

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA
Programa de Pós-Graduação em Química

ERIKA PICCIRILLO

**Busca virtual de inibidores de proteases dos
vírus da Dengue e da Febre Aftosa: construção
de bancos de dados, simulações de dinâmica
molecular e validação experimental.**

Versão corrigida da Tese conforme Resolução CoPGr 5890
O original se encontra disponível na Secretaria de Pós-Graduação do IQ-USP

São Paulo

Data do Depósito na SPG:
09/08/2017

ERIKA PICCIRILLO

**Busca virtual de inibidores de proteases dos
vírus da Dengue e da Febre Aftosa: construção
de bancos de dados, simulações de dinâmica
molecular e validação experimental.**

*Tese apresentada ao Instituto de Química da
Universidade de São Paulo para obtenção do
Título de Doutor em Química*

Orientador (a): Prof (a). Dr (a). Antonia Tavares do Amaral

São Paulo
2017

Ficha Catalográfica
Elaborada pela Divisão de Biblioteca e
Documentação do Conjunto das Químicas da USP

P587b PICCIRILLO, ERIKA
Busca virtual de inibidores de proteases dos
vírus da Dengue e da Febre Aftosa: construção de
bancos de dados, simulações de dinâmica molecular e
validação experimental / ERIKA PICCIRILLO. - São
Paulo, 2017.
405 p.

Tese (doutorado) - Instituto de Química da
Universidade de São Paulo. Departamento de Química
Fundamental.
Orientador: AMARAL, ANTONIA T.DO

1. Busca virtual. 2. Inibidores de protease. 3.
Dengue. 4. Febre Aftosa. 5. Banco de dados. I. T.
II. AMARAL, ANTONIA T.DO, orientador.

Folha de Aprovação

Para meus avós
Alberto G. Valente (in memoriam)
e Teresa S. Piccirillo (in memoriam).

AGRADECIMENTO(S)

Agradeço à Prof.^a Antonia Tavares do Amaral, que me aceitou em seu grupo, quando eu ainda estava no primeiro ano da graduação, e me deu a oportunidade e todas as condições para desenvolver esta tese.

Agradeço aos técnicos e amigos Leandro de Rezende e Leticia Marques Peron por toda a ajuda que me deram com os programas, scripts e análises estatísticas e quimiométricas, sem os quais a parte computacional desta tese não seria possível.

Agradeço aos colegas e amigos Lucas Viviane e Maria Cristiana Peters, pelas discussões científicas, pela agradável companhia em reuniões e congressos científicos e por todos os momentos de distração durante os últimos anos.

Agradeço aos colegas de laboratório Alberto Malvezzi e Noriberto Pradie, pelas discussões científicas e incentivos durante os anos de convivência.

Agradeço ao Prof. Dr. Luiz Juliano do Departamento de Biofísica, Universidade Federal do Estado de São Paulo-SP, por ceder amostras da enzima Lb^{pro} do vírus da Febre Aftosa, fundamentais para a validação experimental dos modelos propostos nesta tese.

Agradeço ao Prof. C. A. Sottriffer e ao Dr. Benjamin Merget, Universidade de Würzburg, Würzburg, Alemanha, por me receberem em seu laboratório durante três meses para realização dos estudos de Dinâmica Molecular apresentados nesta tese.

Agradeço à Prof.^a Wibke E. Diederich e ao aluno William Scanlan, Philipps-Universität de Marburg, Marburg, Alemanha, por me receberem em seu laboratório durante um mês para realização dos ensaios enzimáticos frente à NS2B/NS3^{pro} do vírus da Dengue, fundamentais para a validação experimental dos modelos propostos nesta tese.

Agradeço à Auxiliar Nê pela amizade e pelo enorme carinho durante os últimos 10 anos.

Agradeço à FAPESP que financiou esta tese e meu estágio de pesquisa no exterior (Nº Proc. 2012/06633-2 e 2014/01614-5) e agradeço às agências CAPES e CNPq (Nº Proc. 140997/2012-1) pelo suporte financeiro.

Agradeço à minha mãe Amelia, à minha irmã Rachel e ao meu pai Carlos que sempre me incentivaram, me deram todo seu amor e carinho e nunca deixaram de acreditar em mim.

Agradeço ao meu namorado Guilherme pelo amor, carinho, companhia e, principalmente, pela compreensão durante estes últimos meses.

Por fim, mas principalmente, agradeço a Deus por todas as portas que me abriu e por todas as pessoas que colocou em meu caminho, sem as quais eu jamais teria chegado até aqui.

***Bem aventurado o homem que
acha sabedoria, e o homem
que adquire conhecimento.***
(Provérbios, cap. 3, verso 13)

RESUMO

Piccirillo, E., **Busca virtual de inibidores de proteases dos vírus da Dengue e da Febre Aftosa: construção de bancos de dados, simulações de dinâmica molecular e validação experimental**, 405p. Tese - Programa de Pós-Graduação em Química. Instituto de Química, Universidade de São Paulo, São Paulo.

A Dengue e a Febre Aftosa são infecções virais que ocorrem tanto no Brasil como no mundo, apresentando enorme impacto socioeconômico. As proteases virais são reconhecidas como alvos de grande interesse para o planejamento de antivirais, devido ao seu papel fundamental no ciclo de vida de muitos vírus, incluindo-se os *flavivirus* (vírus da Dengue – DENV) e os *picornavirus* (vírus da Febre Aftosa – FA).

Com o objetivo de buscar e identificar novos inibidores de proteases virais (da NS2B/NS3pro do DENV ou da Lb^{pro} do vírus da FA) propusemos modelos de busca virtual, incluindo diferentes filtros de seleção (ex. farmacofórico, *drug-like*, similaridade e ancoramento) aplicados, sequencialmente, a bancos de dados de compostos comerciais (280×10^3 a 23×10^6 compostos). A seleção final dos compostos foi sempre feita por inspeção visual.

Para NS2B/NS3pro do DENV, construíram-se quatro modelos de busca virtual (Modelo I-DENV a Modelo IV-DENV). O primeiro foi construído, baseando-se na estrutura cristalográfica ligada a um inibidor peptidomimético, e aplicado ao banco *ZINC*. Ao final, dez compostos foram comprados e submetidos a testes de inibição enzimática, frente à NS2B/NS3pro, para validação experimental deste modelo. Dois compostos mostraram alguma atividade inibitória (IC_{50} 150 - 300 μ M).

Visando-se melhorar estes resultados, a flexibilidade da NS2B/NS3pro foi incluída, usando simulações de dinâmica molecular (DM), e um novo banco de dados construído (*ZINC-Curated*). Através de uma análise extensiva do banco *ZINC-Curated*, usando ferramentas estatísticas/quimiométricas, confirmou-se que este foi enriquecido com compostos com características de fármacos.

Outros três modelos de busca virtual foram construídos incluindo-se diferentes informações obtidas nas simulações de DM. O modelo II-DENV foi feito usando ancoramento e aplicado ao banco *ZINC-Curated*, selecionando dezesseis compostos. Nenhum deles apresentou atividade inibitória significativa frente à NS2B/NS3^{pro} do DENV. Os modelos III-DENV e IV-DENV utilizaram modelos farmacofóricos, que tiveram seus desempenhos previamente avaliados usando dados de literatura, e foram aplicados aos bancos *NCI* e *ZINC-Curated*, respectivamente. O modelo III-DENV selecionou quinze compostos, tendo quatro deles apresentado atividade inibitória (IC₅₀ 30 - 100 µM). O modelo IV-DENV selecionou dezoito compostos, sendo quatro ativos frente a esta protease (IC₅₀ 4 - 90 µM), representando uma taxa de acerto de ~22 %. Ainda, uma série de treze análogos estruturais do composto mais ativo foi construída, sendo três deles também ativos. Portanto, as modificações incluídas na busca virtual permitiram melhorar, significativamente, os resultados obtidos.

Para Lb^{pro} do vírus da FA, construíram-se dois modelos de busca virtual (Modelos I-FA e II-FA). O primeiro foi construído usando sua estrutura cristalográfica, sem ligantes, e uma série *in house* de potenciais inibidores covalentes. Seis compostos foram selecionados e testados frente à Lb^{pro}, tendo dois deles baixa atividade inibitória (IC₅₀ 300 - 600 µM).

A partir da disponibilidade da estrutura da Lb^{pro} com ligante, o modelo II-FA foi construído e aplicado ao banco *ZINC-Curated*, selecionando quinze compostos. Estes foram adquiridos e testados frente à Lb^{pro}, não apresentando atividade inibitória significativa. Assim, as modificações incluídas ainda não foram suficientes para selecionar inibidores mais potentes. No entanto, estes modelos/resultados contribuíram para o entendimento da(s) interação(ões) no sítio ativo da Lb^{pro}.

Palavras-chave: Busca virtual, Dengue, Febre Aftosa, Inibidores de Proteases, Dinâmica Molecular e Bancos de dados.

ABSTRACT

Piccirillo, E., **Virtual Screening for protease inhibitors of Dengue and Foot-and-mouth disease virus: database building, molecular dynamics simulations and experimental validation**. 2017, 405p, PhD Thesis - Graduate Program in Chemistry. Instituto de Química, Universidade de São Paulo, São Paulo.

Dengue and Food-and-mouth disease are viral infections that occur in Brazil and in the world, causing a huge socioeconomic impact. Viral proteases are recognized as targets for antiviral design, because they are crucial for the life cycle of many viruses, such as *flavivirus* (Dengue virus – DENV) and *picornavirus* (Food-and-mouth disease virus – FMDV).

In order to discovery novel inhibitors of viral proteases (of NS2B/NS3pro of DENV or of Lb^{pro} of FMDV) virtual screening models were proposed comprising a sequence of different filters (*e.g.* pharmacophore, drug-like, similarity and docking) applied to databases of commercial compounds (280×10^3 to 23×10^6 compounds). In all models, the final selection of compounds was always done by visual inspection.

For DENV NS2B/NS3pro, four virtual screening models were proposed (Model I-DENV to Model IV-DENV). Model I-DENV was built, based on the crystal structure bound to a peptidemimetic inhibitor, and applied to ZINC database. Finally, ten compounds were purchased and submitted to enzymatic assays against this protease to the experimental validation of this model. Two compounds showed some inhibitory activity (IC_{50} 150 - 300 μ M).

In order to improve these results, NS2B/NS3pro flexibility was included, using molecular dynamics (MD) simulations, and a novel database was built (ZINC-Curated). Throughout an exhaustively analysis of ZINC-Curated, using statistical/chemometrics tools, we confirmed that this new database was enriched with drug like compounds.

Other three virtual screening models were built including different information from MD simulations. Model II-DENV was built using docking and applied to the ZINC-Curated database, selecting sixteen compounds. None of them showed a significant inhibitory activity

against DENV NS2B/NS3pro. Models III-DENV and IV-DENV were built using pharmacophore models, which have their performance previously evaluated using literature data, and applied to NCI and ZINC-Curated databases, respectively. Model III-DENV selected fifteen compounds, showing four of them inhibitory activity (IC_{50} 30 - 100 μ M). Model IV-DENV selected eighteen compounds. Four of them were active against this protease (IC_{50} 4 - 90 μ M), representing a hit rate of ~22 %. Moreover, a set of thirteen structural analogues of the most active compound were built, being three of them also active. Thus, the modifications done in the virtual screening procedure really improved our results.

For FMDV Lb^{pro}, two virtual screening models were built (Models I-FMDV and II-FMDV). The first one was based on the crystal structure, without ligands, and used a set of *in house* potential covalent inhibitors. Six of the *in house* compounds were selected and tested against this protease. Two of them showed a weak inhibitory activity (IC_{50} 300 - 600 μ M).

Later on, the Lb^{pro} bound with ligands was available being therefore used to build another model. Model II – FMDV was applied to the ZINC-Curated database, selecting fifteen compounds that were purchased and also tested against the target protease. But none of them showed a significant inhibition. Thus, the incorporated changes were not enough to retrieve active compounds. However, these models/results contributed to better understand Lb^{pro} binding site interactions.

Keywords: *Virtual Screening*, Dengue, Foot-and-mouth disease, Protease inhibitors, Molecular Dynamics and database.

LISTA DE ABREVIATURAS E SIGLAS

A125C	NS2B/NS3pro mutante, com uma Alanina substituindo <i>Cys125</i>
ACP	Análise de Componentes Principais
ADMET	Absorção, Distribuição, Metabolismo, Excreção e Toxicidade
Bz-nKRR-H	benzoil-norleucina-lisina-arginina-arginina-aldeído
CP	Componente Principal
CTE	<i>C-terminal extension</i>
DENV	<i>Dengue virus</i>
DENV1	<i>Dengue virus serotype 1</i>
DENV2	<i>Dengue virus serotype 2</i>
DENV3	<i>Dengue virus serotype 3</i>
DENV4	<i>Dengue virus serotype 4</i>
DM	Dinâmica Molecular
DMSO	Dimetilsulfóxido
E-64	<i>Trans-epoxisuccinil-l-leucilamido-4-guanidino-butano</i>
eIF	Fator de iniciação eucariótico homólogo
FA	Febre Aftosa
<i>HIV</i>	<i>Human Immunodeficiency Virus</i>
<i>HTS</i>	<i>High-Throughput Screening</i>
IC ₅₀	Concentração que inibe 50 % da atividade enzimática
IFN β	Interferon beta
K _m	Constante de Michaelis-Menten
K _d	Constante de dissociação do complexo ligante-receptor
K _i	Constante de dissociação do complexo inibidor-enzima (constante de inibição)
<i>k_{on}</i>	Velocidade de associação do complexo ligante-receptor
<i>k_{off}</i>	Velocidade de dissociação do complexo ligante-receptor
<i>LBDD</i>	<i>Ligand Based Drug Design</i>
Lb ^{pro}	Protease Líder, do inglês <i>Leader protease</i>
LogD	Logaritmo do coeficiente aparente de partição em um dado pH
LogP	Logaritmo do coeficiente de partição
MC	Monte Carlo
NS2B	<i>Non-structural protein 2B</i>
NS2B/NS3pro	<i>Non-structural protein 2B and protease domain of non-structural protein 3</i>
NS3	<i>Non-structural protein 3</i>
<i>P₄, P₃, P₂, P₁, P₁'</i>	Resíduos dos substratos das proteases
pK _a	Cologaritmo da constante de acidez
<i>PAINS</i>	<i>Pan Assay INterference compoundS</i>
<i>PDB</i>	<i>Protein Data Bank</i>
<i>QSAR</i>	<i>Quantitative Structure-Activity Relationship</i>
RMN	Ressonância Magnética Nuclear

<i>RMSD</i>	<i>Root Mean Square Distance</i>
S_I', S_I, S_2, S_3	Cavidades do sítio ativo das proteases
<i>SAR</i>	<i>Structure Activity Relationship</i>
<i>SBDD</i>	<i>Structure Based Drug Design</i>
<i>SFI</i>	<i>Solubility Forecast Index</i>
v_o	Velocidade inicial de formação do produto
V_{max}	Velocidade máxima de formação do produto
Z-FR-MCA	Benzoil-fenilalanina-arginina-7-amido-4-metil-cumarina
Z-RR-MCA	Benzoil-arginina-arginina-7-amido-4-metil-cumarina
λ_{em}	Comprimento de onda de emissão
λ_{ex}	Comprimento de onda de excitação
λ^{max}	Comprimento de onda de máxima absorbância
ΔG	Variação da energia de Gibbs
ΔH	Variação da entalpia
ΔS	Variação da entropia

Nota para a nomenclatura de resíduos e átomos em proteínas

A nomenclatura usada para a identificação de resíduos de aminoácidos em proteínas será a nomenclatura *IUPAC*, de modo análogo ao descrito por Malvezzi (MALVEZZI, 2008). Assim, cada resíduo de aminoácido é descrito por três letras, seguido da posição do resíduo na sequência de aminoácido da estrutura primária da proteína. Por exemplo, o termo “*Ser135*”, representa o resíduo de aminoácido serina na posição 135.

A nomenclatura usada para a identificação de átomos em resíduos de aminoácidos será a derivada da nomenclatura *IUPAC*, em que o átomo de carbono entre o carbono carbonílico e o nitrogênio da amina é denominado *CA* (ou carbono- α). O próximo átomo ligado ao *CA* será denominado *CB* (ou carbono- β), o seguinte será denominado *CG* (ou carbono- γ) e assim por diante, seguindo o alfabeto grego. Após a indicação do átomo, será indicado o resíduo a que este átomo pertence, separado por um hífen. Por exemplo, o termo “*CA-His159*” representa o átomo de carbono- α do resíduo de aminoácido histidina 159.

SUMÁRIO

1. INTRODUÇÃO.....	18
1.1. A EVOLUÇÃO DA DESCOBERTA DE FÁRMACOS	18
1.2. QUÍMICA COMBINATÓRIA E <i>HTS</i> NA DESCOBERTA DE FÁRMACOS	23
1.3. RESULTADOS FALSO-POSITIVOS	25
1.3.1. <i>Frequent hitters e PAINS</i>	26
1.3.2. <i>Compostos com grupos reativos</i>	28
1.3.3. <i>Compostos capazes de formar agregados</i>	29
1.4. BUSCA VIRTUAL.....	32
1.5. INTERAÇÃO LIGANTE-RECEPTOR.....	34
1.5.1. <i>Descrição termodinâmica da interação ligante-receptor</i>	35
1.5.2. <i>Interações não covalentes do complexo ligante-receptor</i>	39
1.5.3. <i>Predição da afinidade ligante-receptor</i>	40
1.5.4. <i>Descrição cinética da interação ligante-receptor</i>	41
1.5.5. <i>Medidas da eficiência da ligação</i>	42
1.6. FILTROS UNI- OU BIDIMENSIONAIS	44
1.6.1. <i>Propriedades drug-like</i>	44
1.6.2. <i>Propriedades estruturais</i>	49
1.6.3. <i>Similaridade 2D</i>	50
1.7. FILTROS TRIDIMENSIONAIS	52
1.7.1. <i>Filtros farmacofóricos</i>	52
1.7.2. <i>Filtros baseados no formato (“shape”)</i>	59
1.8. FILTROS DE ANCORAMENTO.....	60
1.9. BANCOS DE DADOS DE MOLÉCULAS VIRTUAIS.....	67
1.10. SIMULAÇÕES DE DINÂMICA MOLECULAR APLICADAS EM <i>SBDD</i>	71
1.11. DENGUE E FEBRE AFTOSA	76
1.11.1. <i>Dengue</i>	77
1.11.1.1. <i>Doença e Epidemiologia</i>	77
1.11.1.2. <i>Vacina e Tratamento</i>	78

1.11.1.3.	Busca racional de antivirais contra o DENV – a escolha do alvo	79
1.11.1.4.	Protease NS2B/NS3 do DENV	81
1.11.2.	<i>Febre Aftosa</i>	85
1.11.2.1.	Doença e Epidemiologia	85
1.11.2.2.	Vacinas e Tratamentos	87
1.11.2.3.	Busca racional de antivirais contra o vírus FA.....	87
1.11.2.4.	L ^{pro} do vírus da FA.....	88
2.	OBJETIVOS.....	92
3.	MATERIAIS E MÉTODOS	94
3.1.	REAGENTES E SOLVENTES	94
3.1.1.	<i>Ensaio de inibição</i>	94
3.1.2.	<i>Enzimas</i>	94
3.1.3.	<i>Substratos</i>	94
3.1.4.	<i>Compostos testados frente à NS2B/NS3^{pro} do DENV</i>	95
3.1.5.	<i>Compostos testados frente à Lb^{pro} do vírus de FA</i>	98
3.2.	EQUIPAMENTOS	99
3.2.1.	<i>Espectrofotômetro UV-VIS</i>	99
3.2.2.	<i>Fluorímetro e leitor de microplacas</i>	99
3.2.3.	<i>pH-âmetro</i>	99
3.2.4.	<i>Espectrômetro de Ressonância Magnética Nuclear (RMN)</i>	99
3.2.5.	<i>Estações de Trabalho</i>	100
3.3.	BANCO DE DADOS DE MOLÉCULAS VIRTUAIS, SÉRIES DE COMPOSTOS E ESTRUTURAS TRIDIMENSIONAIS DAS PROTEASES VIRAIS.....	101
3.3.1.	<i>Banco de dados de moléculas virtuais utilizados na busca virtual</i>	101
3.3.2.	<i>Bancos de dados de moléculas virtuais utilizados na análise da diversidade química e estrutural</i>	101
3.3.3.	<i>Séries de inibidores conhecidos da NS2B/NS3^{pro} do DENV</i>	102
3.3.4.	<i>Série “in house” de potenciais inibidores covalentes</i>	102
3.3.5.	<i>Estruturas tridimensionais das proteases virais</i>	102
3.4.	PROGRAMAS COMPUTACIONAIS.....	103

3.5. PROCEDIMENTOS COMPUTACIONAIS	106
3.5.1. <i>Construção dos bancos de dados virtuais de moléculas</i>	106
3.5.1.1. Busca virtual	106
3.5.1.2. Análise da diversidade química e estrutural do banco <i>ZINC-Curated</i>	107
3.5.2. <i>Preparação das séries de compostos</i>	108
3.5.2.1. Séries de inibidores conhecidos da NS2B/NS3pro do DENV.....	108
3.5.2.2. Série “ <i>in house</i> ” de potenciais inibidores covalentes.....	110
3.5.3. <i>Análise da diversidade química e estrutural do banco ZINC-Curated</i>	110
3.5.3.1. Descritores moleculares	110
3.5.3.2. Validação da etapa de amostragem	110
3.5.3.3. Propriedades físico-químicas e estruturais.....	111
3.5.3.4. Espaço químico.....	112
3.5.3.5. Similaridade estrutural	113
3.5.4. <i>Busca virtual de inibidores da protease do vírus da Dengue</i>	114
3.5.4.1. Preparação da estrutura da protease do vírus da Dengue.	114
3.5.4.2. Construção e aplicação de modelos farmacofóricos	115
3.5.4.3. Filtro “ <i>drug-like</i> ”	116
3.5.4.4. Filtros de Similaridade.....	116
3.5.4.5. Filtro de “grupos funcionais”	117
3.5.4.6. Ancoramento utilizando o programa <i>GOLD</i>	118
3.5.4.7. Inspeção visual.....	119
3.5.4.8. Estudos da flexibilidade conformacional da NS2B/NS3pro do DENV por simulações de Dinâmica Molecular (DM).....	120
3.5.4.9. Modelos de busca virtual: Modelo I - DENV	126
3.5.4.10. Modelos de busca virtual: Modelo II - DENV	129
3.5.4.11. Modelos de busca virtual: Modelo III - DENV.....	131
3.5.4.12. Modelos de busca virtual: Modelo IV - DENV	134
3.5.4.13. Construção e análise de série de análogos ao composto mais ativo dentre os seleccionados pelo Modelo IV – DENV	137
3.5.5. <i>Busca virtual de inibidores da protease do vírus da Febre Aftosa</i>	137
3.5.5.1. Preparação da estrutura da Lb ^{pro} do vírus da FA sem ligantes.	138

3.5.5.2.	Preparação da estrutura da Lb ^{pro} do vírus da FA complexada com o inibidor E64-R-P-NH ₂	138
3.5.5.3.	Ancoramento utilizando o programa <i>GOLD</i>	139
3.5.5.4.	Cálculo dos campos de interação molecular e identificação de regiões de mínimo de energia.....	139
3.5.5.5.	Construção e aplicação do modelo farmacofórico	140
3.5.5.6.	Filtro de similaridade	141
3.5.5.7.	Modelos de busca virtual: Modelo I – FA.....	141
3.5.5.8.	Modelos de busca virtual: Modelo II - FA.....	142
3.6.	PROCEDIMENTOS EXPERIMENTAIS	144
3.6.1.	<i>Testes de inibição da NS2B/NS3pro do DENV</i>	144
3.6.1.1.	Determinação dos valores de porcentagem de inibição.....	145
3.6.1.2.	Determinação dos valores de IC ₅₀	146
3.6.1.3.	Caracterização do mecanismo de inibição e determinação do valor(es) da constante de inibição (K _i).....	146
3.6.2.	<i>Testes de inibição da sLb^{pro} do vírus da FA</i>	147
3.6.2.1.	Determinação dos valores de IC ₅₀ para os compostos selecionados pelo Modelo I – FA	148
3.6.2.2.	Avaliação da reversibilidade da inibição para os compostos selecionados pelo Modelo I – FA.....	149
3.6.2.3.	Determinação dos valores de porcentagem de inibição para os compostos selecionados pelo Modelo II – FA	150
3.6.3.	<i>Avaliação da estabilidade de cada um dos inibidores nas condições do experimento de inibição</i>	151
3.6.4.	<i>Identificação e/ou caracterização dos compostos adquiridos</i>	152
4.	RESULTADOS E DISCUSSÃO	154
4.1.	INTRODUÇÃO	154
4.2.	BANCO DE DADOS DE MOLÉCULAS VIRTUAIS	156
4.2.1.	<i>Construção dos bancos de dados de moléculas virtuais</i>	156
4.2.2.	<i>Análise da diversidade das propriedades físico-químicas e/ou estruturais no banco ZINC-Curated</i>	159
4.2.2.1.	Validação estatística da etapa de amostragem	162
4.2.2.2.	Distribuição de algumas propriedades físico-químicas e/ou estruturais.....	170
4.2.2.3.	Espaço químico.....	185
4.2.2.4.	Análise da solubilidade em água predita.....	191
4.2.2.5.	Análise da estabilidade metabólica predita	193
4.2.2.6.	Diversidade estrutural	194

4.3. BUSCA VIRTUAL DE INIBIDORES DA PROTEASE DO VÍRUS DA DENGUE	199
4.3.1. <i>Reconhecimento da estrutura e características do sítio ativo da protease do vírus da Dengue para escolha da estrutura 3D</i>	199
4.3.1.1. Reconhecimento da estrutura da proteína por inspeção visual	205
4.3.1.2. Reconhecimento da estrutura da proteína auxiliado por computador.....	208
4.3.1.3. Escolha da conformação do receptor e considerações sobre sua flexibilidade conformacional	213
4.3.2. <i>Modelos de busca virtual: Modelo I - DENV</i>	215
4.3.2.1. Filtro farmacofórico	216
4.3.2.2. Filtro “drug-like”	218
4.3.2.3. Filtro de Similaridade	220
4.3.2.4. Ancoramento.....	223
4.3.2.5. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV225	
4.3.2.6. Validação experimental	227
4.3.3. <i>Estudos da flexibilidade conformacional da NS2B/NS3pro do DENV por simulações de Dinâmica Molecular (DM)</i>	234
4.3.3.1. Agrupamento hierárquico dos <i>snapshots</i> das simulações de DM e abordagem <i>GRID/PCA</i>	236
4.3.3.2. Identificação das interações moleculares conservadas durante as simulações de DM	240
4.3.3.3. Identificação dos modos de ligação do inibidor Bz-nKRR-H.....	243
4.3.4. <i>Modelos de busca virtual: Modelo II - DENV</i>	245
4.3.4.1. Ancoramento.....	247
4.3.4.2. Filtro de “grupos funcionais”	249
4.3.4.3. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV250	
4.3.4.4. Validação experimental	253
4.3.5. <i>Modelos de busca virtual: Modelo III - DENV</i>	257
4.3.5.1. Filtro farmacofórico.....	259
4.3.5.2. Filtro de “grupos funcionais”	270
4.3.5.3. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV271	
4.3.5.4. Validação experimental	274
4.3.6. <i>Modelos de busca virtual: Modelo IV - DENV</i>	281
4.3.6.1. Filtro farmacofórico.....	283
4.3.6.2. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV297	

4.3.6.3.	Validação experimental	300
4.3.7.	<i>Série de análogos ao composto mais ativo dentre os selecionados pelo quarto modelo de busca virtual (Modelo IV – DENV).....</i>	<i>311</i>
4.3.7.1.	Construção da série de análogos	311
4.3.7.2.	Validação experimental	314
4.3.7.3.	Análise preliminar da Relação Estrutura Atividade	322
4.4.	BUSCA VIRTUAL DE INIBIDORES DA PROTEASE DO VÍRUS DA FEBRE AFTOSA	322
4.4.1.	<i>Reconhecimento da estrutura e características do sítio ativo da protease líder do vírus da Febre Aftosa para escolha da estrutura 3D</i>	<i>323</i>
4.4.1.1.	Reconhecimento da estrutura da proteína por inspeção visual	326
4.4.1.2.	Reconhecimento da estrutura da proteína auxiliado por computador.....	329
4.4.1.3.	Escolha da conformação do receptor e considerações sobre sua flexibilidade conformacional	334
4.4.2.	<i>Modelos de busca virtual: Modelo I – FA</i>	<i>334</i>
4.4.2.1.	Ancoramento seguido por Inspeção Visual.....	336
4.4.2.2.	Validação experimental	337
4.4.2.3.	Análise das interações dos inibidores compostos FA-04 e FA-05 com resíduos do sítio ativo da Lb ^{pro} do vírus da FA	341
4.4.3.	<i>Modelos de busca virtual: Modelo II - FA.....</i>	<i>343</i>
4.4.3.1.	Filtro farmacofórico	344
4.4.3.2.	Filtro de similaridade	347
4.4.3.3.	Filtro de inspeção visual	347
4.4.3.4.	Validação experimental	349
5.	CONCLUSÕES.....	353
6.	REFERÊNCIAS.....	356
	ANEXO I.....	368
	ANEXO II	372
	ANEXO III.....	376
	ANEXO IV – PUBLICAÇÃO	380
	SÚMULA CURRICULAR.....	402

1. INTRODUÇÃO

1.1. A evolução da descoberta de fármacos

Os fármacos vêm sendo utilizados, há milhares de anos para os mais diversos fins terapêuticos. Embora não se saiba ao certo quais foram os primeiros fármacos e como estes foram descobertos, existem relatos da utilização de compostos com atividade biológica como medicamentos desde as civilizações antigas (Egípcia, Greco-romana e Chinesa) (MALVEZZI, 2008; RAVINA, 2011; VIEGAS; DA SILVA BOLZANI; BARREIRO, 2006).

Até a metade do século XX, os medicamentos usados pelo homem eram exclusivamente de origem natural (animal, vegetal e/ou mineral) (FALLER et al., 2011; KUBINYI, 2004a). Estes fármacos eram, principalmente, substâncias isoladas de produtos naturais de plantas e metabólitos secundários de microrganismos (FALLER et al., 2011; VIEGAS; DA SILVA BOLZANI; BARREIRO, 2006). Dentre estes, podemos citar: a morfina (extraída da papoula, *Papaver somniferum*), a quinina e quinidina (ambas extraídas de espécies de *Cinchona*) e a penicilina (extraída do fungo *Penicillium sp.*), Figura 1.1. Neste período, a descoberta de fármacos era feita por “tentativa e erro”, em geral, testando-se aleatoriamente diferentes produtos naturais contra as mais diversas doenças e observando sua atividade. Não raro, na literatura, encontram-se compostos descobertos ao acaso “baseados” simplesmente na dedicação/intuição/sorte, como, por exemplo, a penicilina, o clordiazepóxido, a sacarina, a fenoltaleína, entre outros. Com o avanço das ciências naturais, em especial a química e a biologia, o homem passou a estudar sistematicamente as plantas de uso medicinal. (KUBINYI, 1999). Neste contexto, as substâncias isoladas de plantas começavam a ser modificadas, de forma mais ou menos sistemática, em laboratório, visando conseguir compostos com atividade otimizada, ou seja, mais ativos frente ao seu alvo biológico e com menor toxicidade (CUMMING et al., 2013; KUBINYI, 2003; MIGNANI et

al., 2016; ROLLINGER; STUPPNER; LANGER, 2008). O primeiro exemplo de fármaco semissintético baseado em uma substância isolada de planta é o ácido acetilsalicílico (Aspirina®), sintetizado por Felix Hoffmann, em 1897 (Figura 1.1). Este fármaco é uma forma menos tóxica do ácido salicílico, que é o principal metabólito da salicíina (princípio ativo encontrado na casca do salgueiro, *Salix alba*). O ácido acetilsalicílico foi introduzido no mercado em 1899, sendo usado até os dias de hoje não apenas como analgésico e anti-inflamatório, mas também como um inibidor da agregação plaquetária (KUBINYI, 2006; RAVINA, 2011).

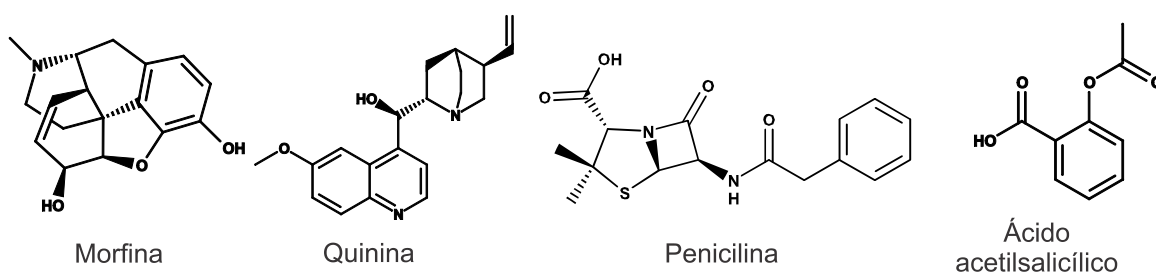


Figura 1.1. Representação da estrutura 2D de alguns fármacos extraídos de produtos naturais (morfina, quinina, penicilina) e do primeiro fármaco semissintético (ácido acetilsalicílico).

A descoberta do ácido acetilsalicílico deu início à era dos fármacos sintéticos (VIEGAS; DA SILVA BOLZANI; BARREIRO, 2006). Assim, novos compostos eram sintetizados e testados em modelos animais e/ou preparações de órgãos. Nesta fase, o processo de desenvolvimento era guiado por hipóteses química e/ou biológica, sendo limitante tanto a etapa de síntese de novos compostos quanto à avaliação de sua atividade biológica.

No começo da década de 1960 (HANSCH; STEWARD, 1964), com o início dos estudos de *QSAR* (do inglês *Quantitative Structure-Activity Relationships* - Relações Quantitativas entre Estrutura química e Atividade biológica), pela primeira vez os conceitos de físico-química orgânica foram usados para explicar a atividade biológica. Estes estudos foram iniciados por Corwin Hansh, que considerava que a ação de um fármaco dependia de dois processos: (i) transporte do fármaco do seu local de aplicação até seu local de ação e (ii)

interações não covalentes entre o fármaco e seu sítio de ligação no receptor alvo (FUJITA, 2011; HANSCH; STEWARD, 1964; KUBINYI, 2007). Nesta tese, o termo “ligante” será usado indiscriminadamente para se referir a qualquer molécula (agonista, antagonista, substrato ou inibidor) que se liga a uma macromolécula biológica (proteína, receptor de membrana, DNA ou RNA). Até os dias de hoje, os modelos de *QSAR* são bastante usados no planejamento de fármacos bem como na predição de algumas de suas propriedades (FUJITA; WINKLER, 2016). Mais tarde, criou-se o termo planejamento de fármacos baseados na estrutura do ligante (*LBDD*, do inglês *Ligand Based Drug Design*), que inclui, entre outras abordagens, os estudos de *QSAR*.

A partir desta abordagem, a síntese de novos compostos passou a ser guiada por modelos de *QSAR*, sendo os testes biológicos uma etapa ainda bastante limitante. Esta limitação só foi superada com a disponibilidade dos testes *in vitro* usando, por exemplo, ensaio de inibição enzimática. Os testes *in vitro* permitiram não só testar de forma mais rápida os análogos sintetizados, mas também aumentaram a confiabilidade dos resultados obtidos (KUBINYI, 2004b). Subsequentemente, o aparecimento da Química Combinatória, da triagem de alta produtividade (*HTS*, do inglês *High-Throughput Screening*) e a evolução da engenharia genética permitiram a síntese de grandes bibliotecas de compostos, que passaram a ser testados frente a seus alvos biológicos, num período de poucas semanas (MAYR; BOJANIC, 2009).

No entanto, embora a Química Combinatória e o *HTS* tenham gerado um vasto número de compostos com diversidade estrutural, possibilitando a seleção de ligantes que apresentassem um ótimo ajuste a um determinado alvo biológico, é consenso na literatura (BEHNEN et al., 2012; KLEBE, 2006; MIGNANI et al., 2016) que estas técnicas não levaram necessariamente à descoberta de novos fármacos, alertando-se, assim, para a necessidade de se considerar, o mais cedo possível, outras características do ligante que o

tornam fármaco (GLEN, 2011; HOPKINS; GROOM, 2002; MIGNANI et al., 2016). Isto significa que, além do ótimo ajuste do ligante a um determinado alvo biológico, deve-se considerar sua absorção, distribuição, metabolismo e eliminação no sistema biológico e sua toxicidade, os chamados parâmetros ADME ou ADMET do ligante. Neste contexto, a evolução dos métodos quimiométricos e a análise das propriedades físico-químicas e estruturais dos ligantes desempenharam um papel fundamental (BOLOGA; OPREA, 2012; CUMMING et al., 2013; KIRCHMAIR et al., 2015; KUBINYI, 2004b; LIPINSKI et al., 2001; MEANWELL, 2011; MIGNANI et al., 2016).

A revolução nas estratégias de descoberta e planejamento de fármacos se beneficiou grandemente do desenvolvimento da cristalografia de raios-X de proteínas e de técnicas multidimensionais em Ressonância Magnética Nuclear (RMN) (CHEN et al., 2016; KIRCHMAIR et al., 2008; KLEBE, 2006; KUBINYI, 2007; ROLLINGER; STUPPNER; LANGER, 2008). Estas técnicas permitiram elucidar grande número de estruturas tridimensionais (3D) de potenciais alvos biológicos, sendo grande parte deles disponíveis no *Protein Data Bank (PDB)*. O conhecimento destas estruturas 3D permitiu avanços consideráveis no planejamento de fármacos, sendo possível incluir informações sobre a estrutura 3D dos alvos biológicos. As estratégias de planejamento de fármacos baseadas na estrutura do alvo biológico são denominadas *SBDD* (do inglês *Structure Based Drug Design*).

