



Universidade de Brasília

Instituto de Biologia

Laboratório de Interação Patógeno-Hospedeiro

Hugo de Almeida Silva

**FERRAMENTAS VIRTUAIS CONTRA DOENÇAS TROPICAIS:
Dengue e doença de Chagas como alvos**

Brasília

2012

Hugo de Almeida Silva

**FERRAMENTAS VIRTUAIS CONTRA DOENÇAS TROPICAIS:
Dengue e doença de Chagas como alvos**

Tese apresentada ao programa de Pós-graduação em Patologia Molecular da Universidade de Brasília como requisito parcial à obtenção do título de Doutor em Patologia Molecular.

Orientador: Prof. Dr. Jaime Martins de Santana
Co-orientadores: Prof. Dr. Bergmann Morais Ribeiro
Prof. Dr. Bernard Maignet

Brasília
2012

Ficha catalográfica elaborada pela Biblioteca Central da Universidade de Brasília. Acervo 1003400.

S586f Silva, Hugo de Almeida.
Ferramentas virtuais contra doenças tropicais : dengue e doença de chagas como alvo / Hugo de Almeida Silva. -- 2012.

176 f. : il. ; 30 cm.

Tese (doutorado) - Universidade de Brasília, Faculdade de Medicina, Pós-Graduação em Patologia Molecular, 2012.
Inclui bibliografia.

Orientação: Jaime Martins de Santana ; Co-orientação: Bergmann Morais Ribeiro, Bernard Maigret.

1. Doenças transmissíveis. 2. Dengue. 3. Chagas, Doença de. 4. Biotecnologia farmacêutica. I. Santana, Jaime Martins de. II. Ribeiro, Bergmann Morais. III. Maigret, Bernard. IV. Título.

CDU 616(213)

Trabalho realizado no Laboratório de Interação Patógeno-Hospedeiro junto ao programa de Pós-Graduação em Patologia Molecular da Faculdade de Medicina da Universidade de Brasília e no *Laboratoire Lorrain de Recherche en Informatique et ses Applications* (LORIA), Université Henri Poincaré, Nancy – França. Apoio financeiro: FAP-DF CAPES e CNPq.

Dedico este trabalho ao meu avô, Ênio Ribeiro de Almeida, que antes de ir-se já previa meu futuro ao me chamar de “Dr. Gungunga”... Saudades!

Igualmente, dedico à minha amada Sol, a quem devo muito do que sou hoje.
Você me faz querer ser sempre melhor.

Aos meus pais, Helcia e Wadyr, que me guiaram até aqui.

AGRADECIMENTOS

Agradeço aos meus pais, à minha família e aos meus amigos, pelo carinho, amor e compreensão nos momentos em que estive ausente devido aos estudos;

À minha esposa Sol, pelo companheirismo, apoio incondicional, força e incentivo – por me fazer sonhar e buscar meus sonhos; e à minha sogra Márcia pelo apoio e acolhimento;

Ao professor Dr. Jaime Santana, pela orientação, por acreditar no meu potencial e me incentivar a seguir o caminho da ciência;

Ao professor Dr. Bergmann Morais, pelos conselhos e boa convivência desde a graduação; e ao meu tio Antônio Chaib, por ter facilitado que nossos caminhos se cruzassem;

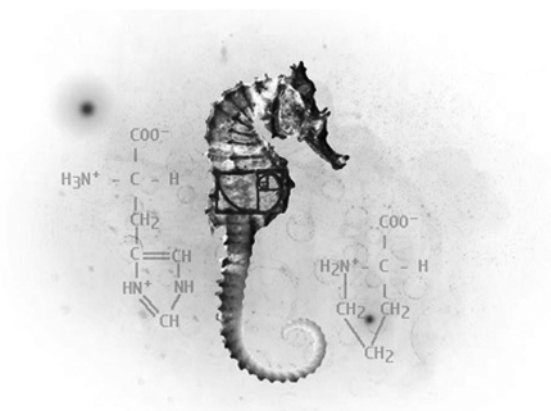
Ao professor Dr. Bernard Maigret, não só por ter me recebido em seu laboratório, mas por ter se mostrado um amigo e se preocupado em tornar a minha estadia na França a melhor possível;

Aos amigos e (ex) parceiros de bancada, Raquel Negreiros, Brina Portugal, Thiago Gastardelo, André França Corrêa, Bruna Guido, Diego Nolasco, Isabel Souza, David Neves, Alessandro Sousa, Paula Beatriz, Clênia Azevedo, e às recém-chegadas Tati, Luz e Cissé, pela amizade, auxílio, apoio e momentos de descontração;

Aos amigos e colegas dos laboratórios “anexos”, de Virologia, Bioquímica e Toxinologia, onde passei parte do tempo seja procurando auxílio para realizar algum experimento, ou apenas procurando momentos de risadas;

Aos colegas e à equipe do LORIA/INRIA em Nancy, França, pelo acolhimento, conselhos, auxílios e também pelos momentos de relaxamento; aos amigos Martín Barrère, Renaud Grisoni e Emmanuel Bresso, pelo companheirismo, festas e brincadeiras;

Aos outros professores do laboratório, Anamélia Bocca, Beatriz de Lima, Cecília Favali, Izabela Dourado, Carla Araújo, e Flávia Nader, por se esforçarem em manter um ambiente de trabalho propício ao desenvolvimento da pesquisa; aos colegas e alunos de iniciação científica, por contribuírem para o avanço dos nossos objetos de estudo.



"Where there's a will, there's a way."

— W. Bishop, PhD.

RESUMO

Apesar de afetarem praticamente um sexto da população mundial, doenças como a dengue e o mal de Chagas continuam sendo negligenciadas pela indústria farmacêutica e pelo governo. Assumindo a responsabilidade da pesquisa de base, as instituições de ensino superior podem tornar o desenvolvimento de fármacos contra as doenças negligenciadas mais atraentes aos olhos da indústria. Com a proposta de diminuir os gastos e o tempo na busca de novos medicamentos contra essas doenças, o presente trabalho buscou aplicar métodos de análise *in silico* para a implementação de uma plataforma de descoberta racional de fármacos voltada para as doenças importantes no Brasil. Dessa forma, duas doenças para as quais não existem vacinas ou tratamento efetivo foram escolhidas para um estudo piloto: a dengue e a doença de Chagas. Um dos alvos mais promissores contra a dengue é a protease viral NS3, que, associada ao seu cofator NS2B, promove o processamento da poliproteína viral e, conseqüentemente, a replicação do vírus. Por meio de técnicas de modelagem comparativa, dinâmicas moleculares e triagem virtual, um possível papel para o cofator NS2B e para o ligante no efeito de ajuste induzido do sítio ativo da NS3 foi proposto. Além disso, quatro famílias de conformações foram encontradas e exploradas em uma campanha de triagem virtual, onde foram obtidos quatro moléculas com atividade inibitória comprovada. Ao mesmo tempo, técnicas similares foram aplicadas a dois alvos de *Trypanosoma cruzi* já estudados pelo nosso grupo: a catepsina B (TcCatB) e a metiltioadenosina fosforilase (TcMTAP). Os modelos construídos se mostraram estáveis durante simulações de dinâmica molecular, e também foram submetidos à triagem virtual. Moléculas com atividade inibitória putativas contra esses alvos foram identificadas. Os resultados deste trabalho apontam para a viabilidade de estabelecer-se uma plataforma de busca racional de fármacos contra outras doenças negligenciadas.

ABSTRACT

Despite affecting about one-sixth of the world's population, diseases like dengue fever and Chagas disease are still overlooked by the government and pharmaceutical industries. By assuming the responsibility over basic research, public institutions of higher education may play a role in making the development of new drugs against neglected diseases more attractive to the big pharma companies. With the aim of cutting costs and time for searching new drugs against those diseases, this work intended to apply *in silico* analysis methods towards the implementation of a rational drug design platform against neglected diseases occurring in Brazil. Thus, two illnesses for which there are no vaccines or effective treatments were chosen for a pilot study: dengue fever and Chagas disease. One of the most promising targets against dengue fever is the viral NS3 protease, which, together with its cofactor NS2B, promotes the viral polyprotein processing and therefore viral replication. By using comparative modeling techniques, molecular dynamics simulations and virtual screening, a possible role of the NS2B cofactor and the ligand binding in the induced fit effect at the NS3 active site was proposed. Also, four conformational families were found and explored in a virtual screening campaign, in which four compounds with proven inhibitory activity were found. At the same time, similar techniques were applied to two targets of *Trypanosoma cruzi* already studied by our team: the cathepsin B (TcCatB) and the methylthioadenosine phosphorilase (TcMTAP). The built models were shown to be stable during molecular dynamics simulations, and were also subjected to virtual screening campaigns. Molecules with putative inhibitory activity against those targets were detected. Together, those results point to the viability in establishing a rational drug discovery platform against neglected diseases.

LISTA DE FIGURAS

Figura 1. Países com transmissão local da dengue nos últimos 25 anos.....	20
Figura 2. Ciclo de replicação do DENV..	22
Figura 3. Organização da poliproteína do DENV e seus sítios de proteólise.....	23
Figura 4. Estruturas cristalográficas da DENV NS3B _{CF} -NS3 _{PRO}	24
Figura 5. Formas de vida do <i>T. cruzi</i> e seu ciclo de vida simplificado.....	29
Figura 6. Modelo teórico da POP Tc80, obtido a partir de modelagem por homologia. ...	31
Figura 7. Parcela de doenças cobertas pelo interesse farmacêutico mundial.....	33
Figura 8. Etapas que constituem a pesquisa básica, antes de uma molécula ser submetida aos testes pré-clínicos.	35
Figura 9. Etapas do desenho racional de drogas assistido por computador.....	41
Figura 10. Funil de estados conformacionais de uma proteína.	52
Figura 11. Amplificação dos fragmentos NS2B _{CF} -gly e gly-NS3 _{PRO} obtidos por PCR em gel de agarose 8%.	76
Figura 12. Padrão de migração do DNA plasmidial de clones do produto de ligação pGEM®-T Easy e do fragmento NS2BCF40GlyNS3PRO.	78
Figura 13. Confirmação dos clones recombinantes de pET-19b-DEN1_NS2B _{CF} NS3 _{PRO} . 79	
Figura 14. Resultados do sequenciamento do plasmídeo pET-19b_ns2b _{cf40} -gly-ns3 _{pro} e do alinhamento da cadeia de aminoácidos traduzidos.	80
Figura 15. Perfil eletroforético do extrato solúvel e insolúvel de clones contendo o plasmídeo de expressão induzidos.	81
Figura 16. Análise de diferentes concentrações de imidazol no tampão de lavagem durante a purificação da NS2B _{CF} -Gly-NS3 _{PRO} recombinante.....	82
Figura 17. Alinhamento manual entre a sequência da NS2B/NS3 do DENV1 e seus homólogos depositados no PDB.	85

Figura 18. Sobreposição estrutural dos <i>templates</i> usados na modelagem por homologia.	86
Figura 19. Modelos baseados em <i>templates</i> de DENV e WNV contribuíram para a construção de um terceiro modelo.	88
Figura 20. Variações de RMSD ao longo do tempo de simulação.	90
Figura 21. Análise da composição de estruturas secundárias ao longo da dinâmica	94
Figura 22. Agrupamento das famílias de conformações amostradas pelos modelos.	96
Figura 23. Mobilidade dos resíduos de cada modelo para os 5 primeiros <i>eigenvectors</i> identificados por PCA.	97
Figura 24. Modelos de cada sistema da protease do vírus da dengue, assim como suas projeções vetoriais de movimento contidas no primeiro <i>eigenvector</i>	99
Figura 25. Forças de interação entre ligante e proteína ou ligante e água.	100
Figura 26. Média das forças de interações entre resíduos do cofator ou da NS3 _{PRO} e cada componente do sistema.	101
Figura 27. Sítio de ligação do cofator e sua variação de volume durante a simulação.	123
Figura 28. Distribuição de escores das diferentes quimiotecas.	125
Figura 29. Processo de filtragem e seleção das moléculas destinadas à etapa de testes de inibição <i>in vitro</i>	126
Figura 30. Preferência dos compostos pelas diferentes conformações utilizadas na busca por inibidores do sítio ativo	128
Figura 31. Ensaios de inibição enzimática a partir dos compostos identificados por vHTS.	129
Figura 32. Poses dos compostos que apresentaram os melhores valores para atividade inibitória.	130
Figura 33. Interações entre o composto 1 e os resíduos de aminoácido da NS3 _{PRO}	131

Figura 34. Interações entre o composto 15 e os resíduos de aminoácido da NS3 _{PRO} . .	131
Figura 35. Interações entre o composto 21 e os resíduos de aminácido da NS3 _{PRO}	132
Figura 36. Interações entre o composto 26 e os resíduos de aminoácido da NS3PRO.	
.....	132
Figura 37. Alinhamento entre a sequência da TcMTAP e de estruturas homólogas depositadas no PDB.	136
Figura 38. Alinhamento da sequência da TcCatB e suas estruturas homólogas depositadas no PDB.	137
Figura 39. Sobreposição das estruturas homólogas utilizadas como guias na construção dos modelos TcMTAP e TcCatB.	138
Figura 40. Modelos da TcCatB e TcMTAP obtidas a partir de estruturas homólogas. ...	139
Figura 41. RMSD entre estruturas de cada quadro da simulação da TcMTAP.	140
Figura 42. RMSD entre estruturas de cada quadro da simulação da TcCatB.	141
Figura 43. Pose e modo de ligação do composto melhor ranqueado na campanha de vHTS contra a TcMTAP.	142
Figura 44. Poses e modos de ligação dos ligantes melhor ranqueados obtidos após vHTS contra a TcCatB.	144

LISTA DE TABELAS

Tabela 1. Principais campos de força e suas utilizações em dinâmicas moleculares.	51
Tabela 2. DNA molde e oligonucleotídeos das reações de PCR.	60
Tabela 3. Códigos de acesso das sequências utilizadas.	68
Tabela 4. Identidade e similaridade de estruturas homólogas à NS3 de DENV1 depositadas no PDB.	84
Tabela 5. Valores de RMSD entre as estruturas <i>template</i> e a estrutura de referência 3L6P.	86
Tabela 6. Compostos selecionados para os testes de inibição <i>in vitro</i>	127
Tabela 7. Identidade e similaridade entre a TcMTAP e estruturas homólogas depositadas no PDB.	135
Tabela 8. Identidade e similaridade entre a TcCatB e estruturas homólogas depositadas no PDB.	135

LISTA DE ABREVIACÕES

AMC	7-Amino-4-metilcumarina
BLAST	<i>Basic Local Alignment Search Tool</i>
DC	Doença de Chagas
DENV 1-4	<i>Dengue virus</i> sorotipos 1 a 4
DH	Dengue hemorrágica
DM	Dinâmica molecular
DMSO	Dimetilsulfóxido
DN	Doenças Negligenciadas
dNTP	Desoxirribonucleotídeos fosfatados
EDTA	Ácido etilenodiamino tetra-acético
HCV	Hepatitis C virus (vírus da hepatite C)
IAA	Álcool isoamílico
IPTG	Isopropil- β -D -1-tiogalactopiranosídeo
JEV	<i>Japanese encephalitis virus</i> (vírus da encefalite japonesa)
LB	Lysogeny Broth (meio de lisogenia), também chamado de meio Luria-Bertani
MVEV	<i>Murray Valley encephalitis virus</i> (vírus da encefalite de Murray Valley)
NDL	N-Benzoil-L-Norleucina-L-Lisil-L-Arginil-L-Arginina (inibidor peptídico da NS3 _{PRO})
NS 1-5	<i>Nonstructural protein</i> (proteína não-estrutural 1 a 5)
NS2B _{CF}	Região do loop hidrofílico da proteína NS2B que age como cofator para a NS3 _{PRO}
NS3 _{PRO}	Domínio de protease da proteína NS3
OPBTc	Oligopeptidase B de <i>T. cruzi</i>
ORF	<i>Open reading frame</i> (fase aberta de leitura)
P1'-4'	Posições de resíduos do substrato peptídico de 1 a 4 posteriores ao sítio de clivagem enzimática (C-terminais)
P4-1	Posições de resíduos do substrato peptídico de 4 a 1 anteriores ao sítio de clivagem enzimática (N-terminais)

PAGE	Eletroforese em gel de poliacrilamida
PCA	<i>Principal component Analysis</i> (Análise de Componentes Principais)
PDB	<i>Protein DataBank</i> (pode-se referir ao formato de arquivo ou ao banco de dados)
POPTc80	Proliloligopeptidase de <i>T. cruzi</i>
RE	Retículo endoplasmático
RMSD	<i>Root mean square deviation</i> (desvio médio quadrático)
S1'-S4'	Subsítios do sítio de ligação responsáveis pelas interações com os resíduos P1'-4' dos substratos peptídicos
S4-S1	Subsítios do sítio de ligação responsáveis pelas interações com os resíduos P1-4 dos substratos peptídicos
SCD	Síndrome do choque da dengue
SDS	Dodecil sulfato de sódio
TcCatB	Catepsina B de <i>Trypanosoma cruzi</i>
TcMTAP	Metiltioadenosina fosforilase de <i>T. cruzi</i>
vHTS	Virtual Highthroughput Screening (varredura virtual de alto rendimento)
VMD	<i>Visual Molecular Dynamics</i>
WNV	<i>West Nile virus</i> (vírus da febre do Nilo ocidental)
X-gal	5-bromo-4-cloro-indolil- β -D-galactopiranosídeo
YFV	<i>Yellow fever virus</i> (vírus da febre amarela)

SUMÁRIO

1	Introdução	17
1.1	Doenças negligenciadas	17
1.1.1	Dengue	18
1.1.2	Doença de Chagas	25
1.2	Medicamentos para doenças negligenciadas.....	33
1.3	Etapas de desenvolvimento do medicamento	34
1.3.1	Pesquisa básica.....	35
1.3.2	Testes pré-clínicos.....	36
1.3.3	Testes clínicos	36
1.4	Descoberta racional de fármacos.....	38
1.4.1	Triagem virtual	41
1.5	Modelagem por homologia.....	48
1.6	Dinâmicas moleculares (DMs).....	49
1.6.1	Princípios	49
1.6.2	Minimização de energia	51
1.6.3	Métodos estocásticos – o exemplo do método de Monte Carlo.....	53
1.6.4	Método determinístico – o exemplo da Dinâmica Molecular	53
1.6.5	Dinâmica Molecular e flexibilidade de proteínas.....	54
2	Justificativa.....	57
3	Objetivo	58

3.1	Objetivos específicos	58
4	Material e Métodos	59
4.1	Clonagem, expressão e purificação da NS2B _{CF} -Gly-NS3 _{PRO}	59
4.1.1	Amplificação e construção do complexo NS2B _{CF} -Gly-NS3 _{PRO}	59
4.1.2	Clonagem nos vetores pGEM®-T easy e pET-19b	61
4.1.3	Expressão em bactérias BL21(DE3) e BL21-AI™	63
4.1.4	Purificação por cromatografia em coluna de Ni ²⁺	64
4.2	Atividade e inibição enzimática	66
4.2.1	Ensaio de atividade enzimática	66
4.2.2	Ensaio de inibição enzimática	66
4.3	Modelagem por Homologia, Dinâmica e Ancoragem Molecular	67
4.3.1	BLAST/Sequências	67
4.3.2	Alinhamento	68
4.3.3	Sobreposição de estruturas	69
4.3.4	Modelagem por homologia	69
4.3.5	Dinâmicas moleculares	69
4.3.6	Ancoragem molecular (docking)	72
5	Resultados e Discussão	74
5.1	Dengue	74
5.1.1	Clonagem e Expressão	76
5.1.2	Modelagem por Homologia	83
5.1.3	Dinâmicas moleculares	89

5.1.4	Manuscrito	102
5.1.5	Acoplamento molecular automatizado.....	123
5.2	Doença de Chagas	134
5.2.1	Modelagem por homologia.....	134
5.2.2	Dinâmicas moleculares	139
5.2.3	Acoplamento molecular automatizado.....	141
6	Conclusões gerais	146
7	Referências Bibliográficas	148

1 Introdução

1.1 Doenças negligenciadas

As doenças negligenciadas (DNs) compõem um conjunto de enfermidades prevalentes nas regiões tropicais, sobretudo em países pobres, onde são responsáveis por altos índices de morbimortalidade. Por atingirem majoritariamente populações marginalizadas, essas doenças contribuem para a manutenção do quadro de desigualdade e exclusão social, principalmente devido aos seus impactos na saúde infantil, na redução da produtividade da classe trabalhadora e na promoção do estigma social (WERNECK *et al.*, 2011). Nesse sentido, essas doenças não só ocorrem com mais frequência em regiões empobrecidas, como também são condições promotoras da pobreza (HOTEZ *et al.*, 2006).

O termo “doenças negligenciadas” foi introduzido pela *The Rockefeller Foundation* em 1977, como sinônimo de doenças ignoradas pelos estabelecimentos de pesquisa biomédica dos países desenvolvidos (THE ROCKEFELLER FOUNDATION, 1978). Já em 2001, a organização não-governamental Médicos Sem Fronteiras (MSF), juntamente com a Organização Mundial de Saúde (OMS), propôs a classificação das doenças em Globais, Negligenciadas e Mais Negligenciadas (MÉDECINS SANS FRONTIÈRES, 2001a). Atualmente, o adjetivo “negligenciadas” indica tanto o descaso das grandes empresas multinacionais, que não obteriam lucro desenvolvendo medicamentos contra doenças que atingem principalmente populações de baixa renda, como a falta de políticas públicas de saúde visando a eliminação e controle dessas doenças (SOUZA, 2010; TROUILLER *et al.*, 2002; WERNECK *et al.*, 2011; YAMEY, 2002).

No Brasil, grande parte do contingente de 40 milhões da população mais pobre está infectada com uma ou mais doenças negligenciadas (HOTEZ, 2008). Essas doenças

perpetuam-se devido a falhas em três grandes áreas do desenvolvimento: “falha de ciência”, quando há escassez de conhecimentos; “falha de mercado”, quando medicamentos ou vacinas existem, porém, a um preço proibitivo; e “falha de saúde pública”, quando medicamentos baratos ou mesmo gratuitos existem, mas não são utilizados devido a um planejamento deficiente (MOREL, 2006). Em um cenário ideal, a primeira falha deveria ser suprida pela pesquisa das instituições de ensino superior (IES), a segunda pelas parcerias público-privadas estabelecidas entre governo federal e indústria farmacêutica, e a terceira pelo governo local. Sendo os detentores de uma patente os grandes responsáveis pelo preço final de um medicamento, a iniciativa das IES na pesquisa de base para o desenvolvimento de novos fármacos é de extrema importância, uma vez que, detendo os direitos sobre a utilização de moléculas, pode controlar e garantir a distribuição dos medicamentos por preços mais baixos à população (MÉDECINS SANS FRONTIÈRES, 2001b).

Dentre as DNs de maior interesse no cenário brasileiro, podemos citar a febre da dengue e a doença de Chagas, além das leishmanioses, hanseníase, filariose linfática, oncocercose, helmintíases transmitidas pelo solo, tracoma e raiva (LINDOSO e LINDOSO, 2009; WERNECK *et al.*, 2011). Atenção especial será dada às duas primeiras, já que pertencem ao escopo deste trabalho.

1.1.1 Dengue

A dengue é uma doença infecciosa causada por quatro sorotipos diferentes do *Dengue vírus* (DENV 1-4), transmitida por mosquitos do gênero *Aedes*. Enquanto que a OMS estima que a doença afete entre 50-100 milhões de pessoas, e que outras 2,5 bilhões estejam em risco (OMS, 2012a), para Beatty e colaboradores esse número é subestimado, podendo chegar a mais de 500 milhões de pessoas afetadas anualmente, com 3,6 bilhões de pessoas em risco nos 124 países onde está distribuída (BEATTY *et al.*, 2010). Além de ser um risco

para pessoas que vivem em regiões tropicais e subtropicais, a dengue é atualmente a doença febril mais comum em pessoas que viajam para essas regiões, ultrapassando a malária e infecções gastrintestinais (SCHWARTZ, 2008).

1.1.1.1 Transmissão e manifestações clínicas

O DENV é transmitido por diversas espécies de *Aedes*, principalmente o *Aedes aegypti* e o *Aedes albopictus*, mosquitos que possuem susceptibilidade e capacidade tanto de veicular horizontalmente, como de transmitir verticalmente (por vias transovarianas) os quatro sorotipos do vírus (MONCAYO *et al.*, 2004). Os hospedeiros vertebrados são o homem e alguns primatas silvestres, que servem de reservatório para o vírus.

A infecção pelo DENV pode levar a uma gama de manifestações clínicas, passando de infecções assintomáticas até casos mais graves de dengue hemorrágica (DH) ou síndrome do choque da dengue (SCD) que podem levar à morte. Vários fatores de risco às formas mais severas da doença já foram estabelecidos, incluindo infecções recorrentes por outros sorotipos (BURKE *et al.*, 1988) e infecções sorotipo- ou genótipo-específicas, entre outros (HALSEY *et al.*, 2012). Manifestações mais brandas incluem febre alta, erupções cutâneas, cefaleia, mialgia, artralgia, dor retro-orbital, fadiga e náuseas (HALSTEAD, 2007; ZELLWEGER *et al.*, 2010), e caracterizam a dengue clássica. Sintomas mais graves se manifestam por encefalite, hepatite, cardiomiopatia e manifestações hemorrágicas (HALSTEAD, 2007).

Apesar de sua alta incidência, severidade e impacto econômico, atualmente não há tratamentos antivirais ou vacinas aprovadas para a dengue. O desenvolvimento de uma vacina anti-dengue tem o desafio de prover proteção para todos os quatro sorotipos ao mesmo tempo (DURBIN e WHITEHEAD, 2010; WHITEHEAD *et al.*, 2007), caso contrário, ela poderia tornar indivíduos imunizados mais suscetíveis à DH (KONISHI, 2011). O protótipo

de vacina em estágio mais avançado, a CYD-TDV (Sanofi-Pasteur), é uma vacina tetravalente atenuada, baseada na vacina de febre amarela. Em estudos clínicos de fase IIb, essa vacina apresentou uma eficiência de apenas 30%, falhando em conferir imunização contra o sorotipo DENV-2 (SABCHAREON *et al.*, 2012).

1.1.1.2 Distribuição geográfica e epidemiologia

Atualmente, casos autóctones de dengue podem ser encontrados em quase todos os continentes (Figura 1) (BEATTY *et al.*, 2010).

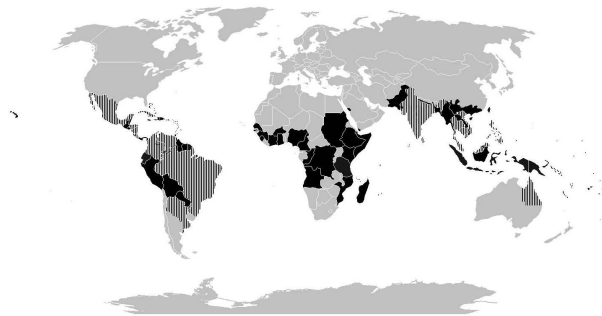


Figura 1. Países com transmissão local da dengue nos últimos 25 anos. As áreas hachuradas correspondem aos locais onde ocorre o maior número de casos, enquanto que as áreas escuras correspondem às áreas onde o risco de dengue existe e um número menor de casos foram relatados. Fonte: (BEATTY *et al.*, 2010).

A expansão em andamento dos habitats dos mosquitos transmissores da dengue (CAMINADE *et al.*, 2012), devido tanto às mudanças climáticas recentes quanto à urbanização de países em desenvolvimento (GUBLER, 2011), tem alertado os centros de controles de doenças internacionais. Casos autóctones na Europa (RUCHE *et al.*, 2010), e surtos recentes nos Estados Unidos (BOURI *et al.*, 2012; FRANCO *et al.*, 2010) demonstram que a dengue já não é mais um problema exclusivo dos países em desenvolvimento. O alastramento da dengue evidencia como a globalização de negócios (com transporte de mosquitos vetores), o aumento de viagens nacionais e internacionais (e movimentação de

indivíduos infectados), a aglomeração urbana, e estratégias ineficazes no combate aos vetores resultaram em uma pandemia na era moderna (SIMMONS *et al.*, 2012).

De acordo com os dados do Sistema de Informação de Agravos de Notificação (SINAN)¹, em 2012 o Brasil apresentou 269.356 casos de dengue clássica, 687 de DH e 29 de SCD, com 85 óbitos no total. A epidemiologia no país varia a cada ano, e está diretamente relacionada às medidas de prevenção da proliferação do vetor *A. aegypti*. É comum após um surto o governo investir em prevenção e divulgação na mídia, o que faz com que no ano seguinte a campanha de prevenção seja geralmente favorável. O Rio de Janeiro, por exemplo, após passar por um surto da doença em 2008, apresentou uma redução de 96% no número de casos suspeitos (SVS/MS, 2010a) em 2009. No entanto, no mesmo período, a região Centro-Oeste registrou um aumento de 92% na incidência dos casos, principalmente nos estados do Mato Grosso e Mato Grosso do Sul. Já o Distrito Federal, que nos anos de 2008 e 2009 apresentou baixa incidência, em 2010 apresentou um aumento de mais de 725% no número de casos notificados (SVS/MS, 2010b).

1.1.1.3 O vírus da dengue

Os DENVs pertencem ao gênero *Flavivirus* da família Flaviridae, um grupo de pequenos arbovírus envelopados de RNA fita-simples positiva (ssRNA⁺). Outros representantes do gênero incluem o *Japanese Encephalitis Virus* (JEV), *West Nile Virus* (WNV), *Yellow Fever Virus* (YFV), *Murray Valley Encephalitis Virus* (MVEV). Dentre estes, todos são responsáveis por endemias e são transmitidos por vetores invertebrados (VASILAKIS e WEAVER, 2008).

O RNA genômico do DENV possui entre 10,5 e 11 quilobases (kb), que codificam uma única fase aberta de leitura (ORF, do inglês *open reading frame*), correspondente a 95% de seu genoma. Durante a infecção (figura 2), o RNA genômico do DENV é traduzido em uma

¹ Disponível em: <http://dtr2004.saude.gov.br/sinanweb/> (acesso em setembro/2012)

poliproteína, que, em associação às membranas do retículo endoplasmático, é hidrolisada por proteases celulares e virais em dez produtos gênicos (figura 3): três componentes estruturais (capsídeo - C, precursor de membrana - prM, e envelope - E) e sete não estruturais (NS1, NS2A/B, NS3, NS4A/B, NS5) (CHAMBERS *et al.*, 1990; RYAN *et al.*, 1998).

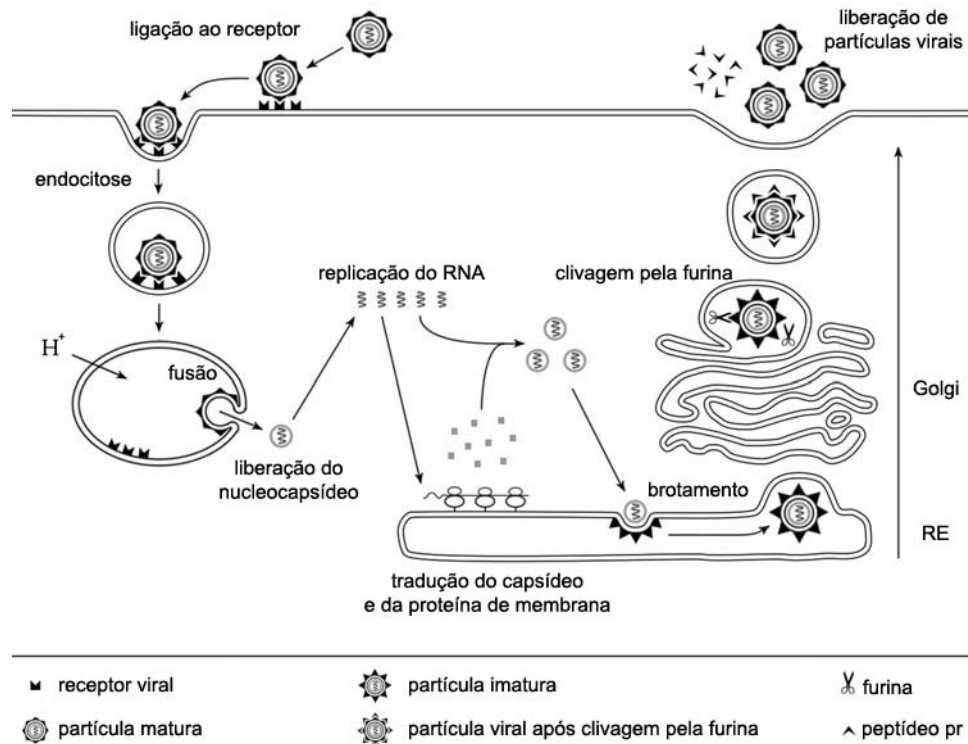


Figura 2. Ciclo de replicação do DENV. O ciclo de replicação viral começa com a entrada do vírus na célula hospedeira, mediada por receptores ainda não elucidados. A acidificação do endossomo leva à fusão de membranas, e, conseqüentemente, ao desencapsulamento. A liberação do nucleocapsídeo no citoplasma é seguida pela tradução da poliproteína viral, transportada para o retículo onde passará por um processamento co- e pós-traducional. Os nucleocapsídeos são montados no citoplasma. Em seguida, seguem para a membrana do RE, onde ocorre o brotamento interno. A partícula imatura está completamente montada, e segue para o complexo de Golgi. Mudanças conformacionais dependentes de pH geradas nas proteínas E e prM possibilitam a clivagem do peptídeo prM pela furina, gerando um peptídeo pr e a proteína M. Esse peptídeo permanece associado, até a liberação da partícula viral infectiva. Adaptado de (RODENHUIS-ZYBERT *et al.*, 2010).

Apenas as proteínas estruturais se tornam parte da partícula viral madura, enquanto que as proteínas não-estruturais estão envolvidas no processamento da poliproteína, síntese do RNA viral, e na morfogênese do vírus (ASSENBERG *et al.*, 2009).

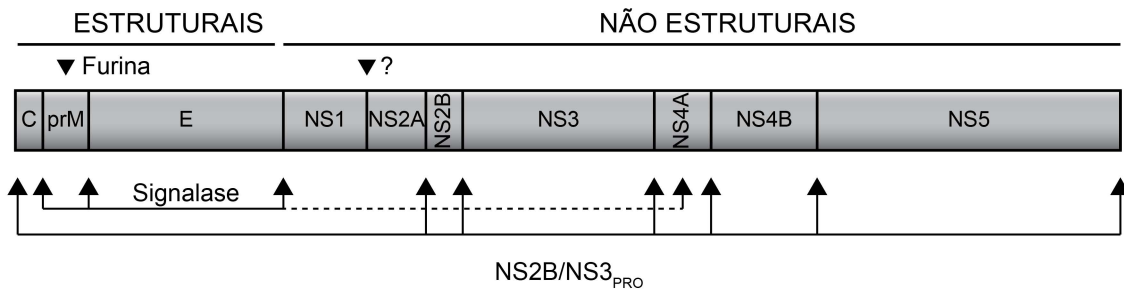


Figura 3. Organização da poliproteína do DENV e seus sítios de proteólise. Voltada para o citoplasma, a protease NS2B/NS3_{PRO} é responsável pelas clivagens N- e C-terminais da poliproteína, além das junções NS2A/NS2B, NS2B/NS3, NS3/NS4A, NS4A/NS4B e NS4B/NS5. No lúmen do RE, a sinalase é responsável pela clivagem nas junções C/prM, prM/E, E/NS1 e uma clivagem interna na região NS4A. Durante a maturação viral no complexo de Golgi, a furina é responsável pela clivagem entre o peptídeo pr e a proteína M. A protease responsável pela clivagem da junção NS1/NS2A ainda não foi identificada.

Durante a tradução do RNA genômico do DENV no citoplasma, o peptídeo precursor da poliproteína é direcionado por pequenas sequências-sinal para o retículo endoplasmático (RE), onde a NS1 e os domínios exógenos da prM e E ficam voltadas para o lúmen, enquanto que a C, NS3 e NS5 se voltam para o citoplasma. NS2A, NS2B, NS4A e NS4B são proteínas transmembrânicas altamente hidrofóbicas, com pequenos fragmentos hidrofílicos (ASSENBERG *et al.*, 2009).

O processamento da poliproteína é requerido para a replicação do vírus, e ocorre de forma co- e pós-traducional (CHAMBERS *et al.*, 1990; FALGOUT e MARKOFF, 1995; FALGOUT *et al.*, 1989, 1991; HORI e LAI, 1990; KUHN *et al.*, 2002). Juntamente com algumas proteases do hospedeiro, a proteína não estrutural 3 (NS3) é uma das responsáveis por esta etapa, e, portanto, um alvo potencial para o tratamento da dengue (LESCAR *et al.*, 2008). Das proteínas não-estruturais, a NS3 e a NS5, ambas proteínas multifuncionais, são melhores caracterizadas. O domínio N-terminal da NS3 contém a protease viral (NS3_{PRO}) (CHAMBERS *et al.*, 1990), que requer como cofator um *loop* hidrofílico da NS2B (NS2B_{CF}) (FALGOUT *et al.*, 1993). Seu domínio C-terminal possui um domínio de helicase (NS3_{HEL}) (WENGLER, 1991). A NS5, por sua vez, possui um domínio de metiltransferase N-terminal e um domínio de RNA-polimerase C-terminal (EGLOFF *et al.*, 2002; TAN *et al.*, 1996). Os

papéis da NS1, NS2A, NS4a e NS4B ainda não estão totalmente elucidados (ASSENBERG *et al.*, 2009).

1.1.1.4 Protease NS3

A NS3_{PRO} (resíduos 1-185 da NS3) é uma serinoprotease tripsina-símile pertencente à família S7, com a tríade catalítica característica dessas enzimas (His51-Asp75-Ser135). Entretanto, ao contrário da tripsina, essa protease possui preferência por resíduos dibásicos (por exemplo, Arg e Lys nas posições P1 e P2), e requer o domínio hidrofílico da NS2B como cofator para a catálise efetiva da poliproteína viral (FALGOUT *et al.*, 1993).

Como dito anteriormente, devido à sua participação fundamental na replicação viral, a NS3_{PRO} constitui um alvo promissor na busca de novas terapias contra a dengue. Formas recombinantes da enzima ativa foram expressas em bactéria *E. coli* (CHAMPREDA *et al.*, 2000; GOUVEA *et al.*, 2007; LEUNG *et al.*, 2001), sua atividade bioquímica foi descrita e comparada com as proteases de outros flavivírus (BERA *et al.*, 2007; GOUVEA *et al.*, 2007; MELINO e PACI, 2007; SHIRYAEV *et al.*, 2007), e seu substrato otimizado foi identificado (LI *et al.*, 2005). Além disso, sua estrutura tridimensional foi elucidada juntamente com seu cofator (NS2B_{CF}) (CHANDRAMOULI *et al.*, 2010; ERBEL *et al.*, 2006), e, mais recentemente com dois inibidores distintos (NOBLE *et al.*, 2012) (figura 4).

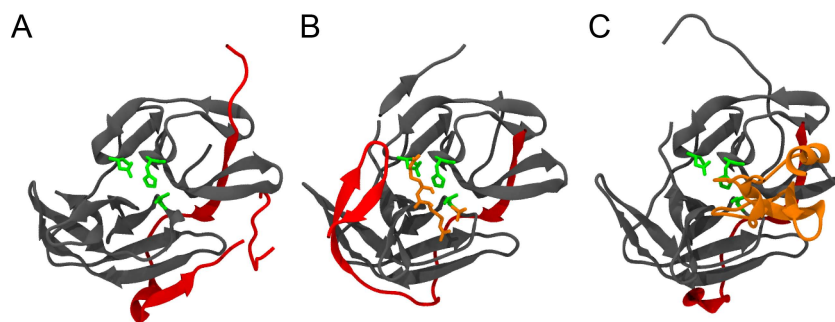


Figura 4. Estruturas cristalográficas da DENV NS3B_{CF}-NS3_{PRO}. A) Apoestrutura (sem ligante – 2FOM). B) NS2B_{CF}-NS3_{PRO} ligada ao inibidor NDL (3U1I). C) NS2B_{CF}-NS3_{PRO} ligada à aprotinina (3U1J). O domínio NS3_{PRO} está representado em cinza, NS2B_{CF} em vermelho, inibidores em laranja e os resíduos catalíticos em verde.

Apesar da busca de inibidores contra a NS3_{PRO} de DENV ser uma linha de pesquisa ativa na luta contra a dengue (CREGAR-HERNANDEZ *et al.*, 2011; NITSCHKE *et al.*, 2011, 2012; SCHÜLLER *et al.*, 2011; TIEW *et al.*, 2012; TOMLINSON e WATOWICH, 2012), poucos avanços foram feitos até o momento (DENG *et al.*, 2012). A flexibilidade da NS2B_{CF}/NS3_{PRO}, como comprovada pelas diferenças entre as estruturas cristalográficas depositadas no *Protein Data Bank* (PDB) (ASSENBERG *et al.*, 2009; CHANDRAMOULI *et al.*, 2010; ERBEL *et al.*, 2006; LUO *et al.*, 2008; NOBLE *et al.*, 2012), pode ser em parte responsável pelos resultados abaixo do esperado nas campanhas de desenho racional de novas drogas (SCHÜLLER *et al.*, 2011). Esse fato deixa aberta a possibilidade de exploração da notável flexibilidade dessa enzima em novas investidas para o descobrimento de moléculas promissoras no combate à dengue.

1.1.2 Doença de Chagas

A doença recebeu o nome do pesquisador brasileiro responsável por sua descoberta, o doutor Carlos Ribeiro Justiniano Chagas, que ainda identificou seus vetores, agente etiológico e descreveu suas características clínicas (MOLINA, 2001). Também conhecida como tripanossomíase americana humana, a doença de Chagas (DC) é causada pelo protozoário hemoflagelado *Trypanosoma cruzi*. Os barbeiros, insetos hematófagos da família Reduviidae, subfamília Triatominae, sobretudo as espécies *Triatoma infestans*, *Rhodnius prolixus* e *Triatoma dimidiata*, são os principais transmissores ao homem (DIOTAIUTI, 2009; ZELEDÓN e RABINOVICH, 1981). A OMS estima que aproximadamente 10 milhões de pessoas estejam infectadas com o *T. cruzi*, e que até 30% dos infectados crônicos possam desenvolver complicações cardíacas, além de 10% também estarem sujeitos a alterações digestivas, neurológicas ou mistas (OMS, 2012b). A doença se tornou endêmica após a domesticação dos vetores selvagens, algo que ocorreu em torno de 200 a 300 anos atrás, por meio da expansão da agricultura e pecuária (COURA, 2007). Atualmente, assim como

acontece com a dengue e outras doenças tropicais, a epidemiologia da DC está passando por mudanças importantes devido à globalização: por se tratar de uma doença crônica, a migração de pessoas infectadas aumenta a distribuição de reservatórios do parasita. Ao mesmo tempo, a importação dos vetores triatomíneos por vias marítimas ou aéreas, devido ao comércio internacional, aumentam os riscos de transmissão fora da América Latina (CASTRO *et al.*, 2009; COURA e VIÑAS, 2010).

1.1.2.1 Transmissão e manifestações clínicas

Além da via entomológica (vetorial), responsável por 80-90% dos casos (CASTRO *et al.*, 2006; YOSHIDA, 2009), outras formas de transmissão da DC incluem: transfusão de sangue ou órgãos de pacientes infectados, transmissão congênita vertical pela placenta, e a via oral (CARDOSO *et al.*, 2006; DIAS, 2006; YOSHIDA, 2009). As transmissões por transplantes preocupam as autoridades de países não endêmicos, como Austrália, Espanha, Canadá e EUA (CASTRO *et al.*, 2009; SCHMUNIS, 2007).

A infecção ocorre durante o repasto do barbeiro contaminado pelo *T. cruzi*, que, ao alimentar-se, libera formas infectantes (metacíclicas) do parasita em suas fezes, que posteriormente entram em contato com as mucosas ou com a lesão do sítio da picada.

Após a entrada do parasito na corrente sanguínea, o indivíduo contaminado pode passar por três fases da doença (aguda, indeterminada ou latente, e crônica), com sintomas diferentes em cada uma delas. A fase aguda é caracterizada por alta parasitemia nos tecidos e no sangue do indivíduo acometido, e é na maioria das vezes assintomática. Quando ocorrem, os sintomas dessa fase podem passar despercebidos, pois se assemelham aos de outras doenças (febre, mialgia e mal-estar). No caso de transmissão vetorial, pode haver uma reação inflamatória com edema no local da picada, apelidada de chagoma ou de sinal de Romaña, caso atinja os olhos. Se o indivíduo não receber tratamento, os sintomas da fase

aguda podem desaparecer naturalmente em um período de 3 a 8 semanas, após o qual a doença pode entrar na fase latente ou indeterminada. Essa fase é caracterizada pela ausência dos sintomas da fase crônica da doença, e normalmente dura de 10 a 20 anos, afetando cerca de 50% dos pacientes chagásicos (MACEDO, 1980).

Cerca de 30% dos acometidos pela DC desenvolvem a fase crônica, com sintomas cardíacos, digestivos, neurológicos ou mistos. Nessa fase, disfunções do sistema nervoso autônomo podem gerar arritmias cardíacas, além da síndrome dos megas (megacólon e megaesôfago) (PY, 2011). Outros sintomas importantes são as cardiopatias graves, como miocardite e infarto do miocárdio, provenientes de lesões inflamatórias crônicas que resultam na hipertrofia e disritmia cardíaca (COURA, 2007).

O benzonidazol, uma droga utilizada na quimioterapia da doença de Chagas, além de ser tóxica, não impede o desenvolvimento e nem elimina o parasito durante a fase crônica da doença (CALDAS *et al.*, 2008). Este cenário indica a necessidade de se buscar novas alternativas para o tratamento desta doença.

1.1.2.2 Distribuição geográfica e epidemiologia

Originalmente, a DC estava confinada às áreas rurais mais pobres das Américas do Sul e Central, onde a transmissão vetorial ao homem ocorre. Entretanto, casos autóctones foram identificados nos Estados Unidos (BERN *et al.*, 2011; CASTRO *et al.*, 2009). Além disso, devido ao aumento na mobilidade entre a população da América Latina e o resto do mundo, a doença vem sendo cada vez mais detectada nos EUA, Canadá, muitos países da Europa e em alguns do Pacífico Ocidental (OMS 2012b; COURA e VIÑAS, 2010; RASSI e MARIN-NETO, 2010). O centro de controle e prevenção de doenças dos EUA (CDC) estima que cerca de 300 mil pessoas infectadas com o *T. cruzi* vivam atualmente dentro de suas fronteiras (COURA e VIÑAS, 2010). Os levantamentos obtidos pela OMS em 2002 e 2007 levantam

discussões válidas a respeito da situação atual nas Américas. Os números em 2002 indicavam que havia aproximadamente 21 mil mortes/ano, enquanto que nas estimativas de 2007 esses números caíram para 12,5 mil mortes/ano. Sabendo que a causa de morte mais frequente na DC são suas complicações da fase crônica, a queda observada nesses números não poderia ser atribuída às campanhas de desalojamento do principal vetor doméstico, como sugerido (MONCAYO, 2003; OMS, 2002; 2007). Espera-se, portanto, que as divergências entre as estimativas venham de diferenças nos métodos de realização desses levantamentos.

1.1.2.3 O *Trypanosoma cruzi*

O *Trypanosoma cruzi* é um protozoário flagelado da família Trypanosomatidae, ordem Kinetoplastidae. Esta família inclui o grupo de protozoários pertencentes aos gêneros *Trypanosoma* e *Leishmania*, de grande importância para a medicina humana e veterinária (TEIXEIRA *et al.*, 2006) Apesar da alta variabilidade genética entre suas linhagens, o *T. cruzi* é atualmente classificado em uma única espécie (ZINGALES *et al.*, 2009). Seu ciclo de vida alterna entre os hospedeiros vertebrados e invertebrados, onde, de acordo com seus aspectos morfológicos e bioquímicos, o parasito apresenta quatro formas celulares distintas (FERNANDES e ANDREWS, 2012) Dentre estas, as formas amastigotas e tripomastigotas sanguíneas podem ser encontradas no hospedeiro vertebrado, enquanto que as formas epimastigotas e tripomastigotas metacíclicas nos insetos vetores (figura 5).

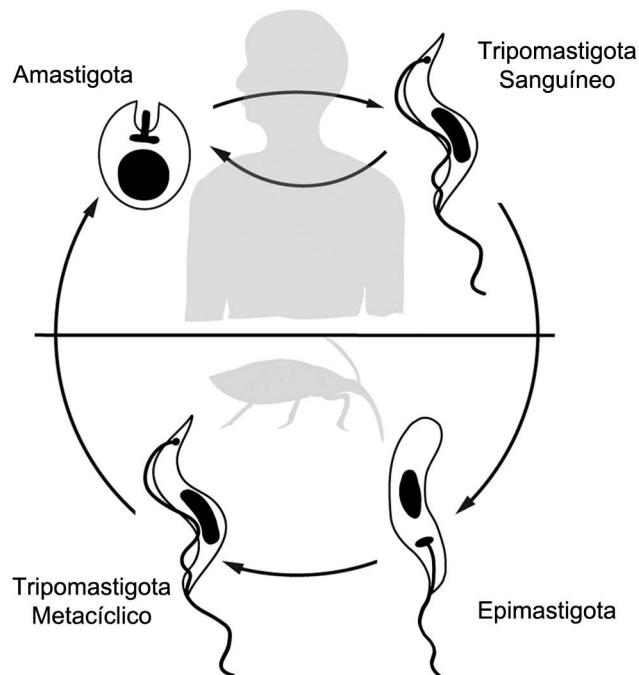


Figura 5. Formas de vida do *T. cruzi* e seu ciclo de vida simplificado. As formas epimastigotas e tripomastigotas metacíclicas são encontradas no inseto vetor; formas amastigotas e tripomastigotas sanguíneas no hospedeiro mamífero. Adaptado de (ALVAREZ *et al.*, 2012).

O ciclo de vida completo do protozoário se inicia quando o barbeiro se alimenta do sangue de um hospedeiro infectado. No intestino médio do inseto, o protozoário se transforma em epimastigota e se multiplica. Em seguida, migram para o intestino posterior, onde passam por uma segunda transformação em tripomastigotas metacíclicas. Durante o novo repasto do inseto vetor, as formas metacíclicas eliminadas em suas fezes infectam o hospedeiro pela mucosa ou pelo sítio da picada. As tripomastigotas invadem as células próximas ao local de infecção por meio de interações que envolvem receptores e moléculas de superfícies específicas tanto no parasito quanto no hospedeiro (SCHENKMAN *et al.*, 1991; YOSHIDA e CORTEZ, 2008). Essas interações culminam na mobilização de Ca^{+2} em ambas as células e, conseqüentemente, na formação da vesícula parasitófora com recrutamento de lisossomos (BURLEIGH *et al.*, 1998; DOCAMPO e MORENO, 1996; TARDIEUX *et al.*, 1992; YOSHIDA, 2006). Mediante fusão de lisossomos com a vesícula parasitófora, as tripomastigotas diferenciam-se em formas amastigotas, evento dependente da acidificação e atividade proteassômica no fagolisossomo (HALL *et al.*, 1992). Após cerca de nove ciclos de

replicação (DVORAK e HYDE, 1973), as amastigotas preenchem o citosol da célula hospedeira. Antes de serem liberadas na corrente sanguínea, passam a se transformar novamente em tripomastigotas. Logo após, começam novos ciclos de infecção em vários tecidos do corpo. Ocasionalmente podem ser sugadas pelo inseto triatomíneo e transmitidas ao próximo hospedeiro (TEIXEIRA *et al.*, 2011).

