

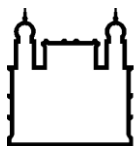
MINISTÉRIO DA SAÚDE
FUNDAÇÃO OSWALDO CRUZ
INSTITUTO OSWALDO CRUZ

Mestrado em Programa de Pós-Graduação Biologia Computacional e Sistemas

PROPOSIÇÃO DE MODELOS TRIDIMENSIONAIS DA REGIÃO COOH-
TERMINAL DA CISTEÍNO-PROTEASE B DE *Leishmania (Leishmania)*
amazonensis

DEBORAH ANTUNES DOS SANTOS

Rio de Janeiro
Junho de 2014



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

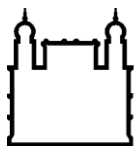
DEBORAH ANTUNES DOS SANTOS

Proposição de modelos tridimensionais da extensão COOH-terminal da cisteíno-protease B de *Leishmania (Leishmania) amazonensis*

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

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

RIO DE JANEIRO
Junho de 2014



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

Programa de Pós-Graduação em Biologia Computacional e Sistemas

AUTOR: DEBORAH ANTUNES DOS SANTOS

PROPOSIÇÃO DE MODELOS TRIDIMENSIONAIS DA EXTENSÃO COOH- TERMINAL DA CISTEÍNO-PROTEASE B DE *Leishmania (Leishmania)* *amazonensis*

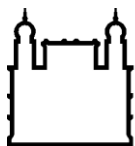
**ORIENTADORES: Prof. Dr. Ernesto Raúl Caffarena
Prof. Dr. Carlos Roberto Alves**

Aprovada em: 04/06/2014

EXAMINADORES:

**Prof. Dr^a Ana Carolina Ramos Guimarães – Presidente (Fiocruz/RJ)
Prof. Dr^a Celia Beatriz Anteneodo de Porto (PUC-Rio)
Prof. Dr. Bernardo Acácio Santini Pereira (Fiocruz/RJ)
Prof. Dr^a Mirian Claudia de Souza Pereira (Fiocruz/RJ) – Suplente
Prof. Dr. Paulo Ricardo Batista (Fiocruz/RJ) – Suplente e Revisor**

Rio de Janeiro, 04 de junho de 2014



Ministério da Saúde

FIOCRUZ
Fundação Oswaldo Cruz

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

*Dedico aos amigos
Aos que se tornaram familiares,
Aos que nasceram familiares
e aos que conheci antes de ontem.*

*Dedico tanto aos que me deixam louco,
Quanto aos que enlouqueço.*

*Aos que me criticam em tudo,
E a um ou outro que atura
Minha "chatura"*

*Aos amigos que correm,
Aos amigos que contemplam.*

*Aos que me consideram muito,
E aos que, com razão, fazem pouco.*

*Aos que conhecem o que penso,
E aos que só conhecem o que faço.*

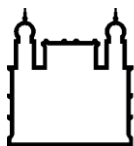
*Aos que passam o dia todo comigo,
e aos que estão todo tempo em mim.*

*Este trabalho é a soma de todos vocês.
E se ele não é melhor,
É por falta de memória,
Mas não por falta de amigos.*

Adaptado de Efraim Rodrigues, Richard B. Primack

AGRADECIMENTOS

- Um agradecimento especial à minha família, pelo constante apoio em todas as fases da minha vida, me fazendo sentir confortável até nos momentos mais difíceis, mas sempre me estimulando a seguir em frente. Muito obrigada, mãe, vó, pai e meu filho canino Gustavo.
- Meus sinceros agradecimentos aos meus queridos orientadores Ernesto Raúl Caffarena e Carlos Roberto Alves pela orientação, incentivo, paciência, disponibilidade e ensinamentos durante a realização deste projeto.
- Ao meu amigo Franklin Souza da Silva por ter me apresentado a esse mundo e ser meu terceiro orientador. Será que devo agradecer ou lamentar? ☺
- Aos amigos do GBCMM e do Grupo Crachá Azul, Amanda Sutter, Lucianna Helene, Vanessa Silva (Ariel), Rafael Ferreira, Artur Brandt, André Torres e Vanessa Sinatti.
- Aos pesquisadores que muito me ajudaram, Ana Carolina Guimarães, Maurício Costa e Paulo Ricardo Batista.
- Aos colegas do Labimdoe pela ajuda no desenvolvimento do projeto, principalmente Raquel de Souza, a quem tanto aporrinhei.
- Ao Dr. Paulo Ricardo Batista, pela revisão crítica da tese.
- Aos membros da banca, por terem gentilmente aceitado o convite para participar da avaliação desse trabalho.
- À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES pelo auxílio financeiro.
- À todos que direta ou indiretamente colaboraram para que esse trabalho se tornasse possível, meus mais sinceros agradecimentos.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

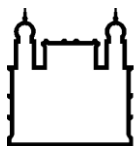
PROPOSIÇÃO DE MODELOS TRIDIMENSIONAIS DA EXTENSÃO COOH- TERMINAL DA CÍSTEINO-PROTEASE B DE *Leishmania (Leishmania)* *amazonensis*

RESUMO

DISSERTAÇÃO DE MESTRADO EM BIOLOGIA COMPUTACIONAL E SISTEMAS

Deborah Antunes dos Santos

A *Leishmania (Leishmania) amazonensis* é um importante agente etiológico da leishmaniose tegumentar em humanos no Brasil. Este parasito apresenta mecanismos de adaptação que podem ser guiados por suas proteases onde se destacam as cisteíno-proteases (CP), sendo a CPB a mais estudada dentre as CPs já descritas em *Leishmania* spp. O foco deste trabalho é a região COOH-terminal da CPB (*cyspep*), por suas reconhecidas propriedades modulatórias sobre o sistema imune de camundongos. No presente estudo busca-se propor um modelo estrutural da *cyspep* visando estabelecer sua relação estrutura-função. O estudo foi conduzido por abordagem *in silico*, desenvolvendo o modelo tridimensional através de técnicas de modelagem comparativa, *threading* e métodos *de novo*, e analisando a estabilidade estrutural por dinâmica molecular. A associação do uso de servidores *online* que utilizam as abordagens de modelagem comparativa, *threading* e métodos *de novo* para predição tridimensional de proteínas foi favorável à resolução de 72 modelos da *cyspep*. Através de informações sobre predição de estrutura secundária e ligações dissulfeto, dos modelos gerados para *cyspep*, 11 foram selecionados para simulações de dinâmica molecular, sendo dois modelos obtidos por modelagem comparativa (*Comp, Fugue*), 5 por *threading* (*IT1, M4, M8, L3 e SP5*) e 4 por métodos *de novo* (*Q5, Q7, D9 e R1*). Os 11 modelos foram submetidos a simulação por dinâmica molecular por um tempo de 200 ns. Os modelos *Fugue* e *IT1* também foram submetidos a simulações de dinâmica molecular sob a influência de diferentes campos de força para melhor avaliar seu conteúdo de estrutura secundária. Nossos resultados sugerem que a *cyspep* pode adotar uma conformação composta principalmente de folhas- β , as quais se mantiveram estáveis após longas simulações de dinâmica molecular sob a influência de diferentes campos de força; enquanto o teor helicoidal foi rapidamente desfeito. Ainda com relação as estruturas em folha- β , não foi possível observar padrões de dobras consenso entre diferentes modelos. Embora os procedimentos usados neste trabalho para predição de ligações de dissulfeto indicarem a possibilidade de 36 ligações entre os resíduos de cisteína (Cys17, Cys21, Cys31, Cys44, Cys52, Cys59, Cys65, Cys73 e Cys85) não foi possível constatar um consenso na predição dessas ligações.



Ministério da Saúde

FIOCRUZ

Fundação Oswaldo Cruz

INSTITUTO OSWALDO CRUZ

PROPOSITION OF THREE-DIMENSIONAL MODELS OF THE COOH-TERMINAL CYSTEINE PROTEASE B EXTENSION OF *Leishmania (leishmania)* *amazonensis*

ABSTRACT

MASTER DISSERTATION IN COMPUTATIONAL SYSTEMS BIOLOGY

Deborah Antunes dos Santos

Leishmania (Leishmania) amazonensis is an important etiological agent of cutaneous leishmaniasis in humans, in Brazil. This parasite has adaptive mechanisms that can be modulated by their proteases, characterizing the cysteine protease (CP) as the most extensively studied CPB among the CPs described for *Leishmania spp.* This work was focused on the COOH -terminal region of CPB (*cyspep*), recognized for their modulatory properties on the immune system of mice. This present study aims to propose a structural model for the *cyspep* trying to determine its structure-function relationship. The present study was conducted by *in silico* to develop a three-dimensional model using comparative modeling, threading and *de novo* methods, and molecular dynamics to analyze its structural stability. The association of many online servers using the approaches of comparative modeling, threading and *de novo* methods for the prediction of the three-dimensional structure of protein was favorable for the achievement of 72 three-dimensional models for *cyspep*. Through secondary structure and disulfide bonds prediction, out of the 72 models generated for *cyspep*, 11 were selected for molecular dynamics simulations, of which 2 models were obtained by comparative modeling (*Comp*, *Fugue*), 5 by threading (*IT1*, *M4*, *M8*, *L3* and *SP5*) and 4 *de novo* methods (*Q5*, *Q7*, *D9* and *R1*). The eleven models were submitted to molecular dynamics simulations for a computational time of 200 ns. The *Fugue* and *IT1* models were also subjected to molecular dynamics simulations under different force fields to better assess their secondary structure content and determine the impact of the force field on the structure. Our results suggest that the *cyspep* might adopt a conformation composed of mostly β -sheets, which remained stable after long molecular dynamics simulations under the influence of different force fields while the helical content was rapidly disrupted. However, it was not possible to recognize a consensus of folding patterns among the various created models. Additionally, although the procedures used in this work for predicting disulfide bonds indicated the possibility of 36 bonds between cysteine residues (Cys17, Cys21, Cys31, Cys44, Cys52, Cys59, Cys65, Cys73 and Cys85) we could not observe a consensus upon the prediction of these bonds.

ÍNDICE

1	INTRODUÇÃO	1
1.1	Considerações Gerais.....	1
1.2	O Parasito e o Ciclo Biológico	4
1.3	Proteases	9
1.4	Proteases de <i>Leishmania</i> spp	10
1.4.1	Aspártico- proteases.....	10
1.4.2	Serino- proteases	11
1.4.3	Metalo- proteases	12
1.4.4	Cisteíno-proteases.....	13
1.4.5	Extensão COOH-terminal da CPB (CTE)	14
1.5	Predição de Estruturas Tridimensionais de Proteínas	21
1.5.1	Modelagem Comparativa.....	22
1.5.2	Abordagem threading	23
1.5.3	Método de novo	25
1.5.4	Método ab initio	25
1.5.5	Acurácia e aplicação de modelos de estrutura de proteínas.....	25
1.6	Dinâmica Molecular	27
1.6.1	Campo de Força	28
1.7	Justificativa.....	31
2	OBJETIVOS	33
2.1	Objetivo Geral	33
2.2	Objetivos Específicos	33
3	MATERIAL E MÉTODOS	34
3.1	Predição de Estrutura Secundária	34
3.2	Predição das Ligações Dissulfeto	35
3.3	Modelagem da Estrutura Tridimensional da Cyspep	36
3.3.1	Modelagem Comparativa.....	36
3.3.2	Threading.....	37
3.3.3	Métodos de novo	40
3.4	Dinâmica Molecular	41
3.4.1	Análise das Simulações de Dinâmica Molecular	42

4	RESULTADOS E DISCUSSÃO	46
4.1	Predição de Estrutura Secundária	46
4.2	Predição das Ligações Dissulfeto	48
4.3	Modelagem da Estrutura Tridimensional da <i>Cyspep</i>	49
4.4	Dinâmica Molecular	67
4.4.1	Simulações de Dinâmica Molecular	67
4.4.2	Influência do campo de força sobre a Dinâmica da estrutura Fugue	75
4.4.3	Influência do campo de força sobre a Dinâmica da estrutura I-Tasser	83
5	DISCUSSÃO	90
6	CONCLUSÕES	97
7	REFERÊNCIAS BIBLIOGRÁFICAS	98
	ANEXO A – ARQUIVOS DE ENTRADA PARA SIMULAÇÕES DE DINÂMICA MOLECULAR	112
	ANEXO B – TABELAS COM DISTÂNCIA ENTRE CISTEÍNAS DAS ESTRUTURAS TRIDIMENSIONAIS PREDITAS	115

ÍNDICE DE FIGURAS

Figura 1: Distribuição das espécies de <i>Leishmania</i> por Estado.	2
Figura 2: Taxonomia de <i>Leishmania</i>	6
Figura 3: Ciclo Biológico de <i>Leishmania</i> spp.	8
Figura 4: Localização da cisteíno-protease em <i>L. (L.) amazonensis</i> usando anticorpo anti-papaína mono específico de coelho.	15
Figura 5: Estrutura predita para a extensão COOH-terminal da CPB de <i>L. (L.) mexicana</i> , onde três potenciais porções de epítomos são atribuídas: <i>PI</i> (Val217 ± Asp226), <i>PII</i> (Val272 ± Glu279) e <i>PIII</i> (Pro287 ± Thr300) (Alves <i>et al.</i> , 2001).	17
Figura 6: Detalhes estruturais dos epítomos selecionados, apresentando com os aminoácidos hidrofóbicos e anéis aromáticos atribuídos	18
Figura 7: Sequências de nucleotídeos e de aminoácidos da <i>cyspep</i>	20
Figura 8: Esquema geral do método teórico da modelagem comparativa.....	24
Figura 9: Acurácia e aplicação de modelos de estrutura de proteínas	26
Figura 10: Exemplo de uma equação empregada para aproximar as forças atômicas que governam o movimento molecular	30
Figura 11: Representação da Superfície Acessível ao Solvente	44
Figura 12: Predição de Estrutura Secundária gerada por quatro diferentes servidores: Jpred, NPS, Sympred e Concord.....	47
Figura 13: Alinhamento múltiplo entre as CTEs de sete espécies de <i>Leishmania</i>	47
Figura 14: Construção do modelo tridimensional da <i>cyspep</i> pelo programa Modeller	51
Figura 15: Construção do modelo tridimensional da <i>cyspep</i> pelo servidor FUGUE.....	52
Figura 16: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor <i>I-Tasser</i>	56
Figura 17: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor <i>LOMETS</i>	57
Figura 18: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor <i>MUSTER</i>	58
Figura 19: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor <i>SPARKS-X</i>	59

Figura 20: Estrutura secundária dos modelos criados por modelagem comparativa e <i>threading</i>	63
Figura 21: Estrutura secundária dos modelos criados por métodos <i>de novo</i>	64
Figura 22: Modelos selecionados para simulações de dinâmica molecular ..	66
Figura 23: Desvios RMS dos átomos da cadeia principal (backbone) em relação à estrutura inicial representado em função do tempo de simulação.	68
Figura 24: RMS de Flutuação representado por resíduo da proteína.	70
Figura 25: <i>Fator-B</i> da estrutura média e respectivos modelos representados em <i>cartoon</i>	71
Figura 26: Estrutura secundária dos modelos submetidos à dinâmica molecular.....	73
Figura 27: Representação da evolução temporal de elementos de estrutura secundária durante as simulações em diferentes modelos.	74
Figura 28: Desvios RMS, Rg e SASA dos átomos da cadeia principal (backbone) do modelo <i>Fugue</i> , através de diferentes campos de força, representados em função do tempo de simulação.....	76
Figura 29: RMS de Flutuação do modelo <i>Fugue</i> simulado através de diferentes campos de força.....	78
Figura 30: Estrutura secundária do modelo <i>Fugue</i> influenciado por diferentes campos de força	79
Figura 31: Representação da evolução temporal de elementos de estrutura secundária durante as simulações através de diferentes campos de força do modelo <i>Fugue</i>	81
Figura 32: Número de clusters ao longo de 200ns de simulação do modelo <i>Fugue</i> através de diferentes campos de força.	82
Figura 33: Desvios RMS, Rg e SASA dos átomos da cadeia principal (backbone) do modelo <i>IT1</i> , através de diferentes campos de força, representados em função do tempo de simulação.....	84
Figura 34: RMS de Flutuação do modelo <i>IT1</i> simulado através de diferentes campos de força.....	85
Figura 35: Estrutura secundária do modelo <i>IT1</i> influenciado por diferentes campos de força	86

Figura 36: Representação da evolução temporal de elementos de estrutura secundária durante as simulações através de diferentes campos de força do modelo *IT1*.....88

Figura 37: Número de clusters ao longo de 200ns de simulação do modelo *IT1* através de diferentes campos de força89

LISTA DE TABELAS

Tabela I: Predição de ligações dissulfeto dos servidores <i>DIANNA</i> e <i>SCRATCH</i>	48
Tabela II: Estruturas encontradas pelo servidor SAS como molde para <i>cyspep</i>	50
Tabela III: Moldes utilizados por cada servidor <i>threading</i> para criação de modelos <i>cyspep</i>	54
Tabela IV: Moldes utilizados para criação dos modelos nos diferentes servidores.....	60
Tabela V: Distância entre os pares de cisteínas dos modelos selecionados para simulações de dinâmica molecular.....	61
Tabela VI: Número de pares de cisteínas que formam ligação dissulfeto nos servidores de predição tridimensional de proteína.	67

LISTA DE SIGLAS E ABREVIATURAS

AIDS	Síndrome da imunodeficiência adquirida
CASP	Avaliação Crítica de Técnicas de Predição de Estrutura de Proteínas
Comp	Modelo da modelagem comparativa
CP	Cisteíno-protease
CPA	Cisteíno-protease A
CPB	Cisteíno-protease B
CPC	Cisteíno-protease C
CTE	Extensão COOH-terminal da cisteíno- protease B
Cyspep (<i>L.</i>) <i>amazonensis</i>	Extensão COOH-terminal da cisteíno- protease B de <i>Leishmania</i>
DiANNA	DiAminoacid Neural Network Application
D9	Modelo 9 do servidor <i>QUARK</i> com restrições do servidor DiANNA
DM	Dinâmica Molecular
DSSP	Dicionário da Estrutura Secundária de Proteínas
EDBCP <i>prediction server</i>	<i>Ensemble-based Disulfide Bonding Connectivity Pattern</i>
Fugue	Modelo do servidor <i>FUGUE</i>
gp63	Glicoproteína de 63 kDa
GPI	Glicosilfosfatidilinositol
HIV	Vírus da imunodeficiência humana
HLA	Antígeno leucocitário humano
I-Tasser	Iterative Threading Assembly Refinement
IUBMB	União Internacional de Bioquímica e Biologia Molecular
L3	Modelo 3 do servidor <i>LOMETS</i>
LOMETS	Local MEta-Threading Server
M4	Modelo 4 do servidor <i>MUSTER</i>
M8	Modelo 8 do servidor <i>MUSTER</i>
MHC	Complexo principal de histocompatibilidade
NMR	Ressonância Magnética Nuclear
NPS	Network Protein Sequence
MUSTER	MUlti-SourcE ThreadER
OMS	Organização Mundial da Saúde
PDB	Protein <i>Data Bank</i>
PPA	Sequence profile–profile alignment
Q5	Modelo 5 do servidor <i>QUARK</i>
Q7	Modelo 7 do servidor <i>QUARK</i>
R1	Modelo 1 do servidor <i>ROBETTA</i>
RG	Raio de Giro
RMSD	Raiz do Desvio Médio Quadrático
RMSF	Flutuação da Raiz Média Quadrática
SAS	<i>Sequence Annotated by Structure</i>
SASA	Área de Superfície Acessível ao Solvente
SCOP	Classificação Estrutural de Proteínas
SNPs	Polimorfismo de nucleotídeo único
SPC	<i>Simple Point Charge</i>
SVM	Máquina de vetor de suporte
TIP3P	<i>Transferable Intermolecular Potentials 3</i>

1 INTRODUÇÃO

1.1 Considerações Gerais

As leishmanioses são consideradas doenças endêmicas em cerca de 98 países do mundo, sendo encontradas nas Américas, África, Europa Oriental, Ásia ocidental e central, Índia e Austrália (WHO,(1990). Constituem um grupo de doenças causadas por diversas espécies de protozoários parasitos do gênero *Leishmania* que são transmitidas durante o repasto sanguíneo de insetos flebotômíneos. Esses parasitos são capazes de produzir alterações na pele, mucosas e cartilagens, caracterizando a forma tegumentar ou acometer órgãos como fígado, baço e medula óssea, configurando a forma visceral da doença (Kato, Gomez et al. 2010).

Essas manifestações clínicas parecem resultar de uma combinação das propriedades do parasito, como infectividade e patogenicidade, e dos diversos fatores do hospedeiro vertebrado, como idade, predisposição genética e estado imunológico. A expressão de fatores de virulência, propriedades antigênicas e a manifestação da doença também variam dependendo da espécie de *Leishmania* (Murray, Berman et al. 2005).

Atualmente nas Américas, são reconhecidas 11 espécies dermatrópicas de *Leishmania* causadoras de doença humana e oito espécies descritas, até o momento, que provocam a doença somente em animais. No Brasil, sete espécies de *Leishmania* causadoras da doença foram identificadas, sendo seis do subgênero *Viannia* e uma do subgênero *Leishmania*. As três principais espécies são: *L. (Viannia) braziliensis*, *L.(V.) guyanensis* e *L. (Leishmania) amazonensis* e, mais recentemente, as espécies *L. (V.) lainsoni*, *L. (V.) naiffi*, *L. (V.) lindenberg* e *L. (V.) shawi* (Figura 1)(Saúde and Epidemiológica 2007).

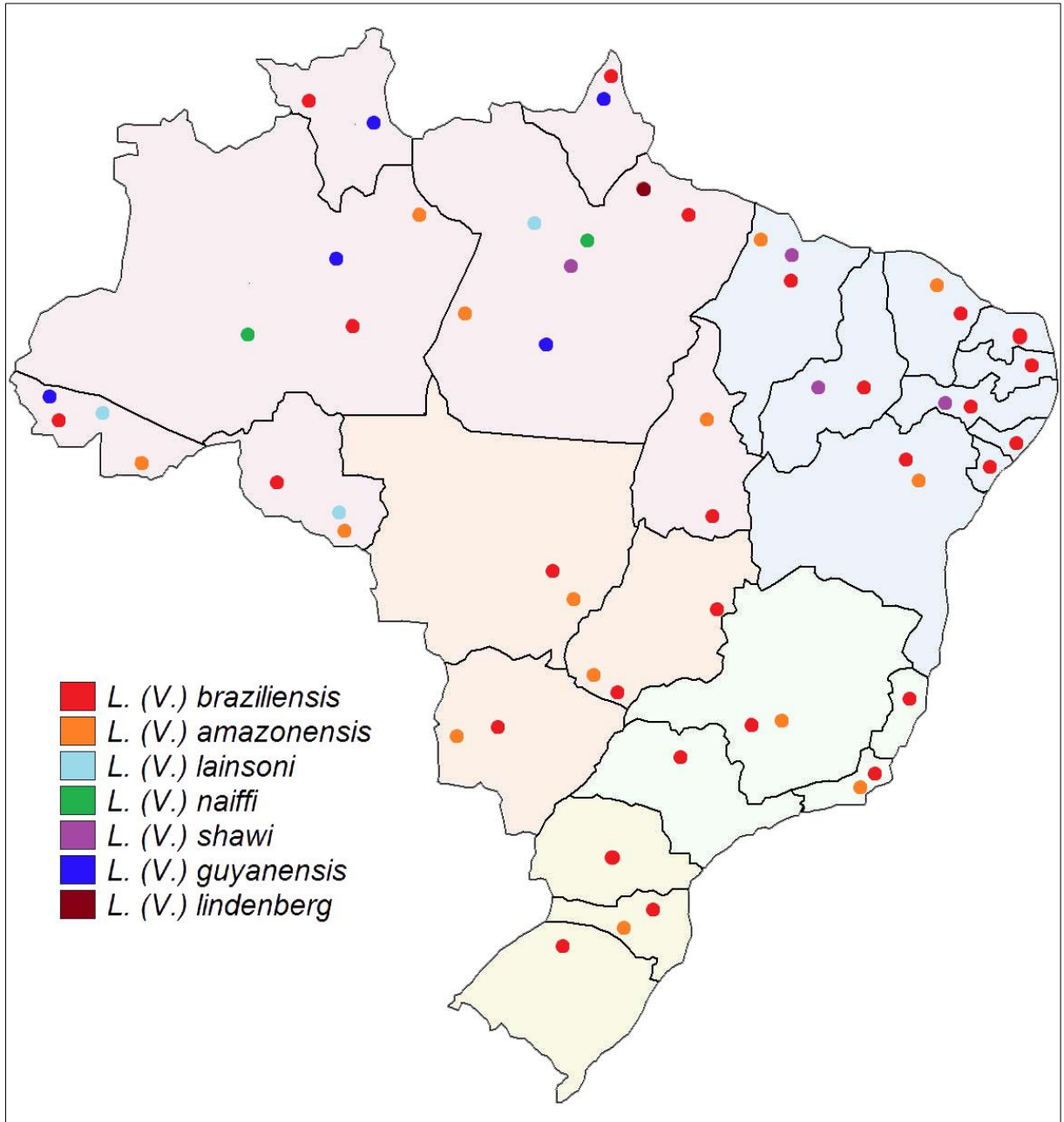


Figura 1: Distribuição das espécies de *Leishmania* por Estado. Adaptado de Ministério da Saúde/Serviço de Vigilância em Saúde/2007.

A leishmaniose também tem aparecido em indivíduos infectados com o vírus da imunodeficiência humana 1 (HIV-1) (Alvar, Aparicio et al. 2008). A co-infecção *Leishmania* / HIV é favorecida uma vez que o parasito e o vírus desenvolvem-se na mesma célula alvo, os macrófagos, e é possível imaginar que uma série de interações aconteça na célula hospedeira. Nesse contexto, estudos demonstraram que a proteína Tat (trans-ativador transcricional) do HIV-1 regula a replicação de *Leishmania* spp em macrófagos humanos. A proteína Tat recombinante é capaz de anular o efeito leishmanicida induzido pelo interferon- γ , o que permite a replicação de *Leishmania* mesmo na presença desta citocina (Barreto-de-Souza, Xavier Medeiros et al. 2008).

Se por um lado o HIV parece facilitar ou promover o desenvolvimento de leishmaniose, por outro, a co-infecção pode potencializar a progressão para a síndrome da imunodeficiência adquirida (AIDS). O parasito é capaz de promover o aumento de transcritos de HIV-1, e conseqüentemente a síntese de vírus, tanto na co-infecção de culturas *ex vivo* de tecido da amígdala de humano, como em culturas de macrófagos, podendo o aumento da produção de HIV-1 estar ligado à alta produção de citocinas inflamatórias induzidas pelo protozoário, como interferon- α e interleucina 1 (Zhao, Papadopoulou et al. 2004).

Mesmo a leishmaniose visceral sendo considerada como a segunda doença parasitária que mais origina fatalidades depois da malária no mundo, de acordo com a Organização Mundial da Saúde (OMS) (World Health 2010) ainda não existem medicamentos eficazes ou vacinas disponíveis contra esta doença. Como as populações mais afetadas pelas leishmanioses estão nos países subdesenvolvidos, indústrias farmacêuticas e agências de financiamento de países desenvolvidos têm pouco interesse no processo de desenvolvimento de medicamentos para combater a doença (Kedzierski, Sakthianandeswaren et al. 2009).

Atualmente, o tratamento das leishmanioses faz uso de antimoniais pentavalentes, e de Pentamidina e Anfotericina B; além de drogas alternativas, quando os pacientes não responderem ao tratamento com os fármacos citados anteriormente (Iwu, Jackson et al. 1994). Tais fármacos apresentam uma série de problemas, que incluem resistência do parasito e indução de efeitos colaterais, que limitam a utilização e, principalmente, a eficácia. Além disso, todos os fármacos que estão disponíveis atualmente são de administração parenteral, o que exige colaboração do paciente. Por esta causa, muitos abandonam o tratamento, fator que favorece o aparecimento de cepas resistentes (Sereno, Guilvard et al. 2001).

Com o objetivo de desenvolver drogas mais eficazes, de baixo custo e com reações adversas menos agressivas para o paciente, novos alvos quimioterápicos de *Leishmania* têm sido investigados e novas abordagens para o desenvolvimento de fármacos devem ser consideradas (Silva-López 2010). Diversas investigações estão centradas sobre o papel dos fatores de virulência que contribuem para a patogênese de *Leishmania* e permitem o parasita invadir e estabelecer a infecção no hospedeiro mamífero (Matlashewski 2001). Embora o impacto exato destes fatores de *Leishmania* nas manifestações clínicas observadas em hospedeiros mamíferos ainda não esteja definida, há evidências de que estes componentes possam modular as interações entre a *Leishmania* e células imunes do hospedeiro (Silva-Almeida, Pereira et al. 2012).

1.2 O Parasito e o Ciclo Biológico

As espécies de *Leishmania*, entre os anos de 1903 a 1950, eram classificadas baseando-se apenas nas manifestações clínicas causadas no homem e em características como espécies de vetores que transmitiam o parasita, distribuição geográfica, tropismo e propriedades antigênicas (Banuls, Hide et al. 2007).

O gênero *Leishmania* foi descrito por Ross em 1903 (Rioux, Lanotte et al. 1985) sendo que a maioria dos membros são parasitas de mamíferos. Lainson & Shaw em 1987 (Lainson and Shaw 1987), dividiram o gênero em dois subgêneros: *Leishmania* e *Viannia*, baseados na localização dos parasitas no intestino do vetor. Membros do subgênero *Leishmania* desenvolviam-se na região pilórica e do intestino médio enquanto as espécies do subgênero *Viannia* desenvolviam-se na região posterior do intestino. Além disso, utilizando-se da análise de isoenzimas por Rioux em 1990 (Rioux, Lanotte et al. 1990), foi possível definir os complexos de espécies dentro de cada subgênero (Revisado por Bañuls e colaboradores, 2007 (Banuls, Hide et al. 2007)).

Espécies como *L. guyanensis* (isolada na Guiana), *L. peruviana* (isolada no Peru), *L. infantum* (isoladas de uma criança na Tunísia) e *L. gerbilli* (isolado a partir de roedores) são exemplos de classificação em relação a distribuição geográfica. Desde os anos de 1970, critérios intrínsecos, tais como dados imunológicos, genéticos e bioquímicos são utilizados para definir as espécies de *Leishmania* (Banuls, Hide et al. 2007). Novos métodos de detecção, isolamento e identificação genética resultaram em um aumento no número de espécies descritas. O uso dessas técnicas moleculares

levou à publicação de um esquema taxonômico pela OMS (1990, World Health 2010),
mostrado na Figura 2.

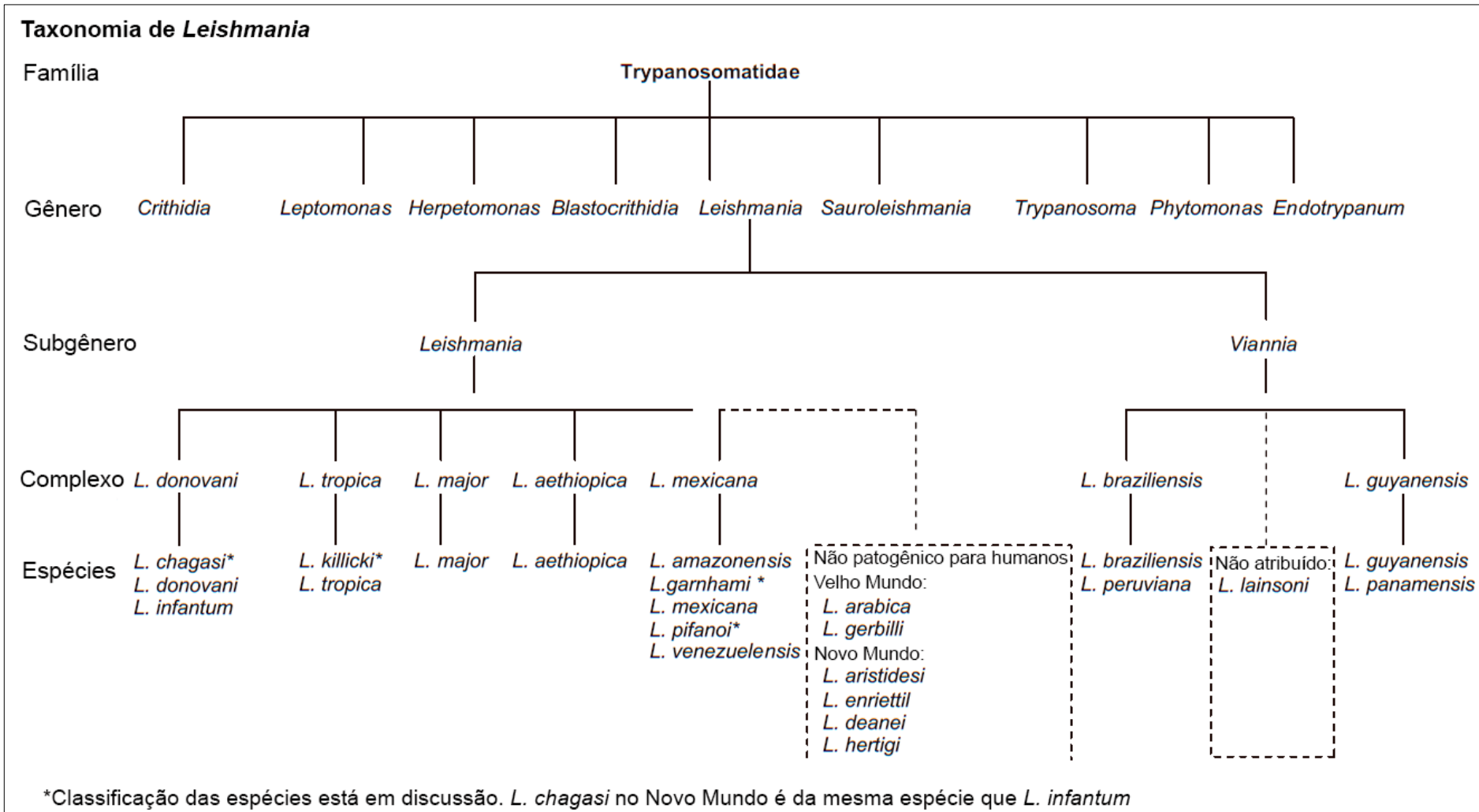


Figura 2: Taxonomia de *Leishmania*. Espécies com (*) têm sido questionadas. Adaptado de OMS, 2010.

No ciclo biológico das espécies de *Leishmania* há duas fases morfológicas distintas: os promastigotas e os amastigotas. Os promastigotas são formas extracelulares com flagelo livre, que desenvolvem-se inicialmente no intestino médio e anterior dos vetores flebotomíneos e posteriormente migram para o aparelho bucal no momento da diferenciação da forma procíclica para a metacíclica – metaciclogênese (Muskus and Marin Villa 2002). Os amastigotas são formas intracelulares sem flagelo livre, que desenvolvem-se preferencialmente nas células do sistema mononuclear fagocitário de mamíferos.

Ambos os promastigotas e amastigotas são capazes de infectar macrófagos, desde que sejam fagocitados quando interagem com as células do hospedeiro vertebrado. O promastigota de *Leishmania* entra em contato com o hospedeiro vertebrado no momento do repasto sanguíneo do inseto vetor sendo rapidamente fagocitado pelos macrófagos, fundindo-se mais tarde com o lisossoma para formar o fagolisossoma, onde ocorre a diferenciação do parasito para as formas amastigotas que residem e se multiplicam dentro dele (Cohen-Freue, Holzer et al. 2007).

Os amastigotas podem ser fagocitados novamente no hospedeiro vertebrado já infectado, quando rompem a célula hospedeira e são liberados. Assim, entram em contato com outros macrófagos e são fagocitados, reiniciando a infecção (Vannier-Santos, Martiny et al. 2002). O ciclo se completa quando um inseto vetor se alimenta no hospedeiro vertebrado ingerindo o sangue contendo macrófagos infectados por amastigotas, que então atingem o tubo digestivo, onde novamente se diferenciam em promastigotas (Mauel 1990) (Figura 3).

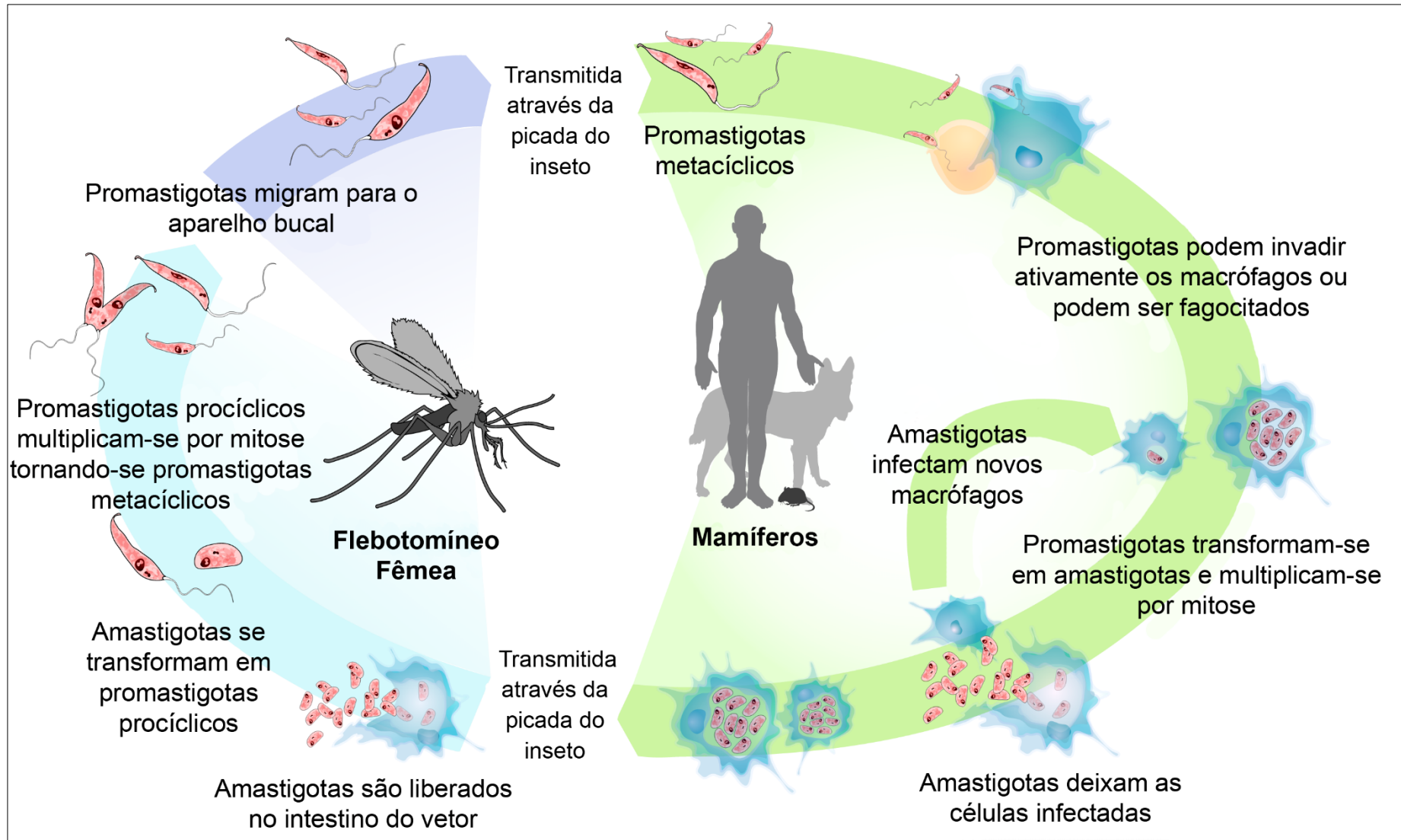


Figura 3: Ciclo Biológico de *Leishmania* spp. Adaptado de Mariana Ruiz Villarreal, 2008.

Tanto os promastigotas quanto os amastigotas apresentam uma série de peculiaridades adaptativas para cada etapa do seu ciclo biológico, as quais permitem que os parasitos escapem das adversidades que encontram nos hospedeiros, invertebrado e vertebrado (Bates and Rogers 2004). A revisão realizada por Silva-Almeida e colaboradores (Silva-Almeida, Pereira et al. 2012) evidencia fatores de virulência que contribuem para patogênese de *Leishmania* e permitem o parasita invadir e estabelecer a infecção no hospedeiro mamífero. Como exemplos citamos os glicoinositolfosfolipídeos (GIPLs) que foram relatados ao auxiliar a *L. (L.) major* sobreviver no interior de macrófagos, inibindo a enzima óxido nítrico sintase (Proudfoot, O'Donnell et al. 1995) e a proteína quinase C (Zufferey, Allen et al. 2003). O lipofosfoglicano é um ligante de macrófagos que está diretamente envolvido nos estágios iniciais da infecção que parece possuir funções não só de estabelecimento do parasito no hospedeiro vertebrado, como também de sobrevivência no inseto vetor e de transmissão do protozoário (Spath, Epstein et al. 2000).

1.3 Proteases

As proteases são importantes fatores de virulência que contribuem para a patogênese de *Leishmania*, uma vez que são enzimas que hidrolisam ligações peptídicas. São enzimas extremamente abundantes, e juntamente com seus homólogos correspondem a aproximadamente 2% dos genes codificados em alguns genomas (Rawlings and Barrett 2000, Puente, Sanchez et al. 2003). Estão distribuídas em uma grande variedade de fontes, como: plantas, animais e micro-organismos. As proteases são vistas como importantes proteínas sinalizadoras que estão envolvidas no correto funcionamento da maioria das funções biológicas. Em humanos, a atividade anormal dessas enzimas pode estar relacionada direta ou indiretamente com diversas patologias cardiovasculares e inflamatórias, cânceres e desordens neurológicas (Turk 2006).

As proteases são classificadas de acordo com a posição de clivagem do substrato. Por este critério, as proteases são divididas em endoproteases, que atuam preferencialmente nas regiões internas das cadeias polipeptídicas, e as exoproteases que clivam ligações peptídicas nas extremidades N ou C terminais das cadeias polipeptídicas e podem ser denominadas aminopeptidases e carboxipeptidases, respectivamente. As aminoproteases podem liberar um simples resíduo de aminoácido (aminopeptidase), um dipeptídeo (dipeptidases) ou um tripeptídeo

(tripeptídeo-peptidases). As carboxiproteases liberam um único aminoácido ou um dipeptídeo (Barrett 1994).

As endoproteases podem ser classificadas conforme: (i) a natureza química do sítio catalítico e; (ii) as relações referentes à sequência de aminoácidos e à estrutura da proteína:

(i) De acordo com o grupo reativo envolvido na catálise, as endoproteases podem ser divididas em quatro principais tipos: aspártico-protease, cisteíno- protease, serino-protease e metalo-protease, que têm, respectivamente, o resíduo de aminoácido ácido aspártico, cisteína ou serina no sítio ativo envolvido na catálise e as metalo-proteases utilizam íon metálico para a atividade catalítica (Rawlings and Barrett 1993).