O primeiro fármaco descoberto usando estratégias *SBDD* é o captopril (Figura 1.2), inibidor da enzima conversora de angiotensina (ECA), descoberto em 1977. A busca de inibidores da ECA como fármacos anti-hipertensivos se iniciou em 1965, com o pesquisador brasileiro Dr. Sergio H. Ferreira, a partir do estudo dos efeitos de peptídeos extraídos de veneno de cobra no sistema renina-angiotensina-aldosterona (RAVINA, 2011). No entanto, os resultados mais promissores só foram obtidos a partir da elucidação da estrutura 3D da carboxipeptidase A (enzima similar à ECA) complexada o ácido L-benzil succínico. Este

complexo foi usado como um modelo para o sítio ativo da ECA, levando à descoberta, numa primeira etapa, do inibidor *N*-succinil-L-prolina ($IC_{50} = 330 \mu M$) e, em seguida, do captopril ($IC_{50} = 23 \text{ nM}$) (KUBINYI, 2006). Ainda, no final dos anos 1990, foi descoberta a dorzolamida (Figura 1.2), inibidor da enzima anidrase carbônica, usada no tratamento do glaucoma. A dorzolamida é o primeiro fármaco desenvolvido a partir da elucidação da estrutura 3D de seu alvo biológico, a anidrase carbônica (KUBINYI, 2006; RONDEAU; SCHREUDER, 2008).

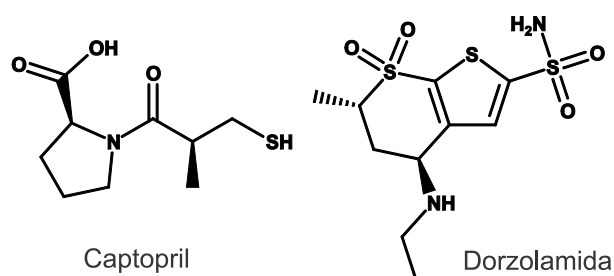


Figura 1.2. Representação da estrutura 2D dos primeiros fármacos desenvolvidos usando estratégias *SBDD*, sendo estes o captopril e a dorzolamida.

Mais recentemente, além de se utilizar a estrutura 3D do alvo biológico, a flexibilidade deste alvo vem sendo cada vez mais incluída nas estratégias de *SBDD*. A importância da flexibilidade conformacional para a interação ligante-receptor é amplamente aceita na literatura. No entanto, somente com os avanços computacionais obtidos nas últimas décadas, a inclusão desta flexibilidade nos protocolos de *SBDD* se tornou factível, através, por exemplo, de simulações de dinâmica molecular (DM) clássicas (DE VIVO et al., 2016; SOTRIFER, 2011; STANK et al., 2016; WIEDER et al., 2017).

Além do captopril e da dorzolamida, outros exemplos de fármacos descobertos empregando abordagens computacionais são: donepezila (KAWAKAMI et al., 1996) inibidor da acetilcolinesterase, usado no tratamento da Doença de Alzheimer; nelfinavir e o amprenavir (KUBINYI, 2006), inibidores da protease do vírus da imunodeficiência humana, usado no tratamento da síndrome da imunodeficiência adquirida; zanamivir e oseltamivir (VON ITZSTEIN et al., 1993; WADE, 1997), inibidores da neuraminidase, antivirais usados

no combate da infecção pelo vírus influenza; imatinibe (CAPDEVILLE et al., 2002), inibidor da proteína quinase, anticancerígeno; aliscireno (MAIBAUM; FELDMAN, 2009), inibidor da renina, usado para tratar hipertensão, e a sitagliptina (WEBER; THORNBERRY, 2007) inibidor da DPP-4 (dipeptidil peptidase 4), usada no tratamento do diabetes Tipo II. Contudo, o número de fármacos provenientes do planejamento auxiliado por computador (*CADD*, do inglês *Computer-Aided Drug Design*) é, ainda, pequeno (IRWIN; SHOICHET, 2016; RAVINA, 2011; SWINNEY; ANTHONY, 2011).

Ainda assim, ferramentas computacionais são amplamente usadas nas indústrias farmacêuticas tanto na etapa de descoberta (fase pré-clínica) como na de desenvolvimento de um fármaco (fase clínica). Dentre estas etapas, podemos citar: a identificação de *hits*¹ e de compostos líderes (do inglês *lead*²) através, por exemplo, da busca virtual; o processo de otimização dos compostos líderes; a predição e otimização de seus parâmetros ADMET e o desenvolvimento de formulação, entre outras (BROWN et al., 2017; KUHN et al., 2016; MCGAUGHEY; PATRICK WALTERS, 2017; MUEGGE; BERGNER; KRIEGL, 2017).

A seguir, os principais aspectos abordados nesta breve introdução da evolução da descoberta dos fármacos serão aprofundados para dar suporte conceitual às técnicas e às abordagens metodológicas utilizadas nesta tese.

1.2. Química Combinatória e *HTS* na descoberta de fármacos

A química combinatória e o *HTS* apareceram como uma forma de superar as limitações na etapa de descoberta de compostos líderes. A química combinatória consiste em preparar bibliotecas com grande número de compostos estruturalmente relacionados

¹ *Hit* é um composto com estrutura identificada/conhecida que possui uma atividade em um ensaio biológico abaixo de determinado valor corte (BUCKLE et al., 2014).

² *Lead* é um composto (ou uma série de compostos) com atividade e seletividade comprovada através de ensaios biológicos que, ainda, preenchem alguns critérios importantes para desenvolvimento de um fármaco como originalidade, patentabilidade e acessibilidade por extração ou síntese (IMMING, 2008).

utilizando as mesmas sequências de reação. Nestas bibliotecas, a diversidade estrutural é obtida usando blocos de construção com diferentes substituintes, sendo estes blocos diferentes tipos de reagente. Assim, n reagentes do tipo A (ex. aminoácidos com a porção *C*-terminal livre) são misturados com m reagentes do tipo B (ex. aminoácidos com a porção *N*-terminal livre) originando $(n \times m)$ produtos do tipo C (diferentes dipeptídeos).

O *HTS* consiste em utilizar robôs acoplados a equipamentos de detecção altamente sensíveis para realizar testes de atividade biológica em larga escala. Atualmente, estes equipamentos são acompanhados de *softwares* utilizados tanto no controle quanto no processamento de dados. A automação dos testes de atividade biológica permitiu testar grande número de compostos oriundos da química combinatória. O uso do *HTS* para identificação de compostos com atividade biológica deu origem ao termo *hit*, se referindo a um composto ativo identificado por *HTS* (BUCKLE et al., 2014).

No entanto, como já apresentado, a junção destas duas técnicas não levou necessariamente à descoberta de novos fármacos. Dentre as principais razões para o insucesso pode-se apontar o fato de as bibliotecas de compostos inicialmente geradas pela química combinatória terem sido baseadas unicamente na acessibilidade sintética e na diversidade estrutural. Deste modo, os compostos presentes nestas bibliotecas apresentavam alta massa molecular, alta lipofilicidade e baixo grau de pureza. Assim, como esperado, os *hits* identificados não apresentavam características físico-químicas adequadas, tendo problemas de biodisponibilidade e farmacocinética (KUBINYI, 2004b; MIGNANI et al., 2016). Outra limitação das primeiras bibliotecas de compostos geradas era conter compostos verdadeiramente dissimilares, uma vez que o conceito de similaridade e dissimilaridade química depende da propriedade e/ou do critério usado para comparação (KUBINYI, 1998). Estas limitações foram em parte superadas priorizando mais a qualidade do que o número de compostos sintetizados, dando ênfase para as bibliotecas de compostos com propriedades de

fármaco-similar (do inglês “*drug-like*”). Estas propriedades serão apresentadas e discutidas no item 1.7 desta Introdução. Neste contexto, a química combinatória passou a ser usada para gerar biblioteca de compostos focados em um determinado alvo biológico bem como para gerar séries de análogos para estudos de QSAR (KUBINYI, 2007).

Em relação ao *HTS*, como limitações iniciais pode-se destacar a falta de parâmetros estatísticos para analisar a qualidade e robustez dos ensaios bem como a constante presença de resultados falso-positivos (ou seja, compostos ativos devido a um mecanismo não específico). Na literatura são apontadas e discutidas diversas razões para ocorrência destes resultados, sendo as principais: o uso de compostos classificados como *frequent hitters* (ROCHE et al., 2002), que podem, ou não, interferir com os ensaios *in vitro* (BAELL; HOLLOWAY, 2010); o uso de compostos com grupos reativos frente às proteínas (RISHTON, 2003) e o uso de compostos capazes de formar agregados moleculares (IRWIN et al., 2015; MALVEZZI et al., 2008; MCGOVERN et al., 2003).

Por fim, mesmo com os avanços na química combinatória e nos ensaios *HTS*, a abordagem “irracional” de se testar, aleatoriamente, um grande número de compostos – vindos de coleções *in house*, disponíveis comercialmente e/ou sintetizados por química combinatória – por *HTS*, não leva, necessariamente, à descoberta de novos candidatos a fármacos. Quando a busca por novos fármacos se inicia sem o conhecimento de qualquer composto líder, centenas de milhares de compostos precisam ser investigados e não há garantias de que será possível encontrar um composto candidato a fármaco. Mesmo quando se conhece um, ou mais, composto(s) líder(es) e se tem informações sobre o alvo biológico de interesse, não há garantias de sucesso (KUBINYI, 2004b; MIGNANI et al., 2016).

1.3. Resultados falso-positivos

A preocupação em se evitar o quanto antes resultados falso-positivos no processo de descoberta de um fármaco é um consenso na literatura (BAELL; HOLLOWAY, 2010;

BOLOGA; OPREA, 2012; JADHAV et al., 2010; MIGNANI et al., 2016; OWENS et al., 2014; RISHTON, 2003). Neste sentido, esta preocupação é bem descrita pelo conceito:

“fail fast, fail cheap” (CLARK; PICKETT, 2000)

A tradução literal deste conceito é “falhe rápido, falhe barato” e seu significado pode ser facilmente entendido quando se considera tanto o tempo investido bem como os altos custos envolvidos nas fases de descoberta/desenvolvimento de fármacos (pré-clínica e clínica). Em 2014, Mullard (MULLARD, 2014) estimou que o custo para se desenvolver um fármaco podem chegar até 2,6 bilhões de dólares. Além do custo exorbitante, este processo é bastante demorado, podendo levar entre 10 a 15 anos. Assim, considerando a importância de identificar falso-positivos na etapa de descoberta (fase pré-clínica) bem como de selecionar apenas compostos com grande chance de sucesso para avançar nas etapas de desenvolvimento (fase clínica), um grande esforço vem sendo feito para entender as razões e evitar, o mais cedo possível, a ocorrência destes falso-positivos e/ou insucessos (BAELL; HOLLOWAY, 2010; IRWIN et al., 2015; MALVEZZI et al., 2008; MEANWELL, 2011; MIGNANI et al., 2016; SCHULZE et al., 2014; WARING et al., 2015). A seguir, serão brevemente abordadas algumas das razões para ocorrência de falso-positivos nos ensaios *in vitro* (fase pré-clínica).

1.3.1. *Frequent hitters* e *PAINS*

O termo *frequent hitters* foi criado por Schneider (ROCHE et al., 2002) se referindo aos compostos identificados como ativos (*hits*) em vários ensaios biológicos frente a diferentes alvos. Schneider e colaboradores analisaram as características dos *hits* identificados em mais de oito ensaios *HTS* com diferentes alvos bem como dos compostos *in house* da Hoffmann-La Roche requeridos por pelo menos seis diferentes projetos de descoberta de fármacos. Dentre as razões para a recorrente identificação destes compostos como *hits* os

autores apontam: a inibição inespecífica dos alvos (ou seja, compostos promíscuos) e a interferência com o ensaio e/ou o método de detecção (ex. moléculas coloridas ou fluorescentes). Seja devido a uma atividade promíscua seja por interferirem no ensaio, os autores concluem que estes *hits* geralmente não são bons compostos líderes (ROCHE et al., 2002). Ainda, neste estudo, os autores puderam separar os *frequent hitters* (479 compostos) dos não *frequent hitters* (423 compostos) utilizando modelos de redes neurais. No entanto, eles não encontraram subestruturas comuns a todos, ou a quase todos, os *frequent hitters* que pudessem ser usadas como filtros na busca virtual (definida no próximo item).

Em 2010, Baell e Holloway (BAELL; HOLLOWAY, 2010) analisaram um conjunto de 93.212 compostos da biblioteca de *HTS* do WEHIMR – *Walter and Eliza Hall Institute of Medical Research* – testados em seis diferentes campanhas de *HTS*, usando a tecnologia *AlphaScreen* para ensaios de inibição de interações proteína-proteína (denominada IPP). Destes, 2.062 (2,2 %) deram resultados positivos em pelo menos quatro campanhas. Além das razões já descritas por Schneider, os autores também sugerem a capacidade de reagir com oxigênio singleto (no caso ensaios tipo *AlphaScreen*) e de formar ligações covalentes como origem da recorrente identificação destes compostos como *hits*. Ainda, a partir da análise da estrutura destes *hits* acompanhada da análise da taxa de enriquecimento, os autores sugeriram uma série de filtros estruturais agrupados em três diferentes classes de acordo com o número de compostos que apresentam a subestrutura sugerida como filtro. Para 15 classes identificadas nestes filtros, os autores ainda compararam seus resultados com os descritos na literatura obtidos por *HTS*, usando diferentes métodos de detecção e outros alvos (não apenas para IPPs). Esta comparação sugeriu 10 classes com compostos capazes de interferir com vários tipos de ensaios, denominados pelos autores como *PAINS* (do inglês, *Pan Assay INterference compoundS*), sendo elas: rodaminas, bases fenólicas de Mannich, hidroxifenil-hidrazonas, alquilidenos de barbituratos, tetra-hidroquinolinas fundidas, 1,2,3-arilpirrois,

benzofuranos ativados, 2-amino-3-carboniltiofenos, catecóis e quinonas. Por fim, outras classes foram identificadas como possíveis *PAINS*, sendo elas: azo, ciano-piridonas, divinilcetonas e alguns indóis. Como via de regra, os autores sugerem evitar compostos contendo estas classes mesmo que algumas delas possam ser encontradas em fármacos já disponíveis no mercado (6,5 %), pois estas também são frequentemente associadas com à toxicidade, reatividade ou à instabilidade metabólica destes fármacos (BAELL; HOLLOWAY, 2010).

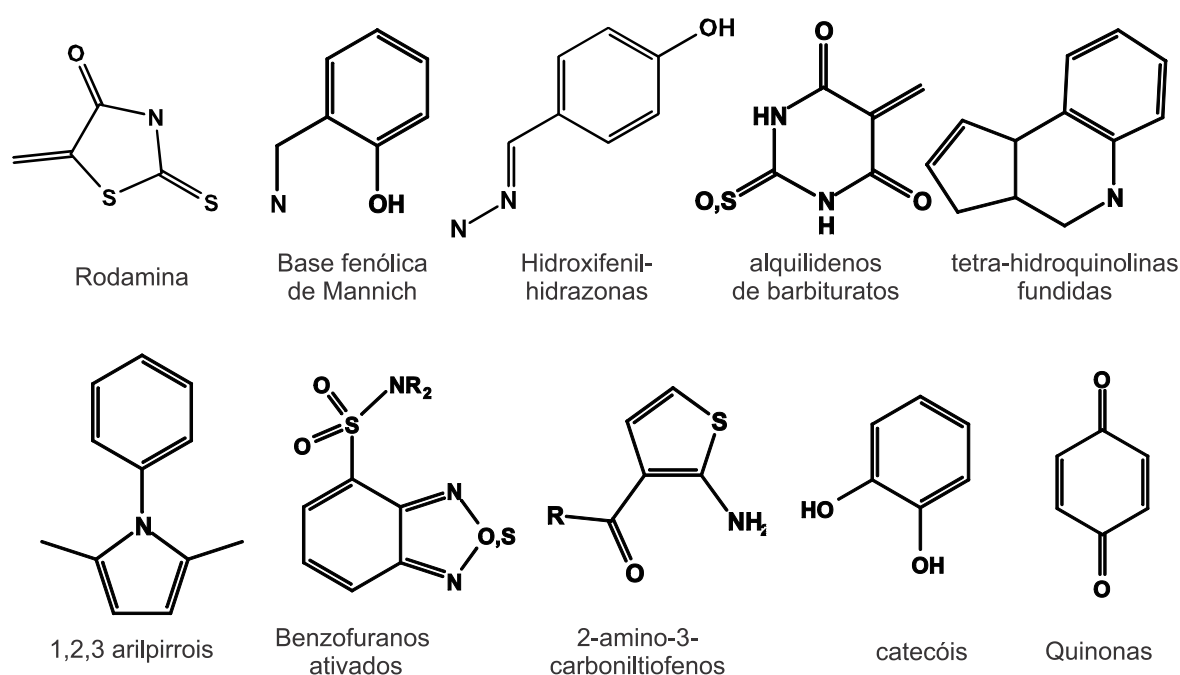


Figura 1.3. Representação da subestrutura 2D das 10 classes com compostos capazes de interferir com vários tipos de ensaios, denominados de *PAINS* (do inglês, *Pan Assay Interference compounds*) (BAELL; HOLLOWAY, 2010), sendo elas: rodaminas, bases fenólicas de Mannich, hidroxifenil-hidrazonas, alquilidenos de barbituratos, tetra-hidroquinolinas fundidas, 1,2,3-arilpirois, benzofuranos ativados, 2-amino-3-carboniltiofenos, catecóis e quinonas.

1.3.2. Compostos com grupos reativos

Compostos contendo grupos reativos frente à proteína são outra razão de resultados falso-positivos, pois a atividade observada para estes compostos se deve mais à sua reatividade do que a sua capacidade de interagir com o alvo de interesse. Os exemplos mais comuns de compostos reativos atuando como falso-positivos formando ligações covalentes são aqueles que possuem em sua estrutura grupos eletrofílicos. Estes grupos são agentes

alquilantes e/ou acilantes reativos frente à nucleófilos biológicos (ex. proteínas plasmáticas e a glutatona). Uma lista de grupos reativos frente à nucleófilos biológicos foi descrita por Rishton (RISHTON, 1997) e atualizada, em 2012, por Bologa e Oprea (2012). A Figura 1.4 apresenta uma lista de alguns destes grupos reativos. Ainda, embora estes grupos estejam presentes em algumas classes terapêuticas, Rischton defende que compostos contendo grupos reativos identificados como “ativos” devem ser avaliados com cautela para verificar se estes resultados não são falso-positivos (RISHTON, 1997).

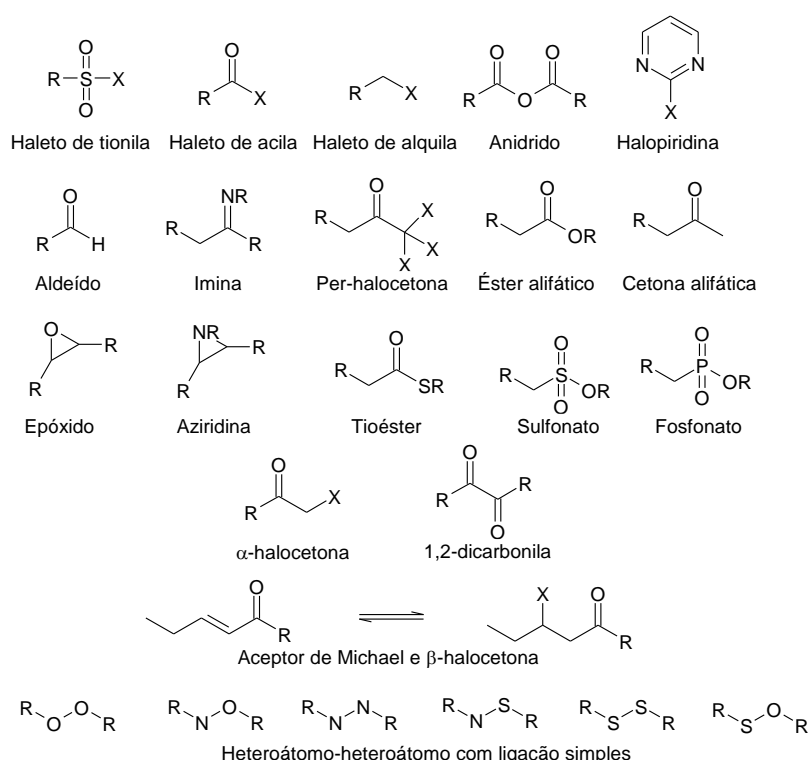


Figura 1.4. Representação da estrutura 2D de alguns grupos reativos responsáveis pelos resultados falso-positivos em experimentos *in vitro*. Grupos retirados das publicações de Bologa e Oprea (2012), Malvezzi (2008) e Rishton (1997).

1.3.3. Compostos capazes de formar agregados

A capacidade dos compostos formarem agregados em ensaios *in vitro* levando a resultados falso-positivos foi descrita pela primeira vez por Schoichet e colaboradores (MCGOVERN et al., 2002). Estes pesquisadores analisaram conjunto de 15 compostos identificados como *hits* em ensaios *HTS* usando diferentes alvos. Ainda, como características comuns a estes compostos, já haviam sido observados mecanismos inibição não competitivos,

correlações estrutura-atividade não significativas, curvas dose-resposta íngremes e baixa seletividade (sendo ativos frente a várias enzimas não relacionadas). Além destas características, estes pesquisadores encontraram que esses compostos apresentavam mecanismo de inibição dependente do tempo sendo, no entanto, reversível. Outra característica identificada foi a diminuição da atividade inibitória na presença de albumina, guanidina ou ureia. Ainda, eles observaram que aumento na concentração da enzima (10 vezes maior) eliminava totalmente a inibição, mesmo usando um excesso de inibidor (concentração 1000 vezes maior). Com base nestes resultados, estes pesquisadores sugeriram que a forma ativa destes inibidores não específicos corresponderia a um agregado de muitas moléculas. Esta hipótese foi confirmada por experimentos de espalhamento de luz (*Dynamic Light Scattering - DLS*) e de microscopia eletrônica, sendo observada a formação de partículas com diâmetro entre 50 a 400 nm. A Figura 1.5 representa o mecanismo dos inibidores não específicos (inibidores “promíscuos”) descrito por Schoichet (MCGOVERN et al., 2002). Neste mecanismo, pequenas moléculas formam partículas com diâmetro entre 50 e 400 nm, que absorvem/adsorvem a enzima alvo, acarretando na sua inibição.

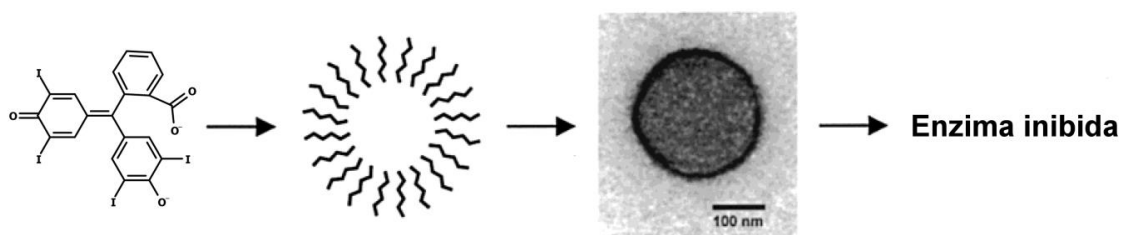


Figura 1.5. Representação do mecanismo dos inibidores não específicos. Neste mecanismo, pequenas moléculas formam partículas com diâmetro entre 50 e 400 nm, que absorvem/adsorvem a enzima alvo inibindo-a (MCGOVERN et al., 2002).

Em outro trabalho, Schoichet e colaboradores (MCGOVERN et al., 2003) verificaram por microscopia que os agregados não levam à desnaturação da enzima alvo. Estes estudos mostraram que a enzima está associada à membrana destes agregados e que a inibição observada é decorrente do “sequestro” da enzima da solução. Mais importante, observou-se

que o uso de Triton X-100 (detergente não iônico) na concentração de 0,01 % v/v consegue reverter ou prevenir esta inibição, sem afetar, significativamente, a atividade da enzima estudada (β -lactamase). A partir deste estudo, o uso de detergentes (ex. Triton X-100, CHAPS, Tween20 ou Tween80) em ensaios *in vitro* vem sendo usado para prevenir e comprovar o mecanismo de inibição não específico (JADHAV et al., 2010; MALVEZZI, 2008; MALVEZZI et al., 2008). No entanto, nestes trabalhos não haviam sido, ainda, sugeridos métodos computacionais para prever este tipo de mecanismo.

Mais recentemente, Irwin e colaboradores (IRWIN et al., 2015) desenvolveram o “*Agregator Advisor*” uma ferramenta on-line para reconhecer compostos que podem formar possíveis agregados. Estes autores analisaram algumas propriedades físico-químicas de um banco de dados com 12.600 compostos reconhecidos por formar agregados. Nesta análise, foi observado que 80% dos compostos apresentavam $\log P > 3$, sendo este considerado o primeiro critério para prever agregação. O segundo critério usado nesta ferramenta foi a similaridade estrutural com um, ou mais, composto(s) presente(s) neste banco. A utilidade da ferramenta foi avaliada selecionando e testando sete moléculas preditas como capazes de formar. Destas sete, cinco apresentaram mecanismo de inibição promíscuo através da formação de agregados.

A diferenciação entre resultados verdadeiro-positivos de falso-negativos se tornou tão importante que a partir de 2014 o *Journal of Medicine Chemistry (JMC)* passou a recomendar que compostos com atividade biológica, independente da origem (por exemplo, comercial, sintetizado e produto natural), devam ser examinados quanto à presença de subestruturas já reconhecidas por interferir nos ensaios *in vitro* e *in vivo* (BAELL; WALTERS, 2014). Dentre as razões para estas interferências além das já descritas anteriormente (agregação, atividade redox, fluorescência, reatividade frente às proteínas e extinção de oxigênio-singlete), também são citadas a presença de impurezas, a alteração de membranas biológicas e falta de estabilidade no tampão do ensaio, para compostos com grupos reativos. Ainda, no caso de

compostos ou classes de compostos já associados a resultados falso-positivos (ex. *PAINS*), o *JMC* recomenda que a atividade descrita seja confirmada por pelo menos dois ensaios diferentes.

Considerando todos os relatos descritos, alguns filtros podem ser aplicados a bibliotecas e/ou bancos de dados de compostos antes que estes sejam submetidos a ensaios biológicos (*in vitro* ou *in vivo*) visando a eliminar aqueles com grupos associados a possíveis resultados falso-positivos e/ou como propriedades indesejadas para um fármaco. Desta forma, o tempo e o custo para se encontrar um composto líder e, subsequentemente, um candidato a fármaco podem ser reduzidos.

1.4. Busca virtual

A busca virtual surgiu em resposta ao insucesso obtido usando a química combinatória em conjunto com os ensaios *HTS*, já que se observou que testar aleatoriamente um grande número de compostos não resultou num avanço significativo na descoberta de novos fármacos. A busca virtual consiste em pré-selecionar os compostos com o auxílio do computador a partir de bancos de dados virtuais com um grande número de moléculas. Desta forma, apenas os compostos pré-selecionados são submetidos aos testes de atividade *in vitro*. Esta pré-seleção é feita prevendo-se, virtualmente, a atividade biológica de interesse através de diferentes abordagens (*LBDD* e/ou *SBDD*). Como vantagem desta estratégia tem-se que os compostos estudados não precisam, necessariamente, existir fisicamente e que neste “teste virtual”, *in silico*, não há consumo de material (KLEBE, 2006).

No entanto, diferentemente dos *HTS*, para se realizar a busca virtual torna-se crucial conhecer as propriedades responsáveis pelas interações entre o composto e o sistema biológico (GOHLKE; KLEBE, 2002; KLEBE, 2006; LIPINSKI, 2000; WENLOCK; BARTON, 2013). Num primeiro momento, não se faz necessário considerar algumas limitações/deficiências experimentais dos compostos selecionados, tais como solubilidade,

formação de agregados, estabilidade, entre outras (KLEBE, 2006). Contudo, como já apresentado e discutido nesta Introdução (item 1.3), a importância de evitar o quanto antes resultados falso-positivos, bem como de priorizar apenas compostos com grandes chances de sucesso na fase clínica, tem feito com que estes aspectos sejam incluídos já nas etapas iniciais. Desta forma, para que os ligantes e/ou compostos selecionados possam, no futuro, se tornar fármacos é desejável considerar na busca virtual determinadas propriedades como sua solubilidade e seus parâmetros de ADMET (GLEN, 2011; HOPKINS; GROOM, 2002; WENLOCK; BARTON, 2013). As informações disponíveis sobre a interação ligante-receptor e sobre as propriedades do ligante são usadas na construção do modelo de busca virtual aplicado a um banco de dados de moléculas virtuais.

A proposição de modelos de busca virtual válidos é dificultada pela complexidade (LEACH; HANN, 2011) inerente às interações ligante-receptor ao nível molecular (discutida em detalhes no item 1.5), além de outros fatores tais como: múltiplos modos de ligação do ligante (MATTOS; RINGE, 1993; SCIOR et al., 2012), múltiplas conformações do ligante e do receptor (KLEBE, 2006; KOCH et al., 2013; SCIOR et al., 2012; STANK et al., 2016), ligação com proteínas plasmáticas e parâmetros ADMET do ligante (GLEESON et al., 2011; HODGSON, 2001; KIRCHMAIR et al., 2015). Tais dificuldades são dificilmente superadas por um único filtro de seleção (OPREA, 2002). Como alternativa na literatura se sugere aplicar diferentes filtros de seleção de modo sequencial (KLEBE, 2006). Os filtros podem ser aplicados em qualquer sequência sem, em princípio, alterar o resultado final da seleção. Entretanto, com o objetivo de se otimizar o uso dos recursos computacionais, em geral, os filtros são aplicados em ordem crescente de complexidade do cálculo. Assim, para o banco de dados de moléculas virtuais se aplicam primeiramente filtros que necessitam de baixos recursos computacionais, deixando os filtros mais complexos e elaborados para serem aplicados nas etapas finais do processo, quando o número de compostos selecionados já for

significativamente reduzido. De modo geral, os filtros podem ser agrupados em três categorias distintas: (i) filtros uni- e bidimensionais (1D e 2D), que se baseiam nas propriedades físico-químicas e na conectividade dos compostos (filtros de propriedades “*drug-like*” e de similaridade 2D); (ii) filtros tridimensionais (3D), que se baseiam na distribuição espacial de características específicas (filtros farmacofóricos e de formato molecular – “*shape*”) e, (iii) filtros de ancoramento. As vantagens, limitações e exemplos de aplicações recentes de sucesso destes filtros de seleção serão apresentados e discutidos a seguir. O sucesso da busca virtual vai depender da qualidade tanto dos filtros de seleção bem como do banco de dados usados (SCIOR et al., 2012).

Ainda, cabe ressaltar que por melhor que seja o poder de previsão do modelo proposto, este deve ser submetido à validação experimental (por exemplo, cristalografia de raios-X e ensaios enzimáticos) visando-se confirmar sua utilidade na seleção de compostos/ligantes bioativos (BRIAN Y FENG et al., 2005; FERREIRA et al., 2010; KLEBE, 2006; MALVEZZI et al., 2011; MALVEZZI; DE REZENDE; AMARAL, 2009; OPREA, 2002). De modo semelhante aos ensaios de *HTS*, os compostos selecionados através da busca virtual que apresentam a atividade biológica de interesse, comprovada por meio da validação experimental, são denominados *hits* (representando resultados positivos).

1.5. Interação ligante-receptor

Como já apresentado a interação ligante-receptor é muito complexa, sendo de extrema importância entendê-la tanto para a compreensão do mecanismo de ação de um composto quanto para prever sua atividade biológica. Para tanto, é preciso conhecer a grandeza e a natureza das forças que determinam esta interação. A maioria dos fármacos disponíveis no mercado se liga ao seu receptor alvo através de interações não covalentes. Por esta razão, estas são as principais interações estudadas e consideradas nas estratégias de *LBDD* e *SBDD*.

1.5.1. Descrição termodinâmica da interação ligante-receptor

Quando um ligante (L) está próximo de seu receptor alvo (R), eles podem interagir formando o complexo ligante-receptor (RL). Esta interação pode ser, inicialmente, mediada por interações eletrostáticas de longo alcance entre o ligante e o receptor e, em seguida, reforçada por interações de curto alcance como, por exemplo, ligações de hidrogênio e interações de *van der Waals*. A força desta interação (ou seja, a afinidade do ligante pelo receptor) pode ser descrita pelos valores das constantes de equilíbrio K_d (constante de dissociação do complexo RL), desde que as condições de equilíbrio termodinâmico tenham sido atingidas. Este equilíbrio pode ser representado pela seguinte equação:



Sendo que,

$$K_d = \frac{[R][L]}{[RL]}$$

onde: K_d é o valor da constante de dissociação para o complexo RL; $[RL]$ é o valor da concentração do complexo RL; $[R]$ é o valor da concentração molar do receptor e $[L]$ é o valor da concentração molar do ligante.

No caso do ligante ser um inibidor enzimático, a constante de dissociação do complexo EI também pode ser denominada constante de inibição (K_i). Quanto menor o valor de K_d/K_i maior é afinidade do ligante/inibidor pelo receptor/enzima alvo. No geral, um ligante potente tem K_d/K_i na ordem de pM a nM. O valor do logaritmo da constante de dissociação ($\log K_d$ ou $\log K_i$) se correlaciona linearmente com o valor da variação da energia padrão de *Gibbs* no processo de formação do complexo *RL/EI*, que pode ser escrito pela seguinte equação:

$$\Delta G^\circ = -2,303 RT \log K$$

onde: ΔG é a variação do valor da energia de *Gibbs* da formação do complexo ligante-receptor, , nas concentrações padrões (1 mol, a 1 atm); R é o valor da constante universal dos gases, T é o valor da temperatura (em Kelvin) e $\log K$ é o valor da constante de equilíbrio da interação entre o ligante e o receptor.

Assim, quando o equilíbrio é alcançado, os valores de K_d/K_i estão na faixa de 10^{-2} a 10^{-12} M, o que corresponde a um valor de ΔG entre -10 e -70 kJ/mol (SCHAEFFER, 2008). A energia de *Gibbs* é composta de um termo entálpico (ΔH) e um termo entrópico ($T\Delta S$), como mostrado na equação a seguir:

$$\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$$

onde: ΔG° é a variação do valor da energia padrão de *Gibbs* da formação do complexo RL; ΔH° é a variação da entalpia; ΔS° é a variação da entropia e T é o valor da temperatura (em Kelvin).

A formação do complexo ligante-receptor pode ser controlada tanto por fatores entrópicos quanto por fatores entálpicos determinando, assim, sua assinatura ou perfil termodinâmico (KLEBE, 2015). No entanto, durante este processo pode ocorrer uma compensação mútua entre os fatores entálpicos e entrópicos, ou seja, o aumento de um fator leva a diminuição do outro. Esta compensação pode ser entendida se considerarmos que uma interação entálpica forte “prende” o ligante dentro do receptor diminuindo, portanto, sua entropia. Por outro lado, um ligante com grande mobilidade no sítio de ligação é entropicamente favorável, uma vez que poucos graus de liberdade são perdidos durante a formação do complexo RL, entretanto as interações formadas no complexo RL serão mais fracas, resultando numa pequena contribuição entálpica (KLEBE, 2015). Devido a esta compensação entrópica-entálpica, nem sempre é possível melhorar a atividade de um composto favorecendo ambos os termos (KLEBE, 2015). Um exemplo desta compensação foi descrito por Klebe (2015) usando série de oito inibidores peptidomiméticos da tripsina (uma serino protease humana envolvida no processo digestivo), diferindo apenas em suas posições P_3 . Os valores de ΔG , ΔH e $T\Delta S$ de cada composto desta série foram determinados por calorimetria de titulação isotérmica (do inglês ITC, *isothermal titration calorimetry*). Os oito inibidores apresentam valores de ΔG comparáveis, entretanto, suas assinaturas termodinâmicas são diferentes (Figura 1.6).

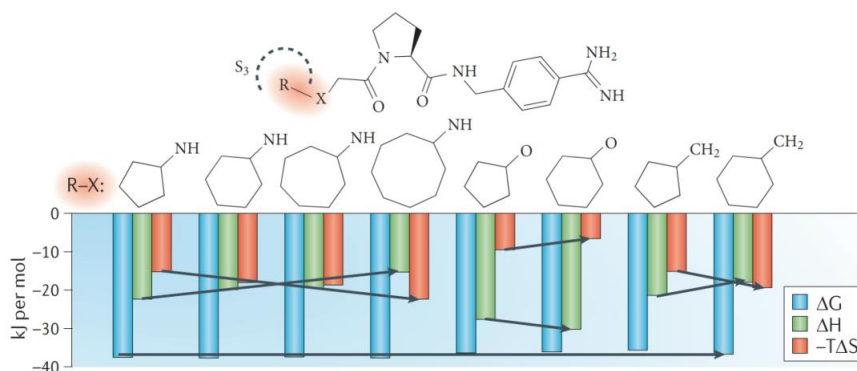


Figura 1.6. Assinaturas termodinâmicas de uma série de oito inibidores peptidomiméticos da tripsina. Os gráficos de barra mostram a contribuição entálpica (ΔH , barras verdes) e entrópica ($-T\Delta S$, barras laranja) para a variação da energia de Gibbs (ΔG , barras azul). Embora os valores da energia de Gibbs (ΔG) sejam comparáveis entre os oito inibidores, estes diferem quanto aos componentes entrópicos ($-T\Delta S$) e entálpicos (ΔH) (KLEBE, 2015).

