteínas globulares são dispostos em alfa-hélices e beta-folhas; a outra metade está disposta em *loops* e *turns*, e outras estruturas conformacionais. *Loops* e *turns* referem-se a curtos segmentos de poucos aminoácidos, que servem para unir duas unidades de estrutura secundária adjacentes (LAZARIDIS & KARPLUS, 1999).

Já uma  $\beta$ -turn envolve quatro resíduos de aminoácidos, onde a ligação de hidrogênio se dá entre o primeiro e o quarto resíduo de aminoácido, resultando numa curva de 180 graus. Os resíduos mais frequentes nestas configurações de  $\beta$ -turn são a prolina e a glicina (Fig.7).

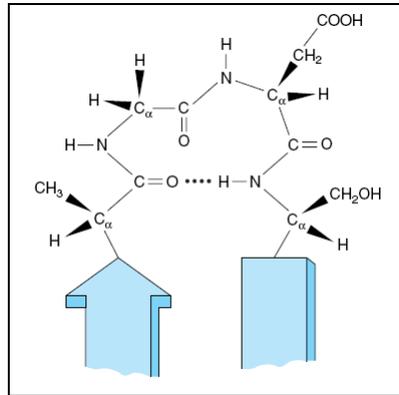


Fig. 7 – Região de  $\beta$ -turn para a união de duas estruturas secundárias de um polipeptídeo (Adaptado KHOLMANN & ROEHM, 2007).

Segundo Ginalski (2006), *loops* (alças) são regiões que contêm um número de resíduos de aminoácidos maior do que um  $\beta$ -turn para conectar regiões adjacentes de estrutura secundária. São bastante perceptíveis na análise conformacional porque apresentam uma disposição espacial de resíduos bastante irregular; no entanto, fazem o papel de chaves biológicas bastante específicas para cada proteína.

Ainda, Strocker & Bode (1995) elucida que em muitas enzimas, os *loops* são a ponte para os domínios responsáveis acoplarem-se aos substratos, que contêm geralmente resíduos de aminoácidos, que são participantes de reações de catálise importantes. Uma vez que muitos *loops* e *turns* residem na superfície de proteínas e são, portanto, expostos aos solventes, eles constituem locais de fácil acesso externo; ou ainda, podem configurar pontos de epítomos para o reconhecimento e ligação de anticorpos específicos.

Nem todas as regiões de proteínas são necessariamente ordenadas. Embora os *loops* sejam estruturalmente irregulares, eles apenas existem em uma determinada conformação estável, graças à interações do tipo ligação de hidrogênio, pontes salinas, e interações hidrofóbicas com

outras regiões da proteína. Essas desordens “anormais” entre os resíduos de aminoácidos podem conferir à proteína uma elevada flexibilidade conformacional. Em muitos casos, estas regiões desordenadas assumem uma certa conformação ordenada mediante o acoplamento de um ligante (XIANG, CSOTO & HONIG, 2002). Esta flexibilidade permite então que estas regiões atuem como reguladores de ligantes – controlando mudanças estruturais e conformacionais que afetam a estrutura tridimensional e funcionalidade biológica da própria proteína.

Geralmente, nestas regiões são encontrados os sítios ativos do peptídeo. As enzimas são muito específicas para os seus substratos. Esta especificidade se deve à existência, na superfície da enzima de um local denominado sítio de ligação do substrato. O sítio de ligação do substrato de uma enzima é um arranjo tridimensional de aminoácidos de uma determinada região da molécula, geralmente complementar à molécula do substrato, e ideal espacial e eletricamente para a ligação do mesmo.

A formação da estrutura terciária da proteína (*fold*ing) é o resultado das interações hidrofóbicas entre a estrutura secundária e o meio aquoso, resultando num empacotamento de estruturas similares em seus formatos, sendo importante para correlacionar a função biológica específica de uma determinada família de proteínas. Este problema recebeu um foco científico no início da segunda metade do século XX, sendo percebido na atualidade como um desafio interdisciplinar, onde são necessárias diferentes ferramentas colaborativas para a sua resolução (ANFENSEN, 1973).

Como possibilidades de análise do enovelamento de proteínas, é levado também em consideração parâmetros termodinâmicos e cinéticos: o enovelamento é um processo espontâneo, onde a proteína vai enovelar tendendo ao estado termodinamicamente mais estável (mais baixo nível de energia livre de Gibbs) e ao estado cineticamente não-aleatório. Existe então uma conformação tridimensional e uma energia livre associada a esta conformação, que seria a conformação mais estável para a proteína (ATKINS & PAULA, 2005).

### 1.3 Termodinâmica de enovelamento das proteínas

#### 1.3.1 Formação da estrutura terciária: enovelamento (*fold*ing)

A estrutura terciária da proteína é o resultado das interações hidrofóbicas entre a estrutura secundária e o meio aquoso, formando assim um empacotamento de estruturas similares em formato (LAZARIDIS, & KARPLUS, 1999). Este arranjo tridimensional é de fundamental importância, já que a conformação dos resíduos de aminoácidos na estrutura terciária da proteína, na maioria das moléculas, é coincidente com a estrutura nativa, ou seja, a estrutura que apresenta uma funcionalidade biológica específica, como pode ser visto na Fig.8.

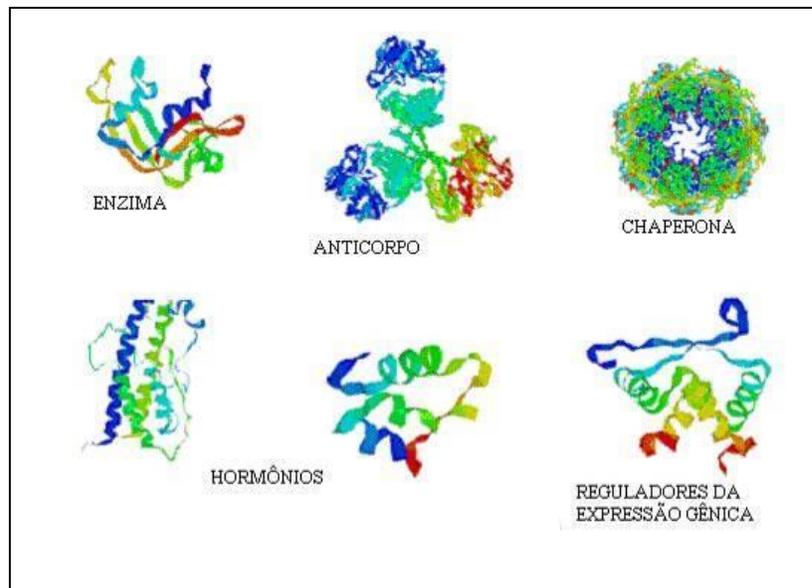


Fig. 8 – Estruturas terciárias coincidentes com a estrutura biologicamente ativa da proteína. (Adaptado MURRAY *et al*, 2003).

Assim, a estrutura nativa de uma proteína caracteriza a sua função biológica. A existência e a funcionalidade correta das proteínas levam à compreensão de como é a possibilidade de formação da estrutura nativa de uma proteína. Uma estrutura funcional significa que a proteína está em seu estado nativo e enovelada corretamente.

### 1.3.2 Hipótese termodinâmica para o enovelamento das proteínas

Christian Anfinsen (1973) pesquisou sobre questões fundamentais para o entendimento do processo de enovelamento das proteínas:

- Por que a proteína enovela para um estado conformacional de estrutura nativa?
- O processo de enovelamento é auxiliado por outros tipos de moléculas?
- Por que há apenas uma probabilidade de que a estrutura nativa seja ativa biologicamente?

O experimento de Anfinsen (1973), para estudar o processo de enovelamento, envolveu a utilização da RNAase A, a qual tem por função facilitar a interação entre o DNA e o RNA em células pancreáticas bovinas. No laboratório, Anfinsen simulou a desnaturação de uma proteína, a partir do desenovelamento induzido por uréia, fazendo com que esta perdesse a função biológica enzimática. Uma vez retirada a uréia do meio, a proteína retornou com sua atividade enzimática.

Segundo Van Holde, Johnson & Ho (1998), a termodinâmica define que um processo espontâneo é acompanhado pela libertação de energia livre, muitas vezes (mas nem sempre), sob a forma de calor, passando o sistema a ocupar um estado de menor energia, sendo considerado então mais estável, como num equilíbrio termodinâmico. Considerando então o *refolding* (renaturação) sendo um processo espontâneo, o estado nativo (proteína enovelada) é o estado termodinamicamente mais estável, isto é, de mais baixa energia. Esta asserção foi a chave para duas das proposições levantadas por Anfinsen (1973):

- O enovelamento acontece devido ao fato de este ser o estado termodinâmico mais estável; este processo de convergência a um estado nativo e funcional é espontâneo.
- Nenhuma outra proteína foi responsável por este processo de re-enovelamento; portanto, não existe a participação de uma molécula externa à proteína.

A partir desta hipótese termodinâmica, foi deduzido por Anfinsen que toda a informação necessária para encontrar o estado nativo está contida na seqüência de aminoácidos, a qual constitui a estrutura primária da proteína.

Em uma análise desta assertiva, como não houve nenhum outro composto químico adicional, supõe-se que o *refolding* da proteína foi executado espontaneamente. Sendo assim, apenas as interações totais entre os próprios aminoácidos da estrutura primária é o que determina a estrutura nativa da proteína. A segunda proposição de Anfinsen (1973) é respondida então pela hipótese termodinâmica: cada estado nativo da proteína é determinado por uma seqüência específica de aminoácidos.

Assim, os resultados destes experimentos de Anfinsen permitem a condução para uma definição sobre o *fold* das proteínas: é o processo espontâneo de autoconstrução da estrutura terciária de uma proteína, a partir do qual a cadeia polipeptídica encontra uma estrutura nativa tridimensional, em certas condições fisiológicas (interações físicas e químicas). Porém, outros parâmetros também são considerados para o *fold* das proteínas, além dos parâmetros termodinâmicos (ATKINS & PAULA, 2005).

### 1.3.3 Hipótese cinética para o enovelamento das proteínas

Em 1968, Cyrus Levinthal (físico que trabalhava em Biologia Molecular), demonstrou que a hipótese termodinâmica de Christian Anfinsen não completava totalmente a discussão sobre o problema do *fold* das proteínas: o estado nativo não pode ser criado de forma aleatória.

As três dimensões ( $x$ ,  $y$  e  $z$ ) são tratadas para a predição de enovelamento das proteínas. A quarta dimensão (o tempo) não é tratada no processo de enovelamento proposto no experimento de Anfinsen, de modo a levar a crer que a hipótese termodinâmica não explica a escala de tempo observada para o processo de *fold*. Uma vez que cada aminoácido pode estar, no mínimo, em dois estados possíveis (isto simplificando o número de conformações das cadeias laterais para 2 estados conformacionais), uma proteína com 100 aminoácidos pode chegar a ter um total de uma ordem de aproximadamente  $2^{100}$  a  $10^{300}$  conformações, representando as possíveis combinações entre os estados a que cada aminoácido tem acesso. Uma dessas combinações é a estrutura nativa, com um estado termodinâmico com energia mais baixa entre todas as conformações possíveis (SALI *et al*, 1994).

Outra observação, prevista pela hipótese cinética, é que a proteína passa de uma conformação a outra em um tempo mínimo de um picossegundo ( $1 \text{ ps} = 10^{-12} \text{ s}$ ). Este é o tempo necessário para uma vibração térmica, indicando uma mudança conformacional da proteína. Assim, seria necessário um tempo muito grande para a proteína explorar exaustivamente (de uma forma aleatória) todo o espaço conformacional, e encontrar alguma conformação correspondente ao estado nativo (XIANG *et al.*, 2002).

Porém, segundo Atkins & Paula (2005), é sabido que o *folding* de uma proteína com 100 aminoácidos não demora mais que alguns segundos, constituindo assim o chamado Paradoxo de Levinthal. O que então é proposto pela hipótese cinética é que antes de a proteína encontrar um estado nativo e com mínima energia, ela antes percorre etapas de enovelamento (*folding pathway*) bem definidas. Todo este processo é composto por estados conformacionais intermediários, onde no final se encontra o estado nativo da proteína, evitando assim a perda de tempo com explorações aleatórias do espaço conformacional.

Os estados intermediários, imaginados por Levinthal, são estados cuja estabilidade termodinâmica é tal que o intervalo de tempo em que a proteína aí permanece é suficientemente longo, para que assim possam ser detectados experimentalmente. No entanto, ao contrário do que acontece na hipótese termodinâmica, o estado nativo não corresponde necessariamente ao estado termodinâmico mais estável, mas sim, ao estado de energia mínima, que é o mais acessível do ponto de vista cinético (ATKINS & PAULA, 2005).

Então, com base na proposta de Levinthal, a investigação experimental em *folding* de proteínas foi, até ao início de 1990, dominada em grande parte pela procura de estados conformacionais intermediários, os quais deveriam ser suficientemente estáveis para que pudessem ser isolados e devidamente caracterizados como um grupo de estados nativos (ATKINS & PAULA, 2005).

### 1.3.4 Termodinâmica vs cinética

A perspectiva clássica do *folding* de proteínas é baseada nas contribuições da dicotomia termodinâmica – cinética, e na abordagem tradicional da Bioquímica, que considera cada molécula um sistema único. No entanto, nem uma nem outra tem a propriedade absoluta de resolverem o problema da conformação nativa das proteínas. Apenas a partir da década de 90 é que o fenômeno do *folding* passa a ser um pouco mais compreendido. Uma das contribuições mais importantes da Física para a compreensão do *folding* de proteínas foi precisamente a reconciliação das perspectivas de Anfinsen e Levinthal, numa teoria unificada que se baseia na natureza estatística do processo (ATKINS & PAULA, 2005).

### 1.3.5 A interação hidrofóbica

Sempre que é analisada a tendência que um evento ocorrer, deve se levar em consideração variação total da entropia do sistema e de seus arredores, e não apenas o sistema sozinho. O agrupamento em conjunto de grupos hidrofóbicos podem resultar em uma contribuição negativa para variação de entropia total do sistema, porque esta agregação corresponde a uma diminuição na desordem do mesmo. À primeira vista, portanto, não se espera que os grupos hidrofóbicos tendam à formação de um aglomerado único (*cluster*). No entanto, não podemos esquecer também o papel do solvente (GINALSKI, 2006).

Lazaridis & Karplus, (1999) afirmam que as moléculas apolares dissolvem-se pouco, ou quase nada, em solventes polares. Assim, fortes interações entre soluto e solvente não são possíveis. Como um resultado, é verificado que cada molécula de soluto é rodeada por uma gaiola de solvatação (Fig.9). Para compreender as conseqüências deste efeito, considera-se a termodinâmica da transferência de energia de um soluto, como em um hidrocarboneto apolar, de um solvente apolar (representação genérica de um meio interno, como o comportamento de uma proteína enovelada) para o meio aquoso polar (representação genérica para o entorno da biomolécula).

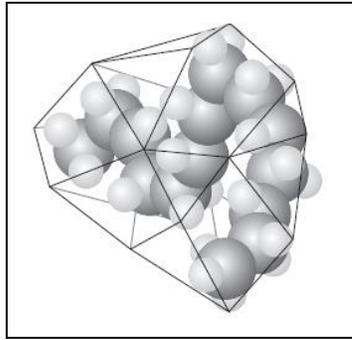


Fig. 9 – Esquema do efeito de solvatação em molécula apolar.

### 1.3.6 Entropia, entalpia e energia livre de Gibbs

Segundo Atkins & Paula (2005), ao se considerar sistemas biológicos, devem ser levados em conta dois tipos de mudança de entropia: a mudança no sistema biológico propriamente dito, e a mudança no meio no qual este sistema encontra-se inserido; assim então pode ser considerada a entropia total do sistema em questão. J. W. Gibbs (1839 – 1903) teorizou sobre como combinar estes dois aspectos em apenas um, combinação a qual hoje é bastante relevante para o desenvolvimento de metodologias de modelagem molecular para proteínas, que é o foco deste trabalho.

Sun (2004) afirma que se faz essencial considerar a entropia de ambos os sistemas: interno e externo à proteína, bem como a espontaneidade do processo. Para simplificar a discussão da contribuição da teoria da energia de Gibbs para sistemas biológicos, é considerada a variação total da entropia, onde a inserção desta energia é uma nova função de estado, para fins de estudos de bioenergética e estrutura biológica. A variação da entropia total, que acompanha o processo, é dada pela Eq.1:

$$\Delta S_{total} = \Delta S + \Delta S_{ambiente} \text{ (Eq.1)}$$

onde  $\Delta S$  é a mudança de entropia para o sistema; para uma variação espontânea,  $\Delta S_{\text{total}} > 0$ .

Berg, Tymoczko & Stryer (2002) indicam que experimentalmente, a variação na energia de Gibbs ( $\Delta G$ ), para o processo de transferência de uma substância apolar para um meio polar, é positiva ( $\Delta G_{\text{transf}} > 0$ ), como esperado em função do aumento da polaridade do solvente; porém, a variação da entalpia ( $\Delta H$ ) é negativa ( $\Delta H_{\text{transf}} < 0$ ). Portanto, existe uma grande diminuição da entropia ( $\Delta S$ ) do sistema ( $\Delta S_{\text{transf}} < 0$ ), o que representa uma variação positiva na energia de Gibbs. Abaixo, um esquema sobre as interações entre as variáveis  $\Delta G$ ,  $\Delta H$  e  $\Delta S$  na Tabela 1.

Tabela 1 – Esquema das influências entre variáveis de espontaneidade, energia e entropia.

<b>Espontaneidade (<math>\Delta G</math>)</b>	<b>Energia (<math>\Delta H</math>)</b>	<b>Entropia (<math>\Delta S</math>)</b>
<b><math>\Delta G &lt; 0</math></b> Enovelamento favorável	<b><math>\Delta H &lt; 0</math></b> Ligações quebradas ou formadas	<b><math>\Delta S &gt; 0</math></b> Nova conformação molecular
<b><math>\Delta G &gt; 0</math></b> Enovelamento desfavorável	<b><math>\Delta H &gt; 0</math></b> Energia compensada pelo meio	<b><math>\Delta S &lt; 0</math></b> Permanência de conformação

Segundo Monk (2004), é observado então que hidrocarbonetos acíclicos saturados tornam-se mais hidrofóbicos com o aumento de tamanho das cadeias carbônicas. Esta tendência pode ser racionalizada pela  $\Delta H$  se tornando mais positiva, e  $\Delta S$  ainda mais negativa, a partir do aumento do número de carbonos da cadeia lateral.

Desta forma, é verificada então a interdependência da entropia e da entalpia com a espontaneidade do enovelamento, considerando a equação que rege então a energia de Gibbs para sistemas em análise termodinâmica, descrita na Eq.2:

$$\Delta G = \Delta H - T\Delta S \text{ (Eq.2),}$$

onde  $\Delta G$  seria a variação da energia de Gibbs,  $\Delta H$  seria a variação de entalpia,  $T$  seria a temperatura considerada, e  $\Delta S$  seria a variação de entropia do sistema.

Monk (2004) ainda afirma que o significado físico de  $G$  é oriundo de sua própria definição, como sendo  $H - TS$ . Por sua vez, a entalpia é uma medida da energia, que pode ser obtida a partir do sistema, como parâmetro de calor. O termo  $TS$  é uma medida da quantidade de energia armazenada no movimento aleatório das moléculas, que compõem uma dada amostra.

Na realização de um trabalho, a energia é transferida de uma forma ordenada, por isso não podemos esperar obter trabalho a partir de uma energia armazenada aleatoriamente. A diferença entre o total da energia armazenada e a energia armazenada aleatoriamente,  $H - TS$ , está disponível para realizar trabalho, e é reconhecida então como variação da energia de Gibbs. Em outras palavras, a energia Gibbs é a energia disponível e armazenada para o movimento uniforme e o arranjo das moléculas do sistema, sob pressão e temperatura constantes (MONK, 2004).

### 1.3.7 Espontaneidade e energia de Gibbs

A diferença de sinal entre  $\Delta G$  e  $\Delta S_{\text{total}}$  implica como uma condição para a ocorrência de um processo espontâneo:  $\Delta S_{\text{total}} > 0$ , e  $\Delta G < 0$ , (considerando pressão e temperatura constantes). Ou seja, em uma mudança espontânea, sob constantes temperatura e pressão constantes, a energia de Gibbs  $\Delta G$  decresce, conforme demonstrado na Fig.10.

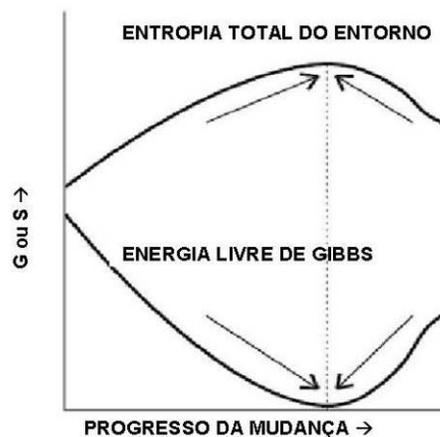


Fig. 10 – Relação das variações da entropia total do entorno ( $\Delta S$ ) e da energia de Gibbs ( $\Delta G$ ).

Naturalmente, é possível achar que um sistema tende a decrescer para uma baixa energia livre total ( $\Delta G_{\text{total}}$ ). No entanto, não deve ser desconsiderado que um sistema total (sistema + meio ambiente), tendendo a energias baixas, também incrementa a entropia total. Assim, o único critério espontâneo de mudança de estado é a entropia total do sistema e seu entorno; a energia livre de Gibbs meramente consiste num meio de expressar a variação total em termos de propriedades de um sistema isolado, e é válida apenas para processos que ocorrem em temperatura e pressão constantes. Para o enovelamento de uma proteína, esta sempre tende à conformação estrutural de menor energia, ou seja, com maior informação biológica disponível para a sua completa funcionalidade ativa.

#### **1.4 Métodos de predição da estrutura tridimensional**

Tsai (2002) afirma que para a predição da estrutura tridimensional de uma proteína, podem ser utilizados métodos *in vitro* (cristalografia de raios-X e espectroscopia de ressonância magnética nuclear) e *in silico* (homologia estrutural, *threading* e dinâmica molecular). Os dois primeiros métodos são realizados em trabalhos de bancada, uma vez que apenas utilizam metodologias experimentais em bancada.

A utilização de ferramentas computacionais, como a Bioinformática, para a análise estatística da similaridade entre seqüências protéicas poupa tempo e dinheiro, já que se faz possível prever substâncias químicas que interajam com um grupo de proteínas similares. A determinação dos domínios funcionais em comum, e sua conseqüente ocorrência em determinada parte da cadeia, pode acelerar o processo de design de novos fármacos e, concomitantemente, sua validação experimental (OLIVEIRA *et al*, 2006).

Um desses casos de proteínas sem um modelo de sua estrutura tridimensional é a metaloprotease VMBMP\_BOTMO (P83512), encontrada na toxina da serpente *Bothrops moojeni*, objeto de estudo deste trabalho.

As metaloproteases, um grupo de proteínas com uma funcionalidade já definida como hemorrágica, embasam a possibilidade de que as proteínas, enquanto homólogas, tendem a

assumir conformações semelhantes. O método de homologia estrutural é realizado computacionalmente, onde a estrutura da proteína alvo é modelada por meio da comparação de sua seqüência com outras seqüências, presentes em banco de dados, que já possuem uma estrutura tridimensional definida por um dos dois primeiros métodos experimentais. A modelagem molecular é uma valiosa ferramenta para a elucidação estrutural de biomoléculas, já que permite de modo rápido e com considerável eficácia, a sugestão de um modelo tridimensional para a proteína que ainda não possui nenhuma predição deste tipo.

A modelagem molecular por homologia, por sua vez, pode ser útil quando as estruturas de proteínas, baseadas em elucidação por cristalografia de raios-X ou por ressonância magnética nuclear (RMN) não estão ainda disponíveis nos bancos de dados (SANTOS FILHO & ALENCASTRO, 2003). Figueiredo *et al* (2003) afirmam que o nível alto da conservação dos domínios protéicos pode ainda ser modelado, segundo uma rotina de algoritmo computacional, de tal forma a observar de modo qualitativo e quantitativo a disposição dessas seqüências ao longo da cadeia protéica, e a conseqüente propriedade de caracterização de determinadas famílias de proteínas em função do domínio protéico.

## **1.5 Aspectos da modelagem molecular por homologia**

A modelagem molecular é uma alternativa bastante eficaz para a investigação de estrutura de biomoléculas, particularmente as proteínas. A estrutura tridimensional de uma proteína pode ser prevista por dois métodos teóricos: método *ab initio* e método com base em homologia. O método por homologia estrutural vem sendo mais aplicado atualmente, devido aos seus resultados serem mais próximos da realidade conformacional (RAMOS, 2007).

### **1.5.1 Breve histórico da modelagem molecular**

As técnicas de modelagem comparativa tornaram-se importantes ferramentas computacionais apenas depois de 1990, quando centenas de estruturas de proteínas já tinham sido depositadas no PDB – Protein Data Bank (<http://www.pdb.org>). Os primeiros estudos de

modelagem datam do final da década de 60, e no início da década de 70 eram realizados com modelos feitos à mão, com artefatos de arame e plástico, e somente mais tarde puderam ser implementados com softwares computacionais (COX & BONANOU, 1969; TOMETSKO, 1970).

Os primeiros modelos por análise comparativa foram construídos simplesmente através de cópias de coordenadas pré-existentes a partir de uma proteína homóloga, e os resíduos não idênticos foram então substituídos em uma nova montagem das cadeias laterais correspondentes (BROWNE *et al*, 1969). Esta abordagem, chamada montagem de corpos-rígidos, é ainda amplamente empregada em estudos atuais, e com considerável sucesso, especialmente quando as proteínas apresentam uma identidade entre suas seqüências acima de 40% (GREER, 1981).

O método de modelagem comparativa foi pioneiro em dois estudos, realizados por Browne *et al* (1969) e Greer (1981). Browne *et al* (1969) publicou informações sobre características estruturais do primeiro modelo obtido por modelagem comparativa, usando uma estrutura obtida por cristalografia de raios X como um molde. Eles modelaram a  $\alpha$ -lactalbumina bovina com uma estrutura tridimensional de lisozima de clara de ovo de galinha, onde as seqüências protéicas tinham 39% de identidade entre si. Esta predição foi comprovada como correta tempos depois, quando a estrutura da  $\alpha$ -lactalbumina bovina foi solucionada (ACHARYA *et al*, 1989).