(ii) O sequenciamento completo do genoma de diversos micro-organismos, levou a uma modernização do sistema de classificação das proteases, pela necessidade de abranger o diversificado repertório catalítico encontrado na natureza, e de catalogar e sistematizar a grande quantidade de informação produzida ao longo dos anos. Dessa forma, foi criado o banco de dados MEROPS (<http://merops.sanger.ac.uk>), que agrupa as enzimas em famílias de acordo com a homologia, e as famílias de mesma origem ancestral são agrupadas em clãs. Este tipo de informação é determinado pela estrutura terciária das enzimas (Rawlings, Waller et al. 2014).

1.4 Proteases de *Leishmania* spp

Em *Leishmania* spp, as proteases desempenham um papel fundamental na interação entre o parasito e o hospedeiro, uma vez que têm o potencial de degradar proteínas e peptídeos que participam de uma gama de funções biológicas. Entre as funções exercidas pelas proteases podemos citar: invasão ao hospedeiro, escape do sistema imune do hospedeiro, resistência do parasito à terapia medicamentosa, metabolismo de proteínas ou peptídeos biologicamente ativos e, assimilação dos aminoácidos como fonte nutricional ou para síntese de substâncias orgânicas (Barrett 1994).

1.4.1 Aspártico- proteases

As aspártico-proteases geralmente atuam em condições ácidas o que limita sua função em locais específicos nos diferentes organismos e as tornam a classe proteolítica menos abundante entre o grupo de proteases (Dash, Kulkarni et al. 2003).

No entanto, essas enzimas têm sido isoladas e estudadas em uma ampla gama de organismos, de vertebrados a retrovírus, como alvos para intervenção terapêutica visto a sua participação em vários processos fisiológicos e patológicos importantes (Eder, Hommel et al. 2007).

Dentro deste contexto, foi relatada a presença de atividade de aspártico-protease em formas promastigotas de *Leishmania mexicana* e *Leishmania infantum*, que são sensíveis a éster metílico de diazoacetil-DL-norleucina (DAN) e a inibidores específicos de aspártico-proteases de HIV, Ac-Leu-Val-Phenilalanina, Nelfinavir (NFV) e Saquinavir. Também verificou-se que na presença de inibidores de aspártico-protease, as formas amastigotas e promastigotas desses parasitos reduziram a atividade proliferativa com uma percentagem significativa de células binucleadas e, em alguns casos, células multinucleadas, indicando um bloqueio da citocinese na divisão celular (Valdivieso, Rangel et al. 2010).

Estudos também verificaram que as aspártico-proteases de *Leishmania major* parecem processar peptídeos de 28 kDa em lisossomas e contribuir para a degradação de cadeias invariáveis em células apresentadoras de antígenos de ratos infectados (Zhang, Maekawa et al. 2000). Ainda foi constatado que sua atividade diminui durante a transformação de promastigotas para amastigotas em *Leishmania amazonensis* (Alves, Corte-Real et al. 2005). Recentemente foi realizado o isolamento da proteína Ddi1-like de *Leishmania major* e verificado que esta possui atividade de aspártico-proteases (Perteguer, Gomez-Puertas et al. 2013).

Até a data, dois genes de aspártico-proteases foram identificados dentro do genoma de *Leishmania*: (i) um gene que codifica para uma aspártico-proteases pertencente à família A22 que apresenta similaridade sequencial com a presenilina 1, e hidrolisa proteínas de membrana do tipo I, sugerindo que esteja envolvida na autofagia (Besteiro, Williams et al. 2007); (ii) o outro gene codifica uma segunda enzima que apresenta identidade sequencial com uma peptidase que cliva peptídeos sinais no interior das membranas e tem distribuição ubíqua em eucariotos (Perteguer, Gomez-Puertas et al. 2013).

1.4.2 Serino- proteases

A atividade desta classe de peptidases em *Leishmania* foi primeiramente identificada por meio da purificação de sua proteína em extrato intracelular hidrossolúvel de promastigotas de *Leishmania (L.) amazonensis* e caracterizada como uma oligopeptidase B. Estudos posteriores sugeriram que a oligopeptidase B, durante

a diferenciação da forma amastigota, é regulada e participa do revestimento de superfície do parasita juntamente com as enzimas enolase e plasminogênio, o que permite aos amastigotas se replicarem sem serem detectados no interior do macrófago (Silva-Almeida, Pereira et al. 2012).

Mutantes de *L. (L.) major* com deficiência de oligopeptidase B mostraram capacidade de crescer normalmente como promastigotas, embora apresentem deficiência na diferenciação para promastigotas metacíclicos. Essas formas foram menos capazes de infectar e sobreviver dentro de macrófagos em experimentos *in vitro*, apesar de manter a sua virulência *in vivo* em camundongos. Estes dados sugerem que a deficiência de oligopeptidase B em *L. (L.) major* não é um fator de virulência prevalente, mas sim faz parte de um conjunto de fatores, indicando que diferenças funcionais entre *Leishmania* influenciam sua interação com o hospedeiro mamífero (Munday, McLuskey et al. 2011).

A subtilisina, outra classe de serino-proteases, de *Leishmania donovani* foi clonada e demonstrou possuir uma importante capacidade catalítica. Quando o gene da subtilisina foi suprimido, a capacidade de diferenciação de promastigota para amastigota, *in vitro*, foi reduzida. Além disso, a atividade desta serino-protease de *Leishmania* é aumentada significativamente em amastigotas quando comparada com formas promastigotas, sugerindo um papel importante dessa enzima quando o parasita habita no hospedeiro vertebrado (Swenerton, Knudsen et al. 2010).

1.4.3 Metallo- proteases

Dentre os fatores moleculares que contribuem para a virulência das *Leishmania* spp destaca-se a glicoproteína de 63 kDa (gp63), uma metalo-protease dependente de zinco. Essa enzima é um dos principais componentes de superfície presente em todas as espécies de *Leishmania*, especialmente em promastigotas, representando mais de 1% do total de proteínas do parasito (Jaffe and Dwyer 2003).

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

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

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

1.4.4 Cisteíno-proteases

As cisteíno-proteases (CPs) são fatores de virulência com propriedades imunomoduladoras durante a infecção de *Leishmania*. De maneira geral, as CPs são enzimas conhecidas por desempenhar papéis essenciais na patogênese de infecções de protozoários parasitas, e o seu potencial como alvos de drogas e vacina tem sido investigado exaustivamente. Neste cenário, a ação de inibidores de CP sobre o controle da infecção pode ser interpretada como uma evidência da importância destas enzimas durante o estabelecimento da infecção no hospedeiro (Selzer, Pingel et al. 1999).

As CPs mais estudadas nas espécies de *Leishmania* são as CPA, CPB e CPC, as quais apresentam relações evolutivas com a papaína (em relação à estrutura molecular e à atividade catalítica) (Silva-Almeida, Pereira et al. 2012). CPA e CPB também possuem características similares com a catepsina L e a CPC com a catepsina B. Vários genes de CP foram caracterizados em *Leishmania*, principalmente em espécies do complexo *L. (L.) mexicana*, tais como *Leishmania (L.) mexicana*

(Robertson and Coombs 1994), *Leishmania (L.) pifanoi* (Marin-Villa, Vargas-Inchaustegui et al. 2008) e *Leishmania (L.) amazonensis*. Também tem sido observado que o polimorfismo de nucleotídeo único (SNPs) que ocorre em CPs, pode variar de acordo com o estágio de vida do parasito, podendo estar relacionado com as características clínicas dermatrópicas em vez de estar relacionado a um estado viscerotrópico (Hide and Banuls 2008).

A CPB apresenta características exclusivas quando comparadas a outras CPs: uma extensão de cerca de 100 aa, na sua região COOH-terminal com presença de múltiplas cópias organizadas em tandem no genoma do parasita. Foi evidenciado que a CPB de *L. (L.) mexicana* e *L. (L.) amazonensis* podem ter um papel na resposta do equilíbrio Th1/Th2 através da clivagem do complexo principal de histocompatibilidade (MHC – *major histocompatibility complex*) de classe II no vacúolo parasitóforo do hospedeiro; clivagem da interleucina IL-2; indução na expressão de receptores IL-4; inibição da produção de IL-12 nos macrófagos e células dendríticas (Pereira, Silva et al. 2011).

Uma evidência da importância das CPBs como um fator essencial de virulência para as espécies do complexo de *L. (L.) mexicana* foi obtida em ensaios de infecções experimentais com camundongos BALB/c utilizando parasitas geneticamente modificados em que os genes para a CPA, CPB e CPC foram deletados (cepas Δ cpa, Δ cpb e Δ cpc) (Mottram, Coombs et al. 2004). Estes ensaios mostraram que a ausência de CPA ou CPC afetaram os parasitas de um modo mais discreto do que a ausência da CPB. Parasitas que não possuíam CPB mostraram uma reduzida capacidade de infectar e induzir lesões do que a estirpe de tipo selvagem (Alexander, Coombs et al. 1998).

1.4.5 Extensão COOH-terminal da CPB (CTE)

A CPB pode atuar sobre o sistema imune do hospedeiro através da sua extensão COOH-terminal. Essa extensão é hidrolisada quando a CPB é processada para a sua forma madura (Robertson and Coombs 1994) no interior do macrófago (Alves, Corte-Real et al. 2005). A CTE e/ou CPB não processada foi observada no citoplasma de células hospedeiras através de marcação com ouro coloidal direcionada com anticorpo para um peptídeo da CTE (Figura 4). Há indícios de que este polipeptídeo atue na modulação do sistema imune do hospedeiro vertebrado (Alves, Corte-Real et al. 2005).

A partir de estudos *in silico* das características físico-químicas, como carga elétrica, hidrofobicidade e ponto isoelétrico, para predição de epítomos de linfócitos T, foi possível a identificação de três sequências de aminoácidos derivadas da CTE de *L. (L.) mexicana* compatíveis para interações com as proteínas do antígeno leucocitário humano (HLA – *human leukocyte antigen*) de classe II (*PI*: VMVEQVICFD; *PII*: VGGGLCFE e; *PIII*: YFLGSIMNTCHYT). Esses epítomos demonstraram a capacidade de estimular a proliferação de linfócitos oriundos da circulação periférica de pacientes com leishmaniose em ensaios *in vitro* (Alves, Pontes de Carvalho et al. 2001).

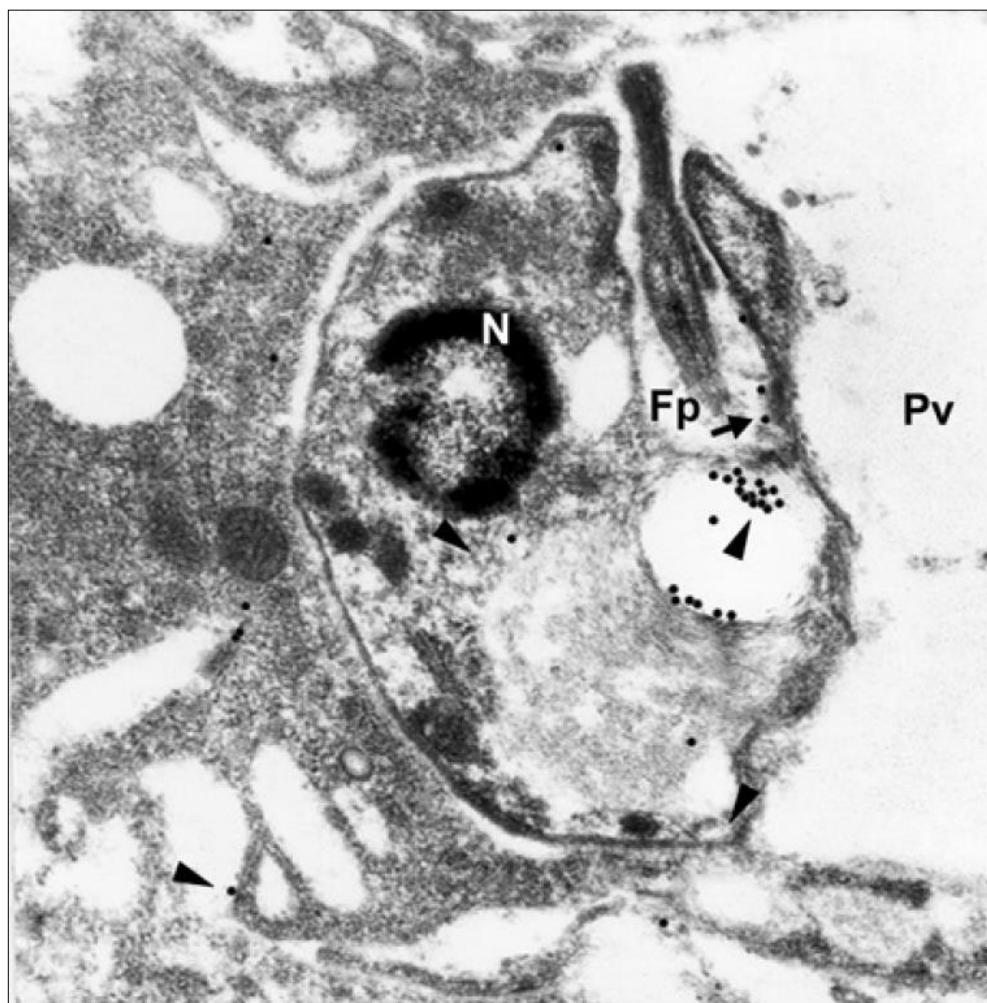


Figura 4: Localização da cisteíno-protease em *L. (L.) amazonensis* usando anticorpo anti-papaína monoespecífico de coelho. Partículas são vistas em associação com a membrana de superfície (seta), vesículas intracelulares e bolso flagelar (ponta de seta). N: núcleo; Fp: bolsa flagelar; Pv: vacúolo parasitóforo (Alves *et al.*, 2005).

Devido à importância imunológica da CTE, foi proposto um estudo *in silico* com o propósito de reconstruir sua estrutura tridimensional de *Leishmania (L.) mexicana*. Os procedimentos de modelagem molecular utilizadas nesse estudo foram guiados a partir da homologia presente na família das cisteíno-proteases. A região madura da cisteíno-protease está intimamente relacionada com papaína e, de fato, a papaína é considerada um arquétipo das cisteíno-proteases apresentando elevada identidade com esta família de enzimas (Alves, Figueiredo et al. 2001).

Como não foram encontrados moldes para CTE para realização da modelagem comparativa, foi construída uma quimera entre o modelo estrutural da papaína combinada com a sequência de aminoácidos da CTE. Essa combinação permitiu modelar a quimera com uma glicosiltransferase. O alinhamento sequencial entre a quimera e o glicosiltransferase mostrou locais de identidade em toda a extensão da proteína apesar da baixa identidade global de 37% (Alves, Figueiredo et al. 2001).

A estrutura tridimensional predita para CTE demonstrou existência de porções organizadas em α -hélice e de porções com estrutura linear. As três sequências de peptídeos derivados da CTE (*PI*, *PII* e *PIII*) estão localizadas em regiões não estruturadas do modelo proposto (Figura 5). Apenas a porção inicial do *PI* encontra-se em região de α -hélice, *PII* e *PIII* apresentaram estrutura *coil* (Figura 6). Em vista da organização espacial e da hidrofobicidade dos aminoácidos dos peptídeos derivados, apoia-se a hipótese que essas porções da CTE possam interagir com moléculas do MHC de classe II (Alves, Figueiredo et al. 2001).

O uso desses mesmos epítomos preditos em ensaios com camundongos infectados com *L. (L.) amazonensis* demonstrou a capacidade de um deles (*PI*) induzir a expressão de citocinas relacionadas à resposta Th2 em células cultivadas a partir de camundongos BALB/c, ao passo que as células de camundongos CBA expostas ao mesmo peptídeo expressam citocinas relacionadas com ambas as respostas Th1 e Th2. A proliferação de linfócitos T também foi estimulada pelo peptídeo sintético, especialmente a de linfócitos T CD8+. Além disso, quando usado para ensaios *in vivo* com animais infectados, o *PI* causou o aumento do crescimento de lesões de camundongos BALB/c, mas não para os camundongos CBA (Alves, Benevolo-De-Andrade et al. 2004).

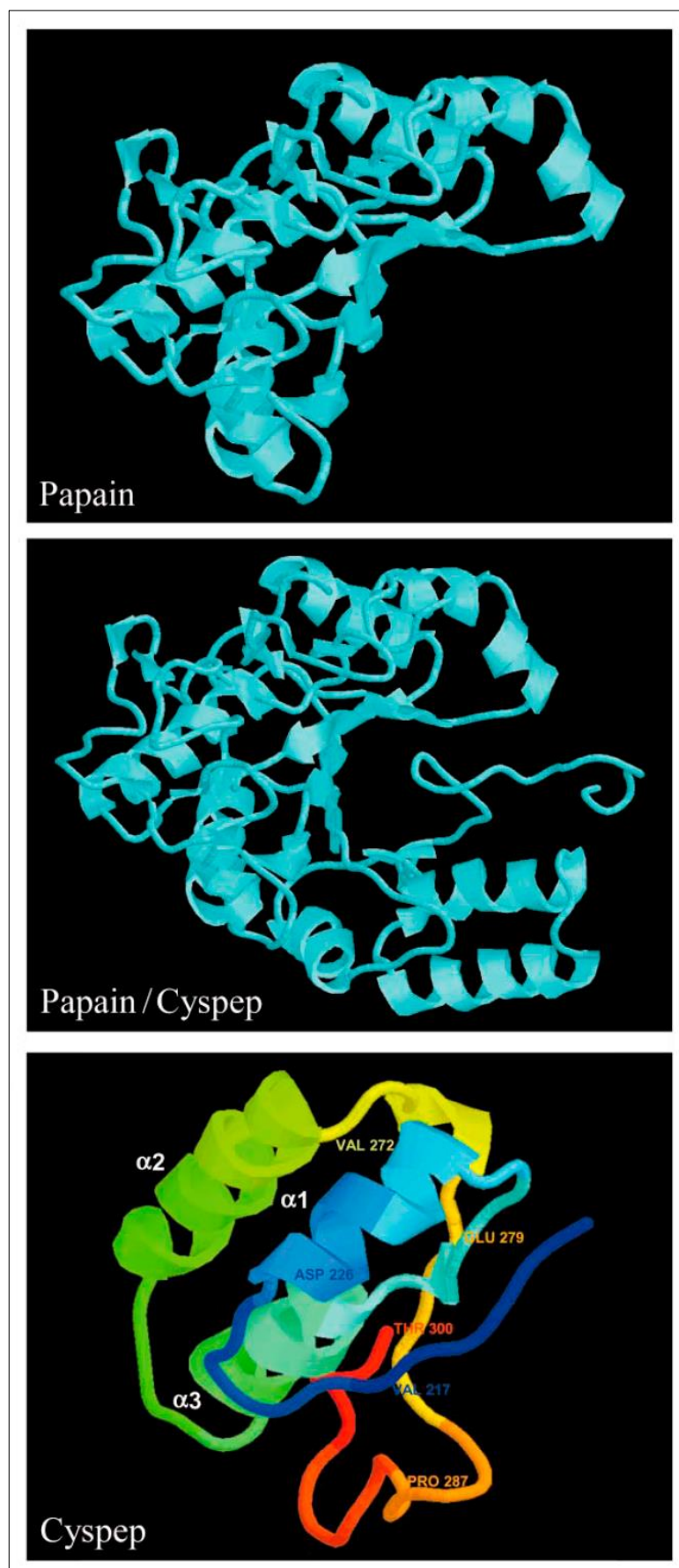


Figura 5: Estrutura predita para a extensão COOH-terminal da CPB de *L. (L.) mexicana*, onde três potenciais porções de epítopos são atribuídas: *PI* (Val217 \pm Asp226), *PII* (Val272 \pm Glu279) e *PIII* (Pro287 \pm Thr300) (Alves et al., 2001).

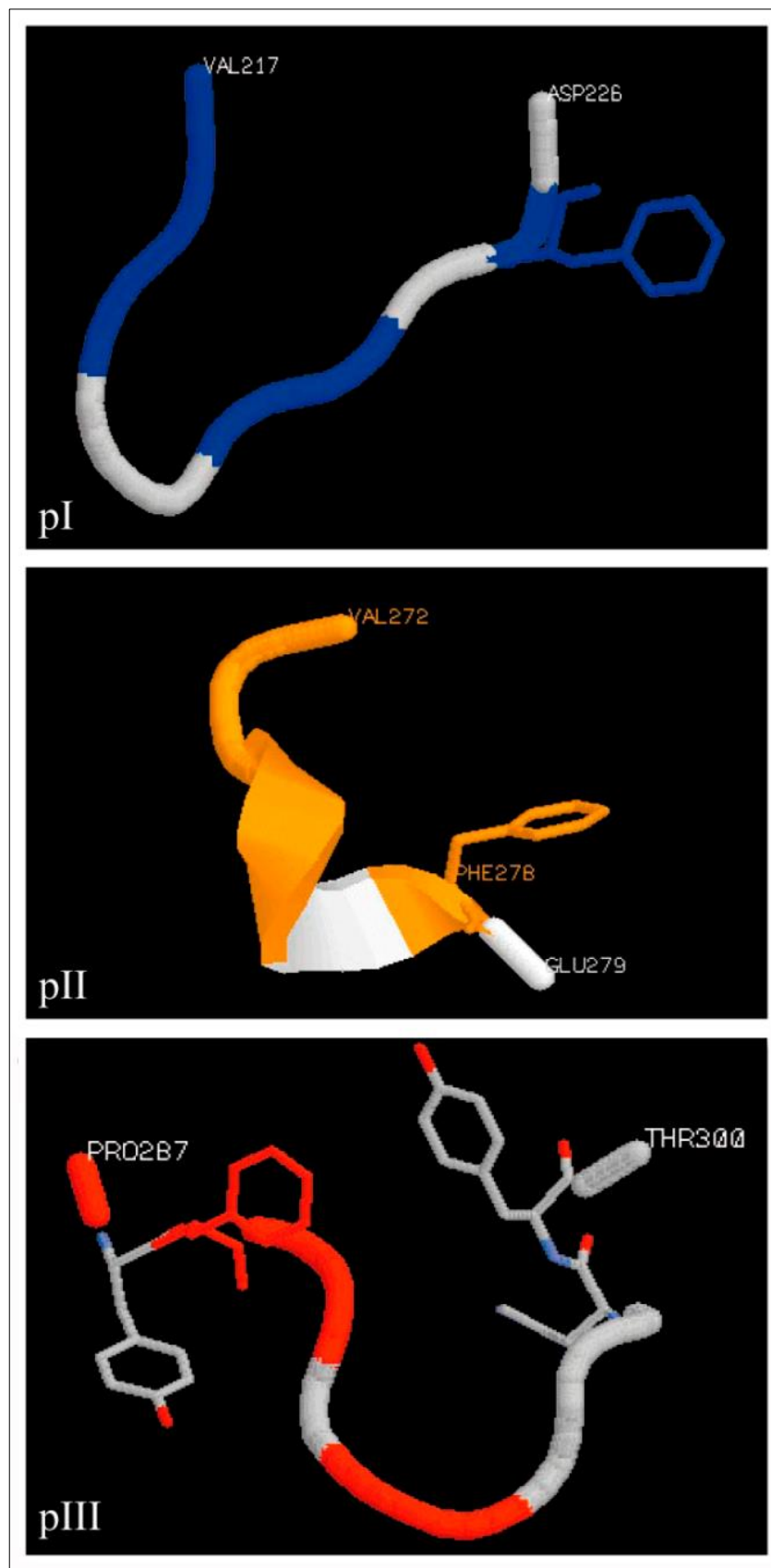


Figura 6: Detalhes estruturais dos epítomos selecionados, apresentando com os aminoácidos hidrofóbicos e anéis aromáticos atribuídos. *PI*, com aminoácidos hidrofóbicos em azul; *PII*, com aminoácidos hidrofóbicos em laranja e; *PIII* com aminoácidos hidrofóbicos em vermelho (Alves *et al.*, 2001).

A CTE pode também desempenhar um papel na infecção de macrófagos com *L. (L.) pifanoi* ou *L. (L.) amazonensis*. A incubação dos parasitas com os anticorpos anti-CTE antes da sua interação com os macrófagos levou a uma redução no número de células infectadas. Este efeito foi mais decisivo com as formas amastigotas do que com promastigotas. A pré-incubação dos parasitas com anticorpo específico para o domínio catalítico da CPB de *L. (L.) pifanoi* não apresentou efeitos na infecção de macrófagos (Marin-Villa, Vargas-Inchaustegui et al. 2008).

Com o intuito de encontrar novos peptídeos imunomoduladores derivados da extensão COOH-terminal da CPB de *L. (L.) amazonensis* (*cyspep*), Pereira e colaboradores (2011) realizaram seu sequenciamento, a sequência nucleotídica e proteica está presente na Figura 7. Essa sequência foi analisada por algoritmos *online* de predição de clivagem proteossômica com o objetivo prever quais sequências seriam originadas a partir da clivagem da *cyspep*. Através deste procedimento foi então possível obter nove epítomos a serem avaliados como candidatos a ligantes do MHC (Pereira, Silva et al. 2011).

Experimentos *in vitro*, com camundongos de diferentes níveis de susceptibilidade à infecção por *L. (L.) amazonensis*, demonstraram que o peptídeo *P6* (FCLGGGLCL) estava relacionado com o aumento nos níveis de interleucina 10 em culturas de células de nódulos linfáticos de camundongos BALB/c infectados com *L. (L.) amazonensis*. Além disso, esse peptídeo também interferiu no equilíbrio de populações de diferentes células sendo consequência de uma redução na percentagem de linfócitos T CD8+ em culturas estimuladas pelo *P6* (Pereira, Silva et al. 2011).

```

5' - gcaccagacccgtgatgggtggagcaggtgatctgcttcgataagaactgcactcaggggtgcaagaaaaccctgatcaaggcgaacgagtgc
Nt- A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C

cacaagaacgggggaggaggcgcctctatgatcaagtgcagtcgcagacgggtgacgatgtgcacgtactcgaacgaattctgcctgggcggggggctg
H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L

tgcctcgagactcctgatggtaagtgcgcgcogtactttttgggctcggtcattaacacctgtcactacacg -3'
C L E T P D G K C A P Y F L G S V I N T C H Y T -Ct

```

Figura 7: Sequências de nucleotídeos (acima) e de aminoácidos (abaixo) da *cyspep*. As sequências estão transcritas nas ordenações de 5' → 3' e NH₃-terminal (Nt) → COOH-terminal (Ct) para nucleotídeos e aminoácidos, respectivamente. Os aminoácidos (representados pelo código de 1 letra) estão organizados abaixo do primeiro nucleotídeo de sua tríade de nucleotídeos (Pereira *et al.*, 2011).

Novos experimentos foram realizados para verificar a estabilidade da ligação dos epítomos *PI* e *P6* da *cyspep* com receptores H-2 de camundongos infectados por *L. (L.) amazonensis* e seus efeitos na resposta imune no hospedeiro mamífero. Foram realizados testes de proliferação e quantificação de linfócitos T reativos aplicando análise de citometria utilizando amostras de drenagem de linfonodos de lesões de camundongos infectados por *L. (L.) amazonensis* (Souza-Silva, Pereira et al. 2014).

A análise computacional das sequências dos peptídeos *PI* e *P6* por meio de algoritmos *online* de predição de clivagem proteossômica identificou sete sequências derivadas que teoricamente são capazes de se ligar às moléculas de MHC. Dessas, cinco peptídeos induziram a blastogênese em células de BALB/c, enquanto apenas dois apresentaram a mesma propriedade em células de camundongos C57BL/6 (Souza-Silva, Pereira et al. 2014).

Além disso, epítomos derivados da *cyspep* podem estar envolvidos na modulação da infecção em camundongos, uma vez que linfócitos T CD8+ reativos aos epítomos foram detectados em linfonodos de drenagem da lesão de camundongos BALB/c e C56BL/6 infectados por *L. (L.) amazonensis*. Observou-se que a taxa de proliferação induzida pelos peptídeos derivados da *cyspep* foi diferente em culturas de células de ambos os haplotipos de camundongos testados. Estes fatos reforçam a hipótese de que a resposta aos diferentes peptídeos derivados da *cyspep* pode variar entre linhagens murinos com níveis distintos de suscetibilidade à infecção (Souza-Silva, Pereira et al. 2014).

1.5 Predição de Estruturas Tridimensionais de Proteínas

A função de uma proteína está intimamente associada à sua estrutura tridimensional. (Pain, Hames et al. 1994) Essa é a afirmativa fundamental que inspira todas as buscas por um método que seja capaz de prever a estrutura nativa de uma proteína a partir da sua sequência de aminoácidos. Tal método poderia ajudar na compreensão e no melhor aproveitamento do potencial contido na grande quantidade de informação biológica, na forma de sequências, que vem sendo gerada graças ao sucesso dos projetos genoma.

A determinação experimental ainda é considerada o melhor processo para se obter a estrutura tridimensional de uma proteína. Entretanto estas técnicas, além de serem financeiramente custosas, podem levar anos e, em alguns casos, a estrutura final pode não ser elucidada. Portanto, o desenvolvimento de métodos computacionais

é tanto uma alternativa mais viável quanto, em alguns casos, a única possibilidade de obtenção de modelos estruturais para algumas proteínas.

A complexidade do estudo das conformações adotadas por uma proteína durante o seu enovelamento até a conformação nativa pode ser ilustrada no chamado de paradoxo de Levinthal. Esse paradoxo estabelece que, se uma proteína fosse se enovelar tentando sequencialmente todas as conformações possíveis, levaria um tempo considerável, da ordem de milhares de anos, mesmo que as conformações se sucedessem a uma taxa muito rápida (Zwanzig, Szabo et al. 1992).

Baseado na observação de que as proteínas se enovelam muito mais rapidamente, Levinthal (Levinthal 1969), propôs que uma busca conformacional aleatória é muito improvável e que uma proteína deveria, portanto, enovelar-se por um processo dirigido. Nesse aspecto, determinar a estrutura tridimensional nativa a partir da sequência de aminoácidos permanece como um dos maiores problemas da ciência moderna, tratando-se de uma questão profundamente multidisciplinar e abrangendo diversas áreas da ciência como engenharia, biologia, física, química e computação científica.

Os métodos desenvolvidos para a predição da estrutura de proteínas podem ser organizados em dois principais grupos: métodos baseados em estruturas molde, que incluem modelagem comparativa e *threading*; e métodos independentes de estruturas molde, chamados de primeiros princípios, que englobam os métodos *de novo* e *ab initio*. O que irá ditar a escolha do método a ser aplicado é a presença ou ausência de estruturas resolvidas experimentalmente que possam ser usadas como molde para a modelagem da sequência alvo. A escolha do método está intrinsecamente relacionada com a porcentagem de identidade obtida a partir do alinhamento entre a sequência alvo e possíveis candidatos a molde.

1.5.1 Modelagem Comparativa

A modelagem comparativa baseia-se no conhecimento de que a conformação estrutural de uma proteína é mais conservada que sua sequência de aminoácidos durante o processo evolutivo, e que pequenas mudanças na sequência, em geral, resultam em, apenas, sutis modificações na estrutura tridimensional (da Silva 2007). Desta forma, a abordagem permite que proteínas que não possuem estruturas tridimensionais determinadas experimentalmente possam ter sua estrutura predita a partir da sua sequência de aminoácidos. Trata-se de um método empírico que utiliza o alinhamento construído entre a sequência da proteína a ser modelada (sequência

alvo) e a sequência de uma proteína relacionada estruturalmente (molde) (Sanchez and Sali 1999).

A modelagem comparativa é o método frequentemente mais empregado, e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. Geralmente, consideram-se como limites mínimos de aplicabilidade do método valor de 30% de identidade entre a sequência primária da proteína de interesse e o molde, para proteínas com mínimo de 100 aminoácidos na sua composição.

O processo de obtenção de um modelo proteico virtual através da execução da estratégia da modelagem comparativa envolve quatro etapas principais (Martí-Renom, Stuart et al. 2000), entre elas: (i) identificação dos moldes; (ii) alinhamento das sequências; (iii) construção do modelo e; (iv) validação dos mesmos (Figura 8) (Sternberg 1996).

1.5.2 Abordagem *threading*

A abordagem *threading* é um método de predição da estrutura de proteína usado com intuito de descobrir homologias remotas que não podem ser descobertas por alinhamento padrão de sequências. A metodologia consiste em identificar se o padrão de enovelamento de uma proteína com estrutura desconhecida é semelhante ao de proteínas já caracterizadas estruturalmente. Se fragmentos da sequência se ajustam bem aos enovelamentos, um alinhamento pode ser deduzido, mesmo que não haja informação suficiente para construir um modelo completo (Gibas and Jambeck 2001).

As limitações dos métodos de *threading* surgem de dois pontos principais: (i) como observado para a modelagem comparativa, se a identidade entre a sequência alvo e a sequência molde for muito baixa, é possível que o enovelamento daquela sequência simplesmente não corresponda a estrutura nativa da sequência de interesse, podendo construir um modelo completamente errado. (ii) Outra limitação é que os modelos gerados por *threading* apresentam uma resolução relativamente baixa, dificultando seu uso em estudos que exigem posicionamento preciso dos átomos como no caso do *docking* molecular de ligantes no sítio da proteína (Gibas and Jambeck 2001).

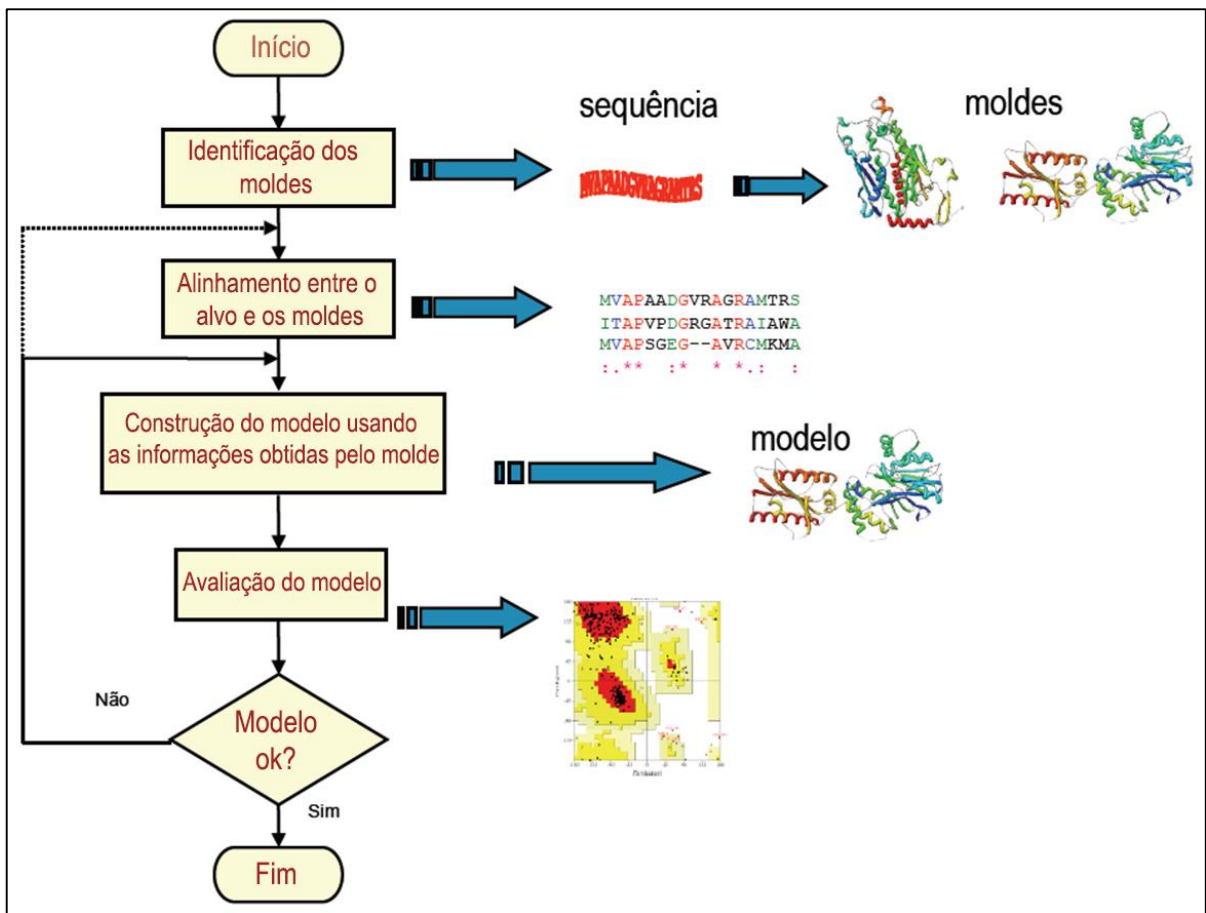


Figura 8: Esquema geral do método teórico da modelagem comparativa. Adaptado de MARTY- RENOM *et al.*, 2000.

1.5.3 Método de novo

Embora a modelagem comparativa e a abordagem *threading* permitam a obtenção de modelos satisfatórios, tais técnicas são inválidas se as proteínas de referência, com estruturas determinadas experimentalmente, não se encontrarem disponíveis. Para possuir a independência de moldes de proteínas homólogas, foram desenvolvidos métodos que usam informações provenientes de bancos de estruturas de proteínas determinadas empiricamente, resultando no tipo de predição chamada *de novo*. Dentre as principais técnicas usadas pela predição *de novo* destacam-se o uso da predição de estruturas secundárias, uso de fragmentos de proteínas, e modificação da função de energia (Dahiyat and Mayo 1997).

1.5.4 Método ab initio

O método *ab initio* se destaca como sendo a tentativa mais ambiciosa para a resolução do problema de predição de estrutura de proteínas. Essa abordagem difere das demais por não usar informações de estruturas conhecidas, relacionadas com a sequência alvo, e por usar funções de energia contendo somente termos de significado físico. Tal estratégia é baseada em dois pressupostos: (i) todas as informações necessárias sobre a estrutura de uma proteína estão contidas em sua sequência de aminoácidos; (ii) e acredita-se que as proteínas enovelam-se para um estado nativo, ou um conjunto de estados nativos, que se encontra no (ou próximo ao) mínimo global de energia livre.

Em geral, o método *ab initio* se restringe a proteínas com número de aminoácidos menor que 100. O procedimento de busca da melhor estrutura da proteína tem como base a hipótese termodinâmica, onde o mínimo de energia livre é o suposto estado nativo da proteína (Xu and Zhang 2012).

1.5.5 Acurácia e aplicação de modelos de estrutura de proteínas

A aplicabilidade de um modelo tridimensional está diretamente relacionada com a acurácia com que este foi gerado (Figura 9). A precisão de um modelo gerado por modelagem comparativa está relacionada com o percentual de identidade sequencial em que baseia-se e o correlaciona com a relação entre a similaridade estrutural e a sequência de duas proteínas (Baker and Sali 2001).

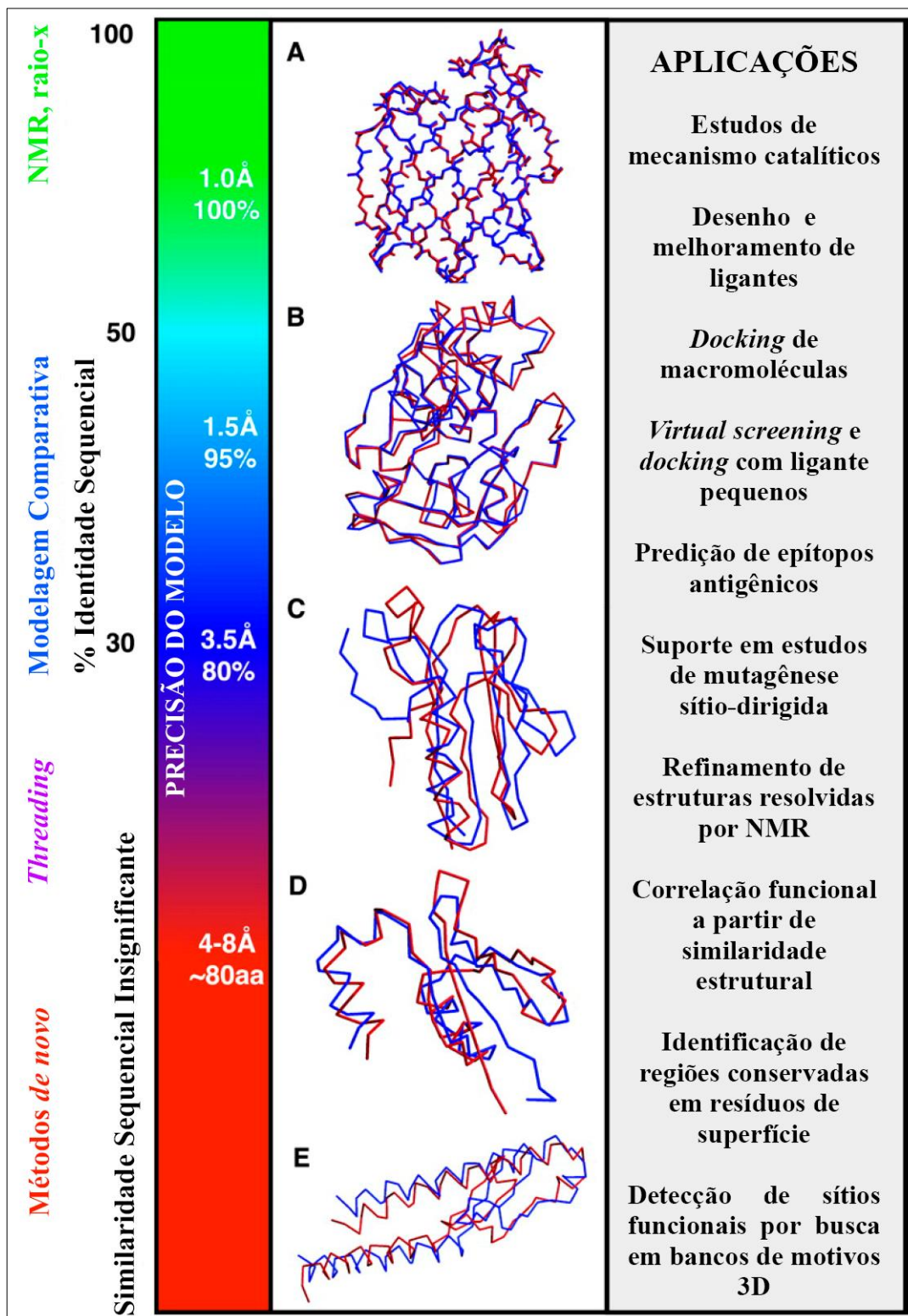


Figura 9: Acurácia e aplicação de modelos de estrutura de proteínas. São mostrados os diferentes intervalos de aplicabilidade da modelagem comparativa, *threading*, e métodos *de novo*; a acurácia correspondente dos modelos estruturais de proteínas; e exemplos de aplicações. (A a C) Exemplos de modelos comparativos baseados em cerca de 60% (A), 40% (B), e 30% (C) de identidade sequencial com a sua estrutura molde. (D e E) Exemplos de predições do método *de novo* com software Rosetta para o experimento CASP4. Predições de estruturas estão em vermelho, e estruturas reais em azul. Adaptado de Baker e Sali, 2001.