Nesta série, quando o aumento do anel na posição R é feito incluindo-se uma amina ($-NH-$) ou um metileno ($-CH_2-$), a contribuição entálpica para a ligação diminui e a contribuição entrópica aumenta. Por outro lado, quando este aumento é feito incluindo um grupo éter ($-O-$), o inverso é observado. A razão desta compensação entálpica-entrópica foi analisada através de cristalografia de raios-X e simulações de DM. Para os ligantes com o grupo amina, duas diferentes conformações foram observadas. A primeira com o anel localizado dentro da cavidade S_3 (fazendo interações predominantemente hidrofóbicas) e a segunda com o anel fora da cavidade (formando uma ligação de hidrogênio). Observou-se que o aumento do tamanho do anel favorece a conformação fora da cavidade (aumentando a contribuição entrópica para ligação). Por outro lado, os compostos com grupo éter ou metileno exibem apenas uma conformação com o anel localizado dentro da cavidade. Ainda, para estes a mobilidade residual dos inibidores com éter é menor, explicando porque para estes a contribuição entálpica é maior. É interessante observar que tanto o tamanho do anel quanto o grupo usado modificam o perfil termodinâmico da ligação enzima-inibidor.

Os valores de variação da entalpia (ΔH) e da entropia ($-T\Delta S$) medidos através, por exemplo, do *ITC* são um somatório de interações favoráveis e/ou desfavoráveis de diversas

naturezas. A Figura 1.7 mostra uma representação das contribuições entálpicas e/ou entrópicas que podem ocorrer durante a formação do complexo RL.

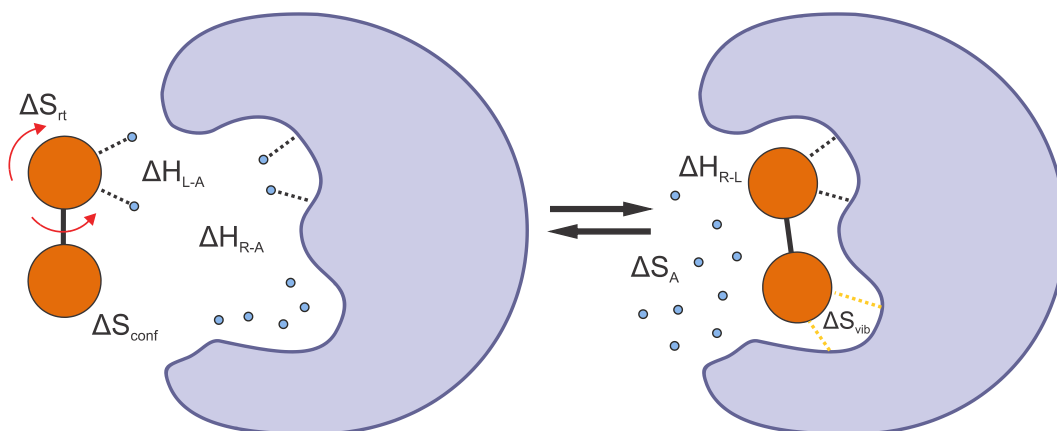


Figura 1.7. Representação esquemática das contribuições entálpicas e/ou entrópicas para a formação do complexo [RL]. Os círculos azuis, as esferas laranja e a superfície roxa representam as moléculas de água, o ligante e o receptor, respectivamente. ΔH_{L-A} e ΔH_{R-A} correspondem às variações das entalpias de dessolvatação do ligante e do receptor, respectivamente. ΔH_{R-L} correspondem à contribuição entálpica das interações LR. ΔS_A corresponde à variação da entropia resultante da dessolvatação de superfícies hidrofóbicas. ΔS_{rt} e ΔS_{conf} correspondem à variação da entropia rotacional e translacional do ligante. ΔS_{vib} corresponde à variação da entropia associada aos novos modos de vibrações de baixa frequência de energia obtidos com a formação de ligações não-covalentes no complexo RL. Linhas pontilhadas pretas e amarelas mostram ligações de hidrogênio e interações de *van der Waals*, respectivamente. Figura baseada na representação descrita por Andrews (ANDREWS; CRAIK; MARTIN, 1984).

Nesta figura, pode-se observar que inicialmente o ligante e o receptor estão solvatados, por moléculas de água. Assim, para a formação do complexo RL, estes devem ser dessolvatados. A dessolvatação de grupos polares do ligante e do receptor durante a formação do complexo RL são entalpicamente desfavoráveis. Estas precisam ser compensadas, em parte, pela formação de ligações de hidrogênio entre o ligante e o receptor. No entanto, a dessolvatação de superfícies hidrofóbicas contribui positivamente para entropia. Este é o chamado efeito hidrofóbico, no qual se observa um ganho entrópico e entálpico quando duas regiões apolares solvatadas interagem entre si. Quando moléculas de água solvatarem uma região apolar, elas formam ligações de hidrogênio mais fortes entre si visando compensar as fracas interações entre elas e a região apolar. Com isso, as moléculas de água ficam num arranjo extremamente organizado, sendo assim, o ganho entrópico associado à dessolvatação

destas regiões é bastante significativo. Ainda, quando duas regiões apolares interagem entre si, elas foram interações de *van der Waals* que são entalpicamente favoráveis.

1.5.2. Interações não covalentes do complexo ligante-receptor

Além de ligações de hidrogênio, interações eletrostáticas e de *van der Waals* podem ser formadas entre o ligante e o receptor. A Tabela 1.1 mostra as principais interações não covalentes encontradas no complexo RL e as respectivas faixas de ΔG envolvidas.

Tabela 1.1. Principais tipos de interações entre um composto e o sistema biológico, um exemplo e as respectivas faixas de energias envolvidas e dependência da distância (r) (SCHAEFFER, 2008).

Tipo de interação	ΔG e dependência da distância (r)	Exemplo
Iônica (longo alcance não direcional)	de -20 a -40 kJ/mol e $1/r$	
Carga-dipolo (depende da orientação do dipolo)	de -12 a -20 kJ/mol e $1/r^2$	
Dipolo-dipolo (depende da orientação mútua dos dipolos)	de -4 a -12 kJ/mol e $1/r^3$	
Carga-dipolo induzido (cátion - π) (depende da polarizabilidade da molécula na qual se induz o dipolo)	de -2 a -10 kJ/mol e $1/r^4$	
Dipolo-dipolo induzido (depende da polarizabilidade da molécula na qual se induz o dipolo)	de -2 kJ/mol e $1/r^6$	
Dispersão (hidrofóbica) (envolve a sincronização mútua da nuvem eletrônica)	de -2 a -4 kJ/mol e $1/r^6$	
Ligação de hidrogênio (iônica e covalente parcial, sendo direcional)	de -4 a -30 kJ/mol	

Na Tabela 1.1 pode-se observar que as interações iônicas são as mais fortes e de longo alcance, mas estas não são direcionais. Já a energia das ligações de hidrogênio varia bastante, no entanto estas são direcionais. Como descrito anteriormente, o ganho entálpico associado à

formação de ligações de hidrogênio no complexo ligante-receptor, em geral, é compensado pela dessolvatação. Sendo assim, estas não levariam a um ganho entálpico líquido. Contudo, o fato de a ligação de hidrogênio apresentar uma geometria específica a torna bastante interessante, pois esta é capaz de conferir especificidade a interação ligante-receptor (BISSANTZ; KUHN; STAHL, 2010). Além da energia destas interações, outro fator importante é a distância em que elas ocorrem. Bissantz, Kuhn e Stahl (2010) fizeram uma análise estatística das principais interações não covalentes usadas em química medicinal através de cristalografia de raios-X. Com esta análise, os autores conseguiram descrever quais as distâncias típicas observadas para estas interações. Para as ligações de hidrogênio, a distância entre o átomo doador e o correspondente átomo aceptor variou entre 2,6 e 3,2 Å. Interações hidrofóbicas mostraram distâncias entre 3,3 e 4,4 Å, enquanto que interações cátion- π apresentaram distâncias entre 3,4 e 4,0 Å (BISSANTZ; KUHN; STAHL, 2010). Por fim, na Tabela 1.1, podemos observar que as interações hidrofóbicas são as mais fracas, no entanto, estas são muito importantes por ocorrem em maior número. Várias vezes relata-se, na literatura, o aumento da afinidade ligante-receptor quando este passa a ocupar um sítio hidrofóbico do receptor (KUHN et al., 2016).

Adicionalmente, na Figura 1.7 é importante observar os termos ΔS_{rt} e ΔS_{Conf} , que representam a perda entrópica rotacional, translacional e conformacional decorrente da formação do complexo RL. Além do ligante, o receptor também pode perder alguns graus de liberdade durante a formação deste complexo. Isto é observado quando o receptor livre adota diferentes conformações e o ligante se liga de forma seletiva a uma destas conformações pré-existentes (STANK et al., 2016).

1.5.3. Predição da afinidade ligante-receptor

Considerando-se a complexidade da interação ligante-receptor e os vários aspectos que a influenciam, prever a afinidade de um ligante pelo seu receptor usando modelos

matemáticos e ferramentas computacionais é uma tarefa bastante difícil. A energia de *Gibbs* é uma função de estado, ou seja, é uma propriedade que depende apenas da estrutura do sistema em equilíbrio. Assim, o cálculo de funções de estado deve ser feito usando um conjunto médio do sistema. Os métodos mais precisos para a determinação de funções de estado são os cálculos de mecânica estatística, como Monte Carlo e Dinâmica Molecular (KITCHEN et al., 2004). Estes métodos permitem fazer uma amostragem significativa do conjunto do sistema, sendo possível estimar, com certa precisão, a entropia do sistema (DE VIVO et al., 2016). Além disso, a inclusão explícita de moléculas de água no sistema melhora o cálculo da sua entropia. Alguns dos métodos usados para se determinar os valores de ΔG e de K_d serão abordados no item 1.10. Embora avanços significativos tenham sido feitos nos cálculos de mecânica estatística, estes ainda são computacionalmente caros, inviabilizando seu uso em larga escala. Com o intuito de tornar estes modelos mais rápidos, várias aproximações são propostas e vários aspectos são desconsiderados (IRWIN; SHOICHET, 2016). Exemplos desta aproximação são as funções de ranqueamento usadas no ancoramento (descritas no item 1.8).

1.5.4. Descrição cinética da interação ligante-receptor

Os conceitos apresentados até agora nesta Introdução são usados para descrever a afinidade da interação do ligante-receptor em condições de equilíbrio termodinâmico. No entanto, mais recentemente tem sido importante considerar também a dinâmica conformacional, que afeta o processo de associação e dissociação de um ligante. Para isto, foi introduzido o conceito do tempo de residência da interação ligante-receptor. O primeiro modelo para descrever o tempo de residência foi introduzido, em 2006, por Copeland (COPELAND, 2015). Neste modelo, o ponto chave para a atividade farmacológica *in vivo* e para duração desta atividade não é somente a afinidade do fármaco pelo receptor, medida “em um sistema fechado” e em condições de equilíbrio termodinâmico (ex. K_i , K_d , IC_{50} e EC_{50}).

Esta consideração veio do fato que o sistema biológico é um “sistema aberto” e que a concentração do fármaco disponível para interagir com seu alvo biológico não é constante. Sendo assim, não seria possível alcançar as condições de equilíbrio necessárias para que estas medidas sejam válidas. Como alternativa, propõe-se que o ponto chave para explicar a interação farmacológica no sistema biológico é o tempo de vida, ou tempo de residência, do complexo RL. Este conceito considera que a atividade biológica só ocorre quando um fármaco se liga ao seu alvo formando o complexo RL e esta só durará enquanto este complexo existir. Uma vez que ocorra a dissociação, o alvo está livre para assumir sua função fisiopatológica normal. Neste conceito, a constante de equilíbrio (K_d) é decomposta em duas constantes de velocidade denominadas k_{on} e k_{off} , que representam a constante de velocidade de associação e de dissociação, respectivamente. O tempo de residência (τ) é definido como sendo o inverso do k_{off} ($\tau = 1/k_{off}$). Assim, o tempo de vida do complexo RL é aumentado quando sua velocidade de associação é rápida e sua velocidade de dissociação é lenta, sendo esta última mais importante. Baseando-se nisso, foi proposto que, uma forma de aumentar a atividade de um fármaco *in vivo* seria aumentando o valor de k_{off} . No entanto, os processo de associação e de dissociação ligante-receptor geralmente incluem várias etapas, que envolvem mudanças conformacionais do receptor. Considerando a importância da constante do k_{off} para se determinar a atividade farmacológica *in vivo*, nos últimos anos vários experimentos e métodos computacionais vêm sido desenvolvidos e usados para determinar e/ou prever o valor destas constantes (COPELAND, 2015; DE VIVO et al., 2016). Dentre os métodos computacionais se destacam as simulações de dinâmica molecular, que serão abordadas no item 1.10.

1.5.5. Medidas da eficiência da ligação

Como já apresentado, existe uma relação direta entre o aumento da afinidade e lipofilicidade do ligante. Esta, no entanto, pode causar um problema, pois ao se aumentar a

afinidade através de interações com regiões hidrofóbicas, pode-se fazer com que as propriedades ADME do ligante fiquem fora da faixa desejada. Como alternativa propôs-se que a afinidade do ligante pelo seu receptor seja corrigida pela sua lipofilicidade, ou massa molecular. Neste contexto, surgiram as medidas de eficiência do ligante (HOPKINS et al., 2014). Dentre estas medidas, se destacam a eficiência do ligante (EL, do inglês *ligand efficiency*), o índice de eficiência do ligante (LEI, do inglês *ligand efficiency index*), eficiência lipofílica do ligante (LLE, do inglês *lipophilic ligand efficiency*) e o índice de eficiência de ligação (BEI, do inglês *binding efficiency index*), apresentadas na Tabela 1.2.

Tabela 1.2. Medidas de eficiência do ligante, definição e valores mínimos sugeridos pela literatura (HOPKINS et al., 2014; MIGNANI et al., 2016).

Nome da medida	Definição ^(a)	Valor mínimo aceitável ^(b)
EL	$-2,3 \frac{\log K_d}{HA} \cong \frac{1,37 pAtividade}{HA}$	> 0,3 kcal por mol por átomo pesado
LEI	$\frac{pAtividade}{HA}$	-
LLE	$pAtividade - clogP$	> 5
BEI	$\frac{pAtividade}{PM}$	27

^(a) Onde K_d é a constante de equilíbrio de dissociação para o complexo RL; HA é o número de átomos pesados do ligante (ou seja, número de átomos sem considerar os átomos de hidrogênio); $pAtividade$ pode ser pIC_{50} (logaritmo negativo do valor da metade da concentração inibitória máxima), pEC_{50} (logaritmo negativo da concentração que induz da metade da resposta máxima), pK_i (logaritmo negativo da constante de inibição) e pK_d (logaritmo negativo da constante de equilíbrio de dissociação); $clogP$ é o valor do coeficiente de partição calculado que pode ser substituído pelo $\log D$ que é o coeficiente de distribuição (este inclui a contribuição da ionização para partição em um determinado pH) e PM é o massa molecular em kDa.

^(b) De acordo com a literatura (HOPKINS et al., 2014; MIGNANI et al., 2016), considerando um ligante com um $K_d > 10$ nM e 38 átomos pesados (~ 500 Da).

A literatura sugere usar medidas de eficiência de ligação para comparar diferentes compostos líderes e selecionar aqueles que devem ser priorizados na fase de otimização (HOPKINS et al., 2014; MIGNANI et al., 2016). No entanto, para que estas comparações

sejam possíveis, os valores de atividade devem ser medidos nas mesmas condições experimentais (HOPKINS et al., 2014).

Os aspectos relacionados à interação ligante-receptor abordados, até aqui, permitem entender algumas das interações e/ou características consideradas nos diferentes filtros de seleção usados na busca virtual. Estes filtros serão apresentados a seguir e a ordem escolhida representa, no geral, um aumento do custo computacional e da complexidade de cada filtro.

1.6. Filtros uni- ou bidimensionais

Como descrito anteriormente os filtros de seleção podem ser uni- ou bidimensionais baseando-se nas propriedades físico-químicas e/ou na conectividade dos átomos.

1.6.1. Propriedades *drug-like*

A ideia de usar propriedades “*drug-like*” (fármaco-similar) como filtro de seleção na busca virtual vem da observação de que fármacos apresentam certas propriedades físico-químicas que o distinguem dos demais compostos. Dentre as características desejadas para um fármaco podemos citar: ser solúvel em água e ser, de preferência, biodisponível por via oral.

Um dos trabalhos pioneiros no sentido de identificar propriedades físico-químicas relacionadas com estas propriedades “*drug-like*” foi feito por Lipinski e colaboradores em 1997. Neste trabalho (LIPINSKI et al., 1997), Lipinski analisou 2.245 fármacos que apresentavam uma boa biodisponibilidade via oral (absorção ou permeabilidade), utilizando quatro propriedades físico-químicas que já haviam sido associadas com medidas de permeabilidade, sendo elas: o massa molecular, o coeficiente de partição octanol/água calculado (usando o programa ClogP), o número de átomos doadores de ligação de hidrogênio (soma dos átomos O-H e N-H) e o número de átomos aceptores de ligação de hidrogênio (soma dos átomos O e N). Cerca de 90% dos compostos analisados apresentam massa molecular ($MM \leq 500$), $ClogP \leq 5$, número de átomos doadores de ligação de hidrogênio

(HBD) ≤ 5 e número de átomos aceptores de ligação de hidrogênio (HBA) ≤ 10 . Ainda, menos de 10% dos compostos analisados apresentam duas propriedades fora da faixa descrita acima. Baseado nestes resultados, Lipinski sugeriu que compostos que apresentem duas, ou mais, propriedades com valores fora destas faixas de corte têm grande chance de serem pouco permeáveis. A aplicação destas quatro propriedades, acompanhadas dos valores de corte descritos acima, é chamada de “Regra dos 5” de Lipinski. Desde sua publicação em 1997, o artigo de Lipinski foi citado mais de 4.900 vezes (segundo o Scopus). Atualmente, a “Regra dos 5” de Lipinski é amplamente difundida na área da química medicinal sendo usada rotineiramente nos protocolos de descoberta de fármacos (MEANWELL, 2011; MIGNANI et al., 2016). Uma série de outras regras para prever a biodisponibilidade oral foram derivadas da regra de Lipinski, sugerindo-se, no entanto, diferentes valores de corte de acordo com os tipos de compostos analisados. A Tabela 1.3 mostra algumas das regras derivadas da “Regra dos 5” de Lipinski (LIPINSKI et al., 1997).

Tabela 1.3. Algumas das regras derivadas da Regra dos 5 de Lipinski (LIPINSKI et al., 1997).

Regra	Crítérios	Predição da biodisponibilidade	Referência
Regra dos 5	$MM \leq 500$; $ClogP \leq 5$, $HBD \leq 5$ e $HBA \leq 10$	Fármacos	Lipinski (LIPINSKI et al., 1997)
Regra dos 4	$MM \leq 400$; $ClogP \leq 4$, $HBD \leq 4$ e $HBA \leq 8$	Compostos líderes	Oprea (TEAGUE et al., 1999)
Regra dos 3	$MM \leq 300$; $ClogP \leq 3$, $HBD \leq 3$ e $HBA \leq 6$	Fragmentos	(MAKARA, 2007)
Regra dos 2	$MM \leq 200$; $ClogP \leq 2$, $HBD \leq 2$ e $HBA \leq 4$	“Blocos de construção”	(GOLDBERG et al., 2015)

Os valores de corte descritos para os compostos líderes são menores quando comparados com os descritos para os fármacos (Tabela 1.3). Esta diferença se deve ao fato que durante a fase de otimização da atividade de um composto líder, existe uma tendência de aumento da complexidade (por exemplo: aumento do MM, aumento do número de anéis e de

ligações com rotação livre) e da lipofilicidade deste composto. Assim, visando-se evitar que os compostos obtidos após este processo tenham baixa biodisponibilidade oral, seus valores de corte são menores. Ainda, como esperado, quanto menor o composto, menor são os correspondentes valores de corte definidos. Além destas propriedades físico-químicas, outras foram analisadas e consideradas como importantes para prever a biodisponibilidade via oral de um fármaco, entre elas: o número de ligações com livre rotação (≤ 10) e a área de superfície polar ($\leq 140 \text{ \AA}^2$ ou $\leq 80 \text{ \AA}^2$), ambas sugeridas por Veber e colaboradores (2002b).

Além da biodisponibilidade via oral, outras propriedades “*drug-like*” relevantes podem ser incluídas nestes filtros, entre elas a solubilidade aquosa, a estabilidade metabólica e a toxicidade, sendo a solubilidade aquosa uma propriedade bastante desejada para um fármaco. No contexto da descoberta de fármacos, é mais importante garantir que o fármaco seja solúvel nas condições experimentais e/ou no sistema biológico do que determinar de forma exata seu grau de solubilidade. Além de ser importante para permitir a avaliação biológica dos fármacos, a solubilidade está associada à sua biodisponibilidade por via oral (LIPINSKI et al., 2001). No entanto, a dificuldade de se prever a solubilidade aquosa de uma molécula é inerente à complexidade dos fatores envolvidos neste processo. Citam-se, entre estes, as propriedades doadoras e/ouceptoras de ligação de hidrogênio do composto bem como da água e o custo energético necessário para romper a estrutura cristalina do estado sólido permitindo sua solvatação. As dificuldades inerentes em prever esta propriedade são um consenso na literatura (HILL; YOUNG, 2010; LIPINSKI, 2000; LIPINSKI et al., 2001; MANNHOLD et al., 2006). Assim, existem diversos modelos descritos na literatura (CHEVILLARD et al., 2012) baseados tanto em descritores moleculares³ bidimensionais ou tridimensionais, que variam desde simples índices até modelos de *QSPR* (*quantitative*

³ Descritores moleculares são termos que caracterizam um aspecto específico de uma molécula. (SANT’ANNA, 2002)

structure propriety relationships). Nenhum destes modelos é capaz de prever com segurança a solubilidade aquosa, considerando todo o espaço químico conhecido. Entre os modelos de predição de solubilidade aquosa descritos na literatura pode-se citar o “*Solubility Forecast Index - SFI*” (HILL; YOUNG, 2010) e o *Soly* (MANNHOLD et al., 2006), calculado pelo programa *Volsurf+* (CRUCIANI et al., 2000).

O *Solubility Forecast Index – SFI* foi proposto por Hill e Young (2010), usando 36.610 compostos disponíveis na GSK. Estes foram divididos em três grupos segundo seu grau de solubilidade aquosa medido experimentalmente: compostos com baixa solubilidade (<30 µM), compostos com solubilidade intermediária (30 – 200 µM) e compostos com alta solubilidade (>200 µM). Os pesquisadores observaram uma boa relação entre a solubilidade experimental e a predita usando o logaritmo da solubilidade com o coeficiente de partição no sistema octanol/água (logP) e o ponto de fusão. Do ponto de vista de um químico medicinal prever o logP é relativamente simples, no entanto, a predição do ponto de fusão é tão complexa quanto a predição da solubilidade. Portanto, embora os valores calculados por este modelo fossem próximos aos determinados experimentalmente, para usá-lo como modelo de predição de solubilidade seria mais útil encontrar outra propriedade que pudesse substituir o ponto de fusão. Após análise, os autores propuseram que a solubilidade aquosa poderia ser descrita usando os valores de $\log D_{pH7,4}$ calculado (que inclui a contribuição da ionização para partição) e o número de anéis aromáticos, dando origem ao *SFI* expresso pela equação:

$$SFI = clogD_{pH7,4} + \#Ar$$

onde, *SFI* é o *Solubility Forecast Index*; $clogD_{pH7,4}$ é o valor do coeficiente de distribuição em pH 7,4 calculado com o programa ACDLabs e *#Ar* é o número de anéis aromáticos.

Analisando os valores de *SFI* obtidos para os compostos com baixa, intermediária e alta solubilidade, os autores sugeriram que compostos com valores de *SFI* > 5 teriam maior probabilidade de apresentarem solubilidade aquosa baixa ou intermediária (HILL; YOUNG,

2010). Assim, este modelo não calcula o grau de solubilidade de um composto, mas o classifica segundo a probabilidade de apresentar, ou não, uma alta solubilidade aquosa. Embora as formas gráficas usadas para apresentar e comprovar as relações entre solubilidade e o *SFI* tenham sido criticadas na literatura (KENNY; MONTANARI, 2013), o uso do *SFI* como critério para selecionar compostos com maior probabilidade de apresentar alta solubilidade é interessante devido à simplicidade deste modelo.

Diferentemente do *SFI*, o modelo de solubilidade aquosa de Mannhold (MANNHOLD et al., 2006) calcula o grau de solubilidade (denominado *Soly*). Outra diferença entre estes modelos é que o último foi obtido usando um modelo de *PLS*⁴ e campos de interação molecular (explicados em detalhes no item 1.7.1). O modelo de *PLS* foi construído usando 1.028 moléculas com valores de solubilidade aquosa determinados experimentalmente, apresentando os seguintes parâmetros estatísticos: $r^2 = 0,74$; $q^2 = 0,73$ (predição interna do modelo) e erro de predição externa de $\sim 0,7$ em unidades de log. Devido ao erro de predição externa, os autores concluem que o modelo não deve ser usado para prever a solubilidade de compostos muito similares, podendo, no entanto, ser usado como filtro para remover compostos com valores de *Soly* abaixo de um determinado corte (MANNHOLD et al., 2006). Entretanto, nenhum valor de corte é sugerido. Assim, mesmo sendo mais complexo o modelo Mannhold (MANNHOLD et al., 2006), também deve ser usado de forma qualitativa visando eliminar compostos preditos com insolúveis

Uma das críticas para a aplicação de filtros de propriedades “*drug-like*” na construção de bancos de dados para serem usados na busca virtual é que estes reduzem o espaço químico⁵

⁴ Mínimos quadrados parciais (do inglês *Partial Least Squares – PLS*) é um método de regressão multivariada generalizada que usa projeções para resumir um grande número de variáveis potencialmente colineares (SANT’ANNA, 2002).

⁵ Espaço multidimensional ocupado por todos os compostos químicos.

levando a possível diminuição na diversidade do banco (HARVEY; EDRADA-EBEL; QUINN, 2015). Neste mesmo sentido, outra crítica feita à aplicação destes filtros é a recorrente observação de que alguns alvos biológicos, ainda não explorados, têm grandes sítios de ligação bastante lipofílicos ou bastante polares. O tamanho e as características destes sítios sugerem que a busca de *hits* e/ou compostos líderes deva ser feita usando compostos fora do “espaço químico de Lipinski” (DOAK et al., 2016). Além disso, algumas classes de compostos (ex. antibióticos, antifúngicos e produtos naturais) não estão dentro do espaço químico definido pela regra de Lipinski, entretanto, apresentam uma boa biodisponibilidade oral (LIPINSKI et al., 1997). Estas classes de compostos têm características estruturais que as tornam substratos de transportadores naturais. Embora algumas vezes criticadas, a aplicação de filtros de propriedades “*drug-like*” é, ainda, bastante recomendada durante as etapas de desenvolvimento de um fármaco (MEANWELL, 2011; MIGNANI et al., 2016).

1.6.2. Propriedades estruturais

A aplicação de filtros de propriedades estruturais tem como objetivo remover compostos com características estruturais indesejadas em um fármaco (BAELL; WALTERS, 2014), sendo comumente denominados de “*garbage filters*” (BOLOGA; OPREA, 2012; KUBINYI, 2004b). Algumas destas características já foram apresentadas e discutidas no item 1.3, pois são as responsáveis por resultados falso-positivos nos ensaios *in vitro*. Dentre elas, a presença de grupos reativos (BOLOGA; OPREA, 2012) bem como de subestruturas já reconhecidas por interferir nestes ensaios (BAELL; HOLLOWAY, 2010). Atualmente, estes filtros estão disponíveis em vários programas para preparação de banco de dados de moléculas virtuais. Entre eles, cita-se o programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]), que contém vários filtros estruturais, incluindo os relativos aos grupos funcionais reativos apresentados na Figura 1.4.

1.6.3. Similaridade 2D

O conceito de similaridade e dissimilaridade química depende da propriedade e/ou do critério usado para comparação (KUBINYI, 1998). Estes filtros se baseiam no conceito de que moléculas semelhantes apresentam atividades biológicas semelhantes. A busca por similaridade usando a estrutura 2D é uma abordagem *LBDD*. Nesta busca, as moléculas são descritas na forma de um conjunto de caracteres denominado *fingerprint* molecular.

Os *Fingerprints* moleculares consistem em uma série de *bits* (0 ou 1), representando a ausência (0) e a presença (1) de uma determinada subestrutura/fragmento em uma molécula. Estes *fingerprints* moleculares podem ser divididos em quatro tipos (WILLETT, 2011):

- Baseados em subestruturas chaves: os *bits* representam a presença, ou a ausência, de determinada subestrutura presente numa lista de subestruturas chaves (Figura 1.8(A)). Este tipo de *fingerprint* só é útil para comparar moléculas que apresentam as subestruturas chave. O número de *bits* é igual ao número de subestruturas chave e cada posição de *bit* representa uma única subestrutura. Exemplos destes *fingerprints* são o *Molecular ACCess System* (MACCS) e o *Barnard Chemical Information Ltd. (BCI)*.
- Topológicos ou *path-based*: cada molécula é analisada e dividida em fragmentos, geralmente, lineares. Em seguida, para cada fragmento se atribui um número *hash* que é usado para gerar o *fingerprint* (Figura 1.8(B)). Desta forma, qualquer molécula irá produzir um *fingerprint* e cada *bit* pode ser ocupado por mais de uma subestrutura. Exemplos destes *fingerprints* são o *The Daylight fingerprint* e o *FP2* do *Open Babel* (O'BOYLE et al., 2011).
- Circulares: estes *fingerprints* são também topológicos e gerados usando número *hash*. No entanto, ao invés de fragmentos lineares, os ambientes de cada átomo são determinados através de esferas. Por isso, estes *fingerprints* não são apropriados para serem usados como subestruturas de referência. Exemplos deste tipo de *fingerprint* são o *Extended-Connectivity Fingerprints (ECFPs)* e o *Functional-Class Fingerprints (FCFP)*.

- Híbridos: que são a combinação de mais de um tipo de *fingerprint*. Como exemplo temos o *UNITY 2D* e o *MP-MFP*.

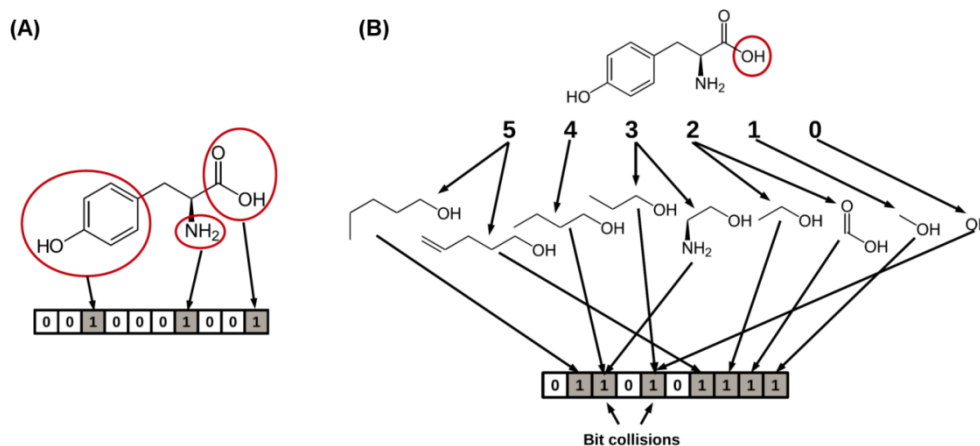


Figura 1.8. Representação de *fingerprints* moleculares obtidos usando a abordagem baseada em subestruturas chaves (A) e abordagem topológica ou *path-based* (B) (CERETO-MASSAGUÉ et al., 2015).

Os *fingerprints* gerados podem ser comparados usando diferentes medidas de similaridade, sendo o coeficiente de Tanimoto (T_c) a mais comumente empregada (CERETO-MASSAGUÉ et al., 2015; WILLETT, 2011), sendo representada pela equação:

$$T_c = \frac{c}{a + b - c}$$

onde, T_c é o valor do coeficiente de Tanimoto entre duas moléculas A e B, “a” é o número de *bits* iguais a 1 na molécula A, “b” é o número de *bits* iguais a 1 na molécula B e “c” é o número de *bits* iguais a 1 nas moléculas A e B.

O valor do T_c varia de 0 a 1, com 0 significando nenhuma sobreposição entre os *fingerprints* e 1 significando uma completa sobreposição dos *fingerprints*. A magnitude do valor do T_c depende do número de *bits* presentes no *fingerprints*, assim quanto maior a molécula de referência, maiores serão os valores de T_c . Numa busca por similaridade, como valor de corte do T_c , sugere-se escolher aquele capaz de diferenciar compostos ativos dos inativos (denominados, geralmente, de *decoys*) (STUMPFE; BAJORATH, 2011).

Algumas limitações em buscas por similaridade 2D citam-se: a presença de “*activity cliffs*”, ou seja, compostos similares com atividades biológicas distintas (STUMPFE et al.,

2014); o fato de compostos semelhantes apresentarem diferentes modos de ligação (KUHNERT; DIEDERICH, 2016); a escolha do descritor molecular e a molécula de referência, pois nem todas as regiões da molécula são relevantes para atividade biológica. Por fim, outra limitação bastante importante é não usar a conformação 3D da molécula (CERETO-MASSAGUÉ et al., 2015).

1.7. Filtros tridimensionais

Os filtros tridimensionais consideram a distribuição espacial das propriedades e/ou a estrutura 3D de cada um dos compostos, ligante ou molécula. Exemplos destes filtros são os farmacofóricos e os baseados no formato (“*shape*”).

1.7.1. Filtros farmacofóricos

Os filtros farmacofóricos selecionam compostos a partir de modelos farmacofóricos. Por definição, farmacóforo é um arranjo espacial das propriedades estéricas e eletrônicas necessárias para garantir interações supramoleculares (não-covalentes) ótimas entre um ligante e seu alvo biológico específico, e, assim, ativar (ou inibir) uma resposta biológica (SIPPL, 2008). O farmacóforo é um conceito abstrato para considerar as características moleculares comuns em uma série de ligantes que são cruciais para sua interação com o alvo biológico. Assim, um farmacóforo não é uma molécula real tão pouco uma associação de grupos funcionais (SIPPL, 2008).

Um modelo farmacofórico é um conjunto de regiões de interação (características químicas) alinhadas no espaço tridimensional. A natureza destas regiões de interação corresponde à das interações não covalentes ligante-receptor mais comumente observadas na química medicinal (ex. ligação de hidrogênio, eletrostática, hidrofóbica, cátion- π e π - π mostradas na Tabela 1.1). Estes modelos podem ser construídos usando informações da estrutura 3D do alvo biológico (abordagem *SDBB*) e/ou a estrutura 3D de uma série de

ligantes com atividade biológica (abordagem *LBDD*) (STUMPFE; BAJORATH, 2011; WIEDER et al., 2017; WOLBER; LANGER, 2005).

Quando os modelos são construídos com base na estrutura 3D do complexo RL, a identificação das regiões de interação importantes para a atividade biológica é feita de modo intuitivo analisando aquelas presentes no complexo (Figura 1.9 (A)). Por outro lado, quando os modelos são construídos com base na estrutura de um ligante, ou de uma série de ligantes, a identificação destas regiões é mais complexa. Neste último caso, a obtenção da conformação bioativa e o alinhamento da série de compostos são etapas cruciais para a construção do modelo. Não existe um número ideal de regiões de interação para se propor um modelo farmacofórico. Acredita-se que quanto maior o número de regiões de interação menos *hits* serão selecionados (WIEDER et al., 2017). Modelos com muitas regiões de interação (ex. ≤ 7) selecionaram um número pequeno de *hits*, que terão uma grande chance de serem similares a estrutura de referência. Por outro lado, modelos com poucas regiões de interação (ex. ≤ 3) irão selecionar *hits* inespecíficos sendo, portanto, de pouca utilidade para busca virtual (WIEDER et al., 2017).

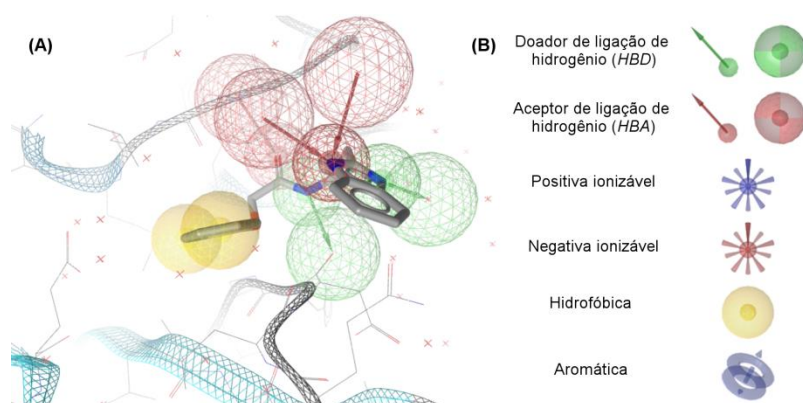


Figura 1.9. Representação de um modelo farmacofórico típico gerado a partir do complexo ligante-receptor: (A) Modelo farmacofórico gerado pelo programa *LigandScout* (WOLBER; LANGER, 2005) a partir do complexo da inibidor-cruzaína (cisteína protease do *Trypanosoma cruzi*) obtido diretamente do *PDB* (código 3KKU, resolução 1,28 Å (FERREIRA et al., 2010)) e (B) alguns dos tipos de interação ligante-receptor reconhecidas pelo *LigandScout* (WOLBER; LANGER, 2005).