McLachlan & Shotton (1971) realizaram a modelagem da  $\alpha$ -protease-lítica de Myxobacteria 495, baseada em estruturas de quimotripsina e elastase, onde a identidade entre essas duas seqüências protéicas era de apenas 18%, e o alinhamento foi seccionado em *gaps* freqüentes. Quando a estrutura da  $\alpha$ -protease-lítica foi elucidada e publicada (BRAYER *et al*, 1979), tornou-se claro que o desalinhamento de trechos das seqüências levaram a regiões incorretas da estrutura 3D, porém a porção de ambos os domínios foram construídos corretamente (DELBAERE *et al*, 1979). Com este modelo, foi demonstrada a dificuldade em alinhar seqüências que apresentem uma similaridade limitada, e uma modelagem variável, principalmente em regiões de *loops*.

Greer (1981) realizou os primeiros trabalhos para demonstrar a importância da modelagem de regiões variáveis. Abstraindo conformações aproximadas, a partir de uma família

de proteínas homólogas de estruturas conhecidas, ele pôde distinguir as regiões conservadas estruturalmente, as quais continham alto índice de homologia, e regiões variáveis estruturalmente, as quais incluíam todas as inserções e deleções. Aplicando a distinção estrutural em novas seqüências, os alinhamentos errôneos entre as seqüências são altamente minimizados. Para cada novo alinhamento de seqüências, as regiões conservadas estruturalmente podem ser reconstruídas a partir de regiões de estruturas conhecidas. A construção de regiões variáveis, contudo, não é um processo simples. As conformações de *loops* de apenas um ou dois resíduos de aminoácidos podem ser extrapoladas a partir de uma das estruturas homólogas mais próximas à estrutura alvo, que se deseja modelar estruturalmente. Esta abordagem foi ainda aplicada para a predição de estruturas de serinoproteases de mamíferos, baseada em um número de proteínas desta família, incluindo uma variedade de soro sanguíneo, proteínas intestinais e pancreáticas, bem como uma enzima bacteriana intimamente relacionada (GREER, 1981).

Em vez de derivar da cadeia principal de uma proteína, a partir de somente uma de suas seqüências homólogas, Taylor (1986) desenvolveu um método de geração de seqüências molde para cada parte da proteína a ser modelada, baseado em padrões de conservação que eram observados em estruturas 3D conhecidas, pertencentes a uma família de proteínas. Os moldes conservados foram originados a partir de um pequeno número de seqüências de estruturas terciárias conhecidas. Os moldes foram então alinhados com outras seqüências de estruturas tridimensionais desconhecidas. Assim, a especificidade das seqüências moldes foi demonstrada pela habilidade de identificar as configurações conservadas em imunoglobulinas conhecidas e em seqüências relacionadas, mas não em outras seqüências. Desta forma, realizando montagens destes padrões conservados, dentro de uma estrutura protéica, requer a aplicação de um campo de força e de uma amostragem de conformação adequada, de modo a refletir uma função biológica.

Devido ao pequeno número de estruturas protéicas disponíveis antes de 1990, a técnica de modelagem comparativa teve seus estudos iniciados no meio da década de 90, com o progresso de projetos em genômica e o crescimento do número de estruturas protéicas resolvidas e depositadas no banco de dados PDB. Com o advento da genômica estrutural, a importância da modelagem comparativa continua em crescimento constante (XIANG, 2002). Apesar de quase 30 anos após o pioneirismo dos trabalhos de Greer, a técnica básica permanece praticamente seguindo o mesmo princípio, sendo utilizada na maioria dos softwares para modelagem

comparativa, através da comparação da seqüência-alvo com uma seqüência-molde, de estrutura 3D já elucidada. Esforços recentes em modelagem comparativa tem sido concentrados na descoberta de seqüências homólogas distantes, na melhoria da acurácia de alinhamentos, e especialmente no refinamento de modelos, através de otimização de funções de energia.

### **1.5.2 Predição por modelagem comparativa**

A metodologia de predição por modelagem comparativa é baseada na evidência de que proteínas com seqüências de aminoácidos idênticos (ou com aminoácidos com propriedades similares) tendem a adquirirem conformações estruturais muito semelhantes. Assim, essas proteínas acabam formando um grupo de biomoléculas que apresentam características funcionais muito parecidas, até mesmo realizando as mesmas funções biológicas dentro de uma reação em cadeia. Assim, é possível transpor e comparar as informações, oriundas de uma proteína com seqüência e estrutura tridimensional conhecidas, com uma seqüência similar e com estrutura 3D desconhecida. A metodologia de modelagem molecular sugerida neste trabalho foi adaptada de Ramos (2007).

### **1.5.3 Banco de dados: comparação de seqüências de aminoácidos**

A busca baseada em seqüências de aminoácidos é uma das operações fundamentais da Bioinformática, servindo como base para outras operações mais complexas. Com o resultados dessa busca, pode se realizar uma análise de quais seqüências são semelhantes e quais são diferentes. Como aplicação, pode-se determinar um ancestral comum a partir de um conjunto de seqüências, partindo do princípio de que seqüências homólogas podem existir em duas espécies distintas, porém com ancestralidade em comum (TICONA, 2008).

Os bancos de dados são importantes fontes de informação, pois neles estão depositados os principais dados sobre seqüências protéicas. Xiong (2006) define um banco de dados como um grande arquivo computadorizado e aplicável para guardar e organizar dados, de forma que estas

informações podem ser resgatadas facilmente via uma variedade de critérios de busca. Neste trabalho, foram utilizados os banco de dados UniProt e PDB, os quais serão descritos abaixo com mais detalhes.

#### 1.5.4 Similaridade entre seqüências

Segundo Claverie & Notredame (2007), seqüências similares freqüentemente derivam da mesma seqüência ancestral. Isto significa afirmar que se a seqüência de aminoácidos de metaloprotease, encontradas em um grupo de um mesmo gênero, são similares, provavelmente elas possuem o mesmo ancestral, compartilham de uma estrutura muito parecida e possuem funções biológicas similares, senão iguais. Este princípio também pode ser aplicável em seqüências protéicas de diferentes gêneros de organismos.

Proteínas ou genes muito similares são classificadas como homólogas, evidenciando assim que possuem funções, estruturas e ancestral em comum. O problema mais interessante a ser solucionado é quanto ao grau de similaridade entre elas. Quando duas proteínas apresentam mais de 25% de identidade entre suas seqüências de aminoácidos, pode ser afirmado que há uma certa similaridade entre as duas. Com um *score* de similaridade total menor que 25%, é colocada em dúvida a ancestralidade comum entre essas seqüências (SANTOS FILHO & ALENCASTRO, 2003).

No entanto, algumas proteínas com apenas 15% de similaridade podem apresentar praticamente a mesma estrutura tridimensional, enquanto algumas proteínas com 20% de identidade são completamente diferentes estruturalmente em 3D. Sendo assim, a homologia não pode ser garantida apenas pela igualdade de aminoácidos em uma determinada seqüência, mas sim pela sua disposição estrutural, conseqüentemente em função biológica (SANTOS FILHO & ALENCASTRO, 2003).

Um outro parâmetro que pode ajudar na identificação de seqüências homólogas é o chamado *e-value*. Este parâmetro que expressa o grau de similaridade entre as seqüências, em função da contra-prova de seqüências obtidas em um banco de dados (XIONG, 2006).

O *e-value* (ou “valor esperado”) é um valor de significância estatística para evidenciar os alinhamentos da seqüência-alvo contra o banco de dados de proteínas. Se o valor da significância estatística atribuído a um alinhamento for maior que o *e-value*, o alinhamento é invalidado. Se este *e-value* for baixo, significa que são poucas as chances de os alinhamentos serem estabelecidos por acaso, e a possibilidade de alinhamentos corretos é bem maior.

São importantes parâmetros a serem também considerados o número de resíduos do aminoácidos e a massa molecular da biomolécula, os padrões de resíduos de aminoácidos conservados e o número de inserções ou deleções na seqüência de aminoácidos.

### **1.5.5 Ferramentas para modelagem da estrutura 3D por homologia estrutural comparativa**

O aplicativo MODELLER é um dos sistemas computacionais mais bem aceitos na atualidade para a execução de inferências simples de estruturas tridimensionais, através da predição por homologia, para aquelas estruturas que possuam uma identidade acima de 75%. O ambiente, que está disponível em <http://www.salilab.org/modeller/>, também se propõe a realizar a predição da estrutura secundária da biomolécula. Por se tratar de um aplicativo livre, pode ser utilizado em qualquer computador, desde que haja uma conexão com a Internet. Não há necessidade de instalação de nenhum pacote ou aplicativo adicional. Após o processo de instalação do aplicativo MODELLER, ele está apto para a aplicação em modelagem de biomoléculas.

Como parâmetro complementar, é utilizado também o servidor SWISSMODEL (SWEDE *et al*, 2006), para previsão da estrutura secundária, disponível em <http://swissmodel.expasy.org/SWISS-MODEL.html>. O SWISSMODEL é capaz de retornar as localizações das estruturas alfa-hélice, beta-folha e *loops*, a partir da seqüência proposta.

### 1.5.6 Ferramentas para validação de modelos

Ao fim da predição de um modelo tridimensional de uma proteína, através da modelagem por homologia, se faz necessária uma validação deste modelo em função de alguns parâmetros físico-químicos. A validação de uma estrutura é uma etapa muito importante da construção de um modelo molecular, podendo ser executada em diferentes níveis de organização estrutural da biomolécula. A qualidade do empacotamento final da proteína deve ser avaliada como um todo, bem como os possíveis erros estruturais localizados.

Após a modelagem, o arquivo de saída gerado possui extensão “.pdb”. Este arquivo comporta as coordenadas espaciais dos átomos constituintes da proteína, sugeridos como o modelo tridimensional. Este arquivo servirá como arquivo de entrada dos programas e servidores web que farão a validação do modelo.

A propriedade que de forma mais taxativa determina a qualidade das estruturas tridimensionais de proteínas geradas por homologia consiste justamente na similaridade e qualidade dos modelos estruturais obtidos de dados experimentais; uma preocupação antiga dos pesquisadores em modelagem molecular de proteínas. As informações de enovelamento acerca da cadeia principal dos aminoácidos são fundamentais na validação de estruturas-modelo, que são preditas com ferramentas computacionais, principalmente quanto ao comportamento energético destas estruturas (SIPPL, 1993).

A qualidade da estrutura construída depende muito do modelo 3D usado como molde, sendo,então, requisitos mínimos que este modelo tenha sido obtido experimentalmente com dados de boa/alta resolução ( $< 2.0$  Angstroms) e fator-R satisfatório ( $< 20\%$ ), no caso onde os modelos tenham sido obtidos por Cristalografia de Proteínas. O parâmetro fator-R é originado experimentalmente em processos de cristalografia a partir de uma função de correlação entre os resultados experimentais e teóricos. Estes últimos, oriundos do modelo tridimensional construído por interpretação experimental. Quanto maior o fator-R, menor a qualidade do modelo gerado. Já a resolução está relacionada com a quantidade de informação experimental foi obtida para a construção do modelo 3D. Assim, quanto maior a resolução, maior o detalhamento da informação

estrutural (TEIXEIRA, 2005). Contudo, dependendo do grau de similaridade entre as seqüências alvo e molde, são toleráveis valores mais baixos para a resolução (JOHNSON *et al*, 1994).

Em se tratando de requisitos estruturais, o modelo deve apresentar uma estrutura terciária satisfatória (LASKOWSKI *et al*, 1998). O alinhamento entre as seqüências alvo e molde também é um aspecto a ser considerado, já que a qualidade da modelagem da seqüência alvo também depende de um molde com uma identidade suficiente àquela proteína que será modelada. As diferenças conformacionais entre as estruturas secundárias (regiões conservadas) das estruturas-molde e -alvo também são pontos determinantes para a qualidade do modelo formado (SANTOS FILHO, 2000).

Logo abaixo, apresentamos uma breve descrição dos principais programas e servidores utilizados para a análise e validação das estruturas tridimensionais de proteínas modeladas: Molprobability (DAVIS *et al*, 2007), Harmony (PUGALENTHI *et al*, 2006), ProSAWeb (WIEDERSTEIN & SIPPL, 2007) e Verify3D (LUTHY *et al*, 1992).

### ***Molprobability***

O modelo produzido deve apresentar também uma qualidade estereoquímica o mais compatível possível, de modo a reproduzir mais . Para tal avaliação de parâmetros estereoquímicos, o servidor utilizado é o Molprobability (DAVIS *et al*, 2007). O servidor encontra-se disponível para acesso no endereço eletrônico <http://molprobability.biochem.duke.edu/index.php>. Neste servidor, são avaliados os comprimentos de ligação química, os ângulos planos, a planaridade dos anéis 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 & SASISEKHARAN, 1968), avaliando assim a qualidade geométrica e estereoquímica do modelo estrutural.

A análise do gráfico de Ramachandran é um recurso bastante útil na modelagem molecular, pois ela possibilita verificar quais os resíduos de aminoácidos que se encontram nas orientações energeticamente mais favoráveis e desfavoráveis, orientando a avaliação da qualidade

de modelos teóricos ou experimentais de proteínas. Na tabela 2, são apresentados os parâmetros de qualidade considerados pelo servidor Molprobit.

Tabela 2 – Interpretação dos resultados para a avaliação do Molprobit.

Parâmetros	IDEAL
Rotâmeros inadequados	<1%
Ramachandran favorável	>98%
desvios de $C_{\beta} > 0.25 \text{ \AA}$	0
Resíduos com ligações inadequadas	<1%
Resíduos com ângulos inadequados	<0.5%

### **HARMONY**

A avaliação estereoquímica da estrutura da proteína examina os parâmetros internos, como comprimento das ligações e ângulos  $\Phi$  e  $\Psi$ . O servidor Harmony <http://caps.ncbs.res.in/harmony/> (PUGALENTHI *et al*, 2006) analisa a compatibilidade entre as estruturas primária e terciária de uma proteína, atribuindo padrões de troca aos resíduos de aminoácidos após considerar seus ambientes locais. Esses ambientes locais são descritos pela conformação da cadeia principal, como também pela acessibilidade do solvente à superfície da estrutura da proteína, e pelos padrões de ligação de hidrogênio (interações).

O servidor permite o envio do modelo em formato “.pdb”, realizando inclusive, quando necessário, o alinhamento de seqüências homólogas. As pontuações são mapeadas sobre a estrutura para posterior exame, o que é útil para reconhecer também as regiões com possíveis erros locais de modelagem molecular por homologia.

### ***ProSA- web (Protein Structure Analysis)***

O programa ProSA web (WIEDERSTEIN & SIPPL, 2007) pode ser utilizado para analisar as condições dos resíduos de aminoácidos na proteína modelada. Cada resíduo de aminoácido é avaliado, tendo como referência um ambiente já esperado, quando comparado aos aminoácidos presentes em proteínas análogas e com alta resolução. A ferramenta pode ser encontrada em <https://prosa.services.came.sbg.ac.at/prosa.php>.

Os cálculos são realizados a partir das posições dos carbonos alfa na cadeia principal da proteína, podendo também ser aplicável a proteínas modeladas a partir de seqüências-molde obtidas com baixa resolução. Com a análise das coordenadas, a energia da estrutura pode ser avaliada, utilizando uma distância de potencial entre os carbonos-alfa (SIPPL, 1990), o qual capta a exposição dos resíduos da proteína ao solvente.

Assim, duas características da estrutura a ser validada são mostradas: *z-score* e uma parcela das energias dos resíduos de aminoácidos. Após a análise do modelo da proteína pelo software, a escala de qualidade *z-score* vai de 1 a 10. Quanto mais o valor de *z-score* se aproximar de 10, melhor será a qualidade do modelo. Tal qualidade está associada à medida do desvio da energia total da estrutura da proteína, no que diz respeito à distribuição de energia derivada de conformações aleatórias. Quando estes *z-scores* encontram-se muito baixos, é um forte indicativo de que a proteína não se encontra ainda num estado nativo.

Segundo Sippl (2006), para facilitar a interpretação do *z-score*, o arquivo de saída é apresentado sob a forma de um gráfico, que contém os mesmos parâmetros para estrutura de proteínas que já foram elucidadas por outros métodos (Cristalografia ou Ressonância Magnética Nuclear), apenas como um comparativo. Essa comparação permite verificar se os parâmetros de *z-score* da proteína em questão se encontram dentro do intervalo de pontuações observadas em proteínas com massas moleculares semelhantes.

O gráfico de energia do modelo mostra a qualidade local pela análise dos pontos de energias em função da posição do aminoácido na seqüência. Em geral, valores positivos correspondem a regiões com algum problema potencial ou trechos modelados de modo errôneo. Algumas flutuações podem ser percebidas em resíduos de aminoácidos isolados, sendo

considerados desprezíveis se não forem em grande quantidade. Desta forma, calcula-se ainda uma média de *z-score* a cada 40 resíduos de aminoácidos (SIPPL, 2006).

### ***Verify 3D***

O modelo-alvo gerado também deve ser avaliado em relação à qualidade do ambiente químico a que ele pode ser submetido, objetivando a determinação do nível de confiabilidade do enovelamento protéico. O software Verify3D ([http://nihserver.mbi.ucla.edu/Verify\\_3D](http://nihserver.mbi.ucla.edu/Verify_3D), LUTHY *et al*, 1992) é uma sugestão para este tipo de análise, onde é possível se determinar os ambientes químicos de cada resíduo do modelo, atribuindo “*scores*” de referência a uma matriz, construída a partir de uma análise estatística e que envolve estruturas de proteínas do banco de dados *Protein Data Bank* (PDB). Nessa matriz estão contidas três coordenadas de posição que cada resíduo apresenta, dentro de cada um dos 18 ambientes químicos previamente definidos.

O gráfico de saída do software demonstra então a compatibilidade de um modelo atômico (3D) com sua própria seqüência de aminoácidos (1D). Cada resíduo é caracterizado com base na classe estrutural, sua localização na cadeia principal e sua região (alfa, beta, *loop*, polar, apolar etc.). A partir daí, é atribuído um *score* para cada resíduo, com base em um banco de dados de estruturas de aminoácidos. A cada 20 resíduos temos uma nova análise de aminoácidos, onde é desejável que o *score* de cada aminoácido deve ficar acima de 0,2 (20%), e considerável que nunca fique abaixo de 0% (LUTHY *et al*, 1992).

Esta ferramenta é bastante útil também na validação das regiões de *loop* do modelo-alvo gerado. A soma dos *scores* obtidos é armazenada em uma média final, onde é comparada a um valor ideal esperado para uma proteína de igual comprimento. A média de adequabilidade da correspondência está por volta de 45%, que é média encontrada considerando-se estruturas de boa qualidade.

## **1.6 Gênero *Bothrops***

### **1.6.1 Ocorrência**

O gênero *Bothrops* ocorre na América do Sul. Há mais de 20 espécies de cobras do gênero *Bothrops* no Brasil. As cobras desse gênero habitam do Sul da Bahia até o Rio Grande do Sul, leste de Mato Grosso e todo o Nordeste. Em geral elas habitam regiões de mata fechada, como a Região Amazônica, mas principalmente campos cultivados onde ficam os roedores que são sua fonte de alimentação. O veneno botrópico é uma fonte de substâncias farmacologicamente ativas que causam intensos efeitos locais e sistêmicos (FRANÇA *et al*, 2006).

### **1.6.2 Domínios e metaloprotease**

Segundo Teixeira (2005), os domínios são segmentos de uma cadeia polipeptídica que apresenta uma conformação tridimensional independente de qualquer outro segmento da cadeia, e que acarreta uma dada função biológica para a proteína. Por sua vez, uma família de proteínas representa um grupo de proteínas que apresentam uma similaridade estrutural, mas não necessariamente uma similaridade de seqüência de resíduos de aminoácidos.

Segundo Kamiguti (*et al*, 1998), os venenos de cobras viperídicas e crotálicas são fontes ricas em componentes que podem afetar a hemostasia, por causa da mudança na função plaquetária e de coagulação do sangue. Os componentes do veneno que afetam o sistema de coagulação incluem enzimas trombolíticas, as quais convertem fibrinogênio em fibrina e ativadores de protrombina. A função plaquetária pode ser afetada pelos componentes do veneno, tal como as metaloproteases hemorrágicas, as fosfolipases e as desintegrinas. Como resultado, é comum encontrar um consumo de fatores de coagulação e incoagulabilidade, acompanhado de hemorragias em vítimas de mordida de cobras (Bjarnason e Fox, 1994).

As metaloproteases são proteínas que contém íons de zinco como um dos ligantes, variando em massa molecular de 20 a 100 kDa, algumas das quais são capazes de induzir

rapidamente sangramento local e/ou sistêmico. As metaloproteases de diferentes massas moleculares estão frequentemente presentes no mesmo veneno.

De acordo com Bjarnason & Fox (1994), as metaloproteases de venenos de serpentes (SVMPs) compreendem quatro classes, definidas de acordo com a constituição dos domínios presentes nas estruturas moleculares: P-I, constituída de enzimas contendo apenas o domínio de metaloprotease; P-II, classe a qual estão inclusas as enzimas apresentando o domínio de metaloprotease, seguida de um domínio de desintegrina ligado (*Dis-like*); P-III, enzimas que possuem um domínio de regiões ricas em cisteínas (*Cys-rich*) em adição aos domínios de metaloprotease e de desintegrina; e P-IV, compreendida pelas enzimas com duas subunidades, uma constituída de três domínios característicos de enzimas do grupo P-III e outra unidade sendo uma C-lectina, ligada através de pontes dissulfeto à primeira unidade.

Nem todas as SVMPs induzem a hemorragia segundo dados experimentais em camundongos. Em termos gerais, SVMP do grupo P-III são toxinas hemorrágicas mais potentes do que as metaloproteases de outros grupos, particularmente do grupo P-I, onde o número de enzimas não-hemorrágicas tem sido caracterizado mais amplamente (GUTIERREZ *et al*, 2005). No entanto, existem SVMPs do grupo P-III que não apresentam uma atividade hemorrágica, porém oferecem potentes efeitos coagulantes, como a classe de ativadores de protrombina A (SILVA *et al*, 2003). A chave para toda esta ação hemorrágica parece ser o mecanismo de hidrólise dos componentes da membrana basal de vasos sanguíneos (OHSAKA, 1979).

### 1.6.3 Mecanismo de catálise das metaloproteases

Segundo Stocker & Bode (1995), o mecanismo de catálise das metaloproteases se dá por meio da quebra das ligações peptídicas do substrato a reagir com a metaloenzima. O zinco catalítico, presente no sítio ativo da metaloprotease, é coordenado por três histidinas, cujos átomos de nitrogênio estão distantes entre si em 2,0Å. Na posição do resíduo Glu, é encontrada uma molécula de água (compondo o *motif* HEXXH), bem próximo ao cátion zinco, funcionando como um quarto ligante (RAMOS, 2007). Esses ligantes interagem com o zinco, sob a forma

trigonal piramidal, em conjunto com o nitrogênio da segunda His ligante (localizada no topo da pirâmide) e com o zinco (que é encontrado coplanar à base da pirâmide).

O resíduo de aminoácido Glu, bem como o zinco bivalente, aumentam a polarização da molécula de água, tornando-a um nucleófilo. Nesta condição, a água ataca o carbono da ligação peptídica do substrato, a qual será rompida ao fornecer a sua hidroxila, e levando o zinco bivalente a um estágio de transição pentacoordenado.

A protonação do nitrogênio da ligação peptídica quebra a ligação entre o carbono e nitrogênio, liberando assim os produtos de clivagem. Por fim, a enzima é regenerada e ocorre a incorporação de uma nova molécula de água.

#### **1.6.4 A espécie *Bothrops moojeni***

A caissaca (*Bothrops moojeni*) é uma serpente de hábito terrícola, sendo a jararaca mais encontrada na Região Centro-oeste (Fig.11). Seu habitat é no cerrado, com hábitos de alimentação básica de roedores e lagartos. Seu tamanho pode chegar até 1,50 metro, possui denteção solenóglifa, e como uma serpente noturna, é uma das mais agressivas do grupo de jararacas (Fig.12). A distância do bote das serpentes corresponde a aproximadamente 1/3 do seu comprimento, e é dado mais no sentido horizontal. A caissaca desfere seu bote mais para o sentido vertical, podendo atingir desta forma partes mais altas do corpo de uma pessoa.

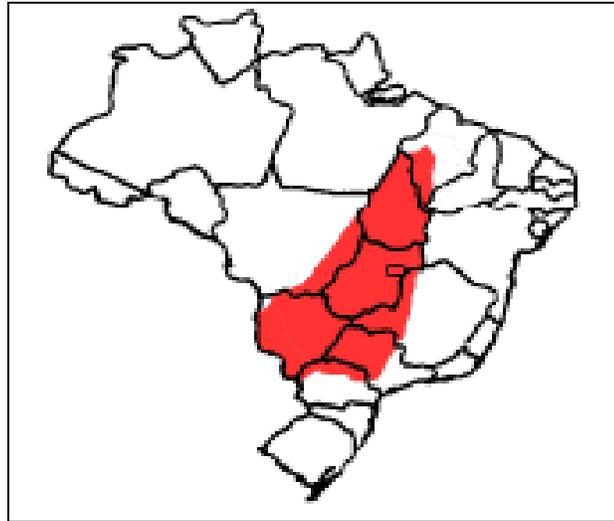


Fig. 11 – Ocorrência da espécie de serpente *Bothrops moojeni* no território brasileiro (BUTANTAN, 2000).



Fig. 12 – Exemplar de serpente *Bothrops moojeni*. Fonte: Bernarde, 2009.

Glândulas de veneno de serpente são uma rica fonte de moléculas bioativas, tais como peptídeos, proteínas e enzimas, que apresentam importante atividade farmacológica localizada e efeitos sistêmicos como dor, edema, hemorragia e necrose muscular. Muitos dos componentes do veneno botrópico são enzimas proteolíticas (BERNARDES *et al*, 2007). A toxina desta espécie pode acelerar o processo de dano e morte celular, quando em contato com os tecidos, numa picada por exemplo. No veneno desta espécie, dentro outros compostos protéicos, são encontrados desintegrinas, metaloproteinase, miotoxinas, NADH desidrogenases, fosfolipases e

trombinas. A metaloprotease (Fig.13) presente no veneno desta espécie hidrolisa preferencialmente a cadeia alfa do fibrinogênio, como também pode hidrolisar a cadeia beta. Hidrolisa azocaseína, e a cadeia beta de fibrina. Contudo, não apresenta atividade hemorrágica significativa em ratos. Abaixo, segue a seqüência de resíduos de aminoácidos da metaloprotease VMBMP\_BOTMO, objeto de estudo deste trabalho. As principais características desta enzima serão descritas na Tabela 03.