Modelos comparativos de alta precisão são baseados em mais de 50% de identidade sequencial com seus moldes. Os erros encontrados nestes casos estão, normalmente, relacionados com pequenas mudanças ou distorções das regiões da cadeia principal, erros na cadeia lateral e em regiões de *loops*. Os modelos de precisão intermediária são baseados em 30 a 50% de identidade sequencial. Nesses modelos, há maior frequência de erros na cadeia lateral e na modelagem de regiões de *loops*, além da existência de ocasionais erros de alinhamento (Baker and Sali 2001).

Já os modelos de baixa precisão, baseados em valores de identidade sequencial menores de 30 %, os erros de alinhamento tornam-se mais frequentes. Além disso, quando um modelo baseia-se num alinhamento muito baixo (abaixo de 25%) para uma estrutura conhecida pode acarretar em enovelamento incorreto. Outros fatores, como a seleção de moldes e a precisão do alinhamento geralmente têm maior impacto sobre a precisão do modelo, especialmente para os modelos baseados em estruturas com menos de 40 % de identidade sequencial para os moldes (Baker and Sali 2001).

A exatidão e confiabilidade dos modelos produzidos por métodos *de novo* são muito mais baixas do que a dos modelos gerados por modelagem comparativa com base em alinhamentos com mais de 30 % de identidade sequencial. Porém a topologia básica de uma proteína ou domínio pode, em alguns casos, ser razoavelmente predita (Figura. 9, D e E) (Baker and Sali 2001).

Os modelos de baixa resolução, produzidos por métodos *de novo*, podem revelar relações estruturais e funcionais entre proteínas com sequências de aminoácidos diferentes e fornecer uma estrutura que permite analisar relações espaciais entre os resíduos conservados ou entre resíduos funcionalmente importantes. Desta forma, predições por métodos *de novo* podem fornecer ideias funcionais importantes que podem ser obtidas a partir de estruturas determinadas experimentalmente (Baker and Sali 2001).

1.6 Dinâmica Molecular

Existe uma complexa rede de entidades químicas que são capazes de criar dinamicamente a vida a nível molecular. Por exemplo, as proteínas e os ácidos nucléicos adotam uma estrutura específica de acordo com a sua função, os íons são transportados através das membranas, enzimas podem desencadear uma cascata de

reações químicas etc. Devido à complexidade dos sistemas biológicos, os métodos computacionais tornaram-se cada vez mais importantes nas ciências da vida. Os sistemas computacionais mais rápidos, poderosos e complexos podem ser explorados através de simulações computacionais (Cai, Li et al. 2012).

A Dinâmica molecular (DM) surgiu como um dos primeiros métodos de simulação aplicados à dinâmica dos líquidos por Alder e Wainwright (Alder and Wainwright 1957, Alder and Wainwright 1959) e por Rahman (Rahman 1964) no final dos anos 1950 e início dos anos 1960. Devido aos avanços na tecnologia e melhorias de algoritmos, a DM tornou-se uma ferramenta valiosa em muitas áreas da física e da química. Desde os anos 1970 a DM tem sido amplamente utilizado para estudar a estrutura e a dinâmica de macromoléculas.

Experimentos desempenham um papel central na ciência. A riqueza de resultados experimentais fornece uma base para a compreensão do mecanismo químico de vida. Técnicas experimentais, tais como a difração de raios X ou NMR, permitem a determinação da estrutura e da elucidação da função de grandes moléculas de interesse biológico. No entanto, o experimento só é possível em conjunto com modelos e teorias. As simulações computacionais têm alterado a interação entre o experimental e a teoria. A essência da simulação é a utilização do computador para modelar um sistema físico (Karplus and McCammon 2002).

Segundo a IUPAC (*International Union of Pure and Applied Chemistry*), a “dinâmica molecular é um procedimento de simulação que consiste na computação do movimento dos átomos em uma molécula ou de átomos individuais ou moléculas em sólidos, líquidos e gases, de acordo com as leis de movimento de Newton” (Van de Waterbeemd, Carter et al. 1997). Em outras palavras, trata-se de um método de simulação que acompanha a evolução temporal dos sistemas estudados em escala atômica e molecular, obtida a partir da integração de equações de movimento (segunda lei de Newton) para cada átomo do sistema. A força que atua sobre cada átomo é obtida a partir das informações presentes nos campos de força clássicos (Kukol 2008).

1.6.1 Campo de Força

De maneira simplificada, atribui-se a denominação de campo de força à descrição de um sistema de muitas partículas pela sobreposição de termos simples. A solução numérica das equações de movimento a partir de coordenadas e velocidades iniciais, para o movimento de cada átomo submetido ao campo de força,

é obtida através de um algoritmo computacional (Allen and Wilson 1989, Frenkel and Smit 1996). Todos eles baseiam-se na resolução das equações de Newton a partir de uma expansão em série de potências da posição de cada partícula e em cada incremento no tempo.

Cada campo de força estabelece um conjunto de equações matemáticas dedicados a reproduzir aspectos do comportamento molecular, como o estiramento de ligações químicas, a deformação de um ângulo de ligação ou a torção de um diedro. Estas equações, por sua vez, são parametrizadas para reproduzir o comportamento dos compostos de interesse (Verli 2014). Os campos de força clássicos são compostos por duas classes de funções potenciais (Figura 10) (Cornell, Cieplak et al. 1995, Durrant and McCammon 2011):

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

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

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

1.6.1.1 Campo de Força AMBER

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

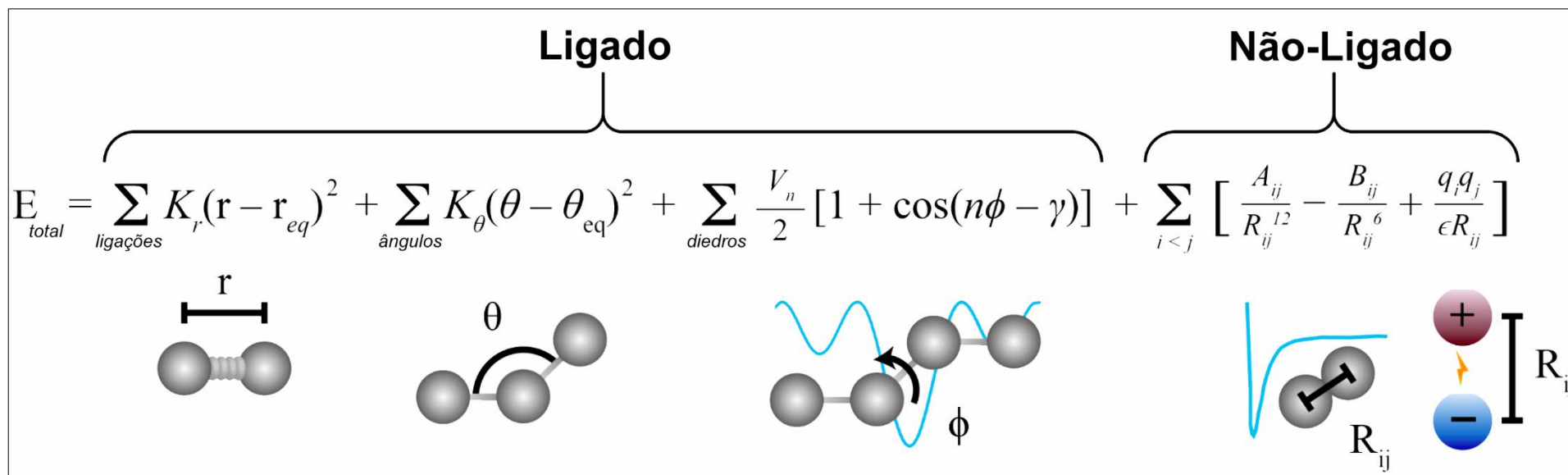


Figura 10: Exemplo de uma equação empregada para aproximar as forças atômicas que governam o movimento molecular. As forças atômicas que governam o movimento molecular podem ser divididas entre aquelas causadas por interações entre átomos que estão quimicamente ligados uns aos outros, e as causadas por interações entre átomos que não estão ligados. As ligações químicas e ângulos atômicos são modelados usando molas simples, e ângulos diedro (ou seja, as rotações sobre uma ligação) são modelados usando uma função sinusoidal que aproxima as diferenças de energia entre conformações eclipsadas e escalonados. Interações não-ligadas surgem devido a forças de *Van der Waals* que são modeladas utilizando o potencial de *Lennard-Jones*; e interações eletrostáticas são modeladas utilizando a lei de *Coulomb*. Adaptado de Durrant e McCammon, 2011.

1.6.1.2 Campo de Força CHARMM

Este campo de força utiliza as mesmas regras de potenciais não ligados que AMBER: potencial de *Lennard-Jones* mais cargas pontuais para átomos soltos ou distantes e as regras de mistura de *Lorentz-Berthelot*. As diferenças são que as interações intramoleculares apresentam dinâmicas do potencial tipo harmônico tanto para estiramento de ligações quanto para torções angulares e ângulos diedros do tipo impróprio, além de termo não escalados do tipo *Coulomb*. Torções angulares do tipo impróprias são utilizadas em casos especiais com fins de restrição de planaridade de resíduos aromáticos (Mackerell, Feig et al. 2004).

1.6.1.3 Campo de Força GROMOS

O campo de força GROMOS é um campo de força baseado na unificação dos átomos. Neste caso, os átomos de hidrogênio apolares, ou seja, aqueles ligados a átomos de carbono, são unidos a este elemento, dando origem a um pseudo-átomo representando as propriedades de grupos CH, CH₂ ou CH₃. Exceção se dá para o grupo CH de anéis aromáticos, que tem os átomos de hidrogênio descritos explicitamente. É parametrizado para usar as mesmas formas funcionais que AMBER, sem a regra geral de mistura de *Lennard-Jones*. As relações intramoleculares são diferenciadas. Flexibilidade das ligações é produzida pelo potencial de estiramento de quarta potência, enquanto ângulos se retorcem com torções do tipo próprias e os diedrais são tratados pela série de cossenos, *Lennard-Jones* e termos de *Coulomb* não escalados (Oostenbrink, Villa et al. 2004).

1.7 Justificativa

A grande quantidade de produtos do processamento das CPs que são lançados em macrófagos parasitados e no ambiente extracelular pode ter um significado relevante nos mecanismos adaptativos das *Leishmania* spp. Entre esses produtos destaca-se a CTE cuja função biológica não está adequadamente definida. Até o presente momento esse polipeptídeo foi verificado quanto à ação de seus peptídeos derivados na modulação da resposta imune do hospedeiro vertebrado. No entanto, também existe a possibilidade da CTE atuar diretamente sobre o hospedeiro vertebrado durante a infecção, influenciando em vários eventos do processo

inflamatório, como: *(i)* aglutinação e adesão celular; *(ii)* migração transendotelial e transepitelial em auxílio à migração de células do sistema imune ao sítio inflamatório; *(iii)* ação lítica direta sobre células do sistema imune ativadas; *(iv)* ativação/inativação de macrófagos; entre outras.

Assim, oportuno se faz investigar a estrutura tridimensional da CTE como uma das estratégias para procurar funções para este polipeptídeo, uma vez que a função de uma proteína pode ser determinada pela sua estrutura. Macromoléculas como proteínas têm uma grande diversidade estrutural, que se reflete nas múltiplas funções que assumem em sistemas vivos.

2 OBJETIVOS

2.1 Objetivo Geral

Identificar possíveis modelos da estrutura tridimensional da região COOH-terminal da cisteino- protease B de *Leishmania (Leishmania) amazonensis* (cyspep).

2.2 Objetivos Específicos

- Construir modelos para estrutura tridimensional desta proteína mediante utilização de diferentes metodologias computacionais.
- Analisar o comportamento dinâmico dos modelos propostos da proteína de interesse.
- Avaliar a estabilidade das estruturas secundárias preditas nos modelos obtidos.

3 MATERIAL E MÉTODOS

3.1 Predição de Estrutura Secundária

A predição da estrutura secundária consenso da *cyspep* foi realizada por meio de quatro distintos servidores:

Jpred <<http://www.compbio.dundee.ac.uk/www-jpred/>> (Cole, Barber et al. 2008): a predição é gerada a partir de seis algoritmos que empregam diferentes métodos, tais como: redes neurais (NNSSP (Salamov and Solovyev 1995)), redes neurais baseadas em decisão (PHD (Rost and Sander 1993)), discriminação linear (DSC (King, Saqi et al. 1997)), métodos combinados de sequências singulares (MULPRED (Barton, 1988, não publicado)), propriedades de ligação de hidrogênio (Predator (Frishman and Argos 1996)), e predição ponderada do número de conservação (ZPRED (Zvelebil, Barton et al. 1987)). O arquivo de saída contém informação sobre propriedades físico-químicas e sobre a acessibilidade do solvente para cada um dos aminoácidos testados.

Sympred <<http://www.ibi.vu.nl/programs/sympredwww/>> (Simossis 2004): é um servidor de predição de estrutura secundária consenso que usa algoritmos de programação dinâmica (PHDpsi (Rost and Sander 1993), PROFsec (Rost, não publicado), SSPro2.01 (Pollastri, Przybylski et al. 2002), YASPIN (Lin, Simossis et al. 2005), JNet (Cuff and Barton 2000) e PSIPred (Jones 1999)), com alto nível de precisão. O servidor vale-se de combinações de algoritmos que utilizam redes neurais para realizar a predição a partir de alinhamentos de múltiplas sequências homólogas.

NPS (*Network Protein Sequence*) <https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_sopma.html> (Combet, Blanchet et al. 2000): realiza a predição de estrutura secundária consenso por meio da seleção de métodos para predição de estrutura secundária individuais como: DPM (Deleage and Roux 1987), DSC (King and Sternberg 1996), GOR I (Garnier, Osguthorpe et al. 1978), GOR III (Gibrat, Garnier et al. 1987), GOR IV (Garnier, Gibrat et al. 1996) MLRC (Guermeur, Geourjon et al. 1999), PHD (Rost and Sander 1993), Predator (Frishman and Argos 1996), SOPMA (Geourjon and Deleage 1995), SYMPA96 (Levin, Robson et al. 1986). Posteriormente, gera a predição de estrutura secundária consenso para cada um dos aminoácidos na proteína, mostrando também a porcentagem da estrutura secundária da proteína obtida em cada um dos servidores de predição.

Condord <<http://helios.princeton.edu/CONCORD/>> (Wei 2011): obtém a predição de estrutura secundária consenso por meio da utilização de vários métodos como PSIPRED (Jones 1999), DSC (King, Saqi et al. 1997), GOR IV (Garnier, Gibrat et al. 1996), Predator (Frishman and Argos 1996), Prof (Ouali and King 2000), PROFphd (Rost and Sander 1994), e SSpro (Cheng, Randall et al. 2005). O servidor apresenta um esquema consenso baseado no método de otimização linear inteiro misto que visa maximizar o número de aminoácidos corretamente preditos e atinge uma melhor precisão na predição. O arquivo de saída contém informações de nível de confiança para cada um dos aminoácidos na organização da estrutura secundária predita. A pontuação de predição consenso é baseada na pontuação de confiança de predições de aminoácidos dos sete métodos que é introduzida para avaliar a probabilidade de um aminoácido possuir um dos elementos de estrutura secundária. A pontuação de confiança é definida como:

$confS(p,i,m)$: a pontuação de confiança predita pelo método m para o aminoácido i da proteína p , $p \in P, i \in subsetPI(P, l), m \in M$

onde l : conjunto de posições dos aminoácidos de uma proteína, $i \in l$; P : conjunto de proteínas de treinamento, $[1 \dots 3000]$, $p \in P$; M : conjunto dos métodos utilizados, $[1,2,3,4,5,6,7]$; $subsetPI(P, l)$: subconjunto indicando o número de aminoácidos de cada proteína p .

3.2 Predição das Ligações Dissulfeto

Para predizer se a *cyspep* possui ligações dissulfeto foram utilizados quatro servidores *online* que implementam técnicas específicas de aprendizagem automatizado:

DiANNA (*DiAminoacid Neural Network Application*) <<http://clavius.bc.edu/~clotelab/DiANNA/>> (Ferre and Clote 2005) realiza a predição do estado de oxidação das cisteínas, sendo então calculada a probabilidade de formação de uma ligação dissulfeto para cada par de cisteínas na sequência. A execução se dá através de uma rede neural treinada que utiliza a predição da estrutura secundária da proteína utilizando o algoritmo PSIPRED (Jones 1999), assim como o perfil proveniente do alinhamento múltiplo gerado por meio de sequências obtidas pelo algoritmo PSIBLAST. DiANNA também realiza uma classificação ternária (Ferre and Clote 2006) através da aplicação de máquina de vetor de suporte (SVM) para determinar se uma cisteína é reduzida (em estado de sulfidrilas livres), meia-cisteína

(envolvida em ligação dissulfeto) ou ligada a um metal ou carbono. Neste último caso, DiANNA prediz se o metal é ferro, zinco, cádmio.

Scratch < <http://scratch.proteomics.ics.uci.edu/> > (Cheng, Randall et al. 2005) utiliza o algoritmo Dipro para predição de cisteínas envolvidas em ligação dissulfeto através da execução de rede neurais. No servidor Scratch, esse algoritmo foi otimizado pela adição de informações sobre estrutura secundária e acessibilidade ao solvente relativa como insumos para aumentar a precisão da predição de ligação dissulfeto, além do uso de SVM para discriminar proteínas com ou sem ligações dissulfeto.

EDBCP (*Ensemble-based Disulfide Bonding Connectivity Pattern prediction server*) < <http://biomedical.ctust.edu.tw/edbcp/> > (Lin and Tseng 2010) fornece a predição da ligação de dissulfeto por conectividade padrão baseada em informações estruturais das proteínas.

DISULFIND <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1538823/>> (Ceroni, Passerini et al. 2006). Nesse servidor, o estado de ligação dissulfeto de cada cisteína é predito pela primeira vez por um classificador binário. Em seguida, as cisteínas que são conhecidas por participar na formação de pontes são combinadas para se obter a conectividade padrão.

3.3 Modelagem da Estrutura Tridimensional da Cyspep

A construção do modelo tridimensional da *cyspep* através da predição estrutural foi proposta por diferentes métodos: Modelagem Comparativa, *Threading* e Métodos *de novo*.

3.3.1 Modelagem Comparativa

A construção de modelos por modelagem comparativa ocorreu de modo manual através do programa *Modeller* (Sali, Potterton et al. 1995), e de forma automatizada utilizando o servidor FUGUE (Shi, Blundell et al. 2001), que realiza busca por homólogos distantes.

No modo manual, a estrutura da *cyspep* foi obtida através da modelagem comparativa, tendo como molde a estrutura da proteína lectina de germe de trigo, depositada no banco de dados de proteínas (PDB – *Protein Data Bank*) sob o código 1WGC (Wright 1990), usando o programa *Modeller* versão 9.11 (Sali, Potterton et al. 1995). A estrutura da lectina foi identificada como um molde adequado através da

ferramenta *Sequence Annotated by Structure* (SAS) (Milburn, Laskowski et al. 1998), que realiza a aplicação de informações estruturais para uma determinada sequência de proteínas. O alinhamento entre a sequência alvo (*cyspep*) e seu molde foi gerado pelo pacote de alinhamento *T-Coffee* (Notredame, Higgins et al. 2000). Foram construídos 50 modelos, dentre os quais o de menor valor de função objetiva, calculada pelo programa Modeller, foi escolhido para posterior análise. A função objetiva é uma medida de quanto o modelo satisfaz as restrições espaciais de entrada, sendo que valores mais baixos indicam um melhor ajuste e, portanto, modelos de maior precisão (Sali, Potterton et al. 1995).

No modo automatizado utilizamos o servidor *FUGUE* <<http://tardis.nibio.go.jp/fugue/prfsearch.html>> que é um programa de reconhecimento de homólogos distantes por comparação de sequência-estrutura (Shi, Blundell et al. 2001). Ele baseia-se nas premissas da modelagem comparativa e pressupõe que cada resíduo em uma estrutura terciária existe em um ambiente particular, podendo ser descrito por características como: conformação da cadeia principal, acessibilidade ao solvente, e estado de ligação de hidrogênio.

O programa *FUGUE* utiliza tabelas de substituição ambiente-específico, derivadas de alinhamentos estruturais na base de dados *HOMSTRAD* (Mizuguchi, Deane et al. 1998), para melhorar a predição de estrutura secundária e o reconhecimento de enovelamento. Além disso, *FUGUE* usa penalidades de *gap*, dependentes de estrutura, que calculam a penalidade para os resíduos de acordo com sua posição em relação aos elementos de estrutura secundária e da acessibilidade ao solvente, resultando em alinhamentos mais precisos.

Em suma, o programa varre um banco de dados de perfis estruturais, calcula a pontuação de compatibilidade da sequência-estrutura e produz uma lista de potenciais homólogos e alinhamentos. No caso da *cyspep*, o molde encontrado foi uma proteína de função desconhecida codificada pelo gene pa3566 de *Pseudomonas aeruginosa* cujo código PDB é 1X7V (Sanders, Walker et al. 2005).

3.3.2 Threading

3.3.2.1 I-Tasser

O servidor *I-Tasser* (*Iterative Threading Assembly Refinement*) <<http://zhanglab.ccmb.med.umich.edu/I-TASSER/>> (Roy, Kucukural et al. 2010) combina os métodos *threading*, *ab initio* e refinamento estrutural para construção de

estruturas tridimensionais. A predição é constituída basicamente por três etapas: (i) Alinhamento multi-fragmentos, a partir da sequência de aminoácidos. O servidor primeiramente tenta recuperar proteínas moldes com dobras similares (superestruturas secundárias) a partir da biblioteca PDB do *LOMETS* (Wu and Zhang 2007), uma abordagem *meta-threading* instalada localmente. (ii) Agrupamento, fragmentos contínuos são extraídos das estruturas moldes e reagrupados em modelos completos por meio de simulações de Monte Carlo do tipo troca entre réplicas (*replica-exchange Monte Carlo*) utilizando *threading*, regiões não alinhadas (principalmente os *loops*) são construídos por método *ab initio*. (iii) Refinamento, a montagem de conformações por fragmentos é executada novamente utilizando as estruturas-modelos (centróides) da etapa anterior. Essa segunda interação é realizada para refinar os centróides eliminando possíveis interferências estéricas. *Clusters* (agrupamentos de estruturas) são gerados novamente e as configurações de menor energia são selecionadas. Por fim, realiza-se uma otimização da energia das ligações de hidrogênio.

3.3.2.2 MUSTER

O *MUSTER* (*MU*lti-*S*ource *T*hreadER) (Wu and Zhang 2008) <<http://zhanglab.ccmb.med.umich.edu/MUSTER/>> é um servidor que usa um algoritmo *threading* empregando uma extensão de “*sequence profile–profile alignment*” (PPA), que combina uma variedade de sequências e recursos estruturais gerados por meio de várias ferramentas como (i) perfis derivados de sequência, (ii) estruturas secundárias, (iii) perfis derivados de estrutura, (iv) acessibilidade ao solvente, (v) ângulos de torção (ângulos psi e phi) e (vi) matriz de pontuação hidrofóbica.

O principal objetivo deste método é otimizar o reconhecimento de enovelamentos de proteínas por meio de diferentes recursos de características estruturais que são combinadas com o método de PPA. A identificação dos melhores alinhamentos entre a sequência de interesse e as sequências moldes é dada por meio do algoritmo de programação dinâmica. A saída do servidor *MUSTER* inclui os dez principais moldes e alinhamentos, os modelos são gerados utilizando o programa Modeller.

3.3.2.3 LOMETS

LOMETS (*Local MEta-Threading Server*) (Wu and Zhang 2007) <<http://zhanglab.ccmb.med.umich.edu/LOMETS/>> é um servidor que utiliza a técnica de *meta-threading*, ou seja, gera as predições de estruturas tridimensionais a partir de modelos consenso de uma variedade de servidores individuais. A predição ocorre a partir de nove diferentes algoritmos individuais que foram selecionados para cobrir diferentes métodos de *threading*:

- FUGUE (Shi, Blundell et al. 2001) , emprega programação dinâmica.
- PROSPECT2 (Xu and Xu 2000), utiliza uma função de pontuação que combina penalizações obtidas no alinhamento com potencial de contato, propensão da estrutura secundária e mutação dos resíduos.
- HHSEARCH (Söding 2005), emprega modelos ocultos de Markov.
- SPARKS2 e SP3 (Zhou and Zhou 2004, Zhou and Zhou 2005), empregam programação dinâmica no alinhamento, sendo que o SPARKS2 emprega um potencial estatístico baseado em conhecimento.
- SAM-T02 (Karplus, Karchin et al. 2003), baseado no alinhamento múltiplo do PSI-Blast, também utiliza modelo oculto de Markov.
- PPA-I (Wu and Zhang 2007), utiliza uma abordagem baseada em alinhamento do tipo PPA com conhecimento da estrutura secundária.
- PPA-II (Wu and Zhang 2007), assemelha-se ao PPA-I, sendo que aqui o perfil é coletado do a partir de alinhamentos de sequências SAMT99.
- PANT (Wu and Zhang 2007), utiliza função de energia que emprega mutação dos resíduos, conhecimento e estrutura secundária e penalizações do alinhamento.

Os modelos selecionados por *LOMETS*, a partir de servidores individuais, são puramente baseados em consenso. Para o melhor desempenho, restrições espaciais são combinadas a partir do consenso dos 20 principais alinhamentos *threading*. As restrições englobam uma série de termos, a saber: (i) contatos de cadeia lateral; (ii) contatos dos átomos do C", que descreve os contatos preditos dos átomos de C α entre pares de resíduos; (iii) distâncias de curto alcance de átomos de carbono que descreve as distâncias dos C α preditos em curto alcance e; (iv) distâncias de C α de longo alcance que descreve as distâncias C α preditas a longo prazo. Ao final, os 10 melhores modelos são exibidos pelo servidor.

3.3.2.4 SPARKS-X

SPARKS-X <<http://sparks-lab.org/yueyang/server/SPARKS-X/>> (Yang, Faraggi et al. 2011) é uma otimização método *threading* SPARKS que altera a função de pontuação no alinhamento e incorpora as técnicas SPINE-X que realizam melhores previsões da estrutura secundária, ângulos de torção e superfície acessível solvente.

3.3.3 Métodos de novo

3.3.3.1 QUARK

O servidor *QUARK* <<http://zhanglab.ccmb.med.umich.edu/QUARK/>> (Xu and Zhang 2012) realiza a previsão da estrutura tridimensional de proteínas sem a utilização de moldes. A criação do modelo é dividida em três passos: (i) Obtenção do perfil da sequência por meio da previsão de estrutura secundária, valores de ângulos de torção, acessibilidade ao solvente e posições de β -turn; seguido da geração de fragmentos da sequência de consulta com 1-20 resíduos de comprimento. (ii) Simulações para construção da estrutura utilizando o método de simulações de Monte Carlo do tipo troca entre réplicas guiada pelo campo de força *QUARK* baseado em conhecimentos de nível atômico e; (iii) Agrupamento dos fragmentos e refinamento *full-atom*.

3.3.3.2 ROBETTA

O servidor *ROBETTA* <<http://www.robetta.org/>> (Chivian, Kim et al. 2005) utiliza métodos *de novo* e a modelagem comparativa para realizar suas previsões. Quando a proteína-alvo não possui estruturas homólogas, o algoritmo Rosetta *de novo* é empregado. O protocolo *de novo* do *ROBETTA* utiliza fragmentos de estruturas existentes no PDB a fim de orientar um conjunto de fragmentos com funções de energia. Esse protocolo chamado de "reconstrói e refina" é composto por uma etapa que emprega um modelo simplificado, chamado modelagem de baixa resolução, seguida de outra etapa que utiliza um modelo *full-atom*, chamado modelagem de alta resolução (Bonneau, Tsai et al. 2001).

Na etapa de baixa resolução, utiliza-se um campo de força reduzido e as regiões da proteína a ser modelada são construídas a partir de bibliotecas de fragmentos, que são gerados a partir de variedade de estruturas diferentes e não-homólogas de tamanhos variados. Na etapa de alta resolução emprega um modelo *full-atom* e a minimização de energia é realizada com um algoritmo de Monte Carlo.

3.4 Dinâmica Molecular

Para a determinação de propriedades dinâmicas e estruturais da *cyspep* segundo características do meio em que se encontra, aplicamos a técnica computacional de Dinâmica Molecular.

O conjunto de programas utilizado para simulações de dinâmica molecular dos modelos gerados por diferentes técnicas de modelagem tridimensional de proteínas foi o pacote GROMACS (Pronk, Pall et al. 2013). As simulações foram realizadas utilizando diferentes campos de força *AMBER99SB ILDN* (Lindorff-Larsen, Piana et al. 2010), *CHARMM22* (Mackerell, Feig et al. 2004) e *GROMOS96 56a6* (Oostenbrink, Villa et al. 2004) com seus respectivos tipos de modelos de água recomendados *Transferable Intermolecular Potentials 3 (TIP3P)* e *Simple Point Charge (SPC)* (Mahoney and Jorgensen 2000).

Todos os resíduos de cisteína envolvidos em pontes dissulfeto foram devidamente assinalados como tais nos arquivos de topologia e as geometrias destas pontes devidamente otimizadas durante as etapas de minimização. Os modelos foram submetidos a 1000 passos de minimização utilizando o algoritmo *steepest descent*, com uma tolerância de 1000 kJ / mol e as coordenadas finais da minimização foram aproveitadas como posições iniciais da dinâmica de restrição.

As simulações foram realizadas em duas etapas: dinâmica de restrição, mediante aplicação de potenciais harmônicos aos átomos pesados da proteína, para acomodação das moléculas de água e a dinâmica molecular para coleta de trajetória. Para o estágio dinâmica de restrição, um tempo de 100 picossegundos (ps) foi simulado com as posições atômicas da proteína restritas mediante a aplicação de potenciais harmônicos considerando as posições iniciais como referência e usando constante de força igual 1000 kJ/mol/nm². O segundo passo foi uma simulação sem restrições, realizada por 200 nanossegundos (ns) ou estendida até 500ns, dependendo da estrutura simulada.

O tratamento eletrostático utilizado em todas as fases da simulação foi *Particle Mesh Ewald (PME)*. Os raios de corte foram 1,0 e 1,4 nm, para o potencial eletrostático e de Lennard-Jones, respectivamente. Condições periódicas de contorno foram aplicadas. O acoplamento térmico de Berendsen modificado (*v-rescale*) foi utilizado para manter a temperatura em 300K, com um tempo de relaxação de 10 fs e acoplamento de pressão de Berendsen isotrópico. Os parâmetros da dinâmica molecular são mostrados no Anexo A.

3.4.1 *Análise das Simulações de Dinâmica Molecular*

3.4.1.1 *Raiz do Desvio Médio Quadrático (RMSD)*

A RMSD é empregada como uma medida para avaliar a diferença estrutural entre moléculas. No caso da dinâmica molecular, o RMSD pode ser empregado para calcular o desvio global de RMSD durante a trajetória de simulação, e avaliar a oscilação do sistema durante o tempo decorrido. Esse procedimento foi realizado por meio da ferramenta *g_rms* do pacote GROMACS (Pronk, Pall et al. 2013). A análise do valor de RMSD é importante, pois valores uniformes e que se mantêm por baixo de certo limite (esse limite depende da proteína em geral) indicam comportamento em que a estrutura se desvia pouco da posição inicial, enquanto que mudanças bruscas na média denotam mudanças importantes na conformação. A RMSD se define matematicamente mediante a expressão:

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

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

3.4.1.2 *Flutuação da Raiz Média Quadrática (RMSF)*

A flutuação do valor de RMS é representada a partir da ferramenta RMSF, uma estratégia complementar à RMSD, que permite obter informação sobre a flexibilidade estrutural local. O RMSF calcula o desvio padrão quadrático médio das posições atômicas ao longo da simulação com relação às posições dos átomos ou resíduos de uma estrutura como referência. Tal cálculo foi obtido com a ferramenta *g_rmsf* do pacote GROMACS (Pronk, Pall et al. 2013). Esta ferramenta também permite converter os valores RMSF para valores de fator Debye-Waller (fator-B) (Levi and Suhl 1979) que são escritos em um arquivo com as coordenadas da estrutura média.

O fator-B correspondente à mobilidade local do resíduo em nível atômico. Este fator representa a diminuição da intensidade na difração por causa da desordem dinâmica, resultante da vibração dos átomos dependentes da temperatura (Kolkman, Stemmer et al. 2003). Tanto maior será a desordem quanto maior for a temperatura da proteína. Seu cálculo é dado por:

$$B_i = 8\pi^2 U_i^2$$

onde U_i^2 é o desvio médio quadrático do átomo i .

3.4.1.3 Dicionário da Estrutura Secundária de Proteínas (DSSP)

O algoritmo DSSP é um método padrão para atribuição de estrutura secundária para os aminoácidos de uma proteína (Kabsch and Sander 1983). Ele faz a leitura da posição dos átomos da cadeia principal (*backbone*) e a partir das informações presentes no arquivo das coordenadas dos átomos em formato PDB, é realizado o cálculo de energia da ligação de hidrogênio entre todos os átomos. As ligações de hidrogênio dos átomos são então usados para determinar a classe mais provável de estrutura secundária para cada resíduo na proteína (Joosten, te Beek et al. 2011).

Com base nisso, oito tipos de estrutura secundária são atribuídos. As estruturas 3_{10} hélice, α -hélice e 5 -hélice são reconhecidas por apresentarem uma sequência repetitiva de ligações de hidrogênio a cada três, quatro ou cinco resíduos, respectivamente. Existem as folhas- β e as β -bridge; a primeira realiza conjuntos de ligações de hidrogênio entre aminoácidos e a segunda apenas uma ligação entre resíduos. Também é possível identificar as voltas (*turn*) que possuem ligações de hidrogênio típicas de hélices; Bend é usado para regiões de alta curvatura e *coil* é empregado quando nenhuma das estruturas anteriores são identificadas.

A evolução dos elementos de estrutura secundária das simulações foi calculada com a ferramenta *do_dssp* do pacote GROMACS (Pronk, Pall et al. 2013) que gera um gráfico que representa o tipo de estrutura secundária, em diferentes cores, para sequência da proteína ao longo do tempo de dinâmica.

A quantificação de estrutura secundária foi empregada para verificar matematicamente a percentagem de cada elemento estrutural envolvido nos modelos propostos para *cyspep*. Para tal foi utilizado o servidor 2Struc <<http://2struc.cryst.bbk.ac.uk/twostruc>> (Klose, Wallace et al. 2010), que realizou a quantificação por meio do algoritmo DSSP (Kabsch and Sander 1983).

3.4.1.4 Raio de Giro (R_g)

O cálculo do R_g de proteínas é indicativo do nível de compactação da estrutura. A medição ocorre entre a massa de cada átomo em relação ao centro de massas da molécula. O raio de giro foi calculado com a ferramenta *g_gyrate* do pacote GROMACS (Pronk, Pall et al. 2013) através da expressão:

$$R_g = \left(\frac{\sum_i |r_i|^2 m_i}{\sum m_i} \right)^{\frac{1}{2}}$$

onde m_i é a massa do átomo i , e r_i é a posição de um determinado átomo i em referência ao centro de massa da molécula.

3.4.1.5 Área de Superfície Acessível ao Solvente (SASA)

A superfície acessível ao solvente calcula as regiões hidrofóbicas, hidrofílicas e a superfície total acessível ao solvente de uma determinada molécula, ou seja, descreve a área sobre a qual um contato entre a proteína e o solvente pode ocorrer (Eisenhaber, Lijnzaad et al. 1995). A SASA é definida como o lugar geométrico dos centros de uma esfera de prova (representando a molécula do solvente) como se esta rolasse sobre a superfície de Van der Waals da proteína (Figura 11). A SASA foi calculada com a ferramenta *g_sas* do pacote GROMACS (Pronk, Pall et al. 2013).

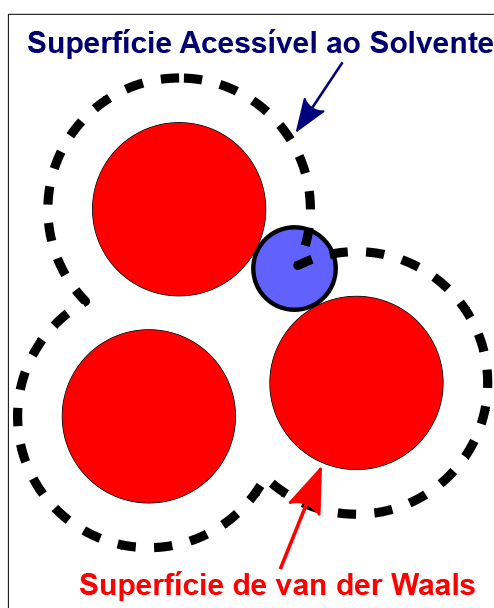


Figura 11: Representação da Superfície Acessível ao Solvente. A superfície de Van der Waals é dada pelo raio atômico, em vermelho; a superfície acessível está representada pela linha tracejada e é criada pelo traçado do centro da esfera da sonda atômica (azul) que desliza ao longo da superfície de Van der Waals. Adaptado de Keith Callenberg, 2010.

3.4.1.6 Clusterização

A clusterização permite agrupar estruturas similares de uma simulação. O agrupamento foi realizado com a ferramenta *g_cluster* do pacote GROMACS (Pronk, Pall et al. 2013) utilizando o algoritmo GROMOS (Daura, Gademann et al. 1999) para determinar a associação de um cluster com RMSD de 0,3nm e 0,35nm no *backbone* para cada grupo de cluster. Todos os grupos são excludentes, por isso uma estrutura isolada só pode ser membro de um único *cluster*.

4 RESULTADOS

4.1 Predição de Estrutura Secundária

O primeiro passo para prever a estrutura espacial da *cyspep* a partir de sua sequência de aminoácidos (Figura 7) foi através da predição de sua estrutura secundária. Nesta etapa do estudo, a sequência da *cyspep* foi submetida para quatro distintos servidores de predição de estrutura secundária consenso: Jpred, Sympred, NPS e Concord.

De maneira geral, estes servidores mostraram uma concordância em relação à presença predominante de estrutura folha- β (Figura 12A), além de apresentar regiões com predições similares entre os resíduos Val5–Val12, Gly38–Cys44 e Val49–Tyr54. Também detectamos que as regiões de maiores variações encontram-se no final da sequência. A predição do servidor Concord foi escolhida dentre os demais como marcação dos modelos tridimensionais criados, pois possui uma informação adicional, o nível de confiança (Figura 12B).

A predição de estrutura secundária consenso também foi realizada com as CTEs de outras espécies de *Leishmania* (*L. mexicana*, *L. pifanoi*, *L. aethiopica*, *L. infantum*, *L. tropica* e *L. major*). Na Figura 13 é possível observar que tanto a sequência quanto a predição de estrutura secundária é conservada entre as CTEs.

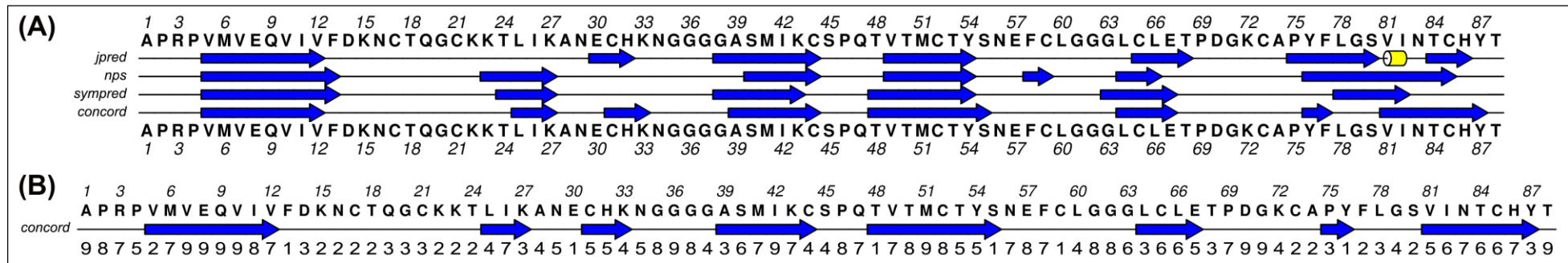


Figura 12: (A) Predição de Estrutura Secundária gerada por quatro diferentes servidores: Jpred, NPS, Sympred e Concord. (B) A predição do servidor Concord foi escolhida para marcação dos modelos tridimensionais gerados. A numeração de 0 a 9 indica o nível de confiança, quanto mais alto o valor maior será a confiança e vice-versa. Cilindro amarelo, 3-10 hélices e setas azuis, folhas- β . A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

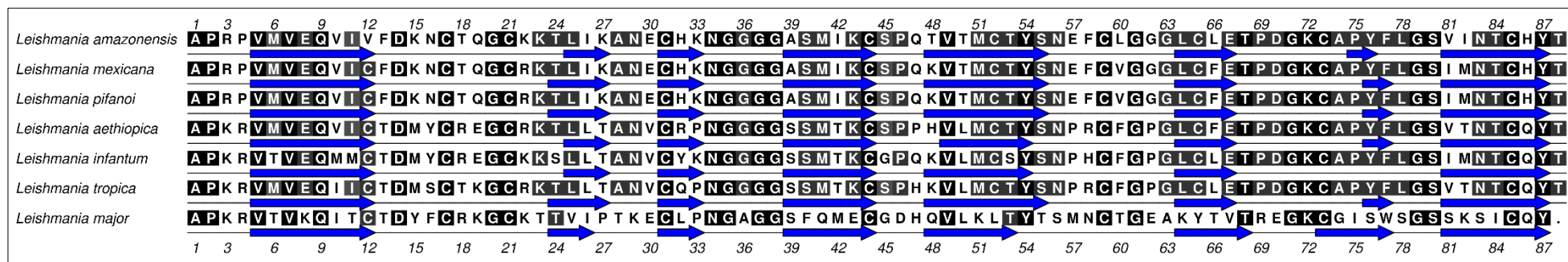


Figura 13: Alinhamento múltiplo entre as CTEs de sete espécies de *Leishmania*. Abaixo de cada sequência encontra-se a predição de estrutura secundária consenso gerada pelo servidor Concord. Setas azuis representam a predição em folha- β . A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

4.2 Predição das Ligações Dissulfeto

A *Cyspep* possui ao longo de sua sequência de aminoácidos 9 resíduos de cisteína. Esses resíduos podem formar ligações dissulfeto que são importantes na manutenção da estrutura terciária das proteínas por influenciarem na regulação de enovelamento, estabilidade termodinâmica e modulação da atividade biológica. Desta forma, foi oportuno considerar uma etapa de predição das possíveis ligações dissulfeto presentes na *cyspep*. Neste estudo utilizamos quatro servidores *online*: *DiANNA*, *Scratch*, *EDBCP* e *DISULFIDE*.