Independente da forma como o modelo farmacofórico tenha sido gerado, este deve ser capaz tanto de selecionar compostos ativos bem como de diferenciar entre os ativos e os

inativos (*decoys*). Assim, a avaliação do desempenho do modelo farmacofórico gerado é bastante importante para o sucesso da busca virtual. Além de se contar o número de compostos ativos e inativos selecionados por um determinado modelo farmacofórico, na literatura são descritos outros critérios para se avaliar a qualidade dos modelos gerados (MCGANN; NICHOLLS; ENYEDY, 2015). Os mais comumente descritos são o fator de enriquecimento e a área sobre a curva (AUC, do inglês *area under the curve*), ambos obtidos a partir da curva *ROC* (do inglês *Receiver Operator Characteristic*, Figura 1.10), que mostra a fração de compostos ativos *versus* a fração de compostos inativos para um dado valor de escore (medida do ajuste do composto ao modelo farmacofórico) partindo-se dos compostos com melhor classificação.

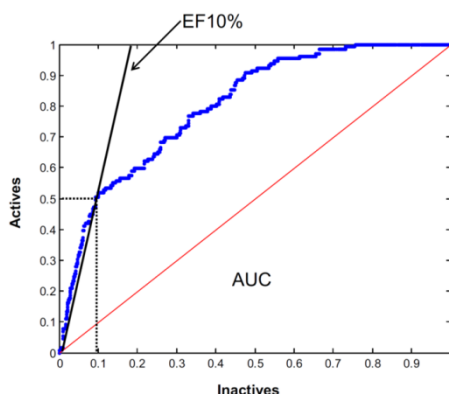


Figura 1.10. Curva ROC típica mostrando a obtenção da AUC e do fator de enriquecimento (*EF*) de 10 %.

Nesta curva, cada ponto representa a fração de ativos/inativos selecionados para um dado valor de escore, partindo-se do maior valor de escore até o menor valor de escore. A linha vermelha representa uma curva com AUC igual a 0,5 (esperada se os compostos fossem selecionados aleatoriamente). Neste exemplo, o valor de $EF_{10\%}$ é igual a 5 (0,5/0,1). (MCGANN; NICHOLLS; ENYEDY, 2015)

O fator de enriquecimento (*EF*) é a fração de compostos ativos (f_a) obtidos quando uma determinada porcentagem de compostos inativos é encontrada (f_i), ou seja, $EF = f_a/f_i$. Assim, o *EF* depende do valor da porcentagem de compostos inativos considerado ($f_i = 1\%$, 5% , 10%), variando de 0 a $1/f_i$ de modo que um modelo perfeito apresentaria um valor de $EF = 1/f_i$, enquanto que um modelo que seleciona compostos ativos e inativos de modo aleatório teria um valor de $EF = 1$. O valor do AUC representa a probabilidade de um composto ativo receber um valor de escore maior que um composto inativo. Este valor varia de 0 a 1, sendo 1

o valor esperado para um modelo perfeito. Curvas com $AUC = 0,5$ indicam que os compostos ativos e inativos são selecionados de forma aleatória. Estas métricas podem ser usadas para analisar a qualidade de outros filtros usados na busca virtual que serão apresentados nos próximos itens (MCGANN; NICHOLLS; ENYEDY, 2015).

Existem diversos programas, comercialmente disponíveis, que permitem reconhecer e gerar automaticamente modelos farmacofóricos, entre eles: *CATALYST* (“Catalyst”, 2005), *FLAP* (BARONI et al., 2007), *LigandScout* (WOLBER; LANGER, 2005) e *MOE* (CHEMICALCOMPUTINGGROUPINC., 2004). Todos estes podem gerar modelos farmacofóricos partindo tanto da estrutura do receptor como do ligante.

Na literatura são encontrados vários exemplos de sucesso do uso de modelos farmacofóricos na busca virtual, permitindo identificar compostos ativos frente a diferentes alvos biológicos (MALVEZZI et al., 2011; MALVEZZI; DE REZENDE; AMARAL, 2009; RECHFELD et al., 2014; WIEDER et al., 2017). Além dos compostos selecionados apresentarem atividade frente ao alvo de interesse, alguns deles têm características estruturais bastante diferentes daquelas observadas nos compostos usados para construir os modelos.

Assim, um exemplo interessante foi observado nos estudos realizados por Rechfeld e colaboradores (2014). Estes pesquisadores buscavam inibidores da interação PKC ϵ /RACK2 (interação proteína-proteína entre a proteína quinase C com seu o receptor de ativação da PKC). No entanto, a ausência de informações sobre inibidores não peptídicos da interação PKC ϵ /RACK2 bem como as características do sítio ativo da PKC ϵ (grande, raso, exposto ao solvente e flexível) tornam esta tarefa bastante desafiadora. Como estratégia, os pesquisadores construíram um modelo farmacofórico baseado nas interações entre a PKC ϵ e o peptídeo EAVSLKPT (inibidor da interação PKC ϵ /RACK2) observadas na estrutura cristalográfica (Figura 1.11). Este modelo foi aplicado a um banco de dados contendo 300.000 moléculas, selecionando 468 (0,2 %) delas. A seleção final foi feita tanto pela classificação do ajuste

geométrico de cada molécula ao modelo bem como por inspeção visual. Por fim, 19 (4,0 %) compostos foram selecionados e adquiridos e testados *in vitro*. Um deles apresentou atividade inibitória significativa ($IC_{50} = 25 \mu M$, compostos número **1** da Figura 1.11). O composto número **1** foi usado como estrutura fundamental para a síntese ou compra de uma série de 19 tienoquinolinas. Vários compostos desta série apresentaram atividade biológica, sendo que o mais ativo apresentou $IC_{50} = 5,9 \mu M$ (compostos número **8** da Figura 1.11) (RECHFELD et al., 2014). Assim, este estudo mostra como modelos farmacofóricos podem ser usados com sucesso na identificação de compostos bioativos mesmo quando o alvo é considerado não *druggable*⁶ e quando as informações sobre outros inibidores são limitadas.

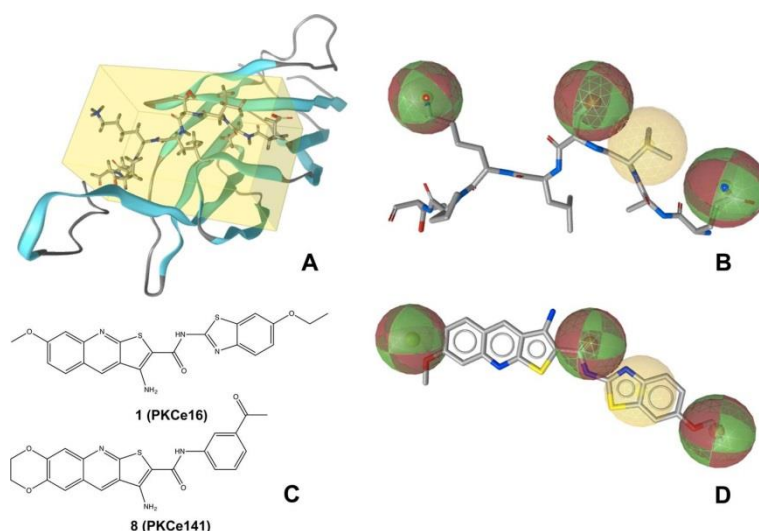


Figura 1.11. Representação do modelo farmacofórico gerado a partir das interações entre a PKCε e o peptídeo EAVSLKPT e da estrutura de dois inibidores identificados por Rechfeld e colaboradores (2014). **(A)** Representação do sítio de interação entre o domínio C da PKCε e o peptídeo EAVSLKPT (PDB 1GMI). **(B)** Modelo farmacofórico gerado pelo programa *LigandScout* (WOLBER; LANGER, 2005) a partir do complexo 1GMI. **(C)** Estrutura de dois inibidores identificados pelo estudo (número **1** e **8**), sendo o número **1** selecionado pelo modelo farmacofórico e o número **8** obtido usando a estrutura de **1** como base. **(D)** Composto número **1** sobreposto ao modelo farmacofórico gerado (RECHFELD et al., 2014).

Modelos farmacofóricos também podem ser gerados usando apenas a estrutura do receptor (ou seja, sem qualquer ligante) (HU; LILL, 2013). Neste caso, não é necessário o

⁶ O termo “*druggability*” se refere à capacidade de um alvo molecular ser modulado de forma favorável (ex. ativado ou inibido) por uma molécula pequena. Estima-se que apenas 10% do genoma humano corresponde a alvos “*druggables*”, portanto, a grande dos alvos seria “não *druggable*” (BUCKLE et al., 2014).

conhecimento prévio de ligantes. Outra vantagem destes modelos é que eles não são restritos ao espaço químico ocupados pelos compostos já descritos como ligantes de um determinado alvo. Nos modelos farmacofóricos, gerados deste modo, as regiões de interação são obtidas mapeando a superfície do sítio de ligação do receptor alvo. Um dos métodos mais comuns de se obter estes modelos é através dos cálculos dos campos de interação molecular (do inglês *Molecular Interaction Fields - MIFs*) usando sondas moleculares de diferentes naturezas (doadora e/ou aceptora de ligação de hidrogênio, hidrofóbica e aromática). Estes campos de interação moleculares podem ser calculados usando, por exemplo, o programa GRID (GOODFORD, 1985). Durante este cálculo o sítio de ligação do receptor alvo é colocado numa caixa com uma grade regularmente espaçada. Em seguida, sondas moleculares (representando átomos ou moléculas de prova) são posicionadas nos vértices desta grade e valores de energia de interação entre a sonda e o receptor são calculados. Os valores de energia de interação calculados são usados para identificar regiões de interação favoráveis (“*hot spots*” (GOHLKE; HENDLICH; KLEBE, 2000)) ao redor da estrutura do receptor biológico. Estes pontos de interação favoráveis podem ser usados para construir um modelo farmacofórico baseado na estrutura do alvo biológico (HU; LILL, 2013).

Como já apresentado, atualmente, além de se utilizar a estrutura 3D do alvo biológico, a sua flexibilidade vem sendo cada vez mais incluída nas estratégias de *SBDD*. Uma forma de incluir a flexibilidade conformacional nos filtros farmacofóricos é gerar modelos farmacofóricos usando diferentes conformações do complexo ligante-receptor. Estas conformações podem ser obtidas experimentalmente (por cristalografia de raios-X ou por RMN) ou computacionalmente (ex. simulações de DM clássicas) (CHOUDHURY; PRIYAKUMAR; SASTRY, 2015; PICCIRILLO et al., 2016).

A importância de se considerar a flexibilidade conformacional foi analisada por Choudhury, Priyakumar e Sastry em 2015, usando como alvo a ciclopropano sintase da

Mycobacterium tuberculosis (CmaA1). Cinco diferentes simulações de DM foram realizadas usando a estrutura da CmaA1 (código PDB 1KPH) complexada com um, ou mais, ligante(s), representando diferentes estágios do ciclo catalítico desta enzima. Cada simulação de DM teve duração de 40 ns, sendo selecionado para cada uma delas um *snapshot* a cada 5 ns. Assim, 40 *snapshots* representando diferentes conformações foram obtidos e usados para gerar modelos farmacofóricos. Ainda, um modelo farmacofórico foi gerado a partir da estrutura cristalográfica (1KPH). Em seguida, todos os modelos foram avaliados usando uma série de 23 compostos ativos frente à esta enzima e 1.398 *decoys*. Como resultado, os autores observaram que os modelos farmacofóricos gerados usando os *snapshots* obtidos nas simulações de DM tiveram um melhor desempenho do que o gerado usando a estrutura cristalográfica. Uma crítica comum deste tipo de abordagem é o pequeno número de *snapshots* usados como representantes das diferentes conformações observadas ao longo de toda a simulação (WIEDER et al., 2017). Além disso, nenhum critério foi usado para escolher estes *snapshots*, sendo esta uma etapa crítica (DE VIVO et al., 2016).

Mais recentemente, Wieder e colaboradores (WIEDER et al., 2017) propuseram uma estratégia para incluir a flexibilidade na criação dos modelos farmacofóricos que não depende da prévia seleção de alguns *snapshots* representativos. Nesta estratégia, todos os *snapshots* obtidos durante a simulação de DM são usados para gerar modelos farmacofóricos. Estes são agrupados de acordo com sua semelhança. Para cada grupo, o valor da energia do ligante em cada um das estruturas presentes neste grupo é calculado. A seguir, o modelo farmacofórico que corresponde à estrutura com valor de energia médio é escolhido como representativo do correspondente grupo. Por fim, os modelos farmacofóricos representativos obtidos para um determinado alvo são usados na busca virtual. Deste modo, para cada alvo várias listas de *hits* são obtidas (uma para cada modelo representativo). Estes *hits* são agrupados gerando uma nova lista na qual cada *hit* é contado uma única vez. A classificação dos *hits* é, então, dada

pelo número de vezes que cada *hit* foi selecionado. A validade desta estratégia foi analisada usando 40 complexos ligante-receptor obtidos do *PDB*. Para cada alvo, foi usado um conjunto de compostos ativos e de compostos preditos como inativos obtidos do banco DUD-E. Como resultado, os autores observaram que o uso desta estratégia permitiu encontrar modelos melhores quando comparados aos usando as estruturas do *PDB* (WIEDER et al., 2017).

A aplicação de filtros farmacofóricos independentemente de se incluir a flexibilidade conformacional do receptor é uma estratégia de sucesso no caminho para a descoberta de novos compostos ativos.

1.7.2. Filtros baseados no formato (“*shape*”)

Os filtros baseados no formato (“*shape*”) selecionam moléculas com base na sua similaridade de “formato”. O formato pode ser definido como coincidência de volumes ou de superfícies. A ideia de se incluir o formato na busca virtual vem do fato de que se duas moléculas tendo um formato similar, talvez possam apresentar propriedades similares (NICHOLLS et al., 2010). O conceito de similaridade baseada no formato não é novo, no entanto, estes passaram a ser usados rotineiramente na busca por novos compostos ativos a partir do surgimento de ferramentas computacionais de busca em larga escala.

De modo análogo ao descrito para os filtros farmacofóricos, existem vários algoritmos de busca baseados no formato. Entretanto, um destaque especial é dado para o programa *ROCS* (HAWKINS; SKILLMAN; NICHOLLS, 2007), pois este é, ainda, reconhecido como o estado da arte na busca baseada no formato (KIRCHMAIR et al., 2009).

Além do formato, o programa *ROCS* (HAWKINS; SKILLMAN; NICHOLLS, 2007) permite incluir informações sobre algumas características químicas na busca virtual, representando átomos ou grupos funcionais como anéis, átomos doadores ou aceptores de ligação de hidrogênio. Desta forma, diferentes tipos de medidas de similaridade podem ser calculadas baseadas: no formato (*Shape Tanimoto*), nas características químicas (*Color*

Tanimoto) e na combinação dos dois (*Combo Tanimoto*). A combinação do formato e dos padrões químicos pode melhorar significativamente o desempenho da busca virtual. De modo análogo ao descrito para os modelos farmacofóricos, para se realizar a busca por similaridade de formato é preciso fornecer uma determinada estrutura 3D da molécula de referência. No entanto, diferentemente dos modelos farmacofóricos baseados no ligante, esta busca parece ser menos sensível à conformação da molécula de referência apresentando, portanto, uma vantagem em relação aos modelos farmacofóricos. Nesta abordagem, torna-se mais importante a escolha da molécula de referência do que a obtenção da conformação bioativa (KIRCHMAIR et al., 2009).

1.8. Filtros de ancoramento

O ancoramento (do inglês *docking*) consiste em posicionar um ligante dentro do seu sítio de ligação e, em seguida, estimar sua afinidade pela proteína alvo (KITCHEN et al., 2004). A aplicação desta estratégia para bibliotecas de compostos permite a seleção de acordo com seu valor de afinidade por determinada proteína alvo. Os estudos de ancoramento podem ser divididos em duas etapas, a primeira envolve a predição correta da pose do ligante (ou seja, sua conformação e orientação dentro do sítio de ligação) e a segunda envolve a predição correta da afinidade do ligante. As duas etapas são complexas e apresentam limitações.

Para que a predição da pose do ligante (ou seja, seu modo de ligação) possa ser feita de forma correta é necessário considerar todos os graus de liberdade conformacionais do ligante (translacional, rotacional e conformacional) de modo rápido e eficiente. Assim, quanto maior e mais flexível o ligante, mais difícil deverá ser prever de forma correta seu modo de ligação. A varredura do espaço conformacional do ligante pode ser agrupada em quatro métodos:

- Baseados em campos de força: usam simulações de DM ou de Monte Carlo (MC) para explorar o espaço conformacional do ligante, buscando-se regiões de mínimos de energia. Um

dos problemas destes métodos é que os ligantes podem ficar “presos” em mínimos locais, caso os tempos de simulação sejam muito curtos. Um exemplo de programa que usa simulações de MC é o *GLIDE* (HALGREN et al., 2004).

- Métodos evolutivos (algoritmo genético): fazem uma busca aleatória do espaço conformacional diminuindo, assim, a probabilidade do ligante ficar preso em mínimos locais. Estes se baseiam na teoria da evolução de Darwin, na qual, uma população de cromossomos (poses do ligante) é gerada. Estes são representados por genes (ângulos diedros e vetores rotacionais e translacionais) que evoluem por mutação de um gene e/ou cruzamento entre dois genes. Por fim, as populações com melhor escore são usadas para gerar novas populações por reprodução. Este processo se repete até que um determinado critério seja atingido. Normalmente, os programas que adotam estes métodos têm melhores desempenhos quando os resultados obtidos são submetidos a uma etapa adicional de minimização, para a localização do mínimo global. Um exemplo de programa que usa algoritmo genético é o *GOLD* (VERDONK et al., 2003).

- Métodos de construção incremental: fazem uma busca sistemática visando a explorar todos os graus de liberdade de um determinado ligante dividindo-o em fragmentos. Em seguida, um fragmento de referência é ancorado usando a complementaridade geométrica. Este fragmento vai sendo, então, repetidamente incrementado até reconstruir totalmente o ligante. Este processo pode ser agilizado considerando apenas alguns ângulos diedros. No entanto, se na conformação bioativa o ligante apresentar ângulos torcionais não usuais, dificilmente estes métodos conseguirão encontrar a pose correta. Um exemplo de programa que usa a ancoragem incremental é o *FlexX* (RAREY et al., 1996).

- Métodos de complementaridade de forma: estes métodos procuram por uma complementaridade estrutural entre as formas do ligante e da correspondente na cavidade da proteína. A flexibilidade do ligante é incluída usando um conjunto de conformações

previamente geradas. Assim, o ancoramento é rígido, sendo apenas rotacionado e translacionado dentro do sítio ativo da proteína. Um problema deste método é que dificilmente a verdadeira conformação bioativa será encontrada dentre as previamente geradas. No entanto, uma vez que as conformações foram geradas, o ancoramento se torna mais rápido. Exemplos de programas que se utilizam deste método são o *DOCK* (MOUSTAKAS et al., 2006) e o *FRED* (SCHULZ-GASCH; STAHL, 2003).

O modo de ligação observado no complexo ligante-receptor obtido por cristalografia de raios-X é considerado, dentro do erro experimental, como uma das possíveis conformações bioativas do ligante. Assim, o desempenho dos programas de ancoramento quanto à obtenção da pose é, normalmente, avaliado observando-se quanto os resultados obtidos reproduzem estes dados experimentais. Em geral, todos os programas de ancoramento, usando qualquer um dos quatro métodos de varredura do espaço conformacional do ligante descritos acima, são capazes de reproduzir satisfatoriamente (com valores de $\text{RMSD}^7 \leq 2\text{\AA}$) as poses dos ligantes observadas nos complexos obtidos por cristalografia (CARLSON et al., 2016). Este tipo de avaliação de desempenho é chamado de *re-docking* (HAWKINS et al., 2008).

A segunda etapa do ancoramento é prever e classificar a afinidade de diferentes poses de um mesmo ligante, e/ou de diferentes ligante(s), por determinada proteína alvo. Esta etapa é fundamental para o sucesso do ancoramento, pois o programa deve ser capaz de diferenciar as poses corretas, das incorretas, bem como os compostos ativos, dos inativos. Como já apresentado anteriormente, métodos de cálculo de energia livre de *Gibbs* usando mecânica estatística foram desenvolvidos para modelar as interações proteína-ligante e, assim, prever afinidade desta ligação. Entretanto, estes métodos ainda são computacionalmente muito caros para serem usados como forma de avaliação de um grande número de complexos ligante-receptor. Como alternativa, os programas de ancoramento preveem a afinidade ligante-

⁷ RMSD (do inglês *Root Mean Square Distance*) é uma medida da distância média entre átomos.

receptor através de funções de ranqueamento. Estas funções fazem várias suposições e simplificações não considerando completamente todos os aspectos envolvidos no reconhecimento ligante-receptor, como, por exemplo, as contribuições entrópicas e a polarizabilidade. De modo geral, as funções de ranqueamento podem ser agrupadas em três classes:

- Funções baseadas em campos de força moleculares: calculam a energia da interação ligante-receptor usando campos de força moleculares da mecânica clássica (descritos em detalhes no item 1.10). A principal desvantagem destes métodos é não considerar os fatores entrópicos associados à interação ligante-receptor. Assim, deve-se ter cuidado para não superestimar as altas afinidades calculadas para moléculas grandes e/ou polares. Um exemplo de função de ranqueamento baseadas em campos de força é a *Goldscore* (KORB et al., 2012).

- Funções empíricas: calculam a energia da interação ligante-receptor usando equação complexa com vários termos, que descrevem propriedades reconhecidas como importantes para a interação ligante-receptor. Séries de treinamento e métodos de regressão multivariada são usados para otimizar os valores dos coeficientes que descrevem a contribuição de cada termo para a equação. Assim, é esperado que as equações obtidas para um conjunto de determinada proteína só possam ser aplicadas para proteínas similares às presentes na série de treinamento. Exemplos de programa que usam funções de ranqueamento empírica são o *FRED* (MCGANN, 2011) e o *GOLD* (VERDONK et al., 2003).

- Funções baseadas no conhecimento de potenciais médios de força: calculam a energia da interação ligante-receptor usando a soma das energias livre de interação de todos os pares de átomos ligante-receptor. Estas energias são calculadas considerando os valores de distância, entre dois pares de átomos, observados em complexos RL obtidos por cristalografia. As vantagens deste método é que não são necessárias séries de testes e que os termos entrópicos e relacionados à dessolvatação são incluídos implicitamente. No entanto, a

definição arbitrária dos tipos de átomos envolvidos na interação é um dos fatores que limita a aplicabilidade destas funções de ranqueamento. Um exemplo de programa que usa as funções baseadas no conhecimento é o *FlexX* (RAREY et al., 1996).

Estudos comparativos mostram que todas as funções de ranqueamento têm enorme dificuldade para identificar a pose mais correta, dentre todas as poses geradas, sendo bastante importante a inspeção visual dos resultados obtidos. Além disso, outra limitação das funções de ranqueamento é a predição absoluta da afinidade do ligante pelo seu alvo. Melhores resultados são obtidos quando se analisa a capacidade das funções classificarem de forma mais, ou menos correta, a atividade relativa de um conjunto de ligantes. A avaliação do desempenho destas funções de ranqueamento quanto à capacidade de diferenciar ativos, de inativos, pode ser feita usando as mesmas métricas usadas para avaliar o desempenho dos modelos farmacofóricos (item 1.7.1) (MCGANN; NICHOLLS; ENYEDY, 2015). No entanto, o desempenho das funções de ranqueamento está de longe de ser ideal e parece ser bastante dependente do alvo molecular usado (CARLSON et al., 2016; IRWIN; SHOICHET, 2016; KITCHEN et al., 2004; KLEBE, 2006; SCHOMBURG; RAREY, 2014; SMITH et al., 2016).

Além das limitações descritas, tanto os algoritmos usados para prever a pose bem como as funções de ranqueamento, outros aspectos relacionados ao insucesso desta abordagem são: (i) a escolha da estrutura tridimensional do alvo molecular a ser usada (baixa qualidade), (ii) a não inclusão da flexibilidade conformacional do alvo e (iii) a ausência de informação sobre águas estruturais. Estes aspectos serão brevemente discutidos a seguir:

- Escolha da estrutura tridimensional do alvo molecular: é uma etapa crucial em qualquer abordagem baseada na estrutura do receptor independentemente de como esta estrutura foi obtida (cristalografia de raios-X, RMN e/ou modelagem por homologia). Em qualquer um destes casos, o conjunto de coordenadas usado sempre será um modelo (WARREN et al., 2012). Assim, a qualidade destes modelos precisa ser verificada (DUNBAR

et al., 2013). Considerando, por exemplo, as estruturas obtidas por cristalografia de raios-X, alguns dos critérios que deveriam ser observados são: a resolução (medida teórica da menor distância observada no cristal, indicando a quantidade de dados coletados); o *R-factor* (medida da diferença entre os dados observados e os preditos usando o modelo, indicador da qualidade do modelo); *R_{free}* (coeficiente de validação cruzada do *R-factor* calculado a partir de dados não usados para construir o modelo); diferença entre *R-factor* e *R_{free}* (medida do *overfit* do modelo); contatos cristalinos com o ligante e/ou com o sítio de ligação e ainda a presença de regiões com densidade incompleta e/ou modelada incorretamente próxima ao ligante e/ou ao sítio de ligação. Pelo exposto, na análise da qualidade de uma estrutura 3D vários critérios deveriam ser analisados. No entanto, não é raro encontrar na literatura estudos *SBDD* usando estruturas de baixa qualidade (HAWKINS et al., 2008; WARREN et al., 2012).

- Inclusão da flexibilidade conformacional: até o momento, não é possível incluir todos os graus de liberdade conformacionais de um alvo molecular. Algumas alternativas para contornar esta limitação são descritas na literatura, entre elas: (i) atribuir flexibilidade para a cadeia lateral de alguns resíduos de aminoácidos usando uma biblioteca de rotâmeros e (ii) usar um conjunto de diferentes conformações do receptor, obtidas tanto experimentalmente quanto por meio de simulações de DM (DE VIVO et al., 2016; KITCHEN et al., 2004).

- Identificar águas estruturais: estas são moléculas de água que fazem ligações de hidrogênio com resíduos da superfície do receptor em uma geometria perfeita, podendo intermediar interações ligante-receptor. Em qualquer interação ligante-receptor ocorre o deslocamento de moléculas de água nas superfícies polares do receptor. A dessolvatação de regiões polares tem um custo entálpico que deve ser compensado pela formação de ligações de hidrogênio no complexo ligante-receptor. No entanto, a perda entálpica associada ao deslocamento de moléculas de água estruturais dificilmente é compensada (BISSANTZ; KUHN; STAHL, 2010). Uma prática comum nas abordagens *SBDD*, incluindo o

ancoramento, é remover todas as moléculas de água presentes na estrutura. No entanto, moléculas de água estruturais deveriam ser mantidas. Exemplos da importância de se considerar moléculas de água, possivelmente estruturais, no ancoramento já foram descritos na literatura (NGUYEN et al., 2014). Nos últimos anos, apesar dos esforços feitos para desenvolver modelos capazes de identificar quais moléculas de água são (ex. *WaterMap* (ABEL et al., 2008)), ou não, estruturais, a identificação e a inclusão destas moléculas de água em estudos de *SBDD* é ainda um desafio (BETZ et al., 2016; MURPHY et al., 2016; NGUYEN et al., 2014; NITTINGER et al., 2015).

Os três aspectos descritos são importantes não só nos estudos de ancoramento, mas em qualquer abordagem *SBDD*. Embora os programas de ancoramento tenham várias limitações, seu uso como ferramenta na busca virtual resultou na identificação de um grande número de ligantes que foram, subsequentemente, validados tanto usando ensaios enzimáticos quanto determinando experimentalmente a estrutura 3D do complexo RL (IRWIN; SHOICHET, 2016). Ainda, o sucesso do uso deste filtro vai depender da qualidade do banco de moléculas virtuais, do tipo de alvo e dos critérios de escolha dos *hits*. Aspectos relacionados à qualidade do banco de dados serão apresentados e discutidos no item 1.9 desta Introdução. A aplicação do ancoramento a alvos moleculares com sítios de ligação amplos e rasos tem menor taxa de sucesso como, por exemplo, para algumas proteases e interações proteína-proteína (IRWIN; SHOICHET, 2016). Na literatura, sugere-se que se pode encontrar entre 2 ou 5 novas classes de compostos com atividade biológica usando o ancoramento, devendo-se testar pelo menos entre 25 a 50 compostos. No entanto, este número pode ser muito maior para proteases e, ainda, sem garantia de sucesso. Por fim, em relação à análise e seleção dos *hits* além do valor de escore sugere-se fazer uma etapa de inspeção visual (IRWIN; SHOICHET, 2016).

Outro exemplo de sucesso do uso do ancoramento na busca de novos ativadores do receptor de opioíde- μ (μ OR) partindo de um banco de dados de 3 milhões de compostos *lead-*

like (líderes similares) (MANGLIK et al., 2016). Estes compostos foram ancorados no sítio de ligação do receptor μ OR usando o programa *DOCK* (MOUSTAKAS et al., 2006). As 2.500 moléculas com melhor valor de escore foram analisadas visualmente considerando sua novidade estrutural, sua capacidade de formar interações com resíduos polares reconhecidos como importantes para formação do complexo ligante-receptor e, excluindo aquelas com conformações energeticamente desfavoráveis. Nesta etapa, 23 moléculas foram selecionadas para serem compradas e testadas frente ao receptor. Destas moléculas, 7 (30 %) mostraram K_i entre 2,3 e 14 μ M, sendo estruturalmente diferentes dos compostos já descritos como inibidores do receptor μ OR. Entretanto, os compostos inicialmente identificados precisavam ser melhorados quanto suas atividades (K_i na faixa de poucos μ M). Para isso, os autores usaram um conjunto de 500 compostos análogos aos mais ativos contendo as principais regiões para ligação com o receptor bem como outros grupos capazes de fazer novas interações com o receptor. Estes 500 análogos foram ancorados e os 15 com melhores escores foram selecionados e testados. Destes, 7 (47 %) foram ativos com K_i entre 42 nM e 4,7 μ M. O composto mais ativo (42 nM) foi, então, modificado por síntese orgânica, resultando em um composto bastante ativo com $K_i = 1$ nM (denominado PZM21). Adicionalmente, PZM21 se mostrou inativo frente a outros tipos de receptores de opioídes sendo, portanto, seletivo.

Outras aplicações do procedimento de ancoramento incluem os estudos de relação estrutura atividade de séries de compostos (FALSINI et al., 2017), a identificação de resultados falsos positivos em estudos de *HTS* (FERREIRA et al., 2010) e a predição da toxicidade e do metabolismo de fármacos (DI MARTINO et al., 2013).

1.9. Bancos de dados de moléculas virtuais

A preparação e/ou escolha do banco de dados de moléculas a ser usado na busca virtual é uma etapa crucial deste processo, sendo bastante importante para o sucesso desta abordagem. Existem vários bancos de dados de moléculas, gratuitos ou comerciais, que

podem ser usados na busca virtual como o ChEMBL (HERSEY, 2015), o ZINC (IRWIN et al., 2012), o PubChem (KIM et al., 2016), o PDBbind (WANG et al., 2004), o WOMBAT (OLAH et al., 2005), o DrugBank (WISHART, 2006), dentre outros. O formato das moléculas varia de acordo com o banco, por exemplo, uni (*SMILE*) ou tridimensionais (*Mol2* e *SDF*).

A sequência de etapas envolvidas na preparação do banco de dados depende de diversos fatores como o alvo molecular em estudo e o(s) tipo(s) de filtro(s) que será(ão) usado(s) na busca virtual (ex. similaridade 2D, farmacofórico, baseado no formato e ancoramento). Entretanto, algumas etapas são, normalmente, empregadas, sendo elas: remoção de compostos com grupos funcionais indesejados (“*garbage filters*”), remoção de compostos com propriedades físico-químicas indesejadas, correção das estruturas e obtenção de diferentes conformações. Cada uma destas etapas será brevemente discutida a seguir.

- Remoção de compostos com grupos funcionais indesejados. Os bancos de dados usados para busca virtual podem conter compostos com grupos funcionais e/ou subestruturas capazes de gerar resultados falso-positivos nos ensaios *in vitro* (ex. como os *PAINS*), compostos com grupos reativos frente a determinados alvos biológicos e compostos capazes de formar agregados. As características e as razões para estes resultados já foram discutidas com detalhes no item 1.3 desta Introdução. Considerando que os *hits* selecionados na busca virtual precisam ser necessariamente validados e que esta validação é, geralmente, feita através de ensaios *in vitro*, remover, já durante a preparação do banco, compostos que podem gerar resultados falso-positivos pode aumentar o sucesso na busca virtual. Esta etapa, geralmente, é denominada de “limpeza” (“*cleaning up*”) e os filtros usados podem ser chamados de “*garbage filters*” (vide item 1.6.2). A aplicação destes filtros só não é recomendada se o objetivo da busca virtual for selecionar compostos que atuem formando uma ligação covalente (MALVEZZI, 2008; MALVEZZI; DE REZENDE; AMARAL, 2009). No entanto, para maioria dos projetos de descoberta de fármacos o objetivo é encontrar

compostos que atuem por um mecanismo não covalente. Por esta razão, a aplicação desta etapa na preparação do banco de dados é bastante recomendada (BOLOGA; OPREA, 2012; MIGNANI et al., 2016; SCIOR et al., 2012).

- Remoção de compostos com propriedades físico-químicas indesejadas.

Diferentemente do descrito anteriormente, não há um consenso na literatura sobre a inclusão de filtros de propriedades “*drug-like*” (vide item 1.6.1) durante a etapa de preparação dos bancos de dados. Como argumento a favor de se incluir estes filtros é poder selecionar, já nas primeiras etapas, compostos com uma maior probabilidade de apresentar bons parâmetros ADMET (BOLOGA; OPREA, 2012; MEANWELL, 2011; MIGNANI et al., 2016; WARING et al., 2015). A principal crítica é que a inclusão destes filtros restringe o espaço químico, diminuindo a chance de se encontrar novas classes de compostos (BENET et al., 2016; DOAK et al., 2016). Outra crítica é que os critérios usados nestes filtros não se aplicam a determinadas classes de compostos (em especial, produtos naturais) bem como a certos tipos de alvo molecular (DOAK et al., 2016; LIPINSKI, 2016).

- Correção das estruturas. Esta etapa é fundamental durante a criação dos bancos de dados e não há críticas a respeito da sua inclusão. A etapa de correção de estruturas visa a garantir que as moléculas presentes no banco de dados estejam descritas de forma correta quanto a: sua quiralidade (quando conhecida); hibridização dos átomos nas ligações químicas; seu estado de protonação (definido para um determinado pH) e sua forma tautomérica. A relevância desta etapa é facilmente compreendida considerando, principalmente, os filtros baseados nas interações ligante-receptor (farmacofórico e ancoragem), pois o padrão de interações de uma molécula depende, por exemplo, de seu estado de protonação (ex. interação iônica) e/ou da sua forma tautomérica (ex. ligação de hidrogênio). Ainda, diferentes estereoisômeros apresentam uma distribuição espacial diferente que poderá modificar tanto sua

forma quanto o padrão de interação em três dimensões (BOLOGA; OPREA, 2012; MARTIN, 2009; SCIOR et al., 2012).

- Obtenção de diferentes conformações. Um dos maiores desafios dos filtros tridimensionais é gerar para cada molécula um conjunto de diferentes conformações que seja adequado para varrer todo seu espaço conformacional (BOLOGA; OPREA, 2012; HAWKINS; NICHOLLS, 2012; SCIOR et al., 2012). Esta etapa tenta garantir que seja possível encontrar a conformação bioativa dentre as conformações geradas. Existem vários programas capazes de gerar conformações 3D como o *Corina* (SADOWSKI; GASTEIGER, 1993), o *Open Babel* (O'BOYLE et al., 2011), o *Omega* (HAWKINS et al., 2010), entre outros. No entanto, nenhum deles é capaz de gerar, em todos os casos, conformações próximas às bioativas. O número de conformações necessárias para se varrer o espaço conformacional vai depender da flexibilidade de cada molécula (medida geralmente pelo número de ligações com livre rotação), sendo maior para moléculas mais flexíveis. Assim, num primeiro momento, poderíamos supor que o ideal seria gerar o maior número de possíveis conformações. Entretanto, quanto maior o número de conformações geradas, maior o espaço em disco necessário para armazenar o banco de dados bem como o custo computacional de se aplicar os filtros de seleção (SCIOR et al., 2012). Não existe, dentro do nosso conhecimento, um número ideal de conformações que devem ser geradas. Além do número de confôrmeros, outro fator importante é evitar a presença daqueles com alta energia. Embora existam na literatura casos de ligantes em complexos que apresentam conformações de alta energia, estes não são a maioria sendo reportado na literatura que 70% dos ligantes formam complexos com energias torcionais inferiores a 3 kcal/mol (SCIOR et al., 2012).

Além das estratégias descritas, com o mesmo objetivo, outros autores sugerem, agrupar as moléculas segundo suas similaridades estruturais e selecionar algumas como representativas de cada grupo (BOLOGA; OPREA, 2012).

1.10. Simulações de Dinâmica Molecular aplicadas em *SBDD*

Um dos primeiros modelos propostos para entender a formação do complexo RL, foi o modelo chave-fechadura criado por Fisher em 1894. Neste modelo, o receptor é a fechadura e o ligante é a chave. As ranhuras da chave corresponderiam às interações ligante-receptor e ambos são considerados como estruturas rígidas (STANK et al., 2016). As limitações deste modelo ficaram evidentes com a disponibilidade de diferentes estruturas 3D de um mesmo alvo molecular complexas com diferentes ligantes permitindo a análise do formato destes sítios de ligação. Em 1958, Koshland propôs o modelo do ajuste induzido (“*induced fit*”) no qual a formação do complexo ligante-receptor induz mudanças conformacionais no receptor, incluindo assim a sua flexibilidade. Posteriormente, surgiu o modelo de seleção conformacional (“*conformational selection*”), no qual o receptor livre (sem o ligante) adota um conjunto de diferentes conformações e o ligante se ligaria, preferencialmente, a uma das conformações pré-existentes. Este modelo foi proposto para explicar as mudanças conformacionais observadas no sítio de ligação após um determinado ligante interagir com um sítio alostérico⁸ (STANK et al., 2016), sendo confirmado experimentalmente. Independente do modelo considerar que a interação ligação-receptor ocorra por ajuste induzido e/ou seleção conformacional, a flexibilidade do ligante (“chave”) e do receptor (“fechadura”) é evidente (COZZINI et al., 2008).