1.1.2.4 Alvos promissores

Devido à ausência de uma vacina e de tratamentos satisfatórios, a pesquisa por novos fármacos para o tratamento da DC tem evoluído sensivelmente nas últimas décadas (DIAS *et al.*, 2009). Por meio do sequenciamento dos genomas do *T. cruzi*, *T. brucei* e *Leishmania sp.*, vários genes existentes apenas no parasito, e não no homem, foram identificados (TIMMERS *et al.*, 2009). Muitos constituem alvos biológicos promissores, sendo a maior parte enzimas (DUSCHAK e COUTO, 2007; MCKERROW *et al.*, 2008; SOEIRO e CASTRO, 2009).

Dentre eles, alguns já foram ou estão sendo estudados pelo grupo do professor Jaime Santana e colaboradores. Podemos citar a catepsina B (TcCatB) (GARCIA *et al.*, 1998; NÓBREGA *et al.*, 1998), a oligopeptidase B (OPBTc) (MOTTA *et al.*, 2012; SANTANA *et al.*, 1992), a prolil-oligopeptidase (POPTc80) (BASTOS *et al.*, 2005; SANTANA *et al.*, 1997), e a metiltioadenosina fosforilase (TcMTAP) (NEVES, 2006).

1.1.2.4.1 Prolil oligopeptidase de *T. cruzi* (POPTc80)

A POPTc80 é uma serinoprotease com atividade máxima em pH neutro, secretada pelas formas infectantes do parasito e degrada colágeno nativo nos moldes das collagenases de bactérias invasivas (SANTANA *et al.*, 1997). Sua função foi conhecida por meio do emprego de inibidores específicos, obtidos por síntese química baseada na sequência do substrato e também por química combinatória. Concentrações micromolares inibem a entrada de tripomastigotas em células mamíferas não fagocíticas, mas não a ancoragem do parasito à

membrana celular, sugerindo um processo ativo de entrada mediado por transdução de sinal (BASTOS *et al.*, 2005; GRELLIER *et al.*, 2001). Modelagem molecular por homologia de sequência sugeriu um mecanismo novo de acesso de grandes substratos, no caso o colágeno, ao sítio ativo dessa classe de serinoproteases (figura 6). Este modelo propõe que os dois domínios da enzima se movimentem através de uma alça flexível, permitindo assim que o colágeno se posicione entre eles, próximo à bolsa catalítica. Em 2005, a estrutura da POP de *Sphingomonas capsulata* foi resolvida em configuração aberta, reforçando ainda mais essa hipótese (SHAN *et al.*, 2005).

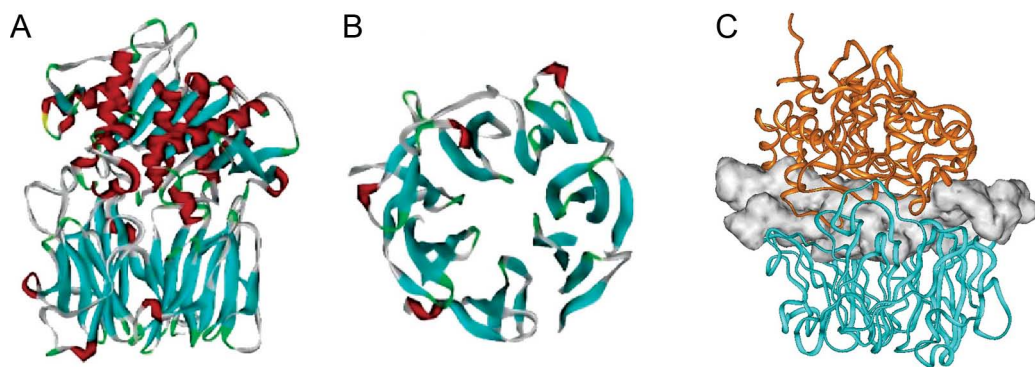


Figura 6. Modelo teórico da POP Tc80 obtido a partir de modelagem por homologia. A) e B) Visões diferentes do modelo POP Tc80. C) Estrutura do colágeno acoplada à POP Tc80. Fonte: (BASTOS *et al.*, 2005).

1.1.2.4.2 Oligopeptidase B de *T. cruzi* (OPBTc)

A OPBTc, descrita pela primeira vez por Santana e colaboradores (SANTANA *et al.*, 1992), gera um fator mobilizador de Ca^{2+} na célula hospedeira, possibilitando a migração de lisossomos para o sítio de ligação do parasito, facilitando sua entrada (BURLEIGH e ANDREWS, 1995; RODRIGUEZ, 1995). Parasitos nocauteados para este gene apresentam baixa capacidade de infecção *in vitro* e *in vivo* (CALER *et al.*, 1998). Dados mais recentes demonstram que essa enzima recombinante é dimérica e estável (MOTTA *et al.*, 2012). Não existe oligopeptidase B em mamíferos (VENÄLÄINEN *et al.*, 2004), o que é extremamente desejável para desenvolvimento racional de drogas.

1.1.2.4.3 *Catepsina B de T. cruzi (TcCatB)*

Bem conservada entre os Kinetoplastidae, a TcCatB é expressa em todas as formas de *T. cruzi* e hidrolisa uma grande variabilidade de substratos (GARCIA *et al.*, 1998). É uma enzima ácida, associada ao lisossomo, e muito provavelmente está associada à importante função de digestão e modelação de forma do parasito. Dados cinéticos indicam que medeia a principal atividade proteolítica ampla de *T. cruzi* (CADAVID-RESTREPO, 2005) ao lado da cruzipaina, um fator de virulência conhecido. Estudos de RNA de interferência com sua homóloga em *T. brucei* foram capazes de resgatar camundongos de uma infecção letal, sugerindo sua importância na persistência desse parasito (ABDULLA *et al.*, 2008). Ainda quanto sua homóloga em *T. brucei*, a TbCatB teve sua estrutura cristalográfica resolvida (KERR *et al.*, 2010), provendo dados suficientes para a criação de um modelo por homologia da TcCatB.

1.1.2.4.4 *Metiltioadenosina fosforilase de T. cruzi*

A TcMTAP cataliza a clivagem fosforolítica reversível da metiltioadenosina (um subproduto da síntese de poliaminas), produzindo metiltioribose-1-fosfato (MTRP) e adenina. O MTRP é reciclado para metionina e a adenina é direcionada para a síntese de nucleosídeos. Uma vez que os tripanosomos não apresentam vias de síntese *de novo* de purinas, e apresentam uma alta demanda de S-adenosilmetionina (AdoMet), um precursor metabólico para as poliaminas (HEBY *et al.*, 2007; PIÑEYRO *et al.*, 2011), a TcMTAP deve permitir o resgate de adeninas e metioninas. Além desse papel importante no metabolismo do parasito, outro fato leva a crer que a TcMTAP seja um promissor alvo de drogas: ao contrário da enzima homóloga em humanos, ela possui alta especificidade a nucleosídeos como a adenosina e seus análogos (GHODA *et al.*, 1988), fato que auxilia o desenvolvimento de inibidores e substratos subversivos específicos. A forma recombinante da TcMTAP já foi obtida ativa e caracterizada bioquímica- e biofísicamente (NEVES, 2006). O uso de

substratos tóxicos foi capaz de impedir a morte de camundongos infectados com *T. brucei* (BACCHI *et al.*, 1997). Além disso, apresenta expressão diferenciada entre formas distintas do protozoário, como verificado por análise proteômica (PARODI-TALICE *et al.*, 2007), sugerindo um papel na mudança das formas epimastigotas e tripomastigotas metacíclicas. O uso de inibidores dessa enzima em culturas de *T. cruzi* teve efeito tripanosomicida (NEVES, 2006).

1.2 Medicamentos para doenças negligenciadas

Apesar do altíssimo fardo das doenças negligenciadas, elas despertam pouco interesse das indústrias farmacêuticas, uma vez que a maioria da população afetada não possui poder aquisitivo suficiente para os tratamentos. Apesar de recentes esforços de iniciativas como o *Drugs for Neglected Diseases initiative* (DNDi), o número de tratamentos para doenças extremamente negligenciadas continua ínfimo em relação às doenças ditas “globais”, ou mesmo em relação aos tratamentos cosméticos (figura 7).

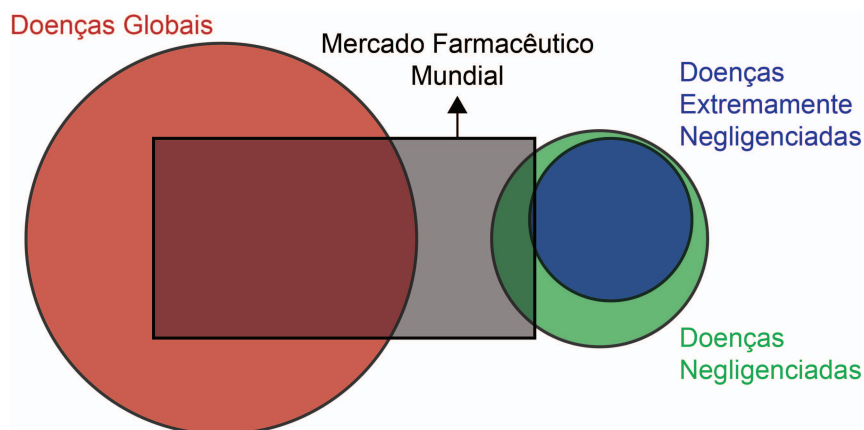


Figura 7. Parcela de doenças cobertas pelo interesse farmacêutico mundial. O retângulo sombreado corresponde ao mercado farmacêutico mundial. A parte do retângulo que não realiza nenhuma interseção corresponde aos produtos cosméticos e anti-calvície, por exemplo. A figura original foi adaptada de forma a incluir as Doenças Extremamente Negligenciadas no grupo de Doenças Negligenciadas, além de simbolizar a pequena parcela de mercado para as DENs desde que o trabalho original foi publicado em 2002. Adaptado de (YAMEY, 2002).

Das 1556 drogas desenvolvidas entre 1975 e 2004, apenas 21 foram desenvolvidas para as doenças negligenciadas, incluindo a tuberculose (CHIRAC e TORRELE, 2006). Boa parte desse número pode ser creditada às iniciativas do setor público, que criaram condições de

mercado favoráveis ao desenvolvimento dessas drogas, principalmente por meio de parcerias público-privadas (PPPs). O problema reside no fato de que para adentrar numa parceria desse tipo, as empresas farmacêuticas precisam de um estímulo. No caso de algumas doenças específicas, como a SIDA, malária e tuberculose, enquanto a população de países endêmicos, como Brasil, Índia e África não possa pagar pelos tratamentos, pessoas sujeitas a essas mesmas doenças em países desenvolvidos podem (YAMEY, 2002). No caso das outras doenças, o governo é responsável por convencer as grandes empresas farmacêuticas a unirem-se nessas PPPs, principalmente por meio da criação de incentivos fiscais e fundos de garantias.

Outra forma de incentivar as empresas farmacêuticas a aderirem às iniciativas do tipo PPP está nas mãos das instituições de ensino superior. Por meio da realização da pesquisa básica, responsável por grande parte dos gastos no desenvolvimento de um novo medicamento, o setor público ganha mais uma moeda para negociação. Ao oferecer às indústrias uma molécula altamente promissora no tratamento de uma doença negligenciada, o governo compartilha custos no desenvolvimento de um medicamento, tornando-o mais atrativo aos olhos das gigantes farmacêuticas.

O processo de desenvolvimento de um novo medicamento, assim como as técnicas de descoberta racional de fármacos, capazes de diminuir o tempo e os gastos nesse processo, e de aumentar as chances de obterem-se moléculas promissoras, será abordado a seguir.

1.3 Etapas de desenvolvimento do medicamento

O caminho percorrido durante a pesquisa de uma molécula para fins farmacêuticos é longo, dispendioso e incerto. De acordo com a PhRMA (*Pharmaceutical Research and Manufacturers of America*), dentre mais de 10.000 moléculas testadas, normalmente apenas uma chega ao mercado, processo que leva em média de 10 a 15 anos para ser finalizado e

pode custar até 1 bilhão de dólares. Além da pesquisa de base, que identifica novos alvos terapêuticos e moléculas promissoras capazes de modulá-los (conhecidas como *leads*, ou protótipos), o processo de desenvolvimento de um novo medicamento inclui os testes pré-clínicos e clínicos (divididos em fase I, II, III e IV), como descritos a seguir.

1.3.1 Pesquisa básica

A pesquisa básica é a fase em que se procura identificar novos compostos que se mostrem promissores no tratamento de alguma doença (figura 8). Para isso, é necessário a identificação das circunstâncias nas quais o desenvolvimento da doença pode ser contido, determinando os genes responsáveis por uma doença humana, ou fatores de virulência nos patógenos que nos acometem, por exemplo. Em seguida, por meio de varreduras de compostos obtidos por síntese química, por prospecção de compostos naturais ou por técnicas virtuais, identificam-se compostos com atividade promissora. Nesse estágio são pesquisados milhares de compostos, e o conhecimento da ciência básica é fundamental. Portanto, parte relevante das pesquisas é realizada nas universidades, institutos e fundações de pesquisa, além de pequenas empresas farmacêuticas e de biotecnologia.

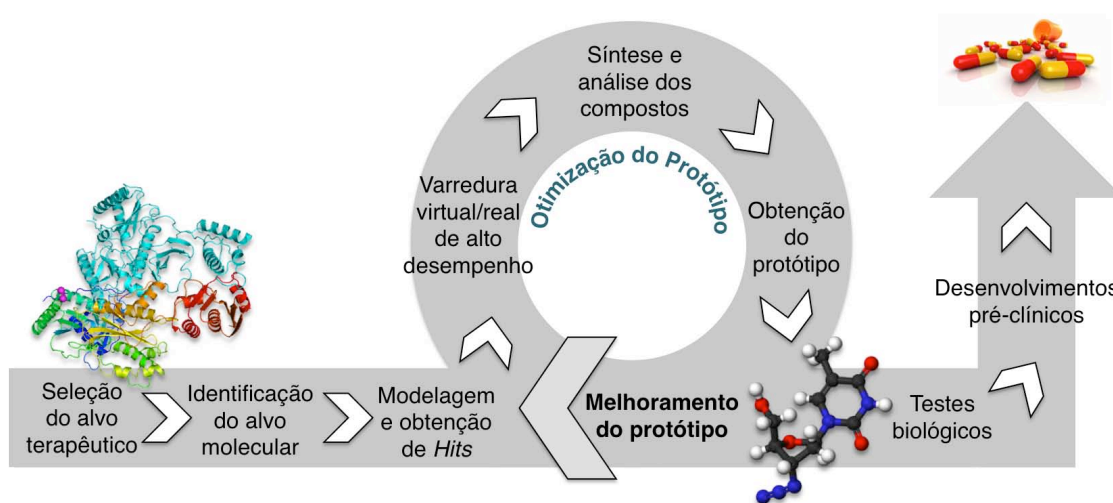


Figura 8. Etapas que constituem a pesquisa básica, antes de uma molécula ser submetida aos testes pré-clínicos. O processo da seleção e validação de um alvo até a identificação de uma molécula promissora pode levar seis anos. Adaptado de ASSES (2011).

1.3.2 Testes pré-clínicos

Os compostos que se mostram promissores na fase de pesquisa básica continuam a ser investigados. Após o isolamento da molécula, passa-se à etapa de desenvolvimento do medicamento com o início dos testes pré-clínicos, considerados o grande filtro da pesquisa e desenvolvimento (P&D) de novas drogas. Cerca de 90% dos compostos testados são eliminados nessa fase, de acordo com a Agência Nacional de Vigilância Sanitária (ANVISA)². Nessa etapa são realizados testes em culturas de células ou tecidos (*in vitro*) e em animais (*in vivo*), com o objetivo de checar os parâmetros de segurança, eficácia e seletividade do novo composto.

Os testes *in vivo* permitem uma análise da eficácia, ou seja, da absorção, distribuição, metabolismo e excreção (ADME) do medicamento dentro do organismo. Além disso, também são realizados testes toxicológicos, a fim de prever os efeitos colaterais e determinar quais são as doses adequadas para os ensaios clínicos em seres humanos.

1.3.3 Testes clínicos

Ainda de acordo com a ANVISA, após passar pelos testes pré-clínicos, a molécula-teste deverá ser avaliada com o objetivo de demonstrar seu caráter inofensivo e eficácia em seres humanos. Somente com base nestes testes é possível a elaboração de um dossiê completo de informações necessárias para a obtenção do registro para a comercialização do medicamento. Os testes clínicos são subdivididos em quatro fases, como descritas abaixo.

- a) **Fase I.** Os primeiros testes são efetuados em voluntários saudáveis, exceto para doenças oncológicas e infecciosas³. Esta fase é também chamada de “fase de segurança”, permitindo comprovar a segurança na utilização de uma determinada

² Disponível em: <http://www.anvisa.gov.br/medicamentos/pesquisa/def.htm>

³ Por razões éticas, a exposição de pacientes saudáveis aos tratamentos anticâncer ou antibióticos não é autorizada devido à atividade dos futuros medicamentos. Neste caso, os testes passam diretamente à fase II.

molécula em humanos, além de determinar a dose mínima ativa e a dose máxima tolerável, bem como os perfis farmacocinéticos e farmacodinâmicos⁴.

- b) Fase II.** Realizada em grupos de pacientes doentes, esta fase permite o ajuste das dosagens a serem prescritas em função das respostas obtidas através de pacientes. A partir desses resultados será possível determinar a posologia do tratamento para a indicação desejada. Além disso, também são avaliados os primeiros efeitos secundários. No caso em que os efeitos secundários se sobressaem ao efeito terapêutico, o desenvolvimento do tratamento é definitivamente suspenso.
- c) Fase III.** Neste momento, os ensaios têm como finalidade avaliar o interesse clínico real e estabelecer a relação risco-benefício do medicamento. Durante esta fase, um grande número de pacientes é submetido tanto à molécula que contém a substância ativa, quanto a um placebo ou outro tratamento de referência. O experimento normalmente é classificado como ‘duplo cego’, onde nem o administrador e nem o paciente têm conhecimento do grupo de estudo a que pertence. Os resultados são então comparados aos dos tratamentos já existentes a fim de confirmar a eficácia da droga.

Com base nos resultados positivos desses testes, passa-se à etapa de registro do novo composto. No Brasil, o registro de um novo medicamento é feito junto à ANVISA. Para obtê-lo, todas as informações sobre o medicamento e suas fases de desenvolvimento devem ser compiladas em dossiês, que são submetidos à agência para aprovação. Em geral, esse processo leva de um a dois anos. É importante ressaltar que o produto deve ser registrado em cada um dos países em que será comercializado. Além dessas três etapas, o medicamento

⁴ Farmacocinética estuda o caminho percorrido pelo medicamento no organismo, desde a sua administração até a sua eliminação. Farmacodinâmica estuda a inter-relação da concentração de uma droga e a estrutura alvo, bem como seu mecanismo de ação.

comercializado continua a ser acompanhado pela empresa e agência reguladora, como descrito abaixo:

- d) **Fase IV.** Esta última fase tem como objetivo identificar os possíveis efeitos secundários tardios, chamada também de “farmacovigilância”.

Uma vez que os testes clínicos e pré-clínicos são cada vez mais rigorosos, os mesmos não podem ser encurtados. A única forma de diminuir o tempo e os gastos no desenvolvimento de novos medicamentos é, portanto, aumentando a eficiência durante a fase de pesquisa destes fármacos. Uma das formas de se aumentar as chances de sucesso é aplicando a chamada “descoberta racional de fármacos”.

1.4 Descoberta racional de fármacos

Os métodos tradicionais de descoberta de drogas são baseados principalmente em tentativa e erro, por meio de testes com compostos químicos em culturas de células ou em animais. Por contrapartida, ao validarmos os alvos responsáveis por uma patologia, sejam eles enzimas humanas ou fatores de virulência de patógenos, a busca de novos medicamentos pode ser feita de maneira racional, direcionada.

A validação de um alvo parte de dois princípios fundamentais. O primeiro é que todo alvo terapêutico deve apresentar propriedades ditas “drogáveis” (HOPKINS e GROOM, 2002). Estas propriedades descrevem a capacidade de um alvo em se ligar a pequenas moléculas que possam modular sua atividade. O segundo é que as modulações de um alvo específico precisam traduzir-se em um efeito terapêutico (FISHMAN e PORTER, 2005; RUSS e LAMPEL, 2005). Para validar essa hipótese, vários testes biológicos de caracterização desses alvos são necessários, seja por meio de mutações ou pela determinação de seus parâmetros bioquímicos.

Existem basicamente duas maneiras de se prever se um alvo é ou não drogável. A primeira se baseia em sua procedência (HOPKINS e GROOM, 2002). Nesse caso, se o candidato fizer parte de uma família de proteínas conhecidas por serem alvos de drogas, ele provavelmente também o será. Esse método também pode ser referenciado como um caso de “prestígio por associação”, e, dessa forma, deve ser analisado com cautela para não representar uma falácia. É sabido, por exemplo, que vários alvos de drogas comercializadas são considerados convencionalmente como não-drogáveis (OWENS, 2007). Dessa forma, a segunda classe de métodos de predição de “drogabilidade” surgiu baseada, principalmente, na análise estrutural da superfície do alvo. Basicamente, estes métodos identificam possíveis sítios de ligação suscetíveis à interação com pequenas moléculas, baseando-se para isso nas características físico-químicas destes sítios (AN *et al.*, 2004; CHENG *et al.*, 2007; HAJDUK *et al.*, 2005).

Após a identificação e validação de um alvo, a purificação de sua forma nativa ou a clonagem e expressão de sua forma recombinante passam a ser extremamente desejáveis, tendo em vista os testes de varredura (virtuais ou não) que se sucederão. Além disso, a expressão recombinante de um alvo algumas vezes permite a resolução de sua estrutura tridimensional, seja por cristalografia de proteínas ou por ressonância magnética nuclear, o que pode ser um fator determinante na descoberta de novos fármacos na ausência de um bom modelo obtido por homologia de sequência.

No que se refere às moléculas promissoras, assim como ocorre com os alvos, existem técnicas preditivas que aumentam as chances de sucesso em uma campanha de descoberta de novos fármacos. Muito popular desde sua publicação, a regra de Lipinski (*Lipinski's rule-of-five*) (LIPINSKI *et al.*, 1997) foi amplamente utilizada para classificar moléculas como *druglike* (fármaco-símiles). O princípio dessas regras é que uma dada molécula deve apresentar características físico-químicas específicas para ser considerada oralmente

disponível. de acordo com Lipinski, para apresentar boa absorção e permeabilidade, o composto deve apresentar menos de cinco doadores de ligações de hidrogênio, menos de 10 receptores de ligações de hidrogênio, massa molecular abaixo de 500 Da e o coeficiente de partição octanol-água (Log P calculado) – fator que indica a hidrofobia ou hidrofília de um composto – abaixo de 5. Entretanto, apenas 51% dos medicamentos aprovados pela agência reguladora dos EUA (FDA) se encaixam nesses parâmetros (ZHANG e WILKINSON, 2007). O problema reside no fato de que a regra-dos-cinco de Lipinski não abrange moléculas derivadas de extratos naturais nem os grandes *blockbusters*⁵, como anticorpos e outras macromoléculas ditas “biológicas”. Desde então, vários autores se propuseram a expandir e modificar essas regras originais para incluir compostos que de outra forma seriam excluídos (para uma ótima revisão, consultar WALTERS, 2012).

A descoberta racional de fármacos pode ser então referenciada como a prospecção de novos medicamentos guiada por predições acerca da “drogabilidade” de um alvo e da fármaco-similaridade de um composto, apresentando diversas vantagens sobre as técnicas comuns de prospecção.

Associada à bioinformática, essa abordagem pode se expandir em uma busca ativa, conhecida por desenho racional de fármacos, onde o pesquisador faz a utilização de filtros para aumentar ainda mais as chances de se encontrar uma molécula biologicamente ativa. As etapas em que a bioinformática se mostra útil no desenho de novos fármacos podem ser visualizadas na figura 9.

⁵ *Blockbusters* são drogas consideradas sucesso de vendas.

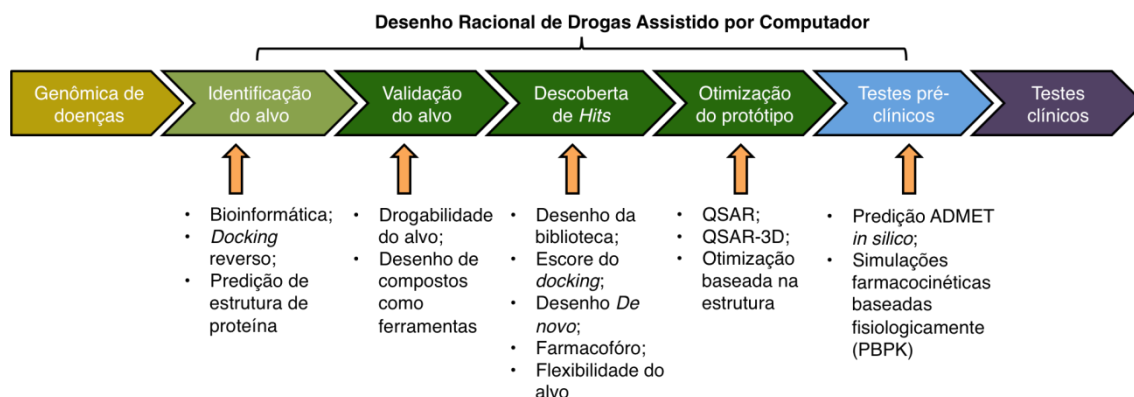


Figura 9. Etapas do desenho racional de drogas assistido por computador. Durante a fase de descoberta, ferramentas podem identificar alvos promissores por meio de buscas por homologia. Além disso, podem-se buscar, através do *docking* reverso, novos alvos de drogas já comercializadas, a fim de descobrir um novo uso para uma droga já registrada. Métodos como modelagem comparativa também são utilizados nessa etapa. Para a validação do alvo, modelos podem ser utilizados para prever efeitos de mutações. Por meio dos modelos também é possível desenhar sondas e outras ferramentas capazes de validar hipóteses. No processo de descoberta de *hits*, além da varredura virtual propriamente dita, a bioinformática é utilizada para filtrar as grandes quimiotecas de forma a enriquecê-las para determinados alvos. Durante a otimização dos *leads*, estudos quantitativos de relação atividade-estrutura (QSAR/QSAR-3D) também são realizados *in silico*. A bioinformática pode auxiliar até mesmo nas fases de testes pré-clínicos, com simulações em modelos de vias metabólicas. Adaptado de ASSES (2011).

Como podemos observar, a bioinformática pode ser empregada em todas as etapas da pesquisa básica de um novo medicamento, diminuindo assim os gastos e o tempo necessários até sua comercialização. O conjunto de ferramentas e metodologias virtuais para o desenvolvimento de novos fármacos constitui a técnica chamada de triagem virtual.

1.4.1 Triagem virtual

A triagem virtual é uma abordagem computacional para a identificação racional de compostos com atividade biológica definida. Contrariamente à triagem de alto rendimento (*High Throughput Screening* – HTS), os compostos químicos não precisam existir fisicamente. Isso representa vantagens no que se diz respeito ao espaço químico amostrado, aumentando a diversidade de moléculas. Dessa forma, o desafio se volta para a filtragem desses compostos, de forma que apenas aqueles com potencial atividade biológica sejam testados. É aí que entram as abordagens racionais, tanto pela aplicação de filtros de “drogabilidade” quanto pela aplicação do conhecimento específico da proteína-alvo. Uma

biblioteca de compostos pode ser reduzida a moléculas com similaridade àquelas com atividade biológica comprovada contra um determinado alvo, por exemplo.

A triagem virtual pode ser iniciada tanto a partir da estrutura tridimensional de um alvo (triagem virtual baseada em estruturas) como a partir de ao menos um ligante conhecido deste alvo (triagem virtual baseada em ligantes). Se ambos o alvo e o ligante possuírem estruturas conhecidas, o conjunto de informações também pode ser usado em uma abordagem combinada.

1.4.1.1 Triagem virtual baseada em ligantes

O princípio fundamental deste método está na suposição de que moléculas semelhantes entre si devem possuir propriedades de ligação similares (ECKERT e BAJORATH, 2007; JOHNSON e MAGGIORA, 1990). Esse tipo de abordagem é utilizado quando se conhece ao menos uma molécula biologicamente ativa contra um alvo específico. Por meio da aplicação de filtros de similaridade, apenas as moléculas com descritores afins são retidas e posteriormente utilizadas no processo de HTS.

No caso de várias moléculas biologicamente ativas serem conhecidas, um método popular de triagem consiste na modelagem de um farmacóforo, que nada mais é do que a determinação das estruturas comuns entre todas as moléculas ativas. Para tanto, é necessário que as moléculas conhecidas atuem no mesmo sítio de ligação. Essa modelagem resulta em um modelo mínimo comum, que pode ser submetido a buscas de similaridade em bancos de dados de compostos químicos, também chamados de quimiotecas, gerando variabilidade química. Porém, esse tipo de abordagem é mais ligado à descoberta de drogas *me-too*⁶, o que pode não ser desejado durante as etapas de descobertas de novas moléculas. Esses métodos, no entanto, são muito eficientes no enriquecimento de bancos de dados para utilização em

⁶ *Me-too* são drogas que não apresentam valor de novidade; correspondem apenas a mais uma alternativa de mercado.

conjunto com as técnicas de triagem virtual baseada em estruturas, principalmente para a busca de compostos contra alvos homólogos que ainda não possuem uma molécula ativa, no refinamento de *leads* e na busca de drogas multialvos (YANG, 2010).

No caso de haver tanto moléculas conhecidamente ativas como inativas contra um alvo, a técnica de aprendizado de máquina pode ser utilizada. Ao final do treinamento, o programa poderá discernir as moléculas com atividade das sem atividade. Essa técnica é importante para evitar falsos positivos (KOEPPEN e KRIEGL, 2011).

1.4.1.2 Triagem virtual baseada em estruturas

Também conhecida como ancoragem molecular (*docking*) ou triagem virtual de alto rendimento (*virtual High Throughput Screening* – vHTS), essa técnica se baseia nos princípios de complementariedade físico-química entre o sítio ativo e o ligante. Com a estrutura 3D de um alvo e com uma biblioteca de compostos promissores disponíveis, os algoritmos de ancoragem molecular testam cada combinação possível, e em seguida ranqueiam os resultados com base em diversas metodologias (ROGNAN, 2011).

A primeira etapa para uma campanha de *docking* bem sucedida é a preparação do banco de dados de ligantes. O enriquecimento do conjunto de compostos a ser testado é fundamental, mas a amostragem de espaços de conformação de cada um destes compostos é de igual importância. Para tanto, existem basicamente três maneiras de lidar com a flexibilidade do ligante.

- **Geração de bibliotecas de ligantes multiconformacionais (flexibases).** Baseia-se na preparação, com o auxílio de programas específicos, de um conjunto de conformémeros (*flexibase*) de todos os ligantes submetidos ao *docking* (KEARSLEY *et al.*, 1994). Nesse caso, a ancoragem é completamente rígida, e depende da

complementariedade exata entre proteína e ligante. Caso a conformação ideal do ligante não tenha sido gerada, o *docking* não apresentará bons resultados.

- **Construção incremental.** Baseia-se na construção do ligante *on-the-fly*, primeiramente dividindo o ligante em seus componentes básicos e ancorando um fragmento de referência com base na complementariedade geométrica. Em seguida, o programa constrói o restante do ligante, respeitando um conjunto de ângulos prováveis entre as ligações atômicas. A vantagem desse método é a de respeitar a conformação da proteína durante o *docking*. Uma desvantagem é que, já que muitos compostos compartilham fragmentos idênticos, pode haver perda de tempo recalculando a ancoragem de fragmentos provenientes de moléculas semelhantes (ROGNAN, 2011).
- **Métodos estocásticos.** São divididos em duas categorias, os baseados no método de Monte Carlo (MCs) ou em algoritmos genéticos (GAs). Os métodos MC estimam a energia das poses geradas para cada ligante. Se a energia for menor ou igual à anterior, ela é mantida. Caso seja maior, uma probabilidade dessa conformação ser mantida é calculada, baseada na temperatura e na diferença de energia entre os dois estados (peso de Boltzmann). É comum nesse tipo de abordagem a diminuição da temperatura para reter o complexo proteína-ligante em um nível de energia baixo (TIETZE e APOSTOLAKIS, 2007). A abordagem de GAs é inspirada na teoria da evolução de Darwin. Neste caso, cada característica de um ligante, como ângulos e torções, é considerada como genes. Cada geração de uma nova pose gera características diferentes, que são passadas adiante caso sejam mais aptas. Esse algoritmo é utilizado nos programas GOLD (JONES *et al.*, 1997) e AutoDock (MORRIS *et al.*, 1998), por exemplo.

Além da amostragem de espaços conformacionais dos ligantes, cada programa de *docking* utiliza um método de ranqueamento para classificar suas predições. Esses métodos podem ser agrupados em três categorias:

- **Funções de escore empíricas.** Usa termos que descrevem propriedades importantes para a ligação de um composto para construir uma equação mestra de predição de afinidade de ligação. Depende de um conjunto de treinamento, e, portanto, deve-se usar uma função diferente para cada classe de alvos (FERRARA *et al.*, 2004).
- **Potencial de força média baseado em conhecimento.** Ao contrário das funções empíricas, esse método baseia-se na média de todas as interações atômicas entre proteína-ligante contidas em um banco de dados (por exemplo, o *Protein Data Bank* – PDB), e não em apenas um conjunto de alvos. Dessa forma, esse sistema funciona similarmente para todas as classes de alvos (ROGNAN, 2011).
- **Campos de força.** Baseia-se num campo de forças de mecânicas moleculares clássicas, como por exemplo, o AMBER ou CHARMM. Levam-se em conta as interações do tipo van der Waals e os componentes eletrostáticos das interações. A principal limitação é que esse método omite o componente entrópico da energia livre de ligação, ou a aleatoriedade das conformações, ranqueando melhor moléculas grandes e polares (ROGNAN, 2011).

Conhecendo os mecanismos por trás dos maiores motores de acoplamento molecular, podemos prosseguir para os outros aspectos importantes da triagem virtual baseada em estruturas. O fluxograma típico de uma campanha de *docking* é composto de quatro etapas, onde cada etapa é considerada igualmente importante e pode influenciar na qualidade dos resultados.

A primeira etapa é a de preparação dos ligantes. É importantíssimo verificar os estados de protonação, tautomeria e estereoisomeria, principalmente ao utilizarem-se funções empíricas para o ranqueamento dos resultados. O formato MOL2 é considerado o padrão-ouro, já que é compatível com a maior parte dos aplicativos e é capaz de lidar com questões críticas (tratamento de grupos simétricos e variedade de tipos atômicos) (ROGNAN, 2011). Recomenda-se gerar configurações 3D confiáveis pelo programa CORINA (Molecular Networks GmbH). A escolha de uma coleção de compostos também é fundamental, e deve ser feita baseada no alvo a ser estudado. Muitas companhias oferecem bibliotecas focadas de compostos comerciais (por exemplo, por meio de similaridade com moléculas ativas).

A etapa seguinte é a de preparação da proteína. É considerado tecnicamente mais simples, pois, ao contrário dos ligantes, trata-se de apenas uma macromolécula. Recomenda-se sempre partir de uma holoestrutura (complexo entre proteína-ligante) ao invés de uma apoestrutura (proteína livre de ligante) (MCGOVERN e SHOICHET, 2003). No caso de múltiplas estruturas disponíveis, é aconselhada a escolha de um conjunto de estruturas que reflitam a flexibilidade conformacional do sítio ativo. Em todos os casos, átomos de hidrogênio, que não estão presentes em estruturas cristalográficas, devem ser adicionados em todos os alvos, com atenção especial para resíduos de serina, treonina e tirosina (hidrogênios polares). Um estado tautomérico para as histidinas também deve ser escolhido, baseado principalmente nos dados biológicos do alvo.

A terceira etapa deve ser a escolha do algoritmo a ser utilizado na campanha de *docking*. Desde o trabalho pioneiro de Kuntz e colaboradores (KUNTZ *et al.*, 1982), mais de 60 programas de acoplamento molecular já foram desenvolvidos (MOITESSIER *et al.*, 2008), e, infelizmente, não existe uma maneira simples de prever qual deles é o melhor em um determinado projeto. Dentre os problemas enfrentados por todos esses algoritmos, estão: a mobilidade tanto dos ligantes como do receptor, o efeito do ambiente proteico na

distribuição de carga no ligante (CHO *et al.*, 2005), e suas interações com moléculas de água circundantes.

Como última etapa, estão as análises pós-*docking*. A maneira mais simples de se acessar os resultados de uma campanha de triagem virtual seria pela simples análise do ranque gerado ao final do processo. Entretanto, devido às limitações do processo de ranqueamento, essa abordagem não fornecerá as informações mais corretas. Uma maneira de melhorar esses resultados seria o de aplicar outras funções de escore (*rescoring*) na esperança de que um escore consenso seja capaz de separar *hits* reais de pistas falsas (CHARIFSON *et al.*, 1999).

Apesar do desenho racional de fármacos aparentar ser um processo simples, muitas campanhas de acoplamento molecular não são bem sucedidas. As dificuldades encontradas no percurso levaram à reavaliação dos princípios envolvidos. Enquanto alguns desenvolvedores de drogas optaram por aperfeiçoar seus métodos experimentais, outros focaram suas atenções no melhoramento dos protocolos computacionais. Esses melhoramentos incluem, entre outros: incorporação da flexibilidade no processo de acoplamento, exploração extensiva das conformações do ligante no sítio de ligação, refinamento e análise da estabilidade dos complexos finais e estimativas das energias livres de ligação. As simulações de Dinâmicas Moleculares (DM) desempenham papel fundamental nessas abordagens (ALONSO *et al.*, 2006).

Outra grande limitação do desenho de fármacos baseado nessa abordagem é a disponibilidade de estruturas 3D em bancos de dados. Principalmente em se tratando de doenças negligenciadas, seria quase impossível a obtenção da estrutura de todos os alvos e fatores de virulência destes patógenos menos visados pela indústria farmacêutica. Apesar de originalmente recomendado apenas em último caso (ROGNAN, 2011), o *docking* molecular

a partir de modelos por homologia já está bem estabelecido (CAVASOTTO, 2011). Todos os estudos de *docking* utilizando receptores do tipo GPCRs humanos são baseados em modelos por homologia, por exemplo, dentre os quais vários apresentaram sucesso (TUCCINARDI, 2009). Nesse sentido, a modelagem por homologia permite a extrapolação de estruturas funcionalmente relacionadas em campanhas de triagem virtual. Considerando a importância da modelagem comparativa no processo de desenvolvimento de novas drogas contra doenças negligenciadas, seus princípios serão expostos logo a seguir.

1.5 Modelagem por homologia

Modelagem por homologia ou comparativa é o processo de gerar um modelo estrutural de uma proteína alvo a partir de uma estrutura guia (*template*) conhecida. Uma vez escolhida a proteína alvo, quatro componentes principais do processo podem ser identificados:

- Seleção de uma ou mais estruturas guias, relacionadas ao alvo, que podem ser utilizadas na transferência de coordenadas atômicas;
- Geração de um alinhamento par-a-par (*pairwise*), ou, em alguns casos, um alinhamento múltiplo entre o alvo e a(s) estrutura(s) guia(s);
- Construção de um modelo preliminar do alvo baseado no *template* e no alinhamento de sequência, que foi utilizado para designar a correspondência entre aminoácidos no alvo e no(s) *template(s)*;
- Refinamento e validação do modelo.

O grau de identidade ou similaridade entre a sequência do alvo e do *template* normalmente é utilizado como predição da qualidade de um modelo. Entretanto, deve-se empregar essa medida com cautela, uma vez que esses valores podem mascarar a identidade real no sítio de ligação. Apesar disso, um alinhamento de sequências correto é necessário,

apesar de não ser uma condição suficiente para o desenvolvimento de um modelo confiável. Em casos de baixa similaridade de sequência, alinhamentos múltiplos podem melhorar a qualidade do alinhamento entre alvo e *template*. Em outros casos, quando um único *template* não é suficiente para produzir modelos precisos, faz-se necessário o uso de múltiplas estruturas guias (CRAIG *et al.*, 2010).

Também há de se notar que a conformação adotada por alguns domínios é dependente do tipo de ligante associado (efeito do ajuste induzido). Desse modo, uma única estrutura pode não representar a conformação final do espaço estrutural do complexo ligante-alvo. As técnicas de acoplamento em conjuntos (*ensemble docking*) se mostraram uma abordagem válida para considerar a flexibilidade de proteínas em estudos de vHTS em vários alvos diferentes, incluindo GPCRs (VILAR *et al.*, 2011). Podemos concluir então que, como proteínas podem se ligar a moléculas estruturalmente diferentes, faz-se necessária a incorporação da flexibilidade do alvo na modelagem de sítios ativos (CAVASOTTO e SINGH, 2008).

Entretanto, como já mencionado anteriormente, nem sempre um alvo possui sua estrutura resolvida, quem dirá várias estruturas. Um método para contornar essa falta de dados estruturais é a utilização de simulações de dinâmicas moleculares para amostrar os espaços conformacionais de uma proteína em solução. Os princípios dessa técnica estão descritos abaixo.

1.6 Dinâmicas moleculares (DMs)

1.6.1 Princípios

Uma simulação de dinâmica molecular clássica (DM) é uma abordagem computacional na qual as equações das leis de movimento de Newton são solucionadas em função de uma representação atomística de um sistema molecular (baseado em mecânica molecular), de

maneira a obter informações sobre suas propriedades ao longo do tempo. Esse tipo de simulação é considerado como uma das técnicas mais versáteis e amplamente utilizadas para o estudo de macromoléculas (KARPLUS e MCCAMMON, 2002). Por meio de simulações de DM pode-se compreender o comportamento dinâmico de proteínas em diferentes escalas de tempo, desde movimentos internos rápidos, passando por mudanças conformacionais lentas, e até mesmo o processo de enovelamento de uma proteína (SNOW *et al.*, 2005).

O princípio da DM se baseia na solução da equação de Newton relativa ao sistema:

$$\vec{F}_i = m_i \vec{a}_i$$

onde m_i é a massa de um átomo i , \vec{a}_i é sua aceleração e \vec{F}_i a soma das forças aplicadas a si, devido a sua interação com os outros átomos do sistema, seguindo os parâmetros contidos no campo de força escolhido.

O campo de força é o conjunto de equações que regem a energia de um sistema e seus parâmetros associados (MONARD, 2003). Ele define a ligação entre a estrutura geométrica dos átomos do sistema de estudo e suas energias potenciais ao curso de uma simulação (MACKERELL, 2004). De fato, ele contém todos os parâmetros que participam da simulação, tanto de dinâmica molecular como de minimização de energia do sistema estudado. Em outros termos, pode-se dizer que o campo de força é um dos ingredientes mais delicados a ser implementado (PEARLMAN e KOLLMAN, 1991). Vários campos de força foram criados em respeito ao sistema estudado e, logo, o campo de força escolhido para uma simulação depende do sistema para o qual ele foi desenvolvido. Alguns dos campos de força mais conhecidos estão descritos na tabela 1.

Tabela 1. Principais campos de força e suas utilizações em dinâmicas moleculares.

Campo de força	Utilização	Referência
CHARMM	Biomoléculas, com ênfase em proteínas	BROOKS <i>et al.</i> , 2009
AMBER	Principalmente proteínas e ácidos nucleicos	WANG <i>et al.</i> , 2004
MM2, MM3 e MM4	Pequenas moléculas orgânicas	PISSURLENKAR <i>et al.</i> , 2009
GROMOS	Biomoléculas	CHRISTEN <i>et al.</i> , 2005

A função de energia potencial contida nesses campos de força podem ser decompostas na soma de potenciais que regem a interação entre átomos ligados e átomos não-ligados. Sua expressão generalizada seria, portanto:

$$V_T = \underbrace{\sum V_{\text{ligações}} + \sum V_{\text{ângulos}} + \sum V_{\text{diedros}} + \sum V_{\text{diedros impróprios}}}_{V_{\text{covalente}}} + \underbrace{\sum V_{\text{vdW}} + \sum V_{\text{eletrostático}}}_{V_{\text{não-covalente}}}$$

Enquanto que a maioria dos campos de força utilizem funções próprias para tratar dos átomos ligados, os átomos não ligados normalmente são regidos pela equação de Lennard-Jones (no caso dos potenciais de van der Waals) e pela equação de Coulomb (para as interações eletrostáticas).

Dentre os campos de força anteriormente listados, o CHARMM, o AMBER e o GROMOS são frutos de muitos anos de pesquisa especializada, orientados principalmente para a reprodução do comportamento dos aminoácidos e ácidos nucleicos. Associados a programas como o NAMD (PHILLIPS *et al.*, 2005) e o GROMACS (HESS *et al.*, 2008), esses campos de força permitem a realização de simulações de dinâmicas moleculares consideradas fiéis, contanto que sejam aplicadas a sistemas biomoleculares (como os complexos proteicos, por exemplo).

1.6.2 Minimização de energia

A minimização de energia potencial é um dos métodos mais comumente utilizados para aperfeiçoar a estrutura de biomoléculas. Essa operação pode ser útil no relaxamento da

conformação inicial de uma molécula (ao eliminar maus contatos atômicos), buscando uma geometria de baixa energia, correspondente a um dos estados estáveis do sistema (figura 10). A minimização de energia é amplamente utilizada no refinamento de estruturas por radiografia cristalográfica ou de espectrometria de ressonância magnética nuclear (NMR), assim como na otimização da estrutura de biomoléculas antes de uma simulação de DM.

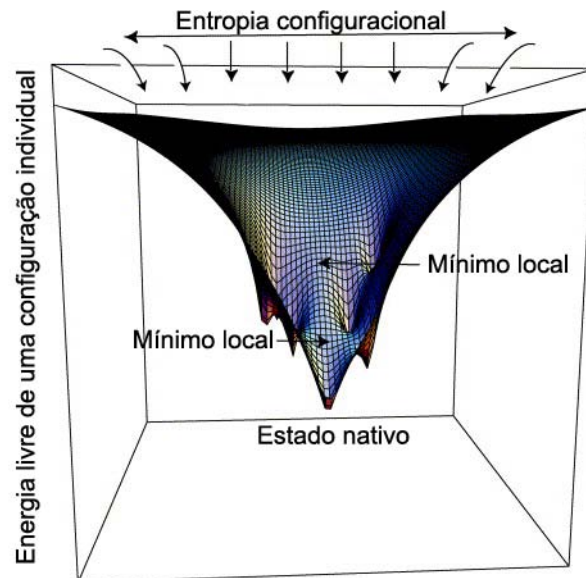


Figura 10. Funil de estados conformacionais de uma proteína. Os espaços amostrados pela proteína antes de atingir o seu estado nativo (de energia mínima global), incluem vários mínimos locais. Adaptado de (ONUCHIC *et al.*, 1996).

Entretanto, uma vez que não há um método matemático geral capaz de descobrir o estado de mais baixa energia de uma molécula (o mínimo global, também chamado de estado nativo), os métodos de minimização de energia tendem a convergir para um dos mínimos locais (figura 10).

Para evitar-se o aprisionamento da estrutura em um mínimo local próximo, recomenda-se amostrar seus espaços conformacionais. Essa busca tem como finalidade a determinação de um conjunto representativo de conformações estáveis da molécula no equilíbrio termodinâmico, ou seja, suscetíveis de existirem em condições experimentais (ALONSO *et al.*, 2006).

Existem duas classes de métodos para obtenção de amostras do espaço conformacional de uma molécula: os métodos estocásticos (Monte Carlo, por exemplo) e os métodos determinísticos (dinâmicas moleculares sistemáticas).

1.6.3 Métodos estocásticos - o exemplo do método de Monte Carlo

São normalmente menos onerosos em se tratando de recursos computacionais, pois se baseiam na geração aleatória de um conjunto de conformações, utilizando critérios de seleção específicos para conservar ou não as novas conformações criadas.

Em uma simulação de Monte Carlo, cada conformação gerada resulta do movimento aleatório de um ou mais átomos, ou da variação de um ou mais diedros. A energia da conformação obtida é então avaliada. Se a nova energia é mais baixa que a anterior, ela é mantida. Normalmente, portanto, os métodos estocásticos não são capazes de garantir a obtenção do mínimo global de energia, mas permitem obter um conjunto de configurações de baixa energia que constituem uma amostragem representativa.

1.6.4 Método determinístico - o exemplo da Dinâmica Molecular

Enquanto que durante uma simulação de Monte Carlo as conformações são geradas de forma aleatória e discretas (dependendo da probabilidade de transição entre o estado atual e o seguinte), aquelas provenientes de uma simulação de DM são conectadas no tempo e formam uma trajetória contínua. A dinâmica molecular é dita determinística, pois todos os estados futuros do sistema podem ser preditos a partir do estado atual.

Durante uma simulação de DM, o sistema sofre mudanças conformacionais e cinéticas que permitem explorar o espaço de fases espaço-temporais acessíveis pelo sistema. Para cada partícula i , em todo tempo t , é associada uma dupla vetorial (posição $\vec{r}_i(t)$ e velocidade $\vec{v}_i(t)$). O conjunto de coordenadas na totalidade do espaço temporal explorado constitui o

que chamamos de trajetória. De acordo com a hipótese ergódica, o estudo de uma trajetória infinitamente longa de um sistema por DM resulta na amostragem de todos os seus espaços de fase. Dessa forma, torna-se possível o acesso às grandezas termodinâmicas, tais como coeficientes de difusão, energia livre do sistema, etc. Isso permite ligar as simulações aos dados experimentais.

A partir de um sistema no qual as condições iniciais foram fixadas, a DM consiste em duas operações: avaliar a força que age em cada átomo no tempo t , e determinar as coordenadas e velocidades de cada átomo no tempo $t + \Delta t$ em função das forças sofridas por eles, onde Δt corresponde ao passo de integração.

1.6.5 Dinâmica Molecular e flexibilidade de proteínas

Atualmente é amplamente aceito que o antigo modelo de “chave e fechadura” da interação entre um ligante e seu receptor não representa uma descrição precisa da maioria dos sistemas biológicos. Ao contrário disso, as interações entre ligante e proteína lembram muito mais uma associação do tipo “mão e luva”, onde ambas as partes são móveis e se adaptam para complementar uma a outra (ajuste induzido). Essa conformação é obtida por meio da mudança de formas entre os componentes do sistema, aumentando contatos favoráveis e diminuindo reações desfavoráveis, maximizando a energia livre de ligação total (VERKHIVKER *et al.*, 2002). Ademais, sabe-se que as regiões do sítio ativo de enzimas possuem tanto áreas de baixa como de alta estabilidade conformacional (LUQUE e FREIRE, 2000).