Nos resultados obtidos os servidores *Disulfind* e *EDBC* não indicaram a presença de ligações dissulfeto, enquanto os servidores *DiANNA* e *Scratch* forneceram uma predição para este tipo ligação entre as 36 combinações possíveis (dado que a *cyspep* possui nove cisteínas e as combinações são tomadas aos pares, o número total de combinações é dado pelo número combinatorial $\binom{9}{2} = \frac{9!}{2!7!} = 36$).

Por outro lado, os servidores *DiANNA* e *Scratch* realizaram a predição da formação de quatro ligações dissulfeto entre combinações distintas dos resíduos de cisteína: Cys17–Cys31, Cys21–Cys52, Cys44–Cys59, Cys73–Cys85 e Cys59–Cys73, Cys31–Cys52, Cys65–Cys85, Cys21–Cys44, respectivamente (Tabela I).

Não houve consenso na predição das ligações dissulfeto entre os quatro servidores utilizados. Isto posto, não levamos em conta essas predições para criação, análise e escolha dos modelos tridimensionais para *cyspep*.

Tabela I: Predição de ligações dissulfeto dos servidores *DiANNA* e *SCRATCH*

Posição da cisteína na sequência	Servidores	
	<i>DiANNA</i>	<i>SCRATCH</i>
17 – 31		
21 – 44		
21 – 52		
31 – 52		
44 – 59		
59 – 73		
65 – 85		
73 – 85		

Em cinza, predição de possível ligação dissulfeto do servidor *DiANNA*.

Em preto, predição de possível ligação dissulfeto da predição do servidor *SCRATCH*.

4.3 Modelagem da Estrutura Tridimensional da *Cyspep*

A modelagem comparativa foi a primeira abordagem empregada para a construção do modelo tridimensional da *cyspep*.

Através do servidor SAS foi possível encontrar o melhor molde para modelagem comparativa, aquele que apresentou maiores valores de identidade e cobertura. Entre as 20 estruturas propostas pelo servidor (Tabela II), o molde selecionado foi a estrutura da lectina de germe de trigo de 170 aminoácidos cujo código PDB é 1WGC (Wright 1990). A lectina de germe de trigo possui identidade sequencial de 30,8% e similaridade de 51,29% com a *cyspep* e sobreposição de 78 aminoácidos, o que constitui uma cobertura de 89%.

A criação do modelo foi realizada através do alinhamento da sequência da *cyspep* com a sequência da lectina de germe de trigo (Figura 14A). Foram obtidos 50 modelos, dos quais selecionamos aquele com menor função objetiva como explicado no item 3.3.1 da metodologia. É possível constatar que a estrutura do modelo escolhido (*Comp*) é semelhante à estrutura do molde, apresentando RMSD igual a 1,336Å (Figura 14B).

Já com a utilização do servidor *FUGUE*, a estrutura selecionada para modelagem comparativa foi uma proteína de função desconhecida codificada pelo gene pa3566 da bactéria gram-negativa *Pseudomonas aeruginosa* cujo código PDB é 1X7V (Sanders, Walker et al. 2005). Neste caso, constatamos uma identidade sequencial de 8% com a *cyspep* e sobreposição de 76 aminoácidos, significando uma cobertura de 86,36% (Figura 15A). O valor de RMSD da estrutura predita foi de 0,495Å, comparado ao seu molde (Figura 15B). Essa baixa identidade se justifica pela técnica de alinhamento para busca de moldes que utiliza o banco de dados HOMSTRAD que fornece alinhamentos que levam em consideração informações estruturais e ambientais de cada resíduo na proteína.

Tabela II: Estruturas encontradas pelo servidor SAS como potenciais moldes para *cyspep*

Identificação	Código PDB	Identidade (%)	Cobertura (%)	E-value
1	1WGC	30,8	89	1,3
2	2UVO	30,8	89	1,3
3	2UWQ	30,8	89	1,3
4	2WGC	30,8	89	1,3
5	2X3T	30,8	89	1,3
6	4AML	30,8	89	1,3
7	9WGA	30,8	89	1,3
8	1ESL	34,9	49	2
9	3WGA	29,9	88	3
10	1JXC	29,7	73	3,6
11	2KCN	37,2	49	4,3
12	2DJ9	52,6	22	8,8
13	1EAI	36,8	43	9,3
14	1LH9	23,8	95	9,4
15	2UWI	33,3	55	9,7
16	1HRE	28,8	59	10
17	1KV0	27,8	82	14
18	2I1W	28,2	89	14
19	2EGH	47,4	22	15
20	1DF2	37,5	36	15

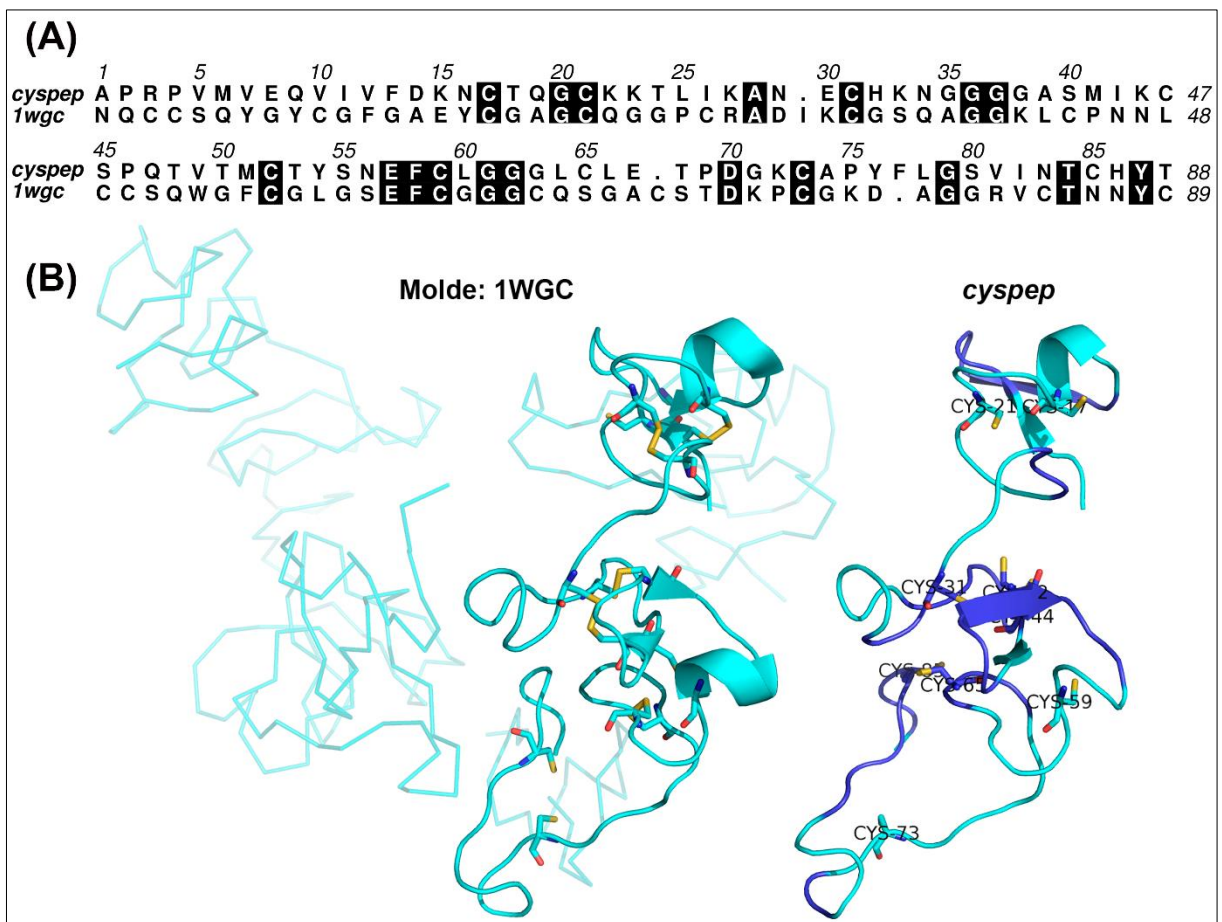


Figura 14: Construção do modelo tridimensional da *cyspep* pelo programa Modeller: (A) Alinhamento da sequência da *cyspep* e a proteína-molde (código 1WGC (Wright 1990) no PDB) obtido através do *T-Coffee*, utilizado para criação do modelo pelo programa *Modeller*. Em destaque estão evidenciados os resíduos idênticos. (B) Modelo por homologia da *cyspep*. Em verde está representada a proteína molde com suas duas cadeias de 170 aminoácidos correspondente à lectina de germe de trigo, código PDB 1WGC, destacado em *cartoon* a região utilizada como molde para criação do modelo. Ao lado, em azul, encontra-se a *cyspep* modelada.

Como os modelos criados pela técnica de modelagem comparativa não se ajustaram à predição de estrutura secundária (Figura 12B), partimos para outras abordagens de modelagem como *threading* e métodos *de novo*.

Os servidores *I-Tasser*, *MUSTER*, *LOMETS* e *SPARKS-X* utilizam a abordagem *threading*. Os moldes utilizados para criação dos modelos e as informações sobre identidades e coberturas são mostradas na Tabela III. O servidor *I-Tasser* gerou cinco modelos utilizando seis diferentes estruturas de proteínas. Os alinhamentos estão exibidos na Figura 16. *LOMETS* também fez uso de seis moldes e dez modelos foram criados a partir deles (Figura 17). Cada um dos servidores *MUSTER* e *SPARKS-X* gerou dez modelos, e cada um dos modelos foi criado por meio de diferentes moldes. Os alinhamentos segundo os servidores *MUSTER* e *SPARKS-X* estão presentes nas Figuras 18 e 19, respectivamente.

Tabela III: Moldes utilizados por cada servidor *threading* para criação de modelos *cyspep*

Servidor	Código PDB	Proteína	Referência	Identidade	Cobertura
I-TASSER	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>	(Antuch, Guntert et al. 1996)	16	92
	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>		13	93
	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>		13	95
	1C01	Peptídeo antimicrobiana de planta	(McManus, Nielsen et al. 1999)	18	82
	1SR4	Toxina citoletal distensora de <i>Haemophilus ducreyi</i>	(Nesic, Hsu et al. 2004)	15	95
	3KT9	Proteína com <i>forkhead</i> - domínio associado de humanos	(Becherel, Jakob et al. 2010)	9	92
	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>		13	95
	1L5B	Cianovirina-N da cianobactéria <i>Nostoc ellipsosporum</i>	(Barrientos, Louis et al. 2002)	10	57
	3SOB	Proteína-6 relacionada com o receptor de lipoproteína de baixa densidade de humanos	(Bourhis, Wang et al. 2011)	15	91
1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>		15	91	
MUSTER	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>	(Antuch, Guntert et al. 1996)	16	92
	1C01	Peptídeo antimicrobiana de planta	(McManus, Nielsen et al. 1999)	16	85
	4BSO	R-espondinas-1 de humano	(Peng, de Lau et al. 2013)	13	93
	3ME2	Fator de necrose tumoral ligado ao membro da superfamília 11 de rato	(Liu, Walter et al. 2010)	11	82
	3QD6	Ligante CD154 de humanos	(An, Kim et al. 2011)	11	81
	4BDW	Fator de coagulação X11A de cadeia pesada de humanos	(Beringer and Kroon-Batenburg 2013)	15	76
	2YN3	Proteína de membrana interna de <i>Salmonella</i>	(Griessl, Schmid et al. 2013)	13	86
	3K51	Membro 15 da superfamília de ligantes de fatores de necrose tumoral de humano	(Zhan, Patskovsky et al. 2011)	19	82
	4LD1	Proteína descaracterizados de <i>Danio rerio</i>	(Hatzopoulos, Erat et al. 2013)	15	90
3HO4	Proteína <i>Hedgehog</i> de humanos	(Bosanac, Maun et al. 2009)	24	70	
LOMETS	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>	(Antuch, Guntert et al. 1996)	13	95
	1SR4	Toxina citoletal distensora de <i>Haemophilus ducreyi</i>	(Nesic, Hsu et al. 2004)	17	65
	1VMJ	Proteína hipotética (TM0723) de <i>Thermotoga marítima</i>	-	19	69
	1N02	Cianovirina-N da cianobactéria <i>Nostoc ellipsosporum</i>	(Barrientos, Louis et al. 2003)	10	53

	1I8N	Proteína antiplaquetária de <i>Haementeria officinalis</i>	(Huizinga, Schouten et al. 2001)	17	66
	1C01	Peptídeo antimicrobiana de planta	(McManus, Nielsen et al. 1999)	16	85
SPARKS-X	1WKT	Toxina <i>killer</i> de levedura <i>Williopsis mrakii</i>	(Antuch, Guntert et al. 1996)	13	99
	1C01	Peptídeo antimicrobiana de planta	(McManus, Nielsen et al. 1999)	14	95
	1H8L	Carboxipeptidase GP180 de pato	(Aloy, Companys et al. 2001)	10	99
	2IJ0	Toxina-1 da síndrome do choque tóxico de <i>Staphylococcus aureus</i>	(Moza, Varma et al. 2007)	12	99
	1ETH	triacilglicerol acilhidrolase de javali	(Hermoso, Pignol et al. 1996)	15	96
	1F02	Domínio c- terminal do receptor translocado de intimina de <i>Escherichia coli</i>	(Luo, Frey et al. 2000)	10	99
	3IYL	Capsídeo externo VP4 de <i>Grass carp reovirus</i>	(Zhang, Jin et al. 2010)	14	94
	3IX0	Beta microseminoproteína de humanos	(Kumar, Jagtap et al. 2010)	12	99
	2PYS	Cianovirina-N da cianobactéria <i>Nostoc ellipsosporum</i>	(Fromme, Katiliene et al. 2007)	7	93
	3G3L	Proteína descaracterizada associada à membrana de <i>Bacteroides fragilis</i>	-	7	99

<i>cyssep</i> <i>1wkt</i>	1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85	A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q . . . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
<i>cyssep</i> <i>1vmj</i>		. G D G Y L I . . . M C K N C D P N T G S C D W K Q N W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E . . .	88
<i>cyssep</i> <i>1wkt</i>		A P R P V M V E Q V I V F D K N C T Q G . C . . K K T L I K A N E C H K N G G G G A S M I K C S P Q . . . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T .	88
<i>cyssep</i> <i>1n02</i>		. G D G Y L I . . . M C K N C D P N T G S C D W K Q N . . . W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E	88
<i>cyssep</i> <i>1i8n</i>		A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C . . . L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
<i>cyssep</i> <i>1wkt</i>	 A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M . . I K C S P Q T V T M C T Y S N E F C L G G G L	64
<i>cyssep</i> <i>1wkt</i>		E T I T A G N E D C W S K R P G W K L P D N L L T K T E F T S V D E C R K M G E E S A V E P S C Y I L Q I N T E T N E C Y R N N E G D V T W S S L Q Y D Q P N V V Q W H L H . . A C S	89
<i>cyssep</i> <i>1wkt</i>		A P R P V M V E Q V I V F D K N C T Q G . C . . K K T L I K A N E C H K N G G G G A S M I K C S P Q . . . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T .	88
<i>cyssep</i> <i>1wkt</i>		. G D G Y L I . . . M C K N C D P N T G S C D W K Q N . . . W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E	88
<i>cyssep</i> <i>1sr4</i>		A P R P V M V E Q V I V F D K N C T Q G . C . . K K T L I K A N E C H K N G G G G A S M I K C S P Q . . . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T .	88
<i>cyssep</i> <i>1wkt</i>		. G D G Y L I . . . M C K N C D P N T G S C D W K Q N . . . W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E	88
<i>cyssep</i> <i>1wkt</i>		A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P . . . Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
<i>cyssep</i> <i>1c01</i>		G D G Y L I M . . . C K N C D P N T G S C D W K Q N W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E . . .	88
<i>cyssep</i> <i>1wkt</i>		A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K . C A P Y F L G S V I N T C H Y T	88
<i>cyssep</i> <i>1wkt</i>	 S A F T V W S G P G C N N . . R A E R Y S K C G C S A I H Q K G G Y D F S Y T G Q . . T A A L Y N Q A G C S G V A H T R F G S S A R A C N P F G W K S I F I Q G	76
<i>cyssep</i> <i>1wkt</i>		88
<i>cyssep</i> <i>1vmj</i>	 K C A P Y F L G S V I N T C H Y T K L A P E K P Y S Q Y K H N D T G E D N A D A H L K R T I M G R E V V I A I T D R K M D L G P W E Q V F Y G E F D G M R P K R V L V K I I G E	88 139
<i>cyssep</i> <i>1wkt</i>		88
<i>cyssep</i> <i>1n02</i>	 N F S Q T C Y N S A I Q G S V L T S T C E R T N G G Y N T S S I D L N S V I E N V D G S L K Y E	88 102
<i>cyssep</i> <i>1i8n</i>	 C L E T P D G K C A P Y F L G S V I N T C H Y T	88 89
<i>cyssep</i> <i>1wkt</i>		88
<i>cyssep</i> <i>1wkt</i>		88
<i>cyssep</i> <i>1sr4</i>	 A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
<i>cyssep</i> <i>1wkt</i>		E Y F R F V N Q S L G T C V E A Y G N G L I H D I C S L D K L A Q E F E L L P T D S G A V V I K S V S Q G R C V T Y N P V S T T F Y S T V T L S V C D G A T E P S R D Q T W Y L A P P	159
<i>cyssep</i> <i>1wkt</i>		88
<i>cyssep</i> <i>1c01</i>		88 76

Figura 17: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor LOMETS.

<i>cyssep</i> <i>1wktA</i>	1 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85	A P R P V M V E Q V I V F D K N C T Q G C . . K K T L I K A N E C H K N G G G G A S M I K C S P Q . . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
		. G D G Y L I . . . M C K N C D P N T G S C D W K Q N . . W N T C V G I G A N V H W M V T G G S T D G K Q G C A T I W E G S G C V G R S T T M C C P A N T C C N I N T G F Y I R S Y R R V E	88
<i>cyssep</i> <i>1c01A</i>		A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
	 S A F T V W S G P G C N N . . R A E R Y S K C G C S A I H Q K G G Y D F S Y T G Q . . T A A L Y N Q A G C S G V A H T R F G S S A R A C N P F G W K S I F I Q C	76
<i>cyssep</i> <i>4bsoA</i>		A P R P V M V . . . E Q V I V F D K . N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F . . . C L G G G L C L E T P D G K C A P Y F L G S V I N . . T C H	86
		. . Q A . . . C A K G C E L C S E V N G C L K . C S P K L F I L L E R N D I R Q V G V C L P S C P P G Y F D A R N P D M N K C I K C K I E H C E A C F S H N F C T K C K E G L Y L H K G R C Y	89
<i>cyssep</i> <i>3me2R1</i>		. . A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
		P C T Q E R H Y E H L G R C C S R . C E P G K Y L S S K C . . T P T S D S V C L P C G P D E Y L D T W N E E D K G L L H K V C D A G K A L V A V D P G N H T A P R R C	80
<i>cyssep</i> <i>3qd6R1</i>		. A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
		C R E K Q Y L I N S Q C C S L C Q P G Q K L V S D C . . T E F T E T E C L P C G E S E F L D T W N R E T H C H Q H K Y C D P N L G L R V Q Q K G T S E T D T I C T	79
<i>cyssep</i> <i>4bdwA</i>		A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E . T P D G K C A P Y F L G S V I N T C H Y T .	88
	 E K C F E P Q L L R F F H K N E I W Y R T E Q A A V A R C Q C K G P D A H C Q R L A S Q A C R T N P C L H G G R C L E V E G H . . . R L C H C P V G Y T G	74
<i>cyssep</i> <i>2yn3A1</i>		. A P R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q . T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H Y T	88
		T P P N A P V M T Y S D I V N D L I I G G T A E A . . K S Q L I I T D S E G N T Y T L T V P D N G K W S A I P V P S E . . . G K F T I T S V D A I G N R S D D V P L D I	79
<i>cyssep</i> <i>3k51B1</i>		. A P R P V M V E . . Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I N T C H	86
		E T P T Y P W R D A E E R L V C A Q C P P G T F V Q R P C . . R R D S P T T C G P C P P R H Y T Q F W N Y L E R C R Y C N V L C G E R E E E A R A . . . C H A T H N R A C R	81
<i>cyssep</i> <i>4d1A1</i>		A P . . R P V M V E Q V I V F D K N C T Q G C K K T L I K A N E C H K N G G G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y F L G S V I	82
		L P D G G R L V V F P N G T R K E L S A D G Q T V K V M F F N G D V K H T M P D Q R V I Y Y Y A E A Q T T H I T Y P D G M E V L Q F P N N . Q T E K H F P D G R K E I T F P D Q T V K	90
<i>cyssep</i> <i>3ho4B2</i>		A P R P V M V E Q V I V F D K N C T Q G C K K . T L I K A N E C H K N G G . G G A S M I K C S P Q T V T M C T Y S N E F C L G G G L C L E T P D G K C A P Y . F L G S V I N T C H Y T	88
	 T S E C S R L C R N G Y C T P T G K C C S P G W E G D . . . F C R T A K . C E P A C R H G G V C V R . P . N K G L . . C K K G Y L G P Q G E H H H	66
<i>cyssep</i> <i>1wktA</i>		88
		88
<i>cyssep</i> <i>1c01A</i>		88
		76
<i>cyssep</i> <i>4bsoA</i>		Y T	88
		P A C P M E C S	97
<i>cyssep</i> <i>3me2R1</i>		88
		A	81
<i>cyssep</i> <i>3qd6R1</i>		88
		79
<i>cyssep</i> <i>4bdwA</i>		88
		P F C D V D T A A	83
<i>cyssep</i> <i>2yn3A1</i>		88
		79
<i>cyssep</i> <i>3k51B1</i>		Y T	88
		81
<i>cyssep</i> <i>4d1A1</i>		N T C H Y T	88
		T L	92
<i>cyssep</i> <i>3ho4B2</i>		88
		66

Figura 18: Alinhamentos entre a sequência alvo e os moldes encontrados pelo servidor *MUSTER*.

Os quatro servidores utilizaram a estrutura depositada sob código PDB 1WKT como molde. O molde de código PDB 1C01 foi utilizado pelos servidores *I-Tasser*, *MUSTER* e *SPARKS-X*, enquanto a estrutura de código 1SR4 foi utilizada pelos servidores *I-Tasser* e *LOMETS*. Apesar das proteínas-moldes 1L5B, 1N02 e 2PYS apresentarem código PDB distintos, todas são pertencentes à classe de cianovirina-N da cianobactéria *Nostoc ellipsosporu* (Tabela IV).

Tabela IV: Moldes utilizados para criação dos modelos nos diferentes servidores

Identificação	I-TASSER	MUSTER	LOMETS	SPARKS-X
1	1wktA	1wktA	1wktA	1wktA
2	1wktA	1c01A	1vmjA	1c01A
3	1wktA	4bsoA	1wktA	1h8IA3
4	1c01A	3me2R1	1n02A	2ij0B1
5	1sr4A	3qd6R1	1i8n_A	1ethB
6	3kt9A	4bdwA	1wktA	1f02I2
7	1wktA	2yn3A1	1wktA	3iyIW5
8	115bA	3k51B1	1sr4A	3ix0B
9	3sobA	4ld1A1	1wkt_	2pysB
10	1wktA	3ho4B2	1h4sA3	3g3IA1

Cores iguais representam mesmos moldes e/ou mesma classe.

Nos modelos selecionados para os ensaios de simulações de dinâmica molecular levamos em conta a presença de ligação dissulfeto ou a menor distância entre os resíduos de cisteínas, além da predição de estrutura secundária. Dos cinco modelos gerados pelo servidor *I-Tasser*, apenas um foi selecionado (*IT1*) para simulações de dinâmica molecular. A estrutura escolhida tem a maior otimização do número de possíveis ligações dissulfeto, considerando a distância entre as cisteínas de até 5,5Å, já que nenhum modelo estava de acordo com a predição dos servidores de ligação dissulfeto (Anexo B, Tabela II). Estabelecemos o raio de corte de até 5,5Å para ter uma maior abrangência, dado que a distância de ligação dissulfeto é de 2,04Å.

MUSTER apresentou nos modelos *M4* e *M8* duas ligações dissulfeto que coincidem com a predição do servidor *DiANNA* (Tabela V), sendo estes os escolhidos para as simulações de dinâmica molecular.

Tabela V: Distância entre os pares de cisteínas dos modelos selecionados para simulações de dinâmica molecular.

Posição	D	S	Comp	Fugue	IT1	M4	M8	L367	SP5	Q5	Q7	D9	R1
17 - 21			7,13	9,2	3,19	9,04	11,82	5,87	7,58	5,47	3,89	8,97	8,66
17 - 31			18,08	24,7	15,41	2,02	2,03	22,81	2,03	17,02	19,33	8,33	27,41
17 - 44			15,19	20,79	14,27	18,12	18,34	5,1	14,92	28,32	15,42	17,13	17,09
17 - 52			13,69	15,05	20,87	15,06	20,31	20,89	9,29	18,9	13,86	8,9	21,56
17 - 59			21,72	33,65	23,26	17,34	17,31	24,19	17,3	16,93	28,34	17,07	31,19
17 - 65			24,46	30,29	23,46	26,98	28,0	20,96	16,3	18,34	18,51	16,05	30,39
17 - 73			37,41	20,39	6,19	36,25	26,31	5,97	21,91	22,51	18,99	15,01	21,37
17 - 85			24,29	9,97	15,21	28,15	24,63	15,12	18,88	23,7	23,07	10,65	15,57
21 - 31			14,09	17,66	13,91	10,33	10,64	18,13	8,42	15,38	18,11	5,56	18,85
21 - 44			12,86	12,19	15,62	9,55	10,61	9,12	8,06	26,43	14,2	11,0	10,06
21 - 52			10,31	8,97	20,37	10,52	13,21	17,93	11,57	17,62	11,33	2,1	14,07
21 - 59			20,58	24,88	22,12	9,19	10,9	19,75	10,63	13,92	25,46	13,38	24,04
21 - 65			19,78	24,13	23,47	21,72	24,82	19,83	9,53	15,7	15,04	12,55	23,46
21 - 73			33,02	14,37	8,89	27,86	17,98	10,38	17,59	19,99	15,3	14,34	13,43
21 - 85			19,94	11,68	13,99	20,22	20,27	11,38	11,98	21,25	19,22	12,52	8,85
31 - 44			8,8	18,63	14,28	19,19	17,52	23,52	16,19	12,31	13,8	11,2	15,92
31 - 52			4,91	22,57	8,72	16,46	20,32	8,08	10,96	9,05	8,83	4,47	11,0
31 - 59			11,16	15,92	9,46	18,23	16,55	2,03	18,19	14,26	16,94	12,1	11,9
31 - 65			7,65	15,42	13,51	27,6	27,83	14,72	17,56	15,17	14,47	8,45	14,38
31 - 73			21,38	9,35	17,12	37,42	25,63	23,96	23,7	16,8	16,24	8,97	9,51
31 - 85			10,9	26,95	3,61	29,53	24,67	8,29	19,83	20,78	22,88	7,08	16,76
44 - 52			5,48	9,95	15,16	10,59	11,65	20,09	15,16	11,36	12,82	9,43	11,2
44 - 59			8,72	18,67	19,96	2,04	2,03	24,71	6,3	18,72	25,88	4,87	25,3
44 - 65			11,56	25,23	16,9	15,46	15,0	18,98	2,02	18,42	17,42	9,68	20,66
44 - 73			22,69	18,66	13,06	18,59	8,19	6,72	11,97	16,84	17,15	15,44	10,11
44 - 85			9,96	20,43	14,82	12,21	11,44	15,88	6,32	21,28	22,62	15,36	2,02
52 - 59			11,05	23,72	5,43	10,66	12,61	7,63	20,05	9,57	15,27	11,4	17,3
52 - 65			10,94	25,23	5,11	15,0	21,14	6,69	16,78	9,21	7,35	10,94	9,86
52 - 73			24,22	18,25	20,0	23,77	13,74	20,57	16,37	9,52	9,04	13,14	2,02
52 - 85			11,93	11,22	7,4	14,83	13,34	7,44	20,94	13,5	15,46	11,25	10,84
59 - 65			11,98	14,82	8,55	14,87	14,14	14,19	4,98	2,06	11,6	9,5	16,4
59 - 73			19,81	17,31	23,18	19,66	9,39	24,94	17,13	6,21	13,37	14,86	16,9
59 - 85			11,55	31,51	8,13	13,59	11,18	9,3	2,02	7,93	16,23	14,17	25,43
65 - 73			13,98	10,72	21,28	17,84	12,95	19,13	12,33	4,43	2,78	6,34	11,11
65 - 85			5,72	27,58	11,48	14,18	8,39	11,46	4,69	6,18	8,67	8,59	20,44
73 - 85			13,43	20,1	15,91	9,58	9,72	15,68	16,34	4,51	6,69	5,15	10,03

A unidade de distância entre os pares de cisteínas é dada em ångström (Å).

Em cinza, predição de possível ligação dissulfeto do servidor *DiANNA*.

Em preto, predição de possível ligação dissulfeto da predição do servidor *SCRATCH*.

Em amarelo, distância que pode formar ligação dissulfeto.

Em azul, distância de até 5,5Å.

LOMETS produziu o modelo *L5* com ligação dissulfeto coincidente com a predição do servidor *DiANNA* (Anexo B, tabela III), porém a estrutura secundária destoa significativamente de sua predição (Figura 20), apresentando α -hélice e ausência de estrutura secundária definida a partir do resíduo Lys33.

Os modelos *L3*, *L6* e *L7* são idênticos e possuem valores de RMSD de 0,75Å e 1,08Å com *L1* e *L9*, respectivamente. Essa semelhança estrutural é consequência da utilização do mesmo molde (1WKT) para a sua criação. Desta forma, os modelos *L367* foram selecionados para dinâmica molecular com a presença de ligação dissulfeto nas Cys31–Cys59 e uma distância de 5,1Å entre as cisteínas Cys17–Cys44 (Tabela V). É importante frisar que o modelo *L3* apresentou o mesmo molde da estrutura *IT1* (1WKT), selecionado do servidor *I-Tasser*, em que os modelos apresentaram um valor de RMSD de 2,44Å quando comparados entre si.

SPARKS-X construiu um modelo (*SP5*) com três ligações dissulfeto, sendo as Cys17–Cys31 correspondentes à predição do servidor *DiANNA* (Tabela V). De todos os modelos obtidos pelas diferentes técnicas, este foi o que apresentou maior número de ligações dissulfeto.

Estruturas criadas utilizando a abordagem de métodos *de novo* foram obtidas pelos servidores *online* *QUARK* e *ROBETTA*. O primeiro apresenta a opção de atribuir restrições de distâncias para a modelagem, de tal modo, a predição tridimensional da *cyspep* foi enviada sem restrições adicionais de distância e também com restrição de distância das cisteínas que potencialmente estariam envolvidas em ligação dissulfeto como predito pelo servidor *DiANNA*. O mesmo procedimento foi realizado separadamente utilizando as restrições obtidas do servidor *SCRATCH*.

Entre os modelos resultantes da modelagem sem restrições adicionais, dois foram selecionados para dinâmica molecular, *Q5* e *Q7* (Figura 21), de acordo com as distâncias entre suas cisteínas (Anexo B, Tabela VI). As estruturas selecionadas possuíram maior número de pares com menor distância entre resíduos de cisteína não excludentes.

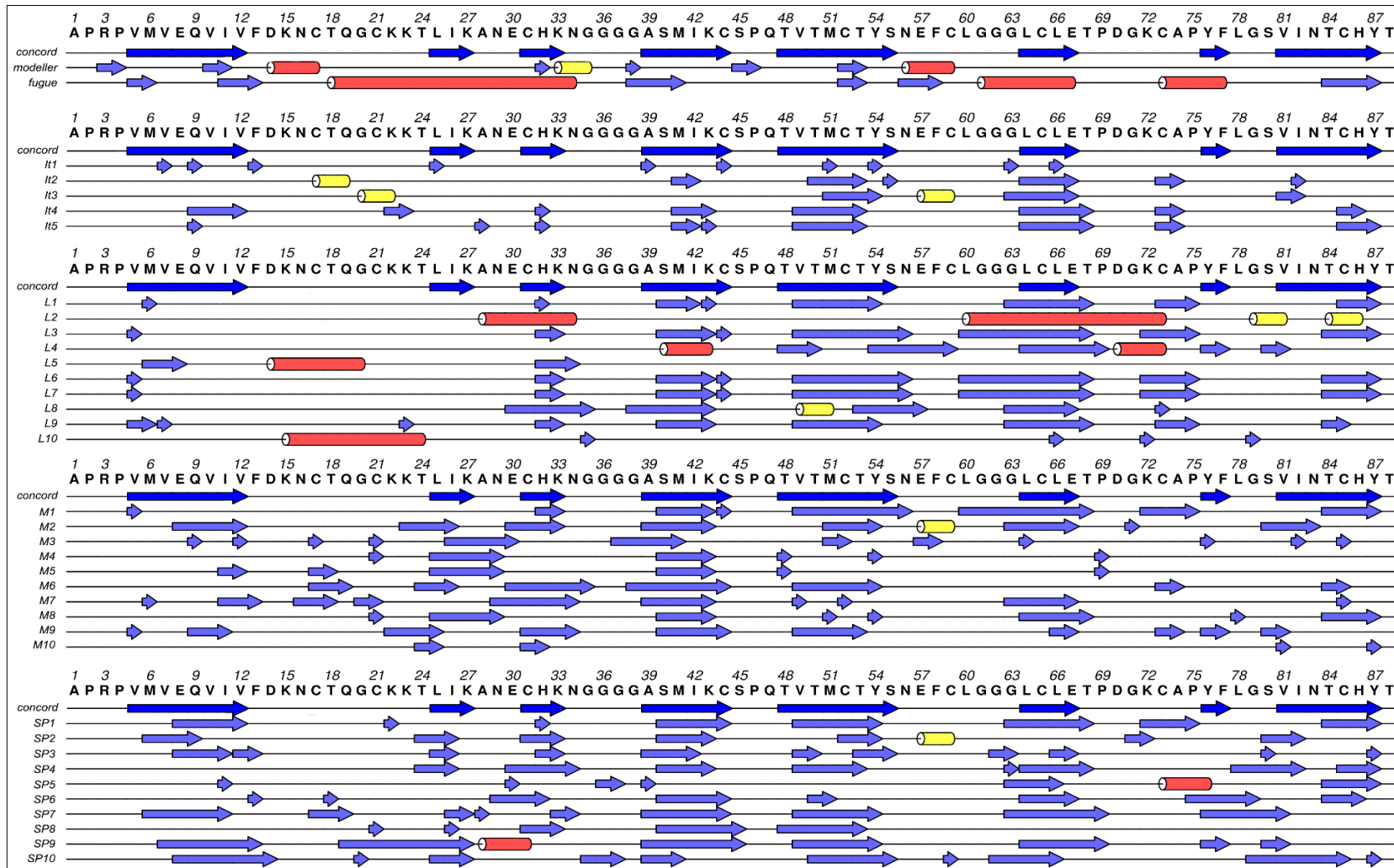


Figura 20: Estrutura secundária dos modelos criados por modelagem comparativa e *threading*. A sequência da *cyspep* é exibida na parte superior da imagem, seguida pela predição de estrutura secundária consenso – concord – em azul escuro. (cilindros vermelhos, α -hélices; cilindros amarelo, 3-10 hélices; setas azuis, folhas- β). O conteúdo da estrutura secundária de cada conformação preditos pelos diferentes métodos/servidores. A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

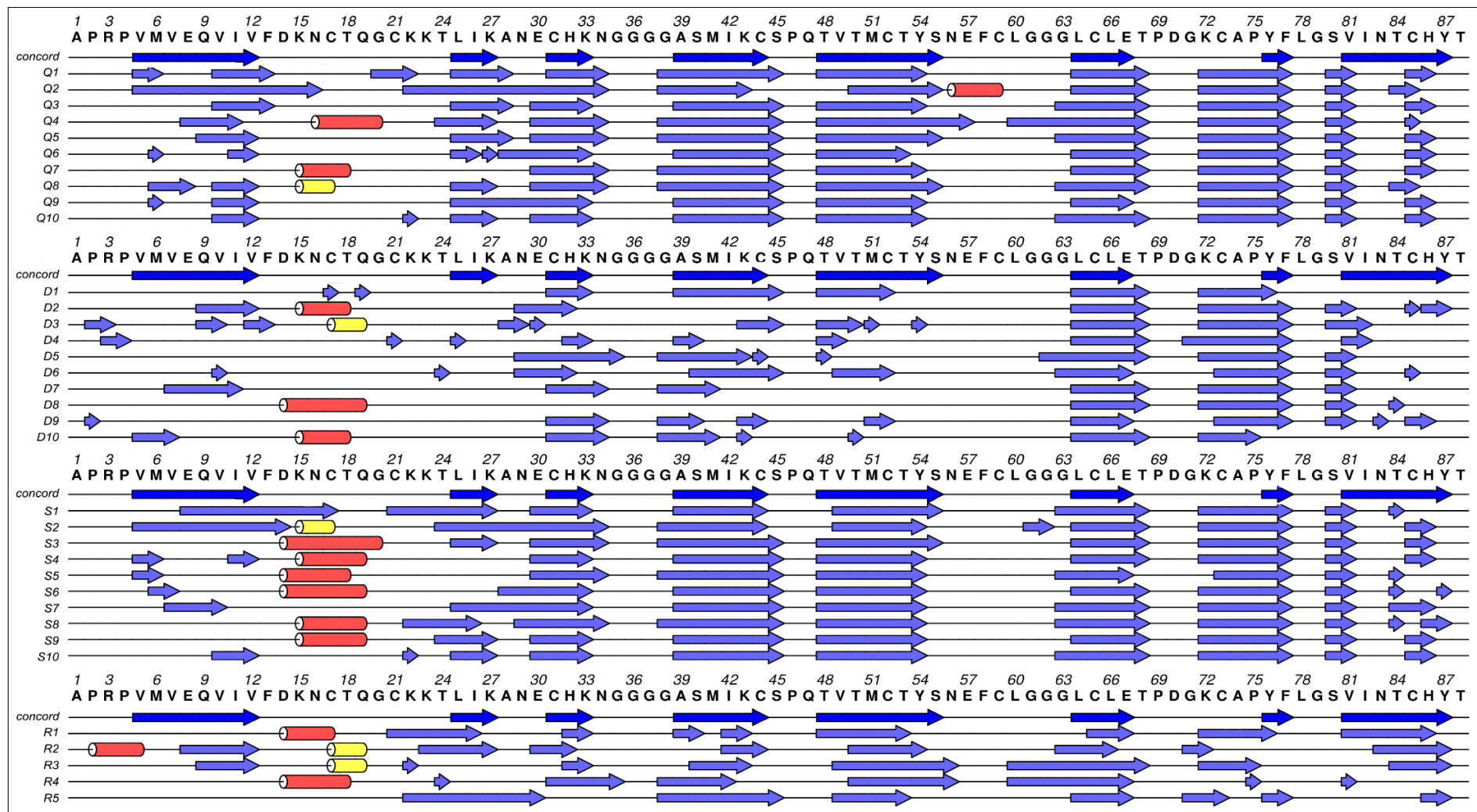


Figura 21: Estrutura secundária dos modelos criados por métodos *de novo*. A sequência da *cyspep* é exibida na parte superior da imagem, seguida pela predição de estrutura secundária consenso – concord – em azul escuro. (cilindros vermelhos, α-hélices; cilindros amarelo, 3-10 hélices; setas azuis, folhas-β). O conteúdo da estrutura secundária de cada conformação preditos pelos diferentes métodos/servidores. A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

Algumas estruturas obtidas com restrição de distância do *DiANNA* tiveram pares de cisteínas concordantes com da predição, porém nenhum modelo apresentou mais que uma ligação dissulfeto (Anexo B, Tabela VII). Destes, o modelo *D9* foi escolhido para realização das simulações por apresentar menor distância entre os demais pares envolvidos na predição, Cys31–Cys52, Cys44–Cys59 e Cys73–Cys85 com suas respectivas distâncias 4,47Å, 4,87Å e 5,15Å. A modelagem com restrições produzida pelo servidor *SCRATCH* não gerou estruturas que corroboraram a predição; assim, nenhum modelo foi escolhido para dinâmica molecular (Anexo B, Tabela VIII).

Como não foi possível construir modelos confiáveis da *cyspep* por modelagem comparativa, *ROBETTA* gerou cinco modelos por métodos *de novo*. O modelo *R1* foi selecionado para dinâmica por conter duas ligações dissulfeto (Tabela V), porém nenhuma destas ajusta-se à predição dos servidores.

Em suma, entre os 72 modelos gerados para *cyspep*, 11 foram selecionados para simulações de dinâmica molecular (Figura 22), sendo dois modelos obtidos por modelagem comparativa (*Comp*, *Fugue*), 5 por *threading* (*IT1*, *M4*, *M8*, *L3* e *SP5*) e 4 por métodos *de novo* (*Q5*, *Q7*, *D9* e *R1*).

Uma análise geral de todos os modelos que apresentam ligações dissulfeto está presente na tabela VI. Nela constam os pares de cisteínas envolvidos na ligação e quantidade de modelos dos servidores que as evidenciam. Dos 36 pares possíveis, apenas metade contêm distância suficiente para ligação dissulfeto. O par Cys17–Cys31 está presente em modelos de quatro servidores, um com abordagem por métodos *de novo* (*QUARK* com restrição *DiANNA*) e três *threading* (*LOMETS*, *MUSTER* e *SPARKS-X*). Os demais pares entre cisteínas possuem somente estruturas provindas de um único servidor, ou seja, apenas um servidor as contém.