De um modo ou de outro, a flexibilidade do ligante já tem sido considerada há algum tempo na busca virtual (IRWIN; SHOICHET, 2016; KITCHEN et al., 2004; SCIOR et al., 2012). No entanto, considerar a flexibilidade do receptor continua sendo um desafio (COZZINI et al., 2008; IRWIN; SHOICHET, 2016; SOTRIFFER, 2011; WIEDER et al.,

⁸ Sítio do receptor distante do seu sítio de ligação, mas ativamente acoplado a ele (BUCKLE et al., 2014). Isso significa que a ligação de uma molécula ao sítio alostérico altera o formato do sítio de ligação, modificando a atividade deste receptor.

2017). Como já apresentado, uma estratégia bastante aceita para se incluir a flexibilidade do receptor na busca virtual é usar um conjunto de conformações deste receptor (DE VIVO et al., 2016; KLEBE, 2006). Este conjunto de conformações pode ser obtido experimentalmente ou computacionalmente usando mecânica estatística (simulações de dinâmica molecular ou Monte Carlo). Considerando que a obtenção experimental deste conjunto de conformações nem sempre é possível, as simulações usando mecânica quântica ganharam uma grande importância, em especial as simulações de dinâmica molecular (DM) (DE VIVO et al., 2016).

Simulações de DM clássica é um método computacional usado para estudar o movimento dos átomos e das moléculas aplicando a equação de movimento de *Newton*, apresentada a seguir.

$$f = ma = \frac{\partial V(x)}{\partial x}$$

onde: f é a força de um átomo i num determinado instante, m é a massa de i , a é a aceleração de i neste instante, V é o potencial e x é a posição deste i no espaço Cartesiano e V é a função de energia potencial empírica.

Um determinado campo de força é usado para estimar a força da interação entre dois átomos e calcular a energia total do sistema (Figura 1.12). Assim, a integração da equação de movimento de *Newton* durante a simulação gera sucessivas configurações do sistema, resultando numa trajetória de coordenadas em função do tempo (DE VIVO et al., 2016). Uma limitação nestes campos de força é não considerar a polarizabilidade dos átomos, pois as cargas parciais destes não podem se alterar durante a simulação. Campos de força polarizáveis foram criados para contornar esta limitação, mas são ainda muito caros computacionalmente. Outra limitação desses campos de força é não permitir que ocorra a quebra, ou formação, de novas ligações, que pode ser contornada usando simulações híbridas com mecânica quântica e clássica (QM/MM). Ainda assim, várias propriedades podem ser calculadas a partir destas trajetórias de DM clássica, entre elas o ΔG , k_{on} , τ e k_{off} (MERGET; SOTRIFFER, 2015; STEINBRECHER et al., 2015).

$$\begin{aligned}
 V = & \sum_i \frac{k_{l,i}}{2} (l_i - l_{0,i})^2 \quad \text{Ligações} \quad \text{Distância de ligação} \quad l_0 \\
 & + \sum_i \frac{k_{\alpha,i}}{2} (\alpha_i - \alpha_{0,i})^2 \quad \text{Ângulo} \quad \text{Ângulo de ligação} \quad \alpha_0 \\
 & + \sum_i \left\{ \sum_k^M \frac{V_{ik}}{2} [1 + \cos(n_{ik}\theta_{ik} - \theta_{0,ik})] \right\} \quad \text{Diedros} \quad \text{Ângulo diedro} \quad \theta_0 \\
 & + \sum_{i,j} \varepsilon_{ij} \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - 2 \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad \text{Pares} \quad \text{Interação de van der Waals} \\
 & + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}} \quad \text{Pares} \quad \text{Interação eletrostática}
 \end{aligned}$$

“Termos para pares de átomos ligados”

“Termos para pares de átomos não ligados”

Figura 1.12. Representação da equação de um campo de força (V) com termos representando interações, par a par, de átomos ligados (ligação, ângulo e diedro) e entre pares de átomos não ligados (interação de *van der Waals* e eletrostática) (DE VIVO et al., 2016). Nos três termos referentes aos átomos ligados, temos: i é um par de átomos; l e α representam distância e ângulo de ligação, respectivamente, sendo ambos descritos por potenciais harmônicos com valores de referência l_0 e α_0 e constantes de força k_l e k_α , respectivamente; θ é o ângulo diedro, descrito por uma série de M cossenos, na qual n_{ik} descreve a multiplicidade do termo k da série, $\theta_{0,ik}$ é a fase do ângulo e V_{ik} é a energia de barreira. Nos dois termos referentes aos pares de átomos não ligados, temos: i e j são átomos; r_{ij} é a distância entre estes átomos; a interação de *van der Waals* é descrita por um potencial 12-6 de *Lennard-Jones*, no qual ε_{ij} é o valor mínimo do vale de energia e σ é a mínima distância de energia (soma dos raios de *van der Waals* dos átomos i e j) e para interação eletrostática se usa um potencial de Coulomb, no qual q_i e q_j são as cargas parciais dos átomos, ϵ_0 é a permissividade do espaço livre e ϵ_r é a permissividade relativa (ou constante dielétrica) (DE VIVO et al., 2016). As figuras foram retiradas do *Scientific American Blog* (“Scientific American Blog”, [s.d.]).

O cálculo da energia de *Gibbs* para descrever a formação do complexo ligante-receptor não é simples, sendo possível usando mecânica estatística (amostragem de *Boltzmann*). Assim, a energia de *Gibbs* pode ser expressa pela equação a seguir:

$$\Delta G^\circ = -kT \ln \left(\frac{\rho_{\text{livre}}(x)}{\rho_{\text{ligado}}(x)} \right) + kT \ln \left(\frac{C_{\text{caixa}}}{C^\circ} \right)$$

onde: ΔG° é a energia de *Gibbs* para formação do complexo LR, nas concentrações padrões (1 M, correspondendo a 1 molécula num volume de 1663 \AA^3); k é a constante de *Boltzmann*; T é a temperatura em Kelvin; $\rho_{\text{livre}}(x)$ $\rho_{\text{ligado}}(x)$ é a probabilidade de se encontrar a proteína nos estados livre e ligado, respectivamente; C_{caixa} é a concentração do par ligante - proteína na caixa e C° é a concentração padrão (~ 1 molécula num volume de 1663 \AA^3).

Para se ter precisão nos valores medidos é necessário uma grande amostragem do sistema. No entanto, a transição entre diferentes estados está associada às barreiras energéticas significativas, podendo ser, então, considerados eventos “raros” com base nas

escalas de tempo usadas nas simulações. Assim, para se calcular a energia livre usando a equação descrita são necessárias simulações longas (feitas repetidas vezes) para que muitos eventos de ligação e dissociação possam ser observados, tornando o resultado estatisticamente significativo. Como alternativa surgiram métodos denominados de *enhanced sampling*, nos quais se explora o fato de que a energia livre é uma função de estado. Isso significa que a diferença de energias do estado A para o estado B, não depende do caminho percorrido para chegar de A para B. De modo geral, isso é feito introduzindo-se um *Hamiltoniano* misto. O cálculo das diferenças de energias livres relativas ($\Delta\Delta G^\circ$), pode ser feito transformando o ligante A em B nos dois estados (ligado e livre). Esta transformação, que não tem sentido físico, não afeta o cálculo do $\Delta\Delta G^\circ$. Depois, a energia livre das duas transformações é cuidadosamente calculada e as contribuições relativas são calculadas, sendo possível determinar os valores de $\Delta\Delta G^\circ$ (DE VIVO et al., 2016). Este método é comumente denominado “*free-energy perturbation*” (FEP) e está representado na Figura 1.13 (A).

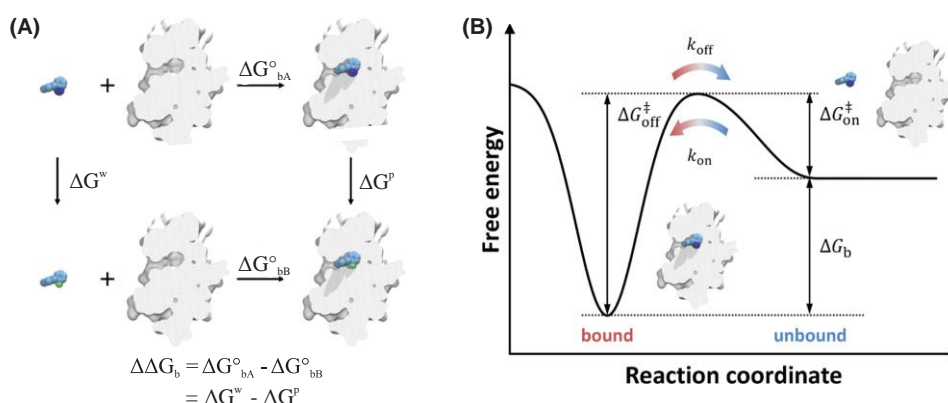


Figura 1.13. Representação do método da FEP usado para calcular os valores da diferença de energia livre relativa ($\Delta\Delta G_b$) entre os ligantes A e B e do método do PMF usado para obter informações termodinâmicas (ΔG_b) e cinéticas (k_{on} e k_{off}). (A) Cálculo do valor da diferença de energia livre relativa ($\Delta\Delta G_b$) entre os ligantes A e B usando os métodos do FEP, onde ΔG_{bA}° e ΔG_{bB}° são as energias livres de ligação de A e B, respectivamente, e ΔG^w e ΔG^p são a energia livre da transformação de A em B, no solvente e, no sítio da proteína. (B) Perfil de energia para um processo simples de formação e dissociação do complexo RL. A energia do estado ligado é menor do que a energia do estado livre. O valor de ΔG_b é obtido a partir da diferença entre os mínimos e os valores das constantes de velocidade são obtidas (k_{on} e k_{off}) a partir das diferenças entre os estados de mínimo e de transição (ΔG_{on}^\ddagger e ΔG_{off}^\ddagger) (DE VIVO et al., 2016).

Na literatura estão descritos, vários exemplos de aplicação deste método tanto para comparar as $\Delta\Delta G$ para formação do complexo RL de séries de análogos bem como para analisar o efeito de mutações dentro do sítio de ligação (ABEL et al., 2017; CIORDIA et al., 2016; STEINBRECHER et al., 2017, 2015). O valor de ΔG° também pode ser obtido, no entanto, são necessários ciclos termodinâmicos mais complexos e mais desafiadores.

Outros métodos de *enhanced sampling*, baseados em caminhos com sentido físico, conseguem calcular a energia livre como uma função de algumas coordenadas de reação, que sejam capazes de considerar graus de liberdade relevantes para a formação e dissociação do complexo RL. Este método pode ser usado para se obter informações termodinâmicas e cinéticas sobre o sistema (Figura 1.13 (B)). Nestes casos, uma intervenção é incluída na simulação de DM de modo que o sistema siga uma coordenada de reação pré-definida. Um dos métodos mais populares de *enhanced sampling* baseado num conjunto de coordenadas é o *umbrella sampling* (DE VIVO et al., 2016).

Além do cálculo do ΔG e das constantes de velocidade, as trajetórias obtidas nas simulações de DM, sem qualquer intervenção, são um conjunto *snapshots* (“fotografias”). Os *snapshots* representam diferentes conformações do receptor ao longo do tempo. Estas conformações podem ser agrupadas selecionando-se algumas estruturas representativas. Estas estruturas podem, então, ser usadas na busca virtual aplicando-se, por exemplo, filtro de ancoramento. Os resultados obtidos em cada ancoragem são agrupados e os melhores *hits* selecionados. A Figura 1.14 mostra uma representação das etapas envolvida neste processo.

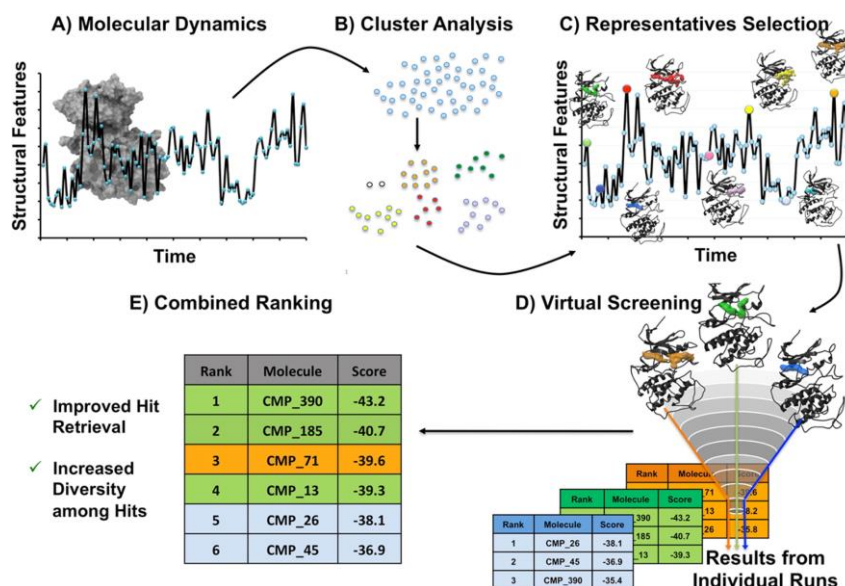


Figura 1.14. Principais etapas da busca virtual aplicando-se o filtro de ancoramento e usando diferentes conformações obtidas por simulações de DM. (A) O espaço conformacional do receptor é explorado usando simulações de DM. (B) Vários *snapshots* são obtidos da trajetória de DM e, subsequentemente, agrupados (ex. usando valores de 2D-RMSD, todos contra todos). (C) Uma estrutura representativa de cada grupo é escolhida (ex. estrutura central). (D) A busca virtual é feita usando, separadamente, a estrutura representativa de cada grupo. (E) Os resultados são combinados e os melhores compostos são selecionados (DE VIVO et al., 2016).

A etapa fundamental neste processo é o agrupamento de todos os *snapshots* e a seleção das estruturas representativas, pois elimina a redundância (ou seja, diferentes estruturas que representam o mesmo estado conformacional). Exemplos de aplicação de simulações de DM para selecionar diferentes conformações relevantes na busca virtual estão descritos na literatura (MISRA et al., 2017; PICCIRILLO et al., 2016). Adicionalmente, como descrito no item 1.7.1, as conformações identificadas nas simulações de DM podem ser usadas para gerar modelos farmacofóricos. Vários exemplos de sucesso do uso de simulações de DM em várias etapas da busca virtual estão descritos na literatura.

1.11. Dengue e Febre Aftosa

A Dengue e a Febre Aftosa são infecções virais que ocorrem tanto no Brasil como no mundo, apresentando significativo impacto socioeconômico, sendo que a busca de medicamentos antivirais se faz necessária quando a vacina ainda não está disponível e/ou confere proteção limitada e principalmente para o controle de surtos.

1.11.1. Dengue

1.11.1.1. Doença e Epidemiologia

A dengue é uma doença viral transmitida pela picada dos mosquitos *Aedes aegypti* e/ou *Aedes albopictus*, sendo o primeiro o principal vetor de transmissão. O vírus da Dengue (DENV, do inglês *Dengue virus*), pertencente à família *Flaviviridae* - gênero *Flavivirus*, possui quatro sorotipos bem caracterizados (DENV-1, DENV-2, DENV-3 e DENV-4). Estes quatro sorotipos estão distribuídos no mundo todo. Mais recentemente, um quinto sorotipo foi descrito (BEHNAM et al., 2016; NITSCHKE et al., 2014; NORMILE, 2013).

A gravidade das infecções causadas pelo DENV varia de forma surpreendente e apenas um quarto das pessoas infectadas apresentam algum sintoma da dengue. Na maioria dos casos nos quais a doença é sintomática, os sintomas são de uma gripe bem forte com febre alta e dores no corpo. A recuperação destes pacientes ocorre depois de alguns dias sem que sequelas graves sejam observadas. Uma pequena parcela de pacientes desenvolve as formas mais graves da doença (a febre hemorrágica da dengue e a síndrome de choque da dengue). Estas formas mais graves de doença podem, no entanto, levar o paciente a óbito (BEHNAM et al., 2016; GUZMAN; HARRIS, 2015; NITSCHKE et al., 2014).

A dengue é causada por qualquer um dos quatro diferentes sorotipos do vírus enquanto as formas mais graves da doença são, geralmente, causadas pela infecção, subsequente, por outro sorotipo diferente (BEHNAM et al., 2016; WHITEHORN; SIMMONS, 2011). Este comportamento é o principal obstáculo para o desenvolvimento de vacinas. Ainda, as formas mais graves da doença estão mais associadas às infecções pelo DENV2, seguida pelos DENV1 e DENV3 e raramente pelo DENV4 (NITSCHKE et al., 2014).

A Organização Mundial da Saúde (OMS) estima que o DENV cause, anualmente, 50 a 100 milhões de infecções no mundo todo das quais ~ 500 mil progridem para formas mais graves da doença, resultando em ~ 25 mil mortes (BHATT et al., 2013; WHITEHORN;

SIMMONS, 2011). Um estudo mais recente estima que as infecções anuais pelo vírus cheguem a 390 milhões. As regiões mais afetadas estão localizadas em áreas tropicais e subtropicais, incluindo o Brasil (Figura 1.15). Os surtos mais recentes desta doença afetaram milhões de pessoas, sendo estes causados pelos quatro sorotipos (BEHNAM et al., 2016).

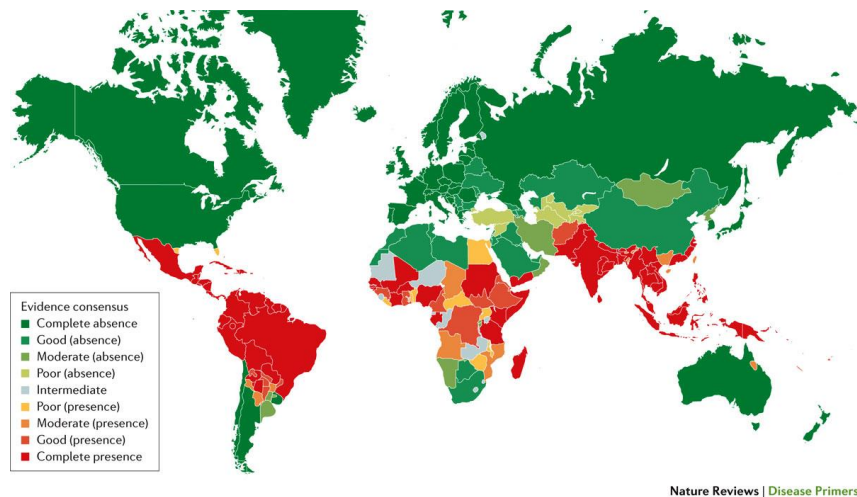


Figura 1.15. Distribuição global da Dengue de acordo com os números de casos (BHATT et al., 2013), variando de totalmente ausente (verde) até totalmente presente (vermelho).

1.11.1.2. Vacina e Tratamento

Uma das dificuldades de se encontrar uma vacina contra a infecção pelo DENV é a necessidade de que esta seja efetiva contra todos os sorotipos do vírus (BEHNAM et al., 2016; GUZMAN; HARRIS, 2015). Assim, embora a busca por uma vacina para prevenir a infecção pelo DENV tenha se iniciado nos anos 1940, a primeira vacina (CYD-TDV, químera tetravalente dos vírus da Febre Amarela e da Dengue, desenvolvida pela indústria farmacêutica *Sanofi Pasteur*) foi licenciada em dezembro de 2015, no México. Estudos clínicos fase III mostram que a vacina CYD-TDV reduz o número de hospitalizações e o aparecimento das formas mais graves da doença (FLIPSE; SMIT, 2015). No entanto, estes estudos ainda mostraram que sua eficácia varia, de acordo com o país e com a região estudada, sendo constantemente relatada uma limitada proteção contra o DENV2 (FLIPSE; SMIT, 2015; NORMILE, 2014). Esta proteção limitada contra a infecção pelo DENV2 pode ser um obstáculo às campanhas de vacinação, principalmente, por este sorotipo estar mais

associado ao desenvolvimento das formas graves da doença. Outras vacinas estão sendo desenvolvidas, mas seus estudos clínicos ainda não foram totalmente concluídos (BEHNAM et al., 2016; FLIPSE; SMIT, 2015; GUZMAN; HARRIS, 2015). Além disso, o aparecimento de um novo sorotipo pode colocar em risco os avanços feitos no sentido de se desenvolver uma vacina tetravalente (NITSCHKE et al., 2014). Considerando todas as dificuldades apontadas, uma estratégia para tratar a Dengue seria o desenvolvimento de uma terapia antiviral.

Até o presente momento, dentro do que nos é conhecido, não existem terapias antivirais aprovadas para tratar e/ou combater a infecção causada pelo DENV. O tratamento usado é apenas para combater as manifestações clínicas da doença. Ainda, vários medicamentos são contraindicados no caso de suspeita de infecção pelo DENV (BEHNAM et al., 2016; BHATT et al., 2013; BOLDESCU et al., 2017; GUZMAN; HARRIS, 2015; NITSCHKE et al., 2014; TIMIRI; SINHA; JAYAPRAKASH, 2016).

1.11.1.3. Busca racional de antivirais contra o DENV – a escolha do alvo

A escolha e identificação de um alvo válido é uma das etapas fundamentais no planejamento de novos fármacos. O DENV é um vírus de RNA fita simples positivo que é traduzido em uma única poliproteína viral, usando o mecanismo de tradução do hospedeiro em conjunto com algumas proteínas virais. Esta poliproteína fica inserida na membrana do retículo endoplasmático e precisa ser clivada em porções menores, maduras, que são denominadas tanto como funcionais/não-estruturais (quando apresentam atividade enzimática) ou como estruturais (quando fazem parte da estrutura do vírus) (CLERCQ, 2004). A clivagem desta poliproteína é feita por proteases do hospedeiro (furina e signalase) e por uma protease viral, dando origem a três proteínas estruturais (C, PrM e E) e 7 funcionais/não-estruturais (NS, do inglês *non structural*). A Figura 1.16 mostra a poliproteína viral e as regiões de clivagem pelas diferentes proteases.

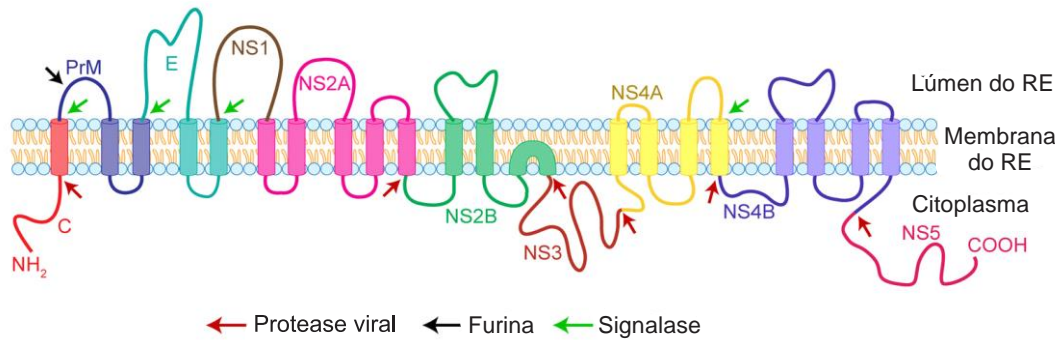


Figura 1.16. Representação da topologia da poliproteína viral na membrana do retículo endoplasmático (RE) bem como dos seus sítios de clivagem. A clivagem desta poliproteína dá origem a três proteínas estruturais – capsídeo (C), proteína precursora de membrana (PrM) e envelope (E) – e sete proteínas não estruturais – NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5. As setas em vermelho, preto e verde indicam as regiões clivadas pela protease viral e pelas proteases do hospedeiro (furina e signalase), respectivamente (DE ALMEIDA et al., 2013).

Dentre os alvos sugeridos para o planejamento de antivirais contra o DENV, os mais explorados são: a glicoproteína do envelope do DENV (atuando no processo de absorção); a polimerase NS5 do DENV, metiltransferase do vírus, a helicase NS3 e a NS4B (todas estas envolvidas no processo de transcrição do RNA); protease viral (responsável pela clivagem da poliproteína viral) e a proteína C do capsídeo viral (responsável pela montagem do vírus) (BEHNAM et al., 2016; NITSCHKE et al., 2014).

As proteases virais são alvos já bastante reconhecidos e de grande interesse, devido ao seu papel fundamental no ciclo de vida de muitos vírus, incluindo-se os *flavivirus* (ex. o DENV e o vírus da Hepatite C), o herpes vírus e o retrovírus (ex. HIV) (CLERCQ, 2004; NOBLE et al., 2010; NOBLE; SHI, 2012) e os *picornavirus* (ex. vírus da Febre Aftosa) (GRUBMAN; BAXT, 2004; TIEW et al., 2012). Um exemplo da importância destas proteases como alvo para o planejamento de fármacos contra *flavivirus* é a protease NS4B/NS5 do vírus da Hepatite C (HCV), para a qual já existem alguns fármacos desenvolvidos e aprovados como tratamento antiviral (DABBOUSEH; JENSEN, 2013; PARKINSON; PRYDE, 2010; TSE, 2013; WYLES, 2013). Pelo exposto, a protease do DENV é reconhecida como um alvo promissor para o desenvolvimento de antivirais (LAI et al., 2013; NITSCHKE et al., 2014; NOBLE et al., 2010, 2012).

1.11.1.4. Protease NS2B/NS3 do DENV

A proteína não-estrutural 3 (NS3) é uma proteína multifuncional com um domínio serino protease (180 aminoácidos na região N-terminal) e um domínio helicase dirigido por ATP e por RNA trifosfatase (180-618 aminoácidos na região C-terminal) (ARAVAPALLI et al., 2012; LESCAR et al., 2008; LIM et al., 2013). O domínio serino protease (NS3pro) contém a tríade catalítica (*His51*, *Asp75* e *Ser135*) e o oxânion (*Gly133*, *Thr134* e *Ser135*). Embora a tríade catalítica se encontre no domínio NS3pro, a ligação, não covalente, com a proteína não-estrutural 2B (NS2B) se faz necessária para que a enzima apresente atividade catalítica. Assim, a protease do DENV em sua forma ativa é denominada de NS2B/NS3pro. NS2B é uma proteína hidrofóbica com 130 aminoácidos, que se insere na membrana do retículo endoplasmático. No entanto, a interação de apenas um segmento conservado de ~40 aminoácidos, com caráter mais hidrofílico, é suficiente para a ativação desta protease (NOBLE et al., 2012; NOBLE; SHI, 2012). Os resíduos da região NS2B participam da formação das cavidades S_2 e S_3 (LESCAR et al., 2008; NOBLE et al., 2012). Outro fator bastante relevante para escolha desta protease como alvo é a alta similaridade da sequência primária das proteases dos diferentes sorotipos do DENV (identidade >60%). O alinhamento da sequência primária das proteases dos quatro sorotipos está apresentado na Figura 1.17.

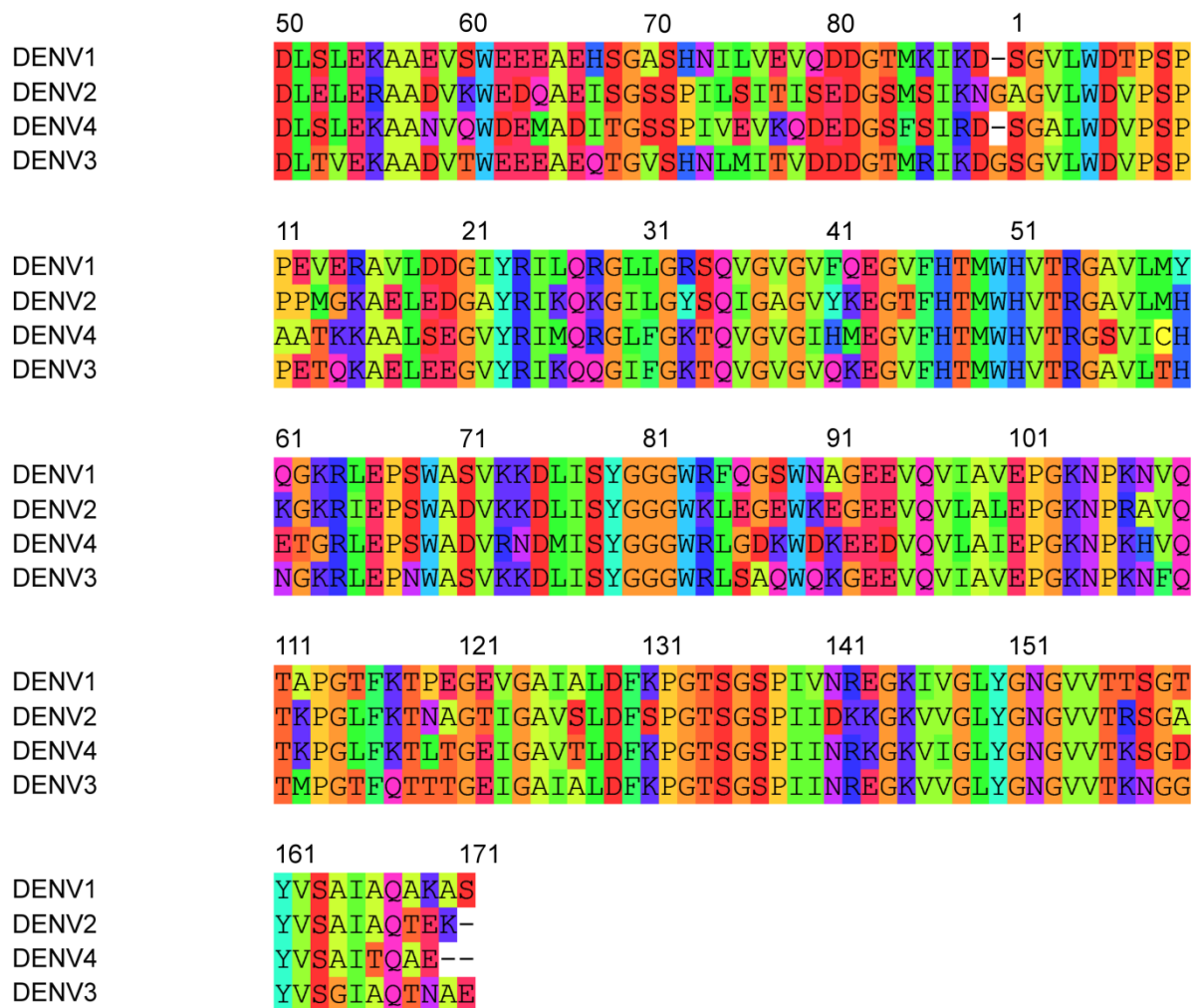


Figura 1.17. Alinhamento da sequência primária da NS2B/NS3pro dos sorotipos DENV1 a DENV4. Os primeiros 37 resíduos correspondem à região NS2B. Alinhamento feito usando o programa Sybyl-X (SYBYL, 2013) e aplicando o esquema de cores de Taylor (TAYLOR, 1997).

O sítio de ligação desta protease é formado por uma sequência de cavidades alinhadas em sentidos opostos, de modo a acomodar os aminoácidos do substrato. Segundo a terminologia usual, as cavidades direcionadas no sentido *N*-terminal do substrato são denominadas $S_1, S_2, S_3, \dots, S_n$, enquanto que as cavidades direcionadas no sentido *C*-terminal do substrato são denominadas $S_1', S_2', S_3', \dots, S_n'$. Analogamente, os resíduos do substrato que ocupam as cavidades correspondentes são denominados $P_n, P_2, P_1, P_1', P_2', P_n'$, respectivamente, conforme apresentado na Figura 1.18.

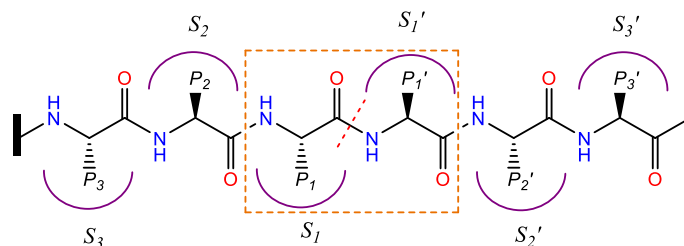


Figura 1.18. Representação esquemática da ligação entre um peptídeo e uma protease, indicando a terminologia usual das cavidades da enzima e dos correspondentes resíduos do substrato. A linha tracejada indica o ponto em que ocorre a hidrólise do peptídeo (TURK, 2006).

No mecanismo de hidrólise do grupo amida proposto para esta classe de enzimas, o oxigênio da cadeia lateral da *Ser135* e o anel imidazólico da *His51* formam um par iônico, aumentando o caráter nucleofílico deste oxigênio (Figura 1.19 (A)). Assim, durante a hidrólise, este nucleófilo (oxigênio da *Ser135*) ataca o carbono carbonílico da ligação peptídica do substrato (Figura 1.19 (B)), gerando um intermediário tetraédrico (Figura 1.19 (C)), estabilizado por ligações de hidrogênio na cavidade oxiânion (*Ser135*, *Gly133* e *Thr134*). Este intermediário tetraédrico perde a porção amina, deixando a enzima acilada (Figura 1.19 (D)). A enzima acilada reage com uma molécula de água (Figura 1.19 (E)), formando ácido carboxílico e restaurando a enzima livre (Figura 1.19A) (BABINE; BENDER, 1997; TURK, 2006).

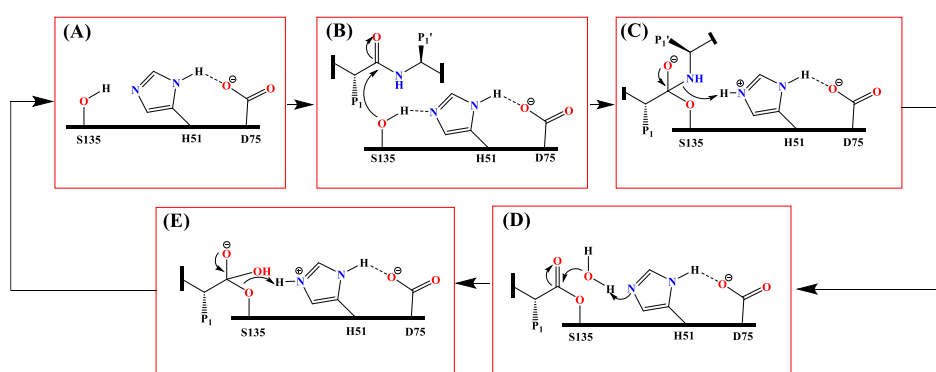


Figura 1.19. Representação esquemática das etapas envolvidas no mecanismo proposto para hidrólise de amidas catalisada por serino proteases: (A) protease livre; (B) ataque do oxigênio da *Ser135* (nucleófilo) ao carbono carbonílico da ligação peptídica do substrato; (C) intermediário tetraédrico formado que é estabilizado por ligações de hidrogênio com resíduos da cavidade oxiânion (*Ser135*, *Gly133*, *Thr134*); (D) perda da porção amina do intermediário tetraédrico, deixando a protease acilada, (E) reação da protease acilada com uma molécula de água formando ácido carboxílico e restaurando a protease livre (A) (BABINE; BENDER, 1997; TURK, 2006).

A estrutura 3D da NS2B/NS3pro do DENV já foi elucidada a partir de dados cristalográficos e de RMN. Estas estruturas estão disponíveis no *PDB*, sendo a maioria obtida na ausência de ligantes e algumas delas complexadas com inibidores peptidomiméticos (GIBBS; STEELE; TOUNGE, 2014, 2014; NOBLE et al., 2012). Ainda, na literatura são descritos outros exemplos de inibidores da NS2B/NS3pro do DENV, dentre os quais se citam: derivados peptídicos (BEHNAM et al., 2015; FRECER; MIERTUS, 2010; NITSCHKE et al., 2012; SCHÜLLER et al., 2011; YANG et al., 2011; YIN et al., 2006), panduratina e 4-hidroxipanduratina (KIAT et al., 2006), derivados da ftalazina (BODENREIDER et al., 2009), α -cetoamidas (STEUER et al., 2011), derivados do antraceno (TOMLINSON et al., 2009; TOMLINSON; WATOWICH, 2011), aminobenzamidas (ARAVAPALLI et al., 2012), aril-cianoacrilamidas (NITSCHKE; STEUER; KLEIN, 2011), derivados quinolínicos (DENG et al., 2012); macrolídeos (TOMLINSON; WATOWICH, 2012); benzotiazóis (LAI et al., 2013) e flavonóides (DE SOUSA et al., 2015).

Os inibidores mais potentes frente à NS2B/NS3pro do DENV descritos na literatura são peptidomiméticos, sendo que o mais ativo deles apresenta um $K_i = 12$ nM (tripeptídeo híbrido (BEHNAM et al., 2015)). No entanto, este inibidor é menos efetivo em inibir, *in vitro*, a replicação do DENV ($EC_{50} = 20,35$ μ M), sendo esta menor atividade atribuída à falta de estabilidade deste composto. Além disso, este inibidor não se mostrou permeável por difusão passiva, o que poderia ser esperado considerando seu caráter polar e a presença de grupos carregados em pH fisiológico (BEHNAM et al., 2015). Assim, embora este inibidor se mostra bastante ativo frente à NS2B/NS3pro do DENV, sua menor potência frente ao vírus, sua baixa estabilidade e permeabilidade não fazem deste composto um promissor e bom candidato a fármaco. Em relação aos inibidores não peptídicos, os compostos mais ativos têm valores de atividade entre 10^{-7} e 10^{-6} M. Muitos destes compostos apresentam um alto grau de flexibilidade e são carregados em pH fisiológico. Ainda mais, alguns foram caracterizados

como inibidores não competitivos (DE SOUSA et al., 2015; WU et al., 2015). Para estes últimos, dois possíveis sítios de inibição alostérica foram sugeridos. Um deles usando ferramentas computacionais, sendo localizado abaixo do sítio de ligação do substrato (WU et al., 2015). O segundo foi sugerido usando cinco sondas de cisteína (YILDIZ et al., 2013). No entanto, ainda não foram obtidas estruturas 3D da NS2B/NS3pro em complexos com qualquer inibidor alostérico, não sendo possível comprovar a existência destes sítios.