Entretanto, a ideia do modelo “chave e fechadura” ainda é passada ao observarmos a estrutura de um complexo obtida por cristalografia: uma estrutura rígida, com duas peças encaixadas como um quebra-cabeças. Podemos inferir, portanto, que enquanto estruturas

de complexos ligante-proteína podem prover resultados satisfatórios durante a triagem virtual, os resultados finais podem ser tendenciosos a um tipo particular de ligante.

Além do efeito do ajuste induzido, proteínas em solução são moléculas móveis. Elas não existem em apenas uma conformação, mas em diversos estados conformacionais separados por barreiras de baixa e alta energia. A distribuição e estabilidade de cada estado conformacional dependerão das características físico-químicas do ambiente e da proteína em si, como por exemplo, a presença de ligantes, íons e cofatores (KUMAR *et al.*, 2000).

Devido à alta flexibilidade contida nas proteínas, um conjunto de conformações do receptor, e não apenas uma estrutura, pode prover uma melhor representação do sistema. A ancoragem utilizando várias conformações de uma estrutura aumentam as chances de encontrar um receptor no estado correto para acomodar um ligante em particular. Entretanto, isso também diminui a seletividade do processo de triagem virtual, já que uma variedade maior de ligantes poderá se ligar nessas representações mais relaxadas da proteína. Dessa forma, o uso de funções de escore refinadas durante o processo final da varredura é fundamental para maximizar a seleção de ligantes ativos.

O papel da DM na amostragem de outras conformações representativas da flexibilidade de um alvo pode ser extremamente útil nesses casos. Schames e colaboradores (SCHAMES *et al.*, 2004) descobriram um novo sítio de ligação na integrase do HIV-1, por meio da ancoragem do inibidor 5CITEP em estruturas derivadas de uma trajetória de 2 ns. O *docking*, nesse caso, identificou dois modos de ligação, um dos quais se aproveitava de um espaço adjacente ao sítio de ligação. Uma série de “compostos borboleta”, com a habilidade de se ligar simultaneamente a ambos os sítios, foram desenhados e demonstrados ligarem-se à conformação aberta da proteína. A descoberta desse novo sítio não teria sido possível sem as simulações de DM do receptor.

Além da análise de movimentos intramoleculares, a DM também possibilita o estudo de efeitos de solventes explícitos na estrutura e estabilidade de uma proteína, de forma a obter propriedades do sistema (por exemplo, densidade, condutividade e momento dipolar), assim como parâmetros termodinâmicos (energias de interação e entropia). Ademais, a DM não é útil apenas para racionalizar propriedades medidas experimentalmente. É sabido que a maioria das estruturas resolvidas por raios X ou métodos de NMR foi refinada utilizando métodos de DM (ALONSO *et al.*, 2006). Portanto, a interação entre técnicas computacionais e experimentais na área de simulações de DM é uma via de mão dupla, com métodos teóricos auxiliando no entendimento e análise de dados experimentais que, por sua vez, são fundamentais para a validação e melhoramento das técnicas e protocolos computacionais.

Finalmente, seja por meio do refinamento e validação de modelos obtidos por homologia ou pela amostragem do espaço conformacional de uma proteína durante uma trajetória, as dinâmicas moleculares se mostram extremamente úteis para o desenho racional de fármacos, principalmente quando aplicadas a alvos de doenças negligenciadas, para as quais faltam estruturas cristalográficas. Associada à diminuição dos custos computacionais, ela se torna ferramenta indispensável nesse processo, capaz de reduzir gastos e tempo necessários para as primeiras etapas da descoberta de novos medicamentos.

2 Justificativa

Tendo em vista as lacunas de medicamentos destinados às doenças negligenciadas, assim como o desinteresse das grandes empresas farmacêuticas em preenchê-las, é papel das entidades públicas de pesquisa se posicionar frente ao desenvolvimento de novas terapias.

Apesar de serem responsáveis por grande parte da pesquisa básica, as IES não possuem estrutura para todas as etapas de desenvolvimento de fármacos, como testes pré-clínicos e clínicos. Resta, portanto, a opção de se firmar parcerias público-privadas para a resolução desse problema. Somente a tríade academia-governo-empresa é capaz de solucionar esse impasse. Mas como tornar um negócio não-rentável atrativo aos olhos das grandes farmacêuticas? A solução pode estar na redução de gastos na pesquisa básica, que se assumidos pelas instituições públicas podem gerar um alívio de até 40% dos gastos da indústria. No entanto, o dinheiro do contribuinte também deve ser economizado e, para isso, novas técnicas capazes de reduzir custos e o tempo necessário para a descoberta de novos fármacos devem ser aplicadas. É nesse ponto que entra a bioinformática como ferramenta contra doenças tropicais. Por meio de técnicas bem estabelecidas, e com a diminuição dos custos de recursos computacionais, uma plataforma de descoberta racional de drogas pode ser instalada com sucesso nas instituições de ensino superior.

Dentro deste contexto, e, tendo em mente a importância da dengue e da doença de Chagas no panorama nacional, fez-se necessário um estudo piloto das técnicas de bioinformática utilizadas no processo da descoberta racional de fármacos.

Os resultados desse estudo, além de contribuírem para o estabelecimento dessa plataforma, revelaram moléculas promissoras com potencial farmacêutico contra a dengue. Além disso, as mesmas técnicas foram capazes de identificar moléculas putativas contra a doença de Chagas, demonstrando que podem servir de base no estudo de outras doenças.

3 Objetivo

Tendo em vista os problemas descritos anteriormente, o objetivo do presente trabalho é aplicar métodos de modelagem e simulações virtuais no estabelecimento de uma plataforma de desenvolvimento racional de fármacos, baseando-se para isso no estudo de dois alvos de importantes doenças no país: a dengue e a doença de Chagas.

3.1 Objetivos específicos

1. Elucidar o papel do cofator e do ligante no efeito de ajuste induzido da protease do vírus da dengue;
2. Investigar as conformações da protease do vírus da dengue durante simulações de dinâmica molecular;
3. Explorar as conformações obtidas na descoberta racional de fármacos;
4. Identificar novos compostos inibidores da protease do vírus da dengue com potencial terapêutico;
5. Desenvolver modelos moleculares de alvos promissores da doença de Chagas;
6. Investigar suas conformações durante uma simulação de dinâmica molecular;
7. Explorar essas conformações na descoberta racional de fármacos;
8. Buscar moléculas com potencial terapêutico contra esses alvos.

4 Material e Métodos

4.1 Clonagem, expressão e purificação da NS2B_{CF}-Gly-NS3_{PRO}

4.1.1 Amplificação e construção do complexo NS2B_{CF}-Gly-NS3_{PRO}

O plasmídeo contendo o segmento de DNA genômico com as sequências de nucleotídeos que codificam as proteínas NS2B e NS3 do *Dengue virus* 1 (isolado DF01-HUB01021093) foi gentilmente cedido pelo professor Dr. Tatsuya Nagata. Os fragmentos de cada domínio (NS2B_{CF} e NS3_{PRO}) foram fusionados por meio de uma reação de sobreposição com uma sequência que codifica um nonapeptídeo maleável não-clivável (Gly_[4]-Ser-Gly_[4]), de maneira que a interação entre o domínio hidrofílico da proteína NS2B e a NS3 pudesse ser promovida (LEUNG *et al.*, 2001). Os oligonucleotídeos sintéticos foram desenhados de modo a inserir um sítio de restrição para a enzima *Xho*I na extremidade 5' do segmento do gene correspondente à NS2B_{CF} e uma sequência não redundante para codificar a dobradiça maleável na extremidade 3'. De maneira semelhante, foi inserida a sequência complementar à da dobradiça maleável na região 5' do segmento correspondente à NS3_{PRO}, e um sítio de restrição para *Bam*HI na região 3'. Dessa forma, as regiões complementares da dobradiça se anelam durante uma reação de PCR envolvendo ambos os fragmentos obtidos de uma primeira reação (NS2B_{CF} e NS3_{PRO}). As sequências dos oligonucleotídeos, com os sítios de restrição sublinhados e as regiões da dobradiça em negrito, estão transcritas abaixo:

```
DEN-1_NS2BF: 5'-TATGCTCGAGGCCGATCTATCATTGGAGAAA-3'  
DEN-1_NS2BR: 5'-CCCGCCTCCACCACTACCTCCGCCCCCGAGTGTGTCATCCCTCTCTTCAT-3'  
DEN-1_NS3F: 5'-GGGGCGGAGGTAGTGGTGGAGGCGGGTCAGGAGTGTATGGGACACACCCA-3'  
DEN-1_NS3R: 5'-ATCGAGGATCCTTTCTTAAACACCTCGTCCTCAATC-3'
```

Para todas as reações, a fim de se evitar mutações indesejadas, foram utilizadas enzimas de alta fidelidade (*Platinum® Taq Pfx*, Invitrogen). A menos que indicado, cada reação era composta também de (com as devidas concentrações entre colchetes): 5 µl de tampão de atividade [10x]; 0,6 µl de mistura de dNTPs [25 mM cada]; 1 µl de MgSO₄ [50 mM]; 1 µl de

cada iniciador [10 µM] (vide tabela 2); 1 µl de DNA molde [20 ng/µl]; 0,4 µl de *Platinum® Taq DNA Polymerase* [2,5 u/µl]; água Milli-Q q.s.p. 50 µl. Em seguida, as amostras foram submetidas a 30 ciclos de amplificação em um termociclador (PTC-100, BioRad) programado como se segue:

01. Dissociação do anticorpo Platinum®⁷ – 94 °C, 2 minutos;
02. Desnaturação – 94 °C, 15 segundos;
03. Anelamento – 57 °C, 30 segundos;
04. Extensão – 68 °C, 45 segundos;
05. Repetir – de 2 a 4, 30 vezes;
06. Extensão final – 68 °C, 4 minutos;
07. Fim da reação – 4 °C, até ser armazenada a -20 °C.

Tabela 2. DNA molde e oligonucleotídeos das reações de PCR.

Reação	DNA molde	Oligonucleotídeos
A	DEN1 DF01	DEN-1_NS2BF + DEN-1_NS2BR
B	DEN1 DF01	DEN-1_NS3F + DEN-1_NS3R
C	A + B	DEN-1_NS2BF + DEN-1_NS3R

Ao final das reações, os produtos amplificados foram examinados por eletroforese em gel de agarose 0,8%. Antes da aplicação das amostras no gel, as mesmas foram diluídas em tampão de amostra 10x (250 mg azul de bromofenol, 250 mg de xilenocianol, 50 mM Tris-HCl pH 7,6, 60% glicerol). A eletroforese foi realizada em tampão TAE (40 mM Tris, 20 mM ácido acético, e 1 mM EDTA), com voltagem constante de 90 V, e a visualização e documentação do gel foi feita sob luz UV em um fotodocumentador (*MultiImage® II*, Alpha Innotech).

Para a obtenção dos fragmentos fusionados (reação C), os produtos previamente obtidos das reações A e B foram separados por eletroforese em gel de agarose 0,8%, e em seguida,

⁷ Essa enzima vem associada a um anticorpo para garantir sua atividade ótima, que evita que ela entre em atividade antes da reação de PCR.

cada uma das bandas com os fragmentos correspondentes foram extraídas do gel e diluídas em água Milli-Q para uma concentração de aproximadamente 20 ng de DNA por μl total. Essa solução foi aquecida até a total dissolução da agarose, e utilizada como DNA molde. Os oligonucleotídeos somente foram adicionados após 10 ciclos de auto-anelamento entre os fragmentos A e B.

A programação dos ciclos da amplificação C foi:

01. Dissociação do anticorpo Platinum® – 94 °C, 2 minutos;
02. Desnaturação – 94 °C, 15 segundos;
03. Anelamento – 62 °C, 45 segundos;
04. Extensão – 68 °C, 30 segundos;
05. Repetir – de 2 a 4, 10 vezes;
06. Adição dos oligonucleotídeos DEN-1_NS2BF e DEN-1_NS3R;
07. Desnaturação – 94 °C, 15 segundos;
08. Anelamento – 57 °C, 30 segundos;
09. Extensão – 68 °C, 45 segundos;
10. Repetir – de 6 a 8, 30 vezes;
11. Extensão final – 68 °C, 4 minutos;
12. Fim da reação – 4 °C, até ser armazenada a -20 °C.

4.1.2 Clonagem nos vetores pGEM®-T easy e pET-19b

O resultado da reação C (ns2b_{CF40}-gly-ns3_{PRO}) foi inicialmente ligado ao vetor pGEM®-T Easy (Promega) com a enzima T4 DNA Ligase, de acordo com instruções do fabricante, em temperaturas de 4 ou 14 °C, por 16 h. A escolha do vetor pGEM®-T Easy se deu porque o mesmo não possui nenhum dos sítios de restrição utilizados no inserto. Isso possibilitou a digestão completa nos sítios específicos, seguida de uma ligação direcional no vetor de expressão desejado.

O produto de ligação foi transformado por choque térmico em bactérias quimicamente competentes DH5 α de acordo com SAMBROOK e RUSSELL (2001). Brevemente, metade da

reação de ligação foi incubada por 30 minutos em uma alíquota de 100 µl de células competentes resfriadas em gelo. Em seguida, os tubos contendo a reação de ligação e as bactérias competentes foram submetidos a um choque térmico de 42 °C por 45 segundos, ao final dos quais foram retornados ao gelo por pelo menos mais 2 minutos antes de se adicionar 800 µl de meio de cultura LB. Após, os tubos foram incubados a 37 °C por uma hora sob agitação constante de 140 RPM. Após a devida recuperação das bactérias, as mesmas foram semeadas em meio LB-ágar contendo 100 µg/ml de ampicilina, 1 mM de isopropil β-D-1-tiogalactopiranosídeo (IPTG) e 40 µl de X-gal em concentração de 40 mg/ml para seleção de colônias positivas.

A análise dos DNAs plasmidiais das colônias recombinantes foi feita por fenol:clorofórmio:álcool isoamílico (iaa) (25:24:1) (Bioagency) e posterior mini preparação de DNA plasmidial. Os clones foram confirmados por PCR com os oligonucleotídeos vetor-específicos T7 promotor (senso) e M13 reverso (antissenso), e por digestão com as enzimas *Bam*HI e *Xho*I, de acordo com instruções do fabricante. Um dos clones positivos (pGEM-DEN-1_ns2b_{CF40}-gly-ns3_{PRO}) foi propagado em meio LB, seguido de uma extração plasmidial em média escala (*PureLink™ HiPure Plasmid Midiprep Kit, Invitrogen*). O resultado da extração foi quantificado em espectrofotômetro e por eletroforese em gel de agarose.

Para ligação no vetor de expressão pET-19b, 6 µg do plasmídeo pGEM-DEN1_ns2b_{CF40}-gly-ns3_{PRO} foram submetidos a uma digestão dupla de 4 horas a 37 °C com as enzimas *Bam*HI e *Xho*I em tampão *REact® 2* (Invitrogen) por conter menos sal. O resultado da digestão foi submetido à eletroforese em gel de agarose 0,8% e a banda correspondente ao inserto liberado foi excisada e submetida à purificação por kit (*illustra™ GFX™ PCR DNA and Gel Band Purification Kit, GE Healthcare*). O fragmento purificado foi ligado ao vetor pET-19b previamente digerido com as mesmas enzimas de forma a manter o gene de interesse em fase com a cauda de poli-histidina do vetor. As condições de ligação, incluindo

razões molares e temperaturas, foram as recomendadas no *pET System Manual* (Novagen, 11^a edição). O produto da ligação foi transformado por choque térmico em bactérias competentes TOP10, da mesma forma como descrito para DH5 α . As bactérias transformadas foram semeadas em placas de Petri com LB-ágar contendo ampicilina (100 μ g/ml) como fator de seleção. Clones foram selecionados, analisados por mini-preparação de DNA plasmidial, PCR com o oligonucleotídeo vetor-específico T7 terminador (antissenso) e com o DEN-1_NS2BF (senso). Um dos clones positivos foi selecionado (pET-19b_DEN1-ns2b_{CF40}-gly-ns3_{PRO}) para preparação em média escala e sequenciamento.

O plasmídeo extraído pelo kit comercial *PureLink™ HiPure Plasmid Midiprep* foi enviado ao Setor de Sequenciamento de DNA do Centro de Estudos do Genoma Humano da Universidade de São Paulo (USP), onde foi submetido a sequenciamento com os oligonucleotídeos DEN-1_NS2BF (senso) e T7 terminador (antissenso). O cromatograma foi analisado e comparado com a sequência previamente depositada do genoma do isolado DEN-1 DF01-HUB01021093. A sequência consenso do sequenciamento da fita senso e da antissenso foi obtida por meio de alinhamento com o programa ClustalW e edição manual.

4.1.3 Expressão em bactérias BL21(DE3) e BL21-AI™

Aproximadamente 100 ng do plasmídeo pET-19b_DEN1-ns2b_{CF40}-gly-ns3_{PRO} foram utilizados para transformar bactérias quimicamente competentes BL21(DE3) por choque térmico como descrito anteriormente. Colônias foram escolhidas e testadas para produção de proteína recombinante mediante indução com 1 mM de IPTG. Brevemente, culturas de células transformadas foram cultivadas em meio LB adicionado de ampicilina (100 μ g/ml) até atingirem uma OD₆₀₀ de aproximadamente 0,6. Aliquotas foram retiradas e armazenadas em glicerol a -80 °C. Em seguida, IPTG foi adicionado em uma concentração final de 1 mM, e as culturas incubadas a 37 °C por 3 horas sob agitação constante de 200

RPM. Células em suspensão foram centrifugadas a 5000 x *g* por 10 minutos, ressuspensas e lisadas com o reagente *BugBuster*[®] (Novagen). As frações solúveis e insolúveis foram analisadas por SDS-PAGE. O mesmo método foi utilizado para transformar bactérias *BL21-AI*TM (exceto que a indução foi realizada quando a OD600 alcançou 0,4 e com a adição de 0,2% de L-arabinose além do IPTG, de acordo com instruções do fabricante) e analisar a expressão de clones individuais, sendo que essa linhagem celular apresentou um maior rendimento do que a primeira. Um dos clones transformados em *BL21-AI*TM foi escolhido, cultivado em meio LB, separado em alíquotas com glicerol e armazenado a -80 °C. Foram testadas várias concentrações de IPTG (1 mM, 0,5 mM, 0,1 mM) nas temperaturas de 37 °C e 20 °C, em intervalos de 3 ou 16 horas.

4.1.4 Purificação por cromatografia em coluna de Ni²⁺

Para a determinação da concentração ideal de imidazol para a purificação da enzima recombinante, extratos celulares obtidos por lise com *BugBuster*[®] foram incubados a 4 °C por 20 minutos em 5 tubos do tipo *ependorf* com 100 µl de resina assentada (*His•Bind*[®] *Purification Kit*, Novagen) previamente lavada com água Milli-Q, carregada com NiSO₄ (*Charge Buffer*) e equilibrada com tampão de lavagem (*Binding Buffer*). Antes da eluição, cada um dos cinco tubos passou por uma etapa de lavagem com 10 volumes de tampão de lavagem (*Wash Buffer*) com concentrações de imidazol que variavam de 10 a 100 mM. Uma vez estabelecida a concentração de lavagem mais eficiente, uma purificação em maior escala foi realizada pela adição de extrato celular obtido por sonicação (3 ciclos de 3 minutos, amplitude 40, pulsos de 7 segundos com intervalos de 1 segundo entre eles) aplicados a coluna de níquel com afinidade por histidina. A coluna, formada por uma matriz de 0,5 ml de resina, foi previamente lavada com H₂O Milli-Q, carregada com NiSO₄ (*Charge Buffer*) e ativada com tampão de ligação (*Binding Buffer*: Tris-HCl 20 mM pH 7,9, NaCl

500 mM, imidazol 5 mM) antes da aplicação do extrato solúvel. Após a passagem da amostra, a coluna foi lavada com 10 volumes de tampão de ligação e 10 volumes de tampão de lavagem (*Wash Buffer* – Tris-HCl 20 mM pH 7,9, NaCl 500 mM, imidazol 100 mM). As amostras contendo a enzima purificada foram eluídas em frações de 0,5 ml com tampão de eluição (*Elute Buffer* – Tris-HCl 20 mM, pH 7,9, NaCl 500 mM, imidazol 300 mM).

Ao fim da purificação, a enzima foi dialisada e concentrada em um filtro *Amicon*[®] *Ultra-15* com membrana *Ultracell-PL 30 kDa* e tampão Tris-HCl pH 8,5. Glicerol estéril foi adicionado antes da armazenagem em freezer -20 °C, em uma concentração final de 50% (v/v). O resultado foi quantificado pelo método de leitura de absorbância direta com subtração de contaminantes de acordo com a fórmula a seguir, onde [P] é a concentração de proteína em g/L, A280 é o valor da absorbância obtido mediante leitura à 280 nm e A260 mediante leitura à 260 nm:

$$[P] = (1,55 \times A280) - (0,76 \times A260)$$

4.1.4.1 Eletroforese desnaturante em gel de poliacrilamida (SDS-PAGE)

Para analisar a separação das proteínas do extrato total, monitorar as etapas de purificação da protease recombinante bem como estimar sua massa molecular, as amostras foram submetidas à *SDS-PAGE* 12%, de acordo com o protocolo descrito por LAEMMLI (1970). O tampão de amostra, composto por 0,1% de SDS e 50 µM de β-mercaptoetanol, foi utilizado para diluir as amostras, que eram fervidas durante 5 min antes da aplicação no gel. A eletroforese dos géis foi realizada com voltagem constante de 90 V na fase de concentração das amostras e 150 V na fase de separação das proteínas. Os géis foram corados com azul de Coomassie para a visualização das bandas e a massa molecular estimada das proteínas foi determinada pela utilização dos marcadores comerciais *BenchMarkTM Protein Ladder* (Invitrogen).

4.2 Atividade e inibição enzimática

4.2.1 Ensaio de atividade enzimática

O peptídeo sintético benzoil-norleucil-lisil-arginil-arginina-7-amino-4-metilcumarina (Bz-Nle-Lys-Arg-Arg-AMC), correspondente ao substrato otimizado para a NS3 do vírus DEN-1 (LI *et al.*, 2005) foi adquirido da empresa Peptide2.0. O peptídeo foi diluído em uma solução estoque (50 mM) com 10% de metanol, de acordo com instruções do fabricante. Aliquotas de subestoque foram diluídas em tampão de atividade (25 mM Tris-HCl pH 8,5) em uma concentração de 500 μM . Para titulação enzimática, uma diluição seriada da enzima foi realizada em triplicata, utilizando 12 dos 96 poços de uma placa. As diluições foram realizadas em tampão de atividade a partir de uma concentração de 1,34 μM de enzima, sucessivamente dividida por 2 até a concentração de 0,168 μM considerando um volume final de 100 μl (volume total da reação). Em seguida, após adição simultânea de 10 μl do substrato (concentração final de 50 μM) com o uso de uma micropipeta multicanal, a hidrólise do substrato foi monitorada a partir da fluorescência emitida pelo AMC (excitação de 360 nm e emissão de 460 nm) em espectrofluorímetro *SpectraMax M5* (Molecular Devices) por 10 minutos em intervalos de 15 segundos. As leituras foram convertidas para valores de velocidade máxima ($V_{\text{Máx}}$) em medidas de unidades de fluorescência por segundo.

4.2.2 Ensaios de inibição enzimática

Os 30 inibidores (21 identificados para o sítio ativo e 9 para o sítio do cofator) foram adquiridos da empresa *Life Chemicals Inc* (Burlington, Canadá). Por se tratarem de compostos não hidrossolúveis, os mesmos foram diluídos em dimetilsulfóxido (DMSO), como instruído pelo fabricante. Já que a atividade enzimática pode ser influenciada pela concentração de DMSO em solução, foi necessária a realização de uma curva de tolerância, evitando assim quaisquer erros induzidos pela adição desse solvente aos inibidores. Para

isso, 6 reações foram realizadas, com concentrações de DMSO variando de 0 a 5% (v/v). Essa curva nos possibilitou definir em 25 mM a concentração estoque dos inibidores, uma vez que a concentração de DMSO não ultrapassaria 1,6% em uma reação de 100 µl com 400 µM do inibidor. Para os testes de triagem dos compostos, a concentração de 200 µM de inibidor foi escolhida, uma vez que moléculas com atividade inibitória acima desse valor não teriam valor terapêutico. As reações foram realizadas em triplicatas, com um branco para cada inibidor para evitar erros originados por autofluorescência. Para cálculos do IC50, efetuou-se uma diluição seriada partindo de 400 µM dos inibidores mais promissores (compostos nº 1, 15, 21 e 26), com fator de diluição 2 a cada poço da série, até uma concentração de aproximadamente 0,78 µM. Para evitar que a inibição proveniente do solvente pudesse interferir nos resultados, a concentração de DMSO foi mantida constante em cada poço (1,6% v/v). As demais condições da reação foram as mesmas dos ensaios de atividade sem inibidores (tampão de atividade Tris-HCl 25 mM, pH 8,5, 50 µM de substrato e 0,168 µM de enzima). A regressão não-linear para determinação do IC50 foi realizada com o programa *Prism 5.0* (GraphPad Inc).

4.3 Modelagem por Homologia, Dinâmica e Ancoragem Molecular

4.3.1 BLAST/Sequências

Todas as buscas de sequências e estruturas homólogas, assim como todas as sequências utilizadas para os alinhamentos foram obtidas do banco de dados do *National Center for Biotechnology Information*⁸ (NCBI). Seus códigos de acesso se encontram na tabela 3 a seguir. Para a modelagem por homologia, as sequências utilizadas para os alinhamentos foram extraídas dos arquivos PDB obtidos do *Protein DataBank*.

⁸ Disponível em: <http://www.ncbi.nlm.nih.gov/>

Tabela 3. Códigos de acesso das sequências utilizadas.

Nome	Fonte	Nº Acesso	Referência
DENV 1, DF01-HUB01021093	GenBank	FJ384655.1	Não publicado
DENV 1 NS2B-NS3 Protease	PDB	3L6P	(CHANDRAMOULI <i>et al.</i> , 2010)
DENV 2 NS2B-NS3 Protease	PDB	2FOM	(ERBEL <i>et al.</i> , 2006)
DENV 4 NS3 Protease-Helicase	PDB	2VBC	(LUO <i>et al.</i> , 2008)
	PDB	2WHX	(LUO <i>et al.</i> , 2010)
MVEV NS3 Protease-Helicase	PDB	2WV9	(ASSENBERG <i>et al.</i> , 2009)
WNV NS2B-NS3 Protease	PDB	2FP7	(ERBEL <i>et al.</i> , 2006)
	PDB	2IJO	(ALESHIN <i>et al.</i> , 2007)
	PDB	3E90	(ROBIN <i>et al.</i> , 2009)
<i>T. cruzi</i> Cathepsin B	GenBank	AF399838.1	(NÓBREGA <i>et al.</i> , 1998)
<i>T. brucei</i> Cathepsin B	PDB	3HHI	(KERR <i>et al.</i> , 2010)
	PDB	3MOR	(KOOPMANN <i>et al.</i> , 2012)
<i>T. cruzi</i> MTAP	GenBank	AY144609.1	Não publicado
<i>S. tokodaii</i> MTAP	PDB	1V4N	Não publicado
<i>A. pernix</i> MTAP	PDB	1WTA	Não publicado
Human MTAP	PDB	1CB0	(APPLEBY <i>et al.</i> , 1999)

4.3.2 Alinhamento

Os alinhamentos automatizados foram realizados com o programa *ClustalW 2* (LARKIN *et al.*, 2007) com definições padrão no servidor EMBL-EBI⁹ (GOUJON *et al.*, 2010). Os alinhamentos manuais e refinamentos foram realizados no *Discovery Studio Visualizer 3.1* (Accelrys), baseado tanto em estruturas secundárias preditas quanto em informações contidas nos arquivos PDB (informações sobre estrutura secundária e lacunas). Estruturas com múltiplas cadeias tiveram suas sequências editadas de forma a unir essas cadeias em apenas uma sequência, considerando as informações biologicamente relevantes (construção

⁹ Disponível em: <http://www.ebi.ac.uk/Tools/msa/clustalw2/>

de complexos, processamento pós-traducional, etc). As figuras dos alinhamentos de sequência foram produzidas com o programa ALINE¹⁰ (BOND e SCHÜTTELKOPF, 2009).

4.3.3 Sobreposição de estruturas

A sobreposição de estruturas foi realizada com o programa *Discovery Studio Visualizer 3.1* (Accelrys Inc), a partir dos alinhamentos de sequência.

4.3.4 Modelagem por homologia

Antes da modelagem por homologia, as estruturas guias foram editadas para a remoção de cadeias não utilizadas, ligantes, água e heteroátomos. A sequência incompleta (com lacunas estruturais) contida nos arquivos PDB foi extraída e alinhada com base nas sequências completas depositadas pelos autores no formato fasta. Baseada nesse alinhamento, uma sobreposição estrutural entre as estruturas foi então realizada, de forma a guiar o alinhamento final a ser submetido ao programa MODELLER (ESWAR *et al.*, 2006). Cinco modelos foram gerados para cada alvo, cada um com mais três modelos com *loops* refinados, gerando um total de 15 modelos. O modelo de energia mais baixa obtido para cada alvo foi então verificado pelo programa PROCHECK (LASKOWSKI *et al.*, 1993). Maiores detalhes sobre a criação dos modelos da protease do vírus da dengue podem ser obtidos no manuscrito (em anexo).

4.3.5 Dinâmicas moleculares

Antes de serem submetidos à dinâmica molecular, os modelos foram adicionados de hidrogênios de acordo com seus estados de protonação em pH 7,0. Além disso, por meio do *plugin* AutoPSF contido no VMD¹¹ (HUMPHREY, 1996) foram atribuídos estados delta-protonados às histidinas. Os sistemas foram solvatados com moléculas do modelo explícito

¹⁰ Disponível em: <http://crystal.bcs.uwa.edu.au/px/charlie/software/aline/>

¹¹ Disponível em: <http://www.ks.uiuc.edu/Research/vmd/>

de água TIP3 em uma caixa com pelo menos 15 Å de distância de qualquer ponto de referência da proteína, onde mais tarde foram adicionados íons de Na⁺ ou Cl⁻ para garantir uma carga eletrostática neutra ao sistema. Parâmetros não existentes para os ligantes foram adicionados usando o campo de forças CHARMM22 (BROOKS *et al.*, 2009). As simulações foram computadas com o programa NAMD¹² (PHILLIPS *et al.*, 2005), usando condições periódicas de controle e uma abordagem do tipo malha de partículas de Ewald (*Particle Mesh Ewald*, ou PME; DARDEN *et al.*, 1993) para tratar interações eletrostáticas de longa distância. A dinâmica foi realizada nas condições normais de temperatura e pressão, usando dinâmicas de Langevin e um pistão para estabelecer temperaturas (300 K) e pressão (1 atm) constantes (FELLER *et al.*, 1995). Cada sistema passou por uma minimização de energia (6400 etapas de gradiente conjugado), após a qual foram equilibrados por 500 ps antes de ter gravada suas trajetória. Os *frames* da trajetória foram gravados em intervalos de 1 ps.

4.3.5.1 RMSD/Estrutura secundária/Agrupamento conformacional

Após as simulações de DM, a estrutura de cada quadro da trajetória foi alinhada com base nos átomos pesados das proteínas (ou de parte delas). O desvio médio quadrático (RMSD) entre cada quadro foi então calculado pelo *plugin RMSD Trajectory Tool*, contido no VMD. Os dados foram exportados como uma planilha ASCII por meio do módulo Multiplot. A evolução da estrutura secundária ao longo do tempo foi monitorada pelo *plugin Timeline*, também contido no VMD. Para o agrupamento de conformações em família, um *script tcl* desenvolvido no *Laboratoire Lorrain de Recherche en Informatique et ses Applications* (LORIA) foi utilizado (ClusterRMSD), com um limite de corte de 3,5 Å. Esse *script* constrói uma matriz bidimensional para identificar famílias de conformações relacionadas entre si dentro do limite estabelecido.

¹² Disponível em: <http://www.ks.uiuc.edu/Research/namd/>

4.3.5.2 PCA

Os modos conformacionais e dados sobre a mobilidade das proteínas foram obtidos com a biblioteca ProDy, juntamente com o *plugin* NMWiz (BAKAN *et al.*, 2011) incluído no VMD. Para tanto, os movimentos translacionais e rotacionais foram removidos por meio do alinhamento estrutural entre as estruturas de cada quadro da trajetória.

4.3.5.3 Interações entre pares

As forças de interação do tipo van der Waals foram computadas aos pares com o programa NAMD. Resumidamente, os arquivos PDB dos modelos de entrada foram editados por *scripts* desenvolvidos no LORIA de modo a selecionar ligante, solvente, cofator, proteína e seus resíduos como grupos, seguido de cálculos de energia do NAMD entre pares desses grupos. Os resultados são dados em $\text{kcal.mol}^{-1} \text{ \AA}^{-1}$, e indicam a “resistência” da interação. Os cálculos envolvendo os solventes são realizados primeiramente. Em seguida, as moléculas de água foram eliminadas dos arquivos de trajetórias dcd por meio de um *script* tcl, permitindo cálculos mais velozes para os pares subsequentes. As interações entre resíduos da proteína ou do cofator foram plotadas como médias, de forma a permitir a visualização das regiões de contatos. Já quando plotadas individualmente, nos permitiram verificar suas flutuações durante a trajetória.

4.3.5.4 Detecção/monitoramento de cavidades

As cavidades foram detectadas pelo programa METAPOCKET 2.0¹³ (ZHANG *et al.*, 2011), um algoritmo que se baseia na combinação de várias ferramentas já estabelecidas na detecção de cavidades para melhorar a detecção. Mudanças no volume e superfície das cavidades foram monitoradas com o programa MDPocket¹⁴ (SCHMIDTKE *et al.*, 2011).

¹³ Disponível em: <http://projects.biotec.tu-dresden.de/metapocket/>

¹⁴ Disponível em: <http://fpocket.sourceforge.net/>

4.3.5.5 Visualização de dados, gráficos e regressões não-lineares

Representações moleculares foram renderizadas pelo *plugin Tachyon Ray Tracer* contidos no VMD, ou ainda pelo *Discovery Studio Visualizer 3.1*. Todos os gráficos (RMSD, interações entre pares, gráficos de mobilidade, volume de cavidades, etc) foram plotados com o *GraphPad Prism 5.0* (GraphPad Software Inc).

4.3.6 Ancoragem molecular (*docking*)

As conformações representativas de cada alvo foram submetidas à ancoragem molecular virtual por meio do programa GOLD 5.1 (CCDC, Cambridge, Reino Unido). No caso da protease do DENV, a flexibilidade do alvo foi explorada por meio das quatro conformações escolhidas para montar um conjunto de estruturas submetido à técnica de *ensemble docking*. Para explorar a flexibilidade dos ligantes, o número de execuções do algoritmo genético do GOLD foi estabelecido em 50. O método de ranqueamento dos resultados escolhido foi o GoldScore, que leva em consideração energias de pontes de hidrogênio e van der Waals entre ligante e proteína, assim como torção e energias de van der Waals do ligante. Para detecção da cavidade a ser usada no *docking*, foi utilizado um algoritmo de inundação em uma esfera de 20 Å de raio, a partir do ponto representando o centro de massa do ligante, que foi removido do modelo antes da execução dos cálculos de acoplamento. Além disso, a função de *early termination* foi ligada, de modo a evitar execuções repetidas desnecessárias.

A coleção de compostos escolhida para nossa campanha veio do catálogo da empresa *Life Chemicals*¹⁵ (Burlington, Canada), e incluiu quimiotecas contra proteases do tipo quimotripsina-, fator Xa-, tripsina-, e trombina-símbios, além de bibliotecas com atividade anti-HCV, anti-malária e inibidores de interação proteína-proteína e de metiltransferases.

¹⁵Acessível em: <http://www.lifechemicals.com>

Elas foram transferidas como arquivos SDF e convertidas para arquivos MOL2 3Ds por meio do programa CORINA (Molecular Networks, SADOWSKI *et al.*, 1994).

4.3.6.1 Tanimoto

Os arquivos de soluções obtidos a partir da campanha de *virtual screening* foram convertidos para códigos SMILES e em seguida submetidos a dois *scripts* desenvolvidos no LORIA para a geração de uma matriz de similaridade entre as moléculas, baseada no coeficiente de Tanimoto. Nessa matriz são atribuídos números de 0 (totalmente diferentes) a 1 (idênticas), que denotam o grau de similaridade estrutural entre duas moléculas. Essa tabela nos auxiliou no processo de escolha dos compostos a serem testados nos ensaios de inibição *in vitro*, evitando testar compostos redundantes.

4.3.6.2 Ligplots

Os modos de interação entre os compostos identificados pelo *virtual screening* e seus alvos foram plotados pelo programa *Discovery Studio Visualizer 3.1*, por meio da função *Analyze complexes → Show 2D diagrams*.

5 Resultados e Discussão

5.1 Dengue

A princípio, o projeto se baseava no desenvolvimento de um modelo da NS2B_{CF}-Gly_[4]-Ser-Gly_[4]-NS3_{PRO} em sua conformação ativa para fins de busca de inibidores competitivos para o sítio ativo. Para tanto, foram desenvolvidos dois modelos preliminares (sem o cofator) da NS3_{PRO}, baseados exclusivamente nas estruturas depositadas para o vírus da dengue (DENV) ou para o vírus da febre do Nilo ocidental (WNV). Nesta primeira etapa, percebemos que existiam características únicas a cada modelo, que foram unidas em um último modelo, no qual também foram adicionados o cofator e o ligante. Para avaliarmos a estabilidade e comportamento do sítio ativo desses modelos, os mesmos foram submetidos a simulações de dinâmica molecular (DM). Entretanto, durante as simulações, foi publicado um artigo (KNEHANS *et al.*, 2011) no qual foi descrito um trabalho muito próximo ao que estávamos propondo. O modelo utilizado no referido trabalho foi bem elaborado, e, ao contrário das tentativas anteriores (FRIMAYANTI *et al.*, 2011, 2012; GANESH *et al.*, 2005; TOMLINSON e WATOWICH, 2011; TOMLINSON *et al.*, 2009), metuculoso. Foi o primeiro trabalho que realizou *virtual screening* em um modelo obtido por homologia na conformação “fechada”, com o cofator modelado baseado no homólogo de WNV. Além disso, realizaram *fragment-based docking*, uma estratégia muito promissora.

Ademais, no início de 2012, foi publicada a estrutura “fechada” da protease do DENV, juntamente com um inibidor peptídico (NOBLE *et al.*, 2012). Tal estudo, apesar de responder a uma das principais questões acerca da protease do vírus da dengue (se ela se comporta ou não como sua contraparte homóloga do WNV), abriu espaço para muitas outras perguntas. Uma delas é porque a protease NS3 apresenta a mesma conformação, independentemente do cofator estar na posição fechada (com o inibidor peptídico) ou aberta (com a aprotinina).

Nossa estratégia se baseou, portanto, em outra abordagem: a exploração da flexibilidade dessa protease para a descoberta de novas moléculas promissoras. Isso é algo inédito, apesar de ter sido abordado em 2009 (EKONOMIUK *et al.*, 2009) para a protease do WNV. Ainda assim, são grandes as diferenças entre nossa abordagem e a de Ekonomiuk e colaboradores: enquanto nós realizamos mais de 30 nanossegundos de dinâmicas moleculares, eles usaram dois nanossegundos; nós buscávamos famílias de conformações, eles buscaram apenas uma conformação que recebesse melhor um fragmento de inibidor. Além disso, na busca por explicações para a formação do ajuste induzido, resolvemos também estudar o papel do cofator e do ligante separadamente na estabilidade da NS3. Para isso, realizamos dinâmicas moleculares com modelos contendo ou não cada um desses componentes, tanto separadamente como associados. Devido à realização dessas DMs, encontramos outra abordagem para o *virtual screening*: em todos os modelos submetidos à DM sem o cofator, o sítio de ligação do mesmo permanecia disponível durante toda a simulação, representando então um possível alvo. Essa hipótese pode ser reforçada se presumirmos que a simulação representa o estado de equilíbrio da protease, onde esse sítio estaria conservado e sujeito à ligação de pequenas moléculas.

Da primeira abordagem extraímos um manuscrito que trata do agrupamento em famílias das conformações percebidas durante a simulação. Por meio de análises como: PCA, interações entre pares e gráficos de interação entre a enzima e seu inibidor (*ligplots*), pudemos demonstrar que cada representante das famílias identificadas correspondia a um estado diferente de conformação, e que por isso estariam sujeitos à técnica de *ensemble docking*. O segundo manuscrito dessa série tratará do *docking* e dos testes *in vitro* dos inibidores encontrados nessa campanha de *virtual screening*, incluindo também o sítio de ligação do cofator como alvo.

Da segunda abordagem, esperamos encontrar correlações entre a presença dos componentes do sistema e a estabilidade do modelo durante a dinâmica. Dessa forma, poderemos sugerir papéis para cada um deles na formação do ajuste induzido. Um dos dados mais interessantes é que a retirada do inibidor não implica na perda da conformação “fechada” da enzima. Para comprovar seria necessário um teste de NMR com inibidores não covalentes (talvez um dos descobertos nesse trabalho?).

5.1.1 Clonagem e Expressão

5.1.1.1 Amplificação e fusão dos fragmentos correspondentes à NS2B_{CF} e NS3_{PRO}

As reações de PCR realizadas a partir de um clone cedido pelo Prof. Tatsuya Nagata com os oligonucleotídeos sintetizados resultaram em produtos com massas moleculares estimadas de aproximadamente 180 pb para o fragmento NS2B_{CF40}-Gly, e de 600 pb para o fragmento Gly-NS3_{PRO} (figura 11A), em concordância com os esperados. Com aproximadamente 750 pb (figura 11B e C), uma terceira reação de PCR, contendo como DNA molde os produtos das outras reações, também apresentou a massa molecular esperada, o que indica que o processo de fusão entre os dois fragmentos foi bem sucedido.

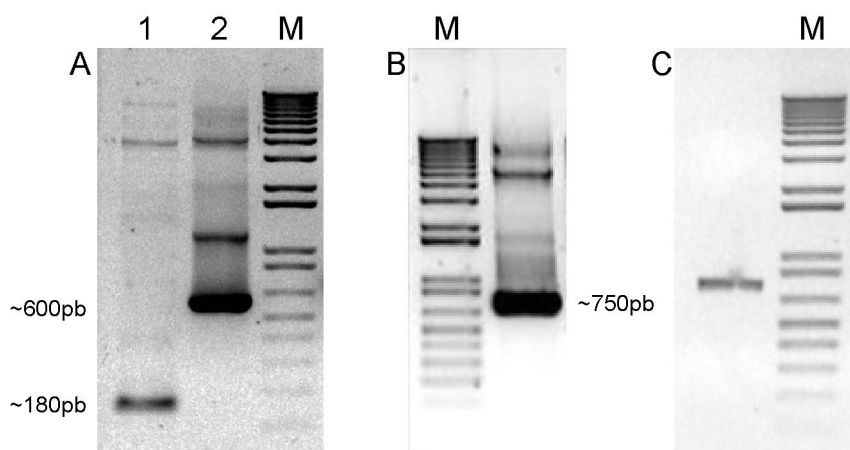


Figura 11. Amplificação dos fragmentos NS2B_{CF}-gly e gly-NS3_{PRO} obtidos por PCR em gel de agarose 8%. A) Fragmentos NS2B_{CF40}gly (1) e glyNS3_{pro} (2). B) Produto de PCR de fusão obtido a partir dos produtos previamente amplificados NS2B_{CF40}gly e glyNS3_{pro}. C) Resultado da eluição da banda correspondente ao produto de fusão NS2B_{CF40}glyNS3_{pro}, posteriormente utilizado para ligação nos vetores de subclonagem. M, marcador de massa molecular 1 kb + DNA Ladder (Invitrogen).

5.1.1.2 Clonagem nos vetores pGEM®-T Easy e pET-19b

O fragmento NS2B_{CF40glyNS3_{PRO}} foi purificado do gel de agarose e em seguida submetido à reação de ligação em vetores comerciais do tipo T/A. Inicialmente, o vetor escolhido foi o pCR®2.1 (Invitrogen), de massa molecular de 3,9 kb, já disponível no laboratório. Entretanto, após algumas tentativas falhas de clonagem no vetor de expressão (dados não mostrados), a estratégia foi alterada e o vetor escolhido passou a ser o pGEM®-T Easy (Promega), de massa molecular de 3015 pb. Esse vetor mostrou-se mais apropriado, uma vez que não possuía os sítios de restrição utilizados para a clonagem no vetor de expressão, facilitando o processo de digestão do inserto nos sítios exclusivamente inseridos pela reação de PCR. Após a ligação no vetor pGEM®-T Easy, colônias brancas e uma colônia azul foram submetidas a triagem por fenol:clorofórmio:iaa (figura 12A), e os clones com plasmídeos de massa molecular compatíveis com a esperada foram selecionados. Seus DNAs plasmidiais foram extraídos por lise alcalina (figura 12B) e posteriormente confirmados por digestão (figura 12C), liberando um inserto de tamanho esperado. Dos clones positivos, um foi escolhido, cultivado em meio LB e submetido à extração plasmidial em média escala (figura 12D), como descrito em material e métodos. Em seguida, aproximadamente 6 µg desse material foi submetido à digestão enzimática pelas enzimas *Xho*I e *Bam*HI, e o inserto digerido purificado a partir do gel (figura 12E).

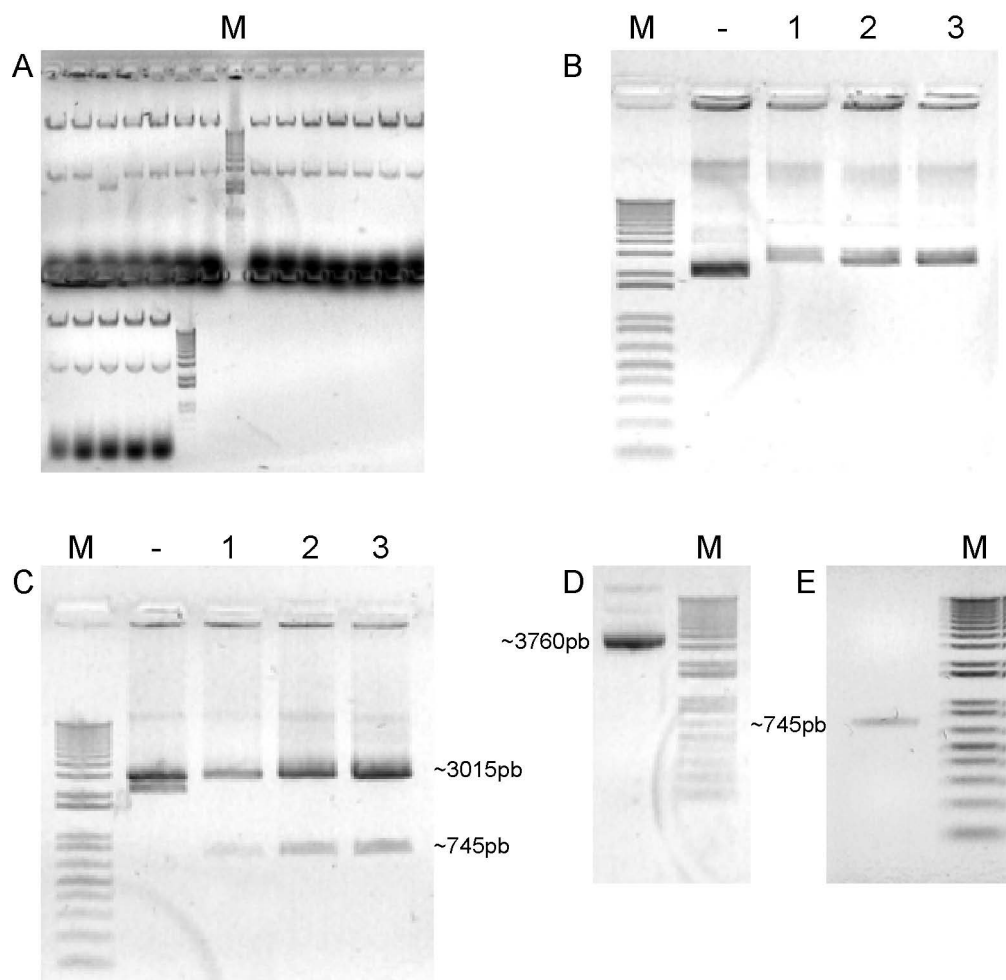


Figura 12. Padrão de migração do DNA plasmidial de clones do produto de ligação pGEM®-T Easy e do fragmento NS2B_{CF40}-Gly-NS3_{PRO}. A) Extração plasmidial por fenol:clorofórmio:iaa. Uma colônia negativa (-) e três colônias positivas (1, 2 e 3) foram escolhidas para purificação por lise alcalina (B) e posterior digestão com a enzima *EcoRI* para confirmação do tamanho de aproximadamente 745 pb do inserto (C). D) Resultado da preparação em média escala linearizada, demonstrando o tamanho correto do vetor+inserto. E) Resultado da eluição da banda correspondente ao inserto liberado da digestão completa pelas enzimas *XhoI* e *BamHI*. M, marcador de massa molecular 1 kb plus DNA ladder (Invitrogen).

O inserto eluído foi então ligado ao vetor pET-19b linearizado. Para facilitar e garantir a linearização total do vetor pET-19b, a digestão foi realizada utilizando um clone de outra enzima recombinante clonada com os mesmos sítios de restrição. Dessa forma, a reação observada no gel corresponderia a uma digestão completa, fato observável pela constatação dos fragmentos correspondentes ao pET-19b e ao inserto previamente clonado nesse vetor. O produto de ligação foi transformado em células TOP10 quimicamente competentes, e, devido a ausência de um sistema de seleção colorimétrico, 5 colônias foram escolhidas aleatoriamente e submetidas à extração plasmidial por lise alcalina (figura 13A). Pelo fato do

vetor pET-19b ser de cópia única por célula, o rendimento excepcional desse tipo de preparação para o clone 3 (figura 13A, poço 3) revelou que esse clone se tratava na verdade do plasmídeo pGEM®-T Easy. Os outros clones, no entanto, apresentaram rendimento compatível com o esperado para um vetor de cópia única por célula. A digestão desses clones para confirmação do tamanho do inserto foi realizada (dados não mostrados), assim como uma PCR com o oligonucleotídeo NS2BF e o oligonucleotídeo vetor-específico T7 terminador (figura 13B), revelando uma banda de tamanho consistente com o esperado. Um dos clones foi escolhido e submetido à extração plasmidial em média escala (figura 13C).