10	20	30	40	50	60
EQQKFSPRYI	ELVVVADHGM	FKKYNSNLNT	IRKVVHEMVN	SMNGFYRSVD	VTASLANLEV
70	80	90	100	110	120
WSKKDLINVQ	KDSRETLKSF	GEWRERDLLP	RISHDNAQLL	TAIVFDGHTI	GRAYTGGMCD
130	140	150	160	170	180
PRHSGVVMD	HSPKNLQVAV	TMAHELGHNL	GMHHDGNQCH	CDAASCIMAD	SLSVVLSYEF
190	200				
SDCSQNQYQT	YLTKHNPQCI	LNEPL			

Fig. 13 – Seqüência de aminoácidos da VMBMP\_BOTMO (P85314), de *Bothrops moojeni*.

Fonte: UniProt, 2010.

Esta metaloprotease também interage seletivamente com íons cálcio ( $\text{Ca}^{+2}$ ) e zinco ( $\text{Zn}^{+2}$ ), apresentando atividade catalítica (metaloendopeptidase) de uma reação de hidrólise, quebrando as ligações alfa-peptídicas em uma cadeia polipeptídica, por um mecanismo no qual a água atua como um nucleófilo. Nesta reação, um ou dois íons metálicos se agregam à molécula de água, onde as cadeias laterais de aminoácidos são pontos desta ligação iônica.

Tabela 3 – Características da metaloprotease P85314 VMBMP\_BOTMO.

Nomenclatura	Zinco-metaloprotease VMBMP_BOTMO
Taxonomic ID (NCBI)	98334
Linhagem taxonômica	Eukaryota › Metazoa › Chordata › Craniata › Vertebrata › Euteleostomi › Lepidosauria › Squamata › Scleroglossa › Serpentes › Colubroidea › Viperidae › Crotalinae › Bothrops
Número de resíduos	205 aminoácidos
Domínio	P1 (8-204aa) Peptidase M12B
Pontos de ligante metálico	Sítio ativo: Glu <sub>145</sub>  Zinco ligante: Hys <sub>144</sub> , Hys <sub>148</sub> , Hys <sub>154</sub> .
Pontes dissulfeto	Cys <sub>119</sub> ↔ Cys <sub>199</sub>  Cys <sub>159</sub> ↔ Cys <sub>183</sub>  Cys <sub>161</sub> ↔ Cys <sub>166</sub>
Cofator	Zinco (Zn <sup>+2</sup> )
Localização subcelular	Secretada
Especificidade de tecido	Encontrada em glândulas de veneno
Função molecular	Atua como hidrolase, metaloprotease (metaloendopeptidase) e toxina. Considerada uma substância patogênica e proteolítica (degrada proteínas).
Massa molecular	23095 Da
Integrada ao UniProt	Dezembro 2007
Última anotação (UniProt)	Junho 2009

## 2 JUSTIFICATIVA

Segundo Meidanis & Setubal (2003), as moléculas bioativas podem inibir ou catalisar reações bioquímicas, algumas vezes objetivando efeitos macroscópicos como a cura ou atenuação de uma doença. Por sua vez, o alinhamento computacional, quando realizado com moléculas biológicas como as proteínas, permite a detecção precoce de moléculas que sejam viáveis economicamente às indústrias farmacêuticas, podendo assim guiar os pesquisadores para as moléculas mais promissoras para novos fármacos. Essa nova aplicação faz com que empresas biotecnológicas estejam familiarizadas com metodologias de busca de identidade e similaridade em seqüências de proteínas.

Atualmente, o alinhamento e a análise estatística de seqüências de proteínas são aplicados em projetos de proteômica, relacionados com ramos econômicos altamente significativos para a Brasil. Destaca-se, por exemplo, o projeto de descrição do genoma do filopatógeno *Xylella fastidiosa*, causador da “praga do amarelinho” em culturas de laranja. Além disto, Meidanis (2003) afirmam ainda que resultados esperados para os problemas de melhoramento genético, de animais e de vegetais, são tratados com menos dispêndios de custo e com maior rapidez através de ferramentas computacionais aplicadas aos problemas da Biologia Molecular e da Química Estrutural de Proteínas.

Para o estudo de sistemas moleculares, é mais viável a pesquisa *in silico*, onde podem ser observados os efeitos de uma perturbação neste sistema, sem efetivamente (em bancada de laboratório) realizar essa perturbação (SANTOS FILHO & ALENCASTRO, 2003). Uma das aplicações do estudo de seqüências especiais, simuladas por computador, é o mecanismo de interação das moléculas bioativas do fármaco com a molécula-alvo, presente no organismo do paciente.

A aplicação destas novas metodologias, com o apoio de ferramentas de Bioinformática, vem se tornando cada vez mais freqüente para o desenvolvimento de fármacos, por ser um recurso mais rápido e complementar ao tipo de pesquisa experimental (XIONG, 2006).

O planejamento racional fármacos, baseado em procedimentos computacionais *in silico*, tem se mostrado como uma eficiente estratégia para o desenvolvimento e validação, sendo capaz de contribuir para etapas de descoberta de protótipos (*lead compounds*), otimização toxicológica (afinidade, especificidade, eficácia e toxicidade), podendo chegar até a contribuir para a elaboração de candidatos a testes clínicos para drogas (SILVA, 2007).

Assim, com as informações sobre a conformação estrutural da proteína-alvo, é possível verificar como um bioreceptor e/ou ligantes interagem entre si e como podem ser formados novos compostos, com complementaridade estérica, hidrofóbica e eletrostática ao seu sítio de ligação. Segundo Marshall (2004), 78% dos fármacos atuais têm como alvo receptor um ou mais pontos de proteínas específicas, o que demonstra a importância do conhecimento da estrutura tridimensional de um alvo molecular para uma potencial aplicação biotecnológica. Muitos dos processos de descoberta e desenvolvimento de novos fármacos são dependentes de resultados gerados por métodos computacionais (PEITISCH, 2004). Neste contexto, enormes são os esforços pelo mundo todo, por instituições públicas e privadas, no sentido de elucidar o maior número de estruturas tridimensionais de proteínas (NORWELL & MACHALEK, 2000).

A modelagem da estrutura tridimensional da metaloprotease P85314 (*B.moojeni*), por sua vez, tem importante sentido no que diz respeito à formulação de soros antiofídicos mais potentes e específicos para situações de emergência médica (HEINEMANN, 2000). Ainda mais na Região Amazônica, onde se faz necessária a aplicação de medidas mais eficazes, a fim de amenizar problemas de saúde como este, diante da dimensão extensa que se tem de vencer para que um tratamento médico chegue ao paciente a tempo. Conhecendo a estrutura de metaloproteases, será possível, posteriormente, realizar a modelagem molecular também de inibidores dos sítios ativos da biomolécula, uma vez que a hemorragia é uma das mais devastadoras características da ação dos venenos de serpentes do gênero *Bothrops*. Estes inibidores poderiam ser utilizados como base de fármacos e novas drogas combinadas. Ainda, modelar a estrutura da metaloprotease do veneno de *Bothrops moojeni*, a fim de conseguir a visão tridimensional da disposição de sítios importantes da biomolécula, permite a compreensão da estrutura de outras metaloproteases similares a esta. Estes estudos destacam, por análise comparativa por homologia, a função de estruturas tridimensionais determinadas e utilizadas na construção de novos fármacos, como os inibidores deste tipo de biomoléculas.

### **3 OBJETIVOS**

#### **3.1 Objetivos Gerais**

Obter o modelo da estrutura terciária de metaloprotease VMBMP\_BOTMO de *Bothrops moojeni* através de técnicas de modelagem molecular por homologia estrutural.

#### **3.2 Objetivos Específicos**

Realizar *in silico* os alinhamentos locais, de modo a verificar o grau de identidade e similaridade entre as seqüências de metaloproteases presentes em venenos de viperídeos.

Obter um grupo de proteínas semelhantes à seqüência-alvo, ou seja, a metaloprotease VMBMP\_BOTMO de *B.moojeni*.

Predizer um modelo tridimensional da metaloprotease VMBMP\_BOTMO, presente no veneno de *Bothrops moojeni*, baseado-se em técnicas para a predição por homologia estrutural, através de seqüências com alta identidade e similaridade entre si.

## 4 MATERIAIS E MÉTODOS

O esquema geral dos métodos utilizados segue a ordem descrita abaixo. As principais etapas do processo de modelagem comparativa por homologia serão descritas detalhadamente em cada subseção: busca de seqüências-molde para a comparação com a seqüência-alvo, alinhamento de seqüências, modelagem comparativa e validação estrutural.

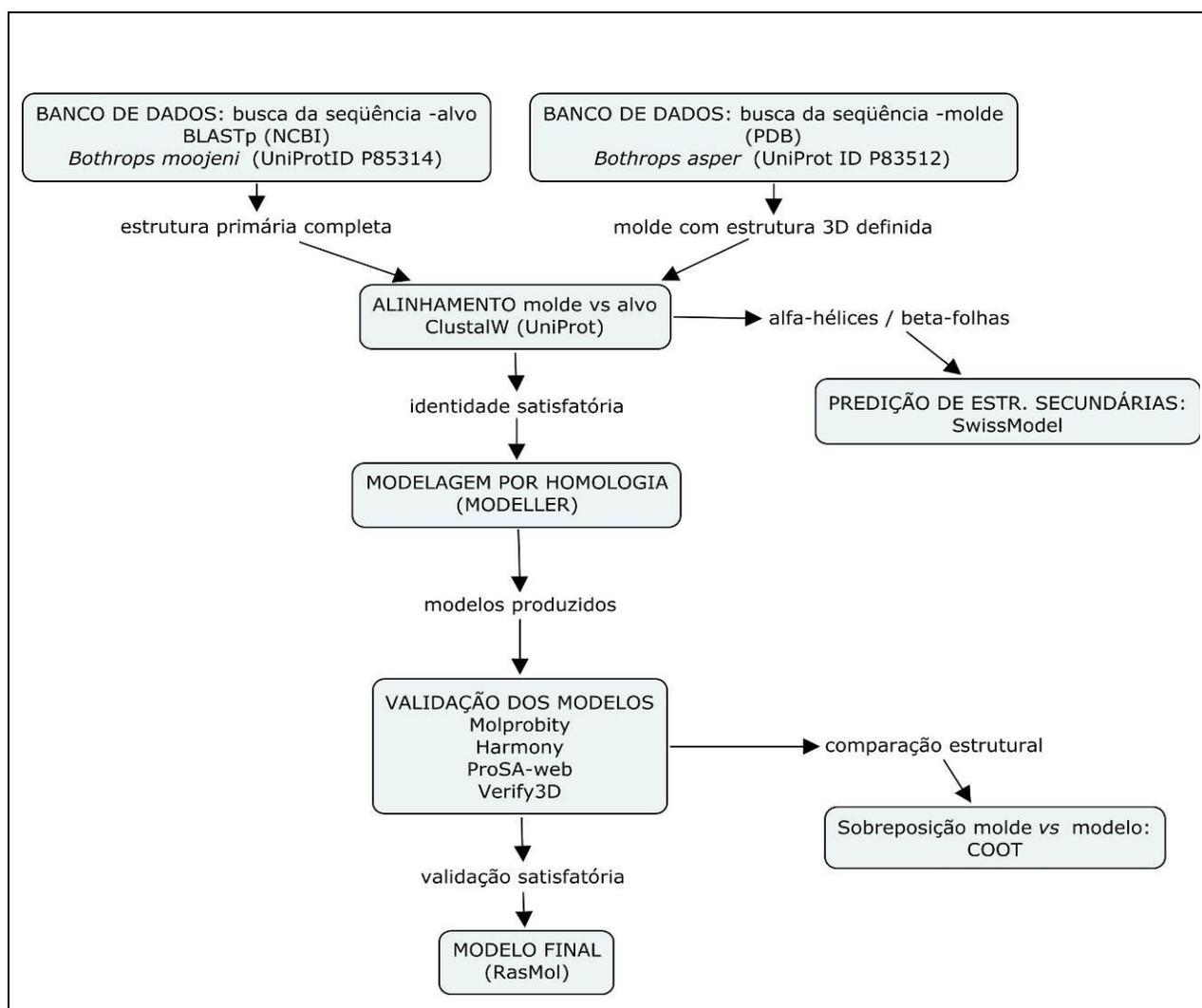


Fig. 14 – Esquema geral do procedimento de modelagem molecular estrutural.

#### **4.1 Obtenção de estruturas similares (seqüências-moldes)**

Inicialmente, foi selecionada a seqüência de aminoácidos da metaloprotease presente na toxina da peçonha da serpente *Bothrops moojeni*, a qual ainda não possui uma estrutura tridimensional conhecida. Esta estrutura será determinada neste trabalho e será denominada, daqui para a frente, de estrutura da “seqüência-alvo”.

As estruturas tridimensionais (e suas respectivas seqüências de aminoácidos) selecionadas como molde foram obtidas através de alinhamento BLAST das seqüências de aminoácidos do gênero *Bothrops*, em banco de dados de proteínas. Foi utilizado neste procedimento o formato Pearson (FASTA) para as seqüências de aminoácidos.

A seqüência de aminoácidos VMBMP\_BOTMO foi utilizada como alvo na busca por seqüências similares, em banco de dados, e com estruturas tridimensionais já determinadas experimentalmente. A obtenção de seqüências similares (moldes) à seqüência alvo foi feita no banco de dados do Uniprot (<http://www.uniprot.org>). O alinhamento BLAST indicou a similaridade entre as seqüências alvo e moldes. Além disto, o banco de dados do PDB (<http://www.pdb.org>) foi consultado para obtenção dos moldes determinados experimentalmente bem como para assegurar que a estrutura terciária da metaloprotease P85314 VMBMP\_BOTMO não tinha sido elucidada por métodos experimentais.

O *score* do alinhamento foi de fundamental importância para a ordenação das seqüências em graus de similaridade. Quanto maior a similaridade entre a metaloprotease alvo e as seqüências moldes sugeridas, menor seria a chance de uma modelagem equivocada ou incompleta.

A estrutura (e a seqüência) molde foi então selecionada, segundo os critérios de:

- Maior porcentagem de identidade e similaridade com relação à seqüência-alvo;
- Menor ocorrência de inserções e deleções na seqüência-molde;

- A seqüência-molde deveria apresentar uma estrutura 3D determinada experimentalmente com parâmetros experimentais satisfatórios (R-fator e resolução).

## **4.2 Alinhamento de seqüência-alvo versus molde**

O alinhamento entre as seqüências-alvo e -molde foi realizado pelo programa ClustalW, no servidor Uniprot, disponível em <http://services.uniprot.org/clustalw>. Através deste alinhamento, foi possível identificar os resíduos de aminoácidos conservados, prevendo a semelhança de características físico-químicas e estruturais da estrutura a ser modelada.

A estrutura escolhida para servir de molde à metaloprotease de *B.moojeni* foi a metaloprotease de *B.asper*, por apresentar 81% de identidade com a seqüência-alvo, e por já dispor de uma estrutura tridimensional depositada no banco de dados Protein Data Bank (PDB).

A metaloprotease de *B.asper* (Fig.14) possui código de entrada como VMBP1\_BOTAS, código UniProt ID P83512, PDB ID 1ND1, código taxonômico 8722 (NCBI Taxonomic database), e foi integrada ao banco de dados UniProt em abril de 2004, tendo sua última modificação em março de 2009.

Como funcionalidade biológica, esta proteína é uma zincoprotease hemorrágica do veneno de serpente, que também ativa o sistema complemento, e induz a quimiotaxia neutrofílica em camundongos (ARCE et al, 2005). Esta proteína apresenta um cofator ligante de íon zinco. Pertence à família de metaloproteases, subfamília P-I (FARSKY, 2000), com uma massa molecular de 22.735 Da, (determinado por MALDI-TOF). Abaixo, segue a estrutura primária da metaloprotease P83512 (*B.asper*) (Fig.15).

10	20	30	40	50	60
MIEVLLVTIC	LAVFPYQGSS	IILESQNVND	YEVVYPRKVT	ELPKGAVQPK	YEDAMQYEFK
70	80	90	100	110	120
VNGEPVVLHL	EKNKGLFSED	YSETHYSPDG	RKIITYPSFE	DHCYYHGRIE	NDADSTASIS
130	140	150	160	170	180
ACNGLKGFHK	LQGETYLIIEP	LKLSQSEAHA	VYKYENVEKE	DEAPKMGVVT	ETNWESYEPI
190	200	210	220	230	240
KKASQSNLTP	EQQRFSPIYI	ELAVVADHGI	FTKYNSNLNT	IRTRVHEMLN	TVNGFYRSVD
250	260	270	280	290	300
VHAPLANLEV	WSKQDLIKVQ	KDSSKTLKSF	GEWRERDLLP	RISHDHAQLL	TAVVFDGNTI
310	320	330	340	350	360
GRAYTGGMCD	PRHSGVGVVD	HSKNNLWVAV	TMAHELGHNL	GIHHDGTGSCS	CGAKSCIMAS
370	380	390	400		
VLSKVLSEYEF	SDCSQNQYET	YLTNHNQPCI	LNKPLLVSG	NELLEAGE	

Fig. 15 – Sequência de aminoácidos da VMBP1\_BOTAS (P83512), de *Bothrops asper*.

Fonte: UniProt, 2010 (<http://www.uniprot.org>).

O cDNA da metaloprotease VMBP1\_BOTAS (P83512) expressa no total 408 aminoácidos. A partir do resíduo de aminoácido 1 ao 20, apresenta um peptídeo sinal; do 21 ao 408, possui em sua estrutura um pró-peptídeo; do 192 ao 394 a porção funcional da proteína nativa após clivagem (total de 203 resíduos de aminoácidos).

A região do domínio peptidase M12B encontra-se entre os resíduos de aminoácidos 198 e 394 e o aminoácido catalítico é o Glu<sub>335</sub>. Os aminoácidos ligantes de zinco estão nas posições His<sub>334</sub>, His<sub>338</sub> e His<sub>344</sub>. Esta proteína apresenta ainda um resíduo modificado na posição Gln<sub>192</sub> para ácido carboxílico pirrolidônico (o primeiro resíduo de aminoácido na proteína funcional), e pontes dissulfeto entre os resíduos de cisteína nas posições 309 e 389, 349 e 375, 351 e 356. Sua estrutura tridimensional foi elucidada por cristalografia de raio-X, em uma resolução de 1,93Å. A estrutura tridimensional da sequência-molde está ilustrada na Fig.16.

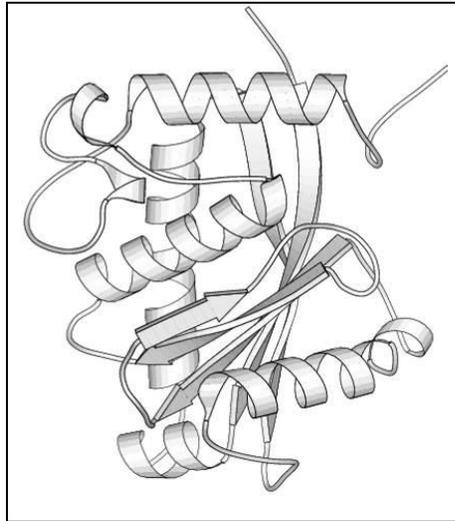


Fig. 16 – Estrutura tridimensional da metaloprotease VMBP1\_BOTAS P83512 (*B.asper*).

### **4.3 Obtenção do modelo primário**

Uma vez obtido o modelo primário da VMBMP\_BOTMO, a partir da modelagem com o molde VMBP1\_BOTAS, este modelo será avaliado qualitativamente, de modo a se fazerem tantas modificações forem necessárias para que um modelo satisfatório seja alcançado, ou seja, para que a estereoquímica do modelo esteja coerente segundo a comparação entre os moldes utilizados.

Para tanto, será utilizado inicialmente o programa MODELLER, disponível em <http://www.salilab.org/modeller/>, de modo a processar as informações estruturais, que serão obtidas a partir dos moldes selecionados, com entrada a partir de um arquivo PDB como molde.

Primeiramente, como rotina de modelagem através do MODELLER, será utilizado um *script* de alinhamento (Fig.17) da seqüência-alvo com a seqüência-molde. O programa executará a rotina, a partir dos arquivos de entrada: seqüência-alvo (formato PIR) e seqüência-molde (formato PDB).

```
from modeller import *  
  
env = environ()  
  
aln = alignment(env)  
  
mdl = model(env, file='1ND1', model_segment=('FIRST:A','LAST:A'))  
  
aln.append_model(mdl, align_codes='1ND1', atom_files='1ND1.pdb')  
  
aln.append(file='prot.ali', align_codes='prot')  
  
aln.align2d()  
  
aln.write(file='prot-1nd1.ali', alignment_format='PIR')  
  
aln.write(file='prot-1nd1.pap', alignment_format='PAP')
```

Fig. 17– Script de alinhamento das seqüências alvo e molde, a ser executado pelo aplicativo MODELLER.

A partir do arquivo de alinhamento gerado *prot-1nd1.ali*, foi possível executar o *script* (Fig.18) para modelagem molecular da estrutura-alvo. A modelagem molecular retornou cinco modelos, em formato de coordenadas PDB, que foram encaminhados para validação antes de serem finalizados como sugestões de estruturas tridimensionais para a metaloprotease em questão.

```

from modeller import *

from modeller.automodel import *

env = environ()

a = automodel(env, alnfile='prot-1ND1.ali',

              knowns='1ND1', sequence='prot',

              assess_methods=(assess.DOPE, assess.GA341))

a.starting_model = 1

a.ending_model = 5

a.make()

```

Fig. 18 – Script para modelagem molecular da seqüência-alvo em função da seqüência-molde. O arquivo de saída oferece cinco modelos selecionados.

Uma vez obtido o modelo primário da VMBMP\_BOTMO, a partir da modelagem com o molde VMBP1\_BOTAS, este modelo foi avaliado quantitativamente, de modo a se fazerem tantas modificações quanto forem necessárias para que um modelo satisfatório fosse alcançado; ou seja, para que a estereoquímica do modelo estivesse coerente segundo a comparação com estruturas obtidas experimentalmente.

#### **4.4 Validação de modelos**

O refinamento dos modelos preliminares visou justamente a verificação da conformação sugerida, tendo como base os resultados de análises computacionais e o retorno de parâmetros de qualidade. Para a visualização dos modelos gerados, foi utilizado o visualizador RasMol, disponível em <http://www.umass.edu/microbio/rasmol/>. Os programas utilizados para a validação dos modelos já foram descritos na Introdução.

## 5 RESULTADOS E DISCUSSÕES

Das seqüências coletadas por alinhamento local de proteínas (BLASTp), foram selecionadas apenas as seqüências que possuíam uma identidade mínima de 70% com a seqüência de metaloprotease P85314 de *B. moojeni*. O resultado é mostrado na Tabela 04, onde as seqüências estão organizadas de acordo com o *e-value*, *score* e a porcentagem de identidade entre os resíduos de aminoácidos.

As seqüências com identidade acima de 70%, indicadas na Tabela 04, também foram comparadas à seqüência-alvo através de um alinhamento múltiplo. Este alinhamento serviu para identificar regiões conservadas em cada seqüência (Fig.19). Neste alinhamento, também é possível visualizar as regiões com certa variação quanto ao tipo de aminoácido que compõe a seqüência. É provável que estas regiões representem as dobras/alças do enovelamento protéico características da metaloprotease de cada espécie, importantes para a conformação tridimensional molecular.

As dobras/alças da estrutura tridimensional a ser definida para a metaloprotease de *B. moojeni* podem ser modeladas a partir de um banco, o qual é formatado a partir de estruturas tridimensionais já conhecidas e postadas no banco de dados PDB. Assim, é possível a previsão da conformação estrutural para os trechos de dobramento, a partir dessas referências, caso o modelo inicial usado como molde não possua informação suficiente para a modelagem destas alças.

Tabela 4 – Amostra de seqüências com identidade acima de 49% com a metaloprotease de *B.moojeni*.