Adicionalmente, até a elaboração desta tese, dentro do que nos é conhecido, não existem relatos de compostos candidatos que estejam em etapas clínicas de desenvolvimento (BHATT et al., 2013; BOLDESCU et al., 2017; LAI et al., 2013; NITSCHKE et al., 2014; NOBLE et al., 2010). Isto significa que o desenvolvimento pré-clínico de compostos ativos ainda não encontrou o candidato ideal, que reúna todas as características de potência, de toxicidade e de biodisponibilidade necessárias para o início dos testes em seres humanos. Além disso, o interesse na descoberta de inibidores da NS2B/NS3pro não se limita ao tratamento da infecção pelo DENV, uma vez que esta protease também é encontrada em outros vírus que causam doenças em humanos como o vírus da Febre do Nilo Ocidental e do Zika (BOLDESCU et al., 2017). A disponibilidade de estruturas 3D da NS2B/NS3pro do DENV bem como informações sobre alguns inibidores fornece subsídios para a aplicação de estratégias modernas (*SBDD* e *LBDD*) para descoberta de novos inibidores desta protease, em especial a busca virtual.

1.11.2. Febre Aftosa

1.11.2.1. Doença e Epidemiologia

A Febre Aftosa (FA), também conhecida como “*foot-and-mouth disease*” (*FMD*), é altamente contagiosa e acomete principalmente ungulados (animais de casco fendido) domésticos e silvestres como bovinos, suínos, ovinos, caprinos e bubalinos sendo caracterizada por causar febre e vesículas na região das patas e da boca (GRUBMAN; BAXT,

2004). Embora a FA apresente uma baixa mortalidade (~1% a 40 % em animais adultos e jovens, respectivamente) sua morbidade é alta (chegando a 100%), levando à perda de produtividade e a prejuízos econômicos para o setor agropecuário (BRITO et al., 2017; GRUBMAN; BAXT, 2004). Do ponto de vista econômico, o Brasil possui ~19% do comércio internacional de carne bovina, exportando ~1,3 milhões de toneladas (~ 5,3 milhões de dólares), em 2016, principalmente para Hong Kong e China (ABIEC, 2016). Ainda, a Organização Mundial da Saúde Animal (OIE), classifica os países como “zonas livre de FA” onde se pratica, ou não, a vacinação (Figura 1.20). Uma grande parte do território brasileiro é classificado “zonas livre de FA com vacinação” (OIE, 2017).

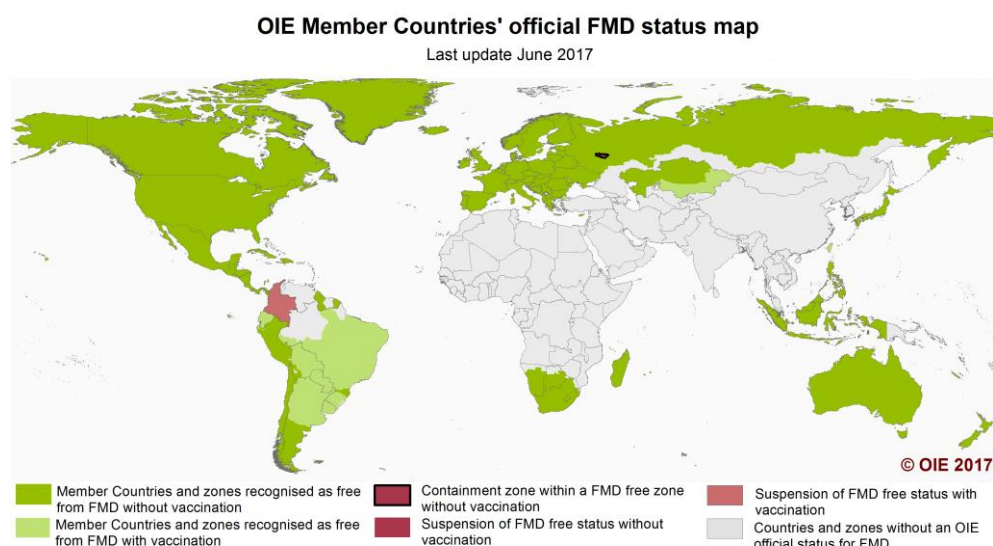


Figura 1.20. Distribuição global da FA de acordo com a classificação feita pela organização mundial da saúde animal (OIE). Países classificados como: zonas livre de FA com (verde claro), ou sem (verde escuro), vacinação; zona contaminada numa área livre sem vacinação (vermelho com borda preta); suspensão do status de zonas livre de FA com (vermelho claro), ou sem (vermelho escuro), vacinação e zonas sem status (cinza) segundo a OIE (OIE, 2017).

O vírus da FA, pertencente à família *Picornaviridae* - gênero *Aphtovirus*, é transmitido principalmente por aerossóis respiratórios e contato direto, ou indireto, com animais infectados e possui sete sorotipos (A, O, C, Ásia 1, SAT1, SAT2 e SAT3, sendo os três primeiros encontrados no Brasil).

1.11.2.2. Vacinas e Tratamentos

O principal tratamento pra FA é a vacinação e o sacrifício dos rebanhos infectados. As vacinas contra o vírus da FA conferem uma proteção sorotipo específica, ou seja, a vacinação contra um sorotipo não confere proteção contra outro. As políticas de vacinação iniciadas no final da década de 1950 conseguiriam erradicar esta doença em algumas regiões do mundo (ex. América do Norte, Europa Ocidental). Entretanto, a ocorrência frequente de surtos tem causado uma grande preocupação quanto aos prejuízos socioeconômicos associados a esta doença. Neste contexto, vários grupos vêm tentando desenvolver novas vacinas (DIAZ-SAN SEGUNDO et al., 2017). Uma das dificuldades é a grande variabilidade genética entre os sorotipos deste vírus. No Brasil, a vacinação de animais se iniciou em 1941, no entanto, ainda existem regiões com foco da doença, principalmente no norte do País (MAYEN, 2003; MINISTÉRIO DA AGRICULTURA, 2014). Até o presente momento, dentro do que nos é conhecido, não existem terapias antivirais aprovadas para tratar e/ou combater a infecção causada pelo vírus da FA. No caso de surto, os animais infectados bem como os próximos a estes são sacrificados.

1.11.2.3. Busca racional de antivirais contra o vírus FA

Como já discutido nesta tese, a escolha e a identificação de um alvo válido é uma das etapas fundamentais no planejamento racional de compostos bioativos. De modo análogo ao DENV, o vírus da FA possui um genoma de RNA fita simples positivo (Figura 1.21) que é traduzido em uma única poliproteína viral. Esta poliproteína é clivada pela ação de proteases virais, gerando quatro proteínas estruturais (VP1, VP2, VP3 e VP4) e sete proteínas não estruturais (2A, 2B, 2C, 3A, 3B, 3C e 3D).

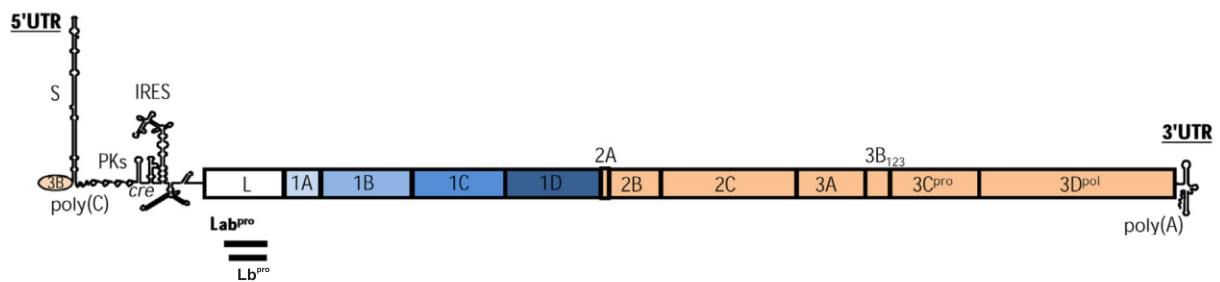


Figura 1.21. Representação do genoma de RNA fita simples do vírus da FA. As caixas representam as regiões codificadoras do genoma de acordo com o nome das proteínas virais que elas codificam. As caixas azuis e laranjas representam as regiões das proteínas estruturais e não estruturais, respectivamente. A caixa branca representa a região que codifica a proteína líder (L^{pro}) do vírus da FA, reconhecida como um fator de virulência. S é um pequeno fragmento do genoma, poly(C) é a cauda de poli citosina, PK é o elemento de replicação, o IRES é o segmento de entrada interno do ribossomo, que permite que o genoma se ligue ao ribossomo sem depender dos fatores de iniciação eucarióticos (DIAZ-SAN SEGUNDO et al., 2017).

Dentre os alvos sugeridos para o planejamento de antivirais contra o vírus da FA, se destacam as proteases líder (L^{pro}) e a $3C^{pro}$ (NORDER et al., 2011; STEINBERGER et al., 2014). Em especial, a protease L^{pro} , que é reconhecida como um alvo bastante promissor por inibir a síntese de proteínas do hospedeiro (DIAZ-SAN SEGUNDO et al., 2017; NOGUEIRA SANTOS et al., 2012; STEINBERGER et al., 2014).

1.11.2.4. L^{pro} do vírus da FA

A L^{pro} corresponde à porção *N*-terminal da poliproteína, sofrendo um processo de auto clivagem no qual sua porção *C*-terminal é separada da proteína VP4 (NOGUEIRA SANTOS et al., 2012). No genoma viral (Figura 1.21), a região que codifica a proteína L^{pro} é precedida por dois códons AUG, sendo que a síntese da proteína se inicia em qualquer um dos dois, levando a duas isoformas da L^{pro} , a Lab e a Lb. Embora estas isoformas apresentem uma diferença de 28 resíduos de aminoácidos, ambas apresentam a mesma atividade enzimática (GUARNE, 1998). Ainda, estudos da literatura de cinética enzimática e de mutagênese sugerem a Lb^{pro} como a isoforma biologicamente ativa da proteína L^{pro} (NOGUEIRA SANTOS et al., 2012; SANTOS et al., 2009). Por estas razões, a isoforma Lb^{pro} é a mais comumente usada e estudada.

A Lb^{pro} pertence à família das papaínas, sendo considerada uma das menores proteínas desta família, com 173 aminoácidos (CENCIC et al., 2007). Contudo, a Lb^{pro} apresenta uma das maiores especificidades desta classe, tendo se identificado três substratos fisiológicos (a própria poliproteína viral e dois fatores de iniciação eucarióticos homólogos (eIF) 4G, eIF4GI e eIF4GII). A clivagem do eIF4GI e do eIF4GII pela Lb^{pro} inibe a síntese de proteínas do hospedeiro sem afetar a síntese de proteínas virais. Ainda mais, Lb^{pro} bloqueia a resposta imune inata através da inibição direta e/ou indireta da indução de interferon beta (IFN β) (GRUBMAN et al., 2008; GUAN; BELSHAM, 2017; STEINBERGER et al., 2014). Estas funções fazem com que a Lb^{pro} seja considerada um fator de virulência⁹ (DIAZ-SAN SEGUNDO et al., 2017; GRUBMAN; BAXT, 2004; GUARNE, 1998; SANTOS et al., 2009; STEINBERGER et al., 2014). Considerando todas estas funções, a literatura mostra que a inibição da Lb^{pro} pode levar à diminuição da taxa de replicação viral, em até 1000 vezes, indicando que esta proteína pode ser considerada um alvo válido para o desenvolvimento da terapia antiviral (GUARNE, 1998; SANTOS et al., 2009; STEINBERGER et al., 2014).

As papaínas são cisteíno proteases e seus sítios de ligação também são formados por uma sequência de cavidades alinhadas em sentidos opostos, que seguem a mesma terminologia apresentada para a NS2B/NS3pro do DENV, vide Figura 1.18. A Lb^{pro} do vírus da FA contém uma tríade catalítica (*Cys51*, *His148* e *Asp163*) e uma cavidade para acomodar um oxianion (*Cys51* e *Asn46*) (GUARNE, 1998; GUARNÉ et al., 2000; OTTO; SCHIRMEISTER, 1997).

O mecanismo de proteólise proposto para as cisteíno proteases é análogo ao já apresentado para as serino proteases com exceção do nucleófilo, que é, neste caso, o íon tiolato (Figura 1.22 (A)). Assim, este nucleófilo ataca o carbono carbonílico da ligação

⁹ Fator que, se ausente, ou comprometido, pode levar à inibição da replicação viral e impossibilitar a propagação da doença.

peptídica, gerando um intermediário tetraédrico (Figura 1.22(B)). Este intermediário tetraédrico perde a porção amina, deixando a enzima acilada (Figura 1.22(C)). A enzima acilada reage com uma molécula de água (Figura 1.22(C)), formando ácido carboxílico e restaurando a enzima livre (Figura 1.22(D)) (BABINE; BENDER, 1997; OTTO; SCHIRMEISTER, 1997; TURK, 2006).

Além da grande especificidade, outro fator que diferencia a Lb^{pro} das proteínas da família das papaínas é a presença de uma extensão C-terminal (CTE). Esta extensão tem cerca de 20 resíduos de aminoácidos e está envolvida no processo de auto clivagem (CENCIC et al., 2007; GUARNE, 1998). Estas diferenças tornam a L^{pro} um alvo ainda mais interessante, para planejamento de antivirais.

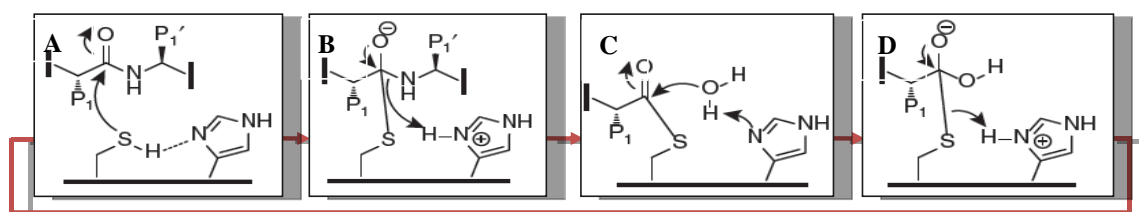


Figura 1.22. Representação esquemática das etapas envolvidas no mecanismo proposto de proteólise para cisteína proteases: (A) Ataque do íon tiolato da Cys51 (nucleófilo) ao carbono carboxílico da ligação peptídica do substrato; (B) Intermediário tetraédrico formado e ligações de hidrogênio com resíduos da cavidade oxânion (Cys51 e Asn46) responsáveis pela estabilização deste intermediário; (C) Perda da porção amina do intermediário tetraédrico, deixando a protease acilada e (D) Reação da protease acilada com uma molécula de água formando ácido carboxílico e restaurando a protease livre (BABINE; BENDER, 1997; TURK, 2006).

As estruturas 3D da Lb^{pro} do vírus da FA foram obtidas por cristalografia de raios-X e RMN, estando disponíveis no PDB. As primeiras estruturas haviam sido obtidas sem ligantes. Em 2014, a estrutura 3D da Lb^{pro} do vírus da FA complexada com um inibidor covalente foi elucidada através de cristalografia de raios-X (STEINBERGER et al., 2014), estando também disponível no PDB. Na literatura, diferentemente da protease do DENV, encontram-se descritos somente alguns poucos inibidores da protease Lb^{pro}, sendo estes o E-64 (N-(trans-epoxisuccinil)-L-leucina 4-guanidinobutilamida) e seus análogos peptídicos (NOGUEIRA

SANTOS et al., 2012; SANTOS et al., 2009; STEINBERGER et al., 2014). Embora a estrutura 3D da Lb^{pro} já tenha sido elucidada e esteja disponível, não há, dentro do nosso conhecimento, relatos da aplicação de estratégias modernas (*SBDD* e *LBDD*) para descoberta de novos inibidores desta protease, em especial a busca virtual. Assim, a disponibilidade de estruturas 3D da Lb^{pro} do vírus da FA fornece subsídios para a aplicação de estratégias modernas de *SBDD* para descoberta de novos inibidores desta protease, em especial a busca virtual.

2. OBJETIVOS

Face ao exposto e considerando a importância da Dengue e da Febre Aftosa, o objetivo principal desta tese foi aplicar estratégias de planejamento racional de fármacos (*LBDD* e *SBDD* - *Ligand and Structure Based Drug Design*), em especial a busca virtual em bancos de dados de compostos comerciais, com o objetivo de gerar modelos visando-se selecionar potenciais inibidores para duas proteases (do vírus da Dengue e do vírus da Febre Aftosa). Ensaio de inibição enzimática foram realizados para validar todos os modelos gerados *in silico*. Em função dos resultados obtidos, visando-se melhorar os modelos de busca virtual, incluíram-se várias modificações. Estas envolveram tanto os filtros de seleção usados (tipos e critérios), bem como a inclusão de outros aspectos importantes para a interação ligante-alvo biológico e para o sucesso da busca virtual.

Adicionalmente, pretendeu-se, a partir dos modelos gerados por busca virtual para as duas proteases, poder contribuir para o entendimento da interação ligante-alvo biológico. Com este objetivo, para a protease do vírus da Dengue, realizaram-se simulações de Dinâmica Molecular visando-se estudar aspectos relacionados à sua flexibilidade conformacional. O conhecimento adquirido permitiu incluir esta flexibilidade nos modelos de busca virtual. Para a protease do vírus da Febre Aftosa, por outro lado, este objetivo foi alcançado usando uma série *in house* de potenciais inibidores covalentes.

Por fim, além de selecionar potenciais inibidores destas proteases, objetivou-se considerar, já nas primeiras etapas da busca virtual, outras características importantes para que um composto bioativo possa, no futuro, se tornar um fármaco. Para isto, construiu-se um banco de dados virtual de compostos enriquecido com compostos com características desejáveis de um fármaco. Visando verificar se este banco foi enriquecido com compostos com características de fármacos sem, no entanto, ter a sua diversidade estrutural significativamente comprometida, ferramentas estatísticas/quimiométricas foram aplicadas.

3. MATERIAIS E MÉTODOS

3.1. Reagentes e Solventes

3.1.1. Ensaios de inibição

- CHAPS: *3-[(3-colamidopropil)-dimetil-amônio]-1-propano-sulfonato* hidratado ($\geq 98\%$ - HPLC, Sigma)
- DMSO: *Dimetilsulfóxido* (P.A., Sigma, para testes biológicos).
- DMSO- d_6 : *Dimetilsulfóxido hexadeuterado* (99, 9 % Sigma, solvente para RMN 1H)
- DTT: *Ditiotreitol* (99%, Sigma), agente redutor de grupos tióis.
- Glicerol: *1, 2, 3 propanotriol* ($\geq 99\%$ Sigma, para biologia molecular)
- NaCl: Cloreto de Sódio (P.A, Synth)
- Borato: *Tetraborato de sódio decahidratado* ($\geq 99, 5\%$ Sigma, reagente ACS)
- Tris base: *tris(hidroximetil)aminometano* ($\geq 99,9\%$ Sigma, para padrão primário e tampões).
- TritonTM-X-100: *4-(1,1,3,3-Tetrametilbutil)phenil-poliethileno glicol* (Sigma)
- Reagente de Purpald®: *4-Amino-3-hidrazino-5-mercapto-1,2,4-triazol* ($\geq 99\%$, Sigma), cortesia de Prof^a. Dr^a. Regina Lúcia Baldini, Departamento de Bioquímica do Instituto de Química da Universidade de São Paulo - SP.

3.1.2. Enzimas

- NS2B/NS3pro do DENV2 (E.C. 3.4.21.98): solução 115 μM , disponível no grupo da Prof^a. Dr^a. Wibke Diederich, Philipps-Universität de Marburg, Marburg, Alemanha.
- NS2B/NS3pro do DENV3 (E.C. 3.4.21.98): solução 197 μM , disponível no grupo da Prof^a. Dr^a. Wibke Diederich, Philipps-Universität de Marburg, Marburg, Alemanha.
- sLb^{pro} do vírus da FA sorotipo Ok1 (E.C. 3.4.22.46): solução 1,69 nM, cortesia de Prof. Dr. Luiz Juliano, Departamento de Biofísica da Universidade Federal de São Paulo - SP.

3.1.3. Substratos

- Phac-KRR-AMC.3TFA, sintetizado e disponibilizado pelo Prof. Dr. Torsten Steinmetzer, Philipps-Universität de Marburg, Marburg, Alemanha.
- Z-LR-AMC: Benzoil-leucina-L-Arginina-L-7-Amido-4-metil-cumarina (Genome, com 98 % de pureza).
- KVQRKLKGAGQSSQ-Abz: sintetizado e disponibilizado pelo Prof. Dr. Luiz Juliano Neto, Departamento de Biofísica da UNIFESP – SP.

3.1.4. Compostos testados frente à NS2B/NS3pro do DENV

- **DV-01:** *N*-(2-cianofenil)-3-(hidrazinocarbonil)propanamida, ZINC01654187 (NCI - NSC99360).
- **DV-02:** 2-cloro-*N*-(4-hidrazinil-6-metilpirimidin-2-il)benzeno-1-sulfonamida ZINC00270493 (Molport - VitasM STK830519).
- **DV-03:** 3-(2-nitrobenzenesulfonamida)benzamida, ZINC6214069 (MolPort - VitasM STK094834).
- **DV-04:** 2-[(2-amino-6-cloropirimidin-4-il)amino]-1-(3,4-difluorofenil) etan-1-ol ZINC29588623 (Mol-Port - Enamine Z397762976).
- **DV-05:** 3-(1,1-dioxo-1λ^σ-tiolan-3-il)-1-{3-[(5-metilpiridin-2-il)amino]propil}urea, ZINC44966262 (MolPort - Enamine Z445985814).
- **DV-06:** 1-[2-(2-fluorofenil)-2-hidroxietyl]-3-(1*H*-pirazol-3-ilmetil)urea, ZINC48260384 (MolPort - Enamine Z511698554).
- **DV-07:** 3-[(4-amino-5-hidrazinil-4*H*-1,2,4-triazol-3-il)sulfanil]-*N*-(2-bromofenil), ZINC20922078 (Scientific Exchange - F295441).
- **DV-08:** (3-[(4-amino-5-hidrazino-1,2,4-triazol-3-il)sulfanil]-*N*-(3-ciano-4,5,6,7-tetrahidrobenzo-tiofen-2-il), ZINC20922301 (Scientific Exchange - F295478).
- **DV-09:** metil 2-[2-[(4-amino-5-hidrazinil-4*H*-1,2,4-triazol-3-il)sulfanil]acetamido]-4,5,6,7-tetrahidro-1-benzotiofene-3-carboxilato, ZINC20922428 (Scientific Exchange - F295499).
- **DV-10:** 2-[(4-amino-5-hidrazinil-4*H*-1,2,4-triazol-3-il)sulfanil]-*N*-[(2-clorofenil)metil]acetamida, ZINC21035463 (Scientific Exchange - M-973226).
- **DV-11:** 2-(1*H*-indol-2-ilformamido)-*N*-(1*H*-pirazol-4-il)acetamida, ZINC06942108 (eMolecules - Enamine Z968599126).
- **DV-12:** 2-cloro-5-[(1*H*-pirazol-3-ilmetil)carbamoil]amino}bezamida, ZINC40540323 (eMolecules - Enamine Z433436158).
- **DV-13:** 4-[(2*E*)-3-fenilprop-2-en-1-il]piperazina-1-carboximidamida, ZINC25605802 (eMolecules - UORSY PB108644332).
- **DV-14:** 2-[(4-cloro-1*H*-pirrol-2-il)formamida]-*N*-(1*H*-pirazol-4-il)acetamida, ZINC58246768 (eMolecules – Enamine Z961789840).
- **DV-15:** *N*-[1-(propan-2-il)piperidin-4-il]naftaleno-1-carboxamida, ZINC12382552 (eMolecules - Vitas M Labs STK492350).
- **DV-16:** *N*-[(2,3-dimetil-1*H*-indol-5-il)metil]-1-(piridin-3-ilmetil)piperidin-4-amina, ZINC12449348 (eMolecules – ChemBridge 40291746).
- **DV-17:** *N*-[(1-benzilpiperidin-4-il)metil]-9*H*-purin-6-amina, ZINC58145695 (eMolecules – Enamine Z838123498).
- **DV-18:** 1-[1-(piridin-2-ilmetil)piperidin-4-il]-3-(tiofen-2-ilmetil)urea, ZINC40549171 (eMolecules – Enamine Z445833352).
- **DV-19:** 1-(1-etilpiperidin-4-il)-3-(1*H*-indol-4-il)urea, ZINC71803380 (eMolecules - Enamine Z1122730913).
- **DV-20:** 4-acetil-*N*-[2-(1*H*-1,3-benzodiazol-2-il)etil]-1*H*-pirrole-2-carboxamida, ZINC69433258 (eMolecules – Enamine Z198287396).
- **DV-21:** 2-fluoro-5-([2-(4-metilpiperidin-1-il)propil]carbamoil)amino}benzamida, ZINC65585171 (eMolecules – Enamine Z902806212).
- **DV-22:** 5-metoxi-*N*-[2-(pirrolidin-1-il)etil]-1*H*-indole-2-carboxamida, ZINC09581802 (eMolecules - Vitas M Labs STK266724).

- **DV-23:** *N*-{3-[(dimetilamino)metil]-1*H*-pirazol-5-il}naftaleno-2-carboxamida, ZINC72258613 (eMolecules - Princeton BioMolecular Research OSSL_717732).
- **DV-24:** *N*-(4-amino-3,5-diclorofenil)-4-[(carbamoilamino)metil] benzamida, ZINC08935785 (eMolecules – Enamine Z167085524).
- **DV-25:** *N*-{3-[(dimetilamino)metil]-1*H*-pirazol-5-il}-4-metoxi-1-benzofuran-6-carboxamida, ZINC72258577 (eMolecules - Princeton BioMolecular Research OSSL_717697).
- **DV-26:** 4-(4-aminobutanamido)-2-clorobenzamida, ZINC37046989 (eMolecules – UORSY PB445569316).
- **DV-27:** 4,6-dimetil-5-[(*E*)-2-fenildiazen-1-il]-2-sulfanilidene-1-[3,4,5-trihidroxi-6-(hidroximetil)oxan-2-il]-1,2-dihidropiridine-3-carbonitrila, (NCI, NSC667731).
- **DV-28:** 4,6-dimetil-5-[(*E*)-2-(4-metilfenil)diazen-1-il]-2-sulfanilidene-1-[3,4,5-trihidroxi-6-(hidroximetil)oxan-2-il]-1,2-dihidropiridine-3-carbonitrila (NCI, NSC667732).
- **DV-29:** 5-[(*E*)-2-(4-clorofenil)diazen-1-il]-4,6-dimetil-2-sulfanilidene-1-[3,4,5-trihidroxi-6-(hidroximetil)oxan-2-il]-1,2-dihidropiridine-3-carbonitrila (NCI, NSC667734).
- **DV-30:** 5-[(*E*)-2-(4-clorofenil)diazen-1-il]-4-metil-6-fenil-2-sulfanilidene-1-[3,4,5-trihidroxi-6-(hidroximetil)oxan-2-il]-1,2-dihidropiridine-3-carbonitrila (NCI, NSC667738).
- **DV-31:** 2-{[9-(3-amino-4-metilfenil)-9*H*-purin-6-il](2-hidroxietyl) amino}etan-1-ol (NCI, NSC100107).
- **DV-32:** 2-[(2-hidroxietyl)(3-nitrofenil)amino]etan-1-ol (NCI, NSC103358).
- **DV-33:** 2-[(2-hidroxietyl)(4-nitrosofenil)amino]etan-1-ol (NCI, NSC104966).
- **DV-34:** 3-metil-3-nitropentano-1,5-diamina hidrocloreto (NCI, NSC116805).
- **DV-35:** (2*R*,3*S*,4*S*,5*R*)-2-(2-amino-6-metoxi-9*H*-purin-9-il)-5-(hidroximetil)oxolane-3,4-diol (NCI, NSC132361).
- **DV-36:** 5-[bis(2-hidroxietyl)amino]pentan-2-ona (NCI, NSC140276).
- **DV-37:** 1-(6-amino-9*H*-purin-9-il)-3-[bis(2-hidroxietyl)amino]propan-2-ol (NCI, NSC371365).
- **DV-38:** *N*-{4-[(4-acetamidofenil)(hidroxi)metil]fenil}acetamida (NCI, NSC4288).
- **DV-39:** 1-(2-hidroxietyl)-3,3-dimetil-octahidro-1*H*-ciclopenta[*c*]piridin-7-ol (NCI, NSC54839).
- **DV-40:** 2-[(2-hidroxietyl){2-[(1-nitroacridin-9-il)amino]etil})amino] etan-1-ol hidrocloreto (NCI, NSC645806).
- **DV-41:** 4-[3-(trifluorometil)fenil]piperidin-4-ol hidrocloreto (NCI, NSC83235).
- **DV-42:** *N*-(2,3-dihidro-1,4-benzodioxin-6-il)-2-(*N*-etil-2-{4-oxo-3*H*,4*H*-tieno[2,3-*d*]pirimidin-3-il}acetamido)acetamida, ZINC9591679 (MolPort, Enamine - Z30787269).
- **DV-43:** 1-(4-metoxifenil)-4-{1-[2-(morfolin-4-il)-2-oxoetil]-1*H*-1,3-benzodiazol-2-il}pirrolidin-2-ona, ZINC15922057 (MolPort, Enamine - Z314606262).
- **DV-44:** *N*-[4-(4-etilpiperazin-1-il)-2-metilfenil]-3-[(oxolan-2-ilmetil)sulfamoil]benzamida, ZINC28731880 (MolPort, Enamine - Z8290229).
- **DV-45:** 2-(4-acetamidofenil)-*N*-[3-(pirrolidin-1-il)butil]quinolina-4-carboxamida, ZINC21662916 (MolPort, ChemDiv. – L036-0422).
- **DV-46:** *N*-(1*H*-imidazol-2-ilmetil)-3-[(4-metilpiperazin-1-il)metil]anilina ZINC80196225 (MolPort, UkrOrgSynthesis PB1323683733).
- **DV-47:** 4-({3*H*-imidazo[4,5-*c*]piridin-2-ilsulfanil}metil)-5-metil-2-(3-metilfenil)-1,3-oxazol, ZINC20865739 (MolPort, ChemDiv. C875-0007).

- **DV-48:** 2-(3-clorofenil)-4-({3H-imidazo[4,5-c]piridin-2-ilsulfanil}metil) -5-metil-1,3-oxazol, ZINC20865752 (MolPort, ChemDiv. C875-0066).
- **DV-49:** 4-(3,4-dietoxifenil)-3-metil-1-[[1,2,4]triazolo[4,3-b]piridazin-6-il]-1H,4H,5H,6H,7H-pirazolo[3,4-b]piridin-6-ona, ZINC33088092 (MolPort, ChemDiv. J081-0806).
- **DV-50:** 1-(4-metoxifenil)-3-{3-oxo-3-[4-(2,4,6-trimetilbenzenesulfonil) piperazin-1-il]propil}urea, ZINC14109553 (MolPort, Enamine Z228049750).
- **DV-51:** 3-[4,6-dimetil-2-(metilsulfanil)pirimidin-5-il]-1-[4-(4-fluorobenzenesulfonil)piperazin-1-il]propan-1-ona, ZINC14070608 (MolPort, Enamine Z90814769).
- **DV-52:** N-(dimetil-1,3-tiazol-2-il)-2-[(5-{[(4-fluorofenil)carbamoil] metil}-4-metil-4H-1,2,4-triazol-3-il)sulfanil]acetamida, ZINC9302446 (MolPort, Vitas-M Laboratory STK219087).
- **DV-53:** ácido 4-{[(2E,5Z)-5-[(3-etoxi-2-hidroxifenil)metilidene]-3-metil -4-oxo-1,3-tiazolidin-2-ilidene]amino}benzóico, ZINC5032850 (MolPort, Vitas-M Laboratory STK698363).
- **DV-54:** ácido 2-cloro-5-{5-[(1E)-2-ciano-2-[(2-feniletil)carbamoil]et-1-en-1-il]furan-2-il}benzoico, ZINC13424134 (MolPort, Vitas-M Laboratory STK506389).
- **DV-55:** ácido 3-[(2E)-2-ciano-3-{4-[(4-fluorofenil)metoxi]-3-metoxifenil}prop-2-enamido]benzoico, ZINC1438749 (MolPort, Vitas-M Laboratory STK522691).
- **DV-56:** N-({3-[2-(dimetilamino)acetamido]fenil}metil)-2-hidroxipiridine-3-carboxamida, ZINC58166454 (MolPort, Enamine Z666062898).
- **DV-57:** 7-[(5-metiltiofen-2-il)](piridin-2-il)amino]metil]quinolin-8-ol, ZINC4123995 (MolPort, Vitas-M Laboratory STK521196).
- **DV-58:** 2-(5-bromo-2-hidroxi-3-metoxifenil)-1,2,3,4-tetrahidroquinazolin-4-ona, ZINC9512483 (MolPort, Enamine Z48978748).
- **DV-59:** 2-(3,5-dicloro-2-hidroxifenil)-1,2,3,4-tetrahidroquinazolin-4-ona, ZINC9609087 (MolPort, Enamine Z48978522).
- **DV-60:** 7-{[(4-metilpiridin-2-il)amino](tiofen-2-il)metil}quinolin-8-ol, ZINC13062711 (MolPort, Enamine Z56842928).
- **DV-61:** 7-[(4-etilfenil)](piridin-2-il)amino]metil]-2-metilquinolin-8-ol, ZINC02073889 (MolPort, Vitas-M Laboratory STK509042).
- **DV-62:** 2-metil-7-{[4-(metilsulfanil)fenil]](piridin-2-il)amino]metil} quinolin-8-ol, CID 2895923 (MolPort, Vitas-M Laboratory STK509186).
- **DV-63:** 7-[(4-etilfenil)](4-metilpiridin-2-il)amino]metil]quinolin-8-ol, CID 2906981 (MolPort, Vitas-M Laboratory, STK509373).
- **DV-64:** 2-metil-7-{[(6-metilpiridin-2-il)amino][4-(metilsulfanil)fenil]metil}quinolin-8-ol, CID 2907327 (MolPort, Vitas-M Laboratory STK521011).
- **DV-65:** 2-metil-7-{[(6-metilpiridin-2-il)amino](tiofen-3-il)metil} quinolin-8-ol, CID 2895602 (MolPort, Vitas-M Laboratory STK520981).
- **DV-66:** 7-{[(4-metilpiridin-2-il)amino](5-metiltiofen-2-il)metil} quinolin-8-ol, CID 3163230 (MolPort, Vitas-M Laboratory STK509557).
- **DV-67:** 7-{[(4-metilpiridin-2-il)amino](piridin-4-il)metil}quinolin-8-ol, CID 2895992 (MolPort, Vitas-M Laboratory STK506802).
- **DV-68:** 2-metil-7-{[(6-metilpiridin-2-il)amino](5-metiltiofen-2-il)metil} quinolin-8-ol, CID 2949438 (MolPort, Vitas-M Laboratory STK522095).
- **DV-69:** 7-[(4-metilfenil)](6-metilpiridin-2-il)amino]metil]quinolin-8-ol, CID 2895523 (MolPort, Vitas-M Laboratory STK325813).

- **DV-70:** 7- $\{[(4\text{-metilpiridin-2-il})\text{amino}](3\text{-metiltiofen-2-il})\text{metil}\}$ quinolin-8-ol, CID 2895553 (MolPort, Vitas-M Laboratory STK325814).
- **DV-71:** 7- $\{[(6\text{-metilpiridin-2-il})\text{amino}](\text{piridin-4-il})\text{metil}\}$ quinolin-8-ol, CID 2895161 (MolPort, Vitas-M Laboratory STK506798).
- **DV-72:** 7- $\{[(4\text{-metilpiridin-2-il})\text{amino}][4\text{-(metilsulfanil)fenil}]\text{metil}\}$ quinolin-8-ol, CID 4354110 (MolPort, Vitas-M Laboratory STK509372)

3.1.5. Compostos testados frente à Lb^{pro} do vírus de FA

- **FA-01:** *N*-(cianometil)-4-etoxibenzamide (MolPort, ChemDiv - Y030-2186).
- **FA-02:** *N*-(cianometil)-3,5-dimetoxibenzamide (MolPort, ChemDiv - Y030-2072).
- **FA-03:** *N*1-(4-ciano-1*H*-pirazol-3-il)-*N*2-(4-metoxiphenetil) oxalamide (MolPort, Vitas M Labs – STK637995).
- **FA-04:** *N*-(cianometil)-8-ethoxi-2*H*-chromeno-3-carboxamide (MolPort, Enamine – Z979002100).
- **FA-05:** 4-acetil-*N*-(cianometil)-3,5-dimetil-1*H*-pirrole-2-carboxamide (MolPort, Enamine – Z979003660).
- **FA-06:** *N*-(1-ciclopropil-5-oxopirrolidin-3-il)-1-metil-2-oxo-1,2-dihidropiridine-3-carboxamide (MolPort, Life Chemicals – F6210-0829).
- **FA-07:** *N*-[2-[1-(2,3-dimetilfenil)tetrazol-5-yl]sulfaniletil] etanosulfonamida, ZINC68044297 (eMolecules – Enamine Z1167063854).
- **FA-08:** 3-metil-5-(9*H*-purin-6-ylsulfanilmetil)isoxazole, ZINC71905440 (eMolecules – Enamine Z1203023757).
- **FA-09:** 5-[(2-fluorophenoxy)metil]-*N*-[(5-isopropil-1,2,4-oxadiazol-3-yl)metil]-1*H*-pirazol-3-carboxamida, ZINC67846966 (eMolecules – ChemBridge 48998181).
- **FA-10:** 2-[metil-[(3-metil-1*H*-1,2,4-triazol-5-yl)metil]amino]-*N*-[2-(*o*-tolil)etil]acetamida, ZINC75107777 (eMolecules – Enamine Z1363072069).
- **FA-11:** 3-morfolinosulfonilbenzenosulfonamida, ZINC69842754 (eMolecules – Enamine Z1137562323).
- **FA-12:** *N*-[[2-(dimetilamino)-4-piridil]metil]-2-metoxi-acetamida (2*E*, 4*E*)-*N*-[3-(2-furil)-1*H*-1,2,4-triazol-5-yl]metil]hexa-2,4-dienamida, ZINC22409626 (eMolecules – Enamine Z237750430).
- **FA-13:** *N*-[[2-(dimetilamino)-4-piridil]metil]-2-metoxi-acetamida, ZINC40512559 (eMolecules – Enamine Z373982642).
- **FA-14:** (5-cloro-2-hidroxi-fenil)-[(2*R*)-2-etilmorfolina-4-yl] metanona, ZINC30597143 (eMolecules – Enamine Z383189730).
- **FA-15:** 2-(1*H*-benzoimidazol-2-ylsulfanil)-*N*-isopentil-acetamida, ZINC03020332 (eMolecules – UORSY PB19727435).
- **FA-16:** *N*-[[3*S*]-1-[(*Z*)-2-metilpent-2-enol]-3-piperidil]metil] metanosulfonamida, ZINC75136897 (eMolecules – Enamine Z438103036).
- **FA-17:** 3-isobutil-*N*-(4*H*-1,2,4-triazol-3-ylmetil)-1*H*-pirazol-5-carboxamida, ZINC69808851 (eMolecules – Enamine Z1175139354).
- **FA-18:** *N*-[(4-cloro-2-metoxi-fenil)metil]metanosulfonamida (eMolecules – Enamine Z1152198122).
- **FA-19:** 2-ciclopropil-5-(9*H*-purina-6-ylsulfanilmetil)-1,3,4-oxadiazol, ZINC71905439 (eMolecules – Enamine Z1203023502).