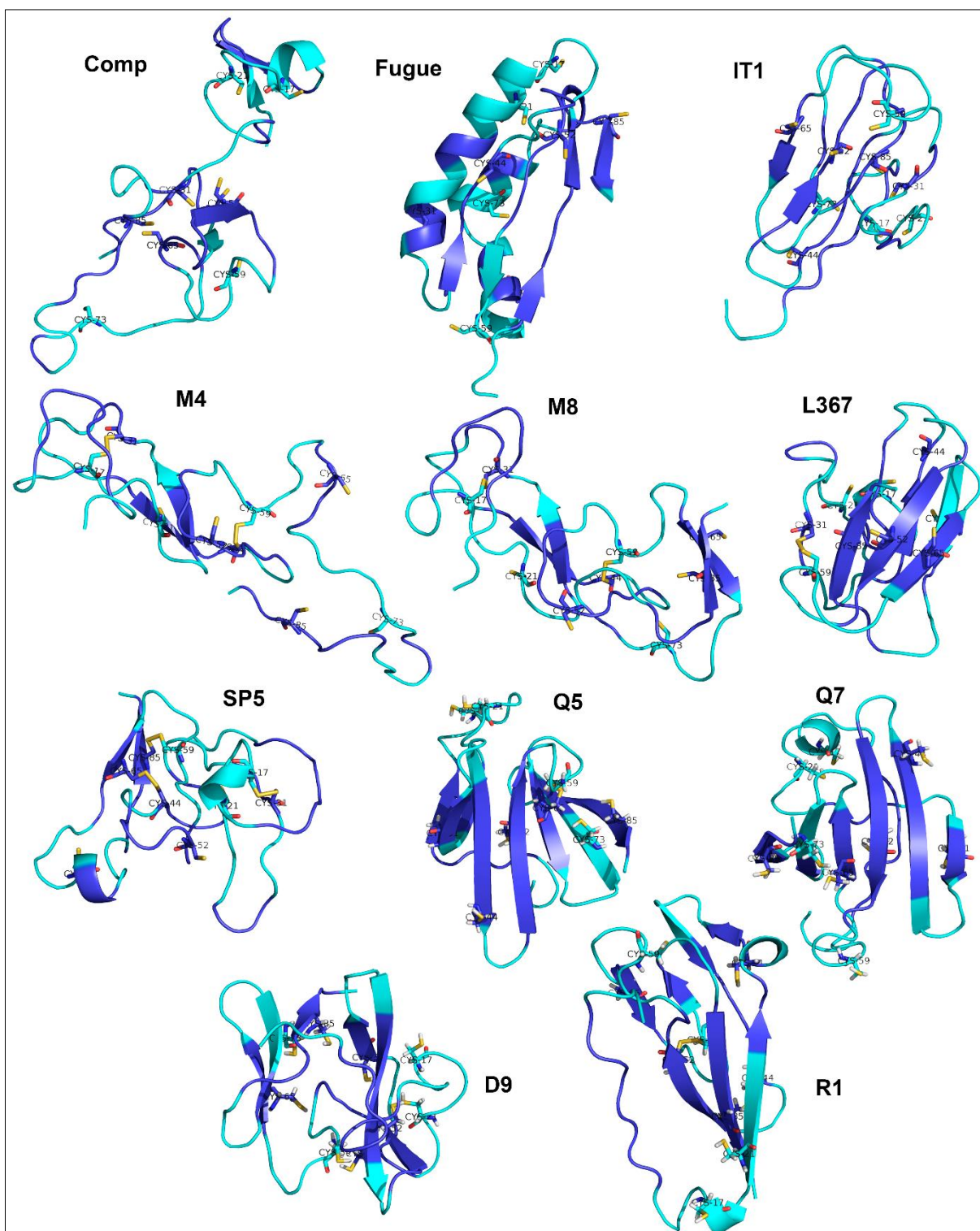


Figura 22: Modelos selecionados para simulações de dinâmica molecular. Em *sticks* estão representadas as cisteínas envolvidas em ligação dissulfeto. Q5, Cys59 – Cys65; D9, Cys21 – Cys52; L3, Cys31 – Cys59; M4, Cys17 – Cys31, Cys44 – Cys59; M8, Cys17 – Cys31, Cys44 – Cys59; R1, Cys44 – Cys85, Cys52 – Cys73; SP5, Cys17 – Cys31, Cys44 – CYS65, Cys59 – Cys85.

Tabela VI: Número de pares de cisteínas que formam ligação dissulfeto nos servidores de predição tridimensional de proteína.

Pares de Cisteínas	QUARK	QUARK (DiANNA)	QUARK (Scratch)	ROBETTA	LOMETS	MUSTER	SPARKS-X
17 - 31	1	3	1	-	1	-	-
17 - 73	-	1	-	-	-	-	-
21 - 44	-	-	1	-	-	-	-
21 - 52	-	-	-	-	2	-	-
31 - 59	5	2	2	-	-	-	-
31 - 65	-	-	-	1	-	-	-
31 - 85	-	-	-	-	-	1	-
44 - 59	-	3	-	-	-	-	-
44 - 65	-	-	1	-	-	-	-
44 - 73	-	-	-	-	1	-	-
44 - 85	-	-	-	-	-	-	1
52 - 65	-	-	-	-	1	1	-
52 - 73	-	-	-	-	1	-	1
59 - 65	-	-	-	1	-	-	-
59 - 73	-	2	-	-	-	-	-
59 - 85	-	-	1	-	-	-	-
65 - 73	-	-	-	6	3	5	-
73 - 85	-	-	-	-	1	-	-

Os números indicam a quantidade de estruturas dos respectivos servidores.

Em cinza, predição de possível ligação dissulfeto do servidor *DiANNA*.

Em preto, predição de possível ligação dissulfeto da predição do servidor *SCRATCH*.

4.4 Dinâmica Molecular

4.4.1 Simulações de Dinâmica Molecular

Dos 11 modelos selecionados para dinâmica molecular, sete (*Q5*, *D9*, *L3*, *M4*, *M8*, *R1* e *SP5*) apresentaram distância entre cisteínas para formar ligação dissulfeto. Todos os modelos foram simulados por um tempo de 200ns com campo de força *GROMOS96 53a6*.

Na Figura 23 é possível observar o valor de RMSD das estruturas simuladas representado em função do tempo. O modelo *Q5* apresentou variação de RMSD nos primeiros 30ns e oscilações ao redor de 0,7nm seguiram até o final da simulação. Os modelos *D9* e *IT1* apresentaram valores de RMSD mais baixos, variando em torno de 0,6nm, enquanto *L3* e *R1* exibiram valores de RMSD em torno de 0,7nm.

O modelo *Fugue* apresentou aumento de RMSD até 110ns, passando por uma estabilização em torno de 0,8nm até 196ns antes de diminuir ao final da simulação em 0,6nm. Os modelos *Q7*, *M4* e *M8* apresentaram maiores valores de RMSD em relação aos demais, sendo que os dois últimos alcançaram o valor máximo de 1,2nm.

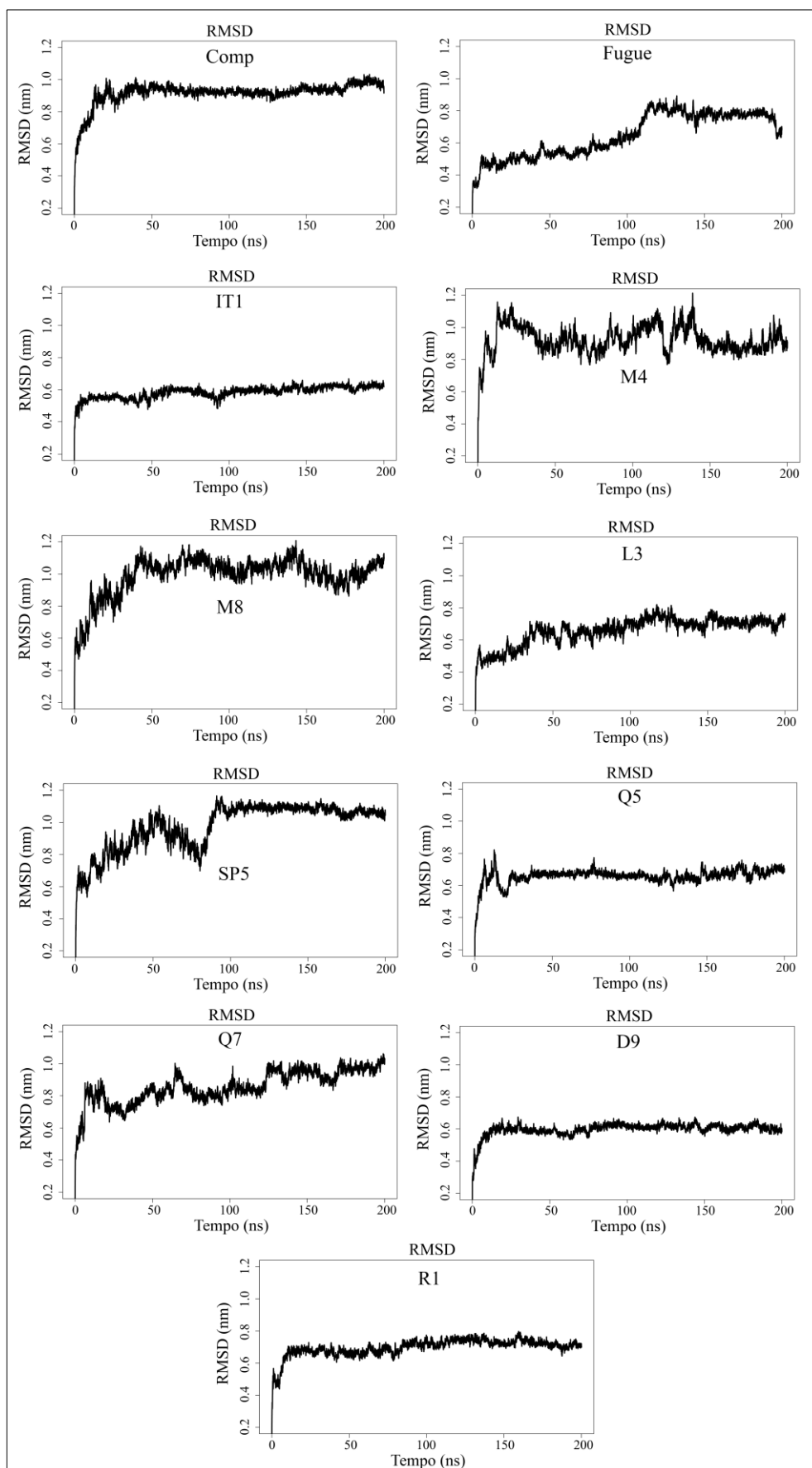


Figura 23: Desvios RMS dos átomos da cadeia principal (backbone) em relação à estrutura inicial representado em função do tempo de simulação.

O modelo *SP5* mostrou variação entre 0,6 e 1,1nm até 100ns seguido de um período de estabilização entorno de 1,1nm. O modelo *Comp* teve aumento de RMSD até 1,0nm nos primeiros 25ns e sua estabilização oscilou em volta de 0.9nm. Coincidentemente as estruturas que apresentaram ligação dissulfeto nos resíduos Cys17–Cys31 (*M4*, *M8* e *SP5*) apresentaram mais altos valores e variações de RMSD entre 0,6nm até 1,2nm, em comparação aos outros modelos.

A Figura 24 mostra os valores de RMSF para os modelos simulados durante a dinâmica molecular. Observamos que grandes flutuações estão associadas com os resíduos que compõem os segmentos N e C terminais, livres em orientação, o que resulta numa maior flexibilidade. A Figura 25 auxilia na melhor visualização das regiões flutuantes através de sua coloração em fator-B, uma escala de cores que indica a mobilidade dos átomos individuais ou resíduos.

Em uma comparação geral entre todas as simulações, as que apresentaram maiores valores de flutuação foram *FUGUE*, *M4*, *M8* e *SP5*. Os modelos *M4* e *M8* variaram em seus 19 primeiros resíduos, que quando comparado com sua estrutura inicial, apresentaram-se como regiões de mudança de estrutura secundária. O modelo *M4* era originalmente constituído de região *coil* e modificou-se para folha- β na estrutura média, enquanto que para *M8* o contrário aconteceu, de folha- β para *coil*. *SP5* apresentou alta flutuação entre os resíduos Met51–Tyr54, que é uma região sem estrutura secundária definida.

Sobre as cisteínas, essas encontram-se altamente estáveis exibindo baixa flutuação, com exceção de Q5 com flutuação na Cys21, Q7 na Cys21 e Cys44, e *SP5* com notável flexibilidade na Cys52. Nenhuma dessas cisteínas estão envolvidas em ligação dissulfeto. Ao avaliar a distância entre os 36 pares possíveis de cisteínas ao longo da simulação, notamos que nenhum par teve uma aproximação menor que 0,3nm, com exceção dos que foram estabelecidos como vínculo ao início da simulação.

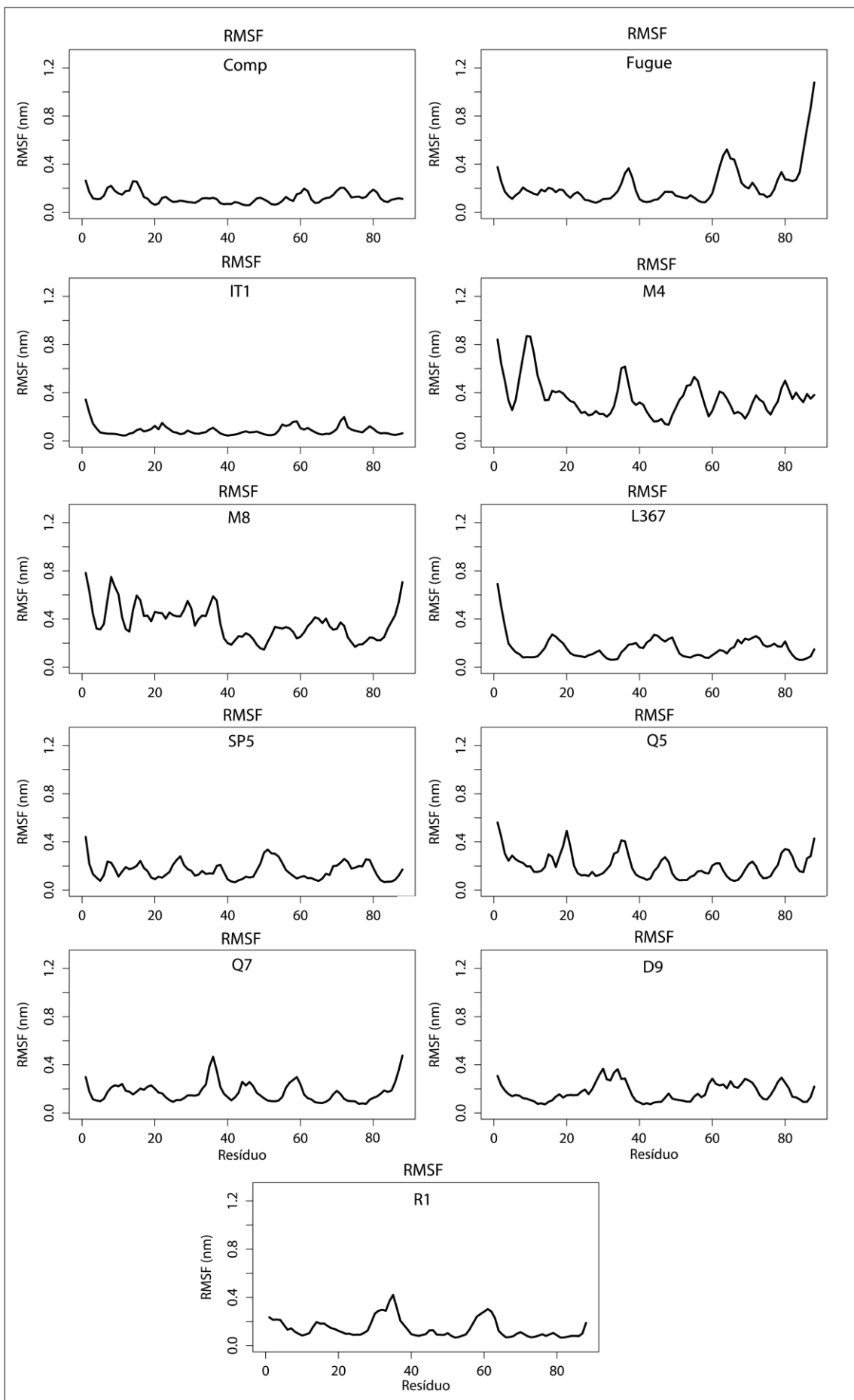


Figura 24: RMS de Flutuação representado por resíduo da proteína.

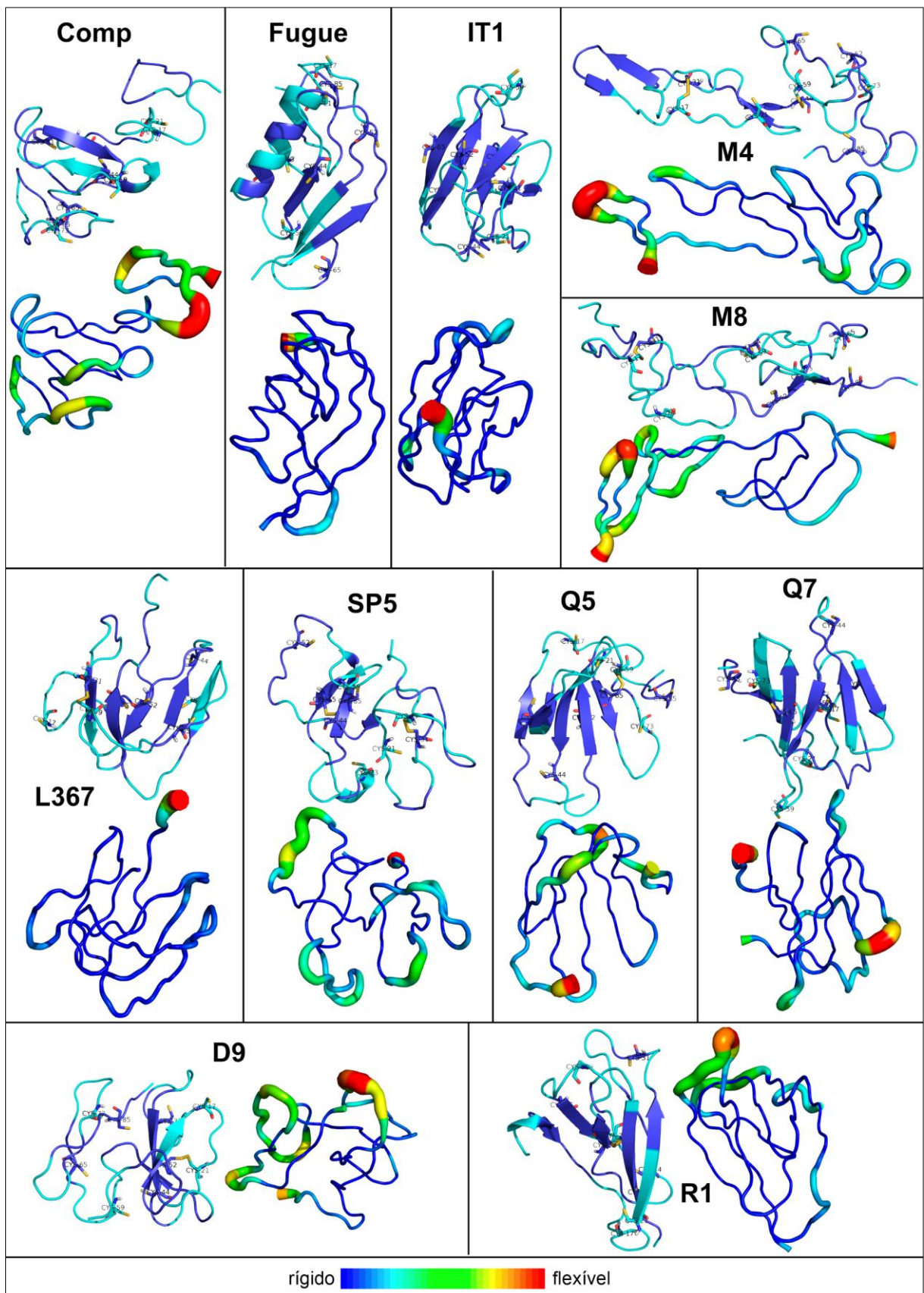


Figura 25: *Fator-B* da estrutura média e respectivos modelos representados em *cartoon*.

A Figura 26 permite avaliar a mudança de elementos de estrutura secundária em diferentes tempos de simulação, inicial e estrutura média, através do cálculo DSSP. O modelo *D9* manteve a quantidade de estrutura secundária entre as duas estruturas. *Q5*, *L3* e *M8* perderam parcialmente estrutura folha- β , enquanto *M4*, *R1* e *IT1* ganharam este elemento, sendo este último modelo, inicialmente pouco estruturado, evoluiu para uma estruturação semelhante à predição de estrutura secundária consenso.

Os modelos *Q7*, *Comp*, *R1*, *SP5* que inicialmente possuíam α -hélice, perderam essa estruturação em suas respectivas estruturas médias. O modelo *Fugue*, que distinguiu-se de todos os outros modelos por apresentar grandes α -hélices, perdeu metade da quantidade dessa estrutura secundária. Todas essas tendências mostram que possivelmente a *cyspep* não apresenta esse elemento de estrutura secundária. A Figura 27 permite visualizar a evolução temporal de elementos de estrutura secundária durante as simulações dos diferentes modelos através do cálculo de DSSP. A α -hélice, quando aparece, se desfaz rapidamente na simulação e não volta a ser estruturada, com ressalva do modelo *Fugue*. Este apresentou três α -hélices inicialmente, sendo que duas delas foram perdidas por completo após 110ns de simulação.

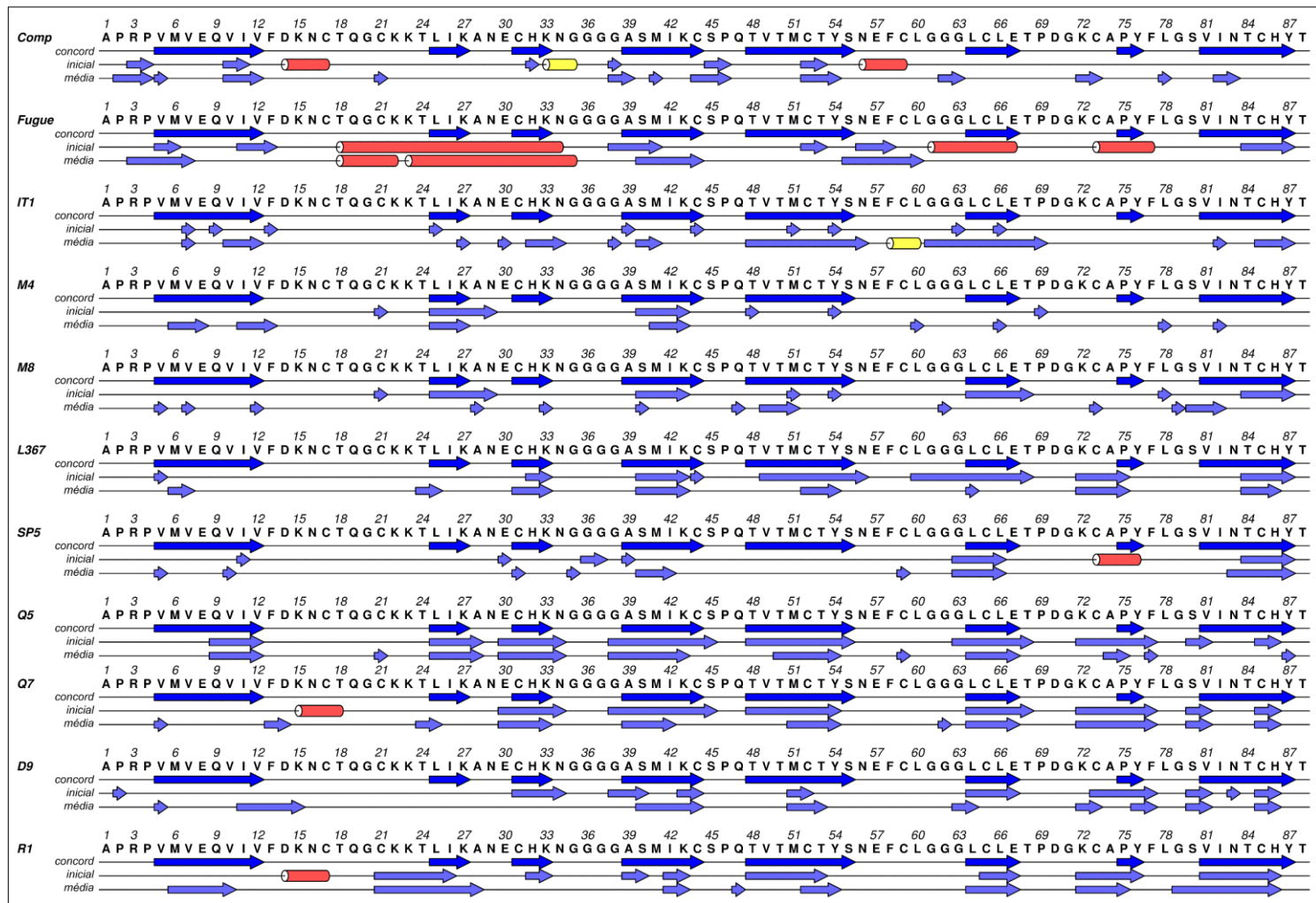


Figura 26: Estrutura secundária dos modelos submetidos à dinâmica molecular. A sequência da *cyspep* é exibida na parte superior da imagem, seguida pela predição de estrutura secundária consenso – concord – em azul escuro (cilindros vermelho, α -hélices; cilindros amarelo, 3-10 hélices; setas azuis, folhas- β). O conteúdo da estrutura secundária de cada conformação inicial e média ao longo de 200ns de simulação. A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

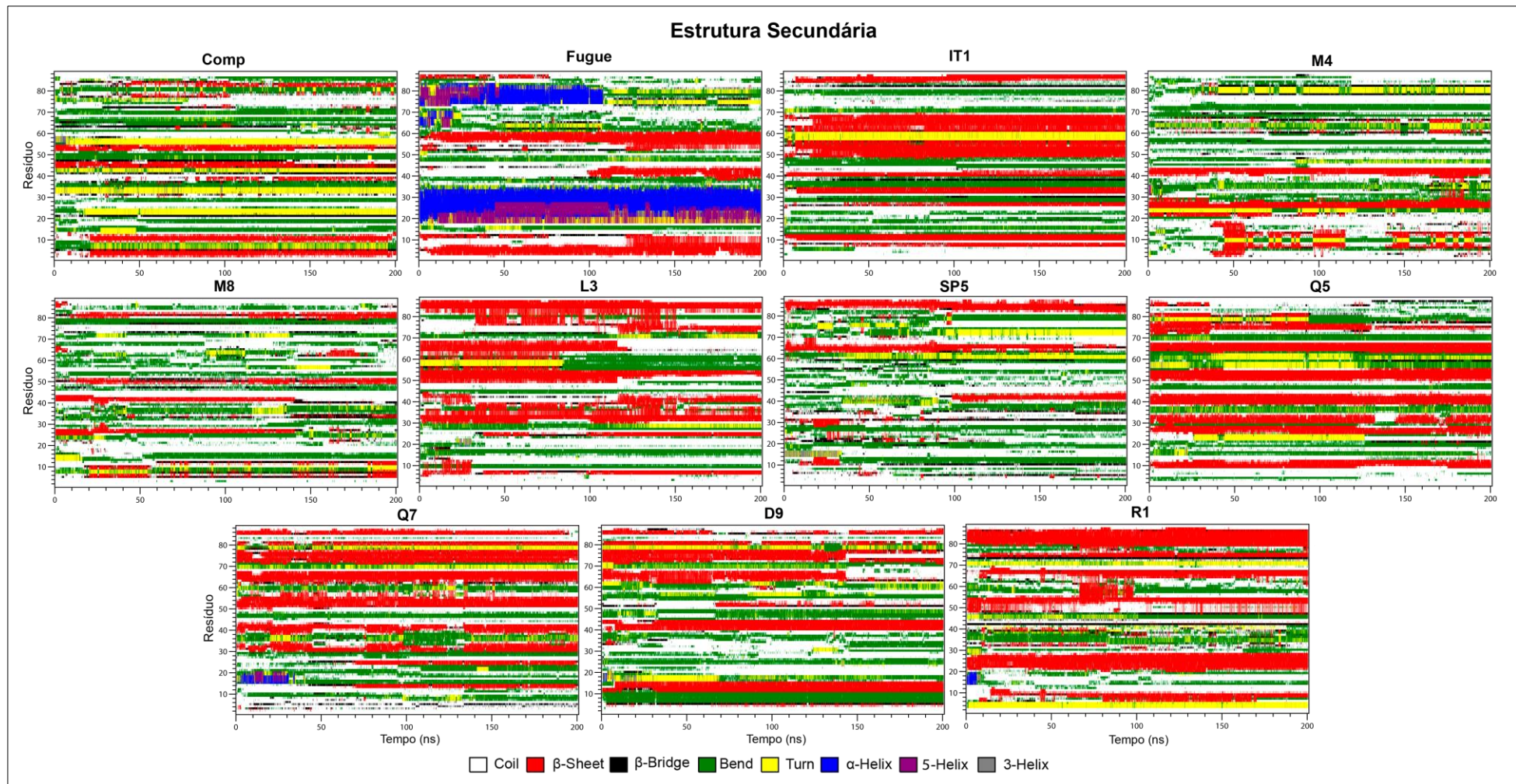


Figura 27: Representação da evolução temporal de elementos de estrutura secundária durante as simulações em diferentes modelos.

4.4.2 Influência do campo de força sobre a Dinâmica da estrutura Fugue

Os resultados das simulações de dinâmica molecular realizadas utilizando o campo de força *GROMOS96 53a6* mostraram uma convergência para ausência de α -hélice na estrutura tridimensional da *cyspep*, e concomitante estruturação de folha- β de acordo com a predição de estrutura secundária. Como a estrutura *Fugue* apresentou maior percentagem de α -hélice, 33% (Figura 26), realizamos simulações utilizando diferentes campos de força: *AMBER99SB ILDN*, *CHARMM22*, *GROMOS96 53a6*, para avaliar a influência dos mesmos na formação de estruturas.

Na Figura 28 é possível observar os valores de RMSD, Rg e SASA do modelo *Fugue* ao longo da trajetória de 200ns em simulações realizadas segundo os diferentes campos de força. A simulação com *AMBER99SB ILDN* mostrou acréscimo de RMSD até 90ns, em seguida as oscilações seguiram em torno de 1,2nm até o final da simulação. Nenhuma grande variação em RMSD foi observada em *CHARMM22*, apresentando o valor mais baixo entre as simulações comparadas, oscilando em torno de 0,5nm. Já *GROMOS96 56a6* mostrou um aumento em RMSD até 110ns, passando por uma estabilização em torno de 0,8nm até 196nm antes de diminuir ao final da simulação para 0,6nm.

Os raios de giro das simulações por *CHARMM22* e *GROMOS96 56a6* apresentaram menor valor em relação a *AMBER99SB ILDN*, oscilando entre 1,2nm e 1,3nm. Com *AMBER99SB ILDN* essa oscilação variou entre 1,5nm e 1,6ns após 90ns de simulação. As informações de RMSD e raio de giro da simulação com campo de força *AMBER99SB ILDN* indicaram que o alto valor de RMSD está relacionado com a baixa compactação que a estrutura adquiriu após 90ns de simulação.

Apesar de *GROMOS96 56a6* possuir baixo valor de Raio de Giro em relação a *AMBER99SB ILDN*, apresentou o maior valor de superfície acessível ao solvente, oscilando entre 30 e 35nm². Isso se deve à maior área exposta entre os resíduos Leu60 a Glu67, que na estrutura inicial e nas simulações por *AMBER99SB ILDN* e *CHARMM22* é um segmento composto de α -hélices, enquanto em *GROMOS96 56a6* tornou-se região formada por *coil*.

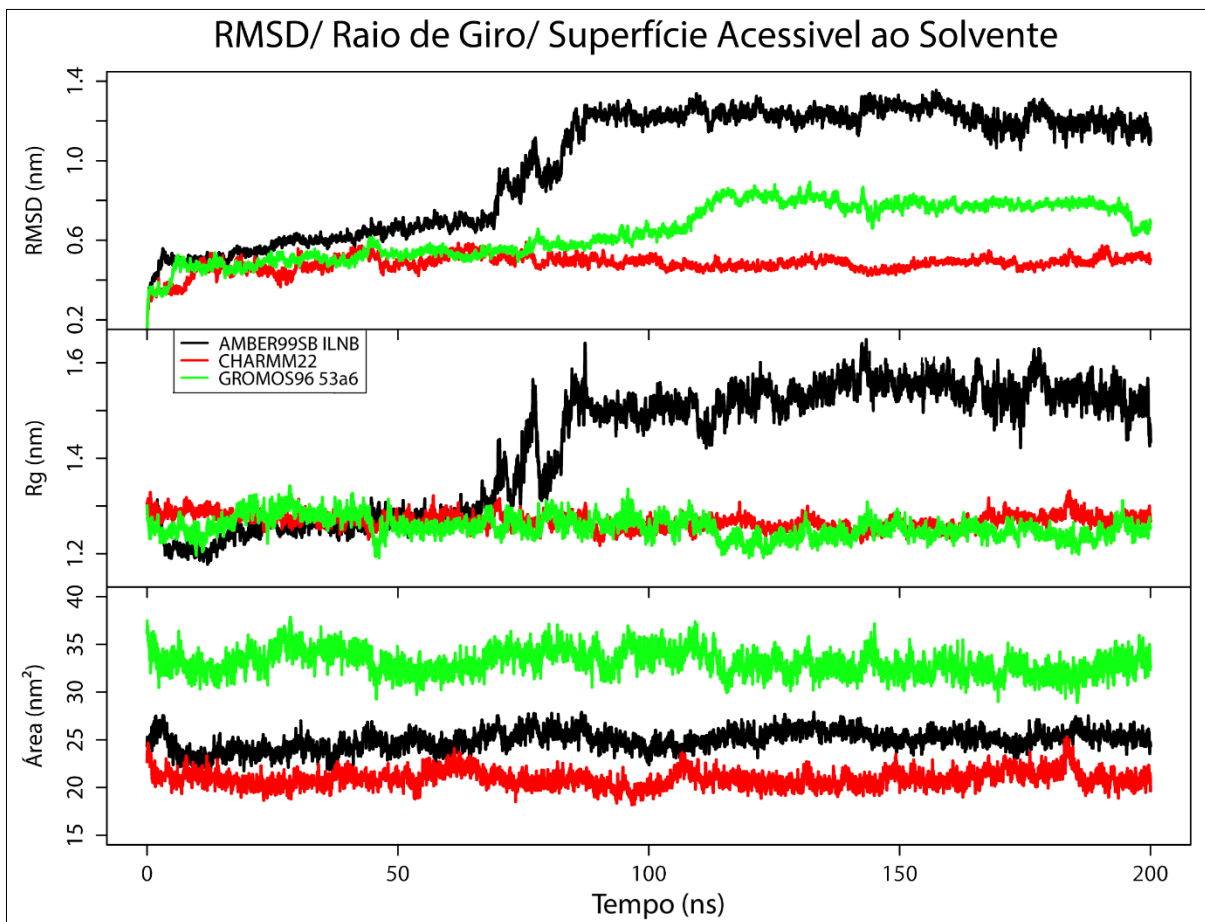


Figura 28: Desvios RMS, Rg e SASA dos átomos da cadeia principal (backbone) do modelo *Figue*, através de diferentes campos de força, representados em função do tempo de simulação.

A Figura 29A mostra o gráfico de RMSF, as flutuações de maiores valores, 0,7nm até 1,0nm, estão associadas com os resíduos que compõem os segmentos N e C terminais. A simulação com o campo de força *CHARMM22* apresentou menor flutuação em seus resíduos em relação aos demais campos de força, o que corrobora os resultados de RMSD.

Para uma melhor visualização, a Figura 29B mostra a análise de fator-B, indicando que a região com maior flexibilidade na simulação por *AMBER99SB ILDN* está compreendida entre os resíduos Asn29 e Met41 e a maiores flutuações ocorrem nos Gly36, Gly37, Gly38 e Ala39, uma região de *coil*. Na simulação por *GROMOS96 53a6* os quatro últimos resíduos (Cys85, His86, Tyr87 e Thr88) estão mais flexíveis; esta região mostrava uma estrutura secundária tipo folha- β no modelo original (Figura 30). Os resíduos de cisteína encontram-se altamente estáveis exibindo menos flutuação, com exceção da simulação com *AMBER99SB* em que a Cys31 mostrou maior flexibilidade.

Em relação à desestruturação de α -hélices das estruturas médias nas simulações (Figuras 29B e 30), o campo de força *GROMOS96 53a6* reduziu de 33% para 14,8%, seguido do *CHARMM22* com 21,6% e *AMBER99SB ILDN* com 25%. Apesar da simulação por *CHARMM22* ter sido a mais estável, *GROMOS96 53a6* evidenciou maior redução de α -hélices, e por tal motivo essa simulação foi estendida até 500ns. Comparando as estruturas médias com 200ns e 500ns de simulação, houve acréscimo de folha- β , de 18,2% para 27,3% (Figuras 29B e 30).

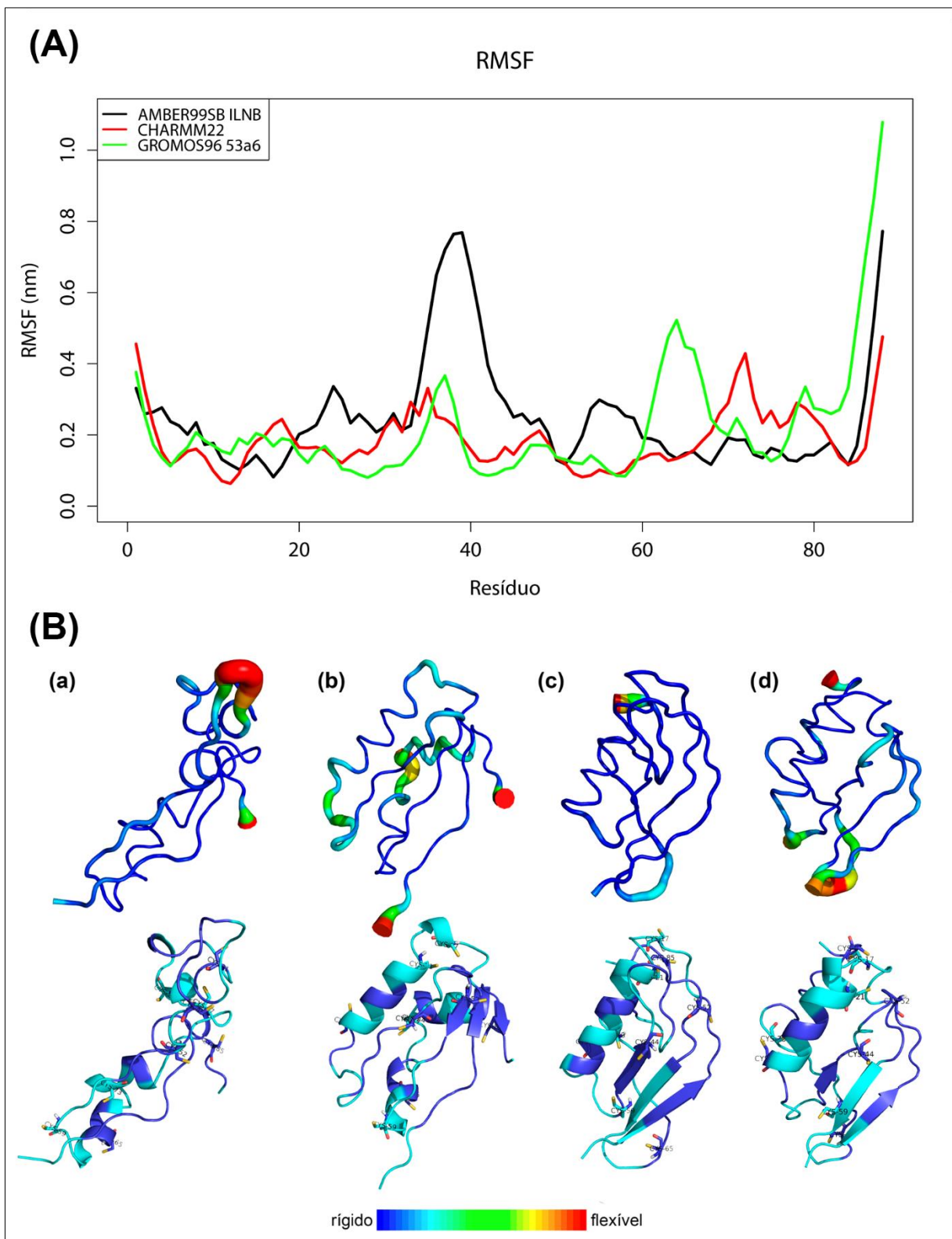


Figura 29: (A) RMS de Flutuação do modelo *Fugue* simulado através de diferentes campos de força. (B) *Fator-B* da estrutura média do modelo *Fugue* em diferentes campos de força (a) AMBER99SB ILDN (b) CHARMM22 (c) GROMOS96 56a6 (d) GROMOS96 56a6 500ns e abaixo seus respectivos modelos representados em *cartoon*.