Acesso	Nome da proteína	Organismo	N. de AA	Ident (%)	E-value
P85314	Zinc metalloproteinase BmooMPalfa-I (EC 3.4.24.-)	Bothrops moojeni (Lance-headed viper) (Caissaca).	200	100.0	1.0E-116
Q7T1T4	Zinc metalloproteinase BjussuMP-2 (EC 3.4.24.-) (BjussuMP-II) (BOJUMET I) (Nonhemorrhagic metalloprotease MP-II) (Fragment)	Bothrops jararacussu (Jararacussu).	369	88.0	1.0E-101
P83512	Zinc metalloproteinase Bap1 (EC 3.4.24.-) (Hemorrhagic metalloproteinase BaP1) (Bap-1) (Precursor)	Bothrops asper (Terciopelo).	408	81.0	2.0E-92
Q072L5	Zinc metalloproteinase/disintegrin [Cleaved into: Metalloproteinase (EC 3.4.24.-) ; Disintegrin ] (Precursor)	Bothrops asper (Terciopelo).	477	82.0	5.0E-92
Q5XUW8	Zinc metalloproteinase/disintegrin [Cleaved into: Metalloproteinase insularinase-A (EC 3.4.24.-) ; Disintegrin insularin ] (Precursor)	Bothrops insularis (Island jararaca) (Queimada jararaca).	476	78.0	5.0E-86
Q1PHZ4	Zinc metalloproteinase-disintegrin BjussuMP-1 (EC 3.4.24.-) (BjussuMP-I) (Precursor) (Fragment)	Bothrops jararacussu (Jararacussu).	547	74.0	3.0E-84
P84907	Zinc metalloproteinase leucurolysin-A (Leuc-A) (EC 3.4.24.-) (Fragment)	Bothrops leucurus (White-tailed jararaca) (White-tailed lancehead).	193	76.0	6.0E-84
Q9I9R4	Zinc metalloproteinase neuwiedase (EC 3.4.24.-) (Fragment)	Bothrops neuwiedi pauloensis (Neuwied's lancehead).	198	68.0	1.0E-72
Q7T1T3	Metalloprotease BOJUMET III (Fragment)	Bothrops jararacussu (Jararacussu).	243	84.0	3.0E-72
Q8QG89	Zinc metalloproteinase BITM02A (EC 3.4.24.-) (Precursor)	Bothrops insularis (Island jararaca) (Queimada jararaca).	409	57.0	4.0E-59
Q98SP2	Zinc metalloproteinase/disintegrin [Cleaved into: Metalloproteinase jararafibrase-2 (EC 3.4.24.-) (Jararafibrase II); Disintegrin bothrostatin (D-BTT) ] (Precursor)	Bothrops jararaca (Jararaca).	477	54.0	9.0E-56
Q8UVG0	Zinc metalloproteinase-disintegrin berythactivase (EC 3.4.24.-) (ery1) (Precursor)	Bothrops erythromelas (Caatinga lance head).	612	53.0	2.0E-55
P85420	Zinc metalloproteinase atroxlysin-I (EC 3.4.24.-) (Atroxlysin-I)	Bothrops atrox (Barba amarilla) (Fer-de-lance).	202	52.0	6.0E-52
Q98UF9	Zinc metalloproteinase-disintegrin HF3 (EC 3.4.24.-) (Precursor)	Bothrops jararaca (Jararaca).	606	52.0	3.0E-51
O93523	Zinc metalloproteinase-disintegrin bothropasin (EC 3.4.24.49) [Cleaved into: Disintegrin bothropasin ] (Precursor)	Bothrops jararaca (Jararaca).	610	50.0	2.0E-49
P30431	Zinc metalloproteinase-disintegrin jararhagin (EC 3.4.24.73) (Jararafibrase I) (JF I) (JG) (HF2-proteinase) [Cleaved into: Disintegrin jararhagin-C (Jaracetin) ] (Precursor) (Fragment)	Bothrops jararaca (Jararaca).	571	49.0	4.0E-49
Q8QG88	Zinc metalloproteinase-disintegrin BITM06A (EC 3.4.24.-) (Precursor)	Bothrops insularis (Island jararaca) (Queimada jararaca).	610	50.0	1.0E-48

Accession	Entry name	Protein names
<input checked="" type="checkbox"/> Q072L5	VM2_BOTAS	<b>Zinc metalloproteinase/disintegrin</b> [Cleaved into: Metalloproteinase (EC 3.4.24.-) ; Disintegrin ]
<input checked="" type="checkbox"/> P83512	VMBP1_BOTAS	<b>Zinc metalloproteinase Bap1</b> (EC 3.4.24.-) (Hemorrhagic metalloproteinase BaP1) (Bap-1)
<input checked="" type="checkbox"/> Q5XUW8	VMINS_BOTIN	<b>Zinc metalloproteinase/disintegrin</b> [Cleaved into: Metalloproteinase insularinase-A (EC 3.4.24.-) ; Disintegrin insularin ]
<input checked="" type="checkbox"/> P84907	VMLEA_BOTLC	<b>Zinc metalloproteinase leucurolysin-A</b> (Leuc-A) (EC 3.4.24.-)
<input checked="" type="checkbox"/> P85314	VMAI_BOTMO	<b>Zinc metalloproteinase BmooMPalfa-I</b> (EC 3.4.24.-)
<input checked="" type="checkbox"/> Q7T1T4	VMMP2_BOTJR	<b>Zinc metalloproteinase BjussuMP-2</b> (EC 3.4.24.-) (BjussuMP-II) (BOJUMET I) (Nonhemorrhagic metalloprotease MP-II)
<input checked="" type="checkbox"/> Q1PHZ4	VMMP1_BOTJR	<b>Zinc metalloproteinase-disintegrin BjussuMP-1</b> (EC 3.4.24.-) (BjussuMP-I)

```

Q072L5 MIEVLLVTICLAVSPYQGSSII LESGNVNDYEVVYPRKVTE LPKGAVQPKYEDAMQYEFK
P83512 MIEVLLVTICLAVFPYQGSSII LESGNVNDYEVVYPRKVTE LPKGAVQPKYEDAMQYEFK
Q5XUW8 MIQVLLVTICLAAFPYQGSSII LESGNVNDYEVVYARKVTE LPKGAVQPKYEDAMQYEFK
P84907 -----
P85314 -----
Q7T1T4 ----- TELPKGAVQPKYEDAMQYEFK
Q1PHZ4 ----- EFK

```

```

Q072L5 VNGEPVVLHLEKNKGLFSEDYSETHYSPDGRKI IITYPSFEDHCYYHGRIENDADSTASIS
P83512 VNGEPVVLHLEKNKGLFSEDYSETHYSPDGRKII IITYPSFEDHCYYHGRIENDADSTASIS
Q5XUW8 VNGEPVVLHLEKNKGLFSEDYSETHYSPDGRQI IITYPPFEDHCYYHGRIENDADSTASIS
P84907 -----
P85314 -----
Q7T1T4 GNGEPVVLHLEKNKGLFSEDYSETHYSPDGRQI IITYPPFEDHCYYHGRIENDADSTASIS
Q1PHZ4 VNGEPVVLHLEKNKGLFSEDYSETHYSPDGRQI IITYPPVEDHCYYHGRIENDADSTASIS

```

```

Q072L5 ACNGLKGFHFKIQGETYLI EPLKLS DSEAHAVYKY ENVEKEDEAPKMCGVTETNWESEY EPI
P83512 ACNGLKGFHFKLQGETYLI EPLKLS DSEAHAVYKY ENVEKEDEAPKMCGVTETNWESEY EPI
Q5XUW8 ACNGLKGFHFKLQGETYLI EPLKLPDSEAHAVYKY ENVEKEDEAPKMCGVTETNWESEY EPI
P84907 -----
P85314 -----
Q7T1T4 ACNGLKGFHFKLQGETYLI EPLKLS DSEAHAVYKY ENVEKEDEAPKMCGVTETNWESEY EPI
Q1PHZ4 ACNGLKGFHFKLQGETYLI EPLKLS DSEAHAVYKY ENVEKEDEAPKMCGVTETNWESEY EPI

```

Q072L5 KKASQSNLTPEQQRFSPRYIELAVVADHGIFTKYNSNLNTIRTRVHEMLNTVNGFYRSVN  
P83512 KKASQSNLTPEQQRFSPRYIELAVVADHGIFTKYNSNLNTIRTRVHEMLNTVNGFYRSVD  
Q5XUW8 EKASQSNLTPEQQKFSRYIELAVVADHGMFTKYNSNLNTIRTRVHEMVNTLNGFFRSVN  
P84907 -----SPRYIELVVVADHGMFKKYNSNLNTIRKVVHEMLNTVNGFYRSMN  
P85314 -----EQQKFSRYIELVVVADHGMFKKYNSNLNTIRKVVHEMVNSMNGFYRSVD  
Q7T1T4 KKASPSNLTPEQQKFSRYIEVAVVADHRMFKKYNSNLNTIRKVVHEMVNSMNGVYRSM  
Q1PHZ4 KKASKLVVTAEQQKFPYRYVEIVVVDRRMVTKYNGDLKKIRKVVYELVNI VNIYRSLN

Q072L5 VTASLASLEVWSKKDLIKVEKDSKTLTSFGEWRRERDLPRI SHDHAQLLTTIVFDNYVI  
P83512 VHAPLANLEVWSKQDLIKVQKDSKTLTSFGEWRRERDLPRI SHDHAQLLTAVVFDGNTI  
Q5XUW8 VDASLANLEVWSKKDLIKVEKDSKTLTSFGEWRRERDLPRI SHDHAQLLTTIVFDQQT  
P84907 VDASLVNLEVWSKKDLIKVEKDSKTLTSFGEWRRERDLPRI SHDHAQLLTVIFLDEETI  
P85314 VTASLANLEVWSKKDLINVQKDSRETLSFGEWRRERDLPRI SHDNAQLLTAIVFDGHTI  
Q7T1T4 VHLSLANLEVWSKKDLINVQKDSRETLSFGEWRRERDLPRI SHDNAQLLTAIVFDQQT  
Q1PHZ4 VHVALVGLLEIWSKGDKITVQPSDYTLNSFGEWRRERDLPRI SHDNAQLLTAIVFDGPTI

Q072L5 GITEFGKMCDPKLSVGVVRDHSSEINLQVAVAMAHELGHNLMGYHDGNQCHCDAASCIMAD  
P83512 GRAYTGGMCDPRHSVGVVRDHSKNNLWVAVTMAHELGHNLGIHHDGSCSCGAKSCIMAS  
Q5XUW8 GLAYTAGMCDPRQSVAVVMDHSSKKNLRVAVTMAHELGHNLGMDHD-DTCTCGAKSCIMAS  
P84907 GIAYTAGMCDLSQMVAVGQDY-----VAVSMAHELGHNLGMRHDGNQCHCNAPSCIMAD  
P85314 GRAYTGGMCDPRHSVGVVMDHSPKNLQVAVTMAHELGHNLGMHHDGNQCHCDAASCIMAD  
Q7T1T4 GRAYIAGMCDPRHSVGVVMDHSENLQVAVTMAHELGHNLGMEHDENQCHCDAPSCVMAS  
Q1PHZ4 GRAYIAGMCDPRHSVGVVMDHSENLQVAVTMAHELGHNLGMEHDENQCHCDAPSCVMAS

\* : . \*\*\* \* . \* : \*\* :\*\*\*\*\*: \* . \* \* . \* : \* : \* : \* : \* : \* : \* : \*

Q072L5 TLREVLVSSEFSDCSQNQYETYLTKHNPQCILNEPLLIVSGNELLEAGEECDCDAPENPCC  
P83512 VLSKVLSSEFSDCSQNQYETYLTNHNPPQCILNKPLLTVSGNELLEAGE-----  
Q5XUW8 TISKGLSFEFSDCSQNQYQTYLTDHNPQCILNKPLTTVSGNELLEAGEECDCGAPENPCC  
P84907 TLSKGLSFEFSDCSQNQYQTYLTKHNPQCILNKP-----  
P85314 SLSVVLSSEFSDCSQNQYQTYLTKHNPQCILNEPL-----  
Q7T1T4 VLSVVLSSEFSDCSQNQYQTYLTKHNPQCILNEPLLTVSGNELLEAGE-----  
Q1PHZ4 VLSVVLSSEFSDCSQNQYQTYLTKHNPQCILNEPLLTVSGNELLEAGEECDCGAPENPCC

: \*\* :\*\*\* . \*\*\*\*\*:\*\*\*\*\* . \*\*\*\*\*:\*

Q072L5 DAATCKLRPGAQCAEGLCCDQCRFKGAGKICRRARGDNPDDRCTGQSADCPRNRFHA---  
P83512 -----  
Q5XUW8 DAATCKLRPRAQCAEGLCCDQCRFKGAGKICRRARGDNPDDRCTGQSADCPRNRFHA---  
P84907 -----  
P85314 -----  
Q7T1T4 -----  
Q1PHZ4 DAATCKLRPGAQCAEGLCCDQCRFKGAGKICRRARGDNPDDRCTGQSADCPRNRFHRNGQ

Q072L5 -----  
P83512 -----  
Q5XUW8 -----  
P84907 -----  
P85314 -----  
Q7T1T4 -----  
Q1PHZ4 PCLYNHGYCYNGKCPIMFYQCYFLFGSNATVAEDDCFNNKKGDKYFYCRKENEKYIPCA

Q072L5	-----		
P83512	-----		
Q5XUW8	-----		
P84907	-----		
P85314	-----		
Q7T1T4	-----		
Q1PHZ4	PCLYNHGYCYNGKCPIMFYQCYFLEFGSNATVAEDDCFNNNKKGDKYFYCRKENEKYI PCA		
Q072L5	-----		
P83512	-----		
Q5XUW8	-----		
P84907	-----		
P85314	-----		
Q7T1T4	-----		
Q1PHZ4	QEDVKCGRLFCDNKKYPCHYNYSEDLDFGMVDHGTKCADGKVC SNRQCVDVNEAYKSTTV		
Q072L5	----	477	VM2_BOTAS
P83512	----	408	VMBP1_BOTAS
Q5XUW8	----	476	VMINS_BOTIN
P84907	----	193	VMLEA_BOTLC
P85314	----	205	VMAI_BOTMO
Q7T1T4	----	369	VMMP2_BOTJR
Q1PHZ4	FSLI	547	VMMP1_BOTJR

Fig. 19 – Alinhamento múltiplo para as seqüências de metaloproteases com identidade acima de 70% com a metaloprotease de *B.moojeni* (P85314). Obtido através do aplicativo ClustalW (<http://services.uniprot.org/clustalw>).

Conforme a porcentagem de identidade, a seqüência de metaloprotease mais próxima da seqüência-alvo VMBMP\_BOTMO de *B.moojeni* seria a VMMP2\_BOTJR, oriunda de *B.jararaca*. Contudo, esta proteína ainda não teve sua estrutura tridimensional elucidada experimentalmente (e disponível no *Protein Data Bank* - PDB) e, portanto, não foi selecionada como molde. A próxima metaloprotease proposta como molde foi a VMBP1\_BOTAS, contida na peçonha de *Bothrops asper*, que, apesar de uma identidade um pouco mais baixa que a seqüência da metaloprotease de *B.jararaca*, possui um modelo tridimensional determinado experimentalmente no banco de dados do Protein Data Bank – PDB, sob o código 1ND1. Esta estrutura foi então selecionada como estrutura-molde para o estudo.

Como a qualidade do modelo a ser gerado depende em grande parte da estrutura da proteína-molde utilizada, é preciso que esta última tenha sido obtida em alta resolução e com um fator-R satisfatório. É preferível utilizar como modelos as estruturas cuja resolução seja igual ou superior a 2 Å e fator-R inferior a 20% (SANTOS FILHO, 2000).

A seqüência de *B.asper* (Uniprot ID P83512.2) possui 408 aminoácidos, sendo um pouco maior em relação ao tamanho da seqüência depositada de *B.moojeni* (Uniprot ID P85314.1), que

possui 205 aminoácidos. Quando alinhadas no aplicativo ClustalW (Servidor Uniprot - <http://services.uniprot.org/clustalw>), ambas tiveram um grande pareamento de resíduos, conforme demonstra a Fig. 19. Na figura 20 é possível visualizar os trechos de resíduos conservados entre as duas seqüências, evidenciando o fato de que, quando maduras, as proteínas apresentam praticamente os mesmos números de resíduos de aminoácidos e a porção N-terminal de VMBP1\_BOTAS é removida por eventos pós-traducionais.

A região N-terminal presente apenas na metaloprotease de *B. asper*, não foi alinhada devido à ausência desta estrutura na metaloprotease de *B. moojeni*, representada no quadro de alinhamento por traços. Estas características de similaridade sugerem que a metaloprotease de *B. moojeni* possa apresentar enovelamento similar ao da metaloprotease de *B. asper madura*. Assim, este recurso de alinhamento torna-se uma importante fonte de informações para a predição da estrutura tridimensional da metaloprotease de *B. moojeni*.

A predição da estrutura secundária da P85314 VMBMP\_BOTMO ajudou na delimitação das regiões de *loops* da seqüência-alvo (em negrito), verificando assim quais resíduos de aminoácidos não são conservados (Fig.21). Essa predição (Fig. 22) foi realizada no servidor SwissModel (ARNOLD *et al*, 2006), disponível em <http://swissmodel.expasy.org>.

B. moojeni	<b>EQQKFS</b> PRYIELVVVADHGMFKKYNLNTIRKVVHEMNVNSMNGFYRSVDVTASLANLEV
B. asper	--ERFSPRYIELAVVADHGIFTKYNSNLNTIRTRVHEMLNTVNGFYRSVDVHAPLANLEV : :*****.*****:*.*****. *****: :***** *.*****
B. moojeni	<b>WSKKDLIN</b> VQKDSRETLKSFGEWRERDLLPRI SHDNAQLLTAIVFDGHTIGRAYTGGMCD
B. asper	<b>WSKQDLIK</b> VQKDSKTLKSFGEWRERDLLPRI SHDHAQLLTAIVFDGNTIGRAYTGGMCD ***:***:***** :*****:*****:*****:*****:*****
B. moojeni	<b>PRHSGVVM</b> DHSPKNLQVAVTMAHELGHNLGMHHDGNQCHCDAASCIMADSLSVLSYEF
B. asper	<b>PRHSGVV</b> RDHSKNNLWVAVTMAHELGHNLGIDHDTGSCSCGAKSCIMASVLSKVLSEYEF ***** ** :** *****:..* ..* *. * *****. ** *****
B. moojeni	<b>SDCSQNQYQ</b> TYLTKHNPQCILNEPL
B. asper	<b>SDCSQNQY</b> ETYLTHNPQCILNKP- *****:****:*****:*

Fig. 20 – Alinhamento da seqüência-alvo (*B. moojeni* P85314.1) e da seqüência-molde (*B. asper* P.83512.2). Obtido através do aplicativo ClustalW (<http://services.uniprot.org/clustalw>).

B.moojeni	<b>EQQKFS</b> PRYIELVVVADHGMEFKKY <b>NSNL</b> NTIRKVVHEMVNSMNGFY <b>RSVDV</b> TASLANLEV
B.asper	--ERFSPRYIELAVVADHGIFTKYNSNLNTIRTRVHEMLNTVNGFYRSVDVHAPLANLEV : :*****.*****:*.******.*****:*: :***** *.*****
B.moojeni	WSK <b>KDLIN</b> VQKDSRET <b>TLK</b> SFG <b>EW</b> RERD <b>LLPR</b> ISHD <b>NAQL</b> LTAIVFD <b>GH</b> TIGRAY <b>TGM</b> CD
B.asper	WSKQDLIKVQKDSKTLK <b>SFG</b> EW <b>RERD</b> LLPRISHDHAQLLTAIVFDGNTIGRAYTGMCD ***:***:***** :*****:*****:*****:*****:*****
B.moojeni	<b>PR</b> HSVGVVMD <b>HSP</b> KNLQVAVTMAHELGHNL <b>GM</b> HHD <b>GN</b> QCH <b>CD</b> AASCIMAD <b>SL</b> SVVLSYEF
B.asper	PRHSVGVVRDHSKNNLWVAVTMAHELGHNLGIDHDTGSCSCGAKSCIMASVLSKVLSYEF ***** ** :* *****:.* ..* *. * *****. ** *****
B.moojeni	<b>SDC</b> SQ <b>NY</b> QTYLT <b>KHN</b> PQ <b>CI</b> LNE <b>PL</b>
B.asper	SDCSQNYETYLTNHNPQCI LNKP- *****:*****:*****:*

Fig. 21 – Delimitação de regiões de loops para a sequência-alvo de *B.moojeni*.

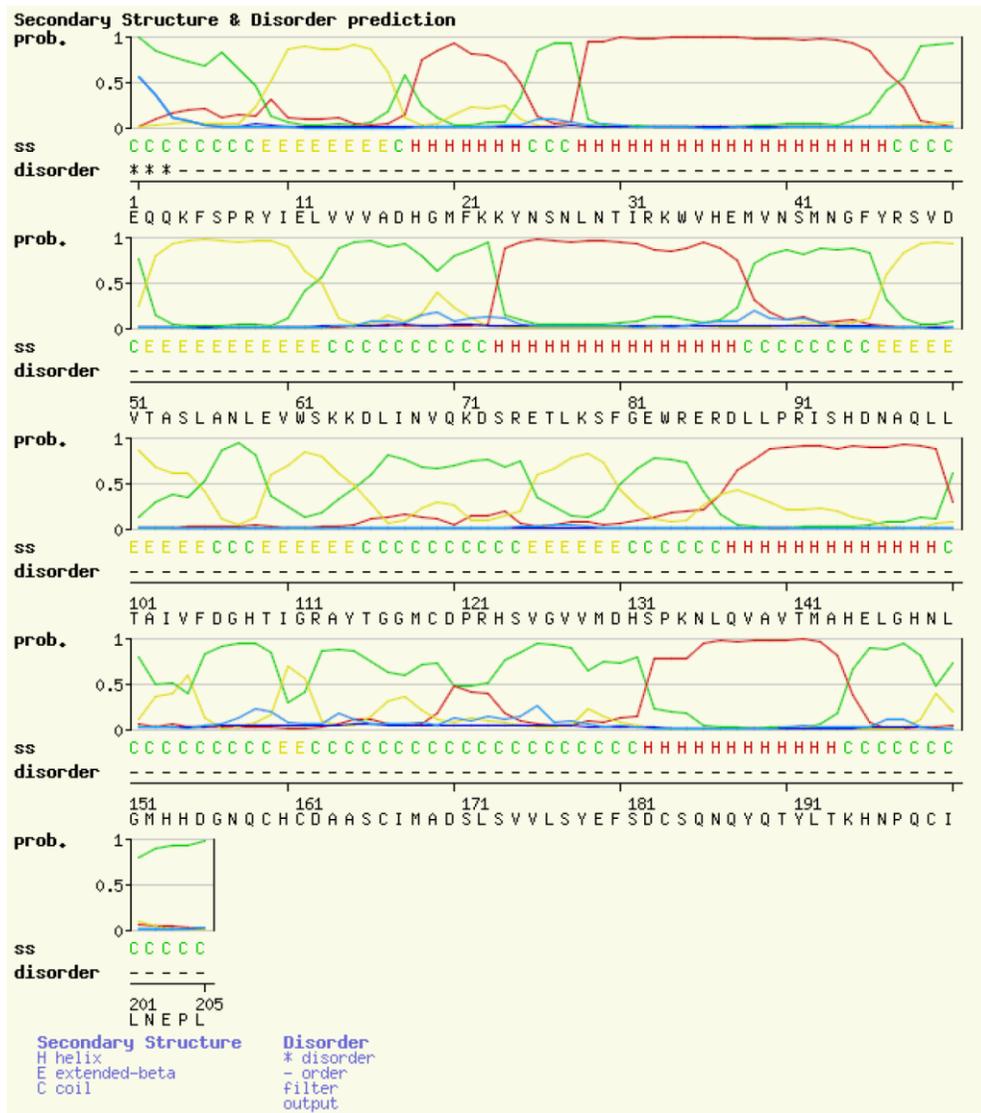
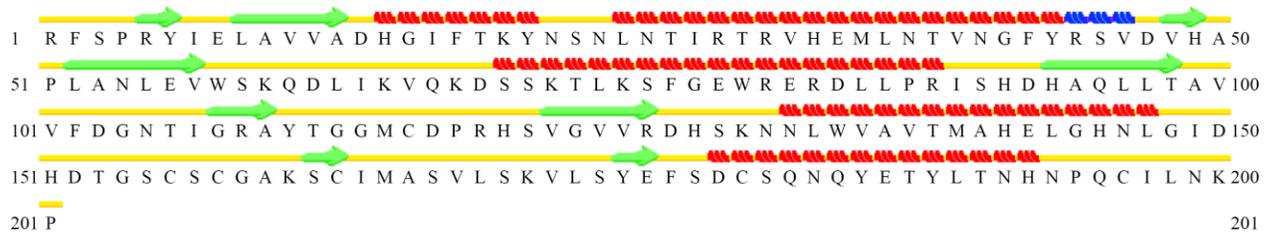
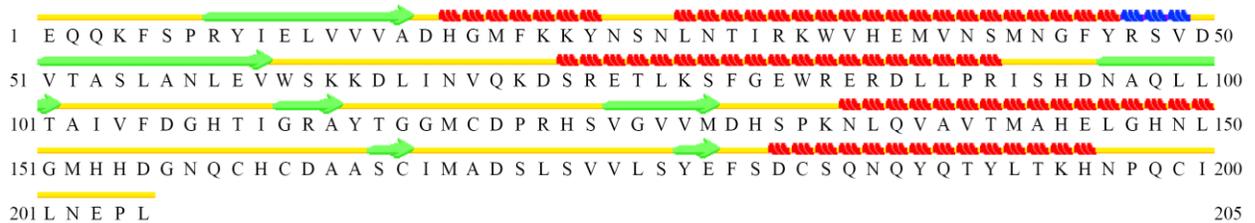


Fig. 22 – Predição da estrutura secundária de VMBMP\_BOTMO de *B.moojeni*.

Sequência Molde: *B. asper*Sequencia Alvo: *B. moojeni*

## Legenda:

	H Alpha-Helix		T Turn
	E Extended Configuration (Beta-sheet)		C or " " Coil
	B Isolated Beta Bridge		G 3-10 Helix
	b Isolated Beta Bridge (Type 3 Fig 4,cd)		I Pi-Helix

A partir da análise de similaridade da sequência de resíduos da metaloprotease de *B. asper* (VMBP1\_BOTAS), foi selecionado o arquivo “\*.pdb” desta sequência: 1ND1.pdb. O arquivo comporta as coordenadas para a estrutura tridimensional da metaloprotease P83512 VMBP1\_BOTAS, que serviu como arquivo de entrada no aplicativo MODELLER para a modelagem tridimensional. Além disto, foi necessário fornecer também ao aplicativo a sequência-alvo da metaloprotease de *B. moojeni*.

Foram obtidos 05 modelos estruturais, a partir do procedimento de modelagem por homologia com o aplicativo MODELLER, em formato de extensão PDB. Estes modelos estruturais obtidos passaram por etapas de validação, antes da escolha definitiva de um modelo estrutural adequado para a representação da metaloprotease estudada.

O reconhecimento dos erros em modelos, experimentais e teóricos, de estruturas de proteínas é um problema importante em biologia estrutural (SIPPL, 1993). Ao construir um modelo de uma proteína, com ou sem o auxílio de informações experimentais, muitas vezes é necessário o uso de uma medida independente, a fim de validar a exatidão do modelo (WALLNER & ELOFSSON, 2007).

Os modelos obtidos neste trabalho, através do aplicativo MODELLER por satisfação de restrições espaciais, foram avaliados quanto a sua estabilidade de enovelamento, visando o refinamento futuro da predição estrutural. A função OBJECTIVE FUNCTION, presente no aplicativo MODELLER, retorna um valor médio de energia do modelo gerado (Tabela 5).

Este valor médio de energia indica quão próximo está o modelo tridimensional teórico do seu respectivo molde. Quanto menor o valor médio de energia atribuído ao modelo da metaloprotease de *B.moojeni*, mais próximo estatisticamente está este modelo tridimensional teórico da estrutura da metaloprotease de *B.asper*. Ainda, considerando as características físico-químicas que cada resíduo apresenta na seqüência-alvo e a sua correspondente estrutura mais estável possível, em conformação tridimensional. A partir deste valor médio de energia, é possível verificar qual dos cinco modelos possui uma menor energia média apresentada, ou seja, uma maior proximidade com o molde (*B.asper*).