- **FA-20:** 2-[4-(*aminoBLAHyl*)phenoxi]acetamida, ZINC09318924 (eMolecules – UORSY PB135241250).
- **E-64:** *N*-(*trans*-epoxisuccinil)-*L*-leucina 4-guanidinobutilamida, inibidor peptídico mimético da Lb^{pro} (Sigma Aldrich).

3.2. Equipamentos

3.2.1. Espectrofotômetro UV-VIS

A estabilidade dos inibidores da NS2B/NS3pro do DENV, nas condições do experimento de inibição, foi analisada espectrofotometricamente registrando os espectros na região UV/VIS obtidos em um espectrofotômetro Hitachi U-2010 (UV/VIS).

3.2.2. Fluorímetro e leitor de microplacas

Os testes de inibição enzimática tanto para a NS2B/NS3pro do DENV bem como para Lb^{pro} do FMDV foram realizados nos seguintes aparelhos:

- Fluorímetro marca SHIMADZU modelo RF-5301 PC, equipado com agitador magnético e banho termostatzado, ajustado para $37,5 \pm 0,1^{\circ}\text{C}$.
- Leitor de placa Tecan modelo Saphire², equipado com controle de temperatura, ajustado para $37,0 \pm 0,1^{\circ}\text{C}$, disponibilizado no laboratório da Prof^a. Dr^a. Wibke Diederich, Philipps-Universität de Marburg, Marburg, Alemanha.

3.2.3. pH-âmetro

Os valores de pH das soluções tampão utilizadas nos testes da inibição enzimática da NS2B/NS3pro do DENV e da Lb^{pro} do vírus da FA foram obtidos em um pH-âmetro marca Accumet, modelo AB15 plus, utilizando-se eletrodos de vidro Ag/AgCl, KCl (sat.) e calibrados com soluções tampão padrão.

3.2.4. Espectrômetro de Ressonância Magnética Nuclear (RMN)

As estruturas tanto de alguns dos compostos adquiridos como potenciais inibidores da NS2B/NS3pro do DENV bem como dos seus possíveis produtos de degradação foram identificadas e caracterizadas por RMN¹H (em DMSO-d₆, 99,9 %) utilizando o

espectrômetro Bruker DPX300 de 300 MHz e/ou Bruker DPX300 de 500 MHz da Central Analítica do IQ-USP, SP.

3.2.5. Estações de Trabalho

Os cálculos e procedimentos computacionais foram realizados nas seguintes estações de trabalho:

- Estação Gráfica ASUS P658D-E, com 1 CPU Intel® Core™ i7 980 @3.33GHz (6 cores, 12 threads), 384kB L1-Cache, 1536kB L2-Cache, 12288kB L3-Cache; 12GB RAM, Placa Gráfica NVIDIA GF104 [GeForce GTX 460], com 1GB de memória, 256-bit;
- Estação Gráfica ASUS P9X79 LE, com 1 CPU Intel® Core™ i7-4930K @3.40GHz (6 cores, 12 threads), 32kB L1-Cache, 256kB L2-Cache, 12288kB L3-Cache; 24GB RAM; Placa Gráfica NVIDIA GK104 [GeForce GTX 760], com 2GB de memória;
- Estação Gráfica Silicon Graphics SGI, com 2 CPUs Quad Core Intel Xeon E5462 @2.80GHz, 6144KB Cache (8 cores); 8GB RAM; Placa Gráfica NVIDIA GPU Quadro FX 3700 com 512MB de memória;
- Estação Gráfica Intel DX79SI, com 1 CPU Intel® Core™ i7-3930K @3.20GHz, 12288KB Cache (6 cores, 12 threads); 48GB RAM; Placa gráfica NVIDIA GF106GL (Quadro 2000), com 1024MB de memória, 128-bit;
- Servidor Silicon Graphics SGI Altix XE320, com 2 CPUs Quad Core Intel Xeon E5472 @3.00Ghz, 6144KB Cache (8 cores); 16GB RAM; 2 HDs 750GB; Unidade de Armazenamento SGI InfiniteStorage 220 com 6 HDs 750GB SATA;
- Servidor Silicon Graphics SGI Altix XE320, com 2 CPUs Quad Core Intel Xeon E5472 @3.00Ghz, 6144KB Cache (8 cores); 16GB RAM; 2 HDs 750GB.

As simulações de dinâmica molecular foram realizadas nas seguintes estações de trabalho, disponíveis no grupo do Prof. C. A. Sotriffer, Universidade de Würzburg, Würzburg, Alemanha:

- Servidor: 16 nodes com 4 GB RAM e processador 2 Dual-Core AMD Opteron, 2214 CPUs, 8 GB RAM e 2 Quad-Core AMD Opterons 2352
- Servidor: 4 nodes com 32 GB RAM cada um com processador 12-core AMD Opteron.
- Servidor: 1 node com 1 AMD Opteron Octa-core, 2 processadores 4 NVIDIA Tesla M2050 GPUs.

Os procedimentos de ancoramento de todo banco de *ZINC-Curated* foram realizados utilizando:

- Cluster Aguia do Laboratório de Computação Científica Avançada (LCCA) da USP, com 48 servidores cada um com 20 CPUs Intel(R) Xeon(R) CPU E7- 2870 @ 2.40GHz, 512 GB de RAM, 22 TB para arquivos temporários.

3.3. Banco de dados de moléculas virtuais, séries de compostos e estruturas tridimensionais das proteases virais

3.3.1. Banco de dados de moléculas virtuais utilizados na busca virtual

Na busca virtual foram utilizados três bancos de dados de composto:

- *ZINC-Original*: ZINC (IRWIN et al., 2012) versão 12, ($\sim 23 \times 10^6$ estruturas), disponíveis para *download* no endereço eletrônico <http://zinc.docking.org/browse/> (em Janeiro de 2011), utilizado como obtido da internet;
- *ZINC-Curated*: ZINC (IRWIN et al., 2012) versão 13, ($\sim 20 \times 10^6$ estruturas, em pH = 7), disponíveis para *download* no endereço eletrônico <http://zinc.docking.org/browse/> (em Janeiro de 2013), submetido a um pré-tratamento;
- *NCI*: NCI ($\sim 281 \times 10^3$ estruturas), disponível para *download* no endereço eletrônico <http://dtp.cancer.gov> (em Outubro de 2014).

3.3.2. Bancos de dados de moléculas virtuais utilizados na análise da diversidade química e estrutural

A análise da diversidade química e estrutural no banco de dados *ZINC-Curated* foi feita utilizando quatro bancos de dados:

- *DB*: DrugBank (Wishart et al., 2006), contendo 1.826 pequenas moléculas aprovadas pelo FDA, disponível para *download* no endereço eletrônico <http://www.drugbank.ca>, (versão 4.3, em 02/2016);
- *ZNP*: ZINC Natural Products, subconjunto de produtos naturais do banco de dados ZINC (Irwin, Sterling, Mysinger, Bolstad, & Coleman, 2012), contendo 166.544 moléculas e incluindo doze catálogos de produtos naturais (AfroDb Natural Products; AnalytiCon Discovery NP; Herbal Ingredients In-Vivo Metabolism; Herbal Ingredients Targets; IBScreen NP; Indofine Natural Products; NPACT Database; Nubbe Natural Products; Princeton NP; Specs Natural Products; TCM Database @ Taiwan e o UEFS Natural Products), disponíveis para *download* no endereço eletrônico <http://zinc.docking.org/browse/catalogs/natural-products> (em 08/ 2015);
- *ZINC-2014*: Subconjunto “*All Purchasable*” do banco de dados ZINC (Irwin et al., 2012), contendo $\sim 23 \times 10^6$ moléculas, disponível para *download* no endereço eletrônico <http://zinc.docking.org/subsets/all-purchasable> (em 11/2014);

- ChEMBL: ChEMBL (HERSEY, 2015), contendo $\sim 1,4 \times 10^6$ moléculas, disponível para *download* no endereço eletrônico <https://www.ebi.ac.uk/chembl/downloads>, (versão 20, em 01/2015).

3.3.3. Séries de inibidores conhecidos da NS2B/NS3pro do DENV

A validação *in silico* de modelos de busca virtual (Modelo III – DENV e Modelo IV – DENV) foi feita utilizando duas séries de inibidores conhecidos da NS2B/NS3pro do DENV (compostos peptídicos e não peptídicos). As estruturas dos inibidores bem como seus valores de atividade foram obtidos das seguintes fontes:

- De artigos da literatura (ARAVAPALLI et al., 2012; BASTOS LIMA et al., 2015; BEHNAM et al., 2014, 2015; BODENREIDER et al., 2009; DE SOUSA et al., 2015; DENG et al., 2012; JÖST et al., 2014; KIAT et al., 2006; KNEHANS et al., 2011; KOH-STENTA et al., 2015; LAI et al., 2013; LAI; SRIDHAR PRASAD; PADMANABHAN, 2013; MUELLER et al., 2008; NGUYEN et al., 2013; NITSCHKE et al., 2012, 2013; NITSCHKE; STEUER; KLEIN, 2011; RAUT et al., 2015; SCHÜLLER et al., 2011; SHIRYAEV et al., 2011; STEUER et al., 2011; TIEW et al., 2012; TOMLINSON; WATOWICH, 2011; VISWANATHAN et al., 2014; YIN et al., 2006; ZHOU et al., 2013)
- Banco de dados ChEMBL (BENTO et al., 2014), disponível para consulta no endereço eletrônico <http://www.ebi.ac.uk/chembl>, acessado no dia 06/12/2016.

3.3.4. Série “*in house*” de potenciais inibidores covalentes

Uma série de 52 compostos comerciais previamente selecionados, por busca virtual (MALVEZZI; DE REZENDE; AMARAL, 2009), como potenciais inibidores covalentes da cruzaina estava disponível em nosso laboratório e foi usada tanto para identificar inibidores da Lb^{pro} do vírus da FA bem como para ter informações sobre as interações dentro do sítio ativo.

3.3.5. Estruturas tridimensionais das proteases virais

As estruturas tridimensionais das proteínas NS2B/NS3pro do DENV (código PDB: 3U1I, resolução 2,30 Å (NOBLE et al., 2012)) e Lb^{pro} do vírus da FA (códigos PDB: 2JQG, obtida por RMN, e 4QBB, obtida por cristalografia de Raio-X com 1,6 Å de resolução (CENCIC et al., 2007; STEINBERGER et al., 2014)), usadas neste trabalho, foram obtidas no banco de dados *Protein Data Bank (PDB)* (BERNSTEIN, F., KOETZLE, T., WILLIAMS, 1977); disponível para consulta no endereço eletrônico <http://www.rcsb.org>.

3.4. Programas Computacionais

Os programas e ferramentas computacionais utilizados nesta tese bem como uma breve descrição da etapa na qual cada um(a) deles(as) foi usada, acompanhada da versão, do fornecedor e do tipo de licença, estão apresentados na Tabela 2.1.

Tabela 2. 1. Programas e ferramentas computacionais utilizadas nesta tese, descrição da etapa na qual cada uma foi usada, versão, fornecedor e tipo de licença.

Programa/Ferramenta	Versão	Descrição da etapa na qual foi usada	Fornecedor/ Tipo de licença
<i>AMBER</i> (SALOMON-FERRER; CASE; WALKER, 2013)	12	Preparação, neutralização, solvatação e minimização das estruturas utilizadas nas simulações de DM	Disponível no grupo do Prof. Dr. C. A. Sotriffer
<i>Checkmol</i> (HAIDER, 2010)	0.5	Cálculo do número de anéis aromáticos de cada molécula do banco ZINC, usado no filtro de solubilidade aplicado ao banco de dados <i>ZINC-Curated</i>	Distribuído gratuitamente no endereço eletrônico http://merian.pch.univie.ac.at/pch/download/chemistry/checkmol
<i>Corina</i> (SADOWSKI; GASTEIGER, 1993)	3.20	Conversão entre os diferentes formatos de arquivos de moléculas e o cálculo de cargas formais	Comercializado pela <i>Molecular Networks</i> (www.molecular-networks.com)
<i>E-dragon</i> (TETKO et al., 2005)	1.0	Cálculo de descritores moleculares	Disponível gratuitamente no endereço eletrônico http://www.vcclab.org/lab/edragon/
<i>Filter</i> (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]), <i>Fred</i> (JONES et al., 1997; JONES; WILLETT; GLEN, 1995)	2.1.1	Cálculo das propriedades físico-químicas e estruturas, utilizadas na redução e na análise da diversidade estrutural do banco de dados <i>ZINC-Curated</i>	Licença educacional fornecida pela <i>OpenEye Scientific Software</i>
	3.0.1	Procedimentos de ancoramento utilizando um grande número de compostos ($> 1 \times 10^6$ compostos)	Licença educacional fornecida pela <i>OpenEye Scientific Software</i>
<i>Gaussian</i> (FRISCH et al., 2009)	09	Cálculo das cargas dos resíduos do inibidor Bz-nKRR	Licença institucional disponibilizada pelo Centro de Computação Eletrônica da USP
<i>GOLD</i> (JONES et al., 1997; JONES; WILLETT; GLEN, 1995)	5.0.1	Procedimentos de ancoramento usando um pequeno número de compostos ($< 1 \times 10^6$ compostos)	Comercializado pela CCDC [CCDC, 2006]

Tabela 2. 2. (Continua)

Programa/Ferramenta	Versão	Descrição da etapa na qual foi usada	Fornecedor/ Tipo de licença
<i>GRID</i> (GOODFORD, 1985)	22c	Obtenção dos valores dos campos de interação moleculares para a proteína Lb ^{pro} do vírus da FA	Comercializado pela Molecular Discovery
<i>Icon</i> disponível no programa <i>LigandScout</i>	4.01	Obtenção das conformações dos inibidores peptídicos, utilizados para validação <i>in silico</i> dos modelos farmacofóricos construídos para DENV	Comercializado pela Inte:Ligand
<i>KNIME</i> (BERTHOLD et al., 2007) com extensões do programa <i>LigandScout</i>	3.2.1/4.0.1	Validação dos modelos farmacofóricos gerados utilizando dados da literatura	Disponível gratuitamente no endereço eletrônico http://www.knime.org e comercializadas pela Inte:Ligand
<i>LigandScout</i> (WOLBER; LANGER, 2005)	3.01/3.1/ 4.1	Conversão entre formatos de arquivos de moléculas (.SDF e .LDB), construção e aplicação dos modelos farmacofóricos aos banco de dados de compostos	Comercializado pela Inte:Ligand
<i>MarvinSketch</i> ("MarvinSketch", 2017)	17.1.2.0	Preparação das imagens das estruturas 2D dos compostos e obtenção dos nomes IUPAC	Licença educacional fornecida pela ChemAxon Ltd. (http://www.chemaxon.com)
<i>Modeler</i> (ESWAR et al., 2006)	9.13	Construção dos resíduos ausentes na estrutura 3U1I da NS2B/NS3pro do DENV	Licença educacional fornecida pelo <i>California Institute for Quantitative Biomedical Research</i> , Universidade da Califórnia, São Francisco, Estados Unidos
<i>MOE</i> (CHEMICALCOMPUTING GROUP INC., 2004)		Preparação da estrutura do inibidor Bz-nKRR utilizadas nas simulações de dinâmica molecular	Disponível no grupo do Prof. Dr. C. A. Sotriffer
<i>Moka</i> (MILLETTI et al., 2007)	2.0/2.6	Cálculo dos valores do logD (em pH 7,4) e obtenção dos tautômeros e dos diferentes estados de protonação dos compostos	Comercializado pela Molecular Discovery
<i>NAMD</i> (PHILLIPS et al., 2005)		Realização das simulações de dinâmica molecular	Disponível no grupo do Prof. Dr. C. A. Sotriffer
<i>Omega</i> (HAWKINS et al., 2010)	2.4.3/2.4.6 /2.5.1	Obtenção das conformações dos compostos dos bancos de dados virtuais e de pequenas moléculas da literatura	Licença educacional fornecida pela OpenEye Scientific Software

Tabela 2. 3. (Continua)

Programa/Ferramenta	Versão	Descrição da etapa na qual foi usada	Fornecedor/ Tipo de licença
<i>Open Babel</i> (O'BOYLE et al., 2011)	2.3.1	Cálculo dos <i>fingerprints</i> moleculares e remoção de moléculas duplicadas na etapa de criação dos bancos de dados virtuais	Distribuído gratuitamente no endereço eletrônico http://openbabel.org
<i>ProteinsPlus Server</i>		Detecção de sítios de ligação da NS2B/NS3pro e cálculo de seus valores <i>drugScore</i> utilizando a ferramenta <i>DoGSiteScorer</i> (VOLKAMER et al., 2012)	Disponível gratuitamente no endereço eletrônico http://proteinsplus.zbh.uni-hamburg.de/
<i>PyMOL</i> (SCHRÖDINGER, 2013)	1.7	Visualização e preparação das imagens dos arquivos <i>PDB</i>	Licença educacional fornecida pela <i>DeLano Scientific LLC</i>
<i>R.E.D.</i> (DUPRADEAU et al., 2010)	III.5	Cálculo das cargas dos resíduos do inibidor Bz-nKRR utilizadas na simulação de dinâmica molecular	Distribuído gratuitamente no endereço eletrônico http://q4md-forcefieldtools.org/RED/ .
<i>R</i> (INDEX; DEVELOPMENT; TEAM, 2009),	3.0/3.2/ 3.3	Análises quimiométricas (agrupamentos hierárquicos, Análise de Componentes Principais, cálculo dos coeficientes de Tanimoto e de índices de similaridade) e determinação dos valores de IC ₅₀	Distribuído gratuitamente no endereço eletrônico http://www.vps.fmvz.usp.br/CRAN/
<i>Sample Size Calculator</i>		Cálculo do tamanho das amostras dos bancos de dados virtuais de moléculas, usado na análise da diversidade química e estrutural do banco <i>ZINC-Curated</i>	Ferramenta <i>on-line</i> disponível gratuitamente no endereço eletrônico http://www.surveysystem.com/sscalc.htm
<i>Sybyl</i> (SYBYL, 2009)/ <i>Sybyl-X</i> (SYBYL, 2013)	8.0/2.1.1	Preparação dos arquivos das proteínas (NS2B/NS3pro e Lb ^{pro}) e dos ligantes (compostos da literatura e selecionados nesta tese de doutorado) e cálculo da superfície do sítio ativo da NS2B/NS3pro	Comercializados pela Certara
<i>VMD</i> (HUMPHREY; DALKE; SCHULTEN, 1996)	1.9.1	Análise das simulações de dinâmica molecular, incluindo o cálculo dos valores de 2D-RMSD, de ocupância da ligação de hidrogênio e de distância das interações iônicas e cátion- π	Distribuído gratuitamente no endereço eletrônico http://www.ks.uiuc.edu/Research/vmd/mainpage
<i>Volsurf+</i> (CRUCIANI et al., 2000)	1.0.7.1	Cálculo dos descritores moleculares	Comercializado pela MolDiscovery

3.5. Procedimentos computacionais

Nesta tese, os métodos utilizados para a busca virtual de inibidores da protease do vírus da Dengue (NS2B/NS3pro do DENV) e da protease líder do vírus da Febre Aftosa (Lb^{pro} do vírus da FA) serão apresentados, sempre, na sequência NS2B/NS3pro do DENV e Lb^{pro} do vírus da FA.

3.5.1. Construção dos bancos de dados virtuais de moléculas

3.5.1.1. Busca virtual

A busca virtual envolveu, no geral, a aplicação de modelos farmacofóricos em diferentes bancos de dados virtuais de moléculas (*ZINC-Original* (IRWIN et al., 2012), *ZINC-Curated* e *NCI*). Para tanto, foi necessário converter estes bancos de dados para o formato *LDB*, requerido pelo programa *LigandScout* (WOLBER; LANGER, 2005). Adicionalmente, o banco *ZINC-Curated* foi submetido a procedimentos de ancoramento sendo necessário converter este banco no formato *oeb.gz*, requerido pelo programa *FRED* (MCGANN, 2011).

ZINC-Original e *NCI*

A conversão foi feita utilizando todas as moléculas dos bancos ZINC (IRWIN et al., 2012) e NCI, obtidas diretamente da internet nos formatos *SDF* ou *SMILE*, e usando em conjunto os programas *Omega* (HAWKINS et al., 2010) e *Idbgen* (pacote *LigandScout*). Nesta etapa, o programa *Omega* gerou diferentes conformações (até 30 para cada molécula) e o programa *Idbgen* removeu entradas duplicadas e salvou o banco de dados no formato *LDB*. Ainda, para as moléculas no formato *SMILE*, as estruturas 3D foram obtidas utilizando o programa *Corina* (SADOWSKI; GASTEIGER, 1993). Todas as etapas envolvidas na criação destes bancos de dados foram realizadas, em *batch*, utilizando as estações gráficas Silicon Graphics SGI e Intel DX79SI.

ZINC-Curated

Como discutido na Introdução, a conversão foi feita após remover compostos com propriedades e/ou estruturas indesejadas para um composto “*drug-like*” usando dois filtros, a saber: filtro de predição solubilidade e filtro de propriedades “*drug-like*” e estruturais. As etapas envolvidas na construção do banco *ZINC-Curated* estão representadas na Figura 3.1.

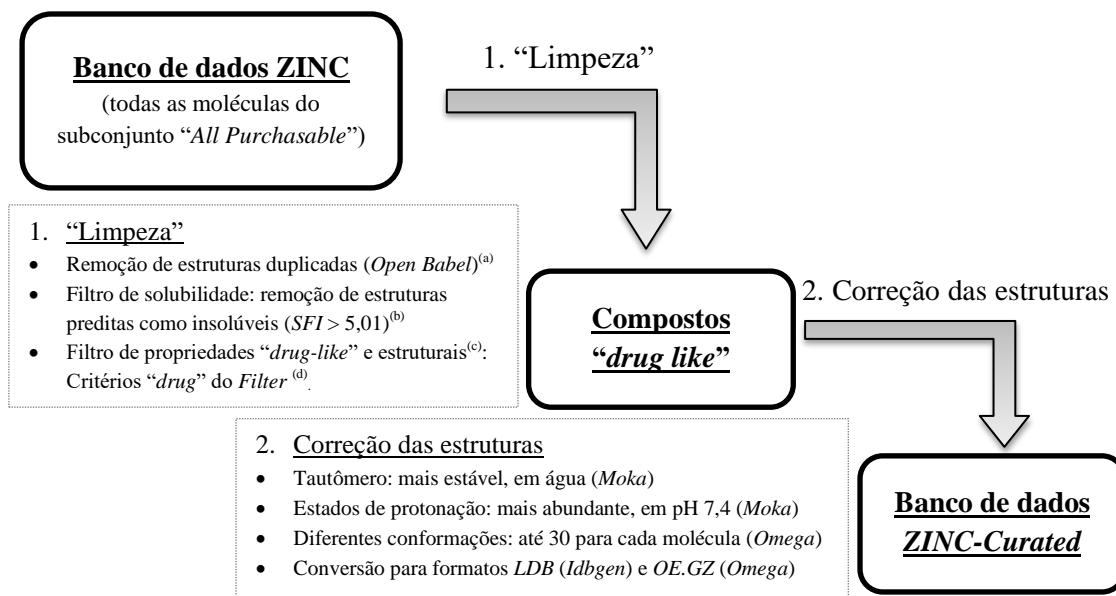


Figura 3.1. Representação esquemática das duas etapas envolvidas na criação do banco de dados *ZINC-Curated*. Todas estas etapas foram feitas, em *batch*, utilizando as estações gráficas ASUS P658D-E e Intel DX79SI.

- ^(a) Compostos apresentando diferentes estados de protonação e/ou diferentes isótopos foram considerados estruturas duplicadas;
- ^(b) O *SFI* (*Solubility Forecast Index* (HILL; YOUNG, 2010)) - discutido em detalhes na Introdução - foi calculado usando os valores de $\text{LogD}_{\text{pH}7,4}$ e o número de anéis aromáticos calculados pelos programas *Moka* (MILLETTI et al., 2007) e *Checkmol* (HAIDER, 2010), respectivamente;
- ^(c) Para aplicar o programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]), as cargas formais dos compostos foram calculadas usando o *Corina* (SADOWSKI; GASTEIGER, 1993);
- ^(d) Todos os critérios padrão - vide Anexo I - foram usados, exceto para o cálculo da solubilidade e da agregação (ambos desabilitados) e para os parâmetros “MAX_2D_PSA”, “MAX_LIPINSK”, “RULE dye” (modificados de 150 para 160, de 1 para 2 e de 0 para 5, respectivamente).

3.5.1.2. Análise da diversidade química e estrutural do banco *ZINC-Curated*

A análise da diversidade química e estrutural do banco de dados *ZINC-Curated* foi feita para os bancos virtuais de moléculas, a saber: DB, ZNP, ZINC-2014 e ChEMBL. Para cada um destes bancos, estruturas duplicadas foram removidas utilizando o programa *Open Babel* (O’BOYLE et al., 2011). Em seguida, as cargas formais de cada composto foram

calculadas usando o programa *Corina* (SADOWSKI; GASTEIGER, 1993). Ainda, para o banco DB, foram removidos todos os compostos classificados como nutracêuticos (ex. vitaminas).

Em seguida, amostras representativas de cada banco foram geradas. Para tanto, o tamanho da amostra de cada um dos bancos foi calculado utilizando a ferramenta *on-line* “*Sample Size Calculator*” e aplicando um intervalo de confiança de 1% com nível de confiança de 95%. Neste cálculo, foi considerado o número de compostos de cada um dos bancos virtuais de moléculas após a remoção das estruturas duplicadas. Por fim, para cada banco virtual de moléculas foram geradas três amostras (denominadas Amostra 1, 2 e 3) selecionando, aleatoriamente, os compostos não repetidos. Todas as análises apresentadas a seguir (item 3.5.3) foram feitas utilizando estas amostras.

3.5.2. Preparação das séries de compostos

3.5.2.1. Séries de inibidores conhecidos da NS2B/NS3pro do DENV

Nesta tese, duas séries de inibidores conhecidos da NS2B/NS3pro do DENV foram construídas, a partir de 1.360 (compostos peptídicos e não peptídicos) e usadas, posteriormente, para a validação *in silico* do desempenho dos modelos de busca virtual (Modelo III – DENV e Modelo IV – DENV). Origem e classificação dos inibidores conhecidos estão apresentadas na Figura 3.2.

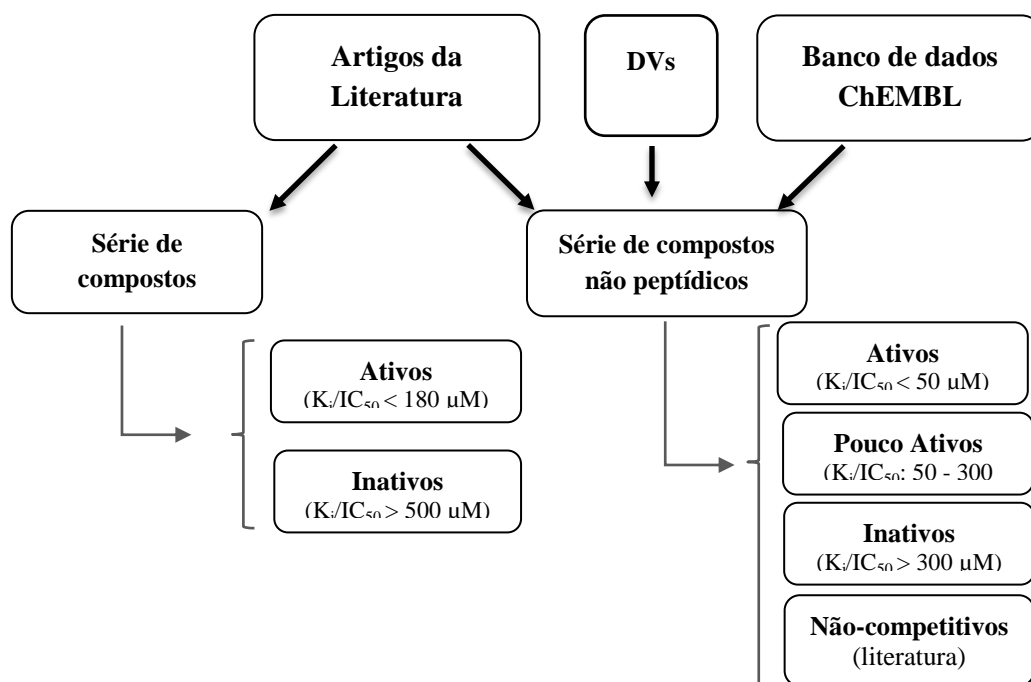


Figura 3.2. Origem e classificação dos inibidores conhecidos da NS2B/NS3pro do DENV. Os inibidores conhecidos foram retirados de artigos da literatura (vide item 3.5.1.2) e/ou do banco de dados ChEMBL (BENTO et al., 2014)^(a). Ainda, os dados de alguns compostos selecionados, por mim, e testados frente à NS2B/NS3pro (denominados DVs) foi incluída. Os inibidores foram agrupados em peptídicos e não peptídicos bem como classificados quanto a sua atividade frente à NS2B/NS3pro do DENV em ativos ($K_i/IC_{50} < 50 \mu M$), pouco ativos (K_i/IC_{50} entre 50 e 300 μM), inativos ($K_i/IC_{50} > 300 \mu M$) e não-competitivos (segundo a literatura).

^(a) Busca realizada, em dezembro de 2016, selecionando todos os compostos testados frente à NS2B/NS3pro do DENV. Em seguida, os resultados foram restringidos para aqueles que apresentavam valores de IC_{50} .

As estruturas 3D dos inibidores da literatura e dos DVs foram obtidas com o auxílio do programa *Sybyl-X* (SYBYL, 2013). Para os inibidores peptídicos, a estrutura tridimensional foi obtida usando, como ponto de partida, o inibidor Bz-nKRR-H na conformação observada na estrutura 3U1I. E, para os inibidores obtidos do banco de dados ChEMBL (BENTO et al., 2014), as estruturas 3D foram geradas utilizando o programa *Corina* (SADOWSKI; GASTEIGER, 1993), a partir do formato *SMILE* obtido diretamente deste banco de dados. Em seguida, para cada um dos compostos foram geradas diferentes conformações (até 900) usando para os compostos pequenos o programa *Omega* (HAWKINS; NICHOLLS, 2012) e para os compostos peptídicos a ferramenta *iCon* (pacote *LigandScout* (Wolber & Langer, 2005)). A conformação do inibidor Bz-nKRR-H encontrada na estrutura cristalográfica 3U1I foi incluída.

3.5.2.2. Série “*in house*” de potenciais inibidores covalentes

A estrutura 3D de cada um dos 52 potenciais inibidores covalentes foi gerada usando o programa *Corina* (SADOWSKI; GASTEIGER, 1993). A seguir, para cada uma das estruturas o tautômero mais estável, em água, foi calculado pelo *Moka* (MILLETTI et al., 2007).

3.5.3. Análise da diversidade química e estrutural do banco *ZINC-Curated*

3.5.3.1. Descritores moleculares

Para cada amostra, descritores moleculares foram calculados, inicialmente, usando o programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]). Nesta etapa, compostos com massa molecular maior que 1000 foram removidos para prevenir erros na próxima etapa. Compostos com peso molecular menor que 1000 foram importados para o programa *Volsurf+* (CRUCIANI et al., 2000), normalizando o estado de protonação (pH 7), utilizando um “*grid spacing*” de 0.5 com “*GRID MIFs parametrization*” estático. Em seguida, o programa *Volsurf+* (CRUCIANI et al., 2000) calculou 128 descritores moleculares (vide Anexo II) utilizando, ou não, os mapas tridimensionais dos campos de interação molecular dos compostos.

3.5.3.2. Validação da etapa de amostragem

A validação da etapa de amostragem foi realizada aplicando testes estatísticos e utilizando 9 propriedades físico-químicas e/ou estruturais foram calculadas pelos programas *Volsurf+* (CRUCIANI et al., 2000) e *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]), a saber:

- Propriedade do *Volsurf+*: logaritmo do coeficiente aparente de partição em pH 7,5 ($\text{LogD}_{\text{pH}7,5}^{\text{Volsurf+}}$)
- Propriedades do *Filter*: peso molecular; logaritmo do coeficiente de partição *n*-octanol/água calculado, usando o algoritmo XlogP (método por fragmento); área de superfície polar topológica (X2d.PSA); número de átomos; número de anéis; número de

ligações com livre rotação; número de doadores e de aceptores de ligação de hidrogênio (do inglês, *HBD* e *HDA*, considerando a definição de Lipinski).

A escolha do teste estatístico a ser utilizado foi feita com base no tipo de distribuição de cada um dos descritores/propriedades. Para tanto, fez-se necessário avaliar se os descritores/propriedades seguem, ou não, uma distribuição próxima da distribuição Normal. Esta avaliação foi feita utilizando gráficos de probabilidade Normal (MAGALHÃES; DE LIMA, 2015; STEHLÍK; STŘELEČEK; THULIN, 2014), construídos usando a função *qqnorm* do programa *R* (R CORE TEAM, 2013). Ainda, foi aplicado o teste *Kolmogorov-Smirnov* (bilateral uma amostra com $\alpha = 0,05$ e $D_{\text{crit}} = 0,03$) para confirmar a hipótese de que os descritores seguem, ou não, uma distribuição próxima da Normal. Com base nos resultados obtidos, foi escolhido e aplicado o teste estatístico de *Kruskal-Wallis* (Análise de variância de uma classificação por postos) utilizando a função *kruskal.test* do programa *R* (R CORE TEAM, 2013) e considerando um nível de significância de 95% ($\alpha = 0,05$). Este teste verificou se as amostras vêm, ou não, de uma mesma população comparando os valores de suas medianas. Por fim, visando verificar se as amostras são, ou não, representativas da população, cada uma destas nove propriedades foram calculadas para todas as moléculas do banco DB. Em seguida, para cada propriedade, a distribuição dos valores calculados usando todo o banco DB foi comparada com os calculados usando as Amostras 1, 2 e 3. Para tanto, foram gerados gráficos de porcentagem cumulativa e de caixa do tipo *Tukey* (vide Anexo III para mais detalhes) usando o programa *R* (R CORE TEAM, 2013). Com base nos resultados obtidos, a etapa de amostragem foi considerada validada e a amostra a ser usada nas análises subsequentes foi escolhida, sorteando-se, randomicamente, um número de 1 a 3 correspondendo ao número da amostra.

3.5.3.3. Propriedades físico-químicas e estruturais

Inicialmente, as nove propriedades físico-químicas e/ou estruturais usadas na etapa de validação da amostragem foram usados para comparar os bancos de dados. Para isso, foram

gerados gráficos de porcentagem cumulativa e de caixa do tipo *Tukey* (vide Anexo III) usando o programa *R* (R CORE TEAM, 2013). Ainda, os valores de mediana calculados para os bancos de dados foram comparados, par-a-par, usando o teste estatístico não paramétrico de *Mann-Whitney U*, usando um $\alpha = 0,05$. Nesta análise foram feitas comparações múltiplas, portanto, fez-se necessário aplicar aos p-valores a correção de *Bonferroni*, como sugerido pela literatura (KRZYWINSKI; ALTMAN, 2014; WASSERMANN et al., 2015). Por fim, valores de logP foram calculados usando o programa *Volsurf+* ($\log P^{Volsurf+}$) e comparados com os valores de $\log D_{pH7,5}^{Volsurf+}$ visando analisar a contribuição da ionização para partição.

3.5.3.4. Espaço químico

A análise e a comparação do espaço químico dos diferentes bancos foram feitas utilizando Análise de Componentes Principais (ACP) dos descritores moleculares gerados pelo programa *Volsurf+* (CRUCIANI et al., 2000). Para isto, os descritores moleculares foram submetidos a um pré-tratamento (Figura 3. 3).