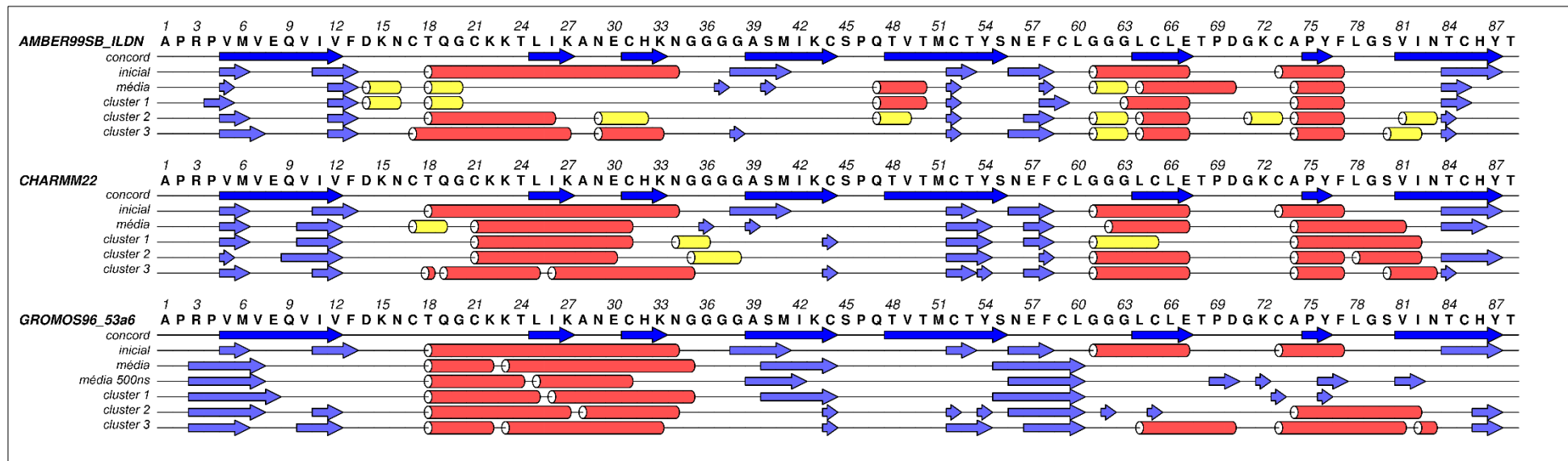


Figura 30: Estrutura secundária do modelo *Fugue* influenciado por diferentes campos de força. A sequência da *cyspep* é exibida na parte superior da imagem, seguida pela predição de estrutura secundária consenso – concord – em azul escuro (cilindros vermelhos, α -hélices; cilindros amarelo, 3-10 hélices; setas azuis, folhas- β). O conteúdo da estrutura secundária da conformação inicial *Fugue* e de cada estrutura média e três maiores clusters ao longo 200ns de simulação, além da estrutura média simulada pelo campo de força *GROMOS96 53a6* ao longo de 500ns de simulação. A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

A Figura 31 mostra que a conformação dos elementos estruturais secundários foi afetada durante o curso das simulações. Observamos que a α -hélice não é preservada em nenhuma das estruturas simuladas, ou pelo menos necessitaria de mais tempo de simulação para analisar a tendência da estrutura quando simulada com o campo de força *CHARMM22*. Dos diferentes campos de força, *AMBER99SB ILDN* apresentou tendência para uma estrutura aleatória composta por *turn* e *coil*, enquanto *CHARMM22* e *GROMOS96 53a6* estabilizaram em 135ns e 110ns, respectivamente.

As poses das estruturas foram clusterizadas com *cutoff* de 0,35nm. *AMBER99SB ILDN* apresentou maior número de clusters, 12, enquanto *CHARMM22* apresentou o menor, 5, seguido do *GROMOS96 53a6* com 8 (Figura 32). Esses valores enfatizam as tendências mostradas pelas análises de RMSD e RMSF (Figuras 28 e 29A).

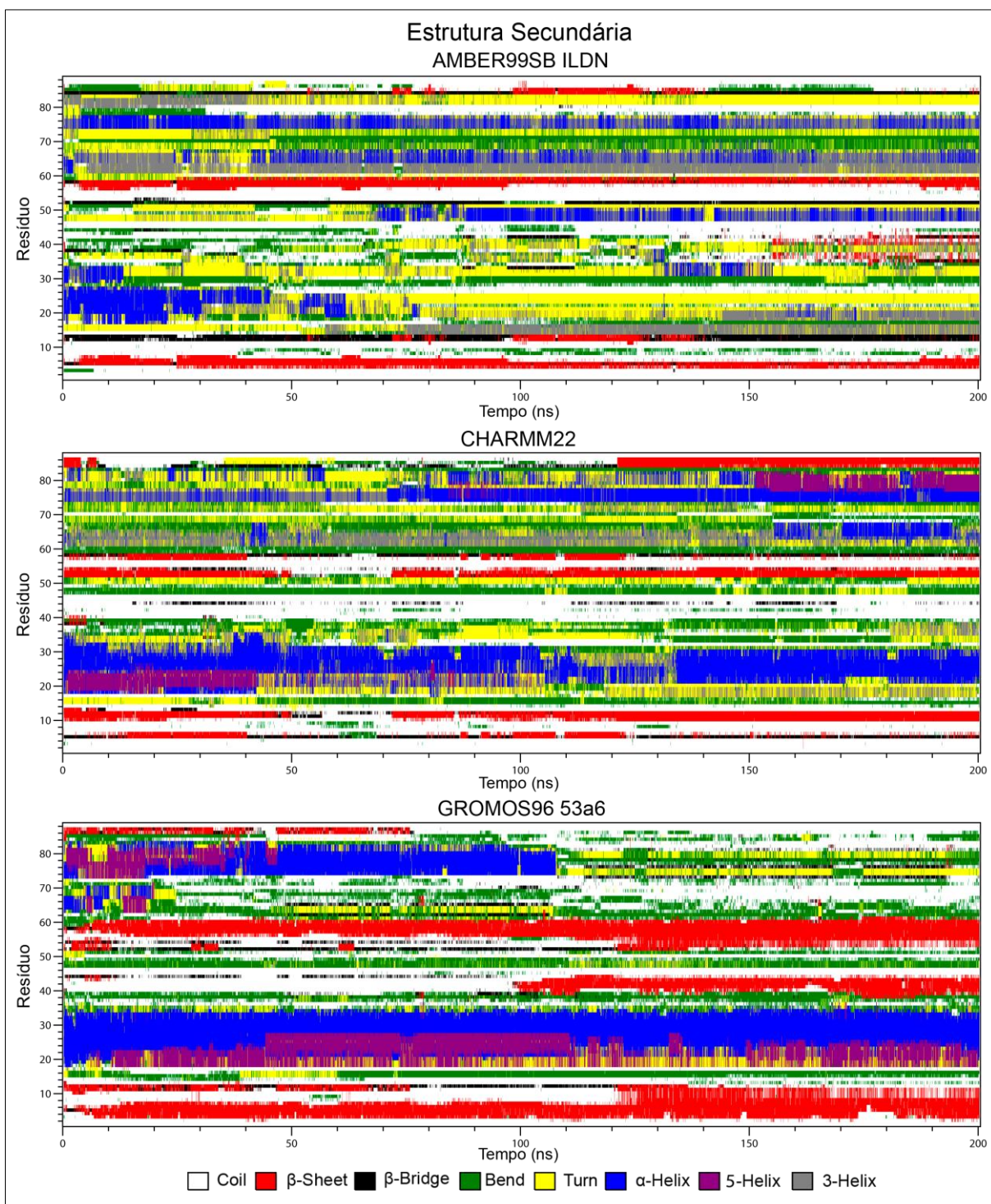


Figura 31: Representação da evolução temporal de elementos de estrutura secundária durante as simulações através de diferentes campos de força do modelo *Fugue*.

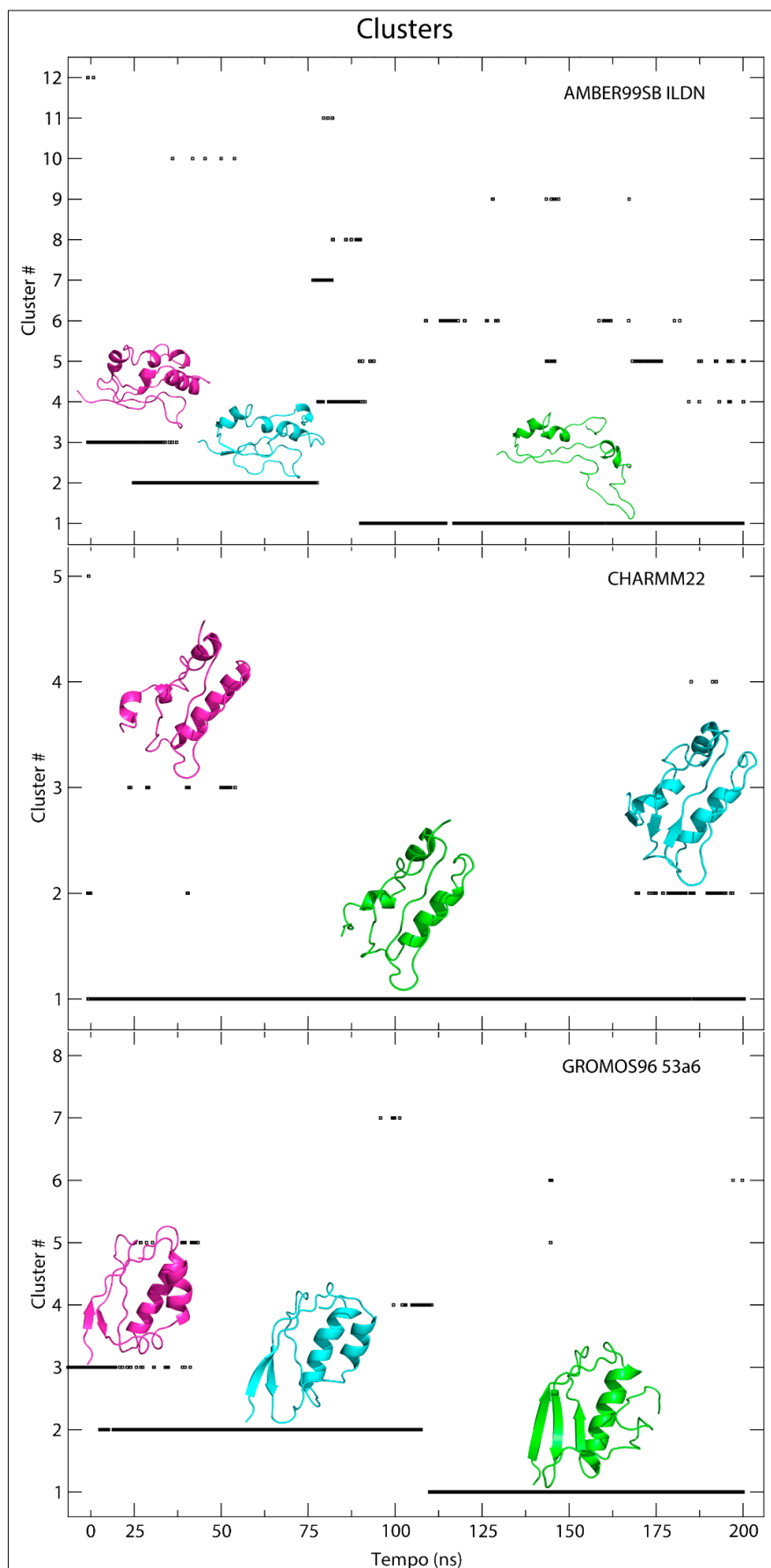


Figura 32: Número de clusters ao longo de 200ns de simulação do modelo *Fugue* através de diferentes campos de força. Em *cartoon* a estrutura representativa dos três maiores clusters (verde, cluster 1; azul, cluster 2; rosa, cluster 3).

4.4.3 Influência do campo de força sobre a Dinâmica da estrutura I-Tasser

Assim como realizamos um estudo para verificar a desestruturação de α -hélice do modelo *Fugue*, uma análise equivalente foi feita para examinar a presença de folha- β , já que a predição de estrutura secundária indica o predomínio desse elemento em relação aos demais. Foram realizadas simulações com diferentes campos de força (*AMBER99SB ILDN*, *CHARMM22*, *GROMOS96 53a6*) com a estrutura *IT1*, que se destacou ao ganhar mais estrutura secundária folha- β entre os modelos simulados.

IT1 foi influenciado de modo diferente pelos campos de força em relação à estrutura *Fugue*. Os resultados de RMSD mostraram menor variação em *AMBER99SB ILDN* e maior em *CHARMM22* (Figura 33). Regiões com maiores oscilações em valores de RMSD coincidiram com as variações de raio de giro, o que sugere que a compactação da proteína influencia em sua mudança estrutural. *GROMOS96 53a6* apresentou maior área acessível ao solvente em relação aos outros campos de força, devido a maior exposição entre os resíduos Ser55 e Gly62.

A Figura 34A mostra as flutuações em cada resíduo. *CHARMM22* apresentou maior flutuação entre os demais, mostrando maior flexibilidade nos resíduos Lys15 e Asn16 e o segmento compreendido entre os resíduos Ala28–Cys31, regiões de *coil*. Os resíduos de cisteína se mostraram altamente estáveis exibindo menor flutuação, com exceção do *CHARMM22* em que Cys31 ficou mais flexível, coincidente com a cisteína que resultou em uma maior flexibilidade quando simulada com o campo de força *AMBER99SB ILDN* para a estrutura *Fugue*.

Em relação à estruturação de folha- β das estruturas médias (Figuras 34B e 35), houve maior ganho no campo de força *GROMOS96 53a6* contendo 34% desse elemento, *AMBER99SB ILDN* e *CHARMM22* ficaram iguais com 22%, porém *CHARMM22* teve um acréscimo de α -hélice, 4,5%.

A estrutura do modelo *IT1* simulada com o campo de força *GROMOS96 53a6* evidenciou aumento mais significativo de folha- β , assim sua simulação foi estendida até 500ns. Comparando as estruturas médias com 200ns e 500ns de simulação, houve perda de 2% de folha- β e do elemento 3_{10} hélice, 3,4%, sendo estes substituídos por *coil* (Figuras 34B e 35).

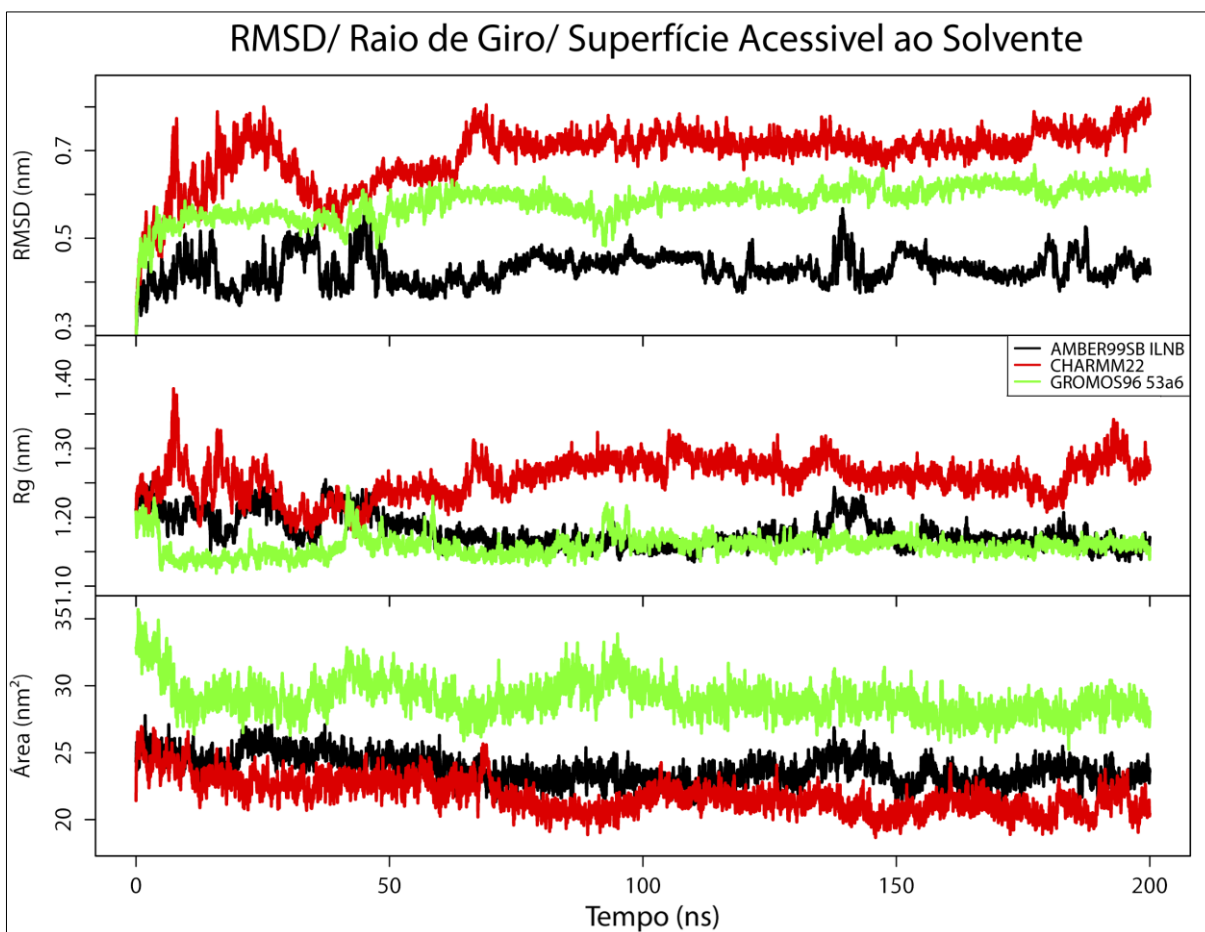


Figura 33: Desvios RMS, Rg e SASA dos átomos da cadeia principal (backbone) do modelo *IT1*, através de diferentes campos de força, representados em função do tempo de simulação.

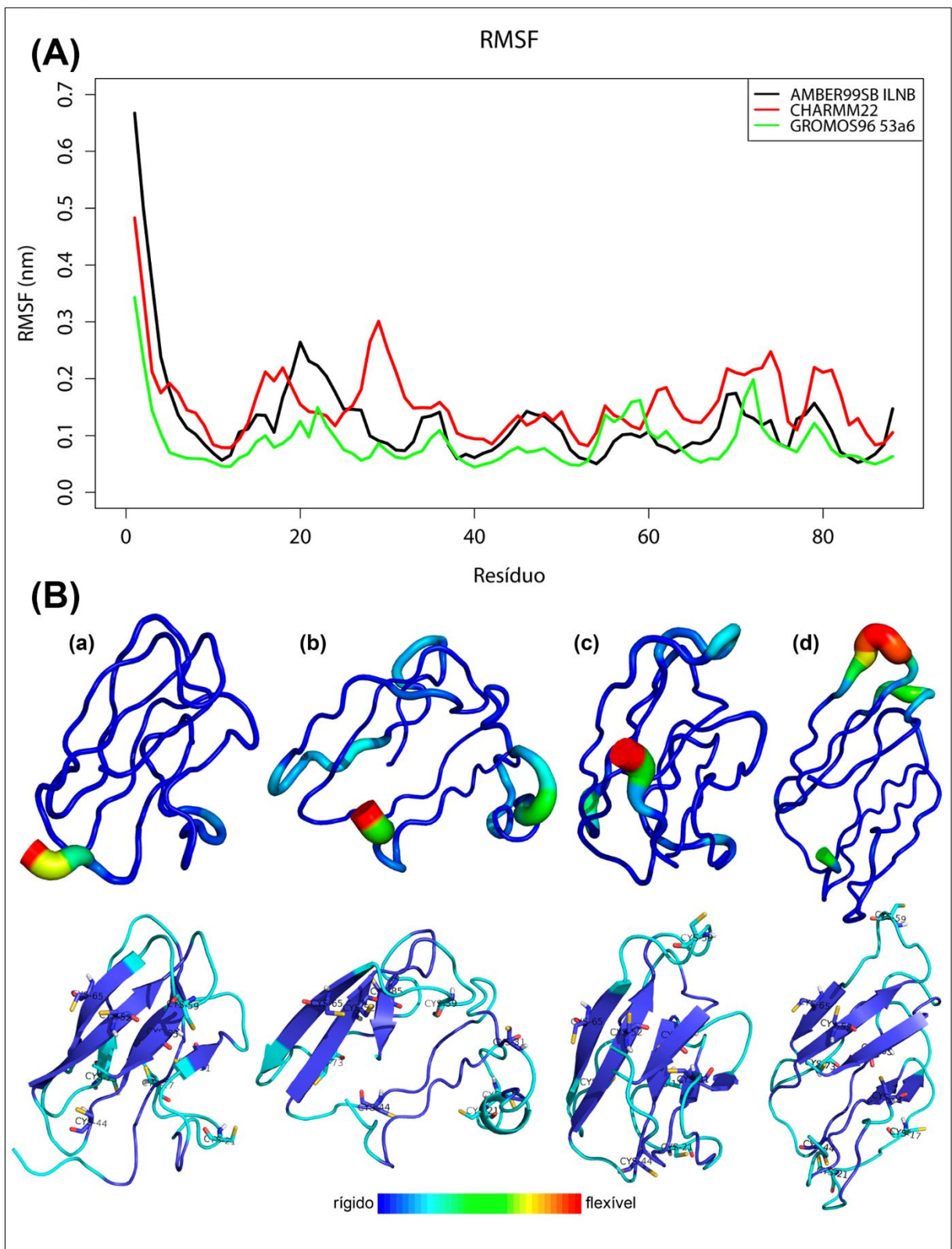


Figura 34: (A) RMS de Flutuação do modelo *IT1* simulado através de diferentes campos de força. (B) *Fator-B* da estrutura média do modelo *IT1* em diferentes campos de força (a) *AMBER99SB ILDN* (b) *CHARMM22* (c) *GROMOS96 56a6* (d) *GROMOS96 56a6 500ns* e abaixo seus respectivos modelos representados em *cartoon*.

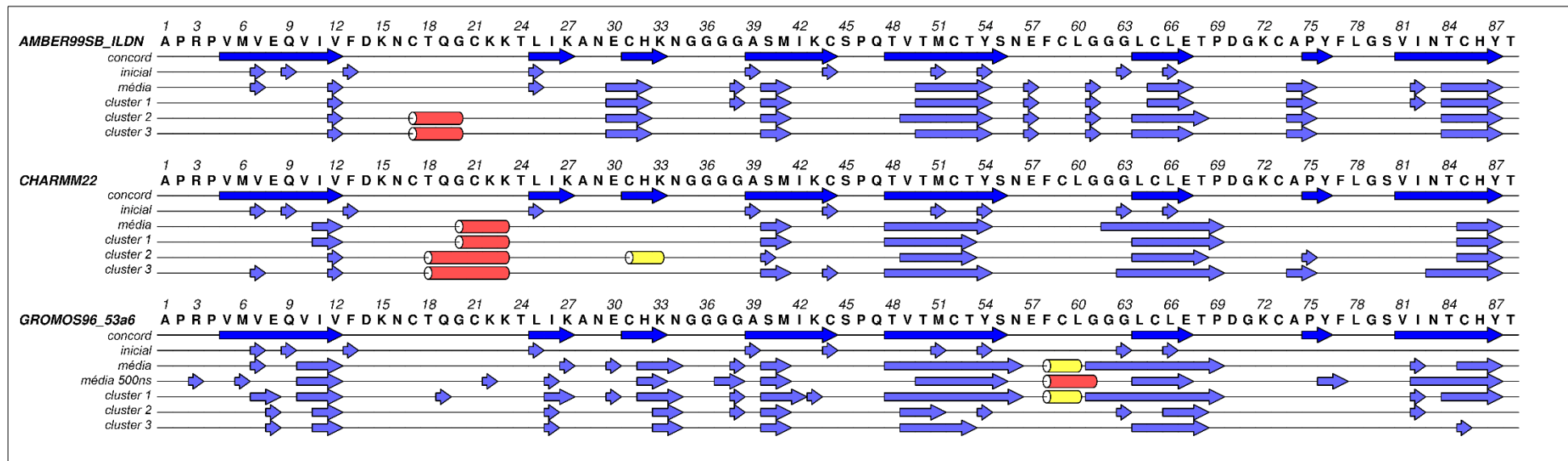


Figura 35: Estrutura secundária do modelo *It1* influenciado por diferentes campos de força. A sequência da *cyspep* é exibida na parte superior da imagem, seguida pela predição de estrutura secundária consenso – concord – em azul escuro (cilindros vermelhos, α -hélices; cilindros amarelo, 3-10 hélices; setas azuis, folhas- β). O conteúdo da estrutura secundária da conformação inicial *It1* e de cada estrutura média e três maiores clusters ao longo 200ns de simulação, além da estrutura média simulada pelo campo de força GROMOS96 53a6 ao logo de 500ns de simulação. A figura foi produzida pelo software ALINE (Bond and Schuttelkopf 2009).

A Figura 36 mostra que a conformação dos elementos estruturais secundários foi afetada durante o curso das simulações. Observamos que a α -hélice não é preservada em nenhuma das estruturas simuladas, apesar de sumir e reaparecer em *CHARMM22*. Apesar das simulações realizadas com *GROMOS96 53a6* possuírem maiores valores de RMSD que as simuladas com *AMBER99SB ILDN*, a evolução dos elementos de estrutura secundária foi mais estável, o que se reflete na mesma quantidade de clusters, idêntica às simulações com *AMBER99SB ILDN* (Figura 37).

As poses das estruturas foram clusterizadas com *cutoff* de 0,3nm. *CHARMM22* apresentou maior número de clusters (14), enquanto *AMBER99SB ILDN* e *GROMOS96 53a6* apresentaram um número mais reduzido (5) (Figura 37). Esses valores estão de acordo com os de RMSD, RMSF e elementos de estrutura secundária.

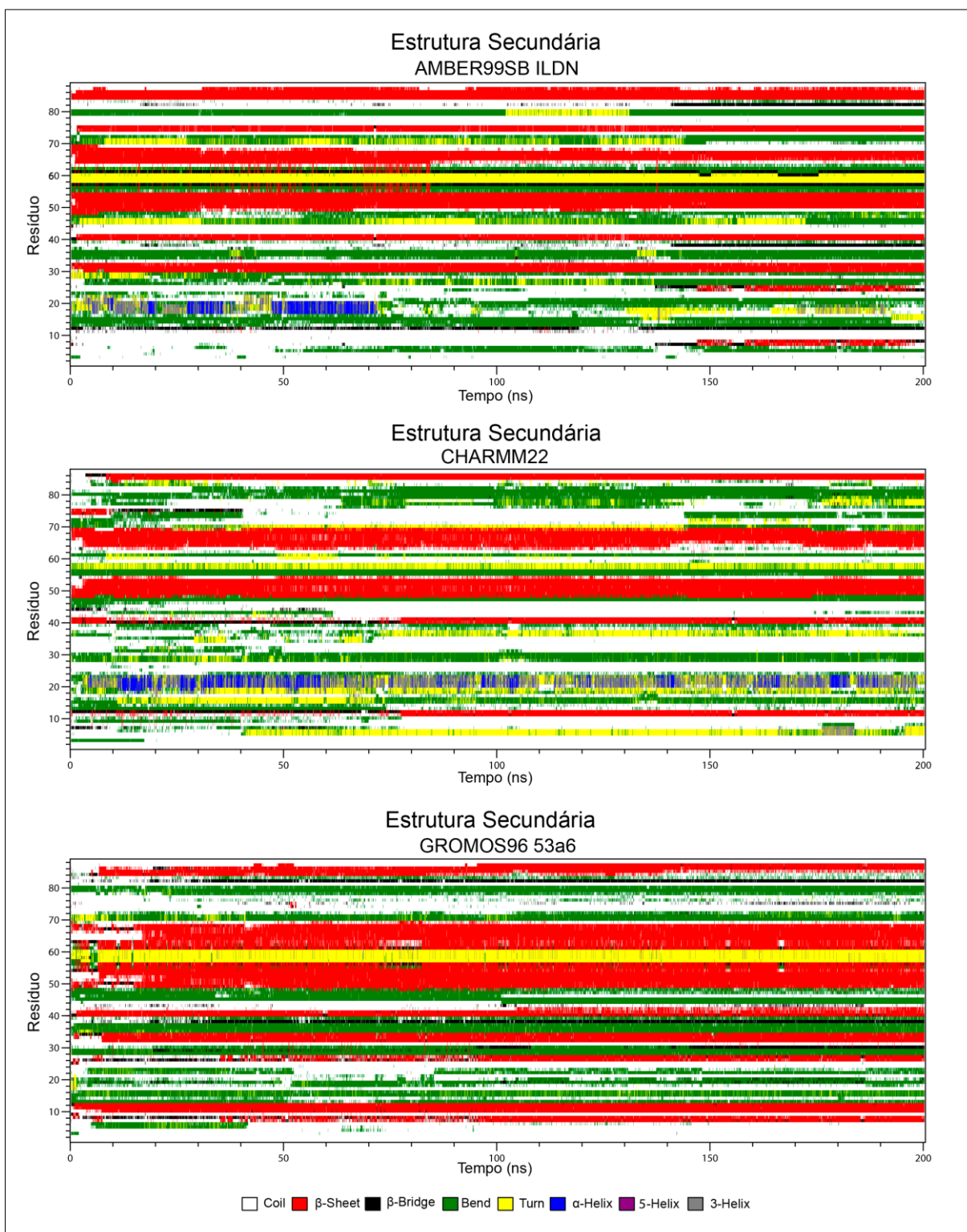


Figura 36: Representação da evolução temporal de elementos de estrutura secundária durante as simulações através de diferentes campos de força do modelo *IT1*.

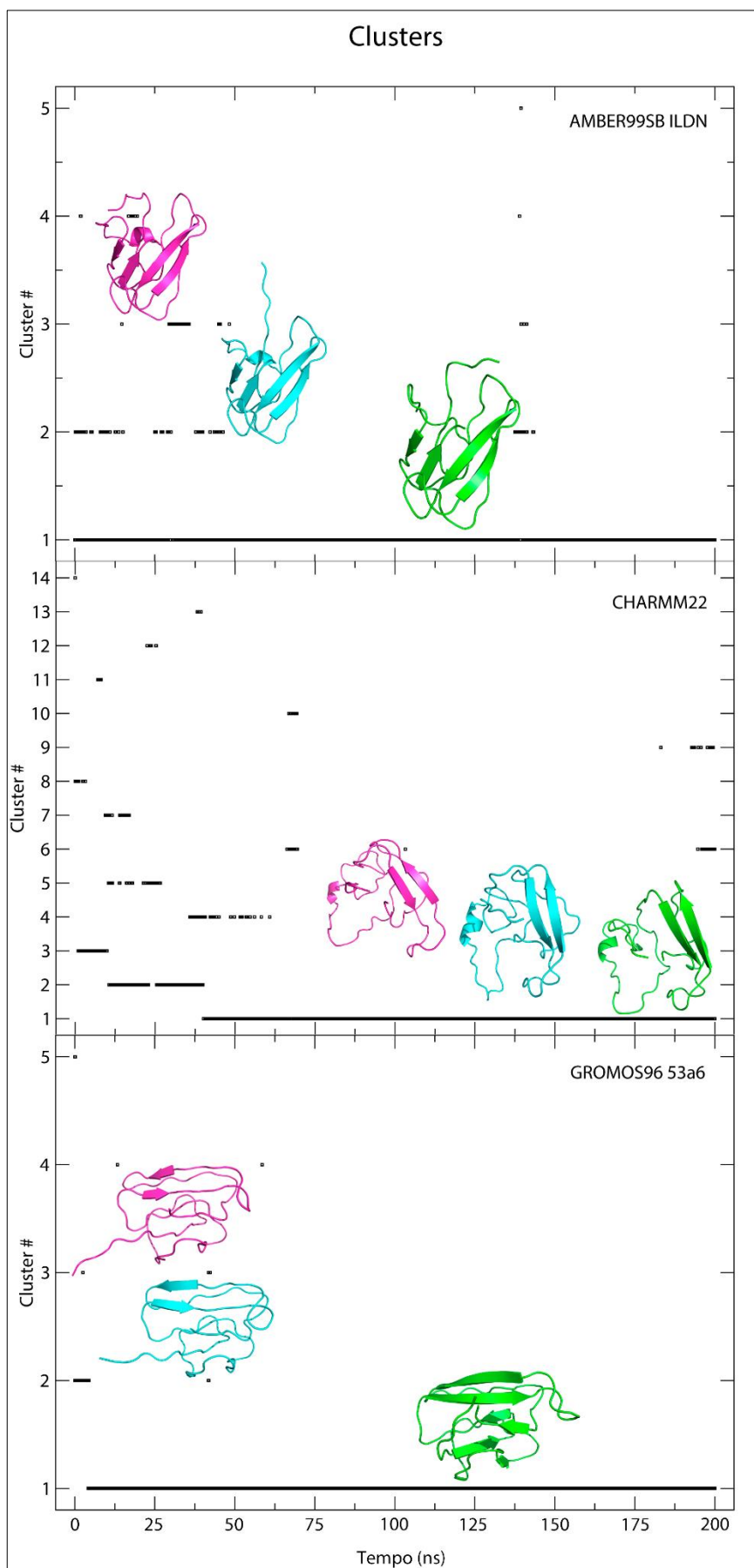


Figura 37: Número de clusters ao longo de 200ns de simulação do modelo *IT1* através de diferentes campos de força. Em *cartoon* a estrutura representativa dos três maiores clusters (verde, cluster 1; azul, cluster 2; rosa, cluster 3).

5 DISCUSSÃO

A infecção por *Leishmania* é um processo complexo, em que tanto o sistema imune do hospedeiro quanto os componentes de virulência do parasita influenciam o resultado da infecção. Alguns fatores de *L. (L.) amazonensis* são capazes de interagir diretamente com o sistema imune do hospedeiro e influenciar uma resposta que beneficia a sobrevivência do parasita (De Souza Leao, Lang et al. 1995).

Nesse contexto, a *cyspep* possui sua sequência de aminoácidos diferente de outras espécies de *Leishmania*, as quais podem indicar existência de epítomos exclusivos para *L. (L.) amazonensis*. Estes epítomos podem estar relacionados com algumas peculiaridades da infecção causada por esse parasito contribuindo para o estímulo de subpopulações de linfócitos relacionadas à patogênese (Soong, Chang et al. 1997).

A atuação desse peptídeo, que é eliminado pelo parasito e pode ser detectado dentro e fora de macrófagos, tende a definir a interação que o mesmo influencia na interação entre o parasito e o hospedeiro, desencadeando respostas favoráveis ao sucesso da vida do parasito, ou seja, contribui ao estilo de vida da *L. (L.) amazonensis*. Desta forma, saber como se organiza a estrutura deste polipeptídeo pode ser útil na descoberta de funções ainda não descritas.

Porém não é trivial a tarefa de se obter a estrutura tridimensional de uma proteína. As técnicas experimentais como a cristalografia de raios-X e ressonância magnética nuclear (NMR) possuem limitações, como qualidade necessária dos cristais obtidos para o trabalho experimental, e a obtenção de amostras em quantidade suficiente para os ensaios, respectivamente. Ainda, o sequenciamento de peptídeos caminha em passos mais largos que a resolução tridimensional de suas estruturas correspondentes, o que gera uma grande defasagem no que diz respeito ao conhecimento sequência-estrutura-função (Santos Filho and Alencastro 2003).

Em consequência, foram desenvolvidos outros métodos de elucidação de estruturas tridimensionais de proteínas baseados em metodologias *in silico* como modelagem comparativa, *threading* e métodos *de novo*. Recentes avanços em algoritmos computacionais para prever a estrutura e função de proteínas têm abrandado o problema, e fornecem informações valiosas sobre as proteínas de interesse.

As proteínas possuem propriedades conformacionais peculiares que lhes permitem enovelar e desenovelar em sua conformação tridimensional nativa. Essas propriedades não são aleatórias, pois existe uma estreita relação entre a função e a estrutura, ou seja, uma proteína só irá exercer sua função eficientemente se estiver em sua estrutura tridimensional nativa, aquela que é alcançada em condições fisiológicas apropriadas (Hegyí and Gerstein 1999). Isto posto, o desenho experimental assumido nesse trabalho – predição da estrutura secundária consenso e de ligação dissulfeto; seguida de modelagem comparativa, *threading* e métodos *de novo*; e dinâmica molecular – foi oportuno para verificar a tendência da estrutura tridimensional dessa proteína.

Desta forma, técnicas *in silico* para predizer conformações estruturais para *cyspep* foram empregadas no presente estudo e os resultados sugeriram que a proteína é formada preferencialmente pela estrutura secundária folha- β . Essa estrutura se manteve presente mesmo quando submetida a longas simulações de dinâmica molecular, da ordem de centenas de nanossegundos. O mesmo não pôde ser observado em relação às α -hélices que se desfizeram durante a simulação. Essas características estruturais podem ser explicadas com base nas propriedades físico-químicas dos aminoácidos que formam a *cyspep*.

A α -hélice está associada com um momento dipolo em que a região amino-terminal da hélice tem uma carga positiva enquanto a carboxi-terminal tem uma carga negativa; assim, os resíduos que são negativamente carregados na extremidade amino e positivamente carregados na extremidade carboxi tendem a estabilizar a α -hélice, já os resíduos com carga oposta tendem a desestabilizar a α -hélice (Schlick 2010).

Estudos mostraram que resíduos com restrições na conformação da cadeia lateral, devido a grupamentos grandes ou volumosos, são menos propensos a formarem α -hélice do que outros resíduos. Por exemplo, glutamina, metionina, leucina favorecem a formação de α -hélice, ao passo que a valina, serina, ácido aspártico, asparagina, prolina e glicina tendem a desestabilizá-las (Schlick 2010). A prolina quebra ou torciona a α -hélice devido ao fato de possuir uma estrutura quimicamente coesa e rígida que não pode doar uma ligação de hidrogênio e porque a sua cadeia lateral interfere estericamente com a cadeia principal do resíduo anterior, no entanto, a prolina pode ser encontrada como o primeiro resíduo de uma hélice. A glicina também tende a perturbar hélices devido a sua elevada flexibilidade conformacional

tornando-se entropicamente desfavorável adotar uma estrutura helicoidal (Pace and Scholtz 1998).

Os resíduos desfavoráveis para formação de α -hélice correspondem a 36,4% dos aminoácidos da sequência da *cyspep*, enquanto os favoráveis correspondem apenas 12,5% desse total, o que não colabora para sua formação e talvez seja a razão destas se desfazerem e não estabilizarem com as simulações de dinâmica molecular, mesmo sob a influência dos diferentes campos de força.

Já em relação às folhas- β , o anel de prolina, como para α -hélices, não se adapta bem, uma vez que não pode participar da rede de ligações de hidrogênio entre *strand*. Os aminoácidos valina, isoleucina e fenilalanina são propensos a formar folha- β (Schlick 2010). Na *cyspep*, as cinco prolinas existentes foram preditas como regiões de *coil* pelo servidor de predição de estrutura secundária consenso e na grande maioria dos modelos criados, as prolinas, também compõem regiões de *coil*. Já os aminoácidos com propensão à folha- β compõem a predição de estrutura secundária consenso e estão presentes em grande parte dos modelos propostos. A dinâmica molecular do modelo *IT1* sob a influência de diferentes campos de força também corrobora a predominância desse elemento estrutural.

Apesar da *cyspep* demonstrar ser um polipeptídeo *all*- β , ou seja, uma classe de domínios estruturais em que a estrutura secundária é composta inteiramente por folhas- β com apenas voltas e loops interligando *strands* adjacentes, não apresenta padrões de dobras consenso entre diferentes modelos. De acordo com o banco de dados de classificação estrutural de proteínas (SCOP - *Structural Classification of Proteins*) (Lo Conte, Ailey et al. 2000), existem 174 tipos de dobras com apenas conformações de folhas- β (<http://scop.berkeley.edu/sunid=48724>).

A construção *in silico* da estrutura tridimensional da CTE de *L. (L.) mexicana* foi proposta no ano de 2001 (Alves, Figueiredo et al. 2001). Essa proteína possui identidade sequencial de 92% com *cyspep*. A abordagem metodológica e a estrutura gerada foram diferentes do utilizado nesse trabalho. A construção do primeiro modelo foi por meio da modelagem comparativa, porém assim como para *cyspep*, não foram encontrados moldes com identidade sequencial suficiente para aplicação da técnica.

Por meio da sequência de aminoácidos da CTE de *L. (L.) mexicana* nenhum molde foi encontrado mediante aplicação da ferramenta computacional BLAST; já com o da *cyspep*, 25 potenciais moldes foram achados, porém apresentaram baixa identidade sequencial com a *cyspep*. Essa diferença pode ser explicada pelo depósito de 83500 novas estruturas no PDB nos últimos 13 anos, o que poderia ter limitado na

época a possibilidade de encontrar moldes mais próximos. Além disso, estruturas tridimensionais determinadas experimentalmente das cisteíno-proteases são sempre obtidas em sua forma madura, ou seja, sem a CTE, isso explica a falta de homólogos com estrutura definida até a atualidade.

O problema da falta de moldes neste trabalho foi solucionado com o uso de outras técnicas de predição tridimensional de proteínas, *threading* e métodos *de novo*. Como essas técnicas não estavam consolidadas em 2001, Alves e colaboradores construíram uma quimera entre a proteína papaína, homóloga à CPB madura, e acresceu a sequência de aminoácidos da CTE ao final dessa sequência. Essa combinação permitiu encontrar como molde uma glicosiltransferase que possui 37% com quimera (Alves, Figueiredo et al. 2001).

A estrutura final foi composta por três regiões de α -hélices (α -hélice 1, CFDKNCTQGC; α -hélice 2, ANECHKNGGGASM e α -hélice 3, SPQKVTMCTYSNEFCV), duas pequenas regiões de *strain* (PAP e LIK) e o restante foram regiões sem estrutura secundária definida. Essa estruturação não coincide com nenhum dos 72 modelos criados neste trabalho. Não há consenso nem com a estrutura *Fugue*, que também é formada por α -hélice, as hélices são menores e encontram-se distribuídas em regiões diferentes da proteína em relação ao primeiro modelo hipotético proposto (Alves, Figueiredo et al. 2001).

Estes 72 modelos da *cyspep* obtidos no presente estudo devem-se à riqueza de informações disponíveis nos bancos de dados e à praticidade da existência de diversos servidores *online* para predição de estruturas tridimensionais de proteínas, o que não foi possível há 13 anos atrás.

Quanto à presença de ligações dissulfeto, a *cyspep* possui nove cisteínas que têm o potencial de formar até quatro dessas ligações. Como essa proteína já foi observada no meio extracelular e no citoplasma de células hospedeiras (Alves, Corte-Real et al. 2005), fortalece a hipótese da presença de ligações dissulfeto na *cyspep*. A maioria das proteínas que tem sua função biológica no meio extracelular contém ligações dissulfeto (Hogg 2003). Essas ligações influenciam na termodinâmica de enovelamento de proteínas, estabilizam a conformação nativa, mantêm a integridade da proteína frente a agentes oxidantes e enzimas proteolíticas no ambiente extracelular. Ou seja, ao estabilizar a estrutura da proteína, ligações dissulfeto podem proteger as proteínas contra danos e aumentar sua meia-vida (Hogg 2003).

Anteriormente, as ligações dissulfeto presentes em proteínas maduras foram pensadas para serem inertes, ou seja, uma vez formadas permanecem inalteradas

durante a vida da proteína (Hogg 2003). No entanto, parece que isso não é necessariamente o caso. Ligações dissulfeto são clivadas em proteínas maduras e, quando isso acontece, tem importantes consequências para a função da proteína (Hogg 2003). Por exemplo, a redução de ligações dissulfeto é um passo fundamental no processamento de antígeno, tanto para permitir o desdobramento de antígenos de proteínas, aumentando o acesso das enzimas proteolíticas de processamento, quanto para expor resíduos de cisteínas livres dentro de epítomos peptídicos lineares reconhecidos por linfócitos T (Collins, Unanue et al. 1991, Fairlie, Stanton et al. 1996, Haque, Hawes et al. 2001).