Tabela 5 – Valores médios de energia para os modelos gerados.

MODELO	P1	P2	P3	P4	P5
MODELLER OBJECTIVE FUNCTION	947.90	995.19	964.42	1004.74	1063.38

Apesar de o modelo P1 ter oferecido o menor valor médio de energia, o que de início indicaria que este modelo seria escolhido como o modelo final, todos os cinco modelos foram submetidos aos servidores de validação estrutural: Molprobitry (DAVIS *et al*, 2007), Harmony

(PUGALENTHI *et al*, 2006), ProSAWeb (WIEDERSTEIN & SIPPL, 2007) e Verify3D (LUTHY *et al*, 1992).

Para fins de discussão dos resultados, apenas o modelo escolhido terá seus resultados expostos nesta seção. Os resultados de validação, obtidos a partir dos outros quatro modelos, estão dispostos na seção de anexos.

Submetendo então os modelos gerados por modelagem molecular ao servidor Molprobitry (<http://molprobitry.biochem.duke.edu/index.php>), foi possível analisar o gráfico de Ramachandran de cada um destes modelos para verificar as conformações adotadas pelas cadeias principais dos resíduos de aminoácidos e identificar possíveis desvios ou erros tendo como base a estereoquímica dos grupos químicos envolvidos (Fig. 23).

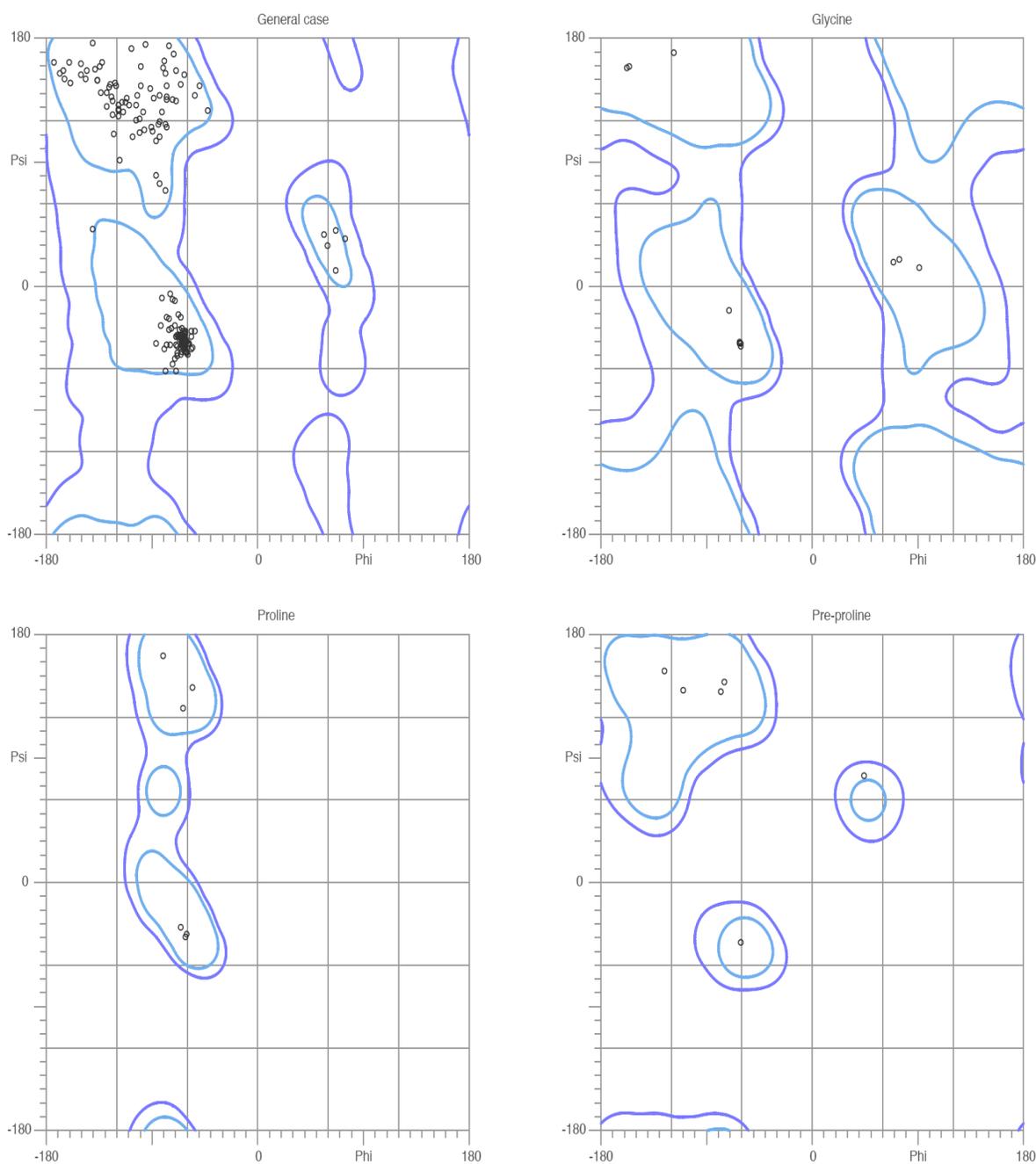


Fig. 23 – Gráficos de Ramachandran fornecidos pelo Molprobability para o modelo P1.

Realizando uma comparação dos resultados obtidos com o uso do Molprobability (Tabela 6), foi verificado que o modelo P3 apresentaria uma maior adequação de seus parâmetros de qualidade, em relação ao intervalo de qualidade ideal requerido para o enovelamento de uma

estrutura protéica. Em virtude de apresentar desvios em dois carbonos de cadeia lateral (carbono beta), este modelo não foi aceito como adequado. Por sua vez, o modelo P4 também apresenta resultados satisfatórios quanto à geometria e quiralidade dos resíduos de aminoácidos da estrutura molecular, porém apresenta um índice de 2,73% para rotâmeros inadequados. Esta porcentagem diminuiu então a possibilidade deste ser o modelo escolhido para representar a metaloprotease em questão. Por último, temos o modelo P1 apresentando um gráfico de Ramachandran favorável, resíduos com ligações e ângulos adequados, apesar de ainda apresentar desvio em um carbono da cadeia lateral (carbono beta). A porcentagem de rotâmeros em conformações inadequadas (1,09%), apesar de estar acima do intervalo ideal (máximo de 1%), também pode ser considerada aceitável. O modelo P1 mostrou-se então o modelo mais aceitável para a conformidade de qualidade, sendo assim um candidato a representar a estrutura tridimensional da metaloprotease.

Tabela 6 – Resultados de qualidade geométrica e estereoquímica, obtidos por Molprobitry.

	IDEAL	P1	P2	P3	P4	P5
Rotâmeros inadequados	<1%	1.09%	1.64%	0.55%	2.73%	2.19%
Ramachandran favorável	>98%	98.52%	97.54%	98.03%	98.52%	97.04%
desvios de $C_{\beta} > 0.25\text{\AA}$	0	1	1	2	1	2
Resíduos com ligações inadequadas	<1%	0.00%	0.00%	0.00%	0.00%	0.00%
Resíduos com ângulos inadequados	<0.5%	0.49%	0.00%	0.00%	0.00%	0.49%

Para a análise de erros grosseiros e locais na conformação estrutural, os cinco modelos foram submetidos ao servidor Harmony (<http://caps.ncbs.res.in/harmony/>). Os arquivos de entrada em cada execução do programa foram justamente um dos modelos em formato “.pdb” e o arquivo de alinhamento de extensão “.ali” entre a seqüência-alvo (*B. moojeni*) e a seqüência-molde (*B. asper*). Este servidor permite também o uso de arquivos de entrada com alinhamentos múltiplos, mas isto não foi utilizado neste trabalho.

No servidor Harmony, cada um dos cinco modelos gerados para a metaloprotease VMBMP\_BOTMO são comparados automaticamente com a estrutura enovelada de outras

proteínas. Essas proteínas que são comparadas aos modelos são de número de resíduos aproximados (cerca de 205 aminoácidos), e foram captadas em banco de dados de proteínas do SwissProt, para a realização do procedimento. A cada um dos enovelamentos das moléculas de comparação, é atribuído um índice de propensão à substituição de seus resíduos de aminoácidos.

A partir da disposição da pontuação de todos os enovelamentos, é estabelecida uma curva de calibração. Nem todas as proteínas de mesmo número de resíduos são comparadas com a seqüência-alvo e são mostradas no gráfico. Apenas as proteínas com maior porcentagem de identidade com o modelo proposto são visualizadas em verde no gráfico. O modelo proposto é visualizado em azul. A boa qualidade de enovelamento de um modelo é determinada se este se encontrar próximo à curva de calibração, conforme demonstrado no modelo P1 (Fig.24).

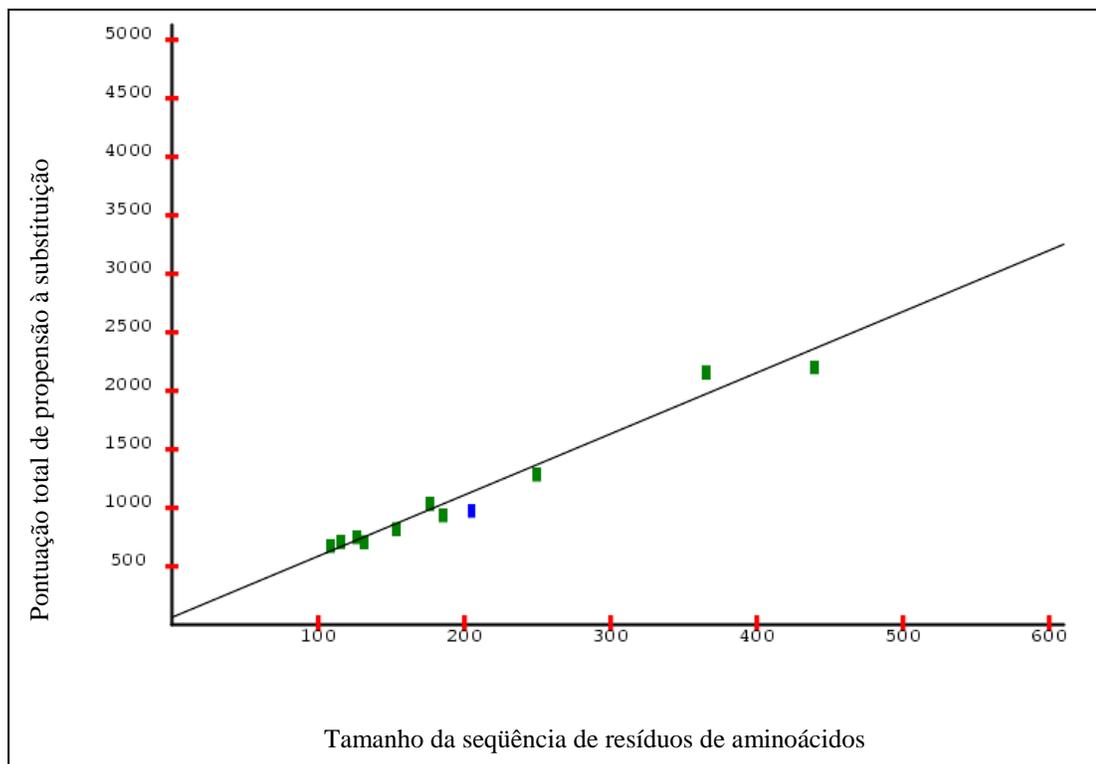


Fig. 24 – Gráfico de detecção de erros grosseiros no enovelamento do modelo P1.

Além disso, uma representação gráfica apresentada pelo HARMONY fornece as pontuações dos modelos com base na seqüência-alvo direta bem como na seqüência-alvo em ordem inversa dos resíduos de aminoácidos. Assim, o resultado obtido com a seqüência inversa pode ser usado como um controle para identificar erros locais nos modelos obtidos. As regiões onde a seqüência inversa (vermelho) adquire maior pontuação de substituição que a seqüência na ordem direta (verde) podem apresentar possíveis erros locais, conforme demonstrado para o modelo P1 obtido (Fig.25).

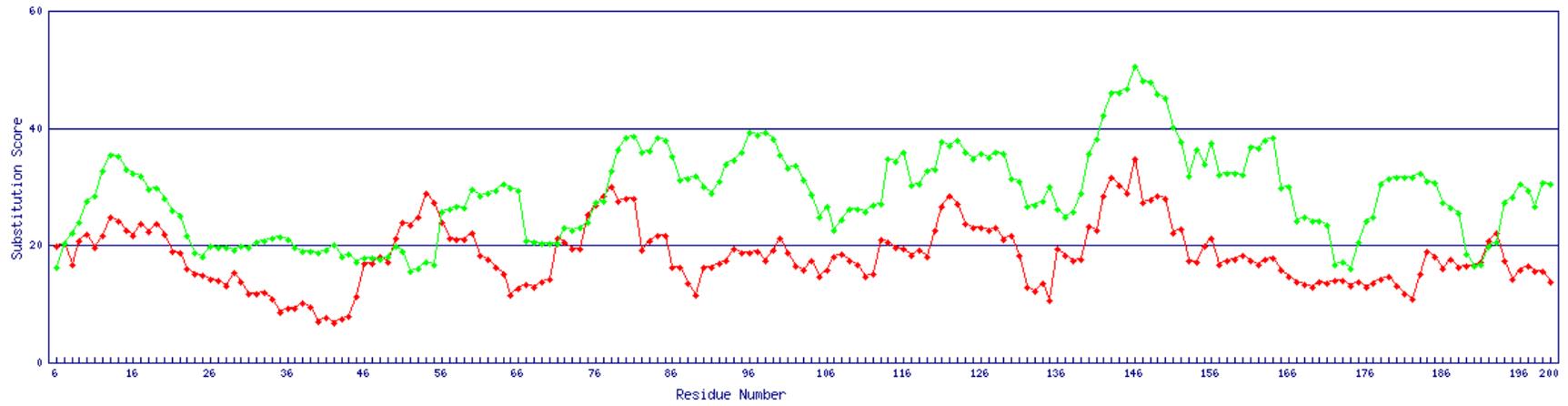


Fig. 25 – Detecção de erros locais no modelo P1 (verde), através do comparativo com a seqüência reversa de aminoácidos (vermelho).

Ainda, na Tabela 7 foi feita uma comparação entre as pontuações de propensão (erros grosseiros no enovelamento) e substituição (erros locais no enovelamento) dos modelos obtidos. Neste comparativo, é verificado que o modelo P4 possui maior valores médios de pontuação de propensão, quando comparado aos valores médios de propensão e substituição encontrados para os outros modelos analisados. Isso indica que o modelo P4 apresenta uma maior tendência a apresentar erros grosseiros no seu enovelamento, como erros de localização espacial das estruturas terciárias (alfa-hélices, beta-folhas e alças).

Por sua vez, o modelo P5 apresenta uma maior propensão à substituição de resíduos de aminoácidos, o que indica que há uma grande possibilidade de em determinadas posições de resíduos de aminoácidos, existam alguns que podem ser substituídos aleatoriamente por um outro resíduo de aminoácido. Este dado é significativo, pois um resíduo mal-posicionado no espaço pode causar a impressão de que outro aminoácido, que eventualmente possa vir a substituí-lo, poderia ser a melhor opção para o enovelamento da proteína.

O modelo que apresenta valores médios mais satisfatórios é o modelo P1, tanto para a pontuação de propensão, quanto para a pontuação de substituição.

Assim, o modelo mais adequado seria o modelo P1, o qual apresenta resultados mais satisfatórios: uma pontuação de propensão mais baixa, o que significa que a estrutura tridimensional proposta apresenta poucos erros de posicionamento de estruturas terciárias, isso em relação a estruturas similares e com estrutura tridimensional já definida; e uma propensão de substituição mais baixa, elucidando que a substituição aleatória de resíduos de aminoácidos na estrutura não necessariamente contribuiriam para um enovelamento de melhor qualidade, que a proposta pelo modelo P1.

Tabela 7 – Comparativo para os resultados dos cinco modelos, obtidos no servidor Harmony.

	<i>B.moojeni</i>				
	P1	P2	P3	P4	P5
Pontuação de propensão	973.49	990.19	994.09	996.79	989.69
Pontuação de substituição	5263.53	5703.74	5317.68	5257.15	5282.68

As pontuações de propensão e de substituição mostraram que as variações de posição dos resíduos de aminoácidos devem ser as mais baixas possíveis. Se houvesse presença de regiões demarcadas em amarelo, nas estruturas dos modelos, indicariam regiões com potenciais problemas estruturais, justamente nas posições dos carbonos alfa (cadeia principal) e beta (cadeias laterais). A representação do modelo P1 foi o que apresentou menores valores de propensão e de substituição (Fig.26). As representações tridimensionais dos outros modelos restantes são visualizadas em anexo.

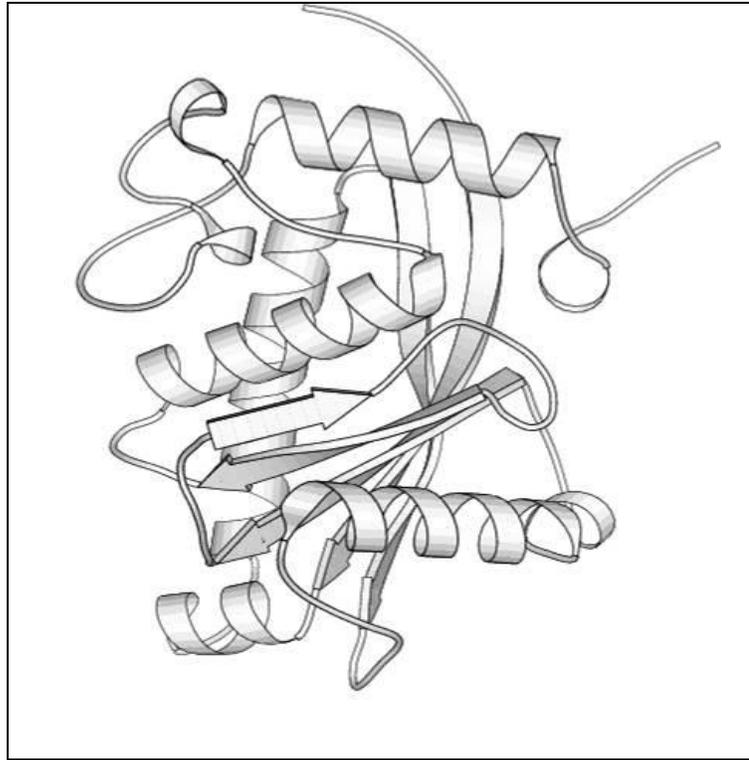


Fig. 26 – Estrutura tridimensional sugerida pelo Harmony para o modelo P1.

Para a análise da qualidade global de energia, os modelos também foram submetidos ao servidor ProSA-web (<https://prosa.services.came.sbg.ac.at/prosa.php>). O parâmetro de qualidade a ser analisado foi o *z-score* (índice de qualidade global), que revela se o modelo estrutural proposto está dentro de um intervalo de minimização de energia aceitável, quando comparado ao *z-score* de proteínas nativas de tamanho semelhante (Fig.27). Os outros modelos analisados, bem

como os resultados fornecidos pelo ProSA-web encontram-se em anexo. As proteínas, que servem de parâmetro de comparação, foram obtidas por cristalografia de raios X (representação pela região em azul-claro) e por ressonância magnética nuclear (representação pela região em azul-escuro).

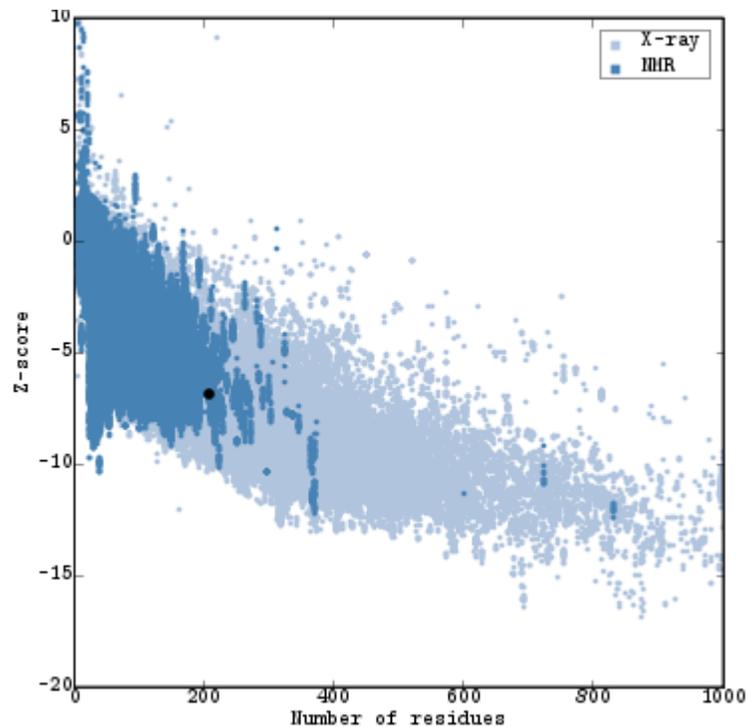


Fig. 27 – Qualidade global estimada pelo ProSA-web para o modelo P1.

A partir de uma análise comparativa dos  $z$ -scores fornecidos pelo servidor ProSA-web, foi possível observar que, considerando o parâmetro de minimização de energia, o modelo P1 obteve o mais próximo valor de  $z$ -score em relação ao molde (Tabela 8).

Tabela 8 – Comparativo dos *z-scores* obtidos pelo ProSA-web.

	<i>B.asper</i>	<i>B.moojeni</i>				
Modelo	Molde	P1	P2	P3	P4	P5
z-score	-6,74	-6,82	-6,84	-6,99	-6,89	-7,00

Outro resultado fornecido pelo servidor é o índice da qualidade local do modelo, também em função da energia da estrutura tridimensional. Porém, a análise agora é em função da energia de cada aminoácido da seqüência primária, em relação à seqüência tridimensional do modelo.

Observando o gráfico (Fig.28), em geral, os valores positivos correspondem a partes problemáticas ou errada da seqüência-alvo. Como há flutuações na parcela individual dos resíduos de aminoácido, a curva é suavizada pelo cálculo da energia média de cada fragmento. Primeiramente, foi calculada a energia média do resíduo de aminoácido, este contido na posição central de uma janela de tamanho 40 (verde escuro). Uma segunda curva, com um tamanho de janela de 10 resíduos, é resultante do cálculo da energia média do resíduo central (verde claro). Os cinco modelos se mostraram satisfatórios em relação aos erros locais.

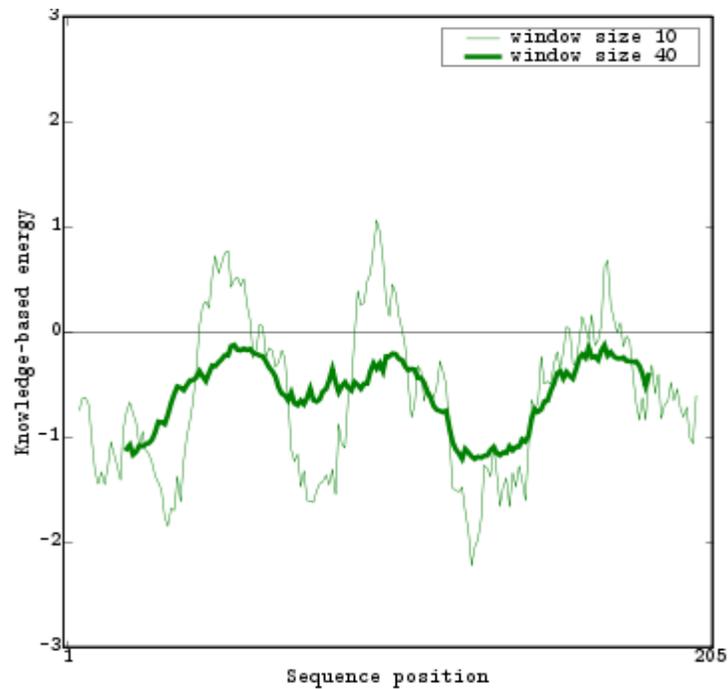


Fig. 28 – Qualidade local estimada para o modelo P1.

Os modelos protéicos também puderam ser avaliados com relação à qualidade dos ambientes químicos, com o objetivo de se determinar o nível de confiabilidade do enovelamento protéico do modelo sugerido (Figs. 54, 55, 56, 57 e 58). As pontuações, fornecidas pela análise do servidor Verify3D, determinam os ambientes químicos de cada resíduo, com referência a uma matriz construída a partir de uma análise estatística envolvendo estruturas de proteínas do PDB, que apresentam alta identidade com a seqüência-alvo. A seqüência molde também foi submetida ao servidor Verify 3D (Fig.29), para efeito comparativo.

Detectando as regiões de baixa qualidade, foi feita então uma correspondência da seqüência primária de resíduos de aminoácidos com a estrutura tridimensional sugerida pela modelagem. Os índices de correspondência das seqüências 3D-1D obtidos foram satisfatórios, chegando a um mínimo de 1% e um máximo de 64%. A média de adequabilidade da correspondência está por volta de 45%, que é média encontrada para estruturas de boa qualidade (LUTHY *et al*, 1992). O molde, por sua vez, apresenta um intervalo de 23% a 66%.

A partir dos resultados gerados pelo Verify3D, o modelo P1 se mostrou o mais adequado quanto à correspondência dos resíduos de aminoácidos de sua estrutura tridimensional, em relação à seqüência primária, uma vez que o intervalo da curva foi de 7% a 63%.

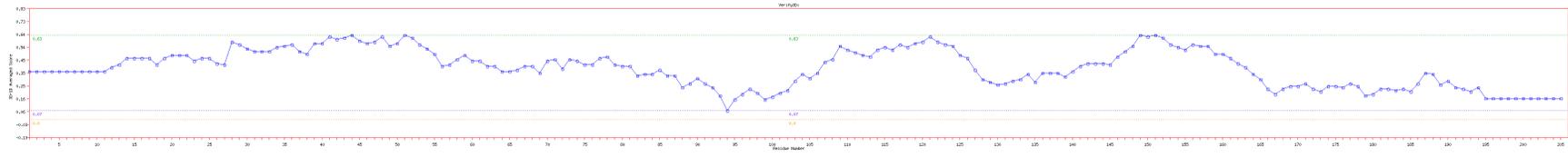


Fig. 29 – Correspondência de seqüências 3D-1D para o modelo P1.