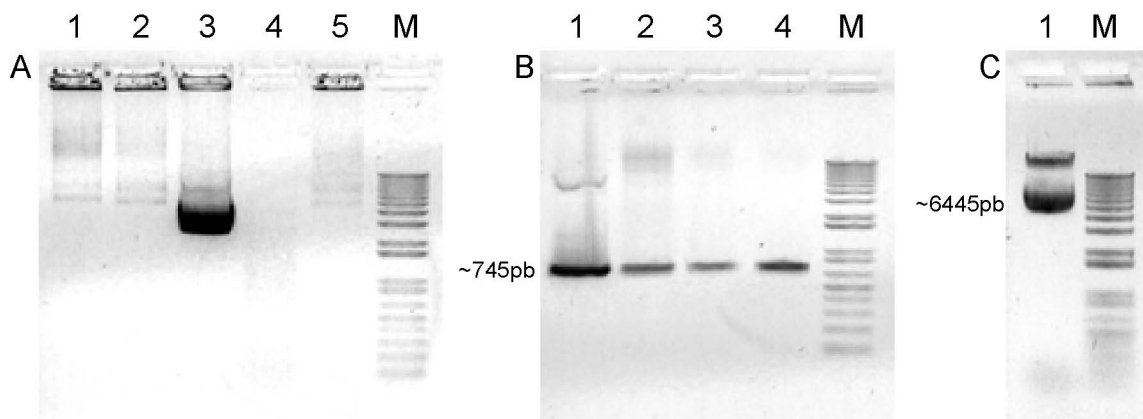


Figura 13. Confirmação dos clones recombinantes de pET-19b-DEN1_NS2B_{CF}NS3_{PRO}. **A)** Resultado da preparação plasmidial por lise alcalina dos clones selecionados (1, 2, 3, 4 e 5). **B)** Confirmação por PCR utilizando um *primer* vetor específico para o pET-19b e outro específico para o gene de interesse. Poços 1, 2, 3 e 4 correspondem aos clones 1, 2, 4 e 5. **C)** Resultado da preparação plasmidial de média escala utilizando kit comercial PureLink™ HiPure Plasmid Midiprep Kit, Invitrogen. M, marcador de massa molecular 1kb plus DNA Ladder, Invitrogen.

5.1.1.3 Sequenciamento do clone pET-19b_NS2B_{CF40}GlyNS3_{PRO}

Após o sequenciamento de cada uma das fitas do clone pET-19b_ns2b_{CF40}-Gly-ns3_{PRO} obtido acima, ambas foram submetidas a um alinhamento, de onde foi extraída a sequência consenso correspondente ao fragmento DEN-1_ns2bCF40-Gly-ns3_{PRO} em fase com a cauda de histidina do vetor pET-19b (figura 14A). O resultado obtido diverge em quatro bases nitrogenadas em relação ao construto baseado na sequência depositada em bancos de dados (nº de acesso GenBank: FJ384655.1), que correspondem a três mutações na sequência de aminoácidos, uma na cadeia NS2B_{CF40} e as outras na NS3_{PRO} (figura 14B).

CCTCTAGAAATAATTTTGTGTTAACTTTAAGAAGGAGATATACCATGGGCCATCATCATCATCATCATCATCATCACAGCAGCG GCCATATCGACGACGACGACAAGCATATGCTCGAGGCCGATCTATCATTTGGAGAAAGCGGCTGAGGTATCCTGGGAAGAAGAAGCA GAACACTCCGGTACCTCACACAACATATTAGTAGAGGTTCAAGATGATGGAACATATGAAAATAAAAGATGAAGAGAGGGATGACAC ACTCGGGGGCGGAGGTAGTGGTGGAGGCGGGTCAGGAGTGTATGGGACACACCCAGCCCTCCGGAAGTGGAAAGAGCGGTTCTTG ATGATGGCATTATAGAATCTTGCAAAGAGGACTGTTGGGCAGGTCCCAAGTAGGAGTGGGAGTTTTCCAAGACGGCGTGTCCAC ACAATGTGGCACGTCACCAGGGGAGCTGTCCCTATGTACCAAGGGAAGAGGCTGGAACCAAGCTGGGCCAGTGTCAAAAAGGACTT GATCTCATATGGAGGAGTTGGAGGTTTCAAGGATCATGGAACACAGGAGAAGAAGTGCAGGTGATAGCTGTTGAACCAGGAAAA ACCCCAAAAATGTACAGACAACGCCGGGTACCTTCAAGACTCCTGAAGCGCAAGTTGGAGCCATAGCTCTAGATTTTAAACCCGGC ACATCTGGATCTCCCATCGTGAACAGAGAGGGAAAAATAGTAGGTCCTTATGGAATGGAGTGGTGACAACAAGTGGAACCTACGT CAGTGCCATAGCCCAAGCTAAAGCATCACAGGAAGGCCCTTACCAGAGATTGAGGACGAGGTGTTTTAAGAAAGGATCCGGTCTGCT AACAAAGCCCCGAAAG		A
DF01-HUB01021093_NS2BglyNS3pro	IDDDDKHMLEADLSLEKAAEVSWEEEAHSGTSH 50	
DEN-1_NS2BglyNS3pro	MGHHHHHHHHHSSGHI DDDDKHMLEADLSLEKAAEVSWEEEAHSGTSH 50 *****	
DF01-HUB01021093_NS2BglyNS3pro	NILVEVQDDGAMKIKDEERDDLGGGGSGGGSGVLWDTSPPEVEGAVL 100	
DEN-1_NS2BglyNS3pro	NILVEVQDDGTMKIKDEERDDLGGGGSGGGSGVLWDTSPPEVERAVL 100 *****.*****	
DF01-HUB01021093_NS2BglyNS3pro	DDGIYRILQRGLLGRSQVGVGFQDGVFHTMWHVTRGAVLMYQGRLEPS 150	
DEN-1_NS2BglyNS3pro	DDGIYRILQRGLLGRSQVGVGFQDGVFHTMWHVTRGAVLMYQGRLEPS 150 *****	
DF01-HUB01021093_NS2BglyNS3pro	WASVKKDLISYGGWRFQGSWNTGEEVQVIAVEPGKNPKNVQTPGTFKT 200	
DEN-1_NS2BglyNS3pro	WASVKKDLISYGGWRFQGSWNTGEEVQVIAVEPGKNPKNVQTPGTFKT 200 *****	
DF01-HUB01021093_NS2BglyNS3pro	PEGEVGAIALDFKPGTSGSPIVNSREGKIVGLYGNVVTTSYVSAIAQA 250	
DEN-1_NS2BglyNS3pro	PEGEVGAIALDFKPGTSGSPIVNSREGKIVGLYGNVVTTSYVSAIAQA 250 *****.*****	
DF01-HUB01021093_NS2BglyNS3pro	KASQEGPLPEIEDEVFKKSGC 272	
DEN-1_NS2BglyNS3pro	KASQEGPLPEIEDEVFKKSGC 272 *****	

Figura 14. Resultados do sequenciamento do plasmídeo pET-19b_ns2b_{cf40}-gly-ns3_{pro} e do alinhamento da cadeia de aminoácidos traduzidos. A) Sequencia consenso obtida por meio do alinhamento entre as fitas senso e antissenso resultantes do sequenciamento. **B)** Alinhamento ClustalW entre a sequencia de aminoácidos da sequencia obtida (DEN-1_NS2B_{CF40}GlyNS3_{PRO}) e a depositada (DF01-HUB01021092, nº de acesso GenBank: FJ384655.1). As substituições podem ser observadas pela falta de continuidade dos asteriscos (*). Espaços () significam substituições não-conservativas; Dois pontos (:), substituições semi-conservativas; e um ponto simples (.), substituições conservativas. Os resíduos cataliticamente ativos (H1526, D1550, S1610) estão destacados por uma caixa preta.

Não se sabe o motivo dessa divergência, uma vez que foram utilizadas enzimas de alta fidelidade para a amplificação e clonagem do construto. Além disso, essas mutações estavam presentes nas duas fitas sequenciadas.

5.1.1.4 Expressão e purificação da NS2B_{CF40}GlyNS3_{PRO} recombinante

Extratos de vários clones de bactérias *E. coli* BL21(DE3) transformados com o vetor pET-19b_NS2B_{CF40}NS3_{PRO} foram analisados em uma varredura por SDS-PAGE para averiguar qual expressava a proteína recombinante mais eficientemente (dados não mostrados). O nível de expressão, normalizado para a OD₆₀₀, variou muito pouco entre os clones testados, assim como a proporção entre proteína presente no extrato solúvel e insolúvel. Um dos

clones foi escolhido e induzido com 1 mM de IPTG, a 37 °C por 3 horas. Os extratos solúveis e insolúveis foram analisados em SDS-PAGE, juntamente com um controle induzido com o vetor pET-19b vazio (figura 15).

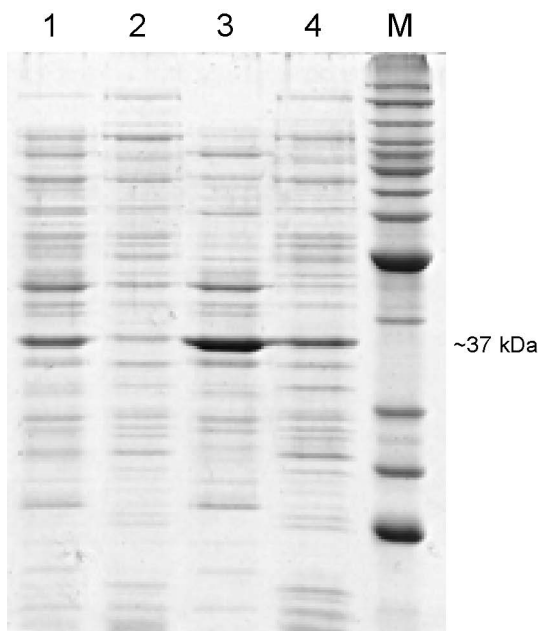


Figura 15. Perfil eletroforético do extrato solúvel e insolúvel de clones contendo o plasmídeo de expressão induzidos. A indução com 1 mM de IPTG resultou num aumento de expressão de uma banda de aproximadamente 37 kDa no extrato solúvel de bactérias transformadas com o vetor pET-19b_NS2B_{CF40}NS3_{PRO} (poço 3), quando comparado ao extrato insolúvel (poço 4) ou com o controle negativo (pET-19b vazio; poço 1, extrato solúvel e poço 2, extrato insolúvel). M, marcador de massa molecular BenchMark™ Protein Ladder, Invitrogen.

Curiosamente, a proteína recombinante apresentou uma massa molecular de aproximadamente 37 kDa em SDS-PAGE, ao invés dos esperados 30 kDa. Entretanto, esses dados são similares aos obtidos por outros trabalhos que utilizaram a mesma técnica de obtenção da enzima fusionada com o cofator NS2B (LEUNG *et al.*, 2001; LI *et al.*, 2005).

Para verificar se o tamanho da proteína estava sendo alterado por algum defeito na linhagem de bactérias BL21(DE3), o vetor de expressão também foi transformado em bactérias competentes da linhagem BL21-AI™ (Invitrogen). A vantagem dessa linhagem de bactérias é que elas possuem o gene da T7 RNA polimerase sob o comando de um promotor forte induzido por L-arabinose. Dessa forma, pode-se controlar a taxa de produção de RNA

mensageiro e, concomitantemente, de expressão da proteína recombinante, algo extremamente útil para expressão de produtos tóxicos.

Uma vez verificado o maior nível de expressão nessas bactérias (dados não mostrados), foi realizada uma varredura das condições de purificação, com diferentes concentrações de imidazol no tampão de lavagem (figura 16A).

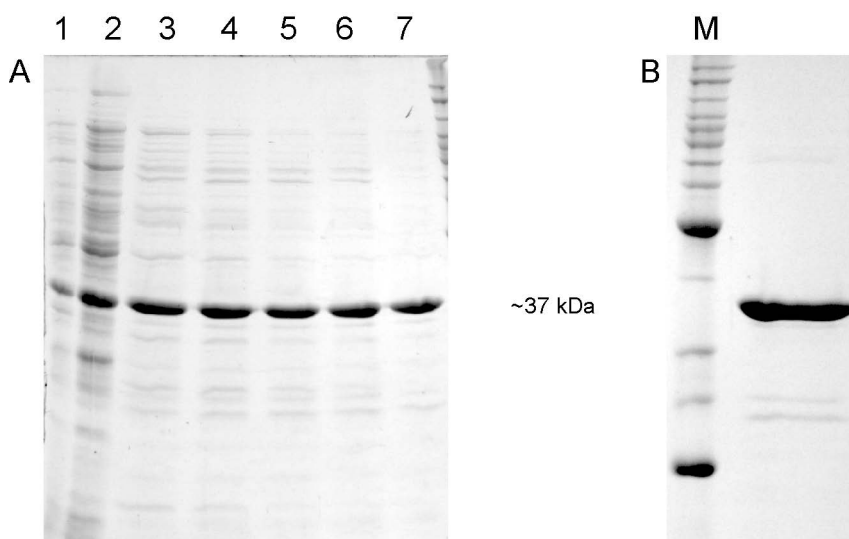


Figura 16. Análise de diferentes concentrações de imidazol no tampão de lavagem durante a purificação da NS2B_{CF}-Gly-NS3_{PRO} recombinante. A) Amostras purificadas pelo método de “batch”, no qual as amostras a serem purificadas foram incubadas com resina em 5 tubos diferentes. 1) Fração insolúvel do extrato celular; 2) fração solúvel; 3) a 7), resultados da eluição, com etapa de lavagem constituída de 10 volumes de tampão de lavagem contendo 10, 20, 30, 50 e 100 mM de imidazol, nessa ordem. **B)** Resultado de purificação em coluna de cromatografia por afinidade a Ni²⁺. Trinta ml de extrato solúvel foram aplicados à coluna, que por sua vez foi lavada com 10 volumes de tampão de lavagem com 100 mM de imidazol. M, marcador de massa molecular BenchMark™ Protein Ladder, Invitrogen.

Nas condições testadas, a proteína apresentou bastante afinidade pela resina mesmo em concentrações de até 100 mM de imidazol, havendo pouca perda de material com lavagens de até 10 volumes de tampão de lavagem com essa concentração. O próximo passo foi o de executar uma purificação em maior escala, utilizando para isso uma coluna de cromatografia ao invés dos tubos *ependorf*. A utilização de um fluxo constante de tampão de lavagem em uma coluna de cromatografia impede que proteínas retiradas da resina voltem a se grudar, promovendo assim uma purificação mais efetiva. Após estabelecer as concentrações de imidazol satisfatórias, 50 ml de extrato celular solúvel de bactérias induzidas foram passados em 0,5 ml de resina previamente lavada, carregada e equilibrada,

como descrito em material e métodos. Ao final de uma etapa de lavagem com 10 volumes de tampão de lavagem com 100 mM de imidazol, a proteína recombinante foi eluída em tampão de eluição contendo 300 mM de imidazol (figura 16B).

Os contaminantes de massa molecular abaixo de 30 kDa foram eliminados durante a etapa de concentração e diálise em filtro tipo Amicon® (Millipore Inc). O processo de concentração promoveu, portanto, o enriquecimento da enzima recombinante em comparação aos contaminantes, gerando uma amostra com elevado grau de pureza, propícia aos estudos de atividade enzimática.

5.1.2 Modelagem por Homologia

5.1.2.1 BLAST

A primeira etapa durante o processo de modelagem por homologia é a escolha de *templates*, ou “estruturas guias”, para a derivação do modelo. De forma a avaliar as estruturas cristalográficas apropriadas à construção dos nossos modelos, executamos uma busca do tipo blastx nos servidores do *National Center for Biotechnology Information* (NCBI), partindo da sequência de nucleotídeos obtida do sequenciamento do vetor de expressão. A busca, restrita ao banco de dados de estruturas cristalográficas PDB, retornou 27 resultados, dentre os quais alguns relacionados ao domínio do cofator NS2B_{CF} e alguns relacionados ao domínio NS3_{PRO}. Removendo estruturas contendo mutações sítio-dirigidas e outras não pertinentes, chegamos a uma lista contendo oito estruturas, sendo quatro relacionadas ao vírus da dengue (DENV), três ao vírus da febre do Nilo ocidental (WNV), e uma relacionada ao vírus da encefalite de Murray Valley (MVEV). Estruturas desprezadas incluíam a protease Spl de *Staphylococcus aureus* e o fator de crescimento epitelial de drosófilas, por exemplo. Os valores de identidade e similaridade relativas à sequência

consultada, assim como o código identificador PDB ID das estruturas homólogas, se encontram na tabela 4.

Tabela 4. Identidade e similaridade de estruturas homólogas à NS3 de DENV1 depositadas no PDB.

Origem	PDB ID	Identidade	Similaridade	Cobertura	Lacunas
DENV1	3L6P	93,09%	93,90%	88,8%	4,88%
DENV2	2FOM	66,38%	82,55%	66,4%	0,43%
DENV4	2WHX	62,84%	81,42%	66,1%	0,55%
DENV4	2VBC	63,68%	81,59%	66,1%	0,50%
WNV	2IJO	52,42%	70,93%	66,4%	0,89%
WNV	3E90	53,67%	72,02%	63,2%	0,46%
WNV	2FP7	52,61%	72,04%	60,2%	0,00%
MVEV	2WV9	53,31%	69,01%	87,4%	2,07%

5.1.2.2 Alinhamento

Com as prováveis estruturas *template* em mãos, a próxima etapa foi a de alinhar manualmente suas sequências (figura 17), baseando-nos em dados sobre a construção dos complexos NS2B_{CF}/NS3_{PRO} de cada espécie. A estrutura 3L6P, por exemplo, possui uma maior porcentagem de lacunas devido a uma deleção dirigida de 10 resíduos, com o intuito de facilitar sua cristalização (CHANDRAMOULI *et al.*, 2010). As estruturas 2VBC e 2WHX correspondem à sequência completa da NS3, incluindo ambos os domínios proteásico e helicásico. Entretanto, o domínio NS2B possui apenas 18 resíduos nessas construções, somente o número suficiente para permitir a expressão da NS3 em sua forma solúvel (LUO *et al.*, 2008, 2010).

Para facilitar as análises de sequência e compreensão do texto, as estruturas que designavam cadeias diferentes aos domínios NS2B_{CF} e NS3_{PRO} foram editadas em uma única cadeia (figura 17, demarcações coloridas).

ligadas ao inibidor (WNV, figura 18, em amarelo) e as estruturas livres desse ligante (DENV, em vermelho, e MVEV em azul).

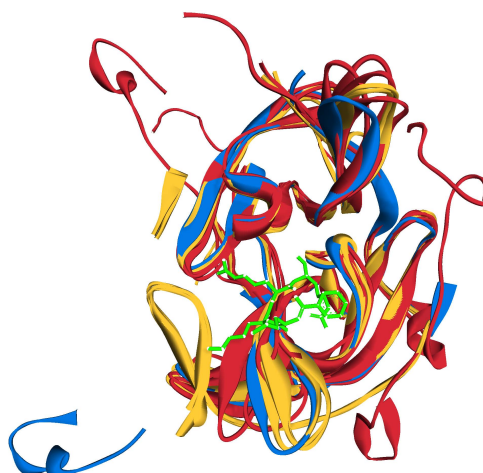


Figura 18. Sobreposição estrutural dos *templates* usados na modelagem por homologia. Estruturas em amarelo: WNV; vermelho: DENV; azul: MVEV. Ficam evidentes: o posicionamento do cofator (*loop* amarelo no canto esquerdo inferior), a mobilidade das estruturas (amarelas são menos flexíveis) e a presença de uma fita- β presente apenas nas estruturas de WNV (faixa amarela no canto esquerdo superior). Em verde, modelo de bolas e varetas do inibidor Bz-Nle-Lys-Arg-Arg-H (NDL).

Além da posição do cofator (fechada em WNV e aberta em DENV e MVEV), outras diferenças observadas incluem um pequeno segmento projetado em fita- β e a posição mais conservada de dois *loops*. Interessantemente, mesmo com as diferenças supracitadas, o RMSD entre os resíduos da NS3_{PRO} são relativamente baixos (tabela 5).

Tabela 5. Valores de RMSD entre as estruturas *template* e a estrutura de referência 3L6P.

PDB ID	RMSD (Å)
2FOM	1,3000
2VBC	0,8525
2WHX	0,9508
2FP7	0,8565
2IJO	0,7619
3E90	0,8158
2WV9	0,9700

5.1.2.3 Modelagem por homologia

Com um caráter a princípio meramente exploratório, e para avaliarmos se as diferenças observadas nas estruturas cristalográficas refletiriam em diferenças comportamentais durante a dinâmica molecular, partimos para a construção de dois modelos distintos: um baseado exclusivamente nas estruturas cristalográficas do DENV, e o outro nas do WNV (figura 19). Ao analisarmos o comportamento dos modelos preliminares após a DM, percebemos que cada um possuía características singulares que foram exploradas na construção de um terceiro modelo.

O modelo derivado das estruturas do DENV possui uma α -hélice C-terminal (na região de dobradiça entre os domínios proteásico e helicásico), enquanto que o modelo derivado do WNV possui uma fita- β projetada em uma das folhas- β da protease. Essas duas características foram incluídas no terceiro modelo, o qual também recebeu o cofator NS2B_{CF} e o inibidor NDL (Bz-Nle-Lys-Arg-Arg-H) da estrutura do WNV em conformação ativa e “fechada” (PDB ID 2FOM; ERBEL *et al.*, 2006). Acreditamos que a fita- β encontrada no modelo baseado em WNV depende da orientação do cofator, já que corresponde à região de transição entre NS3 e NS2B, e, por isso, não foi encontrada no modelo baseado no DENV. Por outro lado, a α -hélice presente apenas no modelo baseado no DENV corresponde à região de transição NS3_{PRO}/NS3_{HEL}, e, por isso, só estava presente no modelo derivado das estruturas da NS3 completa (LUO *et al.*, 2008, 2010). Apesar de não estar presente nos 185 resíduos que compõem o domínio NS3_{PRO}, essa α -hélice nos auxiliou a posicionar a extremidade C-terminal da protease. Além disso, postulamos que o posicionamento desse componente poderia ser importante para a conformação do domínio NS3_{PRO} durante a DM e, dessa forma, foi mantida no terceiro modelo. Como resultado, obtivemos um modelo mais compacto (vide figura 19.3), com as extremidades N- e C-terminais bem posicionadas. A etapa de finalização do modelo a ser submetido à DM constituiu-se da construção do

fragmento correspondente à dobradiça flexível Gly_[4]-Ser-Gly_[4] ligando os domínios NS2B_{CF} ao NS3_{PRO}.

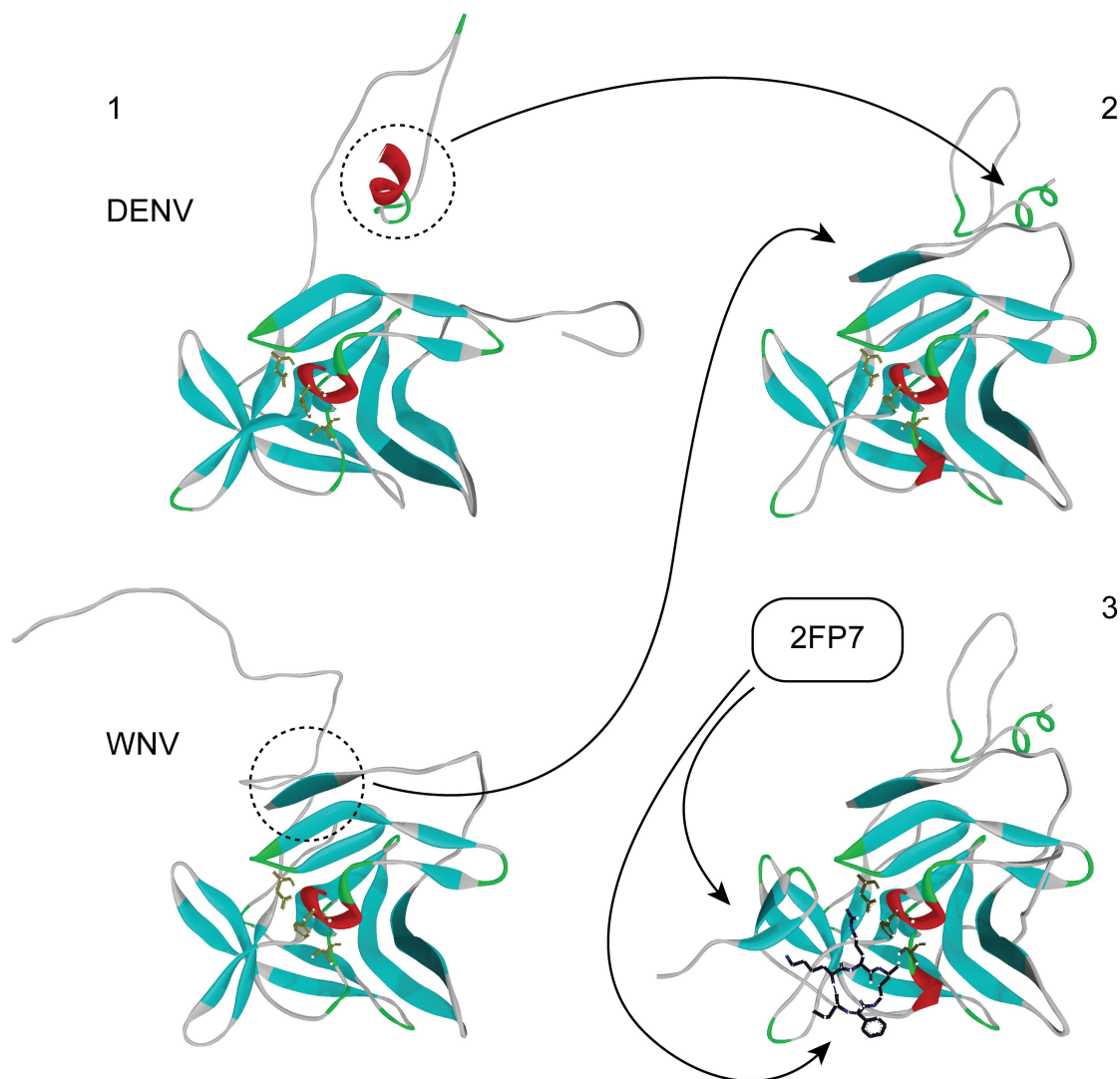


Figura 19. Modelos baseados em *templates* de DENV e WNV contribuíram para a construção de um terceiro modelo. O modelo baseado em estruturas de DENV (canto esquerdo superior) apresenta uma α -hélice não presente no modelo baseado no WNV (canto esquerdo inferior). Este, por sua vez, possui uma fita- β que não está presente no modelo baseado no DENV. Essas estruturas foram levadas em consideração na construção de um terceiro modelo (canto direito superior, que também recebeu o ligante e o cofator da estrutura homóloga 2FP7 (WNV) na conformação ativa.

Até onde sabemos, esse é o primeiro modelo completo da estrutura NS2B_{CF}-Gly_[4]-Ser-Gly_[4]-NS3_{PRO} utilizado em estudos de dinâmica molecular e *virtual screening*. Wichapong e colaboradores, em 2010 (WICHAPONG *et al.*, 2010), realizaram um trabalho de dinâmica molecular com modelos baseados em estruturas retratadas, além de incompletas (os resíduos 43-48 e 89-96 da NS2B, assim como os resíduos 1-18 e 170-185 da NS3 foram desprezados no estudo). Ademais, como bem notado por Knehans e colaboradores

(KNEHANS *et al.*, 2011), no modelo acima mencionado a cavidade do oxianion está fechada, e, portanto, estaria representando uma estrutura inativa.

Com o objetivo de identificar as contribuições individuais do cofator e do inibidor na estabilidade da estrutura durante as DMs, construímos outras três variações do último modelo: (i) apenas com o cofator, (ii) apenas com o ligante e (iii) sem cofator e sem ligante, totalizando quatro modelos submetidos à DM. Ao contrário do trabalho mencionado anteriormente (WICHAPONG *et al.*, 2010), no qual utilizaram DM para comparar os modelos obtidos a partir de diferentes *templates*, o nosso intuito foi de identificar as contribuições individuais de cada componente do complexo na indução da conformação ativa, partindo do pressuposto de que a protease do DENV se comporta da mesma forma que a homóloga do WNV. Os resultados obtidos das análises das simulações desses modelos serão descritos a seguir.

5.1.3 Dinâmicas moleculares

O objetivo de uma dinâmica molecular é simular o estado de equilíbrio de uma proteína em seu ambiente. Seu uso em conjunto com a modelagem por homologia já passou por diversas mudanças de paradigmas. Uma das formas mais simplificadas de sua utilização é a verificação da estabilidade de um modelo ao longo de uma trajetória. Se as estruturas secundárias se mantêm ao longo do tempo, isso significa que o modelo foi bem construído, e que o mesmo não possui erros muito evidentes. Caso contrário, o que se veria seria a perda dessas estruturas, e o conseqüente desenovelamento da proteína. Além disso, uma análise do desvio médio quadrático (RMSD, do inglês *Root Mean Squared Deviation*) entre as estruturas presentes em cada *frame* da trajetória também é um bom indicador de estabilidade e de equilíbrio, como veremos a seguir.

5.1.3.1 RMSDs

Durante uma dinâmica molecular com solvente explícito, a molécula a ser estudada é colocada em uma caixa d'água (ou outro solvente), onde ela pode se mover livremente. Entretanto, dois tipos de movimento não nos vêm ao caso: movimentos translacionais, onde a molécula se movimenta pela caixa d'água em relação à sua posição inicial, como um todo; e movimentos rotacionais, onde a molécula movimenta-se ao redor de um de seus eixos, também como um todo. Os movimentos de nosso interesse são os movimentos mais discretos entre os domínios e cadeias laterais. Para analisá-los separadamente, é necessário alinhar as estruturas de cada quadro da trajetória à estrutura inicial. Dessa forma, somente os movimentos intramoleculares são levados em consideração em análises posteriores.

Tendo isso em mente, a variação do desvio médio quadrático dos átomos pesados (desconsiderando átomos de hidrogênio) de cada modelo foi calculada após um alinhamento prévio dos quadros da trajetória (figura 20).

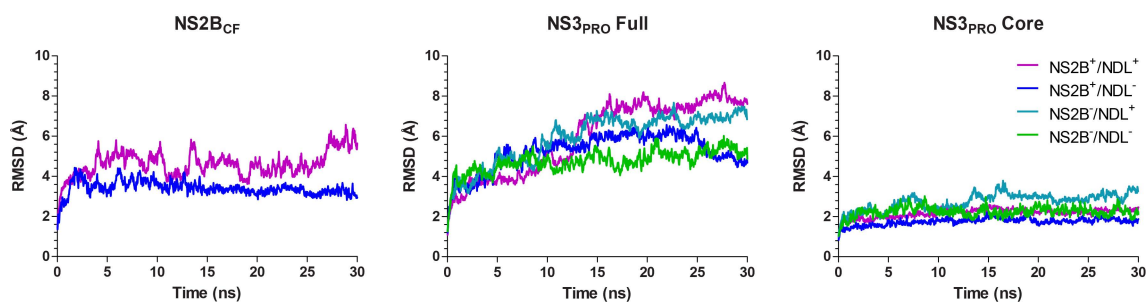


Figura 20. Variações de RMSD ao longo do tempo de simulação. O painel da esquerda corresponde às mudanças de RMSD durante o tempo para o cofator NS2B. É possível notar que na presença do ligante sua estabilidade é diminuída. O painel central corresponde às variações de RMSD da NS3_{PRO}, considerando todos os seus resíduos, incluindo das regiões N- e C-terminais. O painel da direita corresponde às variações dos resíduos centrais, incluindo aqueles do sítio ativo. NS2B⁺: presença do cofator; NS2B⁻: ausência do mesmo; NDL⁺: presença do inibidor; NDL⁻: ausência do mesmo. Novamente é possível observar que a presença do inibidor induz mudanças conformacionais importantes.

Em termos de equilíbrio, pela figura acima podemos notar que todos os modelos tendem a esse estado a partir dos 15 ns da simulação (figura 20, NS3_{PRO} Full). Ao compararmos os desvios do núcleo da enzima (resíduos 20 a 168, NS3_{PRO} Core) com os da proteína inteira, podemos verificar que os desvios foram mais baixos para todos os modelos (RMSD < 4 Å).

Isso é um indício de que o núcleo seria mais estável, e que a maior parte dos deslocamentos viriam dos resíduos 1-20 e 168-208, fato também observado nas estruturas cristalográficas (ASSENBERG *et al.*, 2009; CHANDRAMOULI *et al.*, 2010; ERBEL *et al.*, 2006; LUO *et al.*, 2008; NOBLE *et al.*, 2012; ROBIN *et al.*, 2009).

Ainda em relação ao núcleo da proteína, podemos observar que há diferença quanto à estabilidade nos modelos distintos: os maiores deslocamentos são apresentados por aquele que possui apenas o ligante (NS2B-/NDL+), com um RMSD médio de 2,75 Å. Curiosamente, o modelo com o núcleo mais estável (RMSD médio de 1,75 Å) foi aquele apenas com o cofator (NS2B+/NDL-), sugerindo que a presença do cofator talvez exerça um papel de estabilização do sítio ativo mesmo na ausência do ligante. Esses dados sugerem papéis antagônicos ao ligante e ao cofator: enquanto o primeiro seria responsável pelas mudanças conformacionais relacionadas ao ajuste induzido, o segundo teria o papel de conter grandes variações. Dois fatos que corroboram com essa hipótese são: (i) dadas as mesmas condições para cada simulação, o modelo apenas com o cofator foi o que se mostrou mais estável, o modelo apenas com o ligante foi o menos estável e aquele que possuía os dois componentes permaneceu em um meio-termo; (ii) um trabalho recente (NOBLE *et al.*, 2012) demonstrou que o sítio ativo da proteína possui uma conformação idêntica para pelo menos dois ligantes diferentes, não importando a orientação do cofator.

Voltando a atenção novamente para os desvios da proteína NS3 inteira, também podemos observar diferenças entre os modelos. Neste caso, podemos notar que o modelo NS2B-/NDL- foi o que teve a menor média de RMSD (4,75 Å), seguido dos modelos NS2B+/NDL- (5,24 Å), NS2B-/NDL+ (5,91 Å) e NS2B+/NDL+ (5,97 Å). Isso implica tanto o cofator NS2B em mudanças na periferia da NS3, como o ligante NDL nas maiores mudanças conformacionais. Também é importante ressaltar que, com exceção do modelo completo NS2B+/NDL+, todos os outros tiveram um salto nos valores de RMSD no início da dinâmica.

No caso do modelo NS2B+/NDL+ a transição foi muito mais suave, o que nos leva a crer que as mudanças conformacionais desse modelo não estejam relacionadas à desativação da proteína.

Na mesma figura, podemos ainda observar que a estabilidade do cofator é grandemente perturbada pela adição do ligante NDL (figura 20, painel NS2B_{CF}). Dessa forma, podemos sugerir que o cofator pode atuar para a estabilização da enzima “amortecendo” as interações do ligante com a NS3_{PRO}.

Juntos, esses resultados nos mostram que as maiores mudanças conformacionais procedem do domínio NS3_{PRO}, principalmente de suas extremidades N- e C-terminais, e que, enquanto o núcleo da NS3 pode ser estabilizado pelo cofator, o ligante pode ser o responsável pelas maiores mudanças conformacionais, tanto no domínio NS2B_{CF} quanto no NS3_{PRO}. Infelizmente, os dois artigos que tratam de dinâmicas moleculares com a protease NS3, seja de DENV ou de WNV, foram baseados em modelos com estruturas retratadas (WICHAPONG *et al.*, 2010; ZHOU *et al.*, 2006) e, portanto, não podem nos prover informações precisas para comparações. Desta forma, a proteína mais próxima para a qual foram realizados estudos semelhantes é a NS3_{PRO} do vírus da hepatite C (HCV; ZHU e BRIGGS, 2011). Apesar de tratarem de uma estrutura mais remotamente relacionada, os autores do referido trabalho encontraram resultados muito próximos aos nossos, como por exemplo: (i) em ambos os trabalhos as estruturas possuindo apenas o inibidor foram identificadas como as mais instáveis; e, (ii) o espaço conformacional amostrado pela NS3 é restringido pela ligação do cofator (NS4A, no caso do HCV, e NS2B no caso do DENV). Em contrapartida, mediante ligação do inibidor peptídico, a estrutura do complexo NS3/NS4A do HCV se torna ainda mais rígida, ao contrário do que pudemos notar no caso do DENV. Entretanto, por se tratarem de estruturas mais filogeneticamente distantes, comparações mais aprofundadas devem ser feitas com cautela. É sabido, por exemplo, que a protease do

HCV possui especificidade a substratos diferentes da NS3 do DENV, além de possuir atividade independente do cofator em alguns casos.

5.1.3.2 Estrutura secundária

Uma segunda medida de estabilidade pode ser obtida através da observação da conservação das estruturas secundárias ao longo da dinâmica. Como dito anteriormente, um erro na modelagem por homologia, ou ainda perturbações causadas pela retirada dos componentes dos sistemas, podem refletir na mudança dessas estruturas, e conseqüentemente, do estado de ativação da enzima. Portanto, seja para validar o modelo completo (NS2B+/NDL+) ou para verificar diferenças de estrutura secundária mediante ligação de cada componente do sistema, uma análise detalhada das estruturas secundárias de cada modelo foi realizada (figura 21).

De uma maneira geral, a primeira observação que podemos fazer é que a presença do ligante mais uma vez prova ser responsável pelas perturbações, mas que essas perturbações não são suficientes para gerar mudanças na estrutura secundária. Mais detalhadamente, entretanto, podemos notar que a presença do cofator é fundamental para a estabilização de uma estrutura secundária próxima ao resíduo 49 da NS3 que alterna entre α -hélice (rosa) e hélice-3(10) (azul) (figura 21, NS3_{PRO}, coluna da esquerda vs coluna da direita). O mesmo pôde ser encontrado no trabalho de Zhu e Briggs (ZHU e BRIGGS, 2011), onde os resíduos 55-60 da protease do HCV também adotaram essa conformação na presença do cofator. A formação dessa hélice é importante para a redução da flexibilidade dessa região, e conseqüente estabilização do resíduo catalítico HIS₅₁.

É importante notar também que a presença do ligante NDL aumenta a quantidade de pontes isoladas no cofator (amarelo escuro, figura 21, NS2B_{CF}), na região que compreende os

resíduos 72-78 e 84-86, provavelmente provenientes de interações diretas entre o ligante e o cofator (veja adiante na seção “Interações entre Pares”).

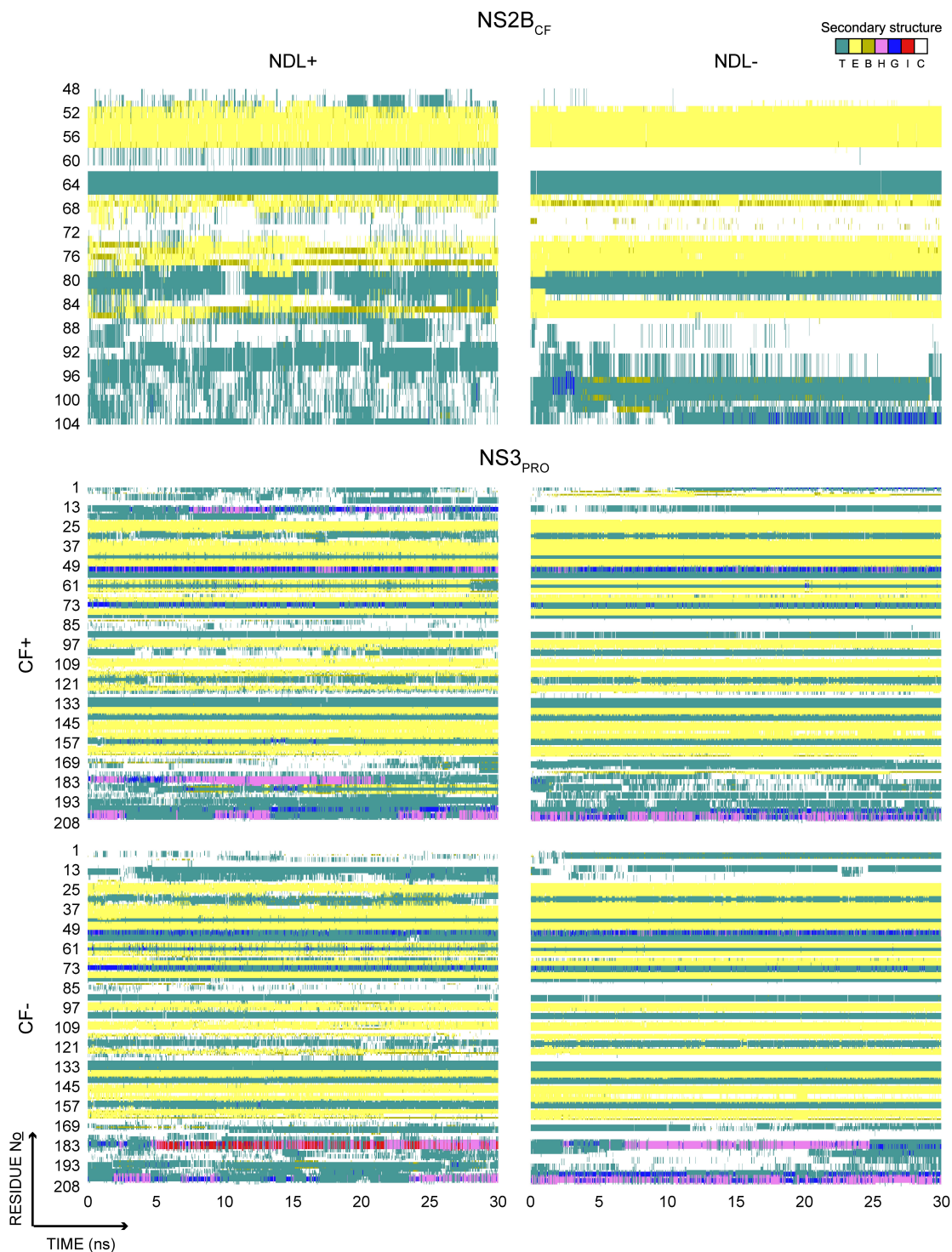


Figura 21. Análise da composição de estruturas secundárias ao longo da dinâmica. NDL+/-: presença ou ausência do inibidor; CF+/-: presença ou ausência do cofator. Legendas no canto superior direito: T = volta (*turn*); E = conformação estendida (componente de folha- β); B = ponte isolada (*isolated bridge*); H = α -hélice; G = hélice-3(10); I = π -hélice. A desorganização do cofator fica mais uma vez evidente na presença do inibidor, além de aumentarem o número de pontes isoladas. No domínio NS3_{PRO}, a maior diferença foi a estabilização de uma hélice próxima ao resíduo 49.

5.1.3.3 Agrupamento conformacional

Apesar de muito importantes para a análise de uma dinâmica molecular, sozinhos os valores de RMSD não são suficientes para nos dar informações a respeito do número de conformações de uma proteína. Para obter esses dados, devemos utilizar métodos de agrupamento conformacionais, dentre os quais podemos citar a matriz de RMSD bidimensional. Nessa matriz simétrica, o RMSD da estrutura de cada frame é comparado com os outros *frames*. Após a construção da matriz, um gráfico é plotado baseado em um limite de diferença máxima pré-estabelecido (*cutoff limit*). Desta forma, podemos agrupar as conformações em famílias que se relacionam dentro de uma faixa específica de variação de RMSD.

Na figura 22, plotamos gráficos dessas matrizes com um limite de 3,5 Å, ou seja, procuramos relacionar os *frames* como membros de uma família, com no máximo 3,5 angstroms de divergência entre eles. Como já constatado anteriormente, a presença do NDL no sistema aumenta os espaços conformacionais experimentados pela proteína, ou seja, a visualização de famílias dentro de um *cutoff* de 3,5 Å é dificultada. Mais uma vez, este fato é atenuado pela presença do cofator. Enquanto que no modelo NS2B-/NDL+ o número de famílias é de aproximadamente seis, no modelo NS2B+/NDL+ esse número é reduzido para quatro. No modelo NS2B+/NDL-, a relação entre as famílias fica mais próxima, com conformações se repetindo entre elas, gerando pelo menos duas grandes famílias. Curiosamente, no modelo NS2B-/NDL- essas relações são estreitadas ainda mais, gerando interseções entre as famílias, unindo-as em um único *cluster*. Analisando o cofator, fica claro o impacto que a presença do ligante NDL exerce no número de espaços conformacionais amostrados: quando o ligante é removido do sistema o cofator adota praticamente a mesma conformação durante toda a dinâmica.

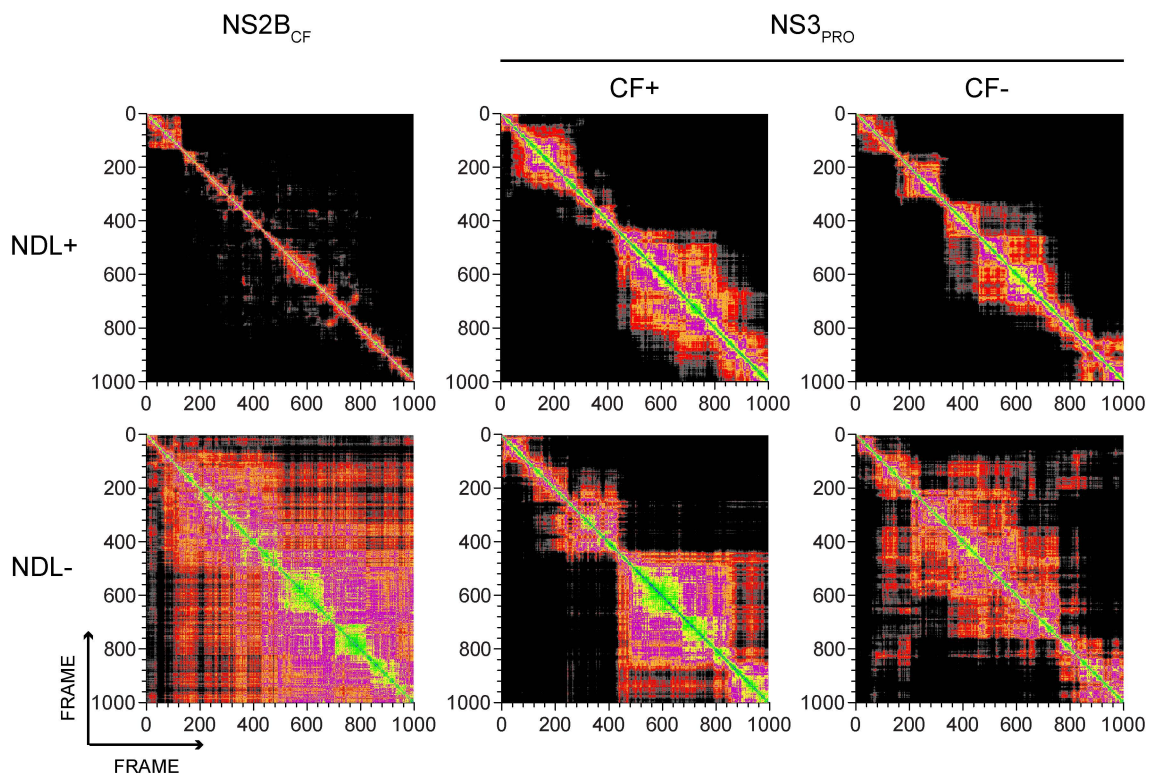


Figura 22. Agrupamento das famílias de conformações amostradas pelos modelos. NDL: inibidor; CF: cofator. O número de famílias conformacionais aumenta na presença do inibidor. Entretanto, o cofator é capaz de conter essas mudanças, reduzindo o espaço amostrado pela NS3_{PRO}. Além disso, na ausência do ligante o cofator apresenta uma única grande conformação dentro dos 3,5 Å estabelecidos como referência.

5.1.3.4 Análise de Componentes Principais (PCA)

Uma vez que as mudanças nas estruturas secundárias durante a dinâmica não poderiam ser responsáveis pelos aumentos nos valores do RMSD, resolvemos então investigar a presença de movimentos entre os domínios do sistema por meio da análise de seus componentes principais. Essa técnica é capaz de delimitar os movimentos mais importantes e agrupá-los em conjuntos (*eigenvectors*, ou autovetores). Como podemos observar na figura 23, os movimentos de maior magnitude partem das extremidades N- e C-terminais da protease. Além disso, esses resultados confirmam mais uma vez que o ligante é o responsável por esses movimentos. O papel do cofator na contenção desses movimentos também fica mais evidente, já que na sua ausência os valores arbitrários de mobilidade extrapolam a escala utilizada no gráfico.

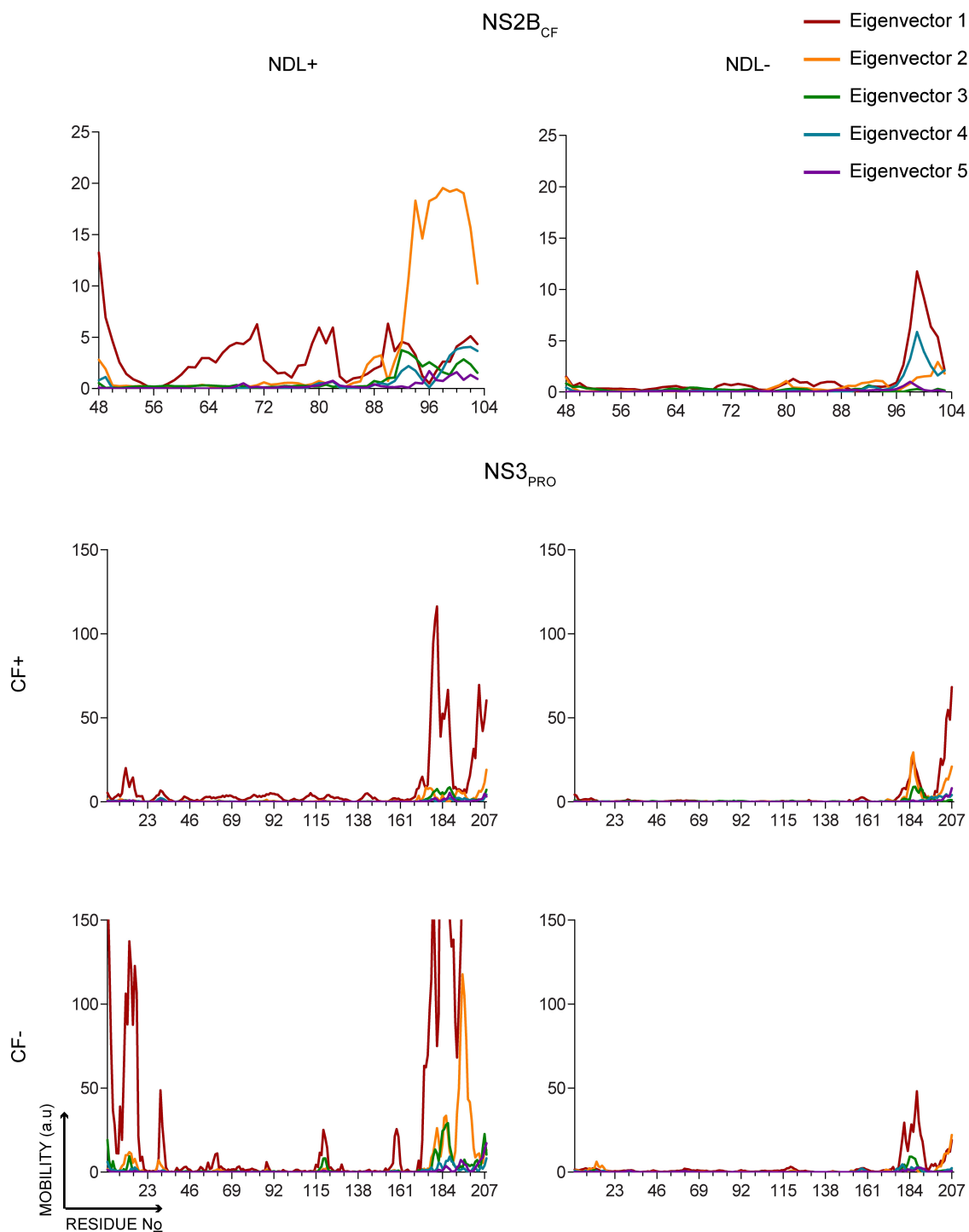


Figura 23. Mobilidade dos resíduos de cada modelo para os 5 primeiros *eigenvectors* identificados por PCA. NDL: inibidor; CF: cofator. A capacidade do cofator em estabilizar os resíduos da região central da proteína, incluindo a região da hélice a partir do resíduo 49, é evidenciada.