A seguir, a Análise de Componentes Principais foi feita utilizando a função *prcomp* do programa *R* (R CORE TEAM, 2013). Os gráficos dos escores e pesos das Componentes Principais (CP) 1 e 2 foram usados para comparar os espaços químicos dos bancos de dados.

Pré-tratamento dos descritores

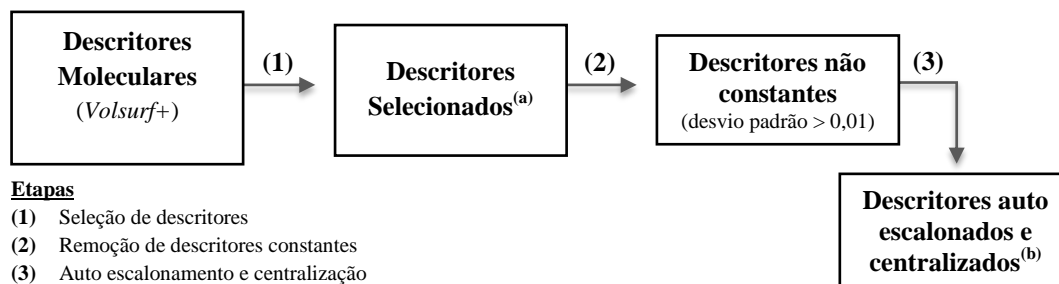


Figura 3. 3. Pré-tratamento dos descritores moleculares usados na análise e comparação do espaço químico dos diferentes bancos através da ACP.

^(a) A seleção de descritores foi feita removendo aqueles representando a mesma propriedade calculada em diferentes condições. Para descritores calculados usando diferentes valores de pH, apenas o descritor com valor calculado no pH = 7,5, ou o mais próximo a 7,5, foi selecionado. Para descritores calculados usando diferentes níveis de energia apenas aqueles representando os valores extremos dos níveis de energia foram considerados.

^(b) Os descritores foram auto escalonados (vide Anexo III) e centralizados, usando a funções *center* e *scale* do programa *R* (R CORE TEAM, 2013).

3.5.3.5. Similaridade estrutural

A similaridade estrutural entre os bancos virtuais de moléculas foi analisada utilizando os correspondentes coeficientes de Tanimoto calculados a partir dos *fingerprints* moleculares. Os *fingerprints* moleculares, do tipo FP2, foram calculados pelo programa *Open Babel* (O'BOYLE et al., 2011). A seguir, estes foram usados para calcular os coeficientes de Tanimoto (Tc), par-a-par, utilizando a função *fpSim* do pacote *ChemmineR* (CAO et al., 2008) disponível no programa *R* (R CORE TEAM, 2013). Para analisar a similaridade estrutural dentro de cada banco virtual, Tc foram calculados considerando todos os possíveis pares de compostos presentes no banco. Para analisar a similaridade entre diferentes bancos, os valores do Tc foram calculados para todos os possíveis pares de compostos de dois bancos diferentes, considerando todas as possíveis combinações. A comparação da similaridade estrutural dentro de cada banco foi feita através da análise de gráficos de caixa do tipo *Tukey*, enquanto que a comparação da similaridade estrutural entre diferentes bancos foi feita contando o número de compostos similares entre dois bancos (compostos com valores de Tc iguais ou maiores que um determinado valor de corte) e calculando a medida de similaridade *Hausdorff-like*, *HLS*, (MAURI et al., 2016). O número de compostos similares foi calculado usando seis diferentes

valores de corte de Tc (1,0; 0,9; 0,8; 0,7; 0,6 e 0,5) e usando o banco DB como referência. O cálculo da medida de Similaridade Hausdorff-like (MAURI et al., 2016), descrita no Anexo III, foi feito para todos os possíveis pares de banco de dados.

3.5.4. Busca virtual de inibidores da protease do vírus da Dengue

Nesta tese, foram construídos e utilizados quatro diferentes modelos de buscas de inibidores da protease do vírus da Dengue (denominados Modelo I – DENV a Modelo IV - DENV). Como descrito na Introdução, estes modelos foram feitos aplicando-se, em sequência, diferentes filtros de seleção. Inicialmente, será apresentada uma descrição geral dos procedimentos computacionais usados na construção destes modelos. Em seguida, será feita uma descrição detalhada dos procedimentos computacionais usando em cada um dos quatro modelos de busca virtual, incluindo os tipos e a ordem dos filtros de seleção aplicados.

3.5.4.1. Preparação da estrutura da protease do vírus da Dengue.

A estrutura tridimensional da NS2B/NS3pro do DENV, código PDB 3U1I e resolução 2,8 Å (NOBLE et al., 2012), obtida diretamente do PDB foi preparada utilizando o programa *Sybyl* (SYBYL, 2009). Nesta etapa, todas as moléculas de água foram retiradas e todos os hidrogênios (polares e apolares) foram adicionados considerando: os resíduos de aminoácidos básicos (arginina e lisina) e o resíduo *N*-terminal protonados (carga formal = + 1); os resíduos de aminoácidos ácidos (ácido aspártico e ácido glutâmico) e o resíduo *C*-terminal desprotonados (carga formal = - 1) e os resíduos de histidina em sua forma neutra (carga formal = 0). Para todos os procedimentos de ancoramento, o inibidor Bz-nKRR-H foi removido e a proteína foi convertida para o formato *MOL2*. Nesta estrutura, as regiões NS2B e NS3 correspondem às cadeias *A* e *B*. Assim, os resíduos dos aminoácidos da região NS2B e NS3 serão denominados como *A* e *B*, respectivamente.

3.5.4.2. Construção e aplicação de modelos farmacofóricos

Os modelos farmacofóricos foram gerados a partir de estruturas da NS2B/NS3pro complexadas com diferentes ligantes, obtidas experimentalmente (complexo 3U1I do PDB) bem como por procedimentos computacionais (ancoramento e simulações de dinâmica molecular, descritos a seguir) preparadas de modo análogo ao descrito no item 3.5.4.1. Todos os modelos farmacofóricos foram gerados utilizando o programa *LigandScout* (WOLBER; LANGER, 2005). Nesta etapa, todas as possíveis regiões de interações entre a protease e o inibidor foram reconhecidas utilizando os valores de distância e de ângulo padrão do programa *LigandScout* (WOLBER; LANGER, 2005). Em seguida, para cada um dos complexos, um modelo farmacofórico foi construído analisando-se cada uma das regiões de interação identificadas e selecionando-se aquelas que deveriam ser mantidas. O número e a natureza da região de interação variam de acordo com o modelo farmacofórico, incluindo regiões: aceptoras de ligação de hidrogênio; doadoras de ligação de hidrogênio, de interação iônica positiva, de interação hidrofóbica e de interação aromática. Cada uma das regiões de interação foi definida como obrigatória ou opcional. Ainda, regiões de exclusão de volume (construídas com base nas posições dos resíduos do sítio ativos) foram incluídas em todos os modelos farmacofóricos. Os raios de tolerância das regiões de interação variam tanto de acordo com o tipo de interação bem como de acordo com modelo farmacofórico, variando de 0,9 a 1,95 Å.

Por fim, os modelos farmacofóricos gerados foram aplicados a um dos bancos de dados de moléculas virtuais previamente construídos (*ZINC-Original*, *ZINC-Curated* e *NCI*) utilizando programa *LigandScout* (WOLBER; LANGER, 2005). Durante esta etapa, o programa selecionou compostos capazes de satisfazer as características mínimas requeridas (ou seja, as regiões de interação obrigatórias) pelo modelo e, calculou para cada composto selecionado um valor de *Pharmacophore Fit* (medida da qualidade do ajuste de cada

composto selecionado ao modelo farmacofórico). Para alguns modelos, estes valores de *Pharmacophore Fit* foram utilizados, posteriormente, no filtro de similaridade. A seguir, os compostos capazes de satisfazer todas as características mínimas requeridas pelo modelo foram selecionados para próxima etapa, independente do seu valor de *Pharmacophore Fit*.

3.5.4.3. Filtro “*drug-like*”

O filtro “*drug-like*” foi baseado na “Regra dos 5” de Lipinski (LIPINSKI et al., 1997). Os compostos que não violaram nenhuma das regras sugeridas por Lipinski (descritas em detalhes na Introdução) foram selecionados. Neste filtro, as/os propriedades/parâmetros físico-químicos usados por Lipinski foram calculados pelo programa *LigandScout* (WOLBER; LANGER, 2005). Ainda, compostos derivados de monossacarídeos (hexoses e pentoses) foram retirados usando o programa *Open Babel* (O’BOYLE et al., 2011). Os compostos restantes foram selecionados para etapa seguinte.

3.5.4.4. Filtros de Similaridade

Filtros de similaridade foram feitos agrupando os compostos previamente selecionados utilizando tanto *fingerprints* moleculares bem como descritores moleculares. Inicialmente, medidas de similaridade (**MS**), par a par, foram calculadas a partir dos *fingerprints* moleculares e dos correspondentes descritores moleculares e convertidas nas correspondentes medidas de dissimilaridades (Figura 3.4). Em seguida, cada medida de dissimilaridade (variando de 0 a 1) foi, separadamente, utilizada para agrupar os compostos previamente selecionados. As etapas de agrupamento foram realizadas pelo programa *R* (R CORE TEAM, 2013), usando algoritmo de agrupamento hierárquico com ligação completa (do inglês, “*complete linkage*”) e a função *hclust*. Em seguida, um valor de corte de dissimilaridade 0,2 foi aplicado (para mais detalhes vide Anexo III). Os compostos que apresentaram mais de 80% de similaridade entre si foram colocados em um mesmo grupo. Por

fim, para cada grupo gerado, o composto apresentando melhor ajuste ao modelo farmacofórico, maior valor de *Pharmacophore Fit* calculado pelo *LigandScout* (WOLBER; LANGER, 2005), foi considerado como representante do respectivo grupo e selecionado para próxima etapa.

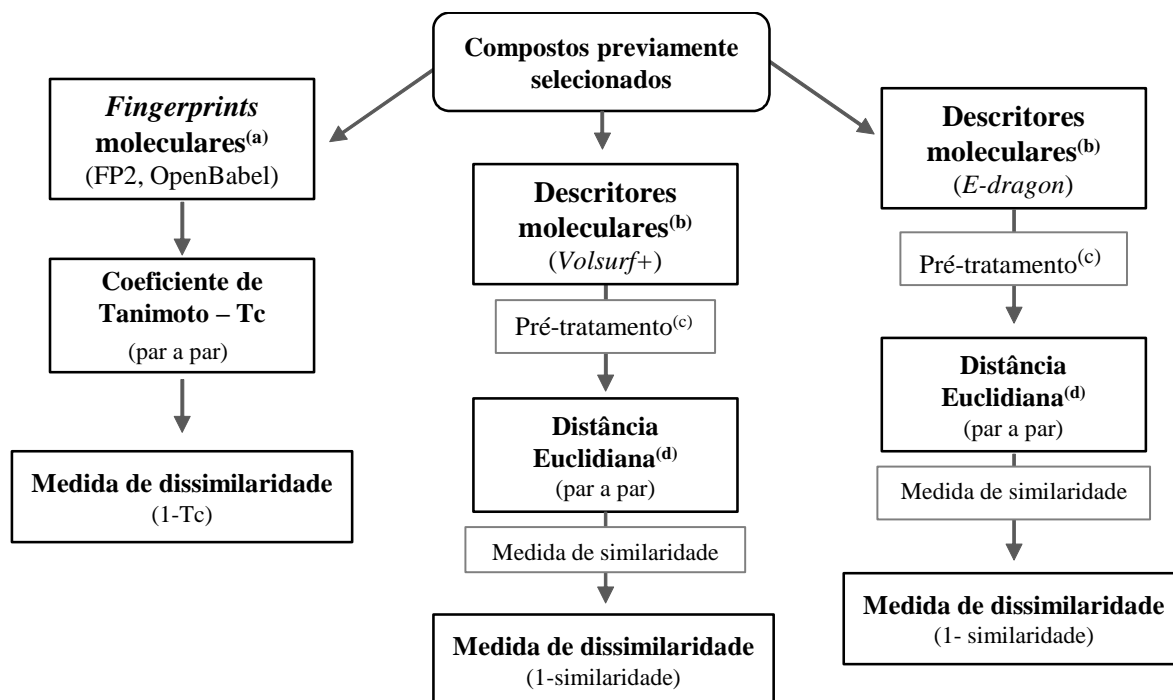


Figura 3.4. Etapas envolvidas no cálculo das medidas de dissimilaridades a partir dos *fingerprints* moleculares^(a) e dos descritores moleculares^(b).

^(a) Os *fingerprints* moleculares, do tipo FP2, foram obtidos como descrito no item 3.5.3.5.

^(b) Os descritores moleculares foram calculados usando *E-dragon* (TETKO et al., 2005) e *Volsurf+* (CRUCIANI et al., 2000). Para o programa *E-dragon* (TETKO et al., 2005), nenhum parâmetro deve que ser definido, já para o programa *Volsurf+* (CRUCIANI et al., 2000), foram usados os mesmo parâmetros descritos no item 3.5.3.1.

^(c) Remoção dos descritores constantes e auto escalonamento e centralização dos descritos não constante (utilizando as funções *center* e *scale* do programa *R* (R CORE TEAM, 2013).

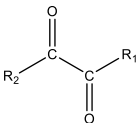
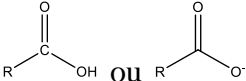
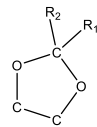
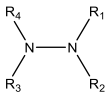
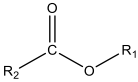
^(d) As distâncias Euclidianas (Anexo III), par a par, foram calculadas usando os descritores moleculares, submetidos a um pré-tratamento^(c) e convertidas em uma medida de similaridade (Anexo III).

3.5.4.5. Filtro de “grupos funcionais”

Filtro de “grupos funcionais”, proposto nesta tese, foi feito em duas etapas. Na primeira etapa, os compostos contendo os seguintes grupos funcionais foram removidos: compostos dicarbonílicos; ácidos carboxílicos; ionizáveis apresentando carga (s) negativa (s) em pH 7,4 (calculadas durante a criação do banco *ZINC-Curated*); grupos protetores (ex. acetal); grupos contendo N_{sp3}-N_{sp3} e ésteres. Para tanto, os grupos funcionais foram descritos

no formato *SMART* (DAYLIGHT INC., 2011) e utilizados numa busca por subestruturas com o auxílio do programa *Open Babel* (O'BOYLE et al., 2011). A representação estrutural e os correspondentes formatos *SMART* utilizados na busca por subestrutura estão apresentados na Tabela 2.2. Os compostos eliminados nesta etapa foram analisados por inspeção visual para confirmar se eles realmente continham um, ou mais, dos grupos funcionais descritos acima.

Tabela 2.4. Grupos funcionais considerados, representação estrutural e correspondente formato *SMART* utilizados como filtro de “grupos funcionais”.

Grupos funcionais	Representação estrutural	Formato <i>SMART</i>
Compostos dicarbonílicos		[CX3](=[OX1])[CX3](=[OX1])
Ácidos carboxílicos		[CX3](=O)[OX2H1,O-]
Ionizáveis com cargas negativas (em pH 7,4)	-	[-1]
Protetores (ex. acetal)		[OX2][CX2][OX2]
Contendo N _{sp3} -N _{sp3}		[N](-N)
Ésteres		[CX3](=O)[OX2][C]

Em uma segunda etapa, para cada um dos compostos mantidos, a espécie ionizável mais abundante (em pH 9,0) foi gerada utilizando o programa *Moka* (MILLETTI et al., 2007). E, aqueles que, ainda, apresentaram espécies com carga negativa foram eliminados.

3.5.4.6. Ancoramento utilizando o programa *GOLD*

Os procedimentos de ancoramento para um número pequeno de compostos (até 10.000 compostos) foram feitos utilizando o programa *GOLD* (JONES et al., 1997; JONES; WILLETT; GLEN, 1995). A estrutura da NS2B/NS3pro do DENV usada foi preparada como

descrito no item 3.5.4.1. E, as estruturas dos compostos a serem ancorados foram preparadas, no formato *SDF*, utilizando o programa *Sybyl/Sybyl-X* (SYBYL, 2009, 2013) considerando-se todos os hidrogênios (polares e apolares), mas sem considerar os valores de carga parcial atômica. Os átomos constituintes do sítio ativo da proteína foram definidos considerando-se todos os átomos em um raio de 6 ou 10 Å (LIEBESCHUETZ; COLE; KORB, 2012), a partir do inibidor Bz-nKRR-H posicionado na cavidade como encontrado no complexo 3U1I. Os parâmetros do algoritmo genético foram ajustados, automaticamente, pelo programa e de acordo com o tamanho do composto, para se obter o máximo de eficiência do ancoramento. A função de *score ChemPLP* (LIEBESCHUETZ; COLE; KORB, 2012) foi usada para determinar o *Fitness*. A seguir, o composto foi ancorado repetidas vezes sobre a estrutura da proteína (10, 20 ou 50 vezes). Ainda, para alguns procedimentos de ancoramento foram introduzidas algumas mudanças na função de ranqueamento (inclusão de *Protein HBond Constraints*) do programa e/ou foi “feita” uma ligação covalente entre o composto e o oxigênio da *Ser135B*. Estas mudanças foram feitas visando reproduzir a pose observada experimentalmente. Para isto, diferentes átomos da NS2B/NS3pro foram indicados como favoráveis para formar ligações de hidrogênio (aceptoras e/ou doadoras) com os compostos a serem ancorados. E, o valor de *Fitness* de cada pose foi penalizado em 10 pontos para cada ligação de hidrogênio que não foi satisfeita. A análise dos resultados do ancoramento bem como os critérios de seleção dos compostos a serem submetidos para as próximas etapas variaram de acordo com o modelo de busca virtual.

3.5.4.7. Inspeção visual

A inspeção visual de cada composto previamente selecionado foi feita verificando seus ajustes no sítio ativo da NS2B/NS3pro. Ainda, para os compostos previamente submetidos a um procedimento de ancoramento usando o programa *GOLD* (JONES; WILLETT; GLEN, 1995), foi verificado se as poses obtidas se mostravam reproduzíveis. Esta

inspeção foi feita com o auxílio dos programas *PyMOL* (SCHRÖDINGER, 2013) e *LigandScout* (WOLBER; LANGER, 2005). Os critérios utilizados na inspeção visual variaram de acordo com o modelo de busca virtual. Os compostos que satisfizeram todos, ou quase todos, os critérios definidos, descritos nos itens 3.5.4.9 a 3.5.4.12, para cada modelo de busca virtual foram selecionados para próxima etapa ou para serem adquiridos e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV usando protocolos experimentais da literatura (GOUVEA et al., 2007; NOGUEIRA SANTOS et al., 2012; WU et al., 2015), descritos em detalhes no item 3.6.1.

3.5.4.8. Estudos da flexibilidade conformacional da NS2B/NS3pro do DENV por simulações de Dinâmica Molecular (DM)

A partir do segundo modelo de busca virtual, com sucesso, foram feitas tentativas para se incluir a flexibilidade conformacional da NS2B/NS3pro do DENV (considerando-se, em especial, os resíduos do sítio ativo e da região NS2B). Como apresentado e discutido na Introdução, simulações de Dinâmica Molecular (DM) podem ser utilizadas como ferramenta para estudar a flexibilidade conformacional de uma proteína. Assim, simulações de DM foram feitas para explorar a flexibilidade conformacional dos resíduos do sítio ativo, em especial os da região NS2B. Para tanto, foram utilizadas três diferentes estruturas iniciais, a saber: a primeira corresponde à estrutura 3U1I inalterada, contendo a *Ser135B* covalentemente ligada ao inibidor Bz-nKRR-H (denominada de “estrutura *covalente*”); a segunda foi obtida removendo completamente o inibidor Bz-nKRR-H (denominada de “estrutura *livre*”) e a terceira foi obtida mantendo o inibidor Bz-nKRR-H no sítio ativo, mas clivando a ligação covalente entre este e a *Ser135B* (denominada de “estrutura *não-covalente*”). Parte do material descrito nos próximos parágrafos faz parte de artigo recentemente publicado no *JCAMD*, do qual sou primeira autora (PICCIRILLO et al., 2016), colocado como anexo.

Preparação da estrutura da NS2B/NS3pro do DENV para as simulações de DM

De modo análogo ao descrito no item 3.5.4.1, a estrutura 3D da NS2B/NS3pro do DENV-3 com o inibidor Bz-nKRR-H foi obtida diretamente do PBD. No entanto, para as simulações de DM fez-se necessário adicionar cinco resíduos ausentes/faltantes na estrutura cristalográfica. Estes cinco resíduos foram adicionados usando o programa *Modeller* (ESWAR et al., 2006) e suas geometrias foram refinadas mantendo-se a proteína rígida, exceto na região destes cinco resíduos. A estrutura da NS2B/NS3pro obtida foi utilizada nas três simulações de dinâmica molecular. As moléculas de água do cristal que apresentaram valores de *B-factor* $\leq 50,08$ foram mantidas nas simulações. Como descrito no item 3.5.4.1, os resíduos básicos e o resíduo *N*-terminal foram deixados protonados, os resíduos de ácidos e o resíduo *C*-terminal foram deixados desprotonados e, os resíduos de histidina foram mantidos na sua forma neutra. A *His51B* foi protonada no nitrogênio- δ de acordo com o mecanismo catalítico proposto para serino proteases, no qual o átomo NE-*His51B* atua como base recebendo o hidrogênio do grupo hidroxila da *Ser135B* (BABINE; BENDER, 1997). As *His47B* e *His60B* foram protonadas no nitrogênio- δ garantindo o maior número de ligações de hidrogênio com os resíduos vizinhos. A *His72A* foi protonada no nitrogênio- ϵ , pois não havia diferença no número de ligações de hidrogênio entre os estados de protonação δ e ϵ .

Simulações de Dinâmica Molecular

As três estruturas foram construídas usando o modulo *tleap* do programa *AMBER* (SALOMON-FERRER; CASE; WALKER, 2013) e assinalando os átomos de acordo com o campo de força ff99SB (SPASIC; SERAFINI; MATHEWS, 2013). Inicialmente, uma etapa de minimização (200 ciclos) foi feita utilizando o modulo *Sander* do programa *AMBER* (SALOMON-FERRER; CASE; WALKER, 2013) e aplicando o modelo geral de solvente implícito de Born (SRINIVASAN et al., 1999; STILL et al., 1990). A estrutura minimizada

foi neutralizada em uma caixa com moléculas de água TIP3P (JORGENSEN et al., 1983) com bordas a, pelo menos, 10 Å da superfície da proteína. Em seguida, o sistema foi equilibrado durante 1 ns em duas etapas. Durante os primeiros 500 ps, restrições harmônicas foram aplicadas a todos os átomos do sistema, exceto aos das moléculas de água, utilizando uma constante de força igual a 0,5 kcal/(mol Å²). Nesta etapa, o sistema foi aquecido de 100 K a 300 K em uma caixa de volume constante com a redução gradativa das restrições harmônicas. Durante os próximos 500 ps, todos os átomos do sistema puderam se mover livremente. O sistema foi tratado com condições periódicas de contorno utilizando interações eletrostáticas do tipo *PME* (do inglês, *particle mesh Ewald*) (DARDEN; YORK; PEDERSEN, 1993) e aplicando valor de corte de 12 Å para interações do tipo *van der Waals*. Em uma caixa com pressão constante, esta (1 atm) foi controlada através de um pistão de controle de pressão de Nosé Hoover Langevin e a temperatura foi mantida constante usando a dinâmica de Langevin. As três simulações usaram tempo de integração de 2 fs e tiveram trajetórias registradas a cada ps durante 200 ns. Ainda, nas simulações das estruturas *covalente* e *não-covalente*, foi necessária a parametrização do inibidor Bz-nKRR-H seguindo protocolo descrito na literatura (CIEPLAK et al., 1995). Os aminoácidos não-canônicos do inibidor (*Bz*, *n*, *Arg-(P₁)* e *Arg(P₁)-Ser135B*) foram protonados utilizando o programa *MOE* (CHEMICALCOMPUTINGGROUPINC., 2004). Em seguida, as porções *N*-terminal e/ou *C*-terminal destes aminoácidos foram bloqueadas adicionando-se grupos acetila e/ou *N*-metila (denominados de grupos bloqueadores), respectivamente (CIEPLAK et al., 1995). Cargas atômicas foram calculadas usando as ferramentas *R.E.D.* (DUPRADEAU et al., 2010) acopladas ao programa *Gaussian* (FRISCH et al.), usando nível HF/6-31G* com restrições Lagrange em cada fragmento para garantir a neutralidade dos grupos bloqueadores (CIEPLAK et al., 1995). Após a etapa de derivatização, os grupos bloqueadores foram retirados, resultando em resíduos não-canônicos livres. Parâmetros ausentes no campo de

força foram estimados com o módulo *parmchk* do *Amber* (WANG et al., 2006). Nas análises das simulações de DM, os resíduos do sítio ativo foram definidos como todos os resíduos da estrutura 3U1I que apresentaram pelo menos 3 átomos distantes 4 Å do inibidor Bz-nKRR-H.

Agrupamento hierárquico dos snapshots e abordagem GRID/PCA

Procedendo-se como descrito na literatura (SCHAEFER; KISKER; SOTRIFFER, 2011), a flexibilidade dos resíduos do sítio ativo foi analisada através do agrupamento hierárquico baseado nos valores de RMSD (do inglês, *Root-Mean-Square Deviation* – vide Anexo III). Inicialmente, as três trajetórias foram centralizadas utilizando o programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996), alinhando-se cada *frame* à estrutura inicial correspondente. A seguir, *snapshots* foram salvos a cada 100 ps, resultando em 2000 coordenadas diferentes para cada simulação. Os 6000 *snapshots* foram sobrepostos e utilizados para calcular valores de RMSD, de todos contra todos, usando o programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996), considerando apenas os resíduos do sítio ativo (todos os átomos exceto os de hidrogênio). Os valores de RMSD, de todos os *snapshots* contra todos os *snapshots*, foram utilizados para agrupar os 6000 *snapshots* usando o programa *R* (R CORE TEAM, 2013). A etapa de agrupamento dos 6000 *snapshots* foi realizada utilizando algoritmo de agrupamento hierárquico com ligação completa (função *hclust*). Neste agrupamento, o valor de corte foi igual a 3 Å. Assim, *snapshots* apresentando valor *RMSD* menor de 3 Å entre si foram colocados em um mesmo grupo. Para cada grupo, o *snapshot* central foi escolhido como estrutura representativa do correspondente grupo e utilizado nas análises subsequentes bem como na inspeção visual.

A seguir, a abordagem de GRID/PCA (PASTOR; CRUCIANI, 1995) foi utilizada para analisar a diferença estrutural entre cada uma das estruturas representativas bem como entre estas e a estrutura 3U1I. Procedendo-se de modo análogo ao descrito na literatura (PASTOR; CRUCIANI, 1995), todas as estruturas representativas foram sobrepostas à estrutura 3U1I e

colocadas em uma caixa com uma grade com espaço regulares de 0,5 Å. As dimensões da caixa foram escolhidas de modo a acomodar todos os resíduos do sítio ativo. Em seguida, os campos de interação molecular foram calculados pelo programa *GRID* (GOODFORD, 1985) utilizando cinco diferentes sondas moleculares (posicionadas nos vértices da grade) de modo a representar os seguintes grupos: hidrofóbico – sonda (DRY); com carga positiva – sonda (N1+); com carga negativa – sonda (O-); doador de ligação de hidrogênio – sonda (N1) e acceptor de ligação de hidrogênio – sonda (O). Os valores de energia calculados pelos campos de interação molecular foram submetidos a um pré-tratamento e, em seguida, analisados através de Análise de Componente Principal (ACP) usando o programa *R* (R CORE TEAM, 2013). No pré-tratamento, os valores de energia representando interações repulsivas (valores positivos) foram convertidos para zero (0 kcal/mol) e, a seguir, todos os valores de energia foram centralizados. A Análise de Componentes Principais foi feita utilizando o algoritmo *Nipals* (função *pca.nipals*) e gerando até 10 Componentes Principais (CP). O resultado foi analisado através do gráfico de escores das CP1, CP2 e CP3.

Identificação das interações moleculares conservadas ao longo das simulações

As interações entre o inibidor Bz-n-KRR-H e os resíduos do sítio ativo foram seguidas durante a simulação de DM das estruturas *covalente* e *não-covalente*. Procedendo-se de modo análogo ao descrito na literatura (MERGET; SOTRIFFER, 2015), as ligações de hidrogênio foram analisadas através de gráficos em barra dos valores de ocupância da ligação durante a simulação (ou seja, porcentagem do tempo de simulação em que se observa determinada interação). Os valores de ocupância foram calculados utilizando o programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996), aplicando os valores padrões de distância e de ângulo. Estes valores de ocupância foram utilizados para classificar as ligações de hidrogênio em estáveis (valor de ocupância ≥ 30 %), com média estabilidade (valor de

ocupância $>10\%$ e $<30\%$) e não estáveis (valor de ocupância $<10\%$). As interações iônicas foram analisadas através de gráficos de caixa dos valores de distância entre os grupos com carga positiva e com carga negativa. Para os resíduos canônicos, as interações iônicas foram identificadas e seus valores de distância medidos usando os parâmetros padrões da ferramenta *Salt-Bridge* disponível no programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996). Para os resíduos não-canônicos, as interações iônicas foram identificadas por inspeção visual e seus valores de distância medidos considerando o centro geométrico do grupo guanidino da *Arg-P_I* (ligada ou não a *Ser135B*) e o centro geométrico dos átomos de oxigênio carboxílico. De modo análogo, as interações cátion- π foram analisadas através de gráficos de caixa dos valores de distância entre sistema π e do grupo catiônico. Estas interações foram identificadas por inspeção visual e suas distâncias medidas utilizando o programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996) e considerando o centro geométrico do sistema π e do grupo catiônico. Todos os gráficos foram gerados utilizando o programa *R* (R CORE TEAM, 2013).

Identificação dos modos de ligação do inibidor Bz-nKRR-H

Procedendo-se como descrito na literatura (SCHAEFER; KISKER; SOTRIFFER, 2011), os diferentes modos de ligação do inibidor Bz-nKRR-H foram analisados através do agrupamento hierárquico baseado nos valores de *RMSD*. Inicialmente, as trajetórias das simulações *covalente* e *não-covalente* foram centralizadas e, em seguida, *snapshots* foram salvos a cada 100 ps, resultando em 2000 coordenadas diferentes para cada simulação. No entanto, diferentemente do descrito anteriormente, cada simulação de DM foi analisada separadamente, uma vez que o número de átomos do inibidor varia de acordo com estrutura usada (presença, ou ausência da ligação covalente com a *Ser135B*). Os 2000 *snapshots* foram sobrepostos e usados para calcular valores de *RMSD*, de todos contra todos, usando o programa *VMD* (HUMPHREY; DALKE; SCHULTEN, 1996) e considerando todos os

átomos do inibidor Bz-nKRR-H, exceto os de átomos hidrogênio. Os valores de *RMSD* foram usados para agrupar os 2000 *snapshots* usando o programa *R* (R CORE TEAM, 2013) e aplicando o algoritmo de agrupamento hierárquico com ligação completa (função *hclust*). Neste agrupamento, o valor de corte foi igual a 5 Å. Em seguida, apenas os grupos mais populosos, ou seja, com maior número de *snapshots* foram considerados. Para cada um destes grupos, o *snapshot* central foi escolhido como estrutura representativa do correspondente grupo e usado nas análises subsequentes e na inspeção visual.

3.5.4.9. Modelos de busca virtual: Modelo I - DENV

O primeiro modelo de busca virtual foi feito aplicando-se, em sequência, cinco filtros (farmacofórico, “*drug-like*”, similaridade, ancoramento e inspeção visual) ao banco de dados *ZINC-Original*, como apresentado na Figura 3. 5.

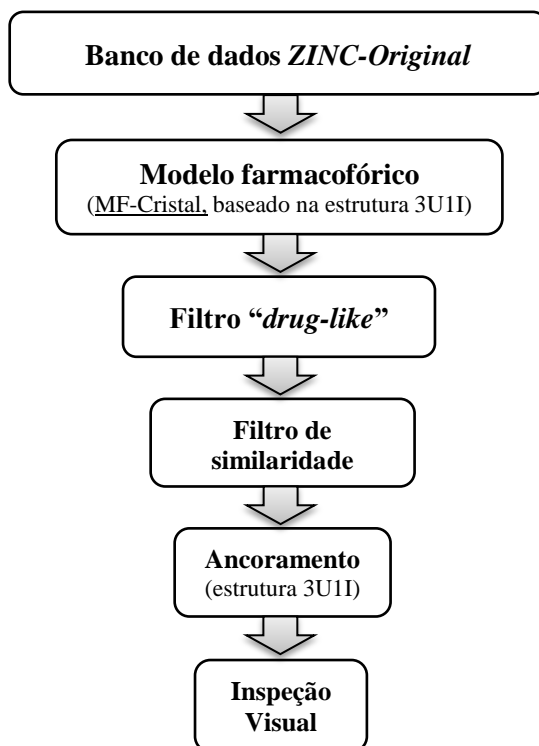


Figura 3. 5. Representação esquemática da construção do primeiro modelo de busca virtual: Modelo I – DENV. Este modelo foi feito aplicando-se, em sequência, cinco filtros (farmacofórico^(a), *drug-like*, similaridade, ancoramento^(b) e inspeção visual^(c)) ao banco *ZINC-Original* (preparado como descrito no item 3.5.1.1). Os filtros *drug-like* e de similaridade foram aplicados como descrito nos itens 3.5.4.3 e 3.5.4.4, respectivamente.

^(a) O modelo farmacofórico foi gerado, como descrito no item 3.5.4.2, a partir das regiões de interações entre a NS2B/NS3pro e o inibidor Bz-nKRR-H observadas na estrutura experimental 3U1I.

^(b) O ancoramento foi feito, como descrito no item 3.5.4.6, incluindo modificações na função de ranqueamento (descritas a seguir). Este procedimento foi validado usando o inibidor Bz-nKRR-H.

^(c) A inspeção visual foi feita como descrita no item 3.5.4.7, usando os critérios descritos a seguir.

Neste primeiro modelo, o modelo farmacofórico foi gerado, como descrito no item 3.5.4.2, utilizando a estrutura experimental 3U1I. Todas as regiões de interação entre a NS2B/NS3pro e o inibidor Bz-nKRR-H foram identificadas e aquelas próximas à cavidade S_I foram utilizadas para construir o modelo farmacofórico, denominado MF-Cristal. Este modelo farmacofórico contém regiões doadoras eceptoras de ligação de hidrogênio (com raios de tolerância de 1,5 Å) e regiões de exclusão de volume. Todas as regiões de interação do modelo foram definidas como obrigatórias. O modelo farmacofórico MF-Cristal foi aplicado, como descrito no item 3.5.4.2, ao banco de dados *ZINC-Original* (preparado como descrito no item 3.5.1.1) e os compostos capazes de satisfazer todas as características deste modelo foram submetidos ao filtro *drug-like* aplicado como descrito no item 3.5.4.3.

Os compostos que passaram pelo filtro *drug-like* foram submetidos ao filtro de similaridade aplicado como descrito no item 3.5.4.4. Neste filtro os compostos foram agrupados e apenas o composto representativo de cada grupo foi selecionado e, em seguida, ancorado na cavidade da NS2B/NS3pro do DENV.

O ancoramento foi feito, como descrito no item 3.5.4.6, utilizando a estrutura experimental 3U1I e o programa *GOLD* (JONES; WILLETT; GLEN, 1995). Este procedimento foi submetido a uma etapa de validação usando o inibidor Bz-nKRR. Inicialmente, a etapa de validação foi feita sem introduzir mudanças na função de ranqueamento. O inibidor Bz-nKRR foi, consecutivamente, ancorado 50 vezes sobre a estrutura da proteína. Para a análise dos resultados, apenas a estrutura da pose com o maior valor de *Fitness* foi considerada e o valor de *RMSD* entre esta pose e a pose do cristal foi calculado. Os critérios para a avaliação dos resultados do ancoramento foram: (i) o valor do *RMSD* calculado; (ii) a presença das interações proteína-inibidor observadas experimentalmente (estrutura 3U1I) e, (iii) a ocupação das cavidades S_1 a S_3 da proteína. Em seguida, foram introduzidas algumas mudanças na função de ranqueamento e foi “feita” uma ligação covalente entre o carbono do aldeído o oxigênio da *Ser135B*. Nesta etapa, os seguintes átomos dos resíduos de aminoácido da NS2B/NS3pro foram indicados como favoráveis para formar ligações de hidrogênio (aceptoras e/ou doadoras) com o inibidor: *OD1-Asp129B*, *O-Phe130B* e *O-Gly151B* e os hidrogênios ligados ao *N-Gly153B* e ao *OH-Try161B*. A avaliação dos resultados deste segundo procedimento de ancoramento, usando as mudanças na função de ranqueamento, foi feita como descrita anteriormente. Em seguida, os compostos previamente selecionados (filtro de similaridade) foram ancorados no sítio ativo da estrutura 3U1I, usando protocolo análogo ao descrito na etapa de validação, exceto pela “presença” da ligação covalente. Todos os compostos foram ancorados 10 vezes. As 10 poses geradas para cada composto foram avaliadas para verificar se estas satisfaziam, ou não, as interações

preditas pelo modelo MF-Cristal. Esta avaliação foi feita usando o descritor de ligação de hidrogênio disponível no programa *GOLD* (JONES et al., 1997; JONES; WILLETT; GLEN, 1995). Em seguida, apenas os compostos que tiveram pelo menos uma pose capaz de respeitar todas as características do modelo farmacofórico foram selecionados e submetidos à inspeção visual.

A inspeção visual foi feita como descrito no item 3.5.4.7 para verificar o ajuste dos compostos selecionados no sítio ativo bem como se suas poses se mostravam reproduzíveis. Esta inspeção