## 6 CONSIDERAÇÕES FINAIS E PERSPECTIVAS FUTURAS

Neste trabalho, foi obtida computacionalmente a estrutura tridimensional preliminar *in silico* para a representação da metaloprotease P85314 de *B.moojeni*. A representação tridimensional permite a visualização e análise de regiões de interesse estrutural, como *loops* e regiões de não-conservação da estrutura, quando comparadas à estrutura que serviu de molde. Essas regiões demarcadas ainda permitem uma análise sobre a possibilidade de acesso ao sítio ativo, de forma a influenciar a atividade não-hemorrágica da biomolécula.

Diante dos resultados obtidos nos processos de validação estrutural, o modelo mais adequado para representar a estrutura tridimensional da VMBMP\_BOTMO é o modelo P1. Apesar dos resultados satisfatórios, esta representação estrutural é um modelo teórico, obtida de modo computacional. Recentemente, Akao (*et al*, 2009) obteve por métodos experimentais de uma isoforma da metaloprotease VMBMP\_BOTMO, realizando estudos estruturais a partir da análise de seu modelo cristalográfico, denominado agora de BmooMPa-I, com 1,76Å de resolução. Este modelo experimental serviu então como parâmetro adicional de validação estrutural para o modelo teórico obtido computacionalmente ao longo do trabalho de pesquisa aqui realizado.

A qualidade geométrica e estereoquímica do modelo teórico de VMBMP\_BOTMO foi então comparada com o modelo experimental BmooMPa-I da mesma biomolécula, servindo então como um estudo mais claro acerca das propriedades estruturais da referida metaloprotease. A sobreposição do modelo teórico com o modelo experimental foi realizada no aplicativo Pymol, ferramenta de visualização de modelos, demonstrando assim a pouca divergência espacial entre o modelo obtido experimentalmente e o modelo obtido por metodologias de modelagem comparativa (Fig.30).

A sobreposição do modelo teórico da estrutura tridimensional VMBMP\_BOTMO contra a seqüência-molde (metaloprotease do veneno de *B.asper*) também foi executada no aplicativo Pymol (Fig.31). Visualmente, é possível perceber que as regiões de *loops* são as regiões que

divergem entre a seqüência-molde (*B. asper*) e a seqüência-modelo final (*B. moojeni*), conferindo então as características estruturais peculiares a cada espécie.

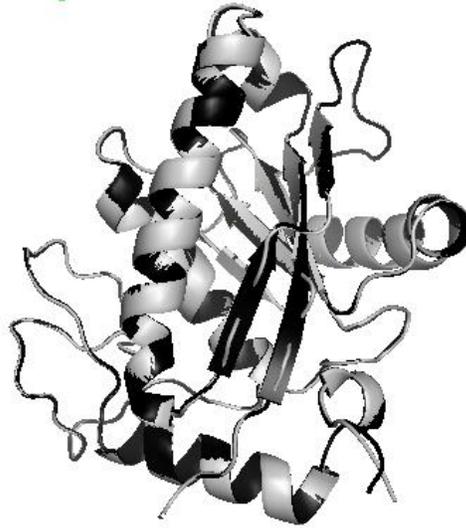


Fig. 30 – Sobreposição de modelo teórico obtido de metaloprotease *B. moojeni* (VMBMP\_BOTMO) e o modelo experimental (BmooMPa-I).

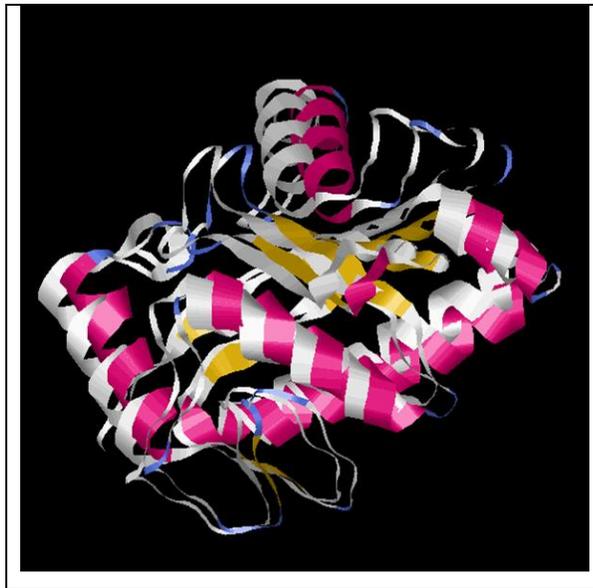


Fig. 31 – Sobreposição de modelo obtido de metaloprotease *B. moojeni* P85314 (colorido) e *B. asper* P83512 (cinza) como molde utilizado.

As posições dos resíduos cataliticamente importante presentes no pacote do sítio ativo, em comparação do modelo teórico e o modelo experimental BmooMPa-I, são encontradas totalmente conservadas, incluindo o *motif* HEXXHXXGXXH (ponto de ligação do zinco) e o “metionina-turn” na parte inferior das histidinas catalíticas. Para efeito comparativo com o modelo obtido computacionalmente neste trabalho, na estrutura de BmooMPa- I, o íon zinco apresenta uma coordenação octaédrica distorcida, formada pelos átomos de nitrogênio das histidinas catalíticas (His<sup>142</sup>, His<sup>146</sup> e His<sup>152</sup>), numeração esta com base na estrutura de adamalysina II (GOMIS-RÜTH *et al.*, 1994) e, adicionalmente, três moléculas de solvente (OW<sup>131</sup>, OW<sup>143</sup> e OW<sup>144</sup>) (Fig. 32). É possível perceber que a estrutura obtida por métodos teóricos é bastante semelhante estruturalmente ao modelo experimental, a partir da visualização na Fig. 33.

Em outras metaloproteases, como H<sub>2</sub>-protease, adamalysina-II e acutolysina A, o íon zinco é coordenado tetraedricamente por pelo mesmos três histidinas catalíticas, e apenas uma molécula de solvente, que é estruturalmente relacionado com o residuo de aminoácido Glu<sup>143</sup>, que tem a função de polarizar este solvente antes do processo químico de ataque nucleofílico sobre a ligação peptídica do substrato de uma cadeia polipeptídica (Fig. 34).

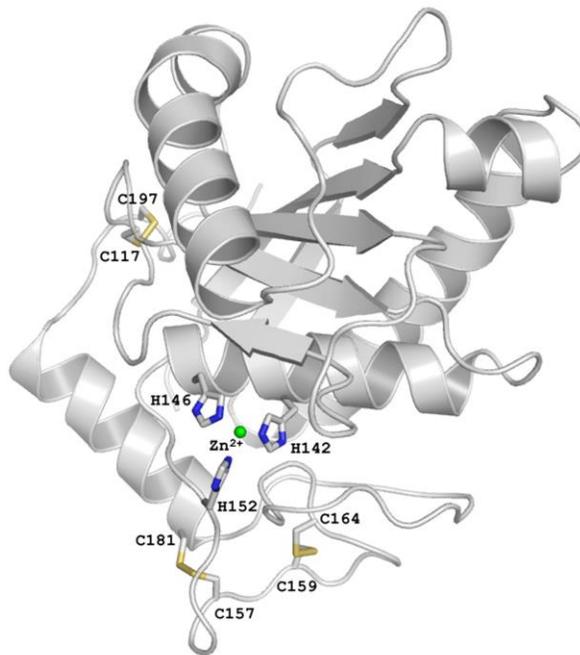


Fig. 32 – Representação esquemática do modelo experimental de BmooMPa-I, elucidando as posições de pontes dissulfeto, histidinas catalíticas e metal zinco.

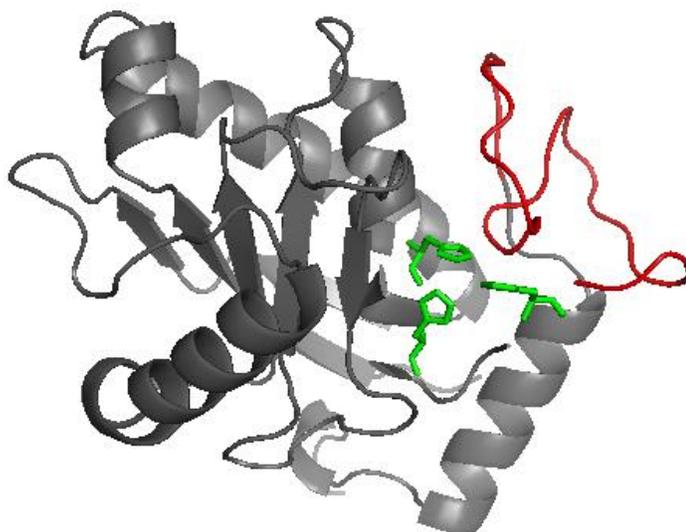


Fig. 33 – Representação esquemática do modelo teórico de VMBMP\_BOTMO: o *loop* catalítico (vermelho), as histidinas do sítio catalítico (verde).

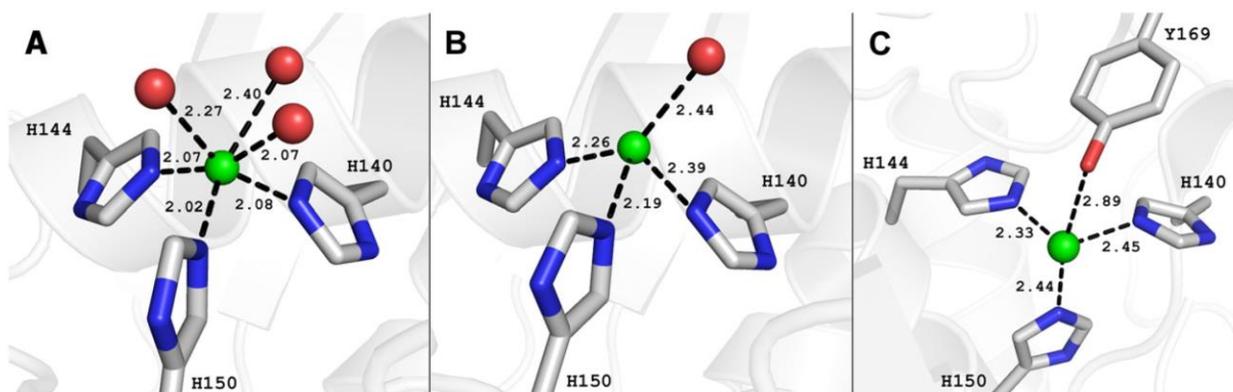


Fig. 34 – Representação esquemática do sítio ativo e a participação do íon zinco nas estruturas da metaloprotease de *B.moojeni* (A), na adamalisina-II (B) e na astacina (C). Os resíduos de aminoácidos interagentes com o íon zinco são representados em esferas. As interações do zinco catalítico com as histidinas são representadas em pontilhado.

A molécula  $OW^{143}$  correspondente no solvente da BmooMPa-I, em efeito comparativo ao modelo obtido computacionalmente neste trabalho, é coordenada por um vínculo de curta duração (distância de  $2,79\text{\AA}$ ), quando comparado com outras metzincinas já descritas, sendo

consideradas então distâncias normais, que variam de 3,52-4,3Å que já foram relatados em casos como este. Um modo diferente da coordenação de zinco também já foi descrito na estrutura de astaceno, em que a interação típica com uma molécula de água passa a ser substituída por um grupo hidroxila da cadeia lateral Tyr<sup>169</sup> (BODE *et al.*, 1992).

Segundo ainda Akao (*et al.*, 2009), a sobreposição entre as estruturas das proteases BmooMPa-I, adamalisina II e estruturas de H<sub>2</sub>-proteases (classe PI não-hemorrágicas de metaloproteases) sobre as estruturas de BAP1, acutolisina A e acutolisina C (Classe P-I hemorrágicas) indicaram que tanto os subgrupos conservam todos os elementos considerados essenciais para o processo de catálise enzimática, incluindo as três histidinas do pacote catalítico, o resíduo de aminoácido polarizador Glu<sup>143</sup> e o “*metionina-turn*” nas mesmas conformações relativas entre as estruturas.

Ainda, uma análise de interfaces de estruturas com o servidor “Protein Interfaces, Surfaces and Assemblies”, realizada no trabalho de Krissinel & Henrick (2007), revela que ambos os subgrupos apresentam áreas de exposição da superfície acessível, com extensão de 9500Å, áreas de exposição estas muito semelhantes entre si, além de um efeito de solvatação com energia livre de cerca de 190 Kcal/mol. As diferenças são apenas significativas em *loops* adjacentes (segmentos de aminoácidos 153-164 e 167-176) para a “*metionina-turn*”, e em ambos os nitrogênios e carbonos terminais, devido a deleções de resíduos de aminoácidos (Fig.35).

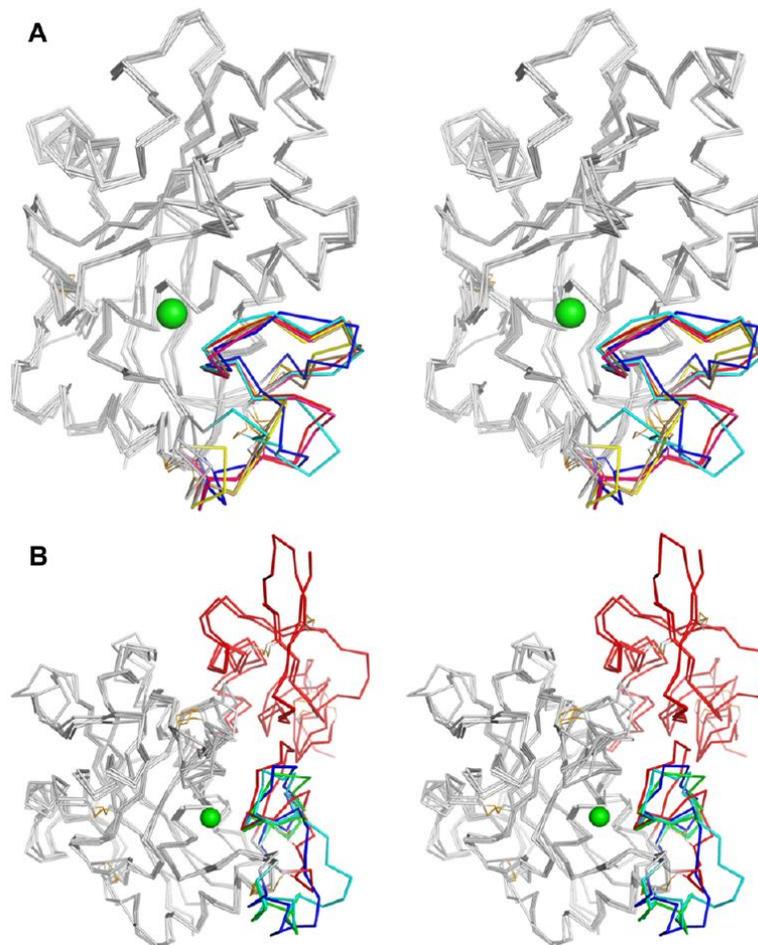


Fig. 35 – Representação esquemática da sobreposição das cadeias principais (*backbones*) das metaloproteases e das adamalysinas, revelando a divergência estrutural no *loop* catalítico quanto a classificação em hemorrágica e não-hemorrágica.

A variabilidade apresentada nas regiões de nitrogênio e carbono terminais, bem como a distribuição de cargas na superfície das estruturas foram analisadas, e não apresentaram qualquer informação que possa diferenciar entre os dois subgrupos (hemorrágicas e não-hemorrágicas). Assim, analisando então a relação da identificação de regiões funcionais das estruturas e comparando os subgrupos entre si, é sugerido que o *motif*, formado pelos resíduos de aminoácidos 153-164 e 167-176, os quais são adjacentes a “metionina-turn”, seja um padrão coerente de composição de resíduos de aminoácidos, o qual permite diferenciar os dois subgrupos.

Conseqüentemente, estes também podem estar envolvidos no desenvolvimento de atividade hemorrágica no subgrupo PI. Nesta hipótese, as estruturas não-hemorrágicas apresentam como diferenciação alguns resíduos de aminoácidos volumosos, que podem então provocar confrontos estéricos, impedindo o reconhecimento do substrato macromolecular. Já no subgrupo de estruturas hemorrágicas, este motivo é composto por resíduos de aminoácidos pequenos e polares, como Ser e Thr, permitindo o processo de vinculação com o substrato e o reconhecimento de metas no processo de indução hemorrágica. Estudos anteriores também já haviam sugerido que a região 153-176, na estrutura principal das metaloproteases, seria a principal diferença entre as manifestações de atividades hemorrágicas (WATANABE et al., 2003), confirmando assim as observações aqui realizadas neste comparativo.

Assim, o modelo proposto neste trabalho, obtido por métodos teóricos de modelagem comparativa, se aproxima muito do modelo obtido experimentalmente, de forma a validar a metodologia como segura e eficaz para a predição de estruturas tridimensionais ainda não elucidadas experimentalmente. O modelo teórico demonstrou boa qualidade de enovelamento e energia, apresentando até uma energia mais baixa que o modelo experimental. Porém, como o modelo experimental sofre influências do entorno, o modelo teórico termina por apresentar uma energia um pouco mais alta que o modelo teórico obtido.

Como perspectivas futuras, são almejadas a análise mais criteriosa das propriedades estruturais da metaloprotease de *B.moojeni* e sua interação com o entorno. Ainda, serão propostas as regiões de epítomos, pela característica da hidrofobicidade da cadeia (potencial exposição da biomolécula à água).

Epítopo é a menor parte de um antígeno, que é potencialmente capaz de estimular resposta imune ao se ligar a um anticorpo. São áreas nas moléculas dos imunógenos, que se ligam aos receptores celulares e aos anticorpos. Um antígeno pode apresentar vários epítomos. Essas regiões também são interessantes para estudos, pois servem como uma nova possibilidade de que a metaloprotease em questão seja utilizada como um indutor da resposta imune em um organismo, quando este for afetado por alguma patologia de difícil tratamento. Uma vez utilizada como fármaco indutor de resposta imune, a metaloprotease de *B.moojeni* poderá ser um potencial auxiliar para os organismos afetados, no processo de cura destas patologias.

## 7 REFERÊNCIAS

- ACHARYA, K. R., STUART, D. I., WALKER, N. P., LEWIS, M. & PHILLIPS, D. C. Refined structure of baboon alpha-lactalbumin at 1,7 Å resolution. Comparison with C-type lysozyme. *Journal Molecular Biology*. N.208: 99-127. 1989.
- AKAO, P. K., TONOLI, C. C. C., NAVARRO, M. S. CINTRA, A. C. O., NETO, J. R. ARNI, R. K. & MURAKAMI, M. T. Structural studies of BmooMPa-I, a non-hemorrhagic metalloproteinase from *Bothrops moojeni* venom. *Toxicon*, 2009.
- ANFINSEN, C. B. Principles that govern the folding of protein chains. *Science* N. 181 (96): 223–230. 1973.
- ARCE, V., AZOFEIFA, G., FLORES M. & ALAPE, A. Molecular cloning and sequence analysis of a type I metalloproteinase from *Bothrops asper* snake venom. EMBL/GenBank/DDBJ databases, 2005.
- ARNOLD, K., BORDOLI, L., KOPP, J., & SCHWEDE, T. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics*, 22,195-201, 2006.
- ATKINS, P. & PAULA, J. *Physical Chemistry for Life Science*. New York: W. H. Freeman, 2005
- BERG, J. M., TYMOCZKO, J. L. & STRYER, L. *Biochemistry*. 5ed. New York: W. H. Freeman & Co., 2002.
- BAKER, D. & SALI, A. Protein structure prediction and structural genomics. *Science* N. 294(5540):93-96, 2001.
- BERNARDE, P. S. *Acidentes ofídicos*. Rio Branco: UFAC, 2009.
- BERNARDES C.P., SANTOS-FILHO, N.A., COSTA, T.R., GOMES, M.S.R., TORRES, F.S., BORGES, M.H., RICHARDSON, M., SANTOS, D.M., PIMENTA, A.M.C., HOMSI-BRANDEBURGO, M.I., SOARES, A.M., OLIVEIRA, F. Isolation of a new metalloproteinase from *Bothrops moojeni* snake venom: functional and structural characterization. UniProtKB Database, 2007.
- BERNARDES C. P., SANTOS-FILHO N. A., MAMEDE C. C. N., COSTA, J. O., OLIVEIRA F. PURIFICATION AND PARTIAL CHARACTERIZATION OF A METALLOPROTEINASE FROM *Bothrops moojeni* (CAIÇACA) VENOM. In: *Anais do IX Symposium of the Brazilian Society on Toxinology*. *J. Venom. Anim. Toxins incl. Trop. Dis.*, , 13, 1, p.343, 2007.
- BJARNASON, J. B. & FOX, J. W. Hemorrhagic metalloproteinases from snake venoms. *Pharmacology and Therapeutics*, 62, 1994.
- BRAYER, G. D., DELBAERE, L. T. & JAMES, M. N. Molecular structure of the alpha-lytic protease from *Myxobacter 495* at 2,8 angstroms resolution. *Journal of Molecular Biology*. N.131: 743-775. 1979.
- BODE, W., GOMIS-RÜTH, F.X., HUBER, R., ZWILLING, R. & STOCKER, W. Structure of astacin and implications for activation of astacins and zinc-ligation of collagenases. *Nature* N. 358, 164–167, 1992
- BROWNE, W. J., NORTH, A. C. T., PHILLIPS, D. C., BREW, K. VANAMAN, T. C. & HILL, R. C. A possible three-dimensional structure of bovine alpha-lactalbumin based on that of hen's egg-white lysozyme. *Journal of Molecular Biology*. N.42:65. 1969.
- CLAVERIE, J. M. & NOTREDAME, C. *Bioinformatics*. Indianapolis: Wiley, 2007.
- COX, R. A. & BONANOU, S. A. A possible structure of the rabbit reticulocyte ribosome. An exercise in model building. *Biochemistry Journal*. N. 114. p.769-774. 1969.

DAVIS, I. W., LEAVER-FAY, A., CHEN, V. B., BLOCK, J. N., KAPRAL, G. J., WANG, X., MURRAY, L. W., ARENDALL, W. B., SNOEYINK, J., RICHARDSON, J. S. & RICHARDSON, D. C. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res.* N. 35 (Web Server issue):W375-83, 2007.

DELBAERE, L. T., BRAYER, G. D. & JAMES, M. N. Comparison of the predicted model of alpha-lytic protease with the x-ray structure. *Nature* 279: 165-168.

Serpentes peçonhentas: família *elapidae* e família *viperidae*. Catálogo do Museu do Instituto BUTANTAN, 2000.

FARSKY, S.H.P., GONÇALVES, L.R.C., GUTIERREZ, J.M., CORREA, A.P., RUCAVADO, A., GASQUE, P., TAMBOURGI, D.V. *Bothrops asper* snake venom and its metalloproteinase BaP-1 activate the complement system. Role in leucocyte recruitment. *Mediators Inflamm.* 9:213-221, 2000.

FIGUEIREDO, J. E. F., ROCHA, W. D. & COELHO, V. T. S. Análise computacional para estudo da genômica funcional: identificação do domínio de ligação de calmodulina em proteínas HSP70 e BiP. Embrapa (comunicado técnico) No. 63, outubro – 2003.

FRANÇA, J. B., HAMAGUCHI, A., RODRIGUES, V. M. & BRANDESBURGO, M. I. H. Estudo Comparativo de Atividades Biológicas Presentes nas Peçonhas de Três Serpentes Brasileiras: *Bothrops moojeni*, *Bothrops neuwiedi pauloensis* e *Crotalus durissus* sp. Disponível em: <[www.propp.ufu.br/revistaelectronica/Edicao%202006\\_1/B/Johara%20Boldrini.pdf](http://www.propp.ufu.br/revistaelectronica/Edicao%202006_1/B/Johara%20Boldrini.pdf)>. Acesso em dez. 2008

FRAUENFELDER, H. & WOLYNES, P. G. Biomolecules: where the physics of complexity and simplicity meet. *Physics Today*, N. 01, 1999.

GINALSKI, K. Comparative modeling for protein structure prediction. *Current Opinion Structural Biology* N. 16(2):172-7, 2006.

GOMIS-RÜTH, F.X., KRESS, L.F., KELLERMANN, J., MAYR, I., LEE, X. & HUBER, R. X-ray crystal structure of the snake venom zincendopeptidase adamalysin II. Primary and tertiary structure determination, refinement, molecular structure and comparison with astacin, collagenase and thermolysin. *J. Mol. Biol.* 239, 513–544, 1994.

GREER, J. Comparative model-building of the mammalian serineprotease. *Journal of Molecular Biology*. N.153:1027. 1981.

GREER, J. Model for haptoglobin heavy chain based upon structural homology. *Process Naturals Academy Scientific. USA.* N.77:3393-3397. 1980.

GUTIÉRREZ, J. M., RUCAVADO, A., ESCALANTEA, T. & DIAZ, C. Hemorrhage induced by snake venom metalloproteinases: biochemical and biophysical mechanisms involved in microvessel damage. *Toxicon* 45, 2005.

HEINEMANN, U.; *Nat. Struct. Biol.* 2000, 7, 940.

HOOFT, R. W. W., VRIEND, G., SANDER, C. & ABOLA, E. E. *Nature.* 381, 272, 1996.

JOHNSON, M. S., SRINIVASAN, N., SOWDHAMINI, R. & BLUNDELL, T. L. Knowledge-based protein modelling. *Crit. Rev. Biochem. Mol. Biol.* 29, 1, 1994.

JONES, D. T. Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292, 1999.

KACZANOWSKI, S. & ZIELENKIEWICZ, P. Why similar protein sequences encode similar threedimensional structures? *Theoretical Chemistry Accounts* N. 125:543-50, 2008.