Outro importante fato, observável pelos picos no gráfico que representa o modelo NS2B-/NDL+, é que o cofator também estabiliza resíduos do núcleo da proteína, em regiões possivelmente importantes para a interação com o ligante (53-69 e 153-161, por exemplo; ver seção “Interações entre Pares”). Quando analisamos os componentes principais do

cofator nos dois modelos (NS2B_{CF+}/NDL₊ e NS2B_{CF+}/NDL₋), nos deparamos com um fato interessante: uma vez na conformação “fechada”, configuração na qual os dois modelos foram produzidos, a retirada do ligante por si só não consegue ser responsável pela reversão em uma conformação “aberta”.

O papel fisiológico do ligante, no caso o substrato, pode ter mais importância durante a produção e enovelamento da poliproteína no retículo endoplasmático (RE), onde induzirá a formação da conformação “fechada” do cofator. Após as primeiras reações *in cis* (intramolecular), a proteína ancorada à membrana do RE pode se desligar do substrato e “navegar” para realizar as reações *in trans* (intermoleculares) (BERA *et al.*, 2007).

A não-passagem do estado fechado para o aberto e, novamente do aberto para o fechado antes da próxima reação, seria um fato extremamente desejável em termos de eficiência cinética na replicação viral. Para comprovar essa hipótese, um trabalho semelhante ao de SU e colaboradores (2009), onde a preferência conformacional do cofator foi avaliada mediante a ausência ou presença do inibidor, poderia ser realizado, mas usando inibidores reversíveis ao invés de um inibidor covalente. Dessa forma, as conformações do cofator poderiam ser acessadas antes da adição do inibidor reversível (conformação aberta), após a adição do inibidor reversível (fechada), e após a retirada do inibidor (teste da hipótese).

Na figura 23, ainda podemos observar que, como esperado, a maioria dos movimentos importantes está no primeiro *eigenvector*, apesar de vários desses movimentos também estarem presentes em outros vetores. Para verificarmos as direções desses movimentos, os valores do primeiro *eigenvector* de cada modelo foram projetados em suas estruturas (figura 24). Nessa representação, podemos notar o papel do cofator na estabilização de três estruturas principais: dois *loops* no canto inferior esquerdo (modelo NS2B₋/NDL₊), e a hélice discutida anteriormente na figura 21.

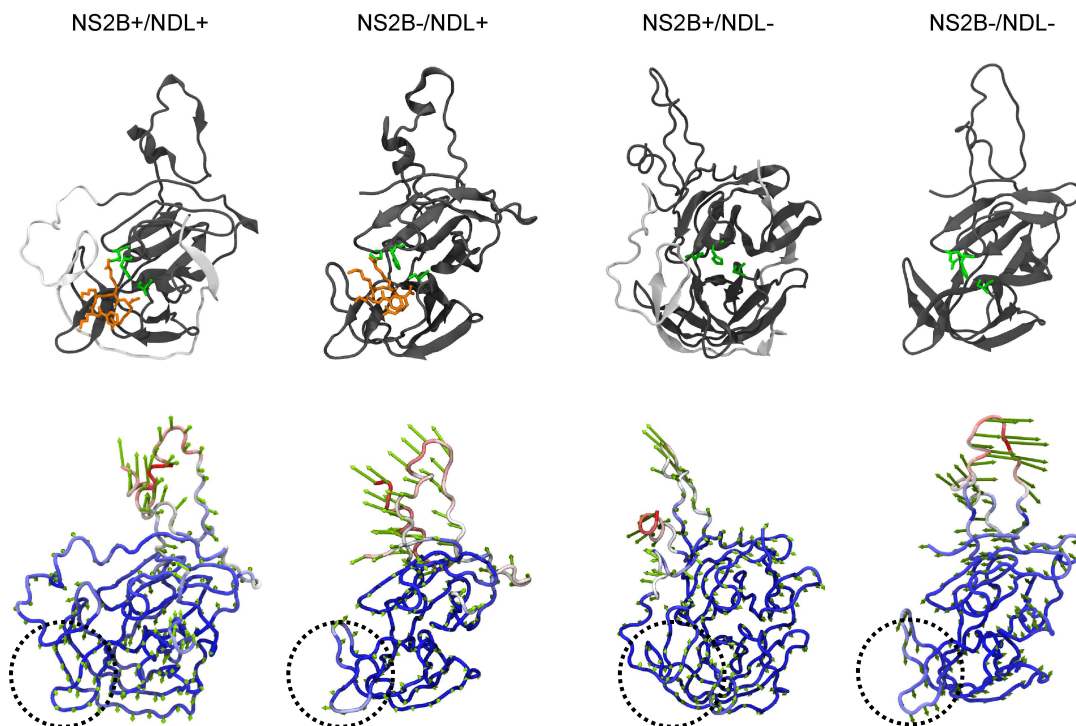


Figura 24. Modelos de cada sistema da protease do vírus da dengue, assim como suas projeções vetoriais de movimento contidas no primeiro *eigenvector*. Em cinza, domínio NS3_{PRO}; em branco, domínio NS2B_{CF}. O ligante NDL está representado em laranja, e os resíduos catalíticos H51, D75 e S135 em verde. Na segunda linha, as cores representam a mobilidade da proteína (vermelho = altamente móvel; azul = conservado). Os vetores indicam a amplitude e direção dos movimentos. Além dos movimentos da região C-terminal, no topo de cada modelo, foram observadas diferenças na mobilidade do *loop* indicado pelo círculo, que ocorrem principalmente na ausência do cofator.

5.1.3.5 Interações entre Pares

De modo a compreendermos melhor o papel dos componentes do sistema nas interações entre eles, primeiramente realizamos uma análise das interações par-a-par entre o ligante e a água ou entre ligante e a proteína ao longo do tempo da dinâmica (figura 25). Dessa forma, fomos capazes de avaliar a afinidade entre o ligante e a proteína, na presença ou ausência do cofator. Com isso, foi possível observar dois fatos importantes: o primeiro é que na presença do cofator o ligante permanece mais tempo sem perder um pouco de sua afinidade caracterizada pelo salto na redução da força após os 5 ns iniciais. Na ausência da NS2B, esse salto ocorre bem antes dos 5 ns. O segundo fato é que a afinidade parece ser recuperada no primeiro caso, e a redução de afinidade seria então reflexo da hidratação do sítio de ligação do inibidor (na medida em que a água entra nesse sítio, ela compete com o ligante pelas interações com a proteína). Já na ausência do cofator, essa afinidade diminui

cada vez mais ao longo do tempo, o que significa que o ligante poderia desligar-se do complexo durante uma simulação mais longa.

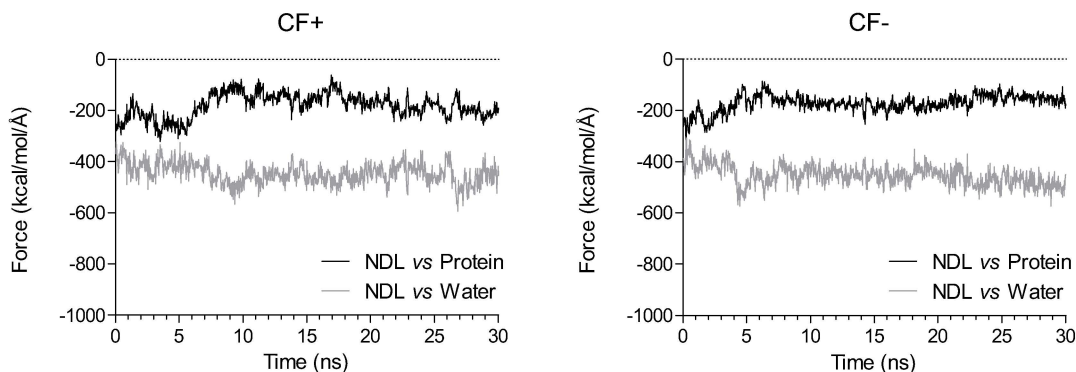


Figura 25. Forças de interação entre ligante e proteína ou ligante e água. CF: cofator. Em preto, força de interação entre a proteína e o ligante; em cinza, entre proteína e água. A hidratação da NS3_{PRO} pode ser um indicativo da diminuição da força de interação entre proteína e ligante. Na ausência do cofator, as duas curvas se afastam, sugerindo essa diminuição. Entretanto, na presença do cofator essa tendência é revertida a partir dos 25 ns de simulação.

Para mapearmos as interações responsáveis por essas afinidades, calculamos as médias das forças entre os pares: cada resíduo do cofator e o ligante, cada resíduo do cofator e a NS3_{PRO} como um todo, cada resíduo da NS3_{PRO} e o cofator como um todo, e, cada resíduo da NS3_{PRO} e ligante (figura 26).

As médias do par “resíduos da NS2B_{CF} vs NDL” nos mostra que os resíduos mais importantes na interação entre eles são os resíduos 83-85 do cofator, fato também observado nas estruturas cristalográficas do WNV (ALESHIN *et al.*, 2007; ERBEL *et al.*, 2006; ROBIN *et al.*, 2009). Por sua vez, o gráfico de interações do par “resíduos da NS2B_{CF} vs NS3_{PRO}” nos mostra que as interações entre os resíduos 50, 52 66, 93, 94 e 95 são diminuídas na presença do ligante, enquanto que as interações entre os resíduos 78 e 80 são reforçadas. Curiosamente, esses dois resíduos fazem parte da região responsável pela alteração do sítio ativo, alinhando os resíduos 151, 153 e 161 da NS3_{PRO} importantes para o reconhecimento do substrato no caso do WNV (ERBEL *et al.*, 2006).

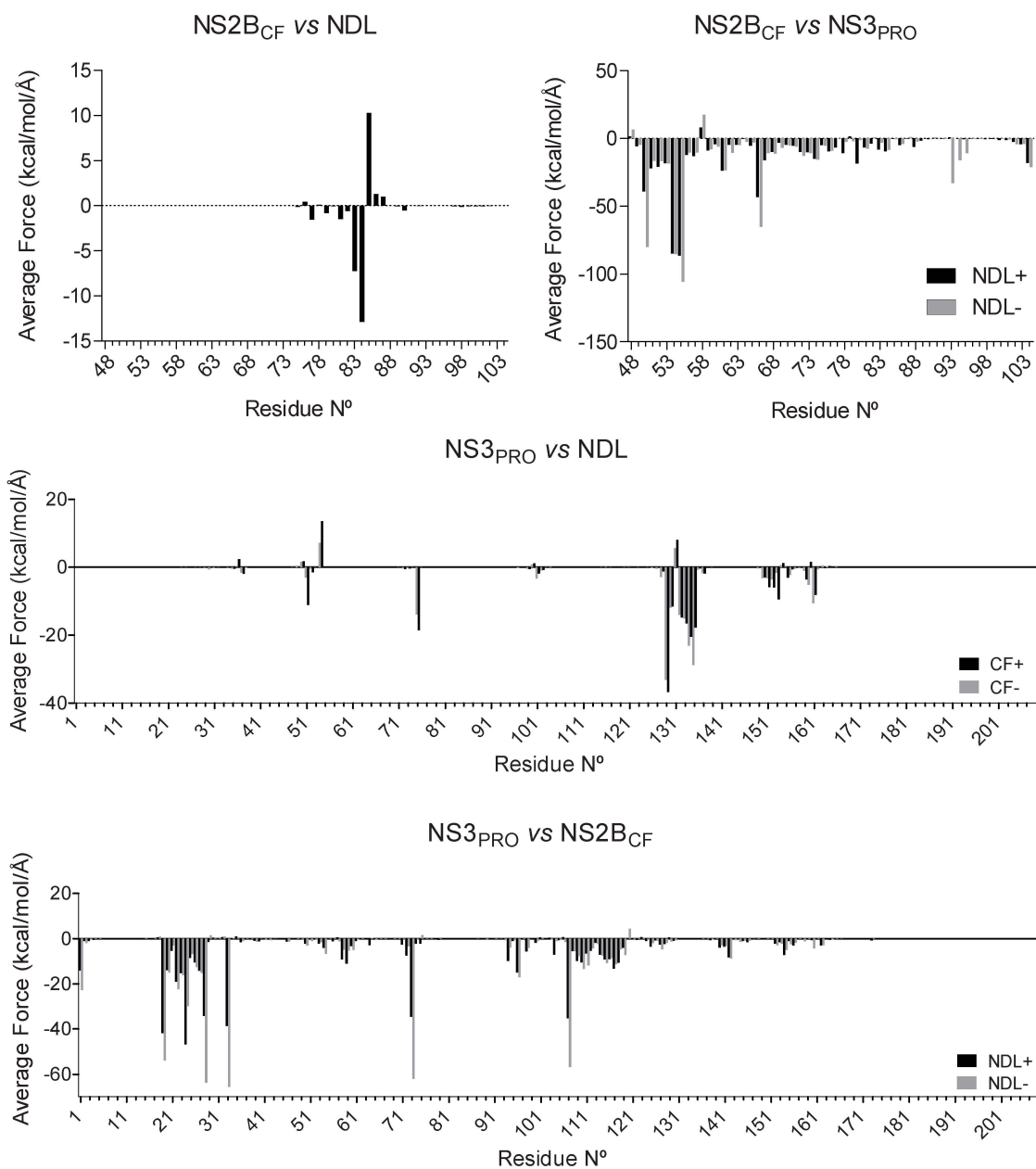


Figura 26. Média das forças de interações entre resíduos do cofator ou da NS3_{PRO} e cada componente do sistema. NDL = inibidor; CF = cofator. Esse método é capaz de mensurar a intensidade da força de interação entre os pares. Os resíduos 73-90 do cofator interagem diretamente com o inibidor. De uma maneira geral, a maior parte das interações entre NS2B_{CF} e NS3_{PRO} diminui na presença do inibidor, enquanto que as interações entre inibidor e NS3_{PRO} aumentam.

Quanto ao par de interações “resíduos da NS3_{PRO} vs NDL”, podemos notar que, de uma maneira geral, para a maior parte dos resíduos envolvidos, houve um aumento da força de atração ao NDL na presença do cofator, provavelmente relacionado à estabilização do sítio ativo mediado por sua ligação à NS3_{PRO}. Já na análise do par “resíduos da NS3_{PRO} vs NS2B_{CF}”, podemos notar uma diminuição da força de interação entre os resíduos 19, 28, 33, 73 e 108 da NS3_{PRO} e o cofator como um todo, de uma maneira ainda mais intensa do que nos outros

pares analisados (notar a diferença nas escalas do eixo y). Apesar de nenhum desses resíduos estar diretamente em contato com o inibidor, as mudanças conformacionais adotadas no sítio ativo podem refletir no restante da protease. Dois desses resíduos, o ASP₁₉ e a LYS₇₃, estão relativamente próximos a uma segunda cavidade dessa enzima, encontrada apenas nas estruturas na conformação ativa e que é conservada em todas as estruturas do WNV, sugerindo que ela seja funcionalmente importante (NOBLE *et al.*, 2012). Essa mesma cavidade já foi indicada como um potencial alvo de drogas (CHAPPELL *et al.*, 2008; NOBLE *et al.*, 2012). Por outro lado, as interações entre o resíduo ARG₂₄ são aumentadas na presença do ligante, em uma região em evidência para a ancoragem de uma fita-β N-terminal do cofator ao complexo.

Além de atribuir papéis a cada um dos três componentes desse sistema, os dados acima apresentados apontam para a estabilidade e autenticidade do nosso modelo completo (NS2B_{CF}-Gly-NS3_{PRO}). O manuscrito a seguir trata da exploração das famílias de conformações desse modelo com a finalidade de utilizar representantes das mesmas em uma campanha de triagem virtual de alto rendimento (VHTs) na busca de novos inibidores com potencial terapêutico.

5.1.4 Manuscrito

Towards the Identification of New Inhibitors of the *Dengue virus* NS3 Protease: New Binding Site Conformations Accessed by Molecular Dynamics Simulation

Hugo de Almeida¹, Izabela M. D. Bastos¹, Bergmann M. Ribeiro², Bernard Maigret³ and Jaime M. Santana^{1*}

¹ Laboratório de Interação Patógeno-Hospedeiro, Departamento de Biologia Celular, Universidade de Brasília, BR-70910-900, Brasília, DF, Brazil.

² Laboratório de Microscopia Eletrônica e Virologia, Departamento de Biologia Celular, Universidade de Brasília, BR-70910-900, Brasília, DF, Brazil.

³ LORIA/UMR 7503, Équipe-projet Orpailleur, Nancy Université, Campus Scientifique, Vandœuvre-Lès-Nancy CEDEX, France.

* Corresponding author: jsantana@unb.br

5.1.4.1 Abstract

Dengue fever is caused by four distinct serotypes of the *Dengue virus* (DENV1-4), and is estimated to affect over 500 million people every year. Presently, there are no vaccines or antiviral treatments for this disease. Among the possible targets to fight this disease is the viral NS3 protease (NS3_{PRO}), which is in part responsible for viral processing and replication. It is now widely recognized that virtual screening campaigns should consider the flexibility of target protein by using multiple active conformational states. The flexibility of the DENV NS3_{PRO} could explain the relatively low success of previous virtual screening studies. In this first work, we explore the DENV NS3_{PRO} conformational states obtained from molecular dynamic (MD) simulations to take into account protease flexibility during the virtual screening/docking process. To do so, a full model of the NS3_{PRO}, including the NS2B cofactor (essential to the NS3_{PRO} activation), a glycine flexible link and the proteolytic domain, was built by multiple template homology modeling. The MD simulations had the purpose to sample, as closely as possible, the protein conformational landscape prior to inhibitor binding. The obtained conformational MD sample was clustered into four families that, together with principal component analysis of the trajectory, demonstrated the presence of protein flexibility. Those results allowed the description of multiple binding modes for the Bz-Nle-Lys-Arg-Arg-H inhibitor, as verified by binding plots and pair interaction analysis. This study allowed us to tackle the protein flexibility in our virtual screening campaign against the *Dengue virus* NS3 protease.

5.1.4.2 Introduction

Dengue fever (DF) is an infectious disease caused by four distinct serotypes of the *Dengue virus* (DENV1-4) transmitted by *Aedes* spp.

Milder manifestations of the disease may include fever, rash, headaches, joint and muscle pain, fatigue and vomiting. Re-infection by different serotypes, however, may cause much more significant clinical conditions, like Dengue

Hemorrhagic Fever (DHF) and Dengue Shock Syndrome (DSS) (1,2) which can cause death.

DF is estimated to affect over 500 million people every year (3) and has also recently been ranked as the most common cause of febrile illness in travelers, surpassing malaria and gastrointestinal infections (4). Together with the ongoing expansion of mosquito habitats (5) due to the recent climate changes and the urbanization of developing countries (6), this fact has drawn attention of sanitary and health centers around the globe. Two autochthonous cases in Europe (7), and recent outbreaks in southern USA (8) have shown that dengue is no longer exclusively a problem for tropical developing countries. Despite its high incidence, severity and economic burden, there are currently no antiviral treatments or vaccines for DF. The development of an efficient anti-DF

vaccine has the challenge to provide protection from all four serotypes at once (9); otherwise it may render immunized individuals more susceptible to DHF (10). Regarding the design of antiviral therapeutics, viral proteases are often proposed as potential targets for treatment due to their essential task of processing viral polyproteins into their functional unities (11). Concerning the *Dengue virus* (DENV) and its close relative *West Nile virus* (WNV), this role is assigned to the multi-domain nonstructural protein 3 (NS3). The NS3 is composed by a protease (NS3_{PRO}) and a helicase (NS3_{HEL}) domain, with the former being responsible for processing the polyprotein in specific sites (Figure 1). The NS3_{PRO} domain (EC 3.4.21.91) belongs to the family S7 of serine proteases, and needs a cofactor, the hydrophilic loop from NS2B (NS2B_{CF}) in the case of DENV and WNV, to become fully active (12,13).

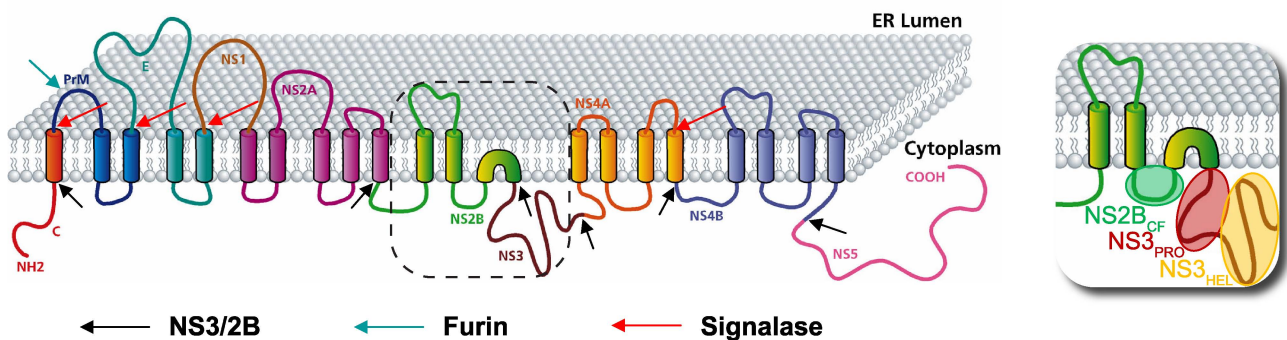


Figure 1. Predicted membrane topology of the Dengue virus polyprotein and its cleavage sites. The polyprotein is composed of three structural subunits: capsid (C), precursor of membrane protein (PrM) and envelope (E), as well as seven nonstructural (NS) subunits: NS1, NS2A, NS2B, NS3, NS4A, NS4B and NS5. The NS3 is a multifunctional protein composed of a helicase domain (NS3_{HEL}, marked in yellow in the detail box) and a protease (NS3_{PRO}, marked in red) domain, which in turn needs the hydrophilic loop of NS2B (NS2B_{CF}, marked in green) as a cofactor to be fully active. Black arrows indicate sites processed by the viral NS2B/NS3 protease. Adapted from Umareddy *et al.*, 2007 (14).

This protease has already been recognized as a valuable target for the design of new antiviral inhibitors against *Dengue virus* (15,16), and,

therefore, designing potent inhibitors against DENV NS3_{PRO} is an active research line in the fight against DF (17–22). The use of

computational drug-design approaches would be useful here to improve the discovery of putative hits and to help obtaining new leads (23–31). However, previous virtual screening campaigns fall short in the identification of new inhibitors, since none was able to find small organic compounds in the submicromolar inhibitory range. As it is now widely accepted that protein flexibility is an important factor to be taken into account to ensure the success of virtual screening campaigns (32), the flexibility of DENV NS3 protease, as highlighted by the differences observed between several x-ray structures (33–36), could explain the relatively low success of previous virtual screening attempts against this target. In fact, flexibility of the DENV NS2B/NS3 protease has already been proposed, but not proved, to explain the poor results of current drug design campaigns (21).

To address this issue, we performed molecular dynamics (MD) simulations towards an ensemble docking campaign. The MD simulations of DENV NS2B/NS3_{PRO} described in the present paper were intended to sample as closely as possible the protein conformational landscape prior to inhibitor binding. The obtained conformational MD sample was therefore clustered into families diverging at least 3.5 Å apart. Thus, a limited protein conformational set, proven to contain structures sufficiently different at their binding sites, was chosen to be employed in a subsequent ensemble docking simulation using focused chemical libraries.

5.1.4.3 Materials and Methods

5.1.4.3.1 Homology Modeling

To construct our 3D model three main factors were considered: (i) it had to be based on the full-length sequence of the expression construct protein that we have, (ii) it had to represent the catalytically active “closed” conformation, and (iii) it needed to be able to take into account the possible influence of N- and C-terminus domains in the NS3 protease active site conformational changes and/or accessibility. This led us to build a 3D model that included the hydrophilic NS2B cofactor region, a glycine flexible link and the NS3 proteolytic domain, as well as the NDL inhibitor.

To obtain this DENV NS2B_{CF}-Gly-NS3_{PRO} model with as much structural information as possible, an alignment among all *Dengue virus* and *West Nile virus* protease 3D structures available in the protein database (PDB) at the time (table 1) was performed.

Table 1. Crystallographic structures used as templates in the homology modeling process and their construct data information.

Virus	PDB id	Structural information	Ref.
DENV2	2FOM	NS2B _{CF} -Gly-NS3 _{PRO}	(33)
DENV4	2VBC	NS2B ₁₈ -Gly-NS3 _{FULL} ^(a)	(34)
DENV4	2WHX	NS2B ₁₈ -Gly-NS3 _{FULL} ^(a)	(37)
DENV1	3L6P	NS2B _{CF} -Gly-NS3 _{PROΔ10} ^(b)	(35)
WNV	2FP7	NS2B _{CF} -Gly-NS3 _{PRO} + NDL	(33)
WNV	2IJO	NS2B _{CF} -Gly-NS3 _{PRO} + BPTI	(38)
WNV	3E90	NS2B _{CF} -Gly-NS3 _{PRO} + NKK	(39)

^(a) Instead of using the full hydrophilic loop from NS2B, a short sequence of 18 residues was used, only to provide solubility to the expressed complex. ^(b) The construct has a 10 residues deletion in the N-terminus of NS3_{PRO}, in order to facilitate crystallization. NDL: Bz-Nle-Lys-Arg-Arg-H; BPTI: bovine pancreatic trypsin inhibitor; NKK: Naph-Lys-Lys-Arg-H.

This allowed us to check the completeness of the PDB structures compared to the

sequences deposited by the same authors. From this analysis we were able to gather information to build three different models of our system: Two preliminary models based either solely on the WNV or solely on the DENV PDB structures and a third one mixing data from both DENV and WNV.

For each model, the PDB templates were prepared by removing unused chains, ligands, water and ions. The cleaned structures were aligned using the Discovery Studio Visualizer 3.1 (Accelrys Inc) based on the sequence alignment of their conserved residues. Prior to submission to the homology software MODELLER (40), the sequence alignment obtained previously was edited to ensure certain features would be found in the resulting model, like the N-terminus region (including NS2B_{CF}) from WNV structures and C-terminus from DENV structures in the final model, e.g. For each set of templates (WNV only, DENV only or WNV/DENV) a total of 5 homology models were generated, each with 3 loop refinements. This generated a total of 45 models. The three lowest energy models (based on each set of templates) were chosen and verified by PROCHECK (41) to ensure their structural quality. We next compared these models and checked their respective qualities so that the best one was retained for further processing. Finally, the atomic coordinates of the NS2B cofactor and those of the NDL inhibitor (Bz-Nle-Lys-Arg-Arg-H) were transferred from the 2FP7 PDB template into the final NS3_{PRO} model obtained

above. After arranging the cofactor in place, we built the glycine linker with the homology module of the INSIGHT II package (Accelrys Inc). The inhibitor was deliberately left covalently unbound to simulate conformational changes prior the formation of the protease/substrate intermediate. Using the AutoPSF Generation Plugin provided with VMD (42), missing hydrogens were added according to their predicted protonation state at pH 7.0 and histidine residues were assigned to their δ -protonated state.

5.1.4.3.2 Molecular Dynamics Simulations

The final model of the complex between the NDL ligand and the NS2B_{CF}-Gly-NS3_{PRO} target obtained above was submitted to MD simulations. The complex was firstly solvated with TIP3P explicit water molecules in a box with at least 15 Å apart from any point in the protein, ending up with a box of 80 x 100 x 85 Å. Eight Na⁺ ions were added to ensure electrostatic neutrality. Missing parameters for the ligand were added using the CHARMM22 forcefield (43). The NAMD software (44) was used to simulate the dynamic behavior of the complex using the parm.prm parameter file from CHARMM22. Long-range electrostatics was evaluated using the particle-mesh Ewald approach (45). Simulations were carried out in the NPT ensemble using Langevin dynamics and piston to fix temperatures (300 K) and pressure (1 atm). Hydrogen-heavy atom bonds were constrained to their equilibrium values with the SHAKE algorithm (46).

The system was first energy minimized (6400 conjugate gradient steps) and then equilibrated (500 ps) before recording the trajectories. All MD trajectory frames were recorded at 1 ps intervals, for a total of 30 ns simulation.

5.1.4.3.3 Clustering

Once the MD simulations were complete, all frames were aligned by taking into account only heavy atoms in the core region of the NS3 protease alone (residue range from 20 to 168). This was done in order to focus the protein conformational analysis on the binding site region only. The RMSDs were calculated by using the appropriate VMD plugin and the corresponding data were exported as an ASCII matrix by the VMD Multiplot module. For clustering, a previously developed in-house Tcl script was used with a cutoff of 3.5 Å. This script builds a 2D RMSD matrix for identifying conformational families by identifying which frames are similar according to a given RMSD.

5.1.4.3.4 Pocket detection and volume analysis

Pockets were detected using METAPOCKET (47), an algorithm using several reliable pocket detection tools to combine their results and improve sensibility. Changes in pocket volume and surface (pocket descriptors) were monitored with MDPocket (48), a software capable of analyzing topological changes in cavities during MD simulations.

5.1.4.3.5 Analysis software

All molecular representations as well as several other analyses (such as distances between catalytic residues through the trajectory and secondary structure maintenance over time, for example) were performed by VMD. All graphics (RMSD, pair interactions, mobility plots, pocket volume evolution, etc) were plotted using GraphPad Prism 5 (GraphPad Software Inc). Two-dimensional binding plots from docking results were generated with Discovery Studio Visualizer 3.1 (Accelrys Inc).

5.1.4.4 Results

5.1.4.4.1 Building the 3D model

5.1.4.4.1.1 PDB Templates Sequence alignment

To build a homology model based on the available homologous PDB structures, we first performed a manual sequence alignment analysis of these templates (Figure 2). This alignment shows highly conserved residues between the DENV and WNV NS3 proteins. One block of 5 residues, numbered 59 to 63, seems to vary according to the DENV serotypes. A lack of structural information at C- and N-terminal regions was found, probably due to their high flexibility, and hence little information from those residues could be utilized for molecular modeling. NS3_{PRO} N-terminus data were more likely to be found in West Nile virus structures (Figure 2, blue box), while those of C-terminus were better provided by full (with both protease and helicase domains) NS3 *Dengue virus* structures (Figure 2, red box).

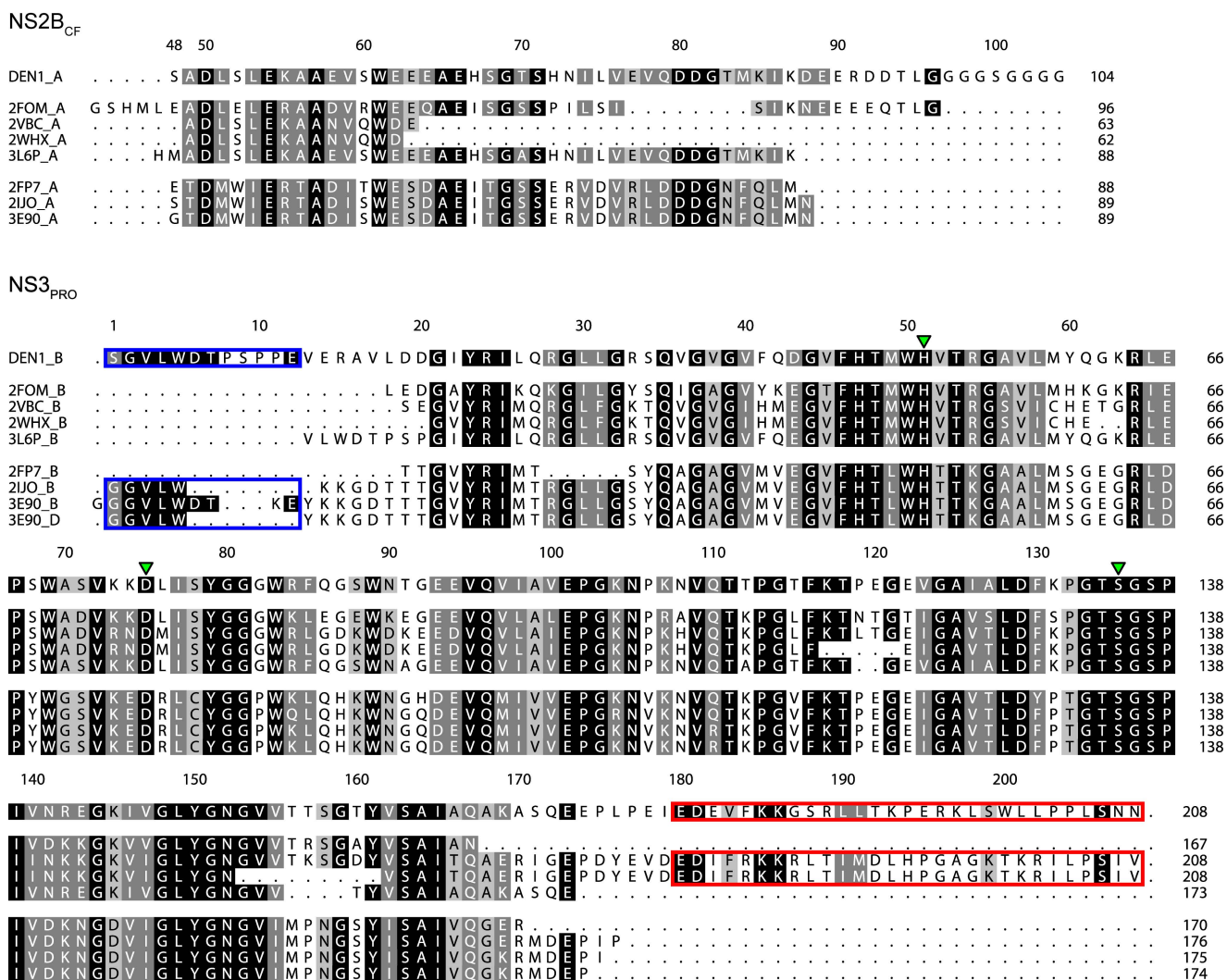


Figure 2. Dengue and West Nile virus NS2B_{CF}-Gly-NS3_{PRO} multiple sequence alignment extracted from the protein databank files. This alignment allowed us to verify the completeness of each deposited structure, as well as to fill structural gaps, direct N- (blue box) and C-terminus (red box) regions and the cofactor itself. Green triangles indicate the catalytic triad (HIS₅₁, ASP₇₅ and SER₁₃₅); sequences are numbered accordingly to the full constructs and shaded based on sequence similarities (black for identical, dark gray for strongly similar and soft gray for weakly similar residues).

5.1.4.4.1.2 Templates structural alignment

Regarding the PDB structures, while much has been said about different conformations related to distinct serotypes (35) and about the influence of ligands and co-factors (33), the protein core structure is well conserved even when compared between DENV and WNV (Figure 3A). The RMSD values for each template main chain atoms compared to the reference protein 2FOM are given in table 2, showing that active site residues position is

similar among them, irrespectively of whether the cofactor was in the open or closed conformation.

Table 2. Homology modeling templates and their catalytic triad main-chain atom RMSD to the reference protein 2FOM, based on the 126 conserved residues from the NS3_{PRO} domain.

PDB id	RMSD (Å)			
	Protein	HIS ₅₁	ASP ₇₅	SER ₁₃₅
2VBC	1.3984	0.448	0.796	0.815
2WHX	1.4121	0.503	0.851	0.764
3L6P	1.3000	0.263	0.323	0.468
2FP7	1.3620	0.412	0.833	0.717
2IJO	1.3330	0.504	0.996	0.784
3E90	1.2913	0.543	0.846	0.696

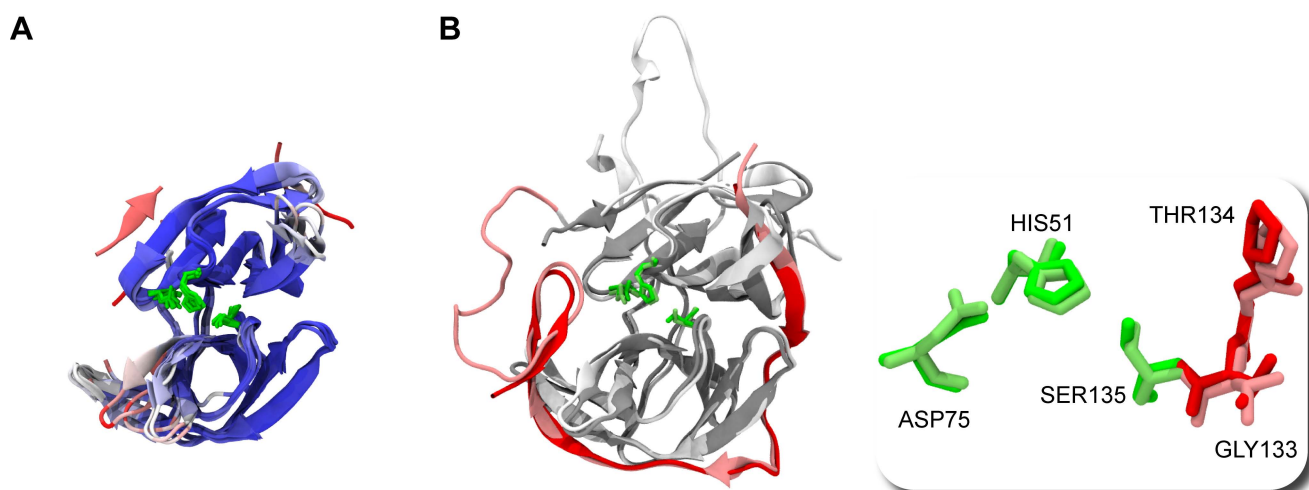


Figure 3. Structural alignments between (A) core region from all templates used in homology modeling, and (B) obtained model and recently available ligand bound DENV NS3_{PRO} structure (PDB id: 3U11) (36). In (A), residues were colored based on their Qres factor, obtained after a STAMP structure alignment performed with the VMD Multiseq plugin (49). Color ranges from dark blue (highly conserved positions) to red (not conserved at all). In (B), crystallographic structure was colored with gray (NS3_{PRO}) and red (NS2B_{CF}); homology model was colored in lighter shades – white (NS3_{PRO}) and pink (NS2B_{CF}). In addition to the highly conserved active site residues position, the oxyanion hole was also preserved in this model (boxed detail).

The situation is different for the C-terminal turn (Figure 3A, residues 151 to 167, marked in shades of red, bottom left corner) that adopted several spatial conformations among DENV structures, but not among WNV, indicating that this conformation may be stabilized by ligand binding.

5.1.4.4.1.3 Homology Modeling

Our strategy was initially to build a robust model of the NS3_{PRO} domain and next to add the NS2B_{CF}, the ligand and the linker pieces to achieve the active conformation of the NS2B_{CF}-Gly-NS3_{PRO} model, which was later used in the MD simulations. For that purpose, we first built two preliminary NS3_{PRO} models based exclusively either on the DENV or on the WNV structures. Besides belonging to different species, the main difference observed was the absence (DENV) or the presence (WNV) of ligands. Among the five models generated by

MODELLER from each template set, those with the lowest energies were analyzed. Overall, the models did not differ much at the protein core, whereas, given the divergence on the amount of information between templates, N- and C-terminus were quite distinct. For the model based on the DENV structures, the C-terminal end was arranged more concisely in an alpha-helix shape, which was not observed in the model based on WNV NS3_{PRO}. On the other hand, the WNV-based model had its N-terminal end projected into a beta strand that contributes to one of the beta-sheets found in the protein (data not shown).

The superposition of these two 3D models based exclusively on the WNV or DENV NS3_{PRO} provided different features that were considered to build a third NS3_{PRO} model. In this last model, the protein core was modeled as the conserved consensus of all template structures, while the N-terminal region came from WNV

proteases and the C-terminal end from the DENV proteases. After comparing the merits of the three models, we finally retained the third one to be used in the next building steps to obtain the final NS2B_{CF}-Gly-NS3_{PRO} model.

After being validated by PROCHECK, the complete NS2B_{CF}-Gly-NS3_{PRO} model was found to be of acceptable accuracy and was chosen as definitive to carry on further studies. Figure 3B shows the structural alignment of the finished homology model with the newly available ligand-bound DENV structure (PDB id: 3U1I) (36). The alignment has a C_α atom RMSD of 0.9 Å for 195 atoms, which gives consistency to our built model. Additionally, as pointed by Knehans *et al.* (24), the oxyanion hole was also present in our model (Figure 3B, boxed detail), even though we did not use restraints during homology modeling. Those features, together with our model completeness, including both inhibitor and cofactor in the active conformations, as well as the gap-filling loop grafts directed by structural evidences found both in DENV and in WNV structures, made our NS2B_{CF}-Gly-NS3_{PRO} model a good starting point for MD simulations and the structure-based virtual screening campaign.

5.1.4.4.2 Protein model overall stability and flexibility

5.1.4.4.2.1 Molecular dynamics

The timeline analysis of the secondary structure pieces during the 30 ns of MD simulations (Figure 4) showed that protein

domains were stable during the trajectory, with the exception of the poly-glycine linker (residues 95 to 104 from the cofactor).

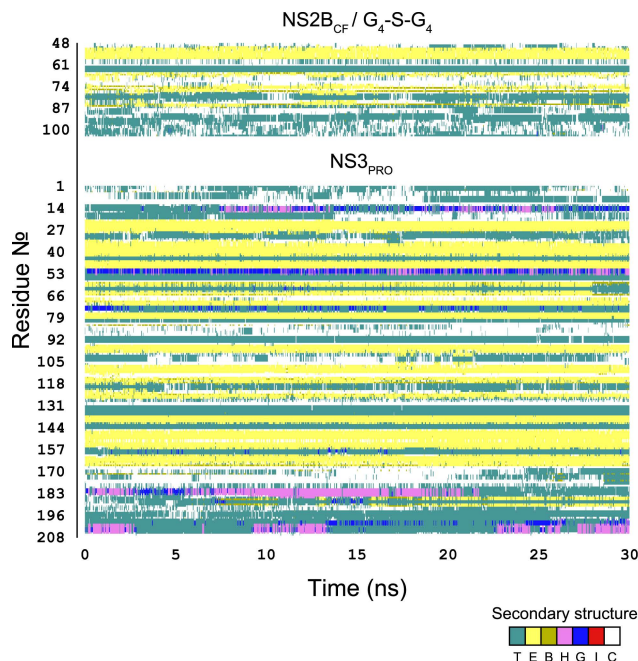


Figure 4. Secondary structure timeline analysis as computed by the timeline plugin contained in VMD. In the graphic, the β -sheet components turn (T) and extended conformation (E) are represented in teal and yellow respectively; isolated bridges are in dark yellow; degrees of helix are in pink (α -helix), and blue (3-10 helix); random coils are in white.

After the MD simulation run, both the ligand and the cofactor remained closely bound to the NS3_{PRO} domain (data not shown). Each frame collected from the trajectory file was then aligned to the first one based on their heavy atoms. RMSD analysis shows a 4-stages curve, with an increasing value in the first 2 ns, after which the model stabilizes for the next 5 ns, followed by a new increase and final stabilization after 15 ns (Figure 5A). This stabilization is achieved much faster, and with much lower RMSD values, when we analyze the protein core alone (residues 20 to 168, figure 5A, gray line). As the secondary structure modifications during MD simulations could not account for such a high deviation when

analyzing the whole model, we supposed that this was rather due to domain flexibility, which

in turn could regulate the access to the binding site.

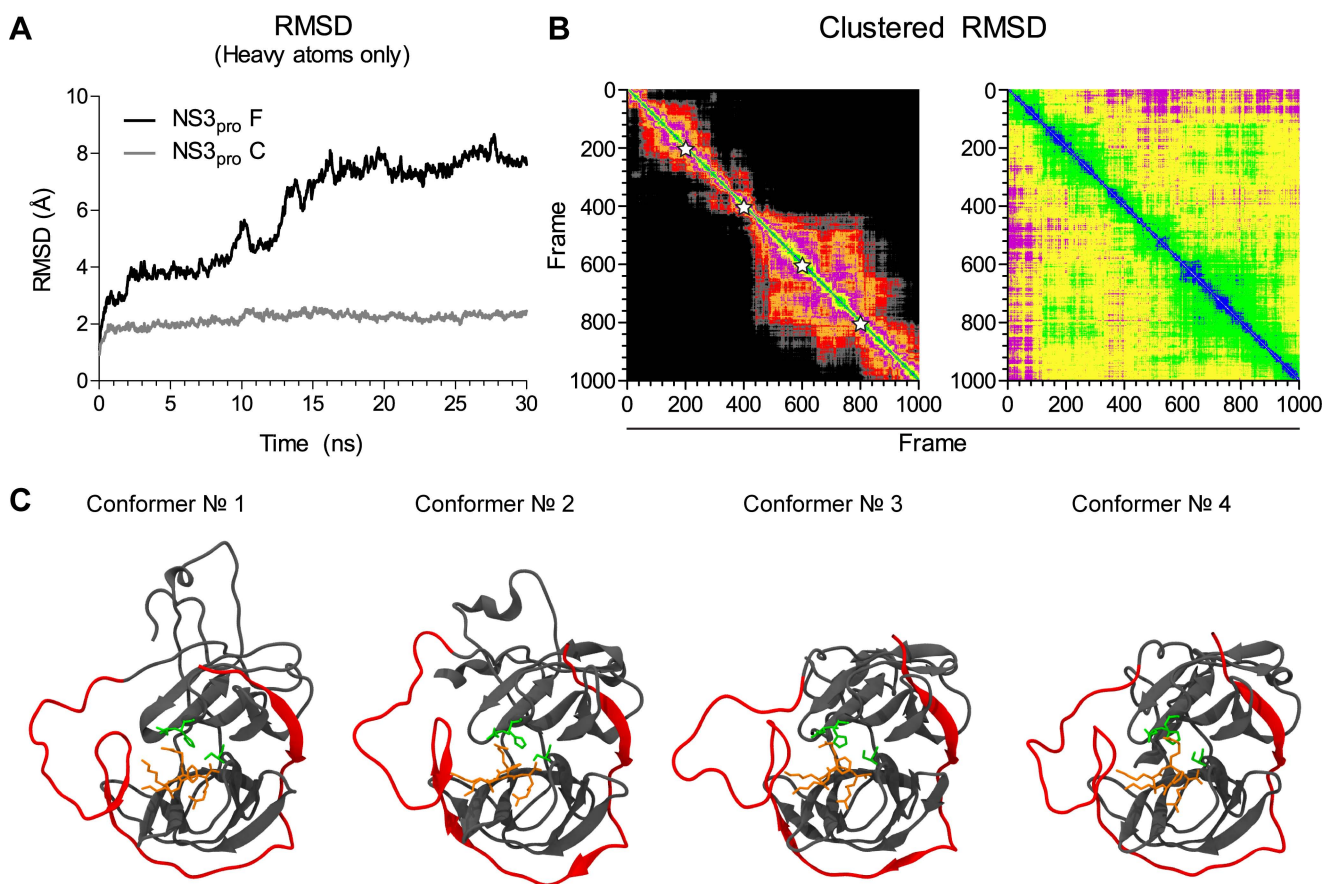


Figure 5. Single- and two-dimensional RMSD plots from the MD simulation trajectory, and their representative conformers. In (A), RMSD values from the whole construct (black line) were compared with RMSD values from the protein core only (gray line). In (B), a two-dimensional RMSD plot was utilized to provide information about conformational families during trajectory. Color ranges from black (as out of the cut-out distance) to white (as in identical structures). (C) Snapshots of representative conformations of each family extracted from the clustered (two-dimensional) RMSD analysis (NS2B_{CF} and linker in red, NS3_{PRO} in gray, active site residues in green and NDL in orange).

As it would be impossible to explore each frame separately in vHTS studies, we decided to use a two-dimensional plot clustering technique capable of investigating the structural relatedness of different parts of a MD trajectory (Figure 5B). By doing so, 4 distinct conformational families were detected when analyzing the whole sample, but only two (and with a high degree of relationship) could be detected while analyzing the protein core alone. By choosing the most conserved spots in the two-dimensional RMSD plot, representatives of each conformational family (snapshots

corresponding to frames 200, 400, 600 and 800, Figure 5C) were selected for a GOLD ensemble docking campaign. Instead of going for the frame 900 as the last conformation, we picked the frame 800, as it was more conserved and appeared to be closely related to the former.

5.1.4.4.2.2 Principal component analysis (PCA)

To verify which motions could account for RMSD variations, we finally performed a principal component analysis (PCA) from the MD trajectory. Results for the first 5 eigenvectors confirm that the greatest motions

indeed come from the C-terminal end of the molecule (Figure 6). This region has been described as highly flexible, as noticed by

differences in the two crystallographic structures for the full (both proteolytic and helicase domains) NS3 construct (34,37,50).

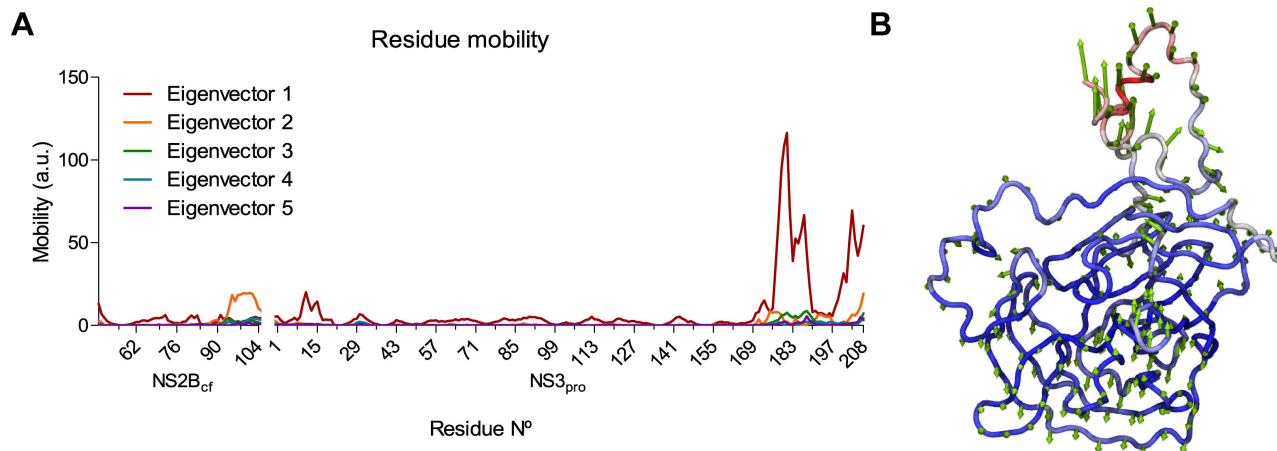


Figure 6. Individual residue mobility in principal component analysis. (A) The greatest eigenvector represents NS3_{PRO} C-terminal domain shifting, and the second eigenvector represents NS2B_{CF} C-terminal and glycine linker shifts. (B) Normal mode visualization from the first eigenvector, also revealing small motions that may be important to binding site plasticity.