Nesse aspecto, a *cyspep* foi verificada quanto à ação de seus peptídeos derivados na modulação da resposta imune do hospedeiro vertebrado (Souza-Silva, Pereira et al. 2014), e relacionando com as ligações dissulfeto, foi analisado que essas ligações podem regular o processamento de antígenos, influenciando assim a seleção de epítomos por moléculas de MHC de classe II. A mutagênese de ligações dissulfeto específicas dentro do antígeno kII de imunoglobulina G demonstrou influenciar a eficiência de apresentação de epítomo por moléculas de MHC de classe II. Este estudo demonstrou que a clivagem de ligações dissulfeto dentro de um antígeno de um epítomo específico pode regular a apresentação e seleção de epítomo (Li, Haque et al. 2002).

Interessantemente, o epítomo *P6* (FCLGGGLCL) da *cyspep*, que foi testado e demonstrou ter interação no sistema imune de camundongos (Silva-Almeida, Pereira et al. 2012, Souza-Silva, Pereira et al. 2014), apresenta cisteínas em sua composição. O *P6* é exclusivo de *L. (L.) amazonensis* e as cisteínas Cys59–Cys65 formam uma ligação dissulfeto em um dos modelos gerados pelo método *de novo*, *Q5*.

Porém, não podemos afirmar se a *cyspep* apresenta ligações dissulfeto, já que não houve consenso dos servidores de predição dessas ligações e em relação aos modelos criados, não apresentaram uma ligação predominante e nem formações adicionais nas simulações de dinâmica molecular, na escala de tempo observada. Para afirmar a sua presença ou ausência e em quais resíduos se encontram serão necessárias análises experimentais, como ensaios de espectrometria de massas (Gorman, Wallis et al. 2002).

Assim como para ligações dissulfeto, não podemos afirmar qual modelo/servidor propôs uma melhor estrutura tridimensional para a *cyspep*. É sabido que a modelagem baseada em modelo torna-se cada vez mais poderosa, pois as dobras de estruturas mais populares, adotadas por várias sequências, são conhecidas (Dai and Zhou

2011). No entanto, o reconhecimento de dobras de estrutura semelhantes na ausência de similaridade sequencial, como no caso da *cyspep*, é um desafio.

Com o intuito de estimular avanços na área, a Avaliação Crítica de Técnicas de Predição de Estrutura de Proteínas (CASP – *Critical Assessment of Techniques for Protein Structure Prediction*) é realizada bienalmente. O principal objetivo dessa avaliação é verificar a qualidade e eficiência dos métodos de predição de estruturas existentes. São disponibilizadas diversas sequências proteicas alvo e ao final uma classificação é divulgada com o desempenho dos grupos participantes (Moult, Fidelis et al. 2014).

CASP possui duas áreas de competição: predição baseada em estrutura e primeiros princípios. Já foi mostrado que as predições de estrutura baseadas em homologia tinham grande destaque na produção de modelos razoáveis nos casos em que se possuía uma estrutura homóloga disponível (Moult, Hubbard et al. 1999), o que não é o caso da *cyspep*. Nos últimos anos, a metodologia de predição *de novo* apresentou-se bem classificada no CASP, tanto que a mesma se mostrou como metodologia provavelmente mais promissora (Moult 2005), em se enquadram os servidores *QUARK* (Xu and Zhang 2012) e *ROBETTA* (Chivian, Kim et al. 2005).

Experimentos CASP também demonstraram vantagens significativas de abordagens compostas na previsão da estrutura de proteínas, que combinam técnicas como *threading*, métodos *de novo* e refinamento estrutural. O servidor *I-Tasser* (Roy, Kucukural et al. 2010) é um exemplo de abordagem composta e foi classificado como o melhor método para a previsão da estrutura de proteínas automatizado no CASP no anos de 2007 e 2009 (Battey, Kopp et al. 2007, Cozzetto, Kryshtafovych et al. 2009). Uma análise da qualidade no alinhamento de sequências para a predição de estruturas demonstrou uma vantagem dominante de algoritmos de alinhamentos baseados em perfil-perfil (Yan, Xu et al. 2013), o mesmo utilizado pelo servidor *MUSTER* (Wu and Zhang 2008).

Em relação à estrutura tridimensional de proteínas, é cada vez mais reconhecido que as relações entre estrutura e função nem sempre são simples, muitas dobras de proteínas/famílias são conhecidas por serem funcionalmente promíscuas (Roy, Srinivasan et al. 2009), e enovelamentos diferentes podem desempenhar a mesma função (Bork, Sander et al. 1993).

De tal modo, não podemos estipular uma relação estrutura-função para a *cyspep* no momento. Para isso são necessárias investigações experimentais para confirmar alguns padrões estruturais da *cyspep*, como o dicroísmo circular para quantificar o

conteúdo de estrutura secundária em diferentes condições fisiológicas, e espectrometria de massas para verificar possíveis ligações dissulfeto.

A combinação de métodos experimentais e *in silico* tem demonstrado ser de grande valia. O hibridismo da espectrometria de massas com a modelagem comparativa (Kamal, Benchaar et al. 2007) e métodos *de novo* (Nili, Mukherjee et al. 2012) foram empregados na modelagem estrutural da proteína cofilina e de seis proteínas ligadoras do fator de crescimento similar a insulina (IGFBPs), respectivamente. O dicroísmo circular também foi utilizado para validar o modelo criado computacionalmente, como um modelo de peptidase CHAP_κ gerado pelo servidor *I-Tasser* que corroborou com os resultados do dicroísmo circular (Fenton, Cooney et al. 2011).

6 CONCLUSÕES

- A associação do uso de servidores *online* que utilizam as abordagens de modelagem comparativa, *threading* e métodos *de novo* para predição tridimensional de proteínas foi favorável para a obtenção de 72 modelos tridimensionais da *cyspep*.
- Das três abordagens para predição da *cyspep*, a modelagem comparativa apresentou os resultados menos satisfatórios, com apenas dois modelos, não concordantes com a predição de estrutura secundária consenso.
- A abordagem de dinâmica molecular clássica aplicada neste trabalho foi pertinente para validação dos 10 modelos tridimensionais propostos para *cyspep* que demonstraram concordância com a predição de estrutura secundária consenso.
- Nossos resultados sugerem que *cyspep* pode adotar uma estruturação composta principalmente de folha- β . Com o uso dos métodos de predição de estrutura seguidos de simulações de dinâmica molecular, sob a influência de diferentes campos de força, foi possível verificar a estabilidade do elemento folha- β , enquanto o teor helicoidal foi rapidamente interrompido.
- Não obstante a estabilidade das estruturas em folha- β após simulações de dinâmica molecular, não foi possível observar padrões de dobras consenso entre diferentes modelos.
- Embora a sequência linear da *cyspep* possua nove resíduos de cisteína, não houve consenso quanto à predição de ligações dissulfeto.

7 REFERÊNCIAS BIBLIOGRÁFICAS

(1990). "Control of the leishmaniasis. Report of a WHO Expert Committee." World Health Organ Tech Rep Ser **793**: 1-158.

Alder, B. and T. Wainwright (1957). "Phase transition for a hard sphere system." The Journal of Chemical Physics **27**(5): 1208-1209.

Alder, B. J. and T. Wainwright (1959). "Studies in molecular dynamics. I. General method." The Journal of Chemical Physics **31**(2): 459-466.

Alexander, J., G. H. Coombs and J. C. Mottram (1998). "Leishmania mexicana cysteine proteinase-deficient mutants have attenuated virulence for mice and potentiate a Th1 response." J Immunol **161**(12): 6794-6801.

Allen, M. P. and M. R. Wilson (1989). "Computer simulation of liquid crystals." J Comput Aided Mol Des **3**(4): 335-353.

Aloy, P., V. Companys, J. Vendrell, F. X. Aviles, L. D. Fricker, M. Coll and F. X. Gomis-Ruth (2001). "The crystal structure of the inhibitor-complexed carboxypeptidase D domain II and the modeling of regulatory carboxypeptidases." J Biol Chem **276**(19): 16177-16184.

Alvar, J., P. Aparicio, A. Aseffa, M. Den Boer, C. Canavate, J. P. Dedet, L. Gradoni, R. Ter Horst, R. Lopez-Velez and J. Moreno (2008). "The relationship between leishmaniasis and AIDS: the second 10 years." Clin Microbiol Rev **21**(2): 334-359, table of contents.

Alves, C. R., T. C. Benevolo-De-Andrade, J. L. Alves and C. Pirmez (2004). "Th1 and Th2 immunological profile induced by cysteine proteinase in murine leishmaniasis." Parasite Immunology **26**(3): 127-135.

Alves, C. R., S. Corte-Real, S. C. Bourguignon, C. S. Chaves and E. M. Saraiva (2005). "Leishmania amazonensis: early proteinase activities during promastigote-amastigote differentiation in vitro." Exp Parasitol **109**(1): 38-48.

Alves, C. R., S. Corte-Real, S. C. Bourguignon, C. S. Chaves and E. M. B. Saraiva (2005). "Leishmania amazonensis: early proteinase activities during promastigote-amastigote differentiation in vitro." Experimental Parasitology **109**(1): 38-48.

Alves, C. R., L. J. O. Figueiredo, F. A. L. O. Saraiva and S. G. De Simone (2001). "Molecular modeling study on a Leishmania cysteine proteinase." Journal of Molecular Structure: THEOCHEM **539**(1-3): 289-295.

Alves, C. R., L. C. Pontes de Carvalho, A. L. Souza and S. G. De Simone (2001). "A strategy for the identification of T-cell epitopes on Leishmania cysteine proteinases." Cytobios **104**(405): 33-41.

An, H. J., Y. J. Kim, D. H. Song, B. S. Park, H. M. Kim, J. D. Lee, S. G. Paik, J. O. Lee and H. Lee (2011). "Crystallographic and mutational analysis of the CD40-CD154 complex and its implications for receptor activation." J Biol Chem **286**(13): 11226-11235.

- Antuch, W., P. Guntert and K. Wuthrich (1996). "Ancestral beta gamma-crystallin precursor structure in a yeast killer toxin." Nat Struct Biol **3**(8): 662-665.
- Baker, D. and A. Sali (2001). "Protein structure prediction and structural genomics." Science **294**(5540): 93-96.
- Banuls, A. L., M. Hide and F. Prugnolle (2007). "Leishmania and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans." Adv Parasitol **64**: 1-109.
- Barreto-de-Souza, V., T. Xavier Medeiros, M. C. Machado Motta, D. C. Bou-Habib and E. M. Saraiva (2008). "HIV-1 infection and HIV-1 Tat protein permit the survival and replication of a non-pathogenic trypanosomatid in macrophages through TGF-beta1 production." Microbes Infect **10**(6): 642-649.
- Barrett, A. J. (1994). "Classification of peptidases." Methods Enzymol **244**: 1-15.
- Barrientos, L. G., J. M. Louis, I. Botos, T. Mori, Z. Han, B. R. O'Keefe, M. R. Boyd, A. Wlodawer and A. M. Gronenborn (2002). "The domain-swapped dimer of cyanovirin-N is in a metastable folded state: reconciliation of X-ray and NMR structures." Structure **10**(5): 673-686.
- Barrientos, L. G., J. M. Louis, D. M. Ratner, P. H. Seeberger and A. M. Gronenborn (2003). "Solution structure of a circular-permuted variant of the potent HIV-inactivating protein cyanovirin-N: structural basis for protein stability and oligosaccharide interaction." J Mol Biol **325**(1): 211-223.
- Bates, P. A. and M. E. Rogers (2004). "New insights into the developmental biology and transmission mechanisms of Leishmania." Curr Mol Med **4**(6): 601-609.
- Batthey, J. N., J. Kopp, L. Bordoli, R. J. Read, N. D. Clarke and T. Schwede (2007). "Automated server predictions in CASP7." Proteins **69 Suppl 8**: 68-82.
- Becherel, O. J., B. Jakob, A. L. Cherry, N. Gueven, M. Fusser, A. W. Kijas, C. Peng, S. Katyal, P. J. McKinnon, J. Chen, B. Epe, S. J. Smerdon, G. Taucher-Scholz and M. F. Lavin (2010). "CK2 phosphorylation-dependent interaction between aprataxin and MDC1 in the DNA damage response." Nucleic Acids Res **38**(5): 1489-1503.
- Beringer, D. X. and L. M. Kroon-Batenburg (2013). "The structure of the Fnl-EGF-like tandem domain of coagulation factor XII solved using SIRAS." Acta Crystallogr Sect F Struct Biol Cryst Commun **69**(Pt 2): 94-102.
- Besteiro, S., R. A. Williams, G. H. Coombs and J. C. Mottram (2007). "Protein turnover and differentiation in Leishmania." Int J Parasitol **37**(10): 1063-1075.
- Bond, C. S. and A. W. Schuttelkopf (2009). "ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments." Acta Crystallogr D Biol Crystallogr **65**(Pt 5): 510-512.
- Bonneau, R., J. Tsai, I. Ruczinski, D. Chivian, C. Rohl, C. E. Strauss and D. Baker (2001). "Rosetta in CASP4: progress in ab initio protein structure prediction." Proteins Suppl 5: 119-126.

- Bork, P., C. Sander and A. Valencia (1993). "Convergent evolution of similar enzymatic function on different protein folds: the hexokinase, ribokinase, and galactokinase families of sugar kinases." Protein Sci **2**(1): 31-40.
- Bosanac, I., H. R. Maun, S. J. Scales, X. Wen, A. Lingel, J. F. Bazan, F. J. de Sauvage, S. G. Hymowitz and R. A. Lazarus (2009). "The structure of SHH in complex with HHIP reveals a recognition role for the Shh pseudo active site in signaling." Nat Struct Mol Biol **16**(7): 691-697.
- Bourhis, E., W. Wang, C. Tam, J. Hwang, Y. Zhang, D. Spittler, O. W. Huang, Y. Gong, A. Estevez, I. Zilberleyb, L. Rouge, C. Chiu, Y. Wu, M. Costa, R. N. Hannoush, Y. Franke and A. G. Cochran (2011). "Wnt antagonists bind through a short peptide to the first beta-propeller domain of LRP5/6." Structure **19**(10): 1433-1442.
- Cai, W., J. Li and S. Yip (2012). *Comprehensive nuclear materials*, Elsevier, Amsterdam.
- Ceroni, A., A. Passerini, A. Vullo and P. Frasconi (2006). "DISULFIND: a disulfide bonding state and cysteine connectivity prediction server." Nucleic Acids Res **34**(Web Server issue): W177-181.
- Cheng, J., A. Z. Randall, M. J. Sweredoski and P. Baldi (2005). "SCRATCH: a protein structure and structural feature prediction server." Nucleic Acids Res **33**(Web Server issue): W72-76.
- Chivian, D., D. E. Kim, L. Malmstrom, J. Schonbrun, C. A. Rohl and D. Baker (2005). "Prediction of CASP6 structures using automated Robetta protocols." Proteins **61 Suppl 7**: 157-166.
- Cohen-Freue, G., T. R. Holzer, J. D. Forney and W. R. McMaster (2007). "Global gene expression in Leishmania." Int J Parasitol **37**(10): 1077-1086.
- Cole, C., J. D. Barber and G. J. Barton (2008). "The Jpred 3 secondary structure prediction server." Nucleic Acids Res **36**(Web Server issue): W197-201.
- Collins, D. S., E. R. Unanue and C. V. Harding (1991). "Reduction of disulfide bonds within lysosomes is a key step in antigen processing." J Immunol **147**(12): 4054-4059.
- Combet, C., C. Blanchet, C. Geourjon and G. Deleage (2000). "NPS@: network protein sequence analysis." Trends Biochem Sci **25**(3): 147-150.
- Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman (1995). "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." Journal of the American Chemical Society **117**(19): 5179-5197.
- Cozzetto, D., A. Kryshchak, K. Fidelis, J. Moutl, B. Rost and A. Tramontano (2009). "Evaluation of template-based models in CASP8 with standard measures." Proteins **77 Suppl 9**: 18-28.
- Cuff, J. A. and G. J. Barton (2000). "Application of multiple sequence alignment profiles to improve protein secondary structure prediction." Proteins **40**(3): 502-511.

- da Silva, V. B. (2007). "Modelagem molecular de proteínas-alvo por homologia estrutural." Revista eletrônica de Farmácia **4**(1).
- Dahiyat, B. I. and S. L. Mayo (1997). "De novo protein design: fully automated sequence selection." Science **278**(5335): 82-87.
- Dai, L. and Y. Zhou (2011). "Characterizing the existing and potential structural space of proteins by large-scale multiple loop permutations." J Mol Biol **408**(3): 585-595.
- Dash, C., A. Kulkarni, B. Dunn and M. Rao (2003). "Aspartic peptidase inhibitors: implications in drug development." Crit Rev Biochem Mol Biol **38**(2): 89-119.
- Daura, X., K. Gademann, B. Jaun, D. Seebach, W. F. van Gunsteren and A. E. Mark (1999). "Peptide Folding: When Simulation Meets Experiment." Angewandte Chemie International Edition **38**(1-2): 236-240.
- De Souza Leao, S., T. Lang, E. Prina, R. Hellio and J. C. Antoine (1995). "Intracellular Leishmania amazonensis amastigotes internalize and degrade MHC class II molecules of their host cells." J Cell Sci **108 (Pt 10)**: 3219-3231.
- Deleage, G. and B. Roux (1987). "An algorithm for protein secondary structure prediction based on class prediction." Protein Eng **1**(4): 289-294.
- Durrant, J. D. and J. A. McCammon (2011). "Molecular dynamics simulations and drug discovery." BMC biology **9**(1): 71.
- Eder, J., U. Hommel, F. Cumin, B. Martoglio and B. Gerhartz (2007). "Aspartic proteases in drug discovery." Curr Pharm Des **13**(3): 271-285.
- Eisenhaber, F., P. Lijnzaad, P. Argos, C. Sander and M. Scharf (1995). "The double cubic lattice method: Efficient approaches to numerical integration of surface area and volume and to dot surface contouring of molecular assemblies." Journal of Computational Chemistry **16**(3): 273-284.
- Fairlie, W. D., P. G. Stanton and M. T. W. Hearn (1996). "Contribution of Specific Disulphide Bonds to Two Epitopes of Thyrotropin β -Subunit Associated with Receptor Recognition." European Journal of Biochemistry **240**(3): 622-627.
- Fenton, M., J. C. Cooney, R. P. Ross, R. D. Sleator, O. McAuliffe, J. O'Mahony and A. Coffey (2011). "In silico modeling of the staphylococcal bacteriophage-derived peptidase CHAP(K)." Bacteriophage **1**(4): 198-206.
- Ferre, F. and P. Clote (2005). "DiANNA: a web server for disulfide connectivity prediction." Nucleic Acids Res **33**(Web Server issue): W230-232.
- Ferre, F. and P. Clote (2006). "DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification." Nucleic Acids Res **34**(Web Server issue): W182-185.
- Frenkel, D. and B. Smit (1996). Understanding molecular simulation : from algorithms to applications. San Diego, Academic Press.

Frishman, D. and P. Argos (1996). "Incorporation of non-local interactions in protein secondary structure prediction from the amino acid sequence." Protein Eng **9**(2): 133-142.

Fromme, R., Z. Katiliene, B. Giomarelli, F. Bogani, J. Mc Mahon, T. Mori, P. Fromme and G. Ghirlanda (2007). "A monovalent mutant of cyanovirin-N provides insight into the role of multiple interactions with gp120 for antiviral activity." Biochemistry **46**(32): 9199-9207.

Garnier, J., J. F. Gibrat and B. Robson (1996). "GOR method for predicting protein secondary structure from amino acid sequence." Methods Enzymol **266**: 540-553.

Garnier, J., D. J. Osguthorpe and B. Robson (1978). "Analysis of the accuracy and implications of simple methods for predicting the secondary structure of globular proteins." J Mol Biol **120**(1): 97-120.

Geourjon, C. and G. Deleage (1995). "SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments." Comput Appl Biosci **11**(6): 681-684.

Gibas, C. and P. Jambeck (2001). Developing bioinformatics computer skills. Beijing Cambridge, O'Reilly.

Gibrat, J. F., J. Garnier and B. Robson (1987). "Further developments of protein secondary structure prediction using information theory. New parameters and consideration of residue pairs." J Mol Biol **198**(3): 425-443.

Gorman, J. J., T. P. Wallis and J. J. Pitt (2002). "Protein disulfide bond determination by mass spectrometry." Mass Spectrom Rev **21**(3): 183-216.

Griessl, M. H., B. Schmid, K. Kassler, C. Braunsmann, R. Ritter, B. Barlag, Y. D. Stierhof, K. U. Sturm, C. Danzer, C. Wagner, T. E. Schaffer, H. Sticht, M. Hensel and Y. A. Muller (2013). "Structural insight into the giant Ca(2)(+)-binding adhesin SiiE: implications for the adhesion of Salmonella enterica to polarized epithelial cells." Structure **21**(5): 741-752.

Guermeur, Y., C. Geourjon, P. Gallinari and G. Deleage (1999). "Improved performance in protein secondary structure prediction by inhomogeneous score combination." Bioinformatics **15**(5): 413-421.

Haque, M. A., J. W. Hawes and J. S. Blum (2001). "Cysteinylation of MHC class II ligands: peptide endocytosis and reduction within APC influences T cell recognition." J Immunol **166**(7): 4543-4551.

Hatzopoulos, G. N., M. C. Erat, E. Cutts, K. B. Rogala, L. M. Slater, P. J. Stansfeld and I. Vakonakis (2013). "Structural analysis of the G-box domain of the microcephaly protein CPAP suggests a role in centriole architecture." Structure **21**(11): 2069-2077.

Hegyí, H. and M. Gerstein (1999). "The relationship between protein structure and function: a comprehensive survey with application to the yeast genome." J Mol Biol **288**(1): 147-164.

Hermoso, J., D. Pignol, B. Kerfelec, I. Crenon, C. Chapus and J. C. Fontecilla-Camps (1996). "Lipase activation by nonionic detergents. The crystal structure of the porcine

- lipase-colipase-tetraethylene glycol monoethyl ether complex." J Biol Chem **271**(30): 18007-18016.
- Hide, M. and A. L. Banuls (2008). "Polymorphisms of cpb multicopy genes in the Leishmania (Leishmania) donovani complex." Transactions of the Royal Society of Tropical Medicine and Hygiene **102**(2): 105-106.
- Hogg, P. J. (2003). "Disulfide bonds as switches for protein function." Trends Biochem Sci **28**(4): 210-214.
- Hsiao, C. H., C. Yao, P. Storlie, J. E. Donelson and M. E. Wilson (2008). "The major surface protease (MSP or GP63) in the intracellular amastigote stage of Leishmania chagasi." Mol Biochem Parasitol **157**(2): 148-159.
- Huizinga, E. G., A. Schouten, T. M. Connolly, J. Kroon, J. J. Sixma and P. Gros (2001). "The structure of leech anti-platelet protein, an inhibitor of haemostasis." Acta Crystallogr D Biol Crystallogr **57**(Pt 8): 1071-1078.
- Iwu, M. M., J. E. Jackson and B. G. Schuster (1994). "Medicinal plants in the fight against leishmaniasis." Parasitol Today **10**(2): 65-68.
- Jaffe, C. L. and D. M. Dwyer (2003). "Extracellular release of the surface metalloprotease, gp63, from Leishmania and insect trypanosomatids." Parasitol Res **91**(3): 229-237.
- Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.
- Joosten, R. P., T. A. te Beek, E. Krieger, M. L. Hekkelman, R. W. Hooft, R. Schneider, C. Sander and G. Vriend (2011). "A series of PDB related databases for everyday needs." Nucleic Acids Res **39**(Database issue): D411-419.
- Kabsch, W. and C. Sander (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features." Biopolymers **22**(12): 2577-2637.
- Kamal, J. K., S. A. Benchaar, K. Takamoto, E. Reisler and M. R. Chance (2007). "Three-dimensional structure of cofilin bound to monomeric actin derived by structural mass spectrometry data." Proc Natl Acad Sci U S A **104**(19): 7910-7915.
- Karplus, K., R. Karchin, J. Draper, J. Casper, Y. Mandel-Gutfreund, M. Diekhans and R. Hughey (2003). "Combining local-structure, fold-recognition, and new fold methods for protein structure prediction." Proteins: Structure, Function, and Bioinformatics **53**(S6): 491-496.
- Karplus, M. and J. A. McCammon (2002). "Molecular dynamics simulations of biomolecules." Nature Structural & Molecular Biology **9**(9): 646-652.
- Kato, H., E. A. Gomez, A. G. Caceres, H. Uezato, T. Mimori and Y. Hashiguchi (2010). "Molecular epidemiology for vector research on leishmaniasis." Int J Environ Res Public Health **7**(3): 814-826.

Kedzierski, L., A. Sakthianandeswaren, J. M. Curtis, P. C. Andrews, P. C. Junk and K. Kedzierska (2009). "Leishmaniasis: current treatment and prospects for new drugs and vaccines." Curr Med Chem **16**(5): 599-614.

King, R. D., M. Saqi, R. Sayle and M. J. Sternberg (1997). "DSC: public domain protein secondary structure predication." Comput Appl Biosci **13**(4): 473-474.

King, R. D. and M. J. Sternberg (1996). "Identification and application of the concepts important for accurate and reliable protein secondary structure prediction." Protein Sci **5**(11): 2298-2310.

Klose, D. P., B. A. Wallace and R. W. Janes (2010). "2Struc: the secondary structure server." Bioinformatics **26**(20): 2624-2625.

Kolkman, J. A., W. P. Stemmer and S. Govindarajan (2003). Combinatorial libraries of monomer domains, US Patent 20,030,157,561.

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

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

Kumar, A., D. D. Jagtap, S. D. Mahale and M. Kumar (2010). "Crystal structure of prostate secretory protein PSP94 shows an edge-to-edge association of two monomers to form a homodimer." J Mol Biol **397**(4): 947-956.

Lainson, R. and J. J. Shaw (1987). Evolution, Classification and Geographical Distribution, Academic Press.

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

Levi, A. and H. Suhl (1979). "Quantum theory of atom-surface scattering: Debye-Waller factor." Surface Science **88**(1): 221-254.

Levin, J. M., B. Robson and J. Garnier (1986). "An algorithm for secondary structure determination in proteins based on sequence similarity." FEBS Lett **205**(2): 303-308.

Levinthal, C. (1969). "Mössbauer Spectroscopy in Biological Systems, Proceedings of a Meeting Held at Allerton House, Monticello, Illinois." P. Debrunner, JCM Tsibris, and E. M nck.

Li, P., M. A. Haque and J. S. Blum (2002). "Role of disulfide bonds in regulating antigen processing and epitope selection." J Immunol **169**(5): 2444-2450.

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

Lin, H. H. and L. Y. Tseng (2010). "DBCP: a web server for disulfide bonding connectivity pattern prediction without the prior knowledge of the bonding state of cysteines." Nucleic Acids Res **38**(Web Server issue): W503-507.

- Lin, K., V. A. Simossis, W. R. Taylor and J. Heringa (2005). "A simple and fast secondary structure prediction method using hidden neural networks." Bioinformatics **21**(2): 152-159.
- Lindorff-Larsen, K., S. Piana, K. Palmo, P. Maragakis, J. L. Klepeis, R. O. Dror and D. E. Shaw (2010). "Improved side-chain torsion potentials for the Amber ff99SB protein force field." Proteins **78**(8): 1950-1958.
- Liu, C., T. S. Walter, P. Huang, S. Zhang, X. Zhu, Y. Wu, L. R. Wedderburn, P. Tang, R. J. Owens, D. I. Stuart, J. Ren and B. Gao (2010). "Structural and functional insights of RANKL-RANK interaction and signaling." J Immunol **184**(12): 6910-6919.
- Lo Conte, L., B. Ailey, T. J. Hubbard, S. E. Brenner, A. G. Murzin and C. Chothia (2000). "SCOP: a structural classification of proteins database." Nucleic Acids Res **28**(1): 257-259.
- Luo, Y., E. A. Frey, R. A. Pfuetzner, A. L. Creagh, D. G. Knochel, C. A. Haynes, B. B. Finlay and N. C. Strynadka (2000). "Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex." Nature **405**(6790): 1073-1077.
- Mackerell, A. D., Jr., M. Feig and C. L. Brooks, 3rd (2004). "Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations." J Comput Chem **25**(11): 1400-1415.
- Mahoney, M. W. and W. L. Jorgensen (2000). "A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions." The Journal of Chemical Physics **112**(20): 8910-8922.
- Marin-Villa, M., D. A. Vargas-Inchaustegui, S. P. Chaves, A. J. Tempone, J. M. F. Dutra, M. J. Soares, T. Ueda-Nakamura, S. C. F. Mendonca, B. Rossi-Bergmann, L. Soong and Y. M. Traub-Cseko (2008). "The C-terminal extension of Leishmania pifanoi amastigote-specific cysteine proteinase Lpcys2: A putative function in macrophage infection." Molecular and Biochemical Parasitology **162**(1): 52-59.
- Martí-Renom, M. A., A. C. Stuart, A. Fiser, R. Sánchez, F. Melo and A. Šali (2000). "Comparative protein structure modeling of genes and genomes." Annual review of biophysics and biomolecular structure **29**(1): 291-325.
- Matlashewski, G. (2001). "Leishmania infection and virulence." Med Microbiol Immunol **190**(1-2): 37-42.
- Mauel, J. (1990). "Macrophage-parasite interactions in Leishmania infections." J Leukoc Biol **47**(2): 187-193.
- McManus, A. M., K. J. Nielsen, J. P. Marcus, S. J. Harrison, J. L. Green, J. M. Manners and D. J. Craik (1999). "MiAMP1, a novel protein from Macadamia integrifolia adopts a Greek key beta-barrel fold unique amongst plant antimicrobial proteins." J Mol Biol **293**(3): 629-638.
- Milburn, D., R. A. Laskowski and J. M. Thornton (1998). "Sequences annotated by structure: a tool to facilitate the use of structural information in sequence analysis." Protein Eng **11**(10): 855-859.

- Mizuguchi, K., C. M. Deane, T. L. Blundell and J. P. Overington (1998). "HOMSTRAD: a database of protein structure alignments for homologous families." Protein Sci **7**(11): 2469-2471.
- Mottram, J. C., G. H. Coombs and J. Alexander (2004). "Cysteine peptidases as virulence factors of Leishmania." Current Opinion in Microbiology **7**(4): 375-381.
- Moult, J. (2005). "A decade of CASP: progress, bottlenecks and prognosis in protein structure prediction." Curr Opin Struct Biol **15**(3): 285-289.
- Moult, J., K. Fidelis, A. Kryshtafovych, T. Schwede and A. Tramontano (2014). "Critical assessment of methods of protein structure prediction (CASP)--round x." Proteins **82 Suppl 2**: 1-6.
- Moult, J., T. Hubbard, K. Fidelis and J. T. Pedersen (1999). "Critical assessment of methods of protein structure prediction (CASP): round III." Proteins Suppl 3: 2-6.
- Moza, B., A. K. Varma, R. A. Buonpane, P. Zhu, C. A. Herfst, M. J. Nicholson, A. K. Wilbuer, N. P. Seth, K. W. Wucherpfennig, J. K. McCormick, D. M. Kranz and E. J. Sundberg (2007). "Structural basis of T-cell specificity and activation by the bacterial superantigen TSST-1." EMBO J **26**(4): 1187-1197.
- Munday, J. C., K. McLuskey, E. Brown, G. H. Coombs and J. C. Mottram (2011). "Oligopeptidase B deficient mutants of Leishmania major." Mol Biochem Parasitol **175**(1): 49-57.
- Murray, H. W., J. D. Berman, C. R. Davies and N. G. Saravia (2005). "Advances in leishmaniasis." Lancet **366**(9496): 1561-1577.
- Muskus, C. E. and M. Marin Villa (2002). "[Metacyclogenesis: a basic process in the biology of Leishmania]." Biomedica **22**(2): 167-177.
- Nesic, D., Y. Hsu and C. E. Stebbins (2004). "Assembly and function of a bacterial genotoxin." Nature **429**(6990): 429-433.
- Nili, M., A. Mukherjee, U. Shinde, L. David and P. Rotwein (2012). "Defining the disulfide bonds of insulin-like growth factor-binding protein-5 by tandem mass spectrometry with electron transfer dissociation and collision-induced dissociation." J Biol Chem **287**(2): 1510-1519.
- Notredame, C., D. G. Higgins and J. Heringa (2000). "T-Coffee: A novel method for fast and accurate multiple sequence alignment." J Mol Biol **302**(1): 205-217.
- Olivier, M., D. J. Gregory and G. Forget (2005). "Subversion mechanisms by which Leishmania parasites can escape the host immune response: a signaling point of view." Clin Microbiol Rev **18**(2): 293-305.
- Oostenbrink, C., A. Villa, A. E. Mark and W. F. van Gunsteren (2004). "A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6." J Comput Chem **25**(13): 1656-1676.
- Ouali, M. and R. D. King (2000). "Cascaded multiple classifiers for secondary structure prediction." Protein Sci **9**(6): 1162-1176.

- Pace, C. N. and J. M. Scholtz (1998). "A helix propensity scale based on experimental studies of peptides and proteins." Biophys J **75**(1): 422-427.
- Pain, R. H., B. Hames and D. M. Glover (1994). Mechanisms of protein folding, IRL Press Oxford.
- Peng, W. C., W. de Lau, F. Forneris, J. C. Granneman, M. Huch, H. Clevers and P. Gros (2013). "Structure of stem cell growth factor R-spondin 1 in complex with the ectodomain of its receptor LGR5." Cell Rep **3**(6): 1885-1892.
- Pereira, B. A. S., F. S. Silva, K. M. Rebello, M. Marin-Villa, Y. M. Traub-Cseko, T. C. B. Andrade, A. L. Bertho, E. R. Caffarena and C. R. Alves (2011). "In silico predicted epitopes from the COOH-terminal extension of cysteine proteinase B inducing distinct immune responses during Leishmania (Leishmania) amazonensis experimental murine infection." Bmc Immunology **12**.
- Perteguer, M. J., P. Gomez-Puertas, C. Canavate, F. Dagger, T. Garate and E. Valdivieso (2013). "Ddi1-like protein from Leishmania major is an active aspartyl proteinase." Cell Stress Chaperones **18**(2): 171-181.
- Pollastri, G., D. Przybylski, B. Rost and P. Baldi (2002). "Improving the prediction of protein secondary structure in three and eight classes using recurrent neural networks and profiles." Proteins **47**(2): 228-235.
- Pronk, S., S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl (2013). "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit." Bioinformatics **29**(7): 845-854.
- Proudfoot, L., C. A. O'Donnell and F. Y. Liew (1995). "Glycoinositolphospholipids of Leishmania major inhibit nitric oxide synthesis and reduce leishmanicidal activity in murine macrophages." Eur J Immunol **25**(3): 745-750.
- Puente, X. S., L. M. Sanchez, C. M. Overall and C. Lopez-Otin (2003). "Human and mouse proteases: a comparative genomic approach." Nat Rev Genet **4**(7): 544-558.
- Rahman, A. (1964). "Correlations in the motion of atoms in liquid argon." Physical Review **136**(2A): A405.
- Rawlings, N. D. and A. J. Barrett (1993). "Evolutionary families of peptidases." Biochem J **290** (Pt 1): 205-218.
- Rawlings, N. D. and A. J. Barrett (2000). "MEROPS: the peptidase database." Nucleic Acids Res **28**(1): 323-325.
- Rawlings, N. D., M. Waller, A. J. Barrett and A. Bateman (2014). "MEROPS: the database of proteolytic enzymes, their substrates and inhibitors." Nucleic Acids Res **42**(Database issue): D503-509.
- Rioux, J. A., G. Lanotte, R. Maazoun and N. Pasteur (1985). "[Electrophoresis of enzymes of the genus Leishmania Ross 1903]." Parassitologia **27**(1-2): 141-156.

- Rioux, J. A., G. Lanotte, E. Serres, F. Pratlong, P. Bastien and J. Perieres (1990). "Taxonomy of Leishmania. Use of isoenzymes. Suggestions for a new classification." Ann Parasitol Hum Comp **65**(3): 111-125.
- Robertson, C. D. and G. H. Coombs (1994). "Multiple High-Activity Cysteine Proteases of Leishmania-Mexicana Are Encoded by the Imcpb Gene Array." Microbiology-Sgm **140**: 417-424.
- Robertson, C. D. and G. H. Coombs (1994). "Multiple high activity cysteine proteases of Leishmania mexicana are encoded by the Imcpb gene array." Microbiology **140 (Pt 2)**: 417-424.
- Rost, B. and C. Sander (1993). "Improved prediction of protein secondary structure by use of sequence profiles and neural networks." Proc Natl Acad Sci U S A **90**(16): 7558-7562.
- Rost, B. and C. Sander (1993). "Prediction of protein secondary structure at better than 70% accuracy." J Mol Biol **232**(2): 584-599.
- Rost, B. and C. Sander (1994). "Combining evolutionary information and neural networks to predict protein secondary structure." Proteins **19**(1): 55-72.
- Roy, A., A. Kucukural and Y. Zhang (2010). "I-TASSER: a unified platform for automated protein structure and function prediction." Nat Protoc **5**(4): 725-738.
- Roy, A., N. Srinivasan and V. S. Gowri (2009). "Molecular and structural basis of drift in the functions of closely-related homologous enzyme domains: implications for function annotation based on homology searches and structural genomics." In Silico Biol **9**(1-2): S41-55.
- Salamov, A. A. and V. V. Solovyev (1995). "Prediction of protein secondary structure by combining nearest-neighbor algorithms and multiple sequence alignments." J Mol Biol **247**(1): 11-15.
- Sali, A., L. Potterton, F. Yuan, H. van Vlijmen and M. Karplus (1995). "Evaluation of comparative protein modeling by MODELLER." Proteins **23**(3): 318-326.
- Sanchez, R. and A. Sali (1999). "Comparative protein structure modeling in genomics." Journal of Computational Physics **151**(1): 388-401.
- Sanders, D. A., J. R. Walker, T. Skarina and A. Savchenko (2005). "The X-ray crystal structure of PA3566 from Pseudomonas aureginosa at 1.8 Å resolution." Proteins **61**(1): 209-212.
- Santos Filho, O. A. and R. B. d. Alencastro (2003). "Modelagem de proteínas por homologia." Química Nova **26**: 253-259.
- Saúde, F. N. d. and C. N. d. E. U. d. V. Epidemiológica (2007). Manual de controle da leishmaniose tegumentar americana, Ministério da Saúde.
- Schlick, T. (2010). Molecular Modeling and Simulation: An Interdisciplinary Guide: An Interdisciplinary Guide, Springer.

Selzer, P. M., S. Pingel, I. Hsieh, B. Ugele, V. J. Chan, J. C. Engel, M. Bogyo, D. G. Russell, J. A. Sakanari and J. H. McKerrow (1999). "Cysteine protease inhibitors as chemotherapy: Lessons from a parasite target." Proceedings of the National Academy of Sciences of the United States of America **96**(20): 11015-11022.

Sereno, D., E. Guilvard, S. Maquaire, M. Cavaleyra, P. Holzmuller, A. Ouaisi and J. L. Lemesre (2001). "Experimental studies on the evolution of antimony-resistant phenotype during the in vitro life cycle of *Leishmania infantum*: implications for the spread of chemoresistance in endemic areas." Acta Trop **80**(3): 195-205.

Shi, J., T. L. Blundell and K. Mizuguchi (2001). "FUGUE: sequence-structure homology recognition using environment-specific substitution tables and structure-dependent gap penalties." J Mol Biol **310**(1): 243-257.

Silva-Almeida, M., B. A. Pereira, M. L. Ribeiro-Guimaraes and C. R. Alves (2012). "Proteinases as virulence factors in *Leishmania* spp. infection in mammals." Parasit Vectors **5**: 160.

Silva-Almeida, M., B. A. S. Pereira, M. L. Ribeiro-Guimaraes and C. R. Alves (2012). "Proteinases as virulence factors in *Leishmania* spp. infection in mammals." Parasites & Vectors **5**.

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

Simossis, V. A. a. H., J. (2004). "Optimally segmented consensus secondary structure prediction." Bioinformatics.

Söding, J. (2005). "Protein homology detection by HMM–HMM comparison." Bioinformatics **21**(7): 951-960.

Soong, L., C. H. Chang, J. Sun, B. J. Longley, Jr., N. H. Ruddle, R. A. Flavell and D. McMahon-Pratt (1997). "Role of CD4+ T cells in pathogenesis associated with *Leishmania amazonensis* infection." J Immunol **158**(11): 5374-5383.

Souza-Silva, F., B. A. Pereira, L. C. Finkelstein, V. Zucolotto, E. R. Caffarena and C. R. Alves (2014). "Dynamic identification of H2 epitopes from *Leishmania* (*Leishmania*) *amazonensis* cysteine proteinase B with potential immune activity during murine infection." J Mol Recognit **27**(2): 98-105.

Spath, G. F., L. Epstein, B. Leader, S. M. Singer, H. A. Avila, S. J. Turco and S. M. Beverley (2000). "Lipophosphoglycan is a virulence factor distinct from related glycoconjugates in the protozoan parasite *Leishmania major*." Proc Natl Acad Sci U S A **97**(16): 9258-9263.

Sternberg, M. J. E. (1996). Protein structure prediction : a practical approach. Oxford ; New York, IRL Press at Oxford University Press.

Swenerton, R. K., G. M. Knudsen, M. Sajid, B. L. Kelly and J. H. McKerrow (2010). "*Leishmania* subtilisin is a maturase for the trypanothione reductase system and contributes to disease pathology." J Biol Chem **285**(41): 31120-31129.

Turk, B. (2006). "Targeting proteases: successes, failures and future prospects." Nat Rev Drug Discov **5**(9): 785-799.