- KAMIGUTI, A.S., ZUZEL M. & THEAKSTON, R. D. G. Snake venom metalloproteinases and disintegrins: interactions with cells. *Brazilian Journal of Medical and Biological Research* 31: 853-862, 1998.
- KOOLMAN, J. & ROEHM, K. H. *Color Atlas of Biochemistry*. New York: Thieme, 2005.
- KRISSINEL, E. & HENRICK, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797, 2007.
- LASKOWSKI, R. A., MACARTHUR, M. W. & THORNTON, J. M. Validation of protein models derived from experiment. *Current Opinion in Structural Biology*, 8, 631, 1998.
- LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures . *J. Appl. Crystallogr.* 26, 283, 1993.
- LAZARIDIS, T. & KARPLUS, M. Effective energy function for proteins in solution. *Proteins. N.* 1;35(2):133-52, 1999.
- LEAL, D. P. *Bioquímica estrutural y aplicada a la medicina*. Ciudad del Mexico: Limusa, 2004.
- LESK, A. *Introduction to Bioinformatics*. New York: Oxford, 2002.
- LEVINTHAL, C. Are the pathways for protein folding? *Journal Chimie Physique* 65, N.01: 44-45, 1968.
- LUTHY, R., BOWIE, J. U. & EISENBERG, D. Assessment of protein models with three-dimensional profiles. *Nature.* v.356, 83, 1992.
- MARSHALL, G. R. Introduction to chemoinformatics in drug discovery – a personal view. In: OPREA, T. I. *Chemoinformatics in drug discovery*. Weinheim: Wiley VHC, 2004.
- McLACHLAN, A. D. & SHOTTON, D. M. Structural similarities between alpha-lytic protease of *Myxobacter* 495 and elastase. *Natural Newer Biology. N.* 229: 202-205. 1971.
- MEIDANIS, J. & SETUBAL, J. *Introduction to Computational Molecular Biology*. Boston: PWS Publishing company, 2003.
- MONK, P. *Physical chemistry : understanding our chemical world*. England: John Wiley & Son's, 2004.
- MURRAY, F. K., GRANNER, D. K., MAYES, P. A. & RODWELL, V. W. *Harper's Illustrated Biochemistry*. 26ed. New York: Lange Medical, 2003.
- NELSON, D. L. & COX, M. M. *Lehninger : Principles of Biochemistry*. 4ed. Sarvier: Wisconsin, 2007.
- NORWELL, J. C. & MACHALEK, A. Z.; *Nat. Struct. Biol.* 2000, 7, 931.
- Óbitos por serpentes. Portal do Ministério da Saúde. Disponível em [http://portal.saude.gov.br/portal/saude/visualizar\\_texto.cfm?idtxt=27623](http://portal.saude.gov.br/portal/saude/visualizar_texto.cfm?idtxt=27623). Acesso em dez. 2008.
- OLIVEIRA, C.T., MACEDO, O. A. C. & OLIVEIRA, T. B. S. Solução do problema da busca de motifs usando Algoritmos Genéticos. 2006. Disponível em <[www.otaviomacedo.eti.br/hp/download/motif\\_relatorio.pdf](http://www.otaviomacedo.eti.br/hp/download/motif_relatorio.pdf)>. Acesso dez/2007.
- OHSAKA, A. Hemorrhagic, necrotizing and edema-forming effects of snake venoms. In: Lee, C.Y. (Ed.), *Handbook of Experimental Pharmacology*, 52. Berlin: Springer-Verlag, 1979.
- PUGALENTI, G, SHAMEER, K., SRINIVASAN, N., SOWDHAMINI, R. HARMONY: a server for the assessment of protein structures. *Nucleic Acids Res. N.* 1; 34 (Web Server issue):W231-4. 2006.

- RAMACHANDRAN, G. N. & SASISEKHARAN, V. Conformation of polypeptides and proteins. *Adv. Prot. Chem.* 23, 283, 1968.
- RAMOS, O. H. P. Predições estruturais com base em homologia. In: *Métodos em Toxinologia: toxinas de serpentes*. São Carlos: EdUFSCar, 2007.
- SALI, A. & BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234(3):779-815, 1993.
- SALI, A., SHAKHNOVICH, E. & KARPLUS, M. How does a protein fold? *Nature* N.369, 1994.
- SANTOS FILHO, O. A. Modelagem da di-hidrofolato Redutase do "Plasmodium falciparum", QSAR-3D pela Metodologia do Campo de Força de Energia Livre e QSAR-4D. Tese de Doutorado, Instituto Militar de Engenharia, Rio de Janeiro, Brasil, 2000.
- SANTOS FILHO, O. A. & ALENCASTRO, R. B. Modelagem de proteínas por homologia. *Quim. Nova*, Vol. 26, No. 2, 2003.
- SILVA, M.B., SCHATTNER, M., RAMOS, C.R.R., JUNQUEIRA-DE-AZEVEDO, I.L.M., GUARNIERI, M.C., LAZZAR, M.A., SAMPAIO, C.A.M., POZNER, R.G., VENTURA, J.S., HO, P.L. & CHUDZINSKI-TAVASSI, A.M. A prothrombin activator from Bothrops erythromelas (jararaca-da-seca) snake venom: characterization and molecular cloning. *Biochemical Journal* 369, 2003.
- SILVA, V. B. Estudos de modelagem molecular e relação estrutura-atividade da oncoproteína hnRNP K e ligantes. Dissertação (Mestrado) – Universidade de São Paulo, Ribeirão Preto, 2007.
- SIPPL, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.*, 213, 859, 1990.
- SIPPL, M. J. Recognition of errors in three-dimensional structures of proteins. *Proteins*, 17, 355, 1993.
- SCHWEDE, T., KOPP, J., GUEX, N. & PEITSCH, M. C. SWISS-MODEL: an automated protein homology-modeling server. *Nucleic Acids Research* 31, 2003.
- STOCKER, W. & BODE, W. Structural features of a superfamily of zinc-endopeptidases: the metzincins. *Current Opinion of Structural Biology*. N.05: 383-390, 1995.
- SUN, S. F. Physical chemistry of macromolecules : basic principles and issues. New Jersey: John Wiley & Son's, 2004.
- TAYLOR, W. R. Identification of protein sequence homology by consensus template alignment. *Journal Molecular Biology*. N. 188:233-258. 1986.
- TEIXEIRA, E. S. Estudo dos processos para modelagem, avaliação, alteração e seleção da estrutura de proteínas. Dissertação (Mestrado): UFPE, 2005.
- TICONA, W. G. C. Algoritmos evolutivos multi-objetivos para a reconstrução de árvores filogenéticas. Tese (doutorado): USP São Carlos, 2008.
- TSAI, C. An introduction to computational biochemistry. New York: Wiley-Liss Inc, 2002.
- TOMETSKO, S. A. Computer approaches to protein structure. II. Model building by computer. *Computational Biomedicine Research*. N.03. p.690-698. 1970.

VAN HOLDE, K. E., JOHNSON, W. C. & HO, P. S. Principles of Physical Biochemistry. New Jersey: Prentice Hall, 1998.

WALLNER, B. & ELOFSSON, A. Quality assessment of protein models. Center for Biomembrane Research and Stockholm Bioinformatics Center, Stockholm University, 2007. Disponível em : <http://bioinfo.se/papers/mqap-book.pdf>. Acesso em agosto2009.

WATANABE, L., SHANNON, J.D., VALENTE, R.H., RUCAVADO, A., ALAPE-GIRON, A., KAMIGUTI, A.S., THEAKSTON, R.D.G., FOX, J.W., GUTIERREZ, J.M. & ARNI, R.K. Amino acid sequence and crystal structure of BaP1, a metalloproteinase from Bothrops asper snake venom that exerts multiple tissue-damaging activities. Protein Sci. 12:2273-2281,2003.

WIEDERSTEIN, M. & SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. Nucleic Acids Research, 35, 2007.

XIANG, Z. X., CSOTO, C. & HONIG, B. Evaluating configurational free energies: the colony energy concept and its application to the problem of protein loop prediction. Process Naturals Academy Scientific. N.99: 7432:7437. 2002.

XIONG, J. Essential Bioinformatics. New York: University Press, 2006.

## **ANEXOS**

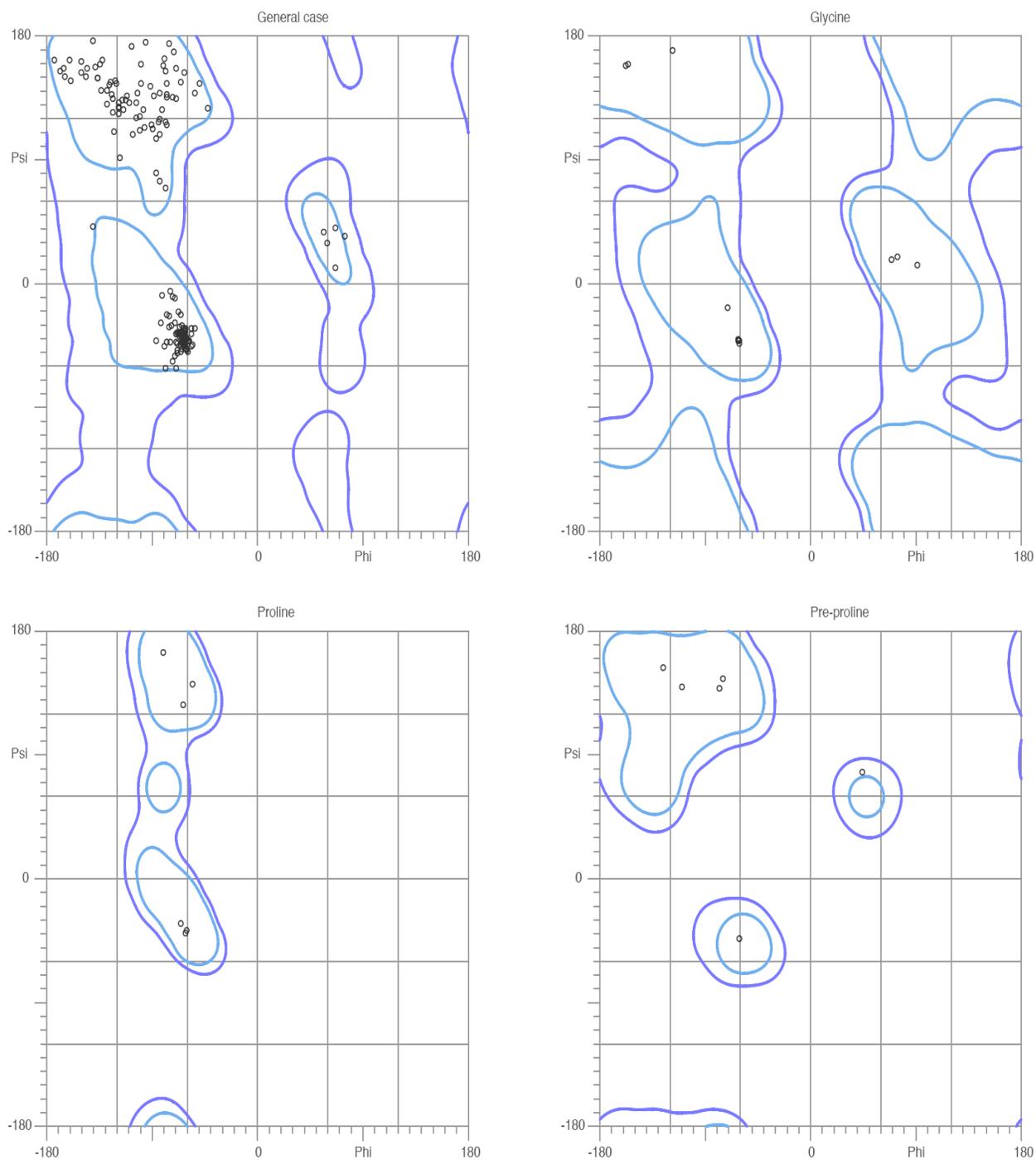


Fig. 36 – Gráficos de Ramachandran fornecidos pelo Molprobitry para o modelo P1.

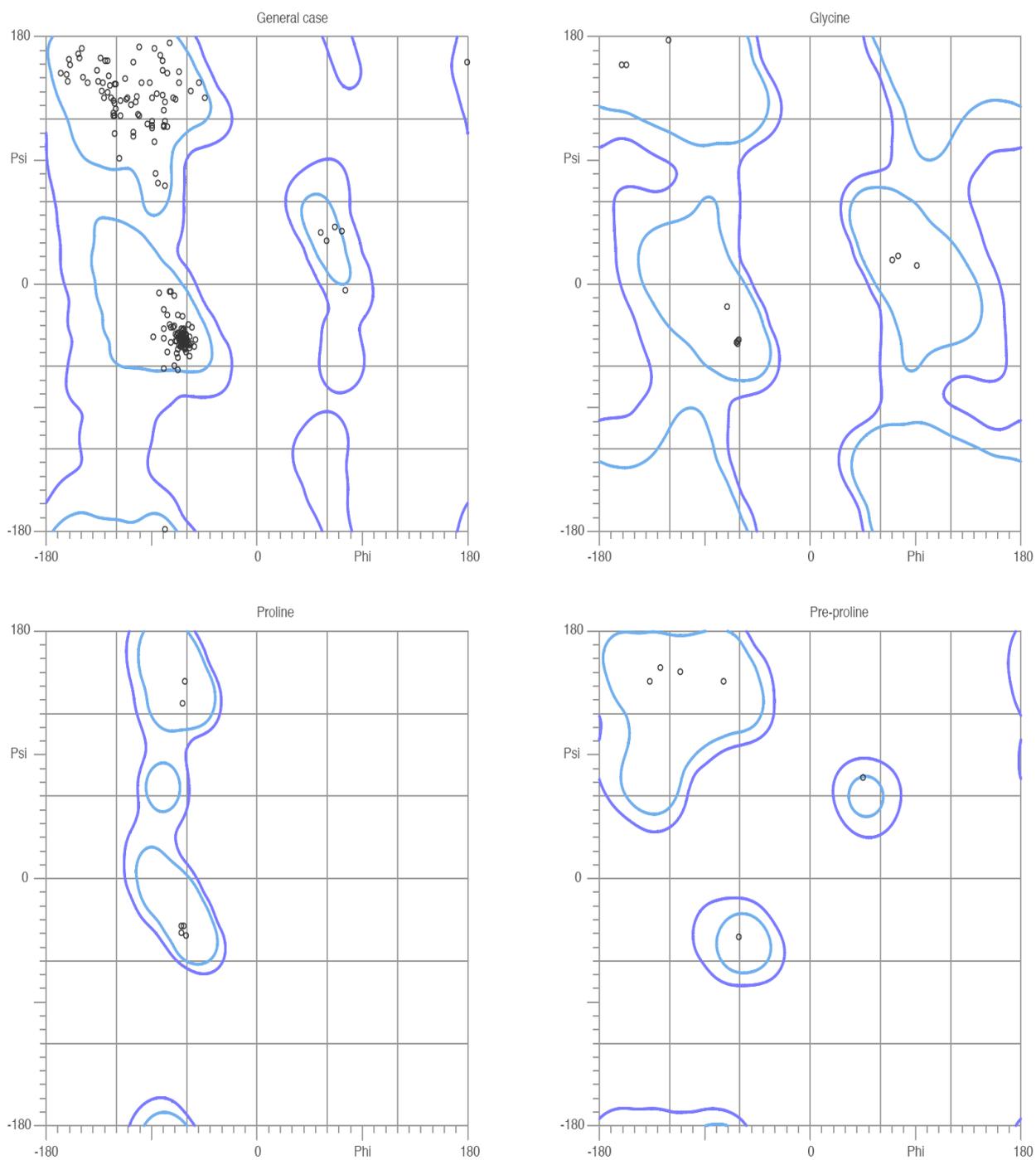


Fig. 37 – Gráficos de Ramachandran fornecidos pelo Molprobitry para o modelo P2.

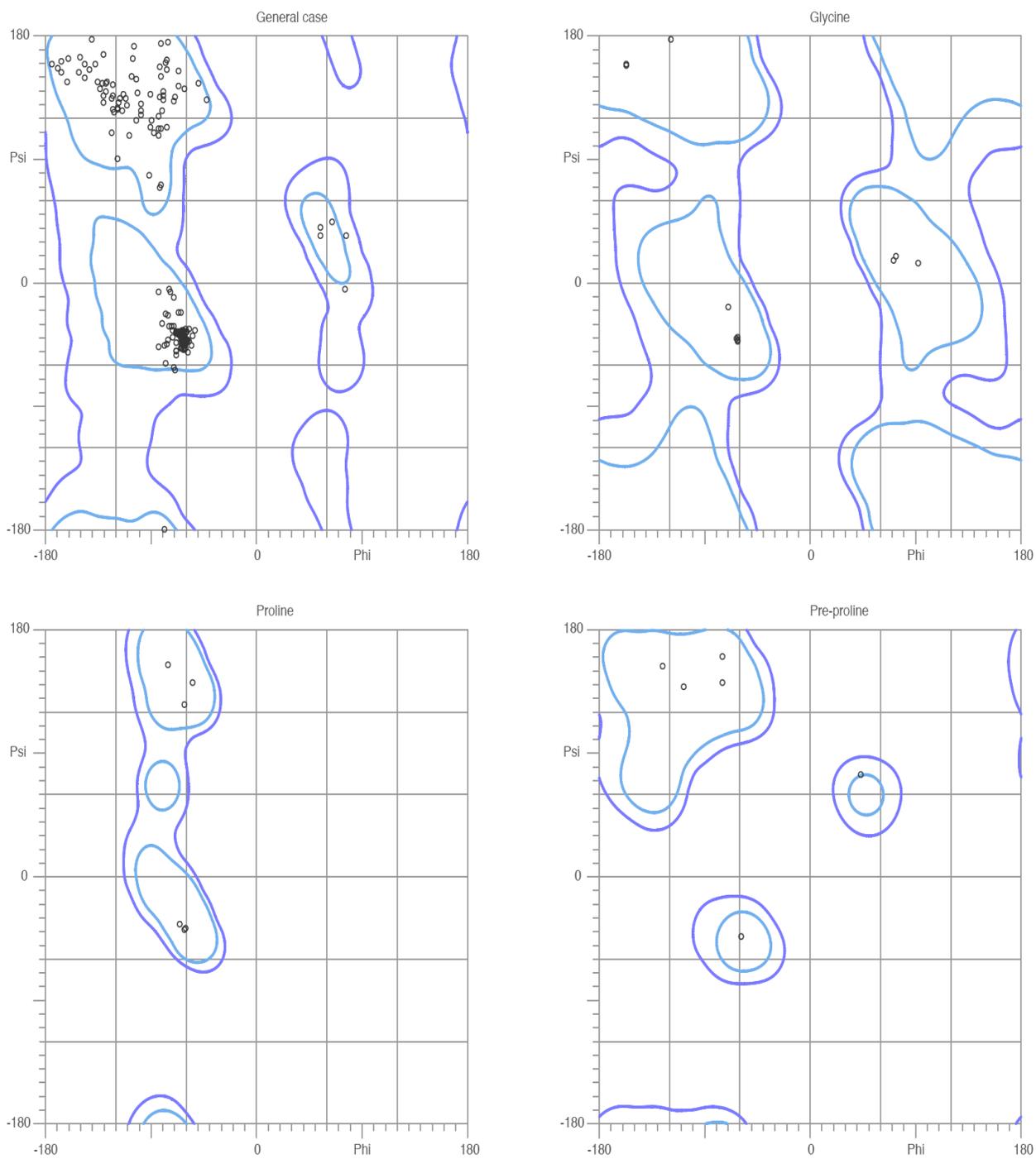


Fig. 38 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P3.

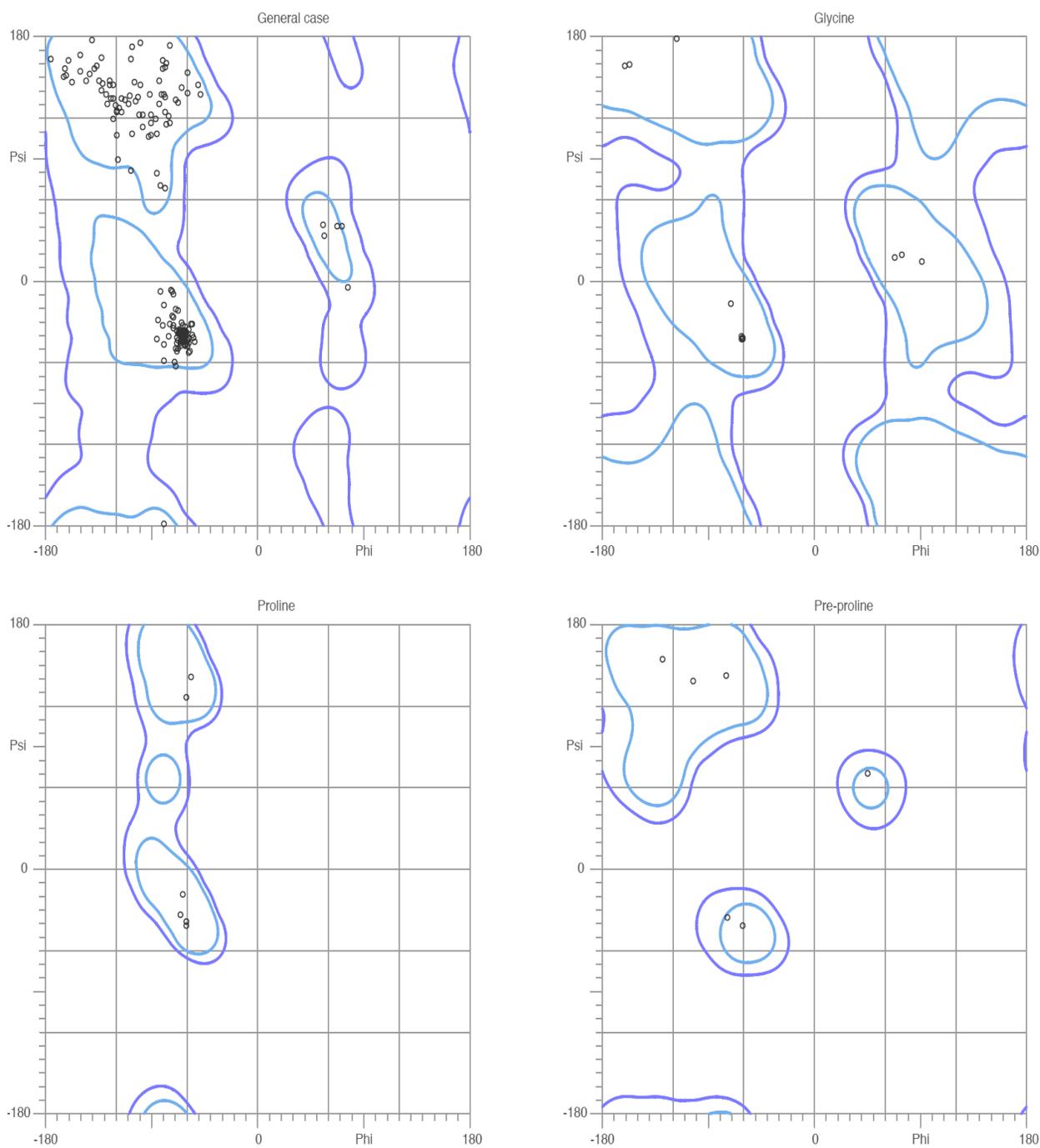


Fig. 39 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P4.

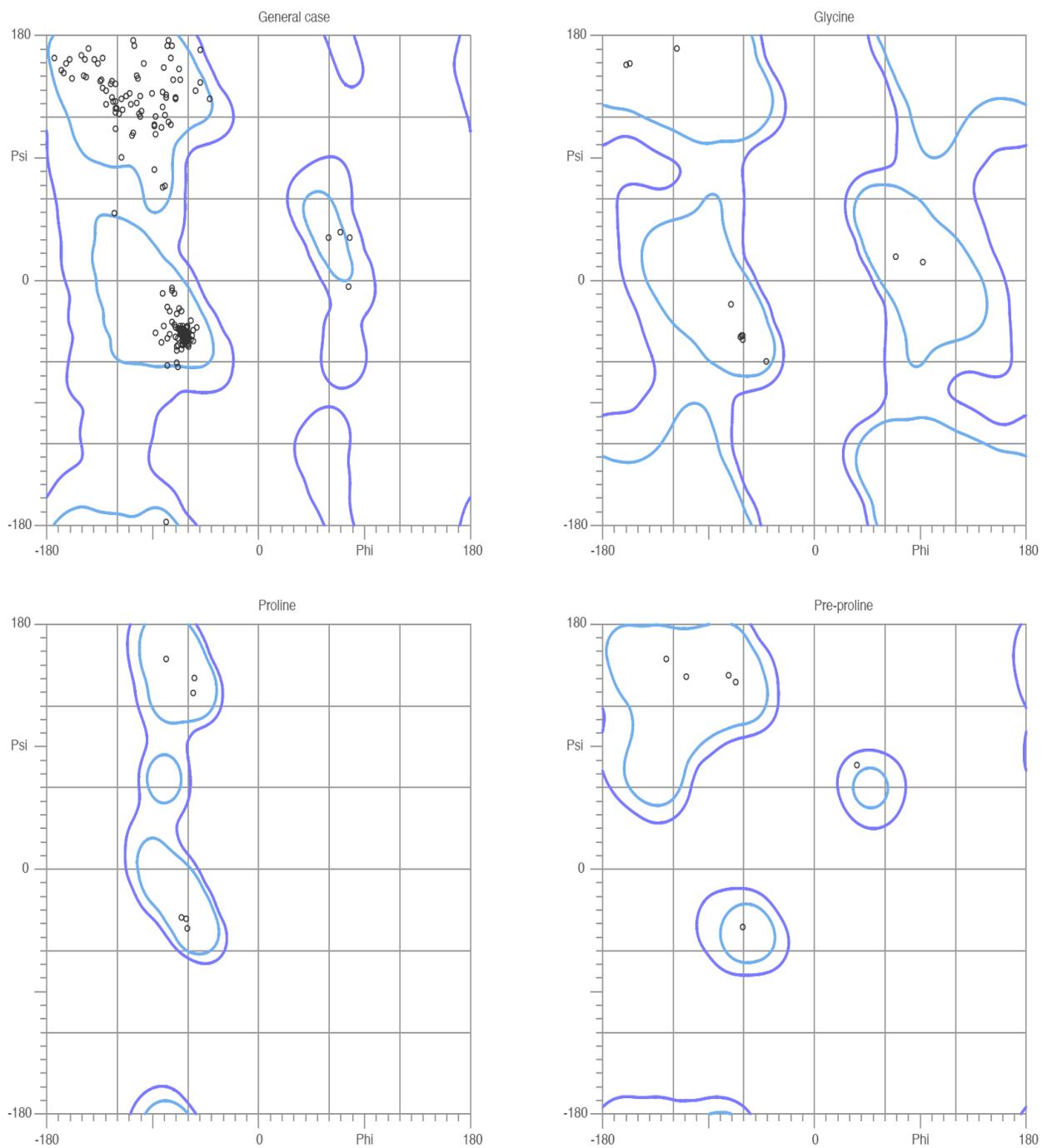


Fig. 40 – Gráficos de Ramachandran fornecidos pelo Molprobit para o modelo P5.