However, this fact does not exclude the possible existence of small but significant motions that may be important for binding site plasticity (Figure 6B). To further investigate this question, we also performed 2D ligand-receptor interaction plots.

5.1.4.4.3 Protein binding site and binding modes flexibilities

5.1.4.4.3.1 Binding pocket detection and evolution

The DENV protease has a very shallow pocket (33,51), a fact that led us to employ the METAPOCKET servers. This server utilizes several redundant already validated geometric algorithms to search for proteins cavities. Thus, where one algorithm may fail, another one can correct the results, providing accurate topological information. This chosen pocket detection strategy showed to be in accordance with experimental data, with almost 70% (15/22,

Table 3) accuracy. Both S1 and S4 sub-pockets were fully detected, with less favorable results for sub-pockets S2 and S3 with 37.5 and 33.3% accuracy, respectively. It is noteworthy to say that this method was capable of detecting several residues not present in the experimental data. Those residues may constitute useful sub-pockets for designing new competitive molecules.

Following detection, the protease pocket coordinates (Figure 7A) were subjected to the MDpocket algorithm to be monitored for volume changes. As in many proteases (52–55), the pocket volume presented a ‘breathing’ behavior, adopting a sinusoidal curve when plotted against time (Figure 7B). This suggested that, even if the protein core remained stable during the same simulation, the binding site accessibility seemed to change, probably due to allosteric modulation provided by the more flexible domains.

Table 3. Pocket residues as detected by METAPOCKET compared to those found experimentally.

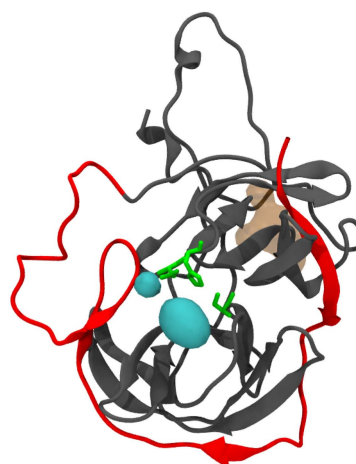
Sub-pocket	NS2B		NS3	
	Detected	Expected	Detected	Expected
S1	-	-	D129	D129
	-	-	F130	F130
	-	-	K131	K131
	-	-	P132	P132
	-	-	G133	G133
	-	-	T134	T134
	-	-	S135	S135
	-	-	Y150	Y150
S2	-	D80	-	V72
	D81	D81	-	K73
	-	G82	-	K74
	T83	T83	D75	D75
S3	M84	M84	-	-
	-	K85	-	-
	-	I86	-	-
S4	-	-	V154	V154
	-	-	V155	V155
N/A*	-	-	Q35	-
	-	-	V36	-
	-	-	W50	-
	-	-	H51	-
	-	-	G151	-
	-	-	N152	-
-	-	G153	-	

Detected pockets were identified by METAPOCKET, as described in Methods. * Detected residues which may be important to ligand binding, but were not previously reported in literature.

5.1.4.4.3.2 Protein/ligand interactions evolution

Ligplots of the protein/ligand binding modes (Figure 8) showed changes in both the protein sub-pockets conformation and in ligand interactions, indicating that the binding site is definitely not static and that this plasticity could be explored for competitive structure based drug design.

A



B

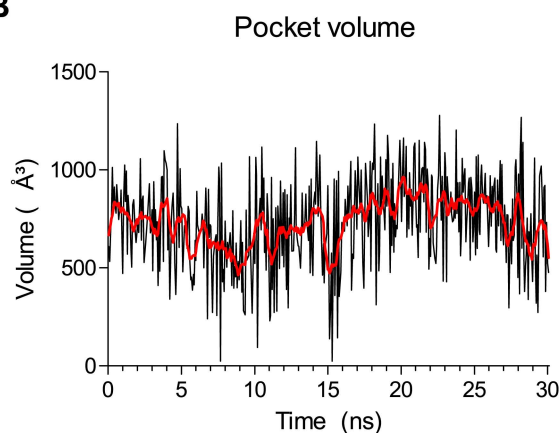


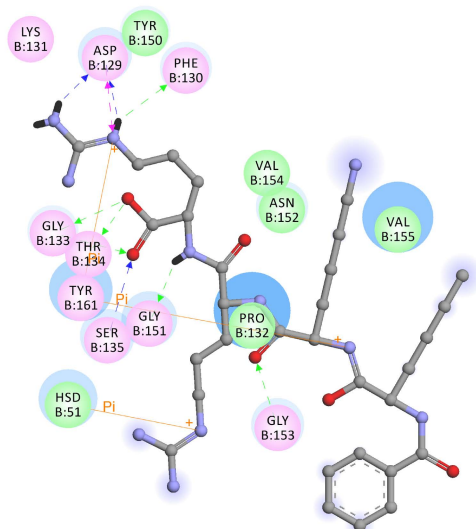
Figure 7. Pocket detection and evolution during MD simulation trajectory. Active site pocket (A) was identified by METAPOCKET and later monitored for changes in volume (B) by MDPocket. In A, DENV NS3_{PRO} is represented in gray; NS2B_{CF} and the glycine linker are in red. Active site residues are represented by green sticks, and the detected pocket is shown in cyan.

Noteworthy was the fact that the residues from the S2 sub-pocket were represented only by THR₈₃ in the conformations 2-4, but this could not be depicted in the conformation 1. In this case, the protein relaxation promoted by the MD simulation may have exposed this residue so it could be detected by this method. The residues from the S3 pocket (MET₈₄, LYS₈₅ and ILE₈₆) showed the same pattern, all absent in conformer 1 but with variable incidence in the other conformations. As for residue interactions, some were more conserved than others. In all conformations, residues GLY₁₃₃ and GLY₁₅₃

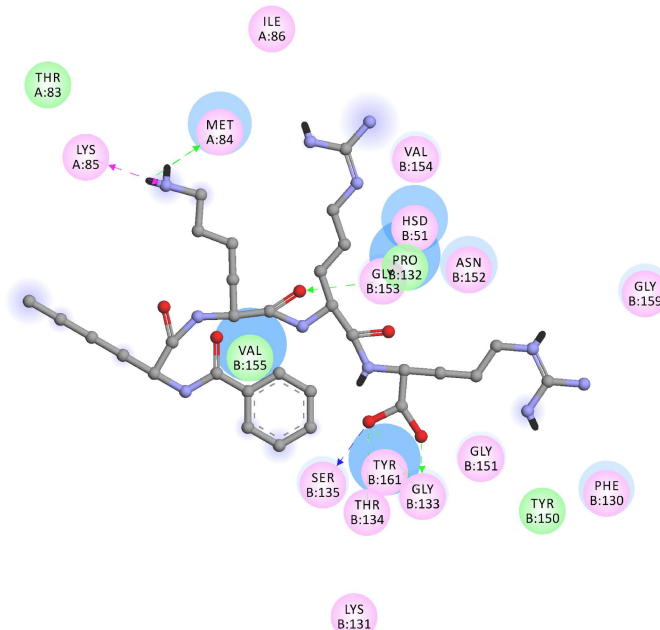
were involved with hydrogen bonds between their main chain atoms and the C-terminal

oxygen from P1 and the backbone oxygen from P2, respectively.

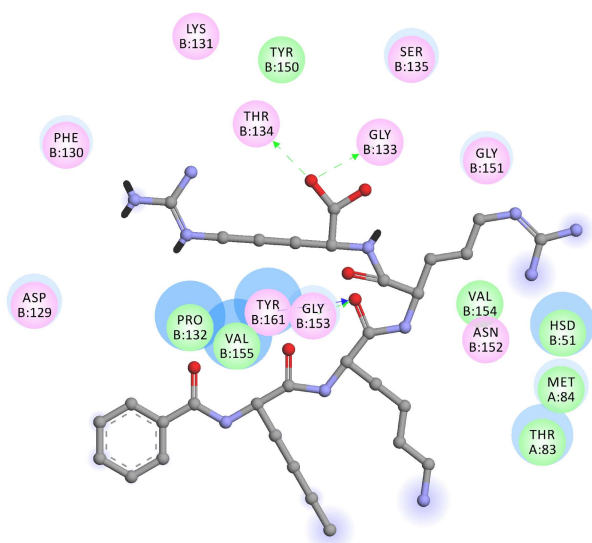
Conformer N° 1



Conformer N° 2



Conformer N° 3



Conformer N° 4

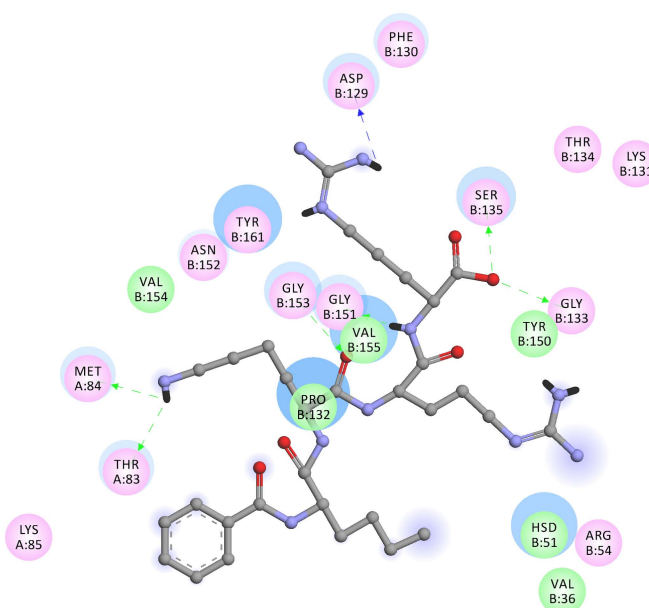


Figure 8. Binding mode fluctuations during the MD simulation. Both ligand and binding site plasticity contributes to changes in the binding mode between the ligand and protein. Pink circles indicate residues involved in hydrogen-bond, polar or charged interactions; green circles indicates residues involved in van der Waals interactions; π -interactions are represented by orange lines; green arrows indicate hydrogen-bonds interactions with amino acid main chains; blue arrows indicate side chain hydrogen-bond interactions. Arrowhead directs towards the electron donor.

On the other hand, π -interactions seemed to need a more constrained conformation, as this was observed only with the aromatic residues TYR₁₆₁ (with both P1 and P3) and HIS₅₁ (with P2) of the first conformation. Hydrogen bonds

coming from SER₁₃₅, GLY₁₅₁ and THR₁₃₄, were less conserved, but also related to the C-terminal oxygen from P1. Interactions between P3 and S3 were observed only in conformations 2 and 4;

side-chain interactions from P1 were observed only in conformation 1 and 4.

To investigate further the plasticity of the DENV NS2B_{CF}NS3_{PRO} binding site and the possible consequences on ligand binding, we analyzed the protein/ligand interactions during

the whole MD trajectory. For that purpose, we calculated the pair interaction energies between the ligand NDL and the protein (Figure 9A, black line), the ligand and surrounding water (Figure 9A, grey line) and for the ligand with each protein residue (Figure 9B).

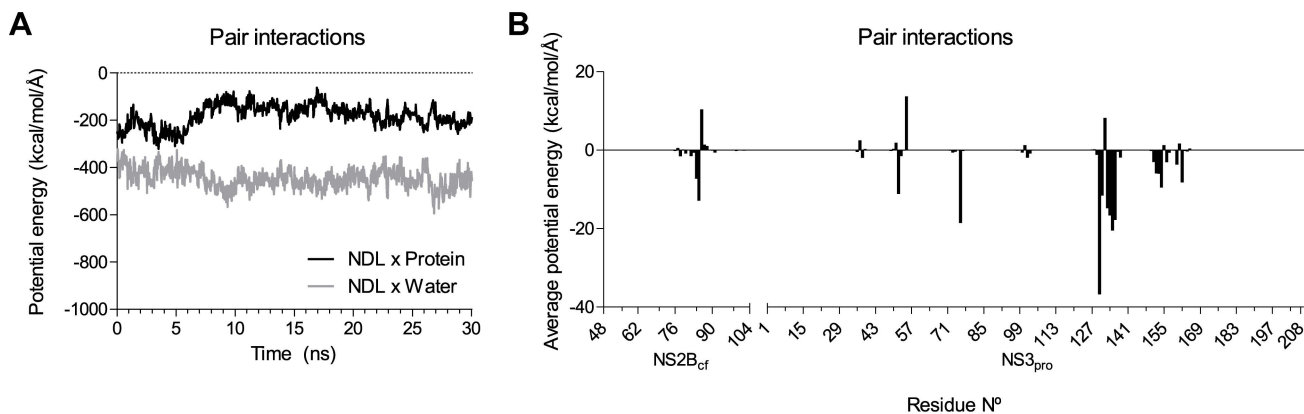


Figure 9. Van der Waals potential energies of pair interactions. (A) Pair interactions from the ligand NDL *versus* protein (black line), and NDL *versus* water (grey line). (B) Average potential energy of pair interactions from the ligand and each protein residue, allowing for contact points mapping.

Contact points between ligand and protein could be mapped by plotting the average van der Waals potential energy of each residue. Similar to results obtained from the binding plots, the main residues involved in the interactions with the ligand were from the sub-pockets S3 (residues MET₈₄, LYS₈₅ and ILE₈₆), and all the residues from S1 and S4. The greatest interaction averages came from ASP₁₂₉, followed by THR₁₃₄ and SER₁₃₅ from S1. However, this technique was more sensitive for detecting residues from the sub-pocket S2, as it was able to spot residues ASP₇₅, LYS₇₃ and VAL₇₂, none of them found within any conformation inspected through the ligand-binding plots.

When analyzing the ligand/protein and the ligand/water interactions together (Figure 9A),

we could see that while the energy of the former increases (thus, diminishing ligand-protein attraction), the latter seems to decrease proportionally (enhancing the ligand-water attraction). This is normally expected, as during MD simulations, the protein-ligand complex becomes more relaxed, allowing water to interact with the residues from the binding site. This results in decreased interactions between ligand and protein. It is interesting to note that, despite the attraction between ligand and protein being diminished at the beginning of the simulation, the complex achieves stabilization after 10 ns. As a matter of fact, one can say that during the last 5 ns protein-ligand interaction actually increases, approaching the initial state.

5.1.4.5 Discussion

One of the limitations in using crystallographic structures in structure-based drug design is the fact that those structures are often related to a particular energy minima state, which is a single conformation among many possible ones. Although this does not represent a significant problem, it may undermine virtual high throughput campaigns. To circumvent this issue, one may use a range of strategies for simulating protein flexibility, among which we can highlight the ensemble docking with structures derived from molecular dynamics simulations.

In the last few years, several DENV NS3_{PRO} automatic docking studies were published based on homology models (23–26,56–58). In our work, besides investigating protein flexibility, we were careful to build our DENV NS3_{PRO} model in the predicted active conformational state (including the oxyanion hole), a concern also present in the study by Knehans *et al.* (24). However, most of the other studies were based either on the retracted apoenzyme structure (26,57), or in the open holoenzyme conformation (23,25,56) and were based on a rigid protein system with a single conformation.

The first homology model for the DENV NS2B/NS3 protease was done by Brookhaven *et al.* in 1999 (59), but as it was built based on the more distantly related hepatitis C virus NS3/NS4A protease (PDB id 1JXP 56) as a template, several structural differences (C α RMSD of 11 Å) were observed when comparing

it to the recently resolved DENV-3 NS2B/NS3 protease (PDB id 3U1I, 35). Also, this work did not take into account the ligand-binding induced fit mechanism, as it was yet to be proposed (33,38), and thus, ligand was absent from their model. Later, in 2007, Kee *et al.* (61) used this same model for automatic docking studies. In 2004 and 2005, respectively, Niyomrattanakit *et al.* (62) and Chanprapaph *et al.* (58) built homology models based on the now retracted apo DENV NS3_{PRO} structure. Similarly, Ganesh *et al.*, in 2005 (57) and Tomlinson *et al.*, in 2009 (26) used homology models based on the retracted structure for automatic docking studies. For studying the importance of domain motion between the NS3_{PRO} and NS3_{HEL} domains through MD simulations (1 ns), Rosales-León *et al.*, in 2007 (63), built a full DENV NS3 model, but also based on this retracted structure. In 2010, two groups (29,64) independently published a similar approach for building active conformational models, where the cofactor and the ligand were extracted from the WNV NS2B_{CF}/NS3_{PRO} structure (PDB id 2FP7) and fused to the DENV NS3_{PRO}. However, the oxyanion hole is not in the active conformation in the work by Wichapong and colleagues (64), and could not be accessed in the work by Freceer and Miertus (29) for further comparisons. Also, neither models presented linker regions nor grafted loops.

Regarding protein flexibility, Ekonomiuk *et al.* (31) also postulated that it would be important to access new conformations for

vHTS in the case of the WNV protease. But instead of performing a long MD simulation, they preferred a short run (1 ns) to look for a conformation able to better accommodate a benzene ring later used for fragment docking. Figure 5A shows the importance of performing a longer simulation to our purposes: even if a protein seems stable during shorter trajectories, they can surpass greater conformational changes after longer simulation times, revealing important motions that otherwise would not be detectable. These motions can be related to important conformational states found during the protein equilibrium, which could be subjected to vHTS. When comparing the one- and two-dimensional RMSD graphs together, we can observe that the variations in the former are related to conformational changes in the latter, which are maintained in a clustered family. This means that conformational changes observed in the one-dimensional RMSD plot endure for some time, which is extremely desirable for finding new inhibitors that may bind to this protein state.

Despite presenting higher conservation during the trajectory, the binding site also presents considerable plasticity to be used in structure-based drug discovery. In the work by Ekonomiuk *et al.* (31), backbone changes as low as 0.8 Å between the crystallographic structure and their model extracted from MD simulations were enough to better accommodate three different fragments used in vHTS. In our case, the average binding site heavy atoms deviation

was approximately 2 Å, reflecting changes both in pocket volume (Figure 7B) and in the ligand binding modes (Figure 8).

In an already mentioned work, Wichapong *et al.* (64) performed 10 ns of MD simulations to examine ligand-protein interactions in three different models. Irrespective of the differences between this current work and Wichapong's and colleagues' (for example: their inhibitor was covalently bound to SER₁₃₅, the absence of the linker region and grafted loops, and the position of the oxyanion hole), our analysis seems to be in accordance with theirs. However, as they did not performed time-wise analysis related to different points in the trajectory, changes in the binding modes through the simulation could not be evaluated.

Besides the fact that almost every binding pocket residue found in the literature was also detected in our study, binding modes were sometimes quite distinct. With the recent publication of the ligand-bound structure of the DENV-3 NS3 protease (36), using the same inhibitor used in our study, we were able to compare the protease sub-pockets. While in the crystallographic data publication the GLY₁₅₁ makes interaction with the P2 residue from the inhibitor, in our study it interacts promptly with the P1 residue. Other changes include the GLY₁₅₃, part of the S3 sub-pocket in the crystallographic structure, but conserved in the S2 in all conformations from our model; and the residue TYR₁₆₁, which was capable of making π -interactions with both P1 and P3 residues in

our studies, while it is described as a S1 sub-pocket residue in the crystallographic structure. Also, LYS₈₅ from the NS2B cofactor was able to produce a charge-charge interaction with the P3 LYS from the inhibitor, a fact not observed in the work from Noble *et al.* (36). Altogether, those changes in binding mode, either between our models and the crystallographic structure or among different conformations extracted from our MD simulations, indicate that the binding site indeed presents significant flexibility that could be explored in structure based drug design.

5.1.4.6 Conclusions

This article describes the analysis of the *Dengue virus* NS3 protease flexibility through molecular dynamics simulations, as a part of an ensemble docking campaign for identifying potential new leads for drug design. As a result, we were able to clearly distinguish four different conformational states, which are being used in a following up virtual high throughput ensemble docking campaign. By doing this, we hope to discover new promising molecules with therapeutic potential to treat this life-threatening disease.

5.1.4.7 Acknowledgments

We would like to thank Dr. Tatsuya Nagata for providing the sequence and the vector used for modeling and expressing the DENV 1 NS2B-NS3 protease, Dr. Dave Ritchie for the critical reading of the manuscript, and the *Laboratoire Lorrain de Recherche en*

Informatique et ses Applications (LORIA) at Nancy, France, for providing computational resources.

5.1.4.8 References

1. Halstead SB, Nimmannitya S, Yamarat C, Russell PK. Hemorrhagic fever in Thailand; recent knowledge regarding etiology. *Japanese journal of medical science & biology.* 1967 Dec;20 Suppl:96–103.
2. Zellweger RM, Prestwood TR, Shresta S. Enhanced infection of liver sinusoidal endothelial cells in a mouse model of antibody-induced severe dengue disease. *Cell host & microbe.* 2010 Feb 18;7(2):128–39.
3. Beatty ME, Stone A, Fitzsimons DW, Hanna JN, Lam SK, Vong S, et al. Best practices in dengue surveillance: a report from the Asia-Pacific and Americas Dengue Prevention Boards. *PLoS neglected tropical diseases.* 2010 Jan;4(11):e890.
4. Schwartz E. Seasonality, Annual Trends, and Characteristics of Dengue among Ill Returned Travelers, 1997~2006. *Emerging Infectious Diseases.* 2008 Jul;14(7):1081–8.
5. Caminade C, Medlock JM, Ducheyne E, McIntyre KM, Leach S, Baylis M, et al. Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. *Journal of The Royal Society Interface.* 2012 Apr 25;
6. Gubler DJ. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21(st) Century. *Tropical medicine and health.* 2011 Dec;39(4 Suppl):3–11.
7. La Ruche G, Souarès Y, Armengaud A, Peloux-Petiot F, Delaunay P, Desprès P, et al. First two autochthonous dengue virus infections in metropolitan France, September 2010. *Euro surveillance : bulletin européen sur les maladies transmissibles = European communicable disease bulletin.* 2010 Sep 30;15(39):19676.
8. Franco C, Hynes N a, Bouri N, Henderson D a. The dengue threat to the United States.

- Biosecurity and bioterrorism : biodefense strategy, practice, and science. 2010 Sep;8(3):273–6.
9. Whitehead SS, Blaney JE, Durbin AP, Murphy BR. Prospects for a dengue virus vaccine. *Nature reviews. Microbiology*. 2007 Jul;5(7):518–28.
 10. Konishi E. Issues related to recent dengue vaccine development. *Tropical medicine and health*. 2011 Dec;39(4 Suppl):63–71.
 11. Steuber H, Hilgenfeld R. Recent advances in targeting viral proteases for the discovery of novel antivirals. *Current topics in medicinal chemistry*. 2010 Jan;10(3):323–45.
 12. Falgout B, Miller RH, Lai CJ. Deletion analysis of dengue virus type 4 nonstructural protein NS2B: identification of a domain required for NS2B-NS3 protease activity. *Journal of virology*. 1993 Apr;67(4):2034–42.
 13. Clum S, Ebner KE, Padmanabhan R. Cotranslational membrane insertion of the serine proteinase precursor NS2B-NS3(Pro) of dengue virus type 2 is required for efficient in vitro processing and is mediated through the hydrophobic regions of NS2B. *The Journal of biological chemistry*. 1997 Dec 5;272(49):30715–23.
 14. Umareddy I, Pluquet O, Wang QY, Vasudevan SG, Chevet E, Gu F. Dengue virus serotype infection specifies the activation of the unfolded protein response. *Virology journal*. 2007 Jan;4:91.
 15. Natarajan S. NS3 protease from flavivirus as a target for designing antiviral inhibitors against dengue virus. *Genetics and molecular biology*. 2010 Apr;33(2):214–9.
 16. Lescar J, Luo D, Xu T, Sampath A, Lim SP, Canard B, et al. Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from Dengue virus as a target. *Antiviral research*. 2008 Nov;80(2):94–101.
 17. Nitsche C, Behnam M a M, Steuer C, Klein CD. Retro peptide-hybrids as selective inhibitors of the Dengue virus NS2B-NS3 protease. *Antiviral research*. Elsevier B.V.; 2012 Mar 26;94(1):72–9.
 18. Tiew K-C, Dou D, Teramoto T, Lai H, Alliston KR, Lushington GH, et al. Inhibition of Dengue virus and West Nile virus proteases by click chemistry-derived benz[d]isothiazol-3(2H)-one derivatives. *Bioorganic & Medicinal Chemistry*. Elsevier Ltd; 2012 Feb;20(3):1213–21.
 19. Tomlinson SM, Watowich SJ. Use of parallel validation high-throughput screens to reduce false positives and identify novel dengue NS2B-NS3 protease inhibitors. *Antiviral research*. Elsevier B.V.; 2012 Feb;93(2):245–52.
 20. Nitsche C, Steuer C, Klein CD. Arylcianoacrylamides as inhibitors of the Dengue and West Nile virus proteases. *Bioorganic & medicinal chemistry*. Elsevier Ltd; 2011 Dec 15;19(24):7318–37.
 21. Schüller A, Yin Z, Brian Chia CS, Doan DNP, Kim H-K, Shang L, et al. Tripeptide inhibitors of dengue and West Nile virus NS2B-NS3 protease. *Antiviral research*. 2011 Oct;92(1):96–101.
 22. Cregar-Hernandez L, Jiao G-S, Johnson AT, Lehrer AT, Wong TAS, Margosiak S a. Small molecule pan-dengue and West Nile virus NS3 protease inhibitors. *Antiviral chemistry & chemotherapy*. 2011 Jan;21(5):209–17.
 23. Frimayanti N, Chee CF, Zain SM, Rahman NA. Design of new competitive dengue ns2b/ns3 protease inhibitors-a computational approach. *International journal of molecular sciences*. 2011 Jan;12(2):1089–100.
 24. Knehans T, Schüller A, Doan DN, Nacro K, Hill J, Güntert P, et al. Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *Journal of computer-aided molecular design*. 2011 Mar;25(3):263–74.
 25. Frimayanti N, Zain SM, Lee VS, Wahab H a, Yusof R, Abd Rahman N. Fragment-based molecular design of new competitive dengue Den2 Ns2b/Ns3 inhibitors from the components of fingerroot (*Boesenbergia rotunda*). *In silico biology*. 2012;11(1-2):29–37.
 26. Tomlinson SM, Malmstrom RD, Russo A, Mueller N, Pang Y-P, Watowich SJ.

- Structure-based discovery of dengue virus protease inhibitors. *Antiviral research*. 2009 Jun;82(3):110–4.
27. Tomlinson SM, Malmstrom RD, Watowich SJ. New approaches to structure-based discovery of dengue protease inhibitors. *Infectious disorders drug targets*. 2009 Jun;9(3):327–43.
 28. Shiryaev S a, Cheltsov AV, Gawlik K, Ratnikov BI, Strongin AY. Virtual ligand screening of the National Cancer Institute (NCI) compound library leads to the allosteric inhibitory scaffolds of the West Nile Virus NS3 proteinase. *Assay and drug development technologies*. 2011 Mar;9(1):69–78.
 29. Frecer V, Miertus S. Design, structure-based focusing and in silico screening of combinatorial library of peptidomimetic inhibitors of Dengue virus NS2B-NS3 protease. *Journal of computer-aided molecular design*. 2010 Mar;24(3):195–212.
 30. Ekonomiuk D, Su X-C, Ozawa K, Bodenreider C, Lim SP, Yin Z, et al. Discovery of a non-peptidic inhibitor of west nile virus NS3 protease by high-throughput docking. *PLoS neglected tropical diseases*. 2009 Jan;3(1):e356.
 31. Ekonomiuk D, Su X-C, Ozawa K, Bodenreider C, Lim SP, Otting G, et al. Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. *Journal of medicinal chemistry*. 2009 Aug 13;52(15):4860–8.
 32. Spyrakis F, BidonChanal A, Barril X, Luque FJ. Protein flexibility and ligand recognition: challenges for molecular modeling. *Current topics in medicinal chemistry*. 2011 Jan;11(2):192–210.
 33. Erbel P, Schiering N, D'Arcy A, Rensus M, Kroemer M, Lim SP, et al. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nature structural & molecular biology*. 2006 Apr;13(4):372–3.
 34. Luo D, Xu T, Hunke C, Grüber G, Vasudevan SG, Lescar J. Crystal structure of the NS3 protease-helicase from dengue virus. *Journal of virology*. 2008 Jan;82(1):173–83.
 35. Chandramouli S, Joseph JS, Daudenarde S, Gatchalian J, Cornillez-Ty C, Kuhn P. Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *Journal of virology*. 2010 Mar;84(6):3059–67.
 36. Noble CG, Seh CC, Chao AT, Shi PY. Ligand-bound structures of the dengue virus protease reveal the active conformation. *Journal of virology*. 2012 Jan;86(1):438–46.
 37. Luo D, Wei N, Doan DN, Paradkar PN, Chong Y, Davidson AD, et al. Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications. *The Journal of biological chemistry*. 2010 Jun 11;285(24):18817–27.
 38. Aleshin AE, Shiryaev SA, Strongin AY, Liddington RC. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. *Protein science : a publication of the Protein Society*. 2007 May;16(5):795–806.
 39. Robin G, Chappell K, Stoermer MJ, Hu S-H, Young PR, Fairlie DP, et al. Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation. *Journal of molecular biology*. Elsevier B.V.; 2009 Feb 6;385(5):1568–77.
 40. Eswar N, Webb B, Marti-Renom MA, Madhusudhan MS, Eramian D, Shen M-Y, et al. Comparative protein structure modeling using Modeller. *Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]*. 2006 Oct;Chapter 5:Unit 5.6.
 41. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *Journal of Applied Crystallography*. 1993 Apr 1;26(2):283–91.
 42. Humphrey W. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*. 1996 Feb;14(1):33–8.

43. Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: the biomolecular simulation program. *Journal of computational chemistry*. 2009 Jul 30;30(10):1545–614.
44. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *Journal of computational chemistry*. 2005 Dec;26(16):1781–802.
45. Darden T, York D, Pedersen L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *The Journal of Chemical Physics*. 1993;98(12):10089.
46. Ryckaert J-P, Ciccotti G, Berendsen HJ. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*. 1977 Mar;23(3):327–41.
47. Zhang Z, Li Y, Lin B, Schroeder M, Huang B. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. *Bioinformatics (Oxford, England)*. 2011 Aug 1;27(15):2083–8.
48. Schmidtke P, Bidon-Chanal A, Luque FJ, Barril X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. *Bioinformatics (Oxford, England)*. 2011 Dec 1;27(23):3276–85.
49. Eargle J, Wright D, Luthey-Schulten Z. Multiple Alignment of protein structures and sequences for VMD. *Bioinformatics (Oxford, England)*. 2006 Feb 15;22(4):504–6.
50. Assenberg R, Mastrangelo E, Walter TS, Verma A, Milani M, Owens RJ, et al. Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. *Journal of virology*. 2009 Dec;83(24):12895–906.
51. Salaemae W, Junaid M, Angsuthanasombat C, Katzenmeier G. Structure-guided mutagenesis of active site residues in the dengue virus two-component protease NS2B-NS3. *Journal of biomedical science*. 2010 Jan;17:68.
52. Miller DW, Agard D a. Enzyme specificity under dynamic control: a normal mode analysis of alpha-lytic protease. *Journal of molecular biology*. 1999 Mar 12;286(1):267–78.
53. Schmidt A, Lamzin VS. Extraction of functional motion in trypsin crystal structures. *Acta crystallographica. Section D, Biological crystallography*. 2005 Aug;61(Pt 8):1132–9.
54. Martin JR, Mulder F a, Karimi-Nejad Y, van der Zwan J, Mariani M, Schipper D, et al. The solution structure of serine protease PB92 from *Bacillus alcalophilus* presents a rigid fold with a flexible substrate-binding site. *Structure (London, England : 1993)*. 1997 Apr 15;5(4):521–32.
55. Liu S-Q, Meng Z-H, Fu Y-X, Zhang K-Q. Insights derived from molecular dynamics simulation into the molecular motions of serine protease proteinase K. *Journal of molecular modeling*. 2009 May 23;17–28.
56. Tomlinson SM, Watowich SJ. Anthracene-based inhibitors of dengue virus NS2B-NS3 protease. *Antiviral research. Elsevier B.V.*; 2011 Feb;89(2):127–35.
57. Ganesh VK, Muller N, Judge K, Luan C-H, Padmanabhan R, Murthy KHM. Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. *Bioorganic & medicinal chemistry*. 2005 Jan 3;13(1):257–64.
58. Chanprapaph S, Saparpakorn P, Sangma C, Niyomrattanakit P, Hannongbua S, Angsuthanasombat C, et al. Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. *Biochemical and biophysical research communications*. 2005 May 20;330(4):1237–46.
59. Brinkworth RI, Fairlie DP, Leung D, Young PR. Homology model of the dengue 2 virus NS3 protease: putative interactions with both substrate and NS2B cofactor. *The Journal of*

general virology. 1999 May;80 (Pt 5):1167–77.

60. Yan Y, Li Y, Munshi S, Sardana V, Cole JL, Sardana M, et al. Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form. *Protein science : a publication of the Protein Society*. 1998 Apr;7(4):837–47.
61. Lee Y, Tan S, Wahab H, Yusof R, Rahman NA. Nonsubstrate Based Inhibitors of Dengue Virus Serine Protease: A Molecular Docking Approach to Study Binding Interactions between Protease and Inhibitors. *Asia Pacific Journal of Molecular Biology and Biotechnology*. 2007;15(2):53–9.
62. Niyomrattanakit P, Winoyanuwattikun P, Chanprapaph S, Angsuthanasombat C, Panyim S, Katzenmeier G. Identification of residues in the dengue virus type 2 NS2B cofactor that are critical for NS3 protease activation. *Journal of virology*. 2004 Dec;78(24):13708–16.
63. Rosales-León L, Ortega-Lule G, Ruiz-Ordaz B. Analysis of the domain interactions between the protease and helicase of NS3 in dengue and hepatitis C virus. *Journal of molecular graphics & modelling*. 2007 Jan;25(5):585–94.
64. Wichapong K, Pianwanit S, Sippl W, Kokpol S. Homology modeling and molecular dynamics simulations of Dengue virus NS2B/NS3 protease: insight into molecular interaction. *Journal of molecular recognition : JMR*. 2010;23(3):283–300.

5.1.5 Acoplamento molecular automatizado

Além da cavidade do sítio ativo das quatro conformações identificadas e discutidas no manuscrito, durante nossas análises de dinâmica molecular outro possível alvo foi detectado na NS3_{PRO}. Em ambos os modelos sem o cofator (NS2B-/NDL+ e NS2B-/NDL-), o sítio de ligação da NS2B_{CF} (figura 27A) permaneceu disponível durante toda a dinâmica. No entanto, ao contrário do sítio de ligação ao substrato, esse sítio não apresentou a mesma variação de volume durante a simulação (figura 27B). A estrutura correspondente ao último *frame* da simulação foi escolhida para representar essa nova cavidade e fazer parte de nossa campanha de varredura virtual.

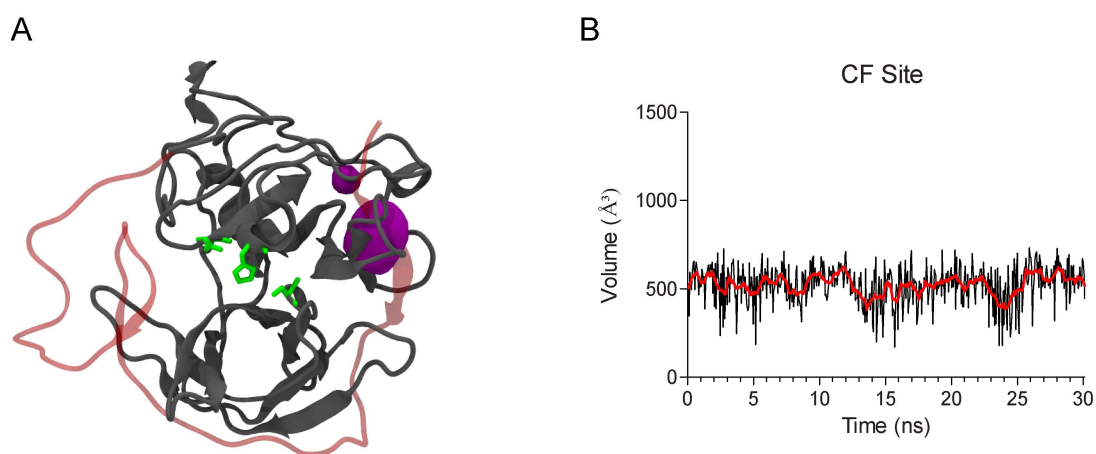


Figura 27. Sítio de ligação do cofator e sua variação de volume durante a simulação. A) Representação do modelo NS2B_{CF}-Gly-NS3_{PRO} com o domínio NS2B_{CF} (em vermelho translúcido) removido para uma varredura de cavidades. A cavidade identificada no sítio de ancoragem da fita-β N-terminal do cofator está representado em roxo. Em verde a tríade catalítica H51, D75 e S135. B) Variação de volume dessa cavidade, com a linha suavizada mostrando que o volume permaneceu constante, sem o comportamento observado para o sítio ativo (vide manuscrito em anexo).

Após a escolha das estruturas que fariam parte do conjunto de alvos de nossa campanha de *ensemble docking* (conformações 1, 2, 3 e 4 do manuscrito, após remoção do ligante NDL), assim como uma conformação para *docking* clássico (conformação final do modelo NS2B₊/NDL₊, após remoção do cofator), finalmente iniciamos a varredura propriamente dita, utilizando para isso o aplicativo GOLD (CCDC; Cambridge, Reino Unido). A técnica de “*ensemble docking*” difere da técnica tradicional ao levar em consideração o mecanismo de

encaixe induzido presente em muitos receptores/moléculas. O programa GOLD, ao invés de tratar as estruturas separadamente, parte de um conjunto de estruturas (obtidas por meio de cristalografia, NMR ou modelagem por homologia), nas quais ele sucessivamente testa cada uma das poses do ligante (geradas por seu algoritmo genético). Dentre todas as combinações possíveis, somente é escolhida uma estrutura do conjunto e uma pose do ligante, correspondentes à combinação com maior escore dentre todas as testadas. O ranque é aplicado imediatamente após esse processo, de forma a poupar tempo de análise. Se o ranque fosse aplicado somente ao final do *docking*, seria necessário classificar e ranquear mais de 12.675.000 moléculas, enquanto que com a técnica de *ensemble docking* esse número é dividido por quatro.

5.1.5.1 Campanha de busca de novos inibidores contra a NS3_{PRO}

Ao total, sete quimiotecas (bibliotecas de compostos químicos) foram testadas, das quais seis direcionadas ao sítio de ligação ao substrato (inibidores de protease e compostos com atividade anti-HCV) e uma ao sítio de ligação do cofator (inibidores de interação proteína-proteína). Após o término do protocolo de acoplamento automatizado, apenas as 50 melhores soluções encontradas foram ordenadas em uma lista de acordo com seu valor GoldScore. Dentre as seis quimiotecas testadas para o sítio de ligação ao substrato, a que obteve melhor resultado foi a de compostos com atividade anti-HCV (figura 28).

Enquanto a mediana de seus 50 melhores escores foi de 86,25, as de outras quimiotecas não passaram de 75,8 (inibidores de protease trombina-símiles). Isso pode ser explicado pelo fato da protease NS3 possuir um sítio de ligação mais raso, diferente de outras serinoproteases convencionais (ERBEL *et al.*, 2006; HUANG, 2012). Outros autores já implicaram esse problema como responsável pelas dificuldades encontradas em campanhas de *virtual screening* contra o HCV (NATARAJAN, 2010). O maior sucesso da quimioteca de

inibidores anti-HCV na nossa investida pode ser explicado pelo fato de ela ter sido construída com base na aplicação da tecnologia de similaridade de “impressão digital” 2D (*2D Fingerprint Technology*, em inglês) em moléculas com atividade antiviral comprovadas. O *docking*, nesse caso, pode ter identificado componentes de amplo espectro dessa biblioteca que teriam como alvo não só a protease NS3 de HCV, mas também a do vírus da dengue.

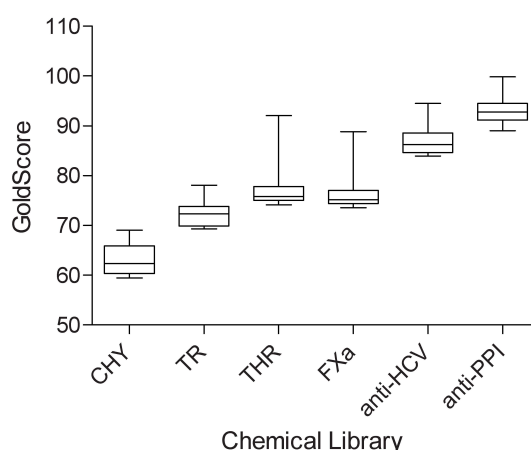


Figura 28. Distribuição de escores das diferentes quimiotecas. CHY: inibidores quimotripsina-símile; TR: inibidores tripsina-símiles; THR: inibidores trombina-símiles; FXa: inibidores fator-Xa-símiles; anti-HCV: vírus da hepatite C; PPI: interação proteína-proteína (inibidores do sítio do cofator). Apesar de apresentar uma estrutura quimotripsina-símile (ERBEL *et al.*, 2006), essa foi a quimioteca que apresentou escores mais modestos.

Os resultados para o banco de dados de inibidores da interação proteína-proteína foram surpreendentes, pois apresentaram praticamente todos os escores acima de 90. Essa não é a primeira vez que o sítio de ligação do cofator é explorado em uma campanha de VHTs. Em 2011, Shiryaev e colaboradores publicaram um trabalho com a mesma abordagem, mas, no entanto, visando à protease do WNV (SHIRYAEV *et al.*, 2011). Essa abordagem oferece vantagens sobre o enfoque mais tradicional de *virtual screening* que visa o sítio ativo da protease, pois os sítios ativos podem ser conservados entre proteases virais e de humanos (SHIRYAEV *et al.*, 2011). Dessa forma, ao buscarmos inibidores com potencial terapêutico que visem o sítio de ligação do cofator estamos diminuindo as chances de causar inibições cruzadas, e, logo, de efeitos colaterais.

Para diminuir ainda mais o número de moléculas a ser testado, executamos um *script* que comparou cada um dos compostos entre si, gerando uma matriz de similaridade baseada no coeficiente de Tanimoto. Essa matriz, composta por valores entre 0 (nenhuma similaridade) e 1 (moléculas idênticas) serviu como filtro para a escolha dos compostos. Apenas moléculas com um coeficiente abaixo de 0,7 foram selecionadas para serem testadas em ensaios de inibição da enzima recombinante. Um esquema do processo de filtragem desses compostos pode ser visualizado a seguir (figura 29).

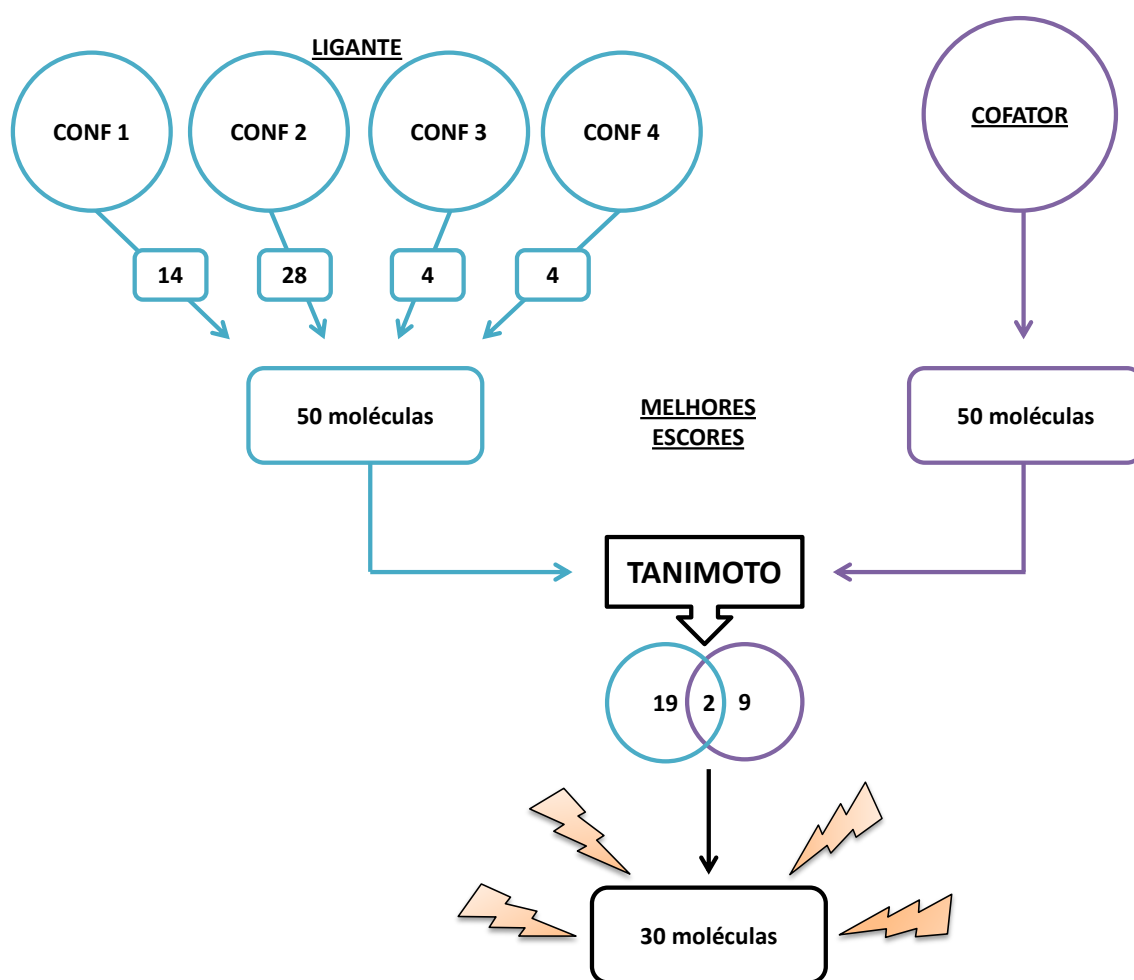


Figura 29. Processo de filtragem e seleção das moléculas destinadas à etapa de testes de inibição *in vitro*. A aplicação do filtro de similaridade baseado no coeficiente de Tanimoto teve o objetivo de diminuir o número de moléculas a ser testadas. Duas moléculas foram identificadas simultaneamente nas duas campanhas de triagem molecular (para o sítio ativo e para o sítio do cofator).

Os nomes das moléculas selecionadas por esse processo podem ser visualizados na tabela 6, juntamente com seus ranques antes e após a filtragem por similaridade.

Tabela 6. Compostos selecionados para os testes de inibição *in vitro*.

Compound Nº	Tanimoto Rank	Gold Rank	GoldScore	Molecule ID	Conformer Nº
09	01	01	94,53	F0520-0770	2
15	02	02	91,16	F0745-1098	1
18	03	04	90,68	F0772-0525	2
17	04	06	90,56	F0772-0522	2
14	05	08	90,19	F0737-0405	3
16	06	09	89,57	F0772-0511	2
11	07	10	89,34	F0715-0300	1
02	08	12	88,95	F0325-0146	4
21	09	13	88,44	F0816-0342	3
27	10	16	87,96	F3394-0640	1
29	11	17	87,82	F5123-0925	2
12	12	21	87,33	F0721-0831	2
30	13	22	86,96	F5295-0032	4
04	14	28	85,94	F0507-2261	2
13	15	31	85,66	F0725-0503	2
23	16	32	85,41	F1598-0004	1
05	17	35	84,89	F0508-1921	2
28	18	40	84,36	F3407-3992	1
22	19	42	84,29	F0829-0004	2
06	20	49	83,91	F0508-3705	2
20	21	50	83,88	F0772-2455	2
19	01	01	99,76	F0772-2132	CF
25	02	14	94,34	F2018-2865	CF
01	03	15	94,31	F0174-0048	CF
27	04	18	93,94	F3394-0640	CF
08	05	21	93,59	F0514-0584	CF
03	06	28	92,44	F0382-0020	CF
10	07	35	91,38	F0520-0786	CF
02	08	41	90,67	F0325-0146	CF
07	09	42	90,36	F0509-4253	CF
26	10	43	89,55	F3165-0782	CF
24	11	50	88,95	F1603-0109	CF

Compostos marcados em **vermelho** foram encontrados nas duas campanhas simultaneamente. Compostos em **verde** foram posteriormente selecionados para os ensaios de IC50. As primeiras 21 fileiras correspondem às classificações da campanha para o sítio ativo, as últimas 11 para o cofator.

Dois compostos (F0325-0146 e F3394-0640, marcados em vermelho na tabela 6) nos chamam a atenção, pois além de terem sido identificados nas duas campanhas

simultaneamente (para o sítio ativo e para o sítio do cofator), foram mantidos mesmo após aplicarmos o coeficiente de Tanimoto.

Uma análise da preferência de conformações (figura 30) observada a partir das soluções da tabela 6 nos mostra a importância da escolha do método de amostragem derivado de dinâmica molecular: está claro o grau de enriquecimento das soluções devido à presença da conformação 2, por exemplo, que está associada a mais de 50% dos resultados.

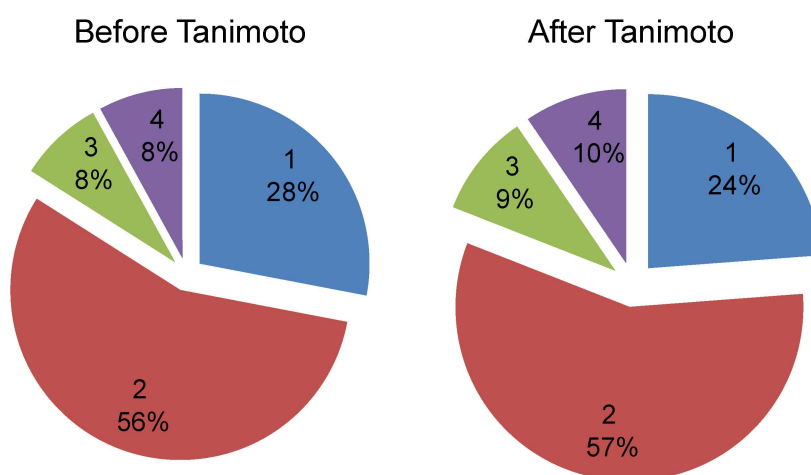


Figura 30. Preferência dos compostos pelas diferentes conformações utilizadas na busca por inibidores do sítio ativo. Conformação 1: azul; conformação 2: vermelho; conformação 3: verde; e conformação 4: roxo. Após exclusão de candidatos pelo coeficiente de Tanimoto, os compostos identificados para a conformação 1 tiveram mais cortes, sugerindo que estes fossem mais semelhantes entre si.

5.1.5.1.1 Ensaio de atividade da enzima recombinante

Após um *screening* prévio dos 30 compostos únicos identificados por ambas as campanhas de vHTS (dados não mostrados), chegamos a 11 candidatos com maior eficácia (figura 31A). Destes, apenas quatro (dois de cada campanha) foram selecionados para ensaios preliminares de IC₅₀ (figura 31B). Os compostos 1 e 21 apresentaram melhores resultados (IC₅₀ de 78,73 ± 0,17 μM e 93,17 ± 0,09 μM, respectivamente) quando comparados com os compostos 15 e 26 (IC₅₀ de 101,6 ± 0,12 μM e 94,95 ± 0,14 μM, respectivamente).