- Valdivieso, E., A. Rangel, J. Moreno, J. M. Saugar, C. Canavate, J. Alvar and F. Dagger (2010). "Effects of HIV aspartyl-proteinase inhibitors on *Leishmania* sp." Exp Parasitol **126**(4): 557-563.
- Van de Waterbeemd, H., R. Carter, G. Grassy, H. Kubinyi, Y. Martin, M. Tute and P. Willett (1997). "Glossary of terms used in computational drug design (IUPAC Recommendations 1997)." Pure and applied chemistry **69**(5): 1146.
- Vannier-Santos, M. A., A. Martiny and W. de Souza (2002). "Cell biology of *Leishmania* spp.: invading and evading." Curr Pharm Des **8**(4): 297-318.
- Verli, H. (2014). "Bioinformática da Biologia à flexibilidade molecular." Porto Alegre.
- Wei, Y. T., J. and Floudas, A. (2011). "CONCORD: A consensus method for protein secondary structure prediction via Mixed Integer Linear Optimization." Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science **468**(2139): 831-850.
- World Health, O. (2010). "Control of the leishmaniasis." World Health Organ Tech Rep Ser(949): xii-xiii, 1-186, back cover.
- Wright, C. S. (1990). "2.2 A resolution structure analysis of two refined N-acetylneuraminyl-lactose--wheat germ agglutinin isolectin complexes." J Mol Biol **215**(4): 635-651.
- Wu, S. and Y. Zhang (2007). "LOMETS: a local meta-threading-server for protein structure prediction." Nucleic Acids Res **35**(10): 3375-3382.
- Wu, S. and Y. Zhang (2008). "MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information." Proteins **72**(2): 547-556.
- Xu, D. and Y. Zhang (2012). "Ab initio protein structure assembly using continuous structure fragments and optimized knowledge-based force field." Proteins **80**(7): 1715-1735.
- Xu, Y. and D. Xu (2000). "Protein threading using PROSPECT: design and evaluation." Proteins: Structure, Function, and Bioinformatics **40**(3): 343-354.
- Yan, R., D. Xu, J. Yang, S. Walker and Y. Zhang (2013). "A comparative assessment and analysis of 20 representative sequence alignment methods for protein structure prediction." Sci Rep **3**: 2619.
- Yang, Y., E. Faraggi, H. Zhao and Y. Zhou (2011). "Improving protein fold recognition and template-based modeling by employing probabilistic-based matching between predicted one-dimensional structural properties of query and corresponding native properties of templates." Bioinformatics **27**(15): 2076-2082.
- Zhan, C., Y. Patskovsky, Q. Yan, Z. Li, U. Ramagopal, H. Cheng, M. Brenowitz, X. Hui, S. G. Nathenson and S. C. Almo (2011). "Decoy strategies: the structure of TL1A:DcR3 complex." Structure **19**(2): 162-171.
- Zhang, T., Y. Maekawa, K. Yasutomo, H. Ishikawa, B. Fawzy Nashed, T. Dainichi, H. Hisaeda, T. Sakai, M. Kasai, T. Mizuochi, T. Asao, N. Katunuma and K. Himeno (2000).

"Pepstatin A-sensitive aspartic proteases in lysosome are involved in degradation of the invariant chain and antigen-processing in antigen presenting cells of mice infected with *Leishmania major*." Biochem Biophys Res Commun **276**(2): 693-701.

Zhang, X., L. Jin, Q. Fang, W. H. Hui and Z. H. Zhou (2010). "3.3 A cryo-EM structure of a nonenveloped virus reveals a priming mechanism for cell entry." Cell **141**(3): 472-482.

Zhao, C., B. Papadopoulou and M. J. Tremblay (2004). "Leishmania infantum promotes replication of HIV type 1 in human lymphoid tissue cultured ex vivo by inducing secretion of the proinflammatory cytokines TNF-alpha and IL-1 alpha." J Immunol **172**(5): 3086-3093.

Zhou, H. and Y. Zhou (2004). "Single-body residue-level knowledge-based energy score combined with sequence-profile and secondary structure information for fold recognition." Proteins: Structure, Function, and Bioinformatics **55**(4): 1005-1013.

Zhou, H. and Y. Zhou (2005). "Fold recognition by combining sequence profiles derived from evolution and from depth-dependent structural alignment of fragments." Proteins: Structure, Function, and Bioinformatics **58**(2): 321-328.

Zufferey, R., S. Allen, T. Barron, D. R. Sullivan, P. W. Denny, I. C. Almeida, D. F. Smith, S. J. Turco, M. A. Ferguson and S. M. Beverley (2003). "Ether phospholipids and glycosylinositolphospholipids are not required for amastigote virulence or for inhibition of macrophage activation by *Leishmania major*." J Biol Chem **278**(45): 44708-44718.

Zvelebil, M. J., G. J. Barton, W. R. Taylor and M. J. Sternberg (1987). "Prediction of protein secondary structure and active sites using the alignment of homologous sequences." J Mol Biol **195**(4): 957-961.

Zwanzig, R., A. Szabo and B. Bagchi (1992). "Levinthal's paradox." Proceedings of the National Academy of Sciences **89**(1): 20-22.

ANEXO A – ARQUIVOS DE ENTRADA PARA SIMULAÇÕES DE DINÂMICA MOLECULAR

```
define = -DFLEXIBLE      ;Define a flexibilidade/mobilidade do solvente, aqui livre
constraints = none      ;Estabelece quaisquer restrições utilizadas no modelo
integrator = steep      ;Tipo de integrador, aqui o "steepest descent"
dt = 0.002 ; ps !      ;Não é necessário para a minimização. Só é necessário para integradores dinâmicas (como
md)
nsteps = 1000          ;Número de passos.
nstlist = 10           ;Frequência de atualização da lista de vizinhos, em passos (atualiza a lista a cada 10
passos)
ns_type = grid         ;Tipo de "Neighbor searching"
rlist = 1.0            ;Raio de corte para vizinhos de baixo alcance
coulombtype = PME      ;Informa ao gromacs a eletrostática do modelo, aqui escolhemos a "Fast Particle-
Mesh Ewald"
rcoulomb = 1.0         ;Raio de corte Coulomb
vdwtype = cut-off      ;Tipo de energia Van der Waals, aqui "cut-off"
rvdw = 1.4            ;Raio de corte para Lennard-Jones ou Potencial de Buckingham
optimize_fft = yes     ;Calcula o plano FFT ideal para a grid de partida
;
;Energy minimizing stuff
;
emtol = 1000.0         ;a minimização converge quando a força máxima é menor do que este valor, em kJ.Mol-
1.nm-1.
emstep = 0.01         ;passo inicial, em nm.
```

Figura 1: Arquivo de entrada para minimização.

```

define = -DPOSRES ; Define restrições de posição
constraints = all-bonds ; Constrições em todas as ligações
integrator = md ; Tipo de integrador, aqui o "md" (outra opção é "sd" para a dinâmica
estocástica).
dt = 0.002 ; ps ! ; "Time-Step" (intervalo de tempo), para integração, em ps
nsteps = 50000 ; total 100 ps. ; Número de passos: nsteps X dt = Tempo da simulação em ps
nstcomm = 10 ; Frequência para o centro de remoção de movimento de massa
nstxout = 500 ; collect data every 1.0 ps ; Frequência a escrever a trajetória em arquivo
nstxtcout = 500 ; Frequência para escrever coordenadas para trajetória xtc
nstvout = 5000 ; Frequência a escrever as velocidades em arquivo
nstfout = 0 ; Frequência para escrever forças para a trajetória de saída.
nstlog = 10 ; Frequência para escrever energias no arquivo de log, as últimas energias são
sempre escritas
nstenergy = 50 ; Frequência a escrever as energias em arquivo
nstlist = 10 ; Frequência de atualização da lista de vizinhos, em passos
ns_type = grid ; Tipo de "Neighbor searching"
rlist = 1.0 ; Raio de corte para vizinhos de baixo alcance)
coulombtype = PME ; Calcula interações eletrostáticas entre os átomos., escolhemos a "Fast Particle
Mesh Ewald"
rcoulomb = 1.0 ; Raio de corte Coulomb, em nm
vdwtype = cut-off ; Tipo de energia Van der Waals, aqui "cut-off"
rvdw = 1.4 ; Raio de corte para Lennard-Jones
pme_order = 4 ; Ordem de interpolação para PME
ewald_rtol = 1e-5
optimize_fft = yes ; Calcula o plano FFT ideal para a grid de partida
DispCorr = no ; Não aplicar qualquer correção
; Berendsen temperature coupling is on
Tcoupl = v-rescale ; Tipo de temperatura de acoplamento onde a velocidade é redimensionada usando um termo
estocástico
tau_t = 0.1 0.1 ; Constante de tempo para a temperatura
tc-grps = protein non-protein ; Grupos separados ao banho de temperatura
ref_t = 300 300 ; Temperatura da proteína e do resto do sistema - em Kelvin
; Pressure coupling is off
Pcoupl = no ; Nenhum acoplamento de pressão é selecionado para esta corrida
Pcoupltype = isotropic ; Acoplamento pressão isotrópica com o tempo constante tau-p (ps)
tau_p = 1.0 ; Constante de tempo de acoplamento (em ps). Valor de Tau, no termostato de Berendsen, quando
ligado
compressibility = 4.5e-5 ; Compressibilidade da água a 1 ATM e 300k, unidade em bar-2
ref_p = 1.0 ; Referencial para o barostato
; Generate velocities is on at 300 K.
gen_vel = yes ; Gerar velocidades iniciais por distribuição de Maxwell
gen_temp = 300.0 ; Temperatura inicial, em Kelvin
gen_seed = -1 ; Usado para inicializar o gerador aleatório para velocidades aleatórias

```

Figura 2: Arquivo de entrada para dinâmica de restrição.

```

define = -DFLEXIBLE           ; Define a flexibilidade/mobilidade do solvente, aqui livre
constraints = all-bonds       ; Condições em todas as ligações
integrator = md               ; Tipo de integrador, aqui o "md"
dt = 0.002 ; ps !            ; "Time-Step" (intervalo de tempo), para integração, em ps
nsteps = 100000000 ; total 200000 ps. ; Número de passos: nsteps X dt = Tempo da simulação em ps
nstcomm = 10                  ; Frequência para o centro de remoção de movimento de massa
nstxout = 5000 ; collect data every 10 ps ; Frequência a escrever a trajetória em arquivo
nstxtcout = 0                 ; Frequência para escrever coordenadas para trajetória xtc
nstenergy = 100               ; Frequência a escrever as energias em arquivo
nstvout = 0                   ; Frequência a escrever as velocidades em arquivo
nstfout = 0                   ; Frequência para escrever forças para a trajetória de saída.
nstlist = 10                  ; Frequência de atualização da lista de vizinhos, em passos
ns_type = grid                ; Tipo de "Neighbor searching"
; rlist = 1.0                 ; Raio de corte para vizinhos de baixo alcance
coulombtype = PME             ; calcula interações eletrostáticas entre os átomos., escolhemos a "Fast Particle
Mesh Ewald"
; rcoulomb = 1.0              ; Raio de corte Coulomb, em nm
vdwtype = cut-off             ; Tipo de energia Van der Waals, aqui "cut-off"
; rvdw = 1.4                  ; Raio de corte para Lennard-Jones
pme_order = 4                 ; Ordem de interpolação para PME
ewald_rtol = 1e-5
optimize_fft = yes           ; Calcula o plano FFT ideal para a grid de partida
DispCorr = no                 ; Não aplicar qualquer correção
; Berendsen temperature coupling is on
Tcoupl = v-rescale           ; Tipo de temperatura de acoplamento onde a velocidade é redimensionada usando um
termo estocástico
tau_t = 0.1 0.1              ; Constante de tempo para a temperatura
tc-grps = protein non-protein ; Grupos separados ao banho de temperatura
ref_t = 300 300               ; Temperatura da proteína e do resto do sistema - em Kelvin
; Pressure coupling is on
Pcoupl = parrinello-rahman
Pcoupltype = isotropic        ; Acoplamento pressão isotrópica com o tempo constante tau-p (ps)
tau_p = 1.0                   ; constante de tempo de acoplamento (em ps). Valor de Tau, no termostato de Berendsen, quando
ligado
compressibility = 4.5e-5      ; Compressibilidade da água a 1 ATM e 300k, unidade em bar-2
ref_p = 1.0                   ; Referencial para o barostato
; Generate velocities is on at 300 K.
; gen_vel = yes               ; Gerar velocidades iniciais por distribuição de Maxwell
; gen_temp = 300.0           ; Temperatura inicial, em Kelvin
; gen_seed = -1              ; Usado para inicializar o gerador aleatório para velocidades aleatórias
; Neighbor Searching Parameters
cutoff-scheme = Verlet
verlet-buffer-drift = 0.005

```

Figura 3: Arquivo de entrada para dinâmica molecular.

ANEXO B – TABELAS COM DISTÂNCIA ENTRE CISTEÍNAS DAS ESTRUTURAS TRIDIMENSIONAIS PREDITAS

Tabela I: Distância entre os pares de cisteínas dos modelos gerados por modelagem comparativa e servidor Fugue

Posição	DiANNA	Scratch	Comp	Fugue
17 – 21			7,13	9,2
17 – 31			18,08	24,7
17 – 44			15,19	20,79
17 – 52			13,69	15,05
17 – 59			21,72	33,65
17 – 65			24,46	30,29
17 – 73			37,41	20,39
17 – 85			24,29	9,97
21 – 31			14,09	17,66
21 – 44			12,86	12,19
21 – 52			10,31	8,97
21 – 59			20,58	24,88
21 – 65			19,78	24,13
21 – 73			33,02	14,37
21 – 85			19,94	11,68
31 – 44			8,8	18,63
31 – 52			4,91	22,57
31 – 59			11,16	15,92
31 – 65			7,65	15,42
31 – 73			21,38	9,35
31 – 85			10,9	26,95
44 – 52			5,48	9,95
44 – 59			8,72	18,67
44 – 65			11,56	25,23
44 – 73			22,69	18,66
44 – 85			9,96	20,43
52 – 59			11,05	23,72
52 – 65			10,94	25,23
52 – 73			24,22	18,25
52 – 85			11,93	11,22
59 – 65			11,98	14,82
59 – 73			19,81	17,31
59 – 85			11,55	31,51
65 – 73			13,98	10,72
65 – 85			5,72	27,58
73 – 85			13,43	20,1

A unidade de distância entre os pares de cisteínas é dada em ångström (Å).

Em cinza, predição de possível ligação dissulfeto do servidor *DiANNA*.

Em preto, predição de possível ligação dissulfeto da predição do servidor *SCRATCH*.

Em amarelo, distância que pode formar ligação dissulfeto.

Em azul, distância de até 5,5Å.

Tabela II: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *I-Tasser*

Posição	D	S	IT1	IT2	IT3	IT4	IT5
17 — 21			3,19	7,16	3,4	10,16	10,79
17 — 31	■		15,41	20,49	19,13	25,18	23,62
17 — 44			14,27	9,65	12,25	6,92	8,58
17 — 52			20,87	20,65	20,21	23,89	23,57
17 — 59			23,26	22,32	20,61	26,03	24,13
17 — 65			23,46	21,76	16,07	23,24	24,24
17 — 73			6,19	9,92	3,2	7,89	6,48
17 — 85			15,21	12,68	20,57	18,68	17,1
21 — 31			13,91	18,67	16,98	21,99	22,83
21 — 44		■	15,62	16,43	11,31	14,07	18,05
21 — 52	■		20,37	22,41	19,41	23,67	27,61
21 — 59			22,12	21,22	18,9	23,45	25,26
21 — 65			23,47	24,46	16,52	25,29	29,39
21 — 73			8,89	14,94	6,51	15,07	14,98
21 — 85			13,99	13,48	19,81	17,78	19,03
31 — 44			14,28	23,83	17,51	21,86	23,65
31 — 52		■	8,72	13,79	9,45	8,29	13,07
31 — 59			9,46	7,98	3,47	5,37	6,85
31 — 65			13,51	18,96	13,73	14,23	18,36
31 — 73			17,12	26,57	21,4	25,19	24,57
31 — 85			3,61	10,23	15,14	8,26	10,2
44 — 52			15,16	21,25	16,96	19,88	21,83
44 — 59	■		19,96	25,68	18,61	23,05	23,91
44 — 65			16,9	21,72	17,81	19,02	22,87
44 — 73			13,06	12,88	13,0	9,63	11,06
44 — 85			14,82	15,99	11,76	15,52	18,27
52 — 59			5,43	8,88	6,79	5,79	7,21
52 — 65			5,11	5,38	8,55	6,05	5,84
52 — 73			20,0	21,4	21,23	21,55	21,41
52 — 85			7,4	9,18	9,81	5,89	9,01
59 — 65			8,55	13,5	12,14	11,37	12,14
59 — 73		■	23,18	25,48	22,51	24,44	23,25
59 — 85			8,13	9,98	14,12	7,6	7,55
65 — 73			21,28	19,99	16,37	19,25	20,59
65 — 85		■	11,48	12,46	14,98	9,01	11,69
73 — 85	■		15,91	16,68	21,12	17,29	15,89

Tabela III: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *LOMETS*

Posição	D	S	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10
17 — 21			10,02	10,74	5,87	13,91	7,0	5,87	5,87	13,5	10,14	4,05
17 — 31			23,81	25,44	22,81	30,38	2,03	22,81	22,81	37,32	22,67	17,46
17 — 44			7,4	31,04	5,1	52,46	23,08	5,1	5,1	35,34	6,33	32,07
17 — 52			22,99	39,47	20,89	72,02	10,54	20,89	20,89	46,71	21,91	34,71
17 — 59			24,91	28,3	24,19	54,0	18,73	24,19	24,19	67,15	24,02	37,34
17 — 65			21,59	29,38	20,96	60,85	31,45	20,96	20,96	50,43	23,07	24,31
17 — 73			8,24	33,48	5,97	61,13	35,81	5,97	5,97	55,42	7,1	4,76
17 — 85			16,03	35,64	15,12	59,94	47,87	15,12	15,12	40,72	15,1	18,33
21 — 31			22,39	16,68	18,13	34,75	7,53	18,13	18,13	25,84	21,07	16,31
21 — 44			15,62	23,88	9,12	58,06	21,94	9,12	9,12	23,84	15,47	34,43
21 — 52			24,57	32,73	17,93	77,29	8,77	17,93	17,93	34,08	24,88	34,3
21 — 59			23,22	27,26	19,75	59,54	17,64	19,75	19,75	55,51	22,86	35,68
21 — 65			25,5	28,04	19,83	67,3	34,03	19,83	19,83	39,39	27,35	22,97
21 — 73			15,74	29,2	10,38	64,71	38,12	10,38	10,38	43,67	13,68	7,46
21 — 85			16,02	28,63	11,38	61,06	48,06	11,38	11,38	32,92	15,5	17,5
31 — 44			23,72	8,48	23,52	24,15	21,34	23,52	23,52	2,03	23,62	27,57
31 — 52			8,84	18,01	8,08	42,6	10,23	8,08	8,08	11,12	10,77	23,8
31 — 59			2,04	21,14	2,03	25,88	19,48	2,03	2,03	30,26	2,02	22,73
31 — 65			13,43	22,6	14,72	34,47	32,12	14,72	14,72	14,34	16,09	14,3
31 — 73			24,68	21,23	23,96	30,92	37,24	23,96	23,96	19,8	22,33	18,23
31 — 85			8,0	22,25	8,29	30,23	49,41	8,29	8,29	16,42	7,93	14,65
44 — 52			22,75	10,64	20,09	21,8	17,91	20,09	20,09	12,61	21,15	25,6
44 — 59			25,22	17,18	24,71	2,03	28,56	24,71	24,71	32,27	24,76	35,05
44 — 65			20,26	18,64	18,98	11,23	41,87	18,98	18,98	16,29	22,37	28,6
44 — 73			12,09	17,38	6,72	13,1	52,55	6,72	6,72	21,66	10,75	29,36
44 — 85			16,78	22,51	15,88	20,93	62,7	15,88	15,88	16,95	16,39	27,41
52 — 59			8,69	21,42	7,63	21,29	12,34	7,63	7,63	23,12	9,88	13,17
52 — 65			5,72	19,27	6,69	20,84	28,6	6,69	6,69	12,42	6,04	12,81
52 — 73			20,98	14,16	20,57	5,26	36,07	20,57	20,57	13,84	19,67	32,13
52 — 85			9,0	20,57	7,44	25,24	45,94	7,44	7,44	23,87	9,72	17,8
59 — 65			13,83	10,47	14,19	9,53	19,55	14,19	14,19	17,01	15,16	14,24
59 — 73			25,35	19,89	24,94	2,97	26,32	24,94	24,94	12,92	23,27	36,26
59 — 85			8,95	32,2	9,3	21,17	34,23	9,3	9,3	31,91	9,03	21,93
65 — 73			18,72	11,4	19,13	8,5	16,27	19,13	19,13	7,97	19,19	22,6
65 — 85			11,22	25,26	11,46	28,24	27,68	11,46	11,46	16,36	13,59	7,69
73 — 85			16,82	15,05	15,68	11,41	17,05	15,68	15,68	22,54	14,49	15,84

Tabela IV: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *MUSTER*

Posição	D	S	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10
17 — 21			5,87	13,41	9,01	9,04	3,01	12,44	16,15	1,82	13,82	3,86
17 — 31			22,81	21,91	16,79	2,02	8,04	8,0	17,47	2,03	22,86	4,74
17 — 44			5,1	17,79	18,5	18,12	20,65	2,03	16,68	18,34	22,56	7,73
17 — 52			20,89	19,75	20,8	15,06	17,01	15,87	4,71	20,31	22,49	15,45
17 — 59			24,19	21,48	18,2	17,34	19,85	28,22	12,49	17,31	22,75	26,1
17 — 65			20,96	17,11	27,73	26,98	22,82	31,38	9,18	28,0	29,35	20,77
17 — 73			5,97	2,02	24,22	36,25	35,16	27,82	32,99	26,31	28,53	24,6
17 — 85			15,12	25,06	21,55	28,15	30,29	22,99	12,68	24,63	36,76	30,05
21 — 31			18,13	24,3	16,99	10,33	12,14	7,42	15,77	10,64	13,43	2,03
21 — 44			9,12	24,4	13,06	9,55	7,93	13,05	5,19	10,61	18,58	5,01
21 — 52			17,93	26,69	11,96	10,52	10,22	16,04	15,09	13,21	16,31	12,05
21 — 59			19,75	24,55	11,51	9,19	7,16	17,4	27,29	10,9	22,12	22,64
21 — 65			19,83	26,76	22,59	21,72	12,74	21,44	11,88	24,82	27,36	17,19
21 — 73			10,38	13,48	19,82	27,86	23,72	16,56	24,85	17,98	28,13	21,1
21 — 85			11,38	26,0	18,7	20,22	18,08	10,91	13,99	20,27	39,72	26,58
31 — 44			23,52	14,92	16,04	19,19	17,88	7,53	12,45	17,52	9,33	3,15
31 — 52			8,08	8,37	22,28	16,46	20,32	14,18	13,65	20,32	12,77	10,99
31 — 59			2,03	2,02	13,3	18,23	17,69	21,74	24,81	16,55	15,8	21,53
31 — 65			14,72	14,18	15,51	27,6	21,7	27,48	9,48	27,83	19,4	16,18
31 — 73			23,96	19,9	12,15	37,42	35,81	21,14	16,66	25,63	21,22	20,09
31 — 85			8,29	11,02	8,08	29,53	29,99	16,16	13,55	24,67	33,59	25,72
44 — 52			20,09	12,34	11,19	10,59	14,38	16,15	14,17	11,65	12,08	8,88
44 — 59			24,71	15,52	8,89	2,04	2,02	28,56	28,1	2,03	7,17	19,06
44 — 65			18,98	14,68	14,68	15,46	9,96	32,49	11,44	15,0	11,71	13,82
44 — 73			6,72	16,44	17,36	18,59	19,86	28,12	20,53	8,19	12,88	17,78
44 — 85			15,88	12,07	17,26	12,21	13,81	23,25	15,39	11,44	24,54	23,6
52 — 59			7,63	7,3	9,64	10,66	12,93	20,6	15,33	12,61	12,44	10,69
52 — 65			6,69	6,65	18,7	15,0	14,98	24,11	7,25	21,14	13,82	5,66
52 — 73			20,57	17,84	19,0	23,77	19,21	21,34	28,84	13,74	15,38	9,17
52 — 85			7,44	15,3	20,46	14,83	16,08	21,1	12,86	13,34	27,9	14,78
59 — 65			14,19	12,78	11,4	14,87	8,35	12,95	17,62	14,14	7,24	5,7
59 — 73			24,94	19,46	10,21	19,66	19,24	2,02	41,43	9,39	6,83	2,02
59 — 85			9,3	12,88	10,98	13,59	13,81	9,16	17,62	11,18	18,19	5,69
65 — 73			19,13	15,49	8,53	17,84	18,73	13,68	25,24	12,95	3,27	4,28
65 — 85			11,46	21,05	10,91	14,18	16,97	17,15	6,69	8,39	15,56	10,23
73 — 85			15,68	23,38	4,43	9,58	8,5	7,47	28,47	9,72	12,96	6,08

Tabela V: Distância entre os pares de cisteínas dos modelos gerados pelo servidor SPARKS-X

Posição	D	S	SP1	SP2	SP3	SP4	SP5	SP6	SP7	SP8	SP9	SP10
17 — 21			4,74	12,39	5,39	7,03	7,58	12,72	11,86	14,54	4,23	6,17
17 — 31			22,8	22,53	23,68	29,35	2,03	12,19	16,22	23,8	26,76	29,72
17 — 44			8,6	19,51	26,73	23,59	14,92	24,46	14,35	14,68	24,56	31,48
17 — 52			21,96	22,98	21,77	22,51	9,29	23,08	13,73	28,67	35,32	18,06
17 — 59			22,81	22,48	6,51	28,87	17,3	27,58	30,48	46,74	49,56	9,58
17 — 65			21,46	21,91	25,92	25,57	16,3	15,34	10,32	45,55	38,42	25,46
17 — 73			7,04	2,04	21,59	16,0	21,91	21,58	14,73	47,89	27,88	29,37
17 — 85			16,26	29,45	20,55	25,5	18,88	21,7	31,73	30,48	30,62	7,69
21 — 31			19,64	24,59	18,46	26,92	8,42	5,88	15,91	10,34	23,16	24,22
21 — 44			11,94	27,17	21,96	19,44	8,06	14,14	17,22	2,03	21,45	27,83
21 — 52			20,93	29,02	17,52	20,52	11,57	14,64	6,33	14,2	31,59	13,73
21 — 59			20,0	25,54	4,27	28,4	10,63	23,47	18,65	33,58	45,81	6,02
21 — 65			21,46	29,75	22,78	22,8	9,53	7,16	3,19	32,74	35,21	21,94
21 — 73			9,8	14,17	17,88	13,12	17,59	9,63	19,23	36,14	24,97	28,86
21 — 85			14,41	33,35	17,56	24,41	11,98	17,05	20,03	19,04	27,13	3,73
31 — 44			25,47	16,11	14,44	9,73	16,19	12,34	17,89	9,35	14,6	18,9
31 — 52			11,07	9,18	14,14	7,95	10,96	11,35	12,76	7,31	15,4	13,3
31 — 59			2,02	2,02	18,27	8,82	18,19	18,25	27,48	26,53	25,51	23,37
31 — 65			15,37	12,77	20,67	10,35	17,56	3,18	17,81	27,93	20,81	12,76
31 — 73			22,03	21,82	20,42	22,59	23,7	10,5	23,17	30,24	6,88	28,9
31 — 85			8,64	15,83	20,99	6,48	19,83	15,2	31,46	16,87	5,87	22,62
44 — 52			21,14	9,97	7,83	7,55	15,16	4,22	13,62	14,18	12,67	16,63
44 — 59			24,92	14,99	22,8	16,27	6,3	14,78	31,88	33,59	26,89	30,77
44 — 65			19,5	10,52	14,11	10,9	2,02	9,17	16,5	33,28	13,89	19,49
44 — 73			10,55	18,34	12,33	15,54	11,97	6,31	28,92	36,16	15,75	17,74
44 — 85			18,84	12,48	11,86	11,97	6,32	15,6	32,79	19,56	20,38	27,79
52 — 59			9,17	7,34	17,24	11,58	20,05	11,38	19,19	21,31	14,91	14,87
52 — 65			4,72	5,03	8,18	12,49	16,78	8,32	8,24	21,69	8,3	9,96
52 — 73			19,34	21,68	14,4	15,2	16,37	8,88	23,73	25,97	18,32	20,23
52 — 85			8,79	12,47	8,22	7,37	20,94	13,22	21,26	13,49	19,03	12,3
59 — 65			13,53	10,84	21,42	11,96	4,98	16,32	20,91	8,87	15,9	21,02
59 — 73			21,62	21,61	20,72	26,17	17,13	20,06	34,01	7,92	27,09	31,61
59 — 85			7,93	15,29	18,28	4,55	2,02	19,4	6,2	17,7	26,13	3,88
65 — 73			17,88	20,3	20,98	23,93	12,33	8,12	18,57	11,58	21,28	20,17
65 — 85			10,76	16,59	11,1	9,5	4,69	14,23	21,52	15,69	24,74	19,31
73 — 85			13,89	28,69	11,29	21,93	16,34	15,17	35,8	18,3	9,08	28,03

Tabela VI: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *QUARK*

Posição	D	S	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
17 — 21			12,81	5,81	4,32	3,45	5,47	7,99	3,89	10,29	7,02	8,2
17 — 31			12,36	16,23	18,42	16,74	17,02	23,08	19,33	13,31	16,1	14,05
17 — 44			16,35	14,84	13,23	24,76	28,32	13,54	15,42	11,69	19,14	9,03
17 — 52			13,78	10,2	21,11	16,35	18,9	20,73	13,86	13,55	6,75	18,64
17 — 59			18,36	13,7	25,41	7,54	16,93	15,99	28,34	19,34	11,25	24,02
17 — 65			20,1	11,79	20,28	11,43	18,34	13,12	18,51	16,71	8,61	15,6
17 — 73			21,42	11,27	20,79	13,63	22,51	11,08	18,99	17,51	7,94	13,97
17 — 85			24,77	15,28	23,83	10,42	23,7	13,65	23,07	25,79	13,02	9,2
21 — 31			16,29	19,98	14,65	17,4	15,38	19,52	18,11	16,95	14,78	7,98
21 — 44			8,54	15,29	11,89	26,27	26,43	14,32	14,2	11,94	12,82	10,66
21 — 52			15,49	15,28	17,55	18,45	17,62	20,57	11,33	11,5	3,31	14,64
21 — 59			20,63	19,1	21,32	10,49	13,92	15,86	25,46	17,9	14,11	17,18
21 — 65			20,41	17,12	16,92	14,04	15,7	13,76	15,04	13,46	11,42	9,16
21 — 73			21,17	16,51	17,7	15,46	19,99	12,24	15,3	14,86	11,83	6,88
21 — 85			25,0	18,78	20,96	12,17	21,25	14,84	19,22	23,12	17,49	3,67
31 — 44			12,78	12,33	12,75	15,63	12,31	14,81	13,8	13,55	15,39	10,8
31 — 52			11,41	7,94	9,42	7,98	9,05	9,52	8,83	8,45	11,72	8,14
31 — 59			13,45	10,51	17,24	14,03	14,26	13,86	16,94	11,34	9,04	10,03
31 — 65			16,24	13,37	15,11	11,32	15,17	14,17	14,47	10,75	9,82	2,05
31 — 73			17,69	14,67	16,2	14,59	16,8	15,37	16,24	10,22	12,01	3,74
31 — 85			21,62	24,31	20,38	17,7	20,78	20,38	22,88	16,28	13,32	6,82
44 — 52			12,66	13,45	12,71	15,42	11,36	12,99	12,82	13,29	13,05	11,39
44 — 59			16,38	19,49	25,75	22,25	18,72	15,54	25,88	20,49	21,14	19,49
44 — 65			16,18	16,87	16,07	15,19	18,42	12,77	17,42	16,1	18,7	11,91
44 — 73			16,88	17,28	15,78	12,5	16,84	12,17	17,15	11,95	19,93	13,09
44 — 85			21,09	23,61	19,95	17,98	21,28	19,16	22,62	18,26	24,88	10,87
52 — 59			5,43	6,73	17,19	10,56	9,57	9,36	15,27	7,45	11,05	10,05
52 — 65			6,38	6,3	8,29	7,28	9,21	9,47	7,35	3,36	8,85	7,31
52 — 73			7,74	7,34	8,6	13,4	9,52	11,07	9,04	6,82	9,72	11,82
52 — 85			11,5	16,84	12,68	16,65	13,5	16,36	15,46	14,96	14,95	14,26
59 — 65			3,93	9,96	14,73	7,21	2,06	3,3	11,6	5,46	5,02	8,44
59 — 73			5,37	10,99	16,78	13,23	6,21	5,1	13,37	11,4	6,53	11,36
59 — 85			8,43	19,94	17,6	12,51	7,93	7,28	16,23	16,58	5,98	16,27
65 — 73			1,55	1,56	2,09	7,47	4,43	2,05	2,78	7,36	2,23	4,83
65 — 85			5,52	11,42	5,33	9,66	6,18	7,48	8,67	14,29	6,44	8,54
73 — 85			4,34	9,93	4,5	5,54	4,51	7,13	6,69	8,68	6,18	5,15

Tabela VII: Distância entre os pares de cisteínas dos modelos gerados pelo servidor QUARK com restrição de distância da predição do DiANNA

Posição	D	S	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
17 — 21			10,65	9,01	6,27	4,22	14,71	15,15	10,63	6,59	8,97	7,4
17 — 31			5,42	6,35	3,13	5,79	4,92	7,24	2,01	5,27	8,33	5,53
17 — 44			16,05	14,14	19,12	12,67	13,97	12,83	17,92	17,09	17,13	14,81
17 — 52			16,05	9,69	8,49	4,07	12,52	14,8	15,55	9,7	8,9	11,14
17 — 59			15,96	12,69	13,89	11,35	10,93	13,92	12,41	16,76	17,07	13,35
17 — 65			17,53	10,9	13,02	14,21	12,32	13,52	14,05	10,94	16,05	21,66
17 — 73			17,15	12,08	12,8	14,39	12,6	11,5	11,87	12,78	15,01	24,31
17 — 85			15,25	10,96	16,58	15,37	10,66	3,78	15,43	17,14	10,65	27,05
21 — 31			6,57	7,79	4,6	9,37	13,17	12,26	11,32	3,47	5,56	10,76
21 — 44			11,98	16,65	13,83	14,7	13,1	14,55	9,21	12,17	11,0	10,09
21 — 52			6,7	6,28	2,34	3,48	4,92	3,0	7,34	4,6	2,1	4,95
21 — 59			14,74	11,02	8,22	11,2	15,84	17,16	10,14	10,75	13,38	10,06
21 — 65			8,52	13,47	7,57	16,3	10,95	18,57	8,68	4,66	12,55	18,28
21 — 73			8,04	16,94	8,22	16,41	12,55	19,57	8,06	7,29	14,34	20,58
21 — 85			8,48	16,99	13,89	17,46	15,14	15,76	13,92	10,78	12,52	22,56
31 — 44			12,1	16,99	16,21	8,32	13,0	12,71	17,9	11,83	11,2	16,08
31 — 52			11,06	8,68	6,85	7,04	9,72	10,99	16,46	6,5	4,47	14,5
31 — 59			11,83	13,74	11,25	10,54	9,33	11,9	11,44	12,04	12,1	16,09
31 — 65			12,42	9,64	11,38	9,71	7,91	10,24	13,66	7,87	8,45	19,84
31 — 73			12,25	9,97	11,35	9,71	7,87	7,76	11,97	10,67	8,97	22,27
31 — 85			11,13	10,17	15,74	10,92	5,94	5,79	15,46	13,54	7,08	25,54
44 — 52			13,54	11,66	12,55	11,77	13,51	12,73	13,43	11,79	9,43	13,12
44 — 59			11,01	7,33	5,83	7,42	6,62	5,86	9,4	6,64	4,87	5,98
44 — 65			14,1	10,93	12,7	1,65	11,19	10,27	5,57	11,93	9,68	11,7
44 — 73			15,04	17,05	14,35	2,3	12,63	15,84	8,5	14,82	15,44	14,98
44 — 85			17,57	16,11	19,97	10,3	16,07	11,23	12,87	12,45	15,36	16,09
52 — 59			13,9	5,78	6,77	9,08	14,14	14,72	16,33	11,2	11,4	13,66
52 — 65			1,89	8,69	6,51	13,41	7,28	16,09	15,0	5,6	10,94	20,63
52 — 73			1,63	14,35	7,8	13,27	8,5	17,75	15,24	6,22	13,14	22,19
52 — 85			5,33	14,53	14,11	14,12	10,52	14,68	21,0	8,39	11,25	23,73
59 — 65			13,7	10,29	8,52	8,47	9,41	5,01	4,38	8,31	9,5	17,27
59 — 73			14,71	17,2	10,37	9,02	9,98	12,23	7,33	11,22	14,86	20,8
59 — 85			15,38	16,98	16,8	14,95	11,95	10,91	11,51	9,01	14,17	22,0
65 — 73			1,56	7,36	2,42	2,1	1,88	7,9	4,95	4,03	6,34	4,17
65 — 85			5,85	7,71	9,01	10,95	6,66	9,87	9,58	7,01	8,59	6,62
73 — 85			4,69	2,43	6,74	9,06	5,3	8,33	6,09	5,86	5,15	3,91

Tabela VIII: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *QUARK* com restrição de distância da predição do *Scratch*

Posição	D	S	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
17 — 21			3,47	7,5	6,61	8,09	8,16	7,9	8,04	7,46	7,39	6,95
17 — 31			18,57	19,67	18,49	11,54	18,88	22,4	20,25	13,15	10,06	12,77
17 — 44			13,28	21,26	20,03	23,45	20,91	19,79	23,31	15,12	12,62	14,14
17 — 52			12,12	18,42	20,2	13,16	9,96	20,32	12,41	18,65	16,19	3,79
17 — 59			15,96	27,18	12,89	8,97	14,82	28,76	22,16	19,93	17,82	11,14
17 — 65			12,05	19,13	15,13	15,8	15,65	17,38	24,56	18,23	18,67	16,9
17 — 73			11,23	18,28	14,38	17,83	17,13	19,59	24,94	18,78	20,28	17,79
17 — 85			15,0	17,31	11,08	19,79	25,39	21,36	26,75	23,06	24,79	17,58
21 — 31			17,97	18,95	14,05	10,56	21,67	24,16	14,03	12,07	14,56	12,68
21 — 44			13,88	17,24	19,79	20,34	21,05	17,83	15,32	17,7	18,29	8,28
21 — 52			12,26	15,97	16,38	12,74	12,64	19,51	5,33	15,56	19,62	3,77
21 — 59			17,55	26,12	8,79	13,94	15,02	31,64	19,63	15,42	18,66	13,95
21 — 65			14,0	17,42	13,52	19,08	12,63	19,08	20,36	15,99	20,04	14,5
21 — 73			13,51	18,42	13,86	20,8	14,04	18,29	19,82	16,17	21,24	14,79
21 — 85			18,23	16,95	12,18	23,37	20,91	17,11	19,57	20,54	25,14	13,58
31 — 44			12,41	11,77	13,91	13,68	12,21	13,03	13,76	14,39	13,3	11,38
31 — 52			8,63	7,74	8,58	9,59	9,29	9,83	8,73	9,92	8,9	11,55
31 — 59			14,68	15,44	13,34	12,71	11,59	9,95	11,47	9,06	10,06	5,75
31 — 65			15,2	6,68	14,43	14,13	13,63	5,93	8,67	6,8	12,25	6,08
31 — 73			16,4	4,12	16,09	15,73	13,68	10,94	6,69	8,73	17,06	7,28
31 — 85			25,62	2,63	18,9	18,88	19,66	18,27	7,49	16,31	19,74	8,83
44 — 52			12,53	13,22	12,0	12,42	13,48	12,26	12,7	13,73	12,83	10,7
44 — 59			17,32	23,95	17,68	26,29	18,3	22,81	24,64	20,73	20,98	15,6
44 — 65			15,19	14,2	12,61	17,23	12,1	12,38	21,9	14,22	17,09	9,59
44 — 73			15,17	15,06	12,6	17,21	12,16	11,64	19,36	13,72	15,83	8,97
44 — 85			21,25	11,38	16,27	20,19	16,06	13,58	12,26	14,1	21,46	6,81
52 — 59			8,28	11,71	10,77	18,8	8,2	15,61	15,8	9,08	10,15	11,75
52 — 65			7,54	2,13	9,0	8,41	9,37	6,8	15,6	4,61	5,66	14,51
52 — 73			8,68	6,94	11,16	9,39	10,42	1,58	14,71	2,19	10,01	15,26
52 — 85			18,05	8,27	15,55	12,31	18,8	9,17	14,46	6,93	12,16	14,81
59 — 65			4,18	10,54	7,15	19,81	9,88	12,82	5,26	7,51	8,55	11,82
59 — 73			5,65	13,08	8,53	21,91	10,13	17,15	8,18	8,47	15,92	12,97
59 — 85			13,68	17,05	9,14	23,97	16,81	24,69	17,33	15,53	15,2	14,18
65 — 73			1,67	5,32	2,43	2,12	1,55	7,53	3,19	2,59	7,77	2,0
65 — 85			10,99	7,58	6,96	4,79	9,89	14,84	12,85	10,7	7,68	4,98
73 — 85			9,51	4,54	5,2	3,32	8,53	7,76	9,72	8,21	6,41	3,27

Tabela XIX: Distância entre os pares de cisteínas dos modelos gerados pelo servidor *ROBETTA*

Posição	D	S	R1	R2	R3	R4	R5
17 — 21			8,66	7,09	7,75	7,52	13,22
17 — 31	■		27,41	27,47	27,1	13,23	40,04
17 — 44			17,09	19,26	7,99	15,43	17,21
17 — 52			21,56	27,12	25,63	4,94	21,75
17 — 59			31,19	29,2	26,73	9,61	33,8
17 — 65			30,39	26,26	24,58	4,87	18,45
17 — 73			21,37	11,09	8,49	12,05	20,52
17 — 85			15,57	27,04	19,36	10,21	26,35
21 — 31			18,85	22,36	19,92	8,37	28,6
21 — 44		■	10,06	17,95	8,84	9,87	7,17
21 — 52	■		14,07	23,26	21,0	12,29	9,34
21 — 59			24,04	23,71	20,39	14,77	22,4
21 — 65			23,46	23,43	21,5	10,93	11,66
21 — 73			13,43	11,71	10,82	18,38	11,98
21 — 85			8,85	21,97	13,3	9,19	13,63
31 — 44			15,92	15,18	23,62	4,16	24,13
31 — 52		■	11,0	9,15	12,62	17,7	21,64
31 — 59			11,9	6,03	5,4	22,24	21,56
31 — 65			14,38	13,39	17,51	14,02	23,58
31 — 73			9,51	24,12	25,3	19,79	20,71
31 — 85			16,76	3,73	9,42	12,15	18,66
44 — 52			11,2	10,99	21,43	20,19	10,78
44 — 59	■		25,3	17,14	22,8	24,15	23,92
44 — 65			20,66	9,16	19,84	16,92	11,44
44 — 73			10,11	13,66	10,85	22,3	5,7
44 — 85			2,02	14,16	16,47	15,71	13,83
52 — 59			17,3	8,09	7,38	7,46	13,47
52 — 65			9,86	4,6	5,59	5,08	10,75
52 — 73			2,02	20,36	20,56	10,1	12,84
52 — 85			10,84	6,32	8,76	12,41	4,65
59 — 65			16,4	12,44	12,56	12,26	20,56
59 — 73		■	16,9	24,15	23,45	17,01	24,57
59 — 85			25,43	3,45	7,63	15,64	10,43
65 — 73			11,11	18,27	18,95	8,08	10,24
65 — 85		■	20,44	10,71	11,87	10,67	14,08
73 — 85	■		10,03	22,34	16,13	18,26	15,06