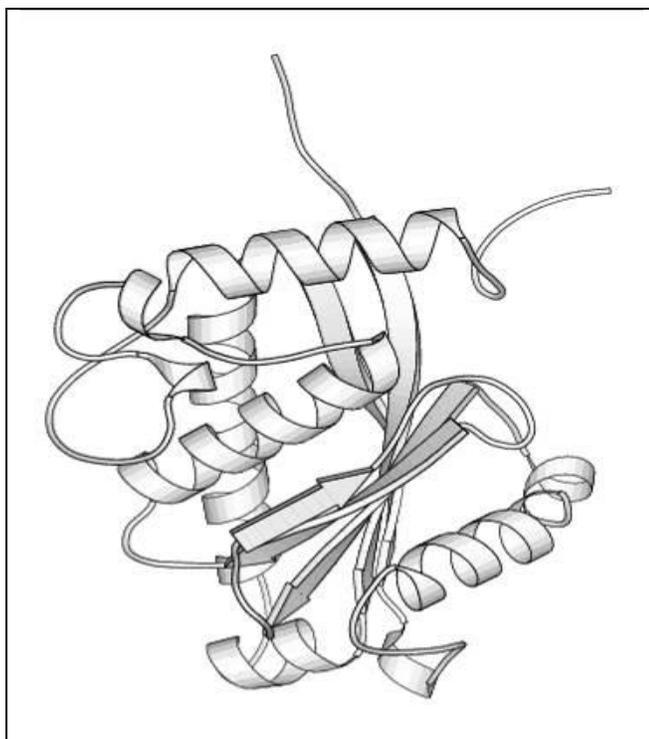


Fig. 41 – Estrutura tridimensional sugerida pelo Harmony para o modelo P2.

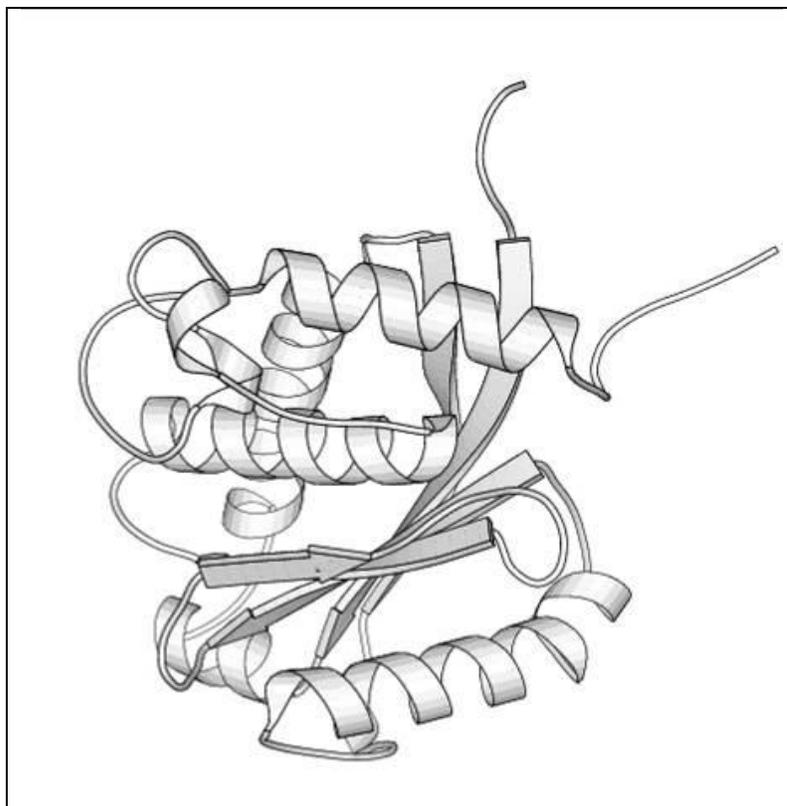


Fig. 42 – Estrutura tridimensional sugerida pelo Harmony para o modelo P3.

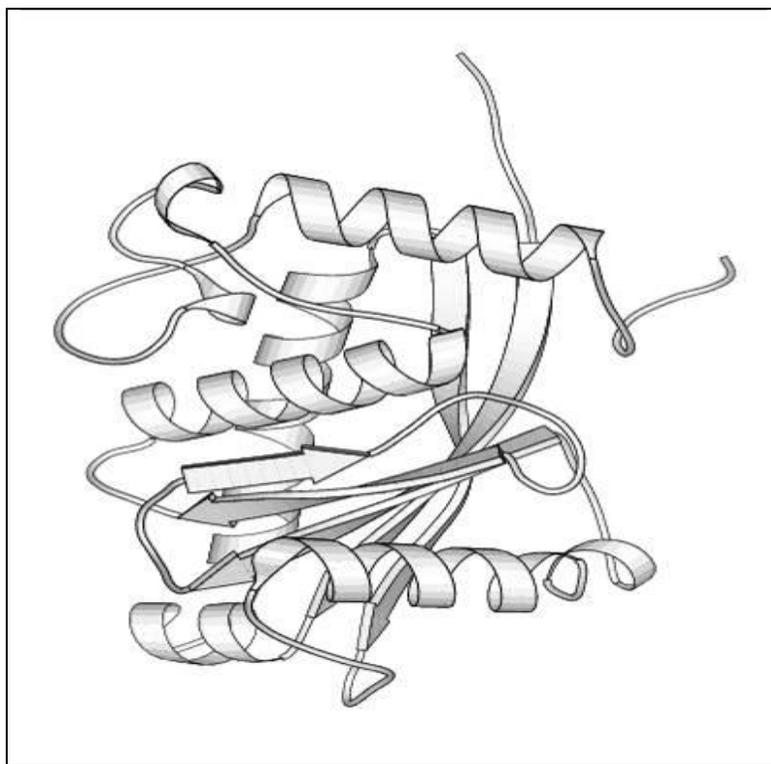


Fig. 43 – Estrutura tridimensional sugerida pelo Harmony para o modelo P4.

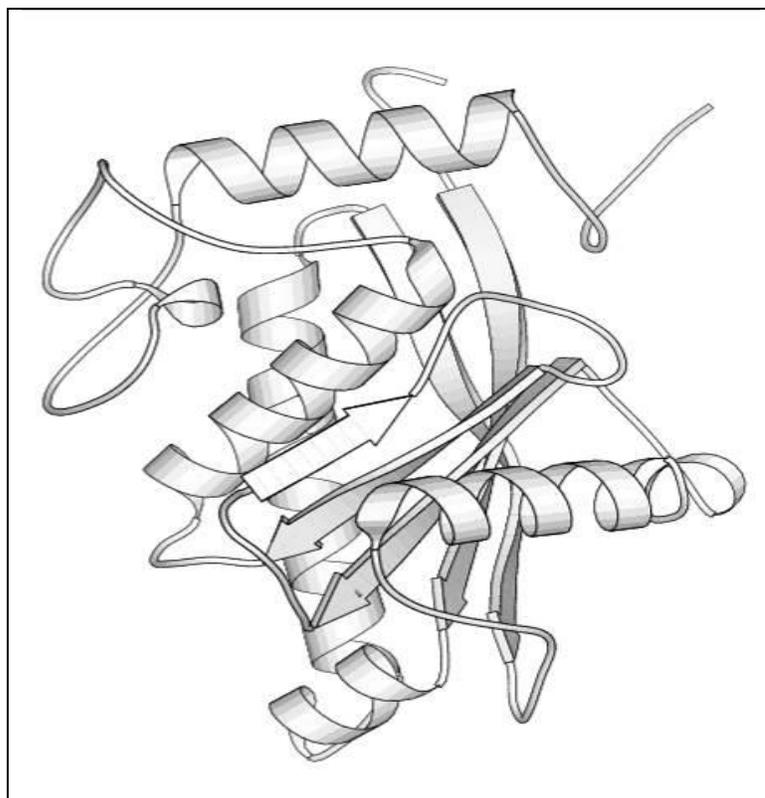


Fig. 44 – Estrutura tridimensional sugerida pelo Harmony para o modelo P5.

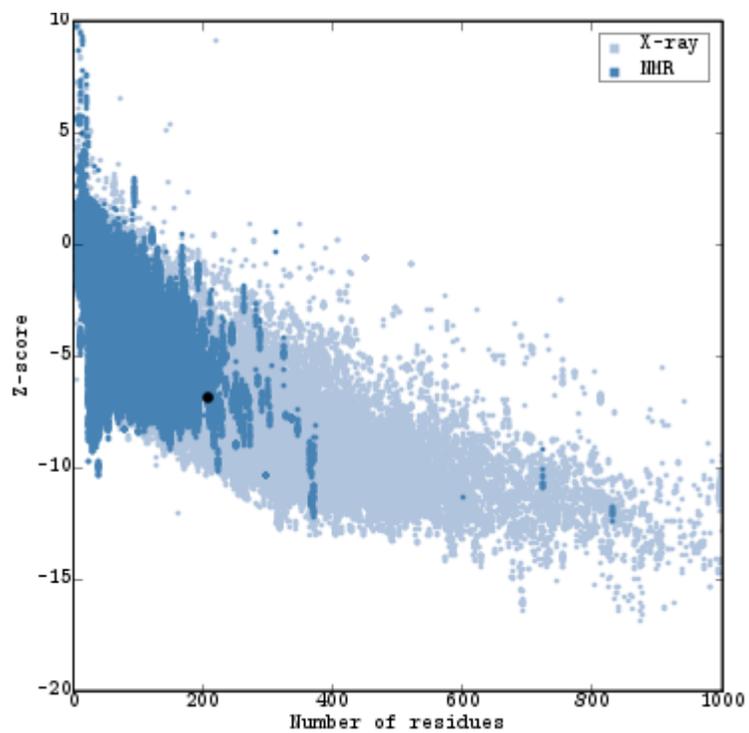


Fig. 45. – Qualidade global estimada pelo ProSA-web para o modelo P2.

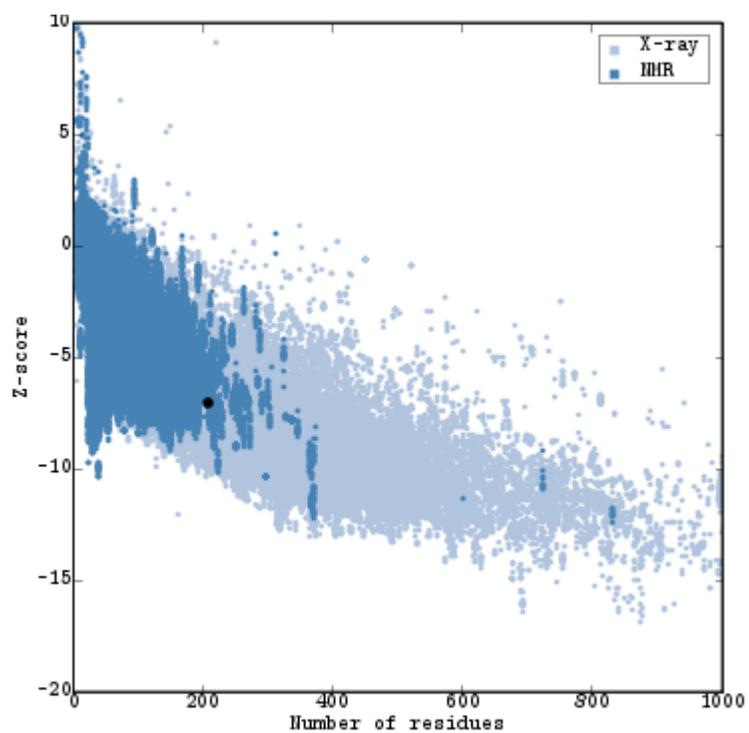


Fig. 46 – Qualidade global estimada pelo ProSA-web para o modelo P3.

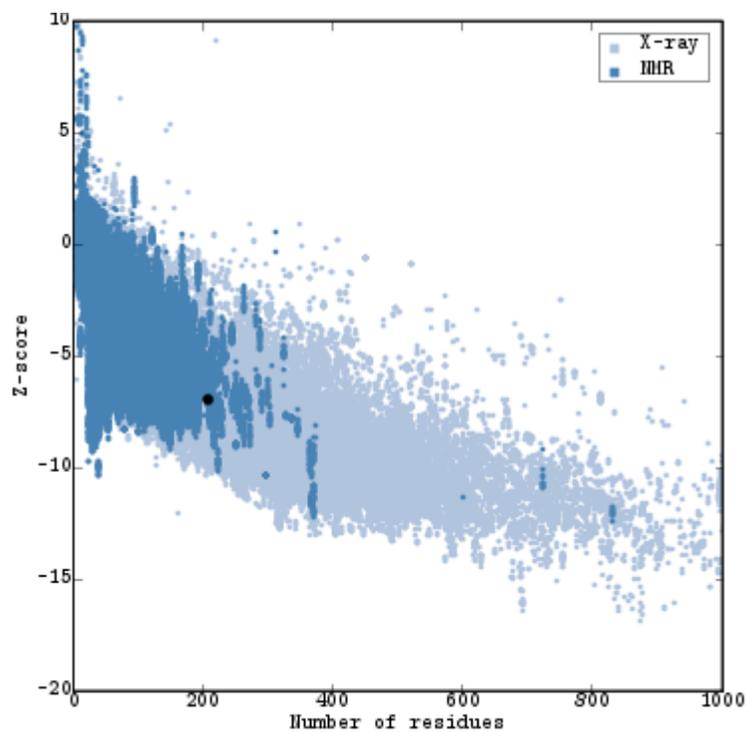


Fig. 47 – Qualidade global estimada pelo ProSA-web para o modelo P4.

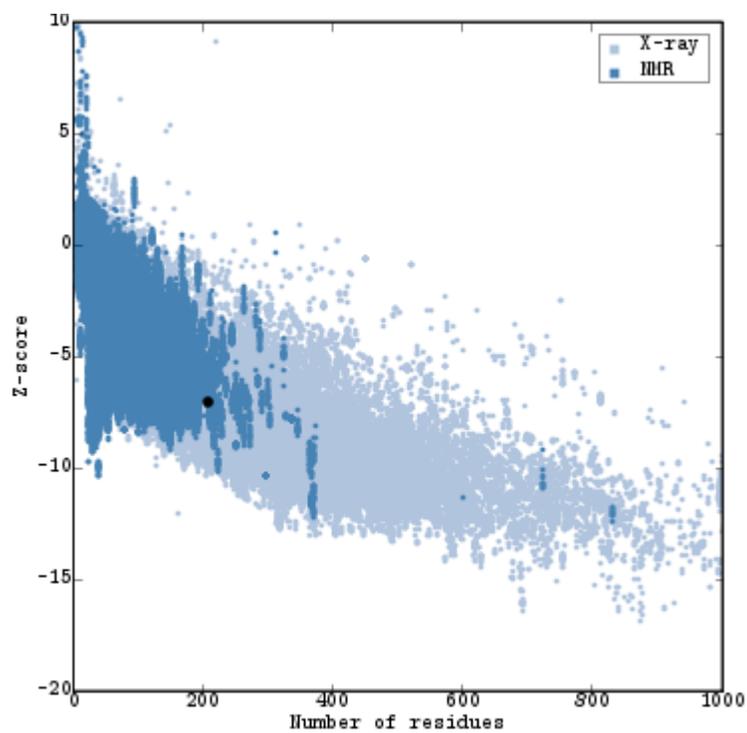


Fig. 48 – Qualidade global estimada pelo ProSA-web para o modelo P5.

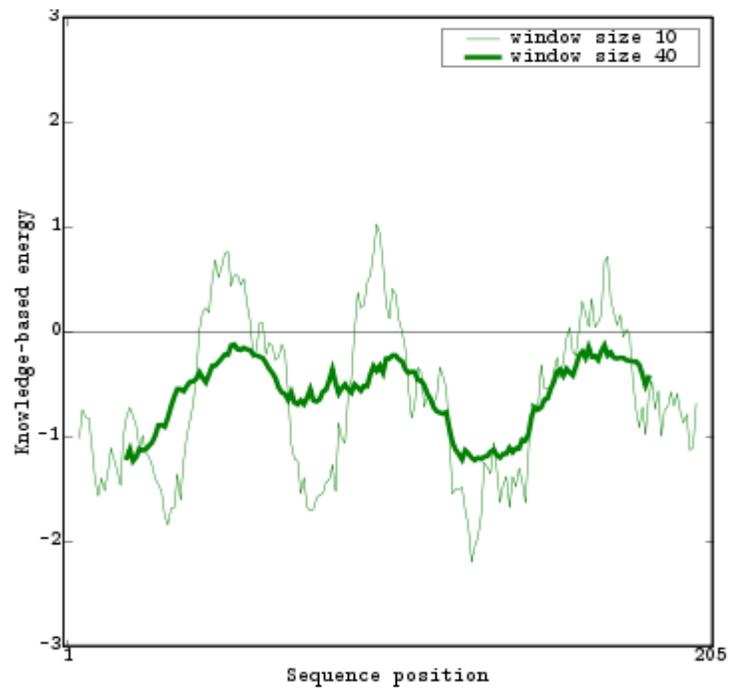


Fig. 49 – Qualidade local estimada para o modelo P2.

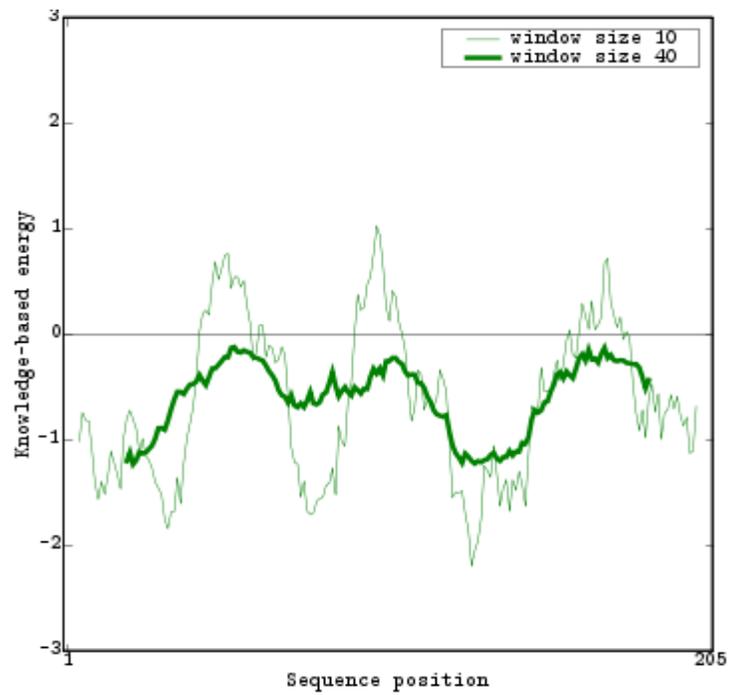


Fig. 50 – Qualidade local estimada para o modelo P3.

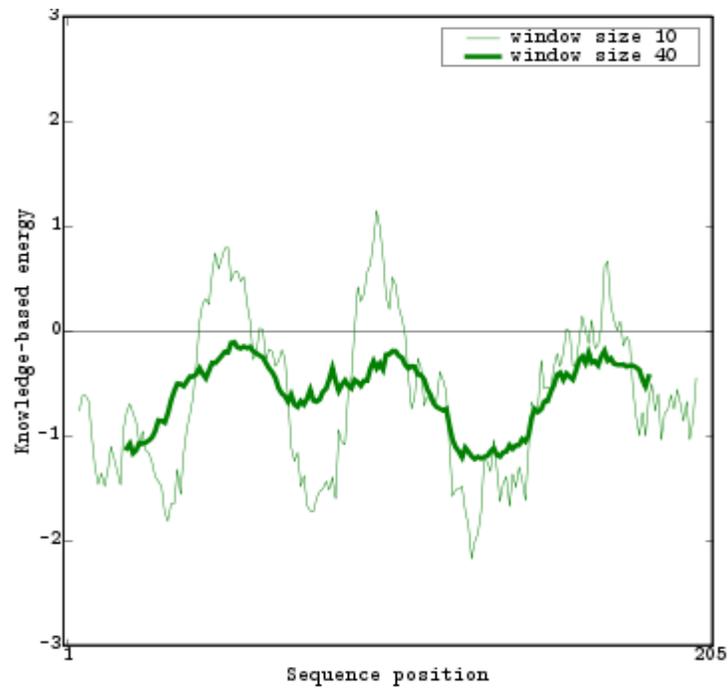


Fig. 51 – Qualidade local estimada para o modelo P4.

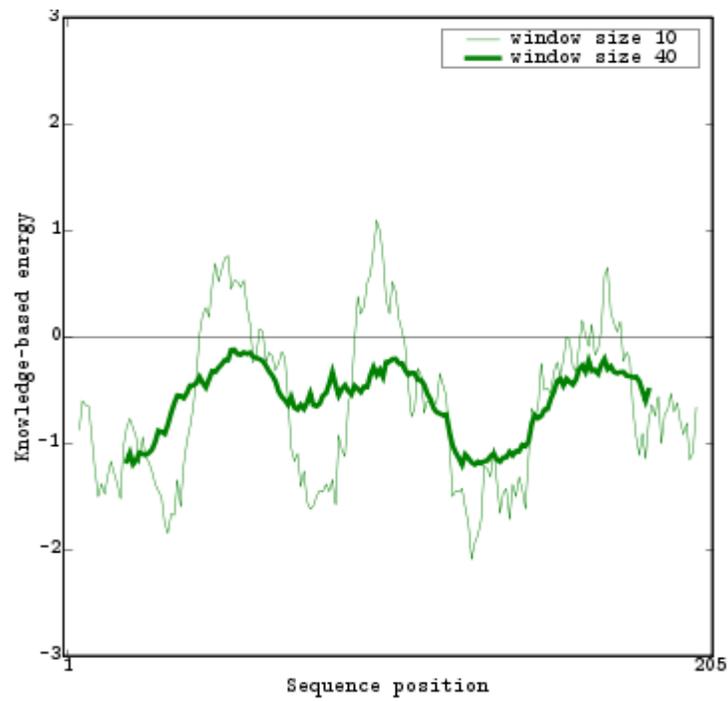


Fig. 52 – Qualidade local estimada para o modelo P5.

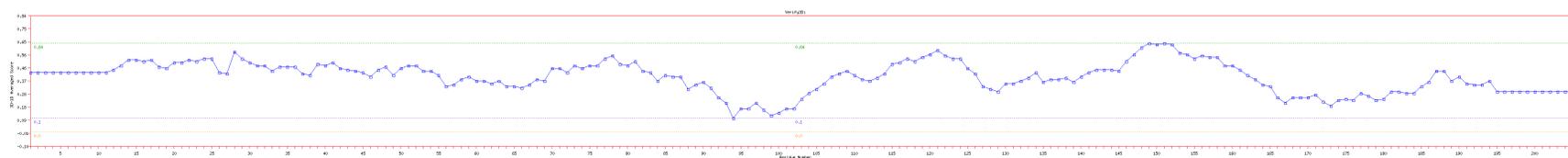


Fig. 53 – Correspondência de seqüências 3D-1D para o modelo P2.

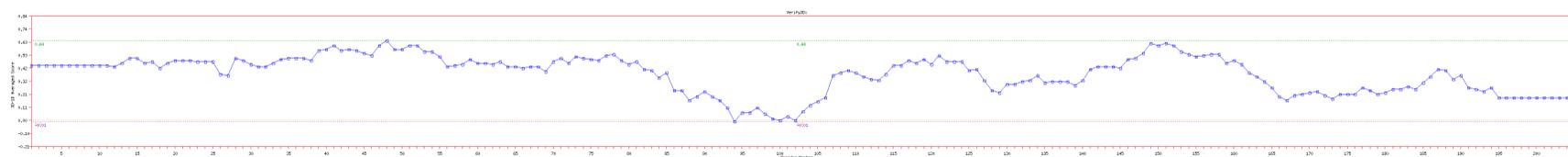


Fig. 54 – Correspondência de seqüências 3D-1D para o modelo P3.

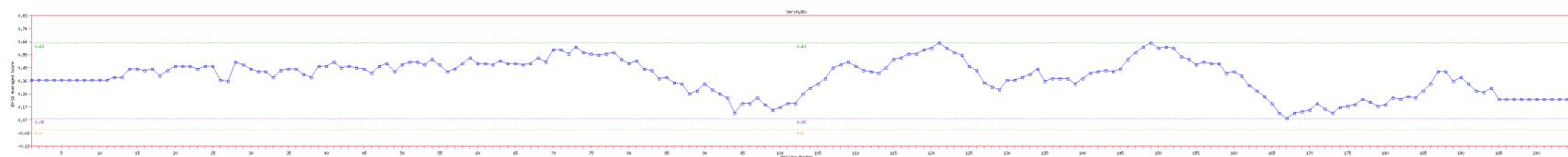


Fig. 55 – Correspondência de seqüências 3D-1D para o modelo P4.

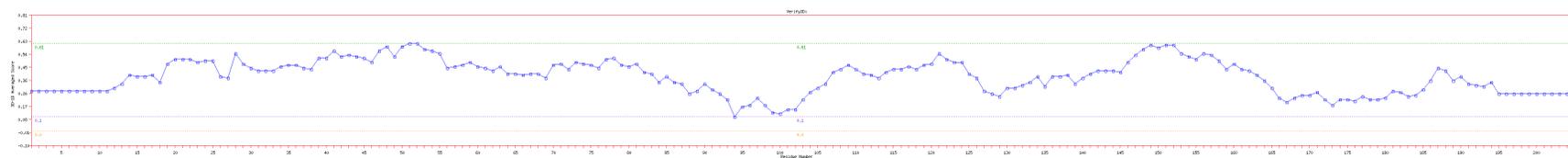


Fig. 56 – Correspondência de seqüências 3D-1D para o modelo P5.



Fig. 57 – Correspondência de seqüências 3D-1D para o molde (*B. asper*).