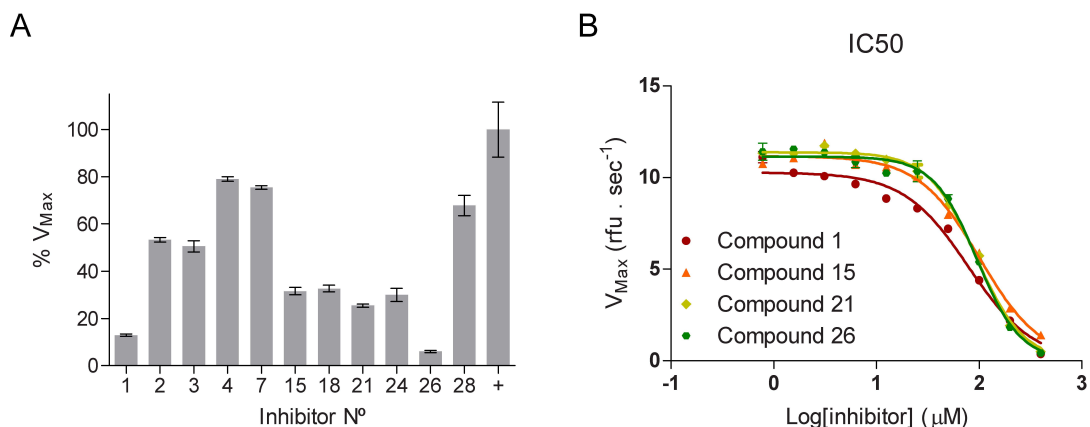


Figura 31. Ensaio de inibição enzimática a partir dos compostos identificados por vHTS. A) Onze melhores inibidores identificados em etapas preliminares. B) IC50 dos 4 compostos mais promissores.

Apesar do resultado mais satisfatório para o composto 1, esses números estão abaixo do esperado para uma campanha de vHTS com fins terapêuticos. No trabalho de Shiryaev e colaboradores (SHIRYAEV *et al.*, 2011), por exemplo, foram encontrados compostos com atividade inibitória muito mais satisfatória para o WNV (valores de 0,23 a 9,83 µM). Entretanto, é importante salientar que os mesmos compostos apresentaram um valor mais elevado contra a protease do DENV (2,04 a >100 µM). Além disso, a quimioteca testada no referido trabalho (*National Cancer Institute compound library*) possui mais de 275.000 compostos, enquanto que neste trabalho foi utilizada apenas uma fração disso (devido à sobreposição das coleções, o número exato de compostos únicos não pôde ser determinado, mas se cada quimioteca possuísse apenas moléculas exclusivas, esse número não ultrapassaria 64.000 compostos). Outro fator que deve ser levado em consideração é a concentração de enzima utilizada nestes ensaios. Enquanto que nós utilizamos 168 nM de enzima, na maioria dos outros artigos foram utilizados somente 10 nM. A razão molar entre a concentração da enzima e dos inibidores pode ser em parte responsável pelos altos valores de IC50 encontrados no presente trabalho.

Quanto ao modo de ação dos inibidores, os compostos 1 e 26 se ligam ao sítio de ligação ao cofator, enquanto que os compostos 15 e 21 ao sítio de ligação ao substrato (figura 32). A maior potência do composto 1, destinados ao sítio do cofator, é extremamente desejável, em virtude das características já descritas anteriormente (diminuição de efeitos secundários).

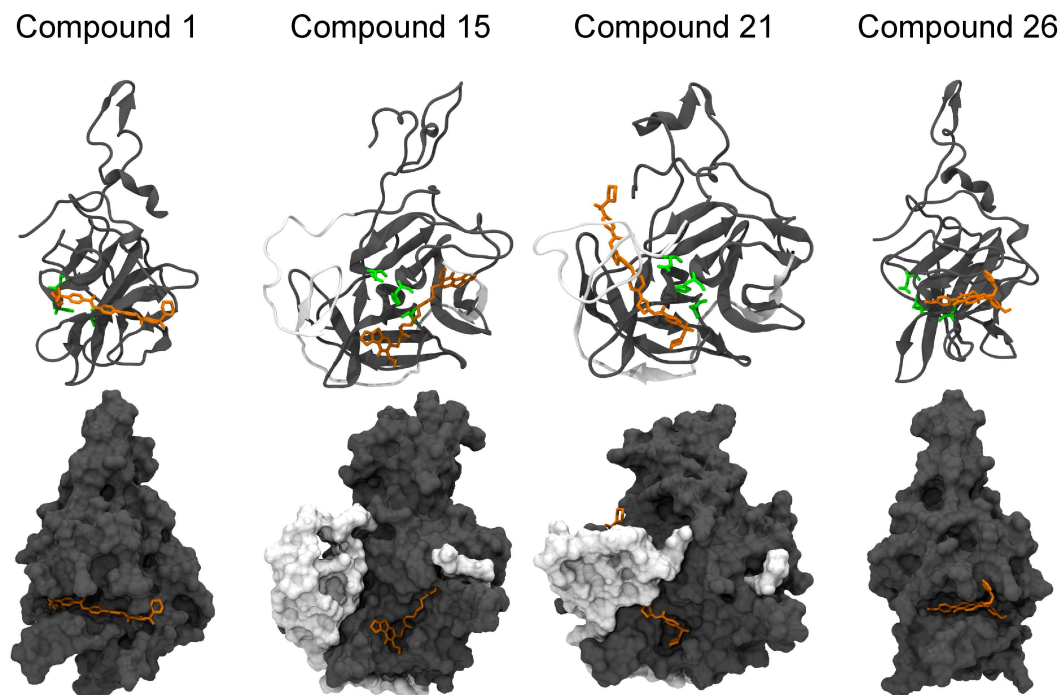


Figura 32. Poses dos compostos que apresentaram os melhores valores para atividade inibitória. Em cinza o domínio NS3PRO, em branco o domínio NS2BCF (removido no caso dos inibidores do sítio do cofator). Inibidores estão representados em laranja, e os resíduos do sítio catalítico em verde.

O composto 1 forma interações do tipo π com a ARG₂₄ da NS3_{PRO}, além de duas pontes de hidrogênio entre a MET₅₉ e ILE₂₂ (figura 33). Devido à sua característica alongada, esse inibidor provavelmente inibe estericamente a ligação do cofator, já que esses resíduos fazem parte da região de ancoragem da NS2B à NS3_{PRO}.

Compound 1

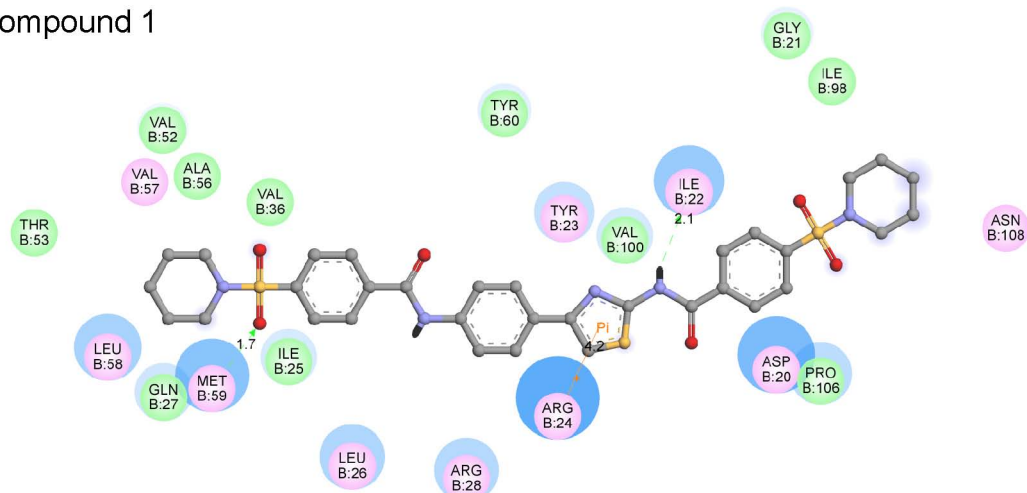


Figura 33. Interações entre o composto 1 e os resíduos de aminoácido da NS3_{PRO}. Ligações de hidrogênio entre os resíduos MET₅₉ e ILE₂₂ e uma ligação tipo π com a ARG₂₄ foram identificadas. Resíduos envolvidos em pontes de hidrogênio, interações de carga ou polar são representadas por círculos cor-de-rosa; resíduos envolvidos em interações van der Waals por círculos verdes; a superfície acessível ao solvente é representada por um halo azul ao redor do resíduo, com diâmetro proporcional à superfície acessível. Setas verdes e azuis representam interações de hidrogênio com a cadeia principal e cadeias laterais, respectivamente. A direção das setas vai de encontro ao doador de elétrons.

Já o composto 15 pode ser um inibidor competitivo pelo sítio ativo, uma vez que pode formar pontes de hidrogênio com os resíduos GLY₁₃₃, GLY₁₅₁ e SER₁₃₅, todos pertencentes ao sítio S1, além de uma ponte de hidrogênio com a ARG₅₄ (figura 34).

Compound 15

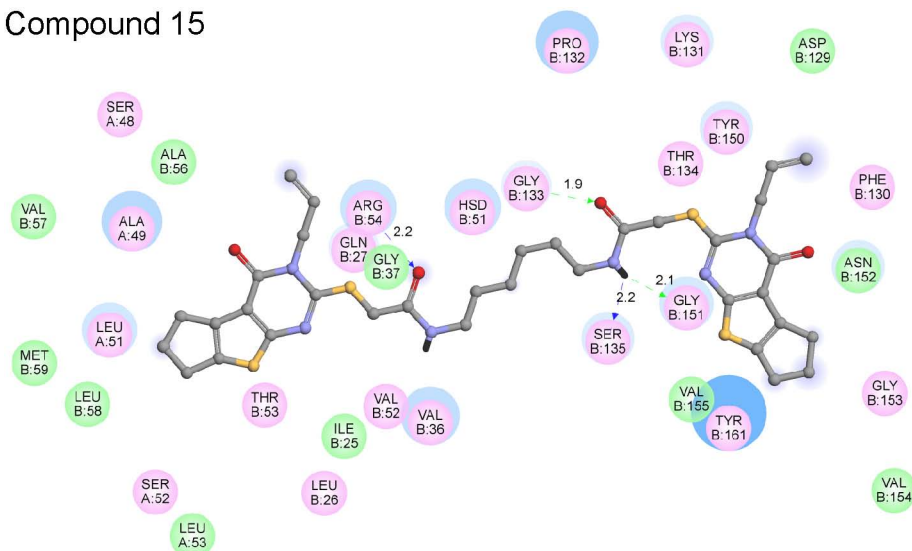


Figura 34. Interações entre o composto 15 e os resíduos de aminoácido da NS3_{PRO}. Ligações de hidrogênio foram identificadas com os resíduos ARG₅₄, SER₁₃₅, GLY₁₅₁, GLY₁₃₃. Outras representações de acordo com a legenda da figura 33.

O composto 21 pode ter um modo de ação misto. Devido à sua longa construção, pode interagir tanto com o sítio ativo por meio de pontes de hidrogênio com os resíduos GLY₁₃₃ e

Esse composto é capaz de formar uma interação do tipo π com a ARG₂₈, além de uma ponte de hidrogênio com o GLN₂₇.

É importante observar que o presente estudo nada mais é do que a primeira etapa no desenvolvimento racional de fármacos, e que a contestação de que as múltiplas conformações amostradas por dinâmica molecular são aptas ao *virtual screening* abre novos caminhos na busca de moléculas com potencial terapêutico contra a NS3_{PRO}. As próximas etapas consistem no refinamento dessas moléculas, seja por química combinatória ou por busca de estruturas similares em outras bibliotecas. Técnicas como o “*scaffold hopping*” (BÖHM *et al.*, 2004) ou “*fragment based vHTS*” (LAW *et al.*, 2009; ZOETE *et al.*, 2009) podem ser aplicadas com base nas poses e modos de ligação dos compostos com maior atividade inibitória. Os métodos de otimização de “*leads*”, como são chamados os compostos que guiam o processo de desenvolvimento racional de fármacos, tiveram efeitos de aprimoramento com compostos até 21.000 vezes mais potentes (800 μM \rightarrow 0,038 μM) do que os compostos originais para a protease do HCV (TSANTRIZOS, 2004). Dessa forma, esperamos ter contribuído não só com a descoberta de novas moléculas com potencial terapêutico, mas também com a aplicação bem sucedida de uma metodologia de obtenção de novas conformações por meio de dinâmicas moleculares, que poderá servir de exemplo na busca de inibidores contra outras doenças.

5.2 Doença de Chagas

Apesar de em menor grau, além da protease do vírus da dengue, dois alvos de *Trypanosoma cruzi* já estudados pelo nosso grupo também foram investigados: a metiltioadenosina fosforilase (TcMTAP; (NEVES, 2006) e a catepsina B (TcCatB; (GARCIA *et al.*, 1998; NÓBREGA *et al.*, 1998). Para ambos os alvos, foram construídos modelos por homologia, seguidos de refinamento por dinâmicas moleculares e acoplagem molecular. No caso da catepsina B (TcCatB), foram encontradas duas famílias de conformações distintas, ambas exploradas em uma campanha de *vHTS* com a colaboração do Prof. Dr. Roosevelt Silva (UFG). É importante ressaltar que, uma vez que uma análise biológica meticulosa dos modelos não foi realizada, os dados apresentados a seguir são preliminares, e que os mesmos foram incluídos no presente trabalho de forma a comprovar a possibilidade da aplicação, em laboratórios brasileiros, das técnicas estudadas durante estágio doutoral na França.

5.2.1 Modelagem por homologia

5.2.1.1 Blast

Assim como para a protease NS3 do DENV, a primeira etapa da modelagem por homologia destes alvos consistiu na busca de estruturas homólogas no banco de dados PDB. Para tanto, uma busca do tipo blastx foi realizada nos servidores do NCBI, restringindo as buscas às moléculas provenientes do PDB. Para o gene da TcMTAP, três estruturas homólogas foram encontradas, incluindo a MTAP humana. Seus valores de identidade, similaridade e porcentagem de lacunas estão na tabela 7. O alto grau de identidade com arquéias termófilas não era inesperado: em 2006, durante a caracterização desta enzima em sua tese de doutorado, Neves constatou que a MTAP de *T. cruzi* exibia identidade considerável entre outras MTAPs hipertermófilas, compartilhando 33% e 31% de identidade

com as enzimas de *Sulfolobus solfataricus* e *Pyrococcus furiosus*, respectivamente (NEVES, 2006).

Tabela 7. Identidade e similaridade entre a TcMTAP e estruturas homólogas depositadas no PDB.

Origem	PDB ID	Identidade	Similaridade	Cobertura	Lacunas
<i>Sulfolobus tokodaii</i> MTAP	1V4N	43%	61%	64%	4%
<i>Aeropyrum pernix</i> MTAP	1WTA	42%	58%	68%	7%
MTAP humana	1CB0	39%	57%	67%	3%

Para a TcCatB, duas estruturas estritamente relacionadas foram identificadas, ambas provenientes de *Trypanosoma brucei* (tabela 8). Devido à distância filogenética entre o nosso alvo e essas estruturas guia, outras candidatas com menor identidade foram desprezadas na construção do modelo.

Tabela 8. Identidade e similaridade entre a TcCatB e estruturas homólogas depositadas no PDB.

Origem	PDB ID	Identidade	Similaridade	Cobertura	Lacunas
<i>Trypanosoma brucei</i> CatB	3MOR	64%	77%	73%	1%
<i>Trypanosoma brucei</i> CatB	3HHI	63%	77%	73%	1%
CatB Humana	3A18	52%	62%	77%	6%

5.2.1.2 Alinhamento

O alinhamento das sequências incompletas extraídas dos arquivos PDB foi realizado novamente com base nas sequências originais depositadas no formato FASTA (figura 37 para a TcMTAP e figura 38 para a TcCatB). Entretanto, em relação à protease NS3 do DENV, essas estruturas homólogas se mostraram mais completas, provendo maior cobertura de coordenadas durante a modelagem por homologia desses alvos.

Mediante o alinhamento de sequência da TcMTAP, foi possível perceber duas diferenças importantes entre as estruturas: (i) a MTAP humana possui uma inserção de 5 resíduos de aminoácidos a partir da posição 131 (considerando a contagem de resíduos pela TcMTAP);

(ii) a TcMTAP possui uma inserção de 9 resíduos de aminoácidos a partir da posição 152, e uma possível inserção de 15 resíduos a partir da posição 281. Apesar disso, grande parte dos resíduos pertencentes ao sítio de ligação ao substrato é conservada.

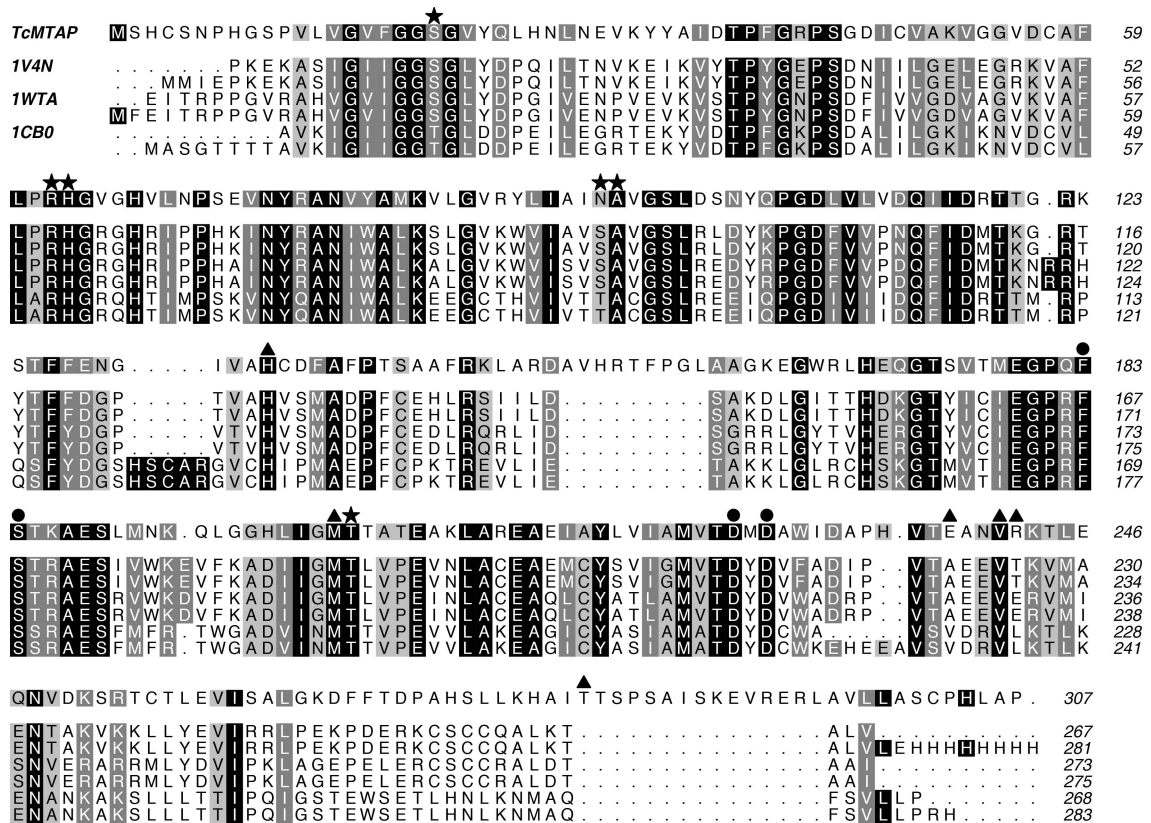


Figura 37. Alinhamento entre a sequência da TcMTAP e de estruturas homólogas depositadas no PDB. O código de conservação é o mesmo da figura 17. As posições dos sítios de ligação ao fosfato (estrela), metiltioribose (triângulo) e base (círculo) estão indicadas.

O mesmo grau de informação estrutural foi verificado nas estruturas guias para a modelagem da TcCatB. Como a maioria das cisteíno proteases, a TcCatB passa por um processo de auto hidrólise para ativar-se (GARCIA *et al.*, 1998), clivando assim um pré-peptídeo (região representada por uma seta completa na figura 38) que foi desconsiderado na modelagem por homologia. Entretanto, sua contraparte homóloga em *T. brucei* possui uma região N-terminal anormalmente longa, com 16 resíduos de aminoácidos além da região de ativação dessa protease (KERR *et al.*, 2010). Esperamos que quaisquer alterações estruturais causadas por esse peptídeo em *T. brucei* possam ter sido revertidas durante o refinamento por dinâmica molecular do nosso modelo.

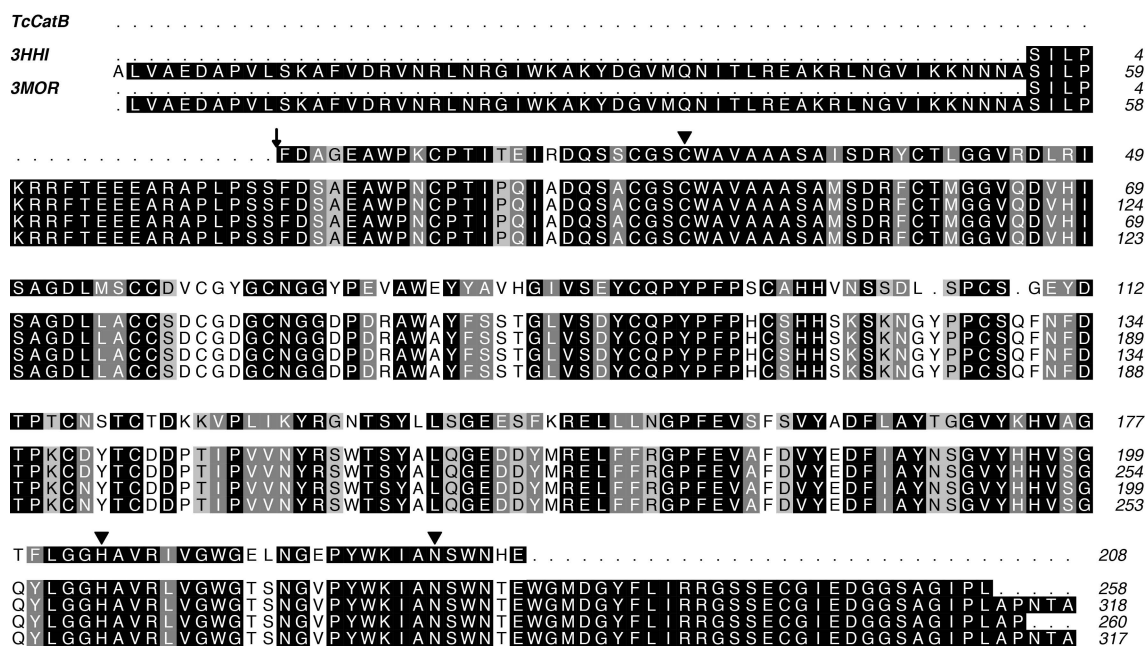


Figura 38. Alinhamento da sequência da TcCatB e suas estruturas homólogas depositadas no PDB. A posição dos resíduos catalíticos Q114, C120, H278 e N298 estão indicadas por triângulos invertidos. O sítio de ativação (clivagem do prépeptídeo) está indicado por uma seta.

5.2.1.3 Sobreposição estrutural dos templates

Quanto à análise estrutural dos *templates* (figura 39), poucas diferenças foram observadas entre eles, com a principal sendo a presença de uma hélice na estrutura da MTAP humana, ausente das estruturas homólogas de *S. tokodaii* e *A. pernix* (figura 39A e B, estrutura vermelha vs azul e amarela, respectivamente). As posições tridimensionais dos resíduos envolvidos na ligação do substrato também se mostraram conservadas. Os RMSDs entre os átomos da cadeia principal das estruturas das MTAPs em relação à estrutura de *S. tokodaii* foram de 0,62 e 2,17 Å, para as estruturas de *A. pernix* e *Homo sapiens*, respectivamente. A divergência mais alta com a MTAP humana é desejável no caso do desenvolvimento de novos fármacos contra essa enzima, uma vez que diminuiria as chances de inibição cruzada e, conseqüentemente, os efeitos colaterais. Entretanto, na ausência de uma estrutura com maior grau de identidade, fica impossível determinar se a TcMTAP seria mais próxima estruturalmente da MTAP humana ou da MTAP dos hipertermófilos.

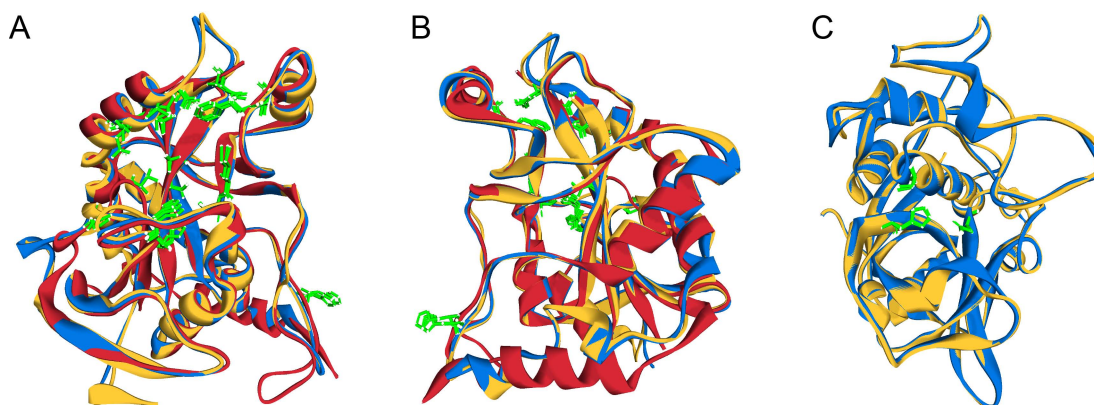


Figura 39. Sobreposição das estruturas homólogas utilizadas como guias na construção dos modelos TcMTAP e TcCatB. A) e B) *Templates* utilizados para a modelagem da TcMTAP. A estrutura que mais se diferencia das outras é a humana (em vermelho), como indicado por uma α -hélice exclusiva. C) *Templates* utilizados para a modelagem da TcCatB, evidenciando a boa conservação estrutural entre as espécies, mesmo na presença (estrutura azul) ou ausência (amarela) de ligante.

A sobreposição das estruturas homólogas à TcCatB (figura 39C), por sua vez, apresentaram um desvio muito baixo entre si (0,26 Å). Esse fato é interessante, pois a única diferença entre essas estruturas é a ligação (3HHI) ou não (3MOR) do inibidor [propilamino-3-hidroxi-butano-1,4-dionil]-isoleucil-prolina (CA-074). Isso significa que a enzima livre do ligante também apresenta uma conformação ativa, o que simplificou o processo de escolha de estruturas guias para a construção do modelo da TcCatB.

5.2.1.4 Modelagem por homologia e detecção de fendas catalíticas

Ao final do processo de modelagem por homologia, a estrutura obtidas para a TcCatB (figura 40A) se sobrepôs bem às estruturas guias, como esperado (RMSD 0,43 Å). Os resíduos de histidina do *loop* de oclusão (HIS₉₄ e HIS₉₅), responsáveis pela estabilização de substratos da atividade carboxipeptidásica presente apenas nessa família de cisteíno-peptidases (QUE *et al.*, 2002), se sobrepuseram perfeitamente às estruturas guia (dados não mostrados). Já a estrutura obtida para a TcMTAP (figura 40B) não se sobrepôs tão bem aos *templates* (RMSD 2,81 Å quando comparado à estrutura 1V4N, 2,75 Å em relação à 1WTA e 3,08 Å em relação à 1CB0). Entretanto, quando alinhadas apenas pelo sítio de ligação ao substrato (resíduos 20, 62, 63, 95, 96, 134, 183, 184, 202, 203, 226, 228, 238, 241, 242 e 281), estes valores são reduzidos para 1,56, 1,61 e 2,38 Å, referentes às estruturas 1V4N,

1WTA e 1CB0, respectivamente. Isso significa que talvez o sítio ativo da TcMTAP se assemelhe mais aos de suas contrapartes em arqueias hipertermófilas do que da enzima homóloga em humanos. Se comprovada experimentalmente, essa hipótese pode sustentar a possibilidade da criação de novas drogas específicas contra o *T. cruzi*, diminuindo as chances de inibição cruzada da MTAP humana do hospedeiro.

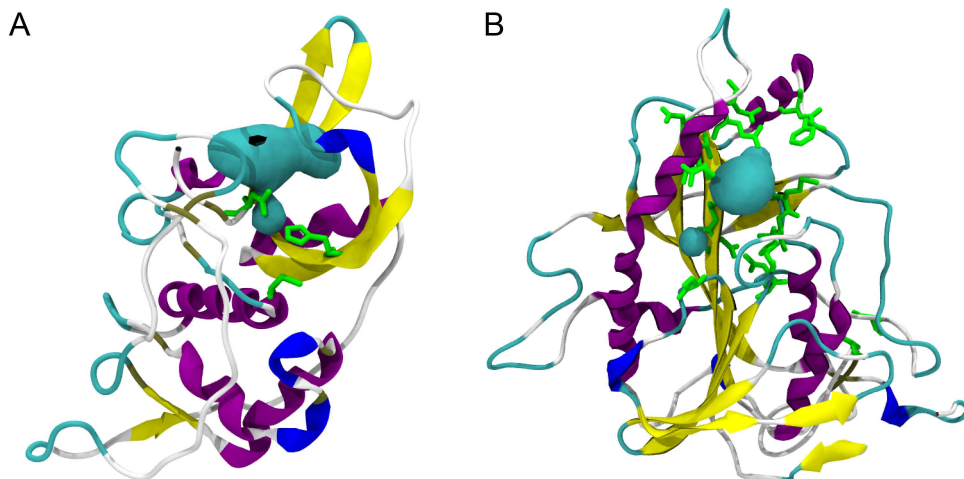


Figura 40. Modelos da TcCatB e TcMTAP obtidas a partir de estruturas homólogas. A) Modelo da TcCatB e a cavidade do sítio ativa, como detectada pelo programa METAPOCKET. B) Modelo da TcMTAP e sua cavidade. Resíduos que participam do sítio de ligação estão representados em verde.

Em ambos os modelos foi possível identificar cavidades correspondentes ao sítio de ligação (figura 40). No entanto, a cavidade do modelo TcCatB pareceu ter sua entrada regulada pelo *loop* de oclusão, restringindo o acesso aos resíduos catalíticos.

5.2.2 Dinâmicas moleculares

Analisando as trajetórias obtidas após as simulações de DM, podemos observar que o modelo da TcMTAP chega a um equilíbrio após aproximadamente 10 ns (figura 41A). A média de RMSD foi de 4,06 Å, o que, levando em consideração a distância filogenética entre o alvo e as estruturas guias, está dentro do esperado. Entretanto, por se tratar de uma estrutura nativamente oligomérica, talvez os contatos entre os monômeros possam estabilizar as mudanças conformacionais responsáveis por esse alto RMSD. Uma nova dinâmica molecular com dímeros da MTAP poderá ser realizada para responder essa

questão. Uma análise de RMSD bidimensional (figura 41B) foi capaz de identificar duas famílias de conformações dentro de um limite de 3,5 Å, com certo grau de sobreposição entre elas. Considerando que a segunda conformação corresponda ao resultado do refinamento obtido pela DM, resolvemos considerar apenas um representante da última família de conformações para experimentos de acoplamento molecular.

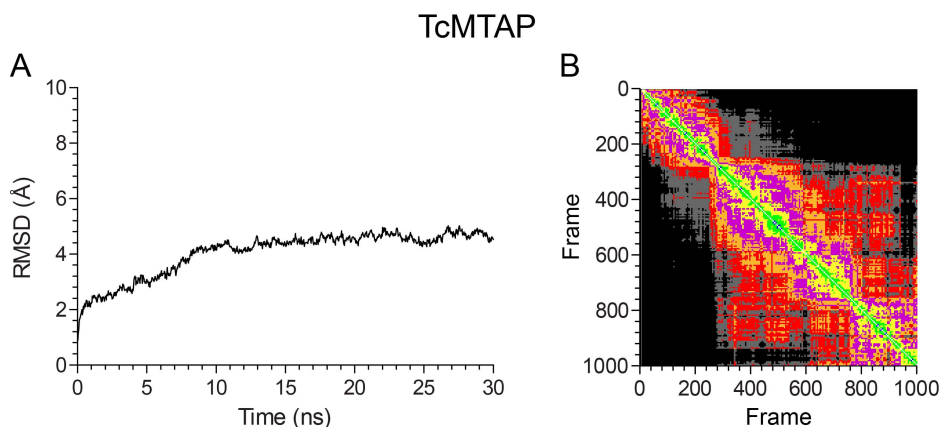


Figura 41. RMSD entre estruturas de cada quadro da simulação da TcMTAP. A) Após um desvio inicial abrupto, a estrutura aparenta se estabilizar após os 10 ns de simulação. B) Apesar de apresentar duas famílias de conformações, a primeira tem vários representantes contidos na segunda, e vice-versa.

Já a análise da trajetória da TcCatB (figura 42A) apresentou pelo menos 3 etapas distintas. A estrutura parece estável entre 5 e 20 ns, quando subitamente o RMSD volta a aumentar, gerando um segundo platô até aproximadamente 28 ns. Após esse período, os valores de RMSD abaixam novamente, representando uma estabilidade da estrutura até o fim da dinâmica. Ao contrário do caso da TcMTAP, onde os desvios da estrutura inicial eram esperados devido à distância filogenética entre as espécies, esses valores de RMSD (média de 3,97 Å) não eram esperados para a TcCatB. Uma possível explicação para esse fenômeno é a presença do pró-peptídeo anormalmente longo nas estruturas de *T. brucei* (KERR *et al.*, 2010), que não está presente na TcCatB nativa (GARCIA *et al.*, 1998).

O método de agrupamento conformacional para esse modelo foi ligeiramente diferente do que para os outros casos. Apesar de também ser baseado nos valores de RMSD de cada *frame*, este método (DAURA *et al.*, 1999) baseia-se no agrupamento de todas as estruturas

dentro de um limite estabelecido (3 Å, nesse caso). Em seguida, esse grupo é removido do *pool*, e o processo se repete para as conformações restantes. Apesar de ter pontos positivos em relação ao outro tipo de análise que realizamos, como a compartimentalização das conformações, esse método não nos apresenta de forma clara o grau de relação entre os membros de famílias diferentes. A partir dessa análise, foram identificadas duas famílias de conformações principais, e outras três pequenas famílias transitórias (figura 42B).

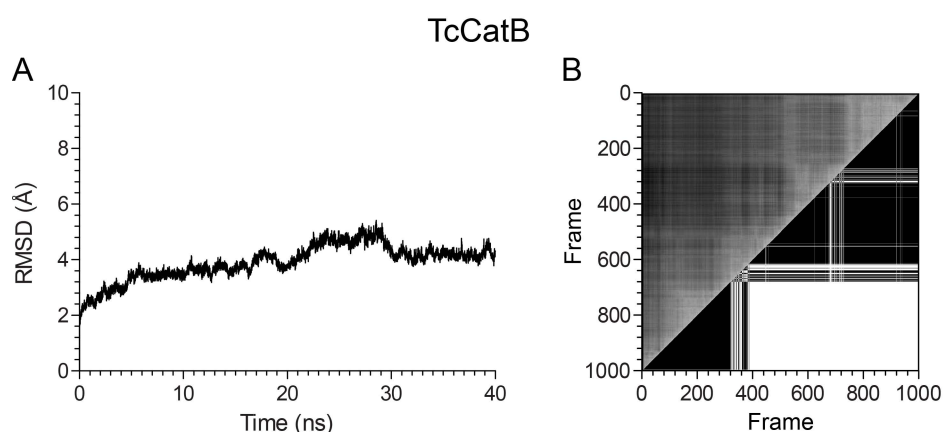


Figura 42. RMSD entre estruturas de cada quadro da simulação da TcCatB. A) A estrutura aparenta 3 estados de estabilização, ou platôs, até sua estabilização. B) Identificação das famílias de conformações da TcCatB. Na diagonal inferior as estruturas relacionadas foram agrupadas em suas respectivas famílias. Duas famílias principais foram identificadas.

As estruturas médias das duas maiores famílias de conformações (ou seja, a estrutura com o maior número de estruturas relacionadas) foram escolhidas como representativas para os experimentos de *docking*.

5.2.3 Acoplamento molecular automatizado

As estruturas obtidas a partir das simulações de dinâmica molecular foram submetidas a uma varredura virtual com compostos de duas bibliotecas diferentes: para a TcMTAP, o banco de dados escolhido foi o de inibidores adenosina-símiles (1.135 compostos, Life Chemicals Inc), e para a TcCatB, escolhemos a biblioteca de compostos anti-malária (23.527 moléculas, Life Chemicals Inc), esperando identificar aqueles com atividade cruzada entre as duas espécies de protozoários.

Os escores obtidos para a varredura da TcMTAP foram muito mais modestos do que para a protease do DENV, por exemplo. Os melhores escores para esse alvo ficaram em torno de 65. Considerando o tamanho reduzido da quimioteca utilizada nesse experimento, esses valores não são tão insatisfatórios, principalmente ao observarmos o resultado do acoplamento para o primeiro ranqueado (figura 43). Além de o composto ter obtido uma boa pose, o modo de ligação do inibidor também parece favorável, já que forma duas pontes de hidrogênio com os resíduos MET₂₀₂ e ALA₉₆, importantes para a ligação do substrato (NEVES, 2006), além de uma interação do tipo π com a HIS₆₇. Isso indica que, caso sua atividade seja confirmada, este composto possa ser um inibidor competitivo.

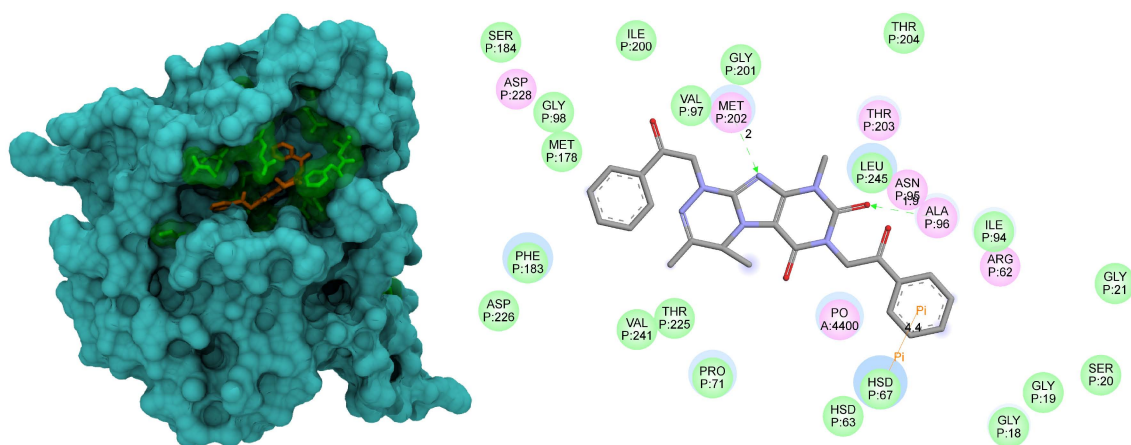


Figura 43. Pose e modo de ligação do composto melhor ranqueado na campanha de vHTS contra a TcMTAP. Representação de superfície da TcMTAP. Os resíduos envolvidos na ligação do substrato estão representados em verde, e o inibidor em laranja. Foram identificadas duas pontes de hidrogênio, entre os resíduos MET₂₀₂ e ALA₉₆, além de uma interação do tipo π com a HIS₆₇.

Quanto à campanha de vHTS contra a TcCatB, os nossos resultados anteriores não puderam ser comparados, já que tanto o programa utilizado para fazer o *docking* (AutoDock Vina *vs* GOLD) como os escores utilizados para ranquear os compostos foram diferentes. Apesar de termos utilizado duas conformações distintas para esse alvo, o protocolo de acoplamento molecular empregado não foi do tipo *ensemble docking* como anteriormente. Como resultado, obtivemos duas tabelas com os 100 melhores resultados para cada configuração da TcCatB, com alguns compostos se repetindo entre elas. A

vantagem de uma campanha de *ensemble docking* é facilmente percebida nesse caso. Ao invés de computar cada estrutura separadamente, ranqueando e gerando tabelas individuais, as melhores combinações entre proteína e ligante são ranqueadas após cada rodada de simulação, reduzindo assim o número de arquivos de saída a serem analisados. O composto 802, por exemplo, aparece na 4^a colocação após o refinamento do ranque por simulações exaustivas para a conformação 1 da TcCatB, mas o mesmo composto aparece na 100^a posição para a conformação 2. O mesmo acontece para outros 16 compostos dessa simulação. Em uma campanha de *ensemble docking*, esses compostos seriam selecionados apenas uma vez, e as colocações ocupadas pelos compostos repetidos seriam preenchidas por outros candidatos, elevando assim as chances de se encontrar moléculas com potencial terapêutico.

Da mesma forma que para a TcMTAP, as poses e modos de ligação dos melhores ligantes para as conformações 1 e 2 da TcCatB podem ser visualizados na figura 44. Apesar de não ter sido detectado pelo *Discovery Studio Visualizer*, o modo de ligação do composto ligado à conformação 1 parece estar relacionado com as interações com os resíduos GLY₂₃, GLY₆₈, TYR₁₃₆, SER₁₅₇ e GLY₁₈₂, apresentando, no entanto, distâncias superiores às de pontes de hidrogênio convencionais (dados não mostrados). No caso da segunda conformação, o modo de ligação do inibidor pôde ser detectado, e envolve pontes de hidrogênio com os resíduos ASN₆₆ e GLY₆₇. Como já discutido anteriormente, o *loop* de oclusão parece ter restringido o acesso dos inibidores, e, portanto, foi sugerido que o modo de ação desses compostos seja pelo impedimento estérico do acesso ao sítio ativo.

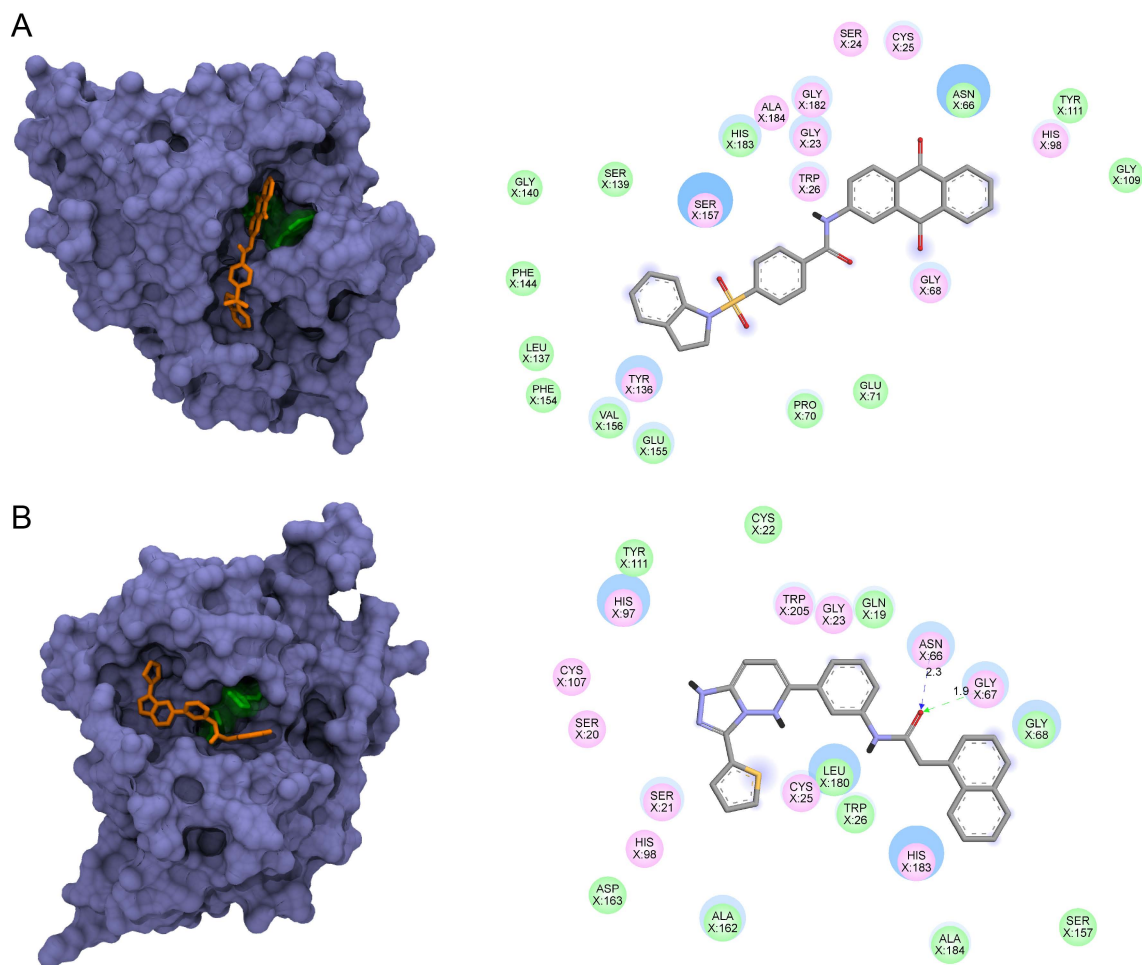


Figura 44. Poses e modos de ligação dos ligantes melhor ranqueados obtidos após vHTS contra a TcCatB. A) Melhor pose computada para a primeira conformação obtida da trajetória de DM. Apesar de apresentar uma boa pose, não foram identificadas ligações entre o alvo e o inibidor. B) Melhor pose computada para a segunda conformação. Nesse caso foram identificadas duas pontes de hidrogênio, com os resíduos ASN₆₆ e GLY₆₇.

Com 26% das pouco mais de 140 estruturas de proteínas de *Trypanosoma cruzi* tendo sido depositadas no PDB nos últimos dois anos, é de se esperar que avanços na área de identificação de novos inibidores por *virtual screening* acompanhem esse aumento de informação estrutural, com resultados cada vez mais refinados. Até lá, as técnicas como as utilizadas nesse trabalho (modelagem por homologia, dinâmicas moleculares e *docking*) devem continuar a ser empregadas por um longo período no combate a doenças tropicais negligenciadas. Em relação ao *T. cruzi*, inibidores contra os alvos dihidrofolato redutase (ZUCCOTTO *et al.*, 2001), tripanotiona redutase (MEIERING *et al.*, 2005; PRIETO *et al.*, 2006), gliceraldeído-3-fosfato desidrogenase (FREITAS *et al.*, 2009), trans-sialidase (NERES

et al., 2009) e cruzaina (ROGERS *et al.*, 2012) já foram identificados por vHTS. Uma vez validados, os inibidores encontrados nesse trabalho poderão aumentar a quantidade de patentes brasileiras no rol de moléculas com potencial terapêutico contra o mal de Chagas. A posse dessas patentes por parte do poder público é de extrema importância no desenvolvimento de fármacos contra doenças negligenciadas, uma vez que garantem o preço acessível e conseqüentemente distribuição àqueles que mais necessitam (HERRLING, 2007; MÉDECINS SANS FRONTIÈRES, 2001b).

6 Conclusões gerais

Com o intuito de estabelecer uma plataforma de desenvolvimento racional de fármacos, duas doenças importantes no cenário nacional foram escolhidas: a dengue e a doença de Chagas.

Por meio de ferramentas de bioinformática, obtivemos uma perspectiva do papel do cofator e do ligante no efeito de ajuste induzido da protease da dengue. Além disso, foram identificadas famílias de conformações durante uma simulação de dinâmica molecular, das quais representantes apresentam dissimilaridades importantes no sítio ativo, mináveis pela abordagem de varredura virtual. Essas conformações foram submetidas ao processo de acoplamento molecular, identificando compostos putativos que mais tarde apresentaram atividade inibitória comprovada em ensaios *in vitro*.

A partir de técnicas similares, dois alvos de *T. cruzi* também foram estudados. Entretanto, por apresentarem características diferentes, a flexibilidade dos mesmos não foi explorada de uma forma tão abrangente como para o vírus da dengue. Contudo, os resultados obtidos tanto para a TcMTAP como para a TcCatB são promissores, e requerem mais análises antes de serem validados.

Espera-se, por meio deste estudo piloto, expandir as técnicas utilizadas para outros alvos em processo de caracterização por nosso grupo, principalmente contra doenças negligenciadas, como leishmaniose, doença de Chagas e tuberculose.

Dos resultados obtidos para o vírus da dengue, uma repetição dos ensaios de determinação de IC₅₀ se faz necessária, já que a quantidade de enzima utilizada nos testes pode não ter sido adequada. A determinação das constantes de inibição também deve ser realizada. Além disso, seria desejável o teste dos inibidores identificados em ensaios de replicação viral, tanto em culturas de células de mosquito como de humanos. Uma rodada

de otimização dos ligantes identificados também pode ser realizada, buscando moléculas cada vez mais eficientes.

Já em relação aos alvos de *T. cruzi*, apesar dos dois modelos terem apresentado resultados putativos satisfatórios, novos testes devem ser realizados, com buscas em quimiotecas mais abrangentes e com modelos mais refinados. Por fim, moléculas selecionadas serão testadas tanto com as enzimas recombinantes quanto com parasitos em cultura.

Com base nos resultados apresentados, acreditamos que a utilização da bioinformática e de descoberta racional de drogas é plausível no cenário nacional, principalmente com a redução dos custos de sistemas de computação poderosos, que utilizam processadores gráficos para o cálculo das interações entre ligante e proteína e das iterações de dinâmicas moleculares. A existência de softwares de ancoragem livres como o AutoDock Vina e o PLANTS, diminuem ainda mais os investimentos necessários para o estabelecimento de uma plataforma de referência.

Juntas, as técnicas aplicadas neste trabalho podem gerar resultados promissores, com custos e tempo de desenvolvimento reduzidos, tornando as parcerias público-privadas mais atrativas aos olhos das indústrias farmacêuticas. Além disso, ao deter as patentes dessas moléculas promissoras, o setor público pode garantir a distribuição dos medicamentos a preço baixo, já que o detentor das patentes dita o preço de mercado de seus produtos.

7 Referências Bibliográficas

- ABDULLA, M.-H.; O'BRIEN, T.; MACKEY, Z. B. *et al.* RNA interference of *Trypanosoma brucei* cathepsin B and L affects disease progression in a mouse model. **PLoS neglected tropical diseases**, v. 2, n. 9, p. e298, jan 2008.
- ALESHIN, A. E.; SHIRYAEV, S. A.; STRONGIN, A. Y.; LIDDINGTON, R. C. Structural evidence for regulation and specificity of flaviviral proteases and evolution of the Flaviviridae fold. **Protein science, a publication of the Protein Society**, v. 16, n. 5, p. 795–806, maio 2007.
- ALONSO, H.; BLIZNYUK, A. A.; GREASY, J. E. Combining docking and molecular dynamic simulations in drug design. **Medicinal research reviews**, v. 26, n. 5, p. 531–68, set 2006.
- ALVAREZ, V. E.; NIEMIROWICZ, G. T.; CAZZULO, J. J. The peptidases of *Trypanosoma cruzi*: digestive enzymes, virulence factors, and mediators of autophagy and programmed cell death. **Biochimica et biophysica acta**, v. 1824, n. 1, p. 195–206, jan 2012.
- AN, J.; TOTROV, M.; ABAGYAN, R. Comprehensive identification of “druggable” protein ligand binding sites. **Genome informatics. International Conference on Genome Informatics**, v. 15, n. 2, p. 31–41, jan 2004.
- APPLEBY, T. C.; ERION, M. D.; EALICK, S. E. The structure of human 5'-deoxy-5'-methylthioadenosine phosphorylase at 1.7 Å resolution provides insights into substrate binding and catalysis. **Structure (London, England : 1993)**, v. 7, n. 6, p. 629–41, 15 jun 1999.
- ASSENBERG, R.; MASTRANGELO, E.; WALTER, T. S. *et al.* Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein

- processing and viral replication. **Journal of virology**, v. 83, n. 24, p. 12895–906, dez 2009.
- ASSES, Y. Concepcion par modélisation et criblage in silico d'inhibiteurs du récepteur c-Met. **Tese de doutorado**. Nancy, França, outubro/2011.
- BACCHI CJ, SANABRIA K, SPIESS AJ, *et al.* In vivo efficacies of 5'-methylthioadenosine analogs as trypanocides. **Antimicrobial Agents Chemotherapy**, v.41, p.2108–2112, 1997.
- BAKAN, A.; MEIRELES, L. M.; BAHAR, I. ProDy: protein dynamics inferred from theory and experiments. **Bioinformatics (Oxford, England)**, v. 27, n. 11, p. 1575–7, 1 jun 2011.
- BASTOS, I. M. D.; GRELLIER, PHILIPPE; MARTINS, N. F. *et al.* Molecular, functional and structural properties of the prolyl oligopeptidase of *Trypanosoma cruzi* (POP Tc80), which is required for parasite entry into mammalian cells. **The Biochemical journal**, v. 388, n. Pt 1, p. 29–38, maio 2005.
- BEATTY, M. E.; STONE, A.; FITZSIMONS, D. W. *et al.* Best practices in dengue surveillance: a report from the Asia–Pacific and Americas Dengue Prevention Boards. **PLoS neglected tropical diseases**, v. 4, n. 11, p. e890, jan 2010.
- BERA, A. K.; KUHN, R. J.; SMITH, J. L. Functional characterization of cis and trans activity of the Flavivirus NS2B–NS3 protease. **The Journal of Biological Chemistry**, v. 282, n. 17, p. 12883–12892, 27 abr 2007.
- BERN, C.; KJOS, S.; YABSLEY, M. J.; MONTGOMERY, S. P. *Trypanosoma cruzi* and Chagas' Disease in the United States. **Clinical microbiology reviews**, v. 24, n. 4, p. 655–81, out 2011.

- BOND, C. S.; SCHÜTTELKOPF, A. W. ALINE: a WYSIWYG protein–sequence alignment editor for publication–quality alignments. **Acta crystallographica. Section D, Biological crystallography**, v. 65, n. Pt 5, p. 510–2, maio 2009.
- BOURI, N.; SELL, T. K.; FRANCO, C. *et al.* Return of epidemic dengue in the United States: implications for the public health practitioner. **Public health reports (Washington, D.C. : 1974)**, v. 127, n. 3, p. 259–66, maio 2012.
- BROOKS, B. R.; BROOKS, C. L.; MACKERELL, A. D.; *et al.* CHARMM: the biomolecular simulation program. **Journal of computational chemistry**, v. 30, n. 10, p. 1545–614, 30 jul 2009.
- BURKE, D. S.; NISALAK, A.; JOHNSON, D. E.; SCOTT, R. M. A prospective study of dengue infections in Bangkok. **The American journal of tropical medicine and hygiene**, v. 38, n. 1, p. 172–80, jan 1988.
- BURLEIGH, B. A. & ANDREWS, N. W. The mechanisms of *Trypanosoma cruzi* invasion of mammalian cells. **Annual review of microbiology**, v. 49, p. 175–200, jan 1995.
- BURLEIGH, B. A.; BARBARA, A.; ANDREWS, N. W.; NORMA W. Signaling and host cell invasion by *Trypanosoma cruzi*. **Current Opinion in Microbiology**, v. 1, n. 4, p. 461–465, ago 1998.
- BÖHM, H.-J.; FLOHR, A.; STAHL, M. Scaffold hopping. **Drug Discovery Today: Technologies**, v. 1, n. 3, p. 217–224, dez 2004.
- BRASIL. Secretaria de Vigilância em Saúde. Ministério da Saúde. Informe Epidemiológico da Dengue. Semanas de 1 a 52 de 2009. 2010a. Disponível em:

http://portal.saude.gov.br/portal/arquivos/pdf/informe_epidemiologico_semana_1a52_09_revisado.pdf Acesso em: agosto/2012

_____. Secretaria de Vigilância em Saúde. Ministério da Saúde. Informe Epidemiológico da Dengue. Análise de situação e tendências – 2010. 2010b. Disponível em:

http://portal.saude.gov.br/portal/arquivos/pdf/informe_da_dengue_ate_a_semana9.pdf

[Acesso em agosto/2012](#)

CADAVID-RESTREPO, G. E. Cinética de atividade e expressão da catepsina B e caracterização bioquímica de uma leucil aminopeptidase de *Trypanosoma cruzi*. **Tese de doutorado**. Universidade de Brasília, dez 2005.

CALDAS, I. S.; TALVANI, A.; CALDAS, S. *et al.* Benznidazole therapy during acute phase of Chagas disease reduces parasite load but does not prevent chronic cardiac lesions. **Parasitology research**, v. 103, n. 2, p. 413–21, jul 2008.

CALER, E. V.; VAENA DE AVALOS, S.; HAYNES, P. A.; ANDREWS, N. W.; BURLEIGH, B. A. Oligopeptidase B-dependent signaling mediates host cell invasion by *Trypanosoma cruzi*. **The EMBO journal**, v. 17, n. 17, p. 4975–86, 1 set 1998.

CAMINADE, C.; MEDLOCK, J. M.; DUCHEYNE, E. *et al.* Suitability of European climate for the Asian tiger mosquito *Aedes albopictus*: recent trends and future scenarios. **Journal of The Royal Society Interface**, 25 abr 2012.

CARDOSO, A. V. N.; LESCANO, S. A. Z.; AMATO NETO, V.; GAKIYA, É.; SANTOS, S. V. Survival of *Trypanosoma cruzi* in sugar cane used to prepare juice. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 48, n. 5, p. 287–289, out 2006.

- CASTRO, A. M.; VINAUD, M. C.; TEIXEIRA, A. R. L. C. Chagas disease: a global health problem. In: TEIXEIRA, A. R. L.; VINAUD, M. C.; CASTRO, A. M. (Eds.). **Emerging Chagas disease**. 1st. ed. Nova Iorque, NY: Bentham Science Publishers, 2009. p. 18–23.
- CASTRO, J. A.; MECCA, M. M. DE; BARTEL, L. C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). **Human & Experimental Toxicology**, v. 25, n. 8, p. 471–479, 1 ago 2006.
- CAVASOTTO, C. N. Homology models in docking and high-throughput docking. **Current topics in medicinal chemistry**, v. 11, n. 12, p. 1528–34, jan 2011.
- CAVASOTTO, C. N. & SINGH, N. Docking and High Throughput Docking: Successes and the Challenge of Protein Flexibility. **Current Computer – Aided Drug Design**, v. 4, n. 3, p. 221–234, 1 set 2008.
- CHAMBERS, T. J.; HAHN, C. S.; GALLER, R.; RICE, C. M. Flavivirus genome organization, expression, and replication. **Annual review of microbiology**, v. 44, p. 649–88, jan 1990.
- CHAMPREDA, V.; KHUMTHONG, R.; SUBSIN, B. *et al.* The two-component protease NS2B–NS3 of *Dengue virus* type 2: cloning, expression in *Escherichia coli* and purification of the NS2B, NS3 (pro) and NS2B–NS3 proteins. **Journal of Biochemistry and Molecular Biology**, v. 33, n. 4, p. 294–299, 2000.
- CHANDRAMOULI, S.; JOSEPH, J. S.; DAUDENARDE, S. *et al.* Serotype-specific structural differences in the protease-cofactor complexes of the *Dengue virus* family. **Journal of virology**, v. 84, n. 6, p. 3059–67, mar 2010.

- CHAPPELL, K. J.; STOERMER, M. J.; FAIRLIE, D. P.; YOUNG, P. R. Mutagenesis of the *West Nile virus* NS2B cofactor domain reveals two regions essential for protease activity. **The Journal of general virology**, v. 89, n. Pt 4, p. 1010-4, abr 2008.
- CHARIFSON, P. S.; CORKERY, J. J.; MURCKO, M. A.; WALTERS, W. P. Consensus scoring: A method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. **Journal of medicinal chemistry**, v. 42, n. 25, p. 5100-9, 16 dez 1999.
- CHENG, A. C.; COLEMAN, R. G.; SMYTH, K. T. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. **Nature biotechnology**, v. 25, n. 1, p. 71-5, jan 2007.
- CHIRAC, P. & TORRELE, E. Global framework on essential health R&D. **Lancet**, v. 367, n. 9522, p. 1560-1, 13 maio 2006.
- CHO, A. E.; GUALLAR, V.; BERNE, B. J.; FRIESNER, R. Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. **Journal of computational chemistry**, v. 26, n. 9, p. 915-31, 15 jul 2005.
- CHRISTEN, M.; HÜNENBERGER, P. H.; BAKOWIES, D. *et al.* The GROMOS software for biomolecular simulation: GROMOS05. **Journal of computational chemistry**, v. 26, n. 16, p. 1719-51, dez 2005.
- COURA, J. R. Chagas disease: what is known and what is needed –a background article. **Memorias do Instituto Oswaldo Cruz**, v. 102 Suppl , n. August, p. 113-122, 2007.
- COURA, J. R. & VIÑAS, P. A. Chagas disease: a new worldwide challenge. **Nature**, v. 465, n. 7301, p. S6-7, 24 jun 2010.

- CRAIG, I. R.; ESSEX, J. W.; SPIEGEL, K. Ensemble docking into multiple crystallographically derived protein structures: an evaluation based on the statistical analysis of enrichments. **Journal of chemical information and modeling**, v. 50, n. 4, p. 511–24, 26 abr. 2010.
- CREGAR–HERNANDEZ, L.; JIAO, G.–S.; JOHNSON, A. T. *et al.* Small molecule pan–dengue and *West Nile virus* NS3 protease inhibitors. **Antiviral chemistry & chemotherapy**, v. 21, n. 5, p. 209–17, jan 2011.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. **The Journal of Chemical Physics**, v. 98, n. 12, p. 10089, 1993.
- DAURA, X.; GADEMANN, K.; JAUN, B. *et al.* Peptide Folding: When Simulation Meets Experiment. **Angewandte Chemie International Edition**, v. 38, n. 1–2, p. 236–240, 15 jan 1999.
- DENG, J.; LI, N.; LIU, H. *et al.* Discovery of Novel Small Molecule Inhibitors of Dengue Viral NS2B–NS3 Protease Using Virtual Screening and Scaffold Hopping. **Journal of medicinal chemistry**, v. 55, n. 14, p. 6278–93 28 jun 2012..
- DIAS, J. C. P. Notas sobre o *Trypanosoma cruzi* e suas características bio–ecológicas, como agente de enfermidades transmitidas por alimentos. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 39, n. 4, p. 370–375, ago 2006.
- DIAS, L. C.; DESSOY, M. A.; SILVA, J. J. N. *et al.* Quimioterapia da doença de Chagas: estado da arte e perspectivas no desenvolvimento de novos fármacos. **Química Nova**, v. 32, n. 9, p. 2444–2457, 2009.

- DIOTAIUTI, L. Triatomine-vector of *Trypanosoma cruzi* infection. In: TEIXEIRA, A. R. L. C.; VINAUD, M. C.; CASTRO, A. M. (Eds.). **Emerging Chagas disease**. 1st. ed. Nova Iorque, NY: Bentham Science Publishers, 2009. p. 24-39.
- DOCAMPO, R.; & MORENO, S. N. J. The role of Ca²⁺ in the process of cell invasion by intracellular parasites. **Parasitology Today**, v. 12, n. 2, p. 61-65, fev 1996.
- DURBIN, A. P. & WHITEHEAD, S. S. Dengue vaccine candidates in development. **Current topics in microbiology and immunology**, v. 338, p. 129-43, jan 2010.
- DUSCHAK, V. G. & COUTO, A. S. An insight on targets and patented drugs for chemotherapy of Chagas disease. **Recent patents on anti-infective drug discovery**, v. 2, n. 1, p. 19-51, jan 2007.
- DVORAK, J. A. & HYDE, T. P. *Trypanosoma cruzi* Interaction with vertebrate cells *in vitro*. **Experimental Parasitology**, v. 34, n. 2, p. 268-283, out 1973.
- ECKERT, H. & BAJORATH, J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. **Drug discovery today**, v. 12, n. 5-6, p. 225-33, mar 2007.
- EGLOFF, M.-P.; BENARROCH, D.; SELISKO, B.; ROMETTE, J.-L.; CANARD, B. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. **The EMBO journal**, v. 21, n. 11, p. 2757-68, 3 jun 2002.
- EKONOMIUK, D.; SU, X.-C.; OZAWA, K. *et al.* Flaviviral protease inhibitors identified by fragment-based library docking into a structure generated by molecular dynamics. **Journal of medicinal chemistry**, v. 52, n. 15, p. 4860-8, 13 ago 2009.

- ERBEL, P.; SCHIERING, N.; D'ARCY, A. *et al.* Structural basis for the activation of flaviviral NS3 proteases from dengue and *West Nile virus*. **Nature structural & molecular biology**, v. 13, n. 4, p. 372–3, abr 2006.
- ESWAR, N.; WEBB, B.; MARTI-RENOM, M. A. *et al.* Comparative protein structure modeling using Modeller. **Current protocols in bioinformatics / editorial board, Andreas D. Baxevanis ... [et al.]**(não entendi isso), v. Chapter 5, p. Unit 5.6, out 2006.
- FALGOUT, B.; CHANOCK, R.; LAI, C. J. Proper processing of *Dengue virus* nonstructural glycoprotein NS1 requires the N-terminal hydrophobic signal sequence and the downstream nonstructural protein NS2a. **Journal of virology**, v. 63, n. 5, p. 1852–60, maio 1989.
- FALGOUT, B. & MARKOFF, L. Evidence that flavivirus NS1–NS2A cleavage is mediated by a membrane-bound host protease in the endoplasmic reticulum. **Journal of virology**, v. 69, n. 11, p. 7232–43, nov 1995.
- FALGOUT, B.; PETHEL, M.; ZHANG, Y. M.; LAI, C. J. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of *Dengue virus* nonstructural proteins. **Journal of virology**, v. 65, n. 5, p. 2467–75, maio 1991.
- FALGOUT, B.; MILLER, R. H.; LAI, C. J. Deletion analysis of *Dengue virus* type 4 nonstructural protein NS2B: identification of a domain required for NS2B–NS3 protease activity. **Journal of virology**, v. 67, n. 4, p. 2034–42, abr 1993.
- FELLER, S. E.; ZHANG, Y.; PASTOR, R. W.; BROOKS, B. R. Constant pressure molecular dynamics simulation: The Langevin piston method. **The Journal of Chemical Physics**, v. 103, n. 11, p. 4613, 1995.

- FERNANDES, M. C. & ANDREWS, N. W. Host cell invasion by *Trypanosoma cruzi*: a unique strategy that promotes persistence. **FEMS microbiology reviews**, v. 36, n. 3, p. 734–47, maio 2012.
- FERRARA, P.; GOHLKE, H.; PRICE, D. J.; KLEBE, G.; BROOKS, C. L. Assessing scoring functions for protein–ligand interactions. **Journal of medicinal chemistry**, v. 47, n. 12, p. 3032–47, 3 jun 2004.
- FISHMAN, M. C. & PORTER, J. A. Pharmaceuticals: a new grammar for drug discovery. **Nature**, v. 437, n. 7058, p. 491–3, 22 set 2005.
- FRANCO, C.; HYNES, N. A.; BOURI, N.; HENDERSON, D. A. The dengue threat to the United States. **Biosecurity and bioterrorism : biodefense strategy, practice, and science**, v. 8, n. 3, p. 273–6, set 2010.
- FREITAS, R. F.; PROKOPCZYK, I. M.; ZOTTIS, A. *et al.* Discovery of novel *Trypanosoma cruzi* glyceraldehyde–3–phosphate dehydrogenase inhibitors. **Bioorganic & medicinal chemistry**, v. 17, n. 6, p. 2476–82, 15 mar 2009.
- FRIMAYANTI, N.; CHEE, C. F.; ZAIN, S. M.; RAHMAN, N. A. Design of new competitive dengue ns2b/ns3 protease inhibitors—a computational approach. **International journal of molecular sciences**, v. 12, n. 2, p. 1089–100, jan 2011.
- FRIMAYANTI, N.; ZAIN, S. M.; LEE, V. S. *et al.* Fragment-based molecular design of new competitive dengue Den2 Ns2b/Ns3 inhibitors from the components of fingerroot (*Boesenbergia rotunda*). **In silico biology**, v. 11, n. 1–2, p. 29–37, 2012.

- GANESH, V. K.; MULLER, N.; JUDGE, K. *et al.* Identification and characterization of nonsubstrate based inhibitors of the essential *Dengue* and *West Nile virus* proteases. **Bioorganic & medicinal chemistry**, v. 13, n. 1, p. 257-64, 3 jan 2005.
- GARCIA, M. P.; NÓBREGA, O. T.; TEIXEIRA, A. R.; SOUSA, M. V.; SANTANA, J. M. Characterisation of a *Trypanosoma cruzi* acidic 30 kDa cysteine protease. **Molecular and biochemical parasitology**, v. 91, n. 2, p. 263-72, 15 mar 1998.
- GHODA, L. Y.; SAVARESE, T. M.; NORTHUP, C. H. *et al.* Substrate specificities of 5'-deoxy-5'(o código não aparece pra mim)-methylthioadenosine phosphorylase from *Trypanosoma brucei brucei* and mammalian cells. **Molecular and Biochemical Parasitology**, v. 27, n. 2-3, p. 109-118, jan 1988.
- GOUJON, M.; MCWILLIAM, H.; LI, W. *et al.* A new bioinformatics analysis tools framework at EMBL-EBI. **Nucleic acids research**, v. 38, n. Web Server issue, p. W695-9, jul 2010.
- GOUVEA, I. E.; IZIDORO, M. A.; JUDICE, W. A. S. *et al.* Substrate specificity of recombinant dengue 2 virus NS2B-NS3 protease: influence of natural and unnatural basic amino acids on hydrolysis of synthetic fluorescent substrates. **Archives of biochemistry and biophysics**, v. 457, n. 2, p. 187-96, 15 jan 2007.
- GRELLIER, P.; VENDEVILLE, S.; JOYEAU, R. *et al.* *Trypanosoma cruzi* prolyl oligopeptidase Tc80 is involved in nonphagocytic mammalian cell invasion by trypomastigotes. **The Journal of biological chemistry**, v. 276, n. 50, p. 47078-86, 14 dez 2001.
- GUBLER, D. J. Dengue, Urbanization and Globalization: The Unholy Trinity of the 21(st) Century. **Tropical medicine and health**, v. 39, n. 4 Suppl, p. 3-11, dez 2011.

- HAJDUK, P. J.; HUTH, J. R.; FESIK, S. W. Druggability indices for protein targets derived from NMR-based screening data. **Journal of medicinal chemistry**, v. 48, n. 7, p. 2518–25, 7 abr 2005.
- HALL, B. F.; WEBSTER, P.; MA, A. K.; JOINER, K. A.; ANDREWS, N. W. Desialylation of lysosomal membrane glycoproteins by *Trypanosoma cruzi*: a role for the surface neuraminidase in facilitating parasite entry into the host cell cytoplasm. **The Journal of Experimental Medicine**, v. 176, n. 2, p. 313–325, 1992.
- HALSEY, E. S.; MARKS, M. A.; GOTUZZO, E. *et al.* Correlation of Serotype-Specific *Dengue virus* Infection with Clinical Manifestations. **PLoS Neglected Tropical Diseases**, v. 6, n. 5, p. e1638, 1 maio 2012.
- HALSTEAD, S. B. Dengue. **Lancet**, v. 370, n. 9599, p. 1644–52, 10 nov 2007.
- HEBY, O.; PERSSON, L.; RENTALA, M. Targeting the polyamine biosynthetic enzymes: a promising approach to therapy of African sleeping sickness, Chagas' disease, and leishmaniasis. **Amino acids**, v. 33, n. 2, p. 359–66, ago 2007.
- HERRLING, P. Patent sense. **Nature**, v. 449, n. 7159, p. 174–5, 13 set 2007.
- HESS, B.; KUTZNER, C.; SPOEL, D. V. D.; LINDAHL, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 3, p. 435–447, 2008.
- HOPKINS, A. L. & GROOM, C. R. The druggable genome. **Nature reviews. Drug discovery**, v. 1, n. 9, p. 727–30, set 2002.
- HORI, H. & LAI, C. J. Cleavage of *Dengue virus* NS1–NS2A requires an octapeptide sequence at the C terminus of NS1. **Journal of virology**, v. 64, n. 9, p. 4573–7, set 1990.

- HOTEZ, P. J.; MOLYNEUX, D. H.; FENWICK, A. *et al.* Incorporating a rapid-impact package for neglected tropical diseases with programs for HIV/AIDS, tuberculosis, and malaria. **PLoS medicine**, v. 3, n. 5, p. e102, jan 2006.
- HOTEZ, P. J. The giant anteater in the room: Brazil's neglected tropical diseases problem. **PLoS neglected tropical diseases**, v. 2, n. 1, p. e177, jan 2008.
- HUANG, D. High-Throughput Virtual Screening Lead to Discovery of Non-Peptidic Inhibitors of *West Nile virus* NS3 Protease. In: BARON, R. (Ed.). **Methods**. New York, NY: Springer New York, 2012. v. 819p. 615-623.
- HUMPHREY, W. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33-38, fev 1996.
- JOHNSON, M. & MAGGIORA, G. **Concepts and applications of molecular similarity**. [S.l. s.n.], 1990. p. 393
- JONES, G.; WILLETT, P.; GLEN, R. C.; LEACH, A. R.; TAYLOR, R. Development and validation of a genetic algorithm for flexible docking. **Journal of molecular biology**, v. 267, n. 3, p. 727-48, 4 abr 1997.
- KARPLUS, M. & MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature structural biology**, v. 9, n. 9, p. 646-52, set 2002.
- KEARSLEY, S. K.; UNDERWOOD, D. J.; SHERIDAN, R. P.; MILLER, M. D. Flexibases: a way to enhance the use of molecular docking methods. **Journal of computer-aided molecular design**, v. 8, n. 5, p. 565-82, out 1994.
- KERR, I. D.; WU, P.; MARION-TSUKAMAKI, R.; MACKEY, Z. B.; BRINEN, L. S. Crystal Structures of TbCatB and rhodesain, potential chemotherapeutic targets and major

- cysteine proteases of *Trypanosoma brucei*. **PLoS neglected tropical diseases**, v. 4, n. 6, p. e701, jan 2010.
- KNEHANS, T.; SCHÜLLER, A.; DOAN, D. N. *et al.* Structure-guided fragment-based in silico drug design of dengue protease inhibitors. **Journal of computer-aided molecular design**, v. 25, n. 3, p. 263-74, mar 2011.
- KOEPPEN, H. & KRIEGL, J. Ligand-Based Virtual Screening. *In: Virtual Screening, Principles, Challenges and Practical Guidelines*. 2011.
- KONISHI, E. Issues related to recent dengue vaccine development. **Tropical medicine and health**, v. 39, n. 4 Suppl, p. 63-71, dez 2011.
- KOOPMANN, R.; CUPELLI, K.; REDECKE, L. *et al.* *In vivo* protein crystallization opens new routes in structural biology. **Nature methods**, v. 9, n. 3, p. 259-62, mar 2012.
- KUHN, R. J.; ZHANG, W.; ROSSMANN, M. G. *et al.* Structure of *Dengue virus* implications for flavivirus organization, maturation, and fusion. **Cell**, v. 108, n. 5, p. 717-25, 8 mar 2002.
- KUMAR, S.; MA, B.; TSAI, C. J.; SINHA, N.; NUSSINOV, R. Folding and binding cascades: dynamic landscapes and population shifts. **Protein science : a publication of the Protein Society**, v. 9, n. 1, p. 10-9, jan 2000.
- KUNTZ, I. D.; BLANEY, J. M.; OATLEY, S. J.; LANGRIDGE, R.; FERRIN, T. E. A geometric approach to macromolecule-ligand interactions. **Journal of molecular biology**, v. 161, n. 2, p. 269-88, 25 out 1982.
- LARKIN, M. A.; BLACKSHIELDS, G.; BROWN, N. P. *et al.* Clustal W and Clustal X version 2.0. **Bioinformatics (Oxford, England)**, v. 23, n. 21, p. 2947-8, 1 nov 2007.

- LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, v. 26, n. 2, p. 283–291, 1 abr 1993.
- LAW, R.; BARKER, O.; BARKER, J. J. *et al.* The multiple roles of computational chemistry in fragment-based drug design. **Journal of computer-aided molecular design**, p. 459–473, 17 jul 2009.
- LESCAR, J.; LUO, D.; XU, T. *et al.* Towards the design of antiviral inhibitors against flaviviruses: the case for the multifunctional NS3 protein from *Dengue virus* as a target. **Antiviral research**, v. 80, n. 2, p. 94–101, nov 2008.
- LEUNG, D.; SCHRODER, K.; WHITE, H. *et al.* Activity of recombinant dengue 2 virus NS3 protease in the presence of a truncated NS2B co-factor, small peptide substrates, and inhibitors. **The Journal of biological chemistry**, v. 276, n. 49, p. 45762–71, 7 dez 2001.
- LI, JUN; LIM, SIEW PHENG; BEER, D. *et al.* Functional profiling of recombinant NS3 proteases from all four serotypes of *Dengue virus* using tetrapeptide and octapeptide substrate libraries. **The Journal of biological chemistry**, v. 280, n. 31, p. 28766–74, 5 ago 2005.
- LINDOSO, J. A. L. & LINDOSO, A. A. B. P. Neglected tropical diseases in Brazil. **Revista Do Instituto De Medicina Tropical De São Paulo**, v. 51, n. 5, p. 247–253, 2009.
- LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 23, n. 1–3, p. 3–25, 13 jan 1997.

- LUO, D.; WEI, N.; DOAN, D. N. *et al.* Flexibility between the protease and helicase domains of the *Dengue virus* NS3 protein conferred by the linker region and its functional implications. **The Journal of biological chemistry**, v. 285, n. 24, p. 18817–27, 11 jun 2010.
- LUO, D.; XU, T.; HUNKE, C. *et al.* Crystal structure of the NS3 protease–helicase from *Dengue virus*. **Journal of virology**, v. 82, n. 1, p. 173–83, jan 2008.
- LUQUE, I. & FREIRE, E. Structural stability of binding sites: consequences for binding affinity and allosteric effects. **Proteins. Structure, Function, and Bioinformatics**. v. 41, n. S4, p. 63–71, set 2000.
- MACEDO, V. Forma indeterminada da doença de Chagas. **Jornal Brasileiro de Medicina**, v. 38, p. 34–40, 1980.
- MACKERELL, A. D. Empirical force fields for biological macromolecules: overview and issues. **Journal of computational chemistry**, v. 25, n. 13, p. 1584–604, out 2004.
- MCGOVERN, S. L. & SHOICHET, B. K. Information decay in molecular docking screens against holo, apo, and modeled conformations of enzymes. **Journal of medicinal chemistry**, v. 46, n. 14, p. 2895–907, 3 jul 2003.
- MCKERROW, J. H.; ROSENTHAL, P. J.; SWENERTON, R.; DOYLE, P. Development of protease inhibitors for protozoan infections. **Current opinion in infectious diseases**, v. 21, n. 6, p. 668–72, dez 2008.
- MEIERING, S.; INHOFF, O.; MIES, J. *et al.* Inhibitors of *Trypanosoma cruzi* trypanothione reductase revealed by virtual screening and parallel synthesis. **Journal of medicinal chemistry**, v. 48, n. 15, p. 4793–802, 28 jul 2005.

- MELINO, S. & PACI, M. Progress for *Dengue virus* diseases. Towards the NS2B-NS3pro inhibition for a therapeutic-based approach. **The FEBS journal**, v. 274, n. 12, p. 2986-3002, jun 2007.
- MOITESSIER, N.; ENGLEBIENNE, P.; LEE, D.; LAWANDI, J.; CORBEIL, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. **British journal of pharmacology**, v. 153 Suppl, p. S7-26, mar 2008.
- MOLINA, J. Cure of experimental Chagas' disease by the bis-triazole DO870 incorporated into "stealth" polyethyleneglycol-poly lactide nanospheres. **Journal of Antimicrobial Chemotherapy**, v. 47, n. 1, p. 101-104, 1 jan 2001.
- MONARD, G. Introduction à la Modélisation Moléculaire. **CNRS**. Formação contínua. (2003).
- MONCAYO, A. Chagas disease: current epidemiological trends after the interruption of vectorial and transfusional transmission in the Southern Cone countries. **Memórias do Instituto Oswaldo Cruz**, v. 98, n. 5, p. 577-591, jul 2003.
- MONCAYO, A. C.; FERNANDEZ, Z.; ORTIZ, D. *et al.* Dengue emergence and adaptation to peridomestic mosquitoes. **Emerging infectious diseases**, v. 10, n. 10, p. 1790-6, out 2004.
- MOREL, C. M. Inovação em saúde e doenças negligenciadas. **Cadernos de Saúde Pública**, v. 22, n. 8, p. 1522-1523, ago 2006.
- MORRIS, G. M.; GOODSSELL, D. S.; HALLIDAY, R. S. *et al.* Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **Journal of Computational Chemistry**, v. 19, n. 14, p. 1639-1662, 15 nov 1998.

- MOTTA, F. N.; BASTOS, I. M. D.; FAUDRY, E. *et al.* The *Trypanosoma cruzi* virulence factor oligopeptidase B (OPBTc) assembles into an active and stable dimer. **PloS one**, v. 7, n. 1, p. e30431, jan 2012.
- MÉDECINS SANS FRONTIÈRES. **Fatal imbalance: the crisis in research and development for drugs for neglected diseases**. Geneva. MSF. [S.l. s.n.], 2001a
- MÉDECINS SANS FRONTIÈRES. A matter of life and death: The role of patents in access to essential medicines. n. November, 2001b.
- NATARAJAN, S. NS3 protease from flavivirus as a target for designing antiviral inhibitors against *Dengue virus*. **Genetics and molecular biology**, v. 33, n. 2, p. 214-9, abr 2010.
- NERES, J.; BREWER, M. L.; RATIER, L. *et al.* Discovery of novel inhibitors of *Trypanosoma cruzi* trans-sialidase from in silico screening. **Bioorganic & medicinal chemistry letters**, v. 19, n. 3, p. 589-96, 1 fev 2009.
- NEVES, D. Metiltioadenosina Fosforilase De *Trypanosoma cruzi*, Um Alvo Potencial Para Quimioterapia Da Doença De Chagas, Apresenta Ampla Especificidade A Substratos E Elevada. **Tese de doutorado**. Universidade de Brasília, 2006.
- NITSCHKE, C.; BEHNAM, M. A. M.; STEUER, C.; KLEIN, C. D. Retro peptide-hybrids as selective inhibitors of the *Dengue virus* NS2B-NS3 protease. **Antiviral research**, v. 94, n. 1, p. 72-79, 26 mar 2012.
- NITSCHKE, C.; STEUER, C.; KLEIN, C. D. Arylcianoacrylamides as inhibitors of the *Dengue* and *West Nile virus* proteases. **Bioorganic & medicinal chemistry**, v. 19, n. 24, p. 7318-37, 15 dez 2011.

NOBLE, C. G.; SEH, C. C.; CHAO, A. T.; SHI, P. Y. Ligand-bound structures of the *Dengue virus* protease reveal the active conformation. **Journal of virology**, v. 86, n. 1, p. 438–46, jan 2012.

NÓBREGA, O. T.; SANTOS, S. M. A; TEIXEIRA, A. R.; SANTANA, J. M. Cloning and sequencing of tccb, a gene encoding a *Trypanosoma cruzi* cathepsin B-like protease. **Molecular and biochemical parasitology**, v. 97, n. 1–2, p. 235–40, 30 nov 1998.

ONUCHIC, J. N.; SOCCI, N. D.; LUTHEY-SCHULTEN, Z.; WOLYNES, P. G. Protein folding funnels: the nature of the transition state ensemble. **Folding and Design**, v. 1, n. 6, p. 441–450, dez 1996.

Organização Mundial de Saúde. Control of Chagas disease. Second report of a WHO expert committee. p. 905 2002.

_____. Reporte sobre la enfermedad de Chagas. Grupo de trabajo científico. 2005 -
Atualizado em 2007.

_____, Dengue and severe dengue. Fact Sheet N° 117. Janeiro/2012. 2012a.

<http://www.who.int/mediacentre/factsheets/fs117/en/> Acessado em: setembro/2012.

_____, Chagas disease (American trypanosomiasis). Fact Sheet N° 340. Agosto/2012.

2012b. <http://www.who.int/mediacentre/factsheets/fs340/en/> Acessado em:
setembro/2012.

OWENS, J. Determining druggability. **Nature Reviews Drug Discovery**, v. 6, n. 3, p. 187–187, mar 2007.

- PARODI-TALICE, A.; MONTEIRO-GOES, V.; ARRAMBIDE, N. *et al.* Proteomic analysis of metacyclic trypomastigotes undergoing *Trypanosoma cruzi* metacyclogenesis. **Journal of mass spectrometry, JMS**, v. 42, n. 11, p. 1422-32, nov 2007.
- PEARLMAN, D. A. & KOLLMAN, P. A. Evaluating the assumptions underlying force field development and application using free energy conformational maps for nucleosides. **Journal of the American Chemical Society**, v. 113, n. 19, p. 7167-7177, set 1991.
- PHILLIPS, J. C.; BRAUN, R.; WANG, W. *et al.* Scalable molecular dynamics with NAMD. **Journal of computational chemistry**, v. 26, n. 16, p. 1781-802, dez 2005.
- PISSURLENKAR, R. R. S.; SHAIKH, M. S.; IYER, R. P.; COUTINHO, E. C. Molecular Mechanics Force Fields and their Applications in Drug Design. **Anti-Infective Agents in Medicinal Chemistry (Formerly 'Current Medicinal Chemistry - Anti-Infective Agents)**, v. 8, n. 2, p. 128-150, 1 abr 2009.
- PIÑEYRO, M. D.; PARODI-TALICE, A.; PORTELA, M. *et al.* Molecular characterization and interactome analysis of *Trypanosoma cruzi* trypanredoxin 1. **Journal of proteomics**, v. 74, n. 9, p. 1683-92, 24 ago 2011.
- PRIETO, J. J.; TALEVI, A.; BRUNO-BLANCH, L. E. Application of linear discriminant analysis in the virtual screening of antichagasic drugs through trypanothione reductase inhibition. **Molecular diversity**, v. 10, n. 3, p. 361-75, ago 2006.
- PY, M. O. Neurologic manifestations of Chagas disease. **Current neurology and neuroscience reports**, v. 11, n. 6, p. 536-42, dez 2011.

- QUE, X.; NGO, H.; LAWTON, J. *et al.* The cathepsin B of *Toxoplasma gondii*, toxopain-1, is critical for parasite invasion and rhoptry protein processing. **The Journal of biological chemistry**, v. 277, n. 28, p. 25791-7, 12 jul 2002.
- RASSI, A. & MARIN-NETO, J. A. Chagas disease. **Lancet**, v. 375, n. 9723, p. 1388-402, 17 abr 2010.
- ROBIN, G.; CHAPPELL, K.; STOERMER, M. J.; *et al.* Structure of *West Nile virus* NS3 protease: ligand stabilization of the catalytic conformation. **Journal of molecular biology**, v. 385, n. 5, p. 1568-77, 6 fev 2009.
- RODENHUIS-ZYBERT, I. A.; WILSCHUT, J.; SMIT, J. M. *Dengue virus* life cycle: viral and host factors modulating infectivity. **Cellular and molecular life sciences: CMLS**, v. 67, n. 16, p. 2773-86, ago 2010.
- RODRIGUEZ, A. A trypanosome-soluble factor induces IP3 formation, intracellular Ca²⁺ mobilization and microfilament rearrangement in host cells. **The Journal of Cell Biology**, v. 129, n. 5, p. 1263-1273, 1 jun 1995.
- ROGERS, K. E.; KERÄNEN, H.; DURRANT, J. D. *et al.* Novel cruzain inhibitors for the treatment of Chagas' disease. **Chemical biology & drug design**, v. 80, n. 3, p. 398-405, set 2012.
- ROGNAN, D. Docking Methods for Virtual Screening: Principles and Recent Advances. *in* **Virtual Screening: Principles, Challenges, and Practical Guidelines**, 2011.
- RUCHE, G. L. A.; SOUARÈS, Y.; ARMENGAUD, A. *et al.* First two autochthonous *Dengue virus* infections in metropolitan France, September 2010. **Euro surveillance : bulletin**

- européen sur les maladies transmissibles – European communicable disease bulletin, v. 15, n. 39, p. 19676, 30 set 2010.
- RUSS, A. P.; LAMPEL, S. The druggable genome: an update. **Drug discovery today**, v. 10, n. 23-24, p. 1607-10, dez 2005.
- RYAN, M. D.; MONAGHAN, S.; FLINT, M. Virus-encoded proteinases of the Flaviviridae. **The Journal of general virology**, v. 79 (Pt 5), p. 947-59, maio 1998.
- SABCHAREON, A.; WALLACE, D.; SIRIVICHAYAKUL, C. *et al.* Protective efficacy of the recombinant, live-attenuated, CYD tetravalent dengue vaccine in Thai schoolchildren: a randomised, controlled phase 2b trial. **The Lancet**, v. 6736, n. 12, p. 1-9, set 2012.
- SADOWSKI, J.; GASTEIGER, J.; KLEBE, G. Comparison of Automatic Three-Dimensional Model Builders Using 639 X-ray Structures. **Journal of Chemical Information and Modeling**, v. 34, n. 4, p. 1000-1008, 1 jul 1994.
- SAMBROOK, J. & RUSSELL, D. **Molecular cloning: a laboratory manual**. 3rd editio ed. [S.l.]: Cold Spring Harbor Laboratory Press, 2001. p. 2344
- SANTANA, J.; GRELLIER, P.; SCHRÉVEL, J.; TEIXEIRA, A. R. L. A *Trypanosoma cruzi*-secreted 80 kDa proteinase with specificity for human collagen types I and IV. **Biochemical Journal**, v. 137, p. 129-137, 1997.
- SANTANA, J. M.; GRELLIER, P.; RODIER, M.-H.; SCHREVEL, J.; TEIXEIRA, A. Purification and characterization of a new 120 kDa alkaline proteinase of *Trypanosoma cruzi*. **Biochemical and Biophysical Research Communications**, v. 187, n. 3, p. 1466-1473, set 1992.

- SCHAMES, J. R.; HENCHMAN, R. H.; SIEGEL, J. S. *et al.* Discovery of a novel binding trench in HIV integrase. **Journal of medicinal chemistry**, v. 47, n. 8, p. 1879–81, 8 abr 2004.
- SCHENKMAN, S.; DIAZ, C.; NUSSENZWEIG, V. Attachment of *Trypanosoma cruzi* trypomastigotes to receptors at restricted cell surface domains. **Experimental Parasitology**, v. 72, n. 1, p. 76–86, jan 1991.
- SCHMIDTKE, P.; BIDON-CHANAL, A.; LUQUE, F. J.; BARRIL, X. MDpocket: open-source cavity detection and characterization on molecular dynamics trajectories. **Bioinformatics (Oxford, England)**, v. 27, n. 23, p. 3276–85, 1 dez 2011.
- SCHMUNIS, G. A. Epidemiology of Chagas disease in non endemic countries: the role of international migration. **Memórias do Instituto Oswaldo Cruz**, v. 102, p. 75–86, out 2007.
- SCHWARTZ, E. Seasonality, Annual Trends, and Characteristics of Dengue among Ill Returned Travelers, 1997~2006. **Emerging Infectious Diseases**, v. 14, n. 7, p. 1081–1088, jul 2008.
- SCHÜLLER, A.; YIN, Z.; BRIAN, C. C. S. *et al.* Tripeptide inhibitors of *Dengue* and *West Nile virus* NS2B-NS3 protease. **Antiviral research**, v. 92, n. 1, p. 96–101, out 2011.
- SHAN, L.; MATHEWS, I. I.; KHOSLA, C. Structural and mechanistic analysis of two prolyl endopeptidases: role of interdomain dynamics in catalysis and specificity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 10, p. 3599–604, 8 mar 2005.
- SHIRYAEV, S. A.; CHELTSOV, A. V.; GAWLIK, K.; RATNIKOV, B. I.; STRONGIN, A. Y. Virtual ligand screening of the National Cancer Institute (NCI) compound library leads to the

- allosteric inhibitory scaffolds of the *West Nile virus* NS3 proteinase. **Assay and drug development technologies**, v. 9, n. 1, p. 69–78, mar 2011.
- SHIRYAEV, S. A.; KOZLOV, I. A.; RATNIKOV, B. I. *et al.* Cleavage preference distinguishes the two-component NS2B–NS3 serine proteinases of *Dengue* and *West Nile viruses*. **The Biochemical journal**, v. 401, n. 3, p. 743–752, 2007.
- SIMMONS, C. P.; FARRAR, J. J.; NGUYEN, VAN V. C.; WILLS, B. *Dengue*. **The New England journal of medicine**, v. 366, n. 15, p. 1423–32, 12 abr 2012.
- SNOW, C. D.; SORIN, E. J.; RHEE, Y. M.; PANDE, V. S. How well can simulation predict protein folding kinetics and thermodynamics? **Annual review of biophysics and biomolecular structure**, v. 34, p. 43–69, jan 2005.
- SOEIRO, M. N. C.; CASTRO, S. L. DE. *Trypanosoma cruzi* targets for new chemotherapeutic approaches. **Expert opinion on therapeutic targets**, v. 13, n. 1, p. 105–21, jan 2009.
- SOUZA, W. DE. **Doenças negligenciadas**. Rio de Janeiro: Academia Brasileira de Ciências, 2010. p. 56
- SU, X.-C.; OZAWA, K.; QI, R. *et al.* NMR analysis of the dynamic exchange of the NS2B cofactor between open and closed conformations of the *West Nile virus* NS2B–NS3 protease. **PLoS neglected tropical diseases**, v. 3, n. 12, p. e561, jan 2009.
- TAN, B. H.; FU, J.; SUGRUE, R. J. *et al.* Recombinant dengue type 1 virus NS5 protein expressed in *Escherichia coli* exhibits RNA-dependent RNA polymerase activity. **Virology**, v. 216, n. 2, p. 317–25, 15 fev 1996.

- TARDIEUX, I.; WEBSTER, P.; RAVESLOOT, J. *et al.* Lysosome recruitment and fusion are early events required for trypanosome invasion of mammalian cells. **Cell**, v. 71, n. 7, p. 1117-1130, dez 1992.
- TEIXEIRA, A. R.; HECHT, M. M.; GUIMARO, M. C.; SOUSA, A. O.; NITZ, N. Pathogenesis of Chagas' disease: parasite persistence and autoimmunity. **Clinical microbiology reviews**, v. 24, n. 3, p. 592-630, jul 2011.
- TEIXEIRA, A. R.; NASCIMENTO, R. J.; STURM, N. R. Evolution and pathology in Chagas disease: a review. **Memórias do Instituto Oswaldo Cruz**, v. 101, n. 5, p. 463-491, ago 2006.
- THE ROCKEFELLER FOUNDATION. **Annual report**. . [S.l: s.n.] , 1978
- TIETZE, S. & APOSTOLAKIS, J. GlamDock: development and validation of a new docking tool on several thousand protein-ligand complexes. **Journal of chemical information and modeling**, v. 47, n. 4, p. 1657-72, 2007.
- TIEW, K.-C.; DOU, D.; TERAMOTO, T. *et al.* Inhibition of *Dengue virus* and *West Nile virus* proteases by click chemistry-derived benz[d]isothiazol-3(2H)-one derivatives. **Bioorganic & Medicinal Chemistry**, v. 20, n. 3, p. 1213-1221, fev 2012.
- TIMMERS, L. F. S. M.; PAULI, I.; BARCELLOS, G. B. *et al.* Genomic Databases and the Search of Protein Targets for Protozoan Parasites. **Current Drug Targets**, v. 10, n. 3, p. 240-245, 1 mar 2009.
- TOMLINSON, S. M.; MALMSTROM, R. D.; RUSSO, A. *et al.* Structure-based discovery of *Dengue virus* protease inhibitors. **Antiviral research**, v. 82, n. 3, p. 110-4, jun 2009.

- TOMLINSON, S. M. & WATOWICH, S. J. Anthracene-based inhibitors of *Dengue virus* NS2B-NS3 protease. **Antiviral research**, v. 89, n. 2, p. 127-35, fev 2011.
- TOMLINSON, S. M. & WATOWICH, S. J. Use of parallel validation high-throughput screens to reduce false positives and identify novel dengue NS2B-NS3 protease inhibitors. **Antiviral research**, v. 93, n. 2, p. 245-52, fev 2012.
- TROUILLER, P.; OLLIARO, P.; TORREELE, E. *et al.* Drug development for neglected diseases: a deficient market and a public-health policy failure. **Lancet**, v. 359, n. 9324, p. 2188-94, 22 jun 2002.
- TSANTRIZOS, Y. S. The design of a potent inhibitor of the hepatitis C virus NS3 protease: BILN 2061--from the NMR tube to the clinic. **Biopolymers**, v. 76, n. 4, p. 309-23, jan 2004.
- TUCCINARDI, T. Docking-based virtual screening: recent developments. **Combinatorial chemistry & high throughput screening**, v. 12, n. 3, p. 303-14, mar 2009.
- VASILAKIS, N. & WEAVER, S. C. **The history and evolution of human dengue emergence.** [S.l.]: Elsevier Inc., 2008. v. 72p. 1-76
- VENÄLÄINEN, J. I.; JUVONEN, R. O.; MÄNNISTÖ, P. T. Evolutionary relationships of the prolyl oligopeptidase family enzymes. **European journal of biochemistry / FEBS**, v. 271, n. 13, p. 2705-15, jul 2004.
- VERKHIVKER, G. M.; BOUZIDA, D.; GEHLHAAR, D. K. *et al.* Complexity and simplicity of ligand-macromolecule interactions: the energy landscape perspective. **Current Opinion in Structural Biology**, v. 12, n. 2, p. 197-203, abr 2002.

- VILAR, S.; FERINO, G.; PHATAK, S. S. *et al.* Docking-based virtual screening for ligands of G protein-coupled receptors: not only crystal structures but also in silico models. **Journal of molecular graphics & modelling**, v. 29, n. 5, p. 614-23, mar 2011.
- WALTERS, W. P. Going further than Lipinski's rule in drug design. **Expert opinion on drug discovery**, v. 7, n. 2, p. 99-107, mar 2012.
- WANG, J.; WOLF, R. M.; CALDWELL, J. W.; KOLLMAN, P. A.; CASE, D. A. Development and testing of a general amber force field. **Journal of computational chemistry**, v. 25, n. 9, p. 1157-74, 15 jul 2004.
- WENGLER, G. The carboxy-terminal part of the NS 3 protein of the *West Nile flavivirus* can be isolated as a soluble protein after proteolytic cleavage and represents an RNA-stimulated NTPase. **Virology**, v. 184, n. 2, p. 707-15, out 1991.
- WERNECK, G. L.; HASSELMANN, M. H.; GOUVÊA, T. G. Panorama dos estudos sobre nutrição e doenças negligenciadas no Brasil. **Ciência & Saúde Coletiva**, v. 16, n. 1, p. 39-62, jan 2011.
- WHITEHEAD, S. S.; BLANEY, J. E.; DURBIN, A. P.; MURPHY, B. R. Prospects for a *Dengue virus* vaccine. **Nature reviews. Microbiology**, v. 5, n. 7, p. 518-28, jul 2007.
- WICHAPONG, K.; PIANWANIT, S.; SIPPL, W.; KOKPOL, S. Homology modeling and molecular dynamics simulations of *Dengue virus* NS2B/NS3 protease: insight into molecular interaction. **Journal of molecular recognition ; JMR**, v. 23, n. 3, p. 283-300, 2010.
- YAMEY, G. The world's most neglected diseases. **Bmj**, v. 325, n. July, p. 176-177, 2002.

- YANG, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug discovery today**, v. 15, n. 11-12, p. 444-50, jun 2010.
- YOSHIDA, N. Molecular basis of mammalian cell invasion by *Trypanosoma cruzi*. **Anais Da Academia Brasileira De Ciencias**, v. 78, n. 1, p. 87-111, 2006.
- YOSHIDA, N. Molecular mechanisms of *Trypanosoma cruzi* infection by oral route. **Memórias do Instituto Oswaldo Cruz**, v. 104, p. 101-107, jul 2009.
- YOSHIDA, N.; CORTEZ, M. *Trypanosoma cruzi* parasite and host cell signaling during the invasion process. **Molecular Mechanisms of Parasite Invasion**, p. 82-91, 2008.
- ZELEDÓN, R.; RABINOVICH, J. E. Chagas' disease: an ecological appraisal with special emphasis on its insect vectors. **Annual review of entomology**, v. 26, p. 101-33, jan 1981.
- ZELLWEGER, R. M.; PRESTWOOD, T. R.; SHRESTA, S. Enhanced infection of liver sinusoidal endothelial cells in a mouse model of antibody-induced severe dengue disease. **Cell host & microbe**, v. 7, n. 2, p. 128-39, 18 fev 2010.
- ZHANG, M.-Q.; WILKINSON, B. Drug discovery beyond the "rule-of-five". **Current opinion in biotechnology**, v. 18, n. 6, p. 478-88, dez 2007.
- ZHANG, Z.; LI, Y.; LIN, B.; SCHROEDER, M.; HUANG, B. Identification of cavities on protein surface using multiple computational approaches for drug binding site prediction. **Bioinformatics (Oxford, England)**, v. 27, n. 15, p. 2083-8, 1 ago 2011.
- ZHOU, H.; SINGH, N. J.; KIM, K. S. Homology modeling and molecular dynamics study of *West Nile virus* NS3 protease: a molecular basis for the catalytic activity increased by the NS2B cofactor. **Proteins**, v. 65, n. 3, p. 692-701, 15 nov 2006.

ZHU, H.; BRIGGS, J. M. Mechanistic role of NS4A and substrate in the activation of HCV NS3 protease. **Proteins**, v. 79, n. 8, p. 2428-43, ago 2011.

ZINGALES, B.; ANDRADE, S.; BRIONES, M. *et al.* A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051-1054, nov 2009.

ZOETE, V.; GROSDIDIER, A.; MICHIELIN, O. Docking, virtual high throughput screening and in silico fragment-based drug design. **Journal of cellular and molecular medicine**, v. 13, n. 2, p. 238-48, fev 2009.

ZUCCOTTO, F.; ZVELEBIL, M.; BRUN, R. *et al.* Novel inhibitors of *Trypanosoma cruzi* dihydrofolate reductase. **European journal of medicinal chemistry**, v. 36, n. 5, p. 395-405, maio 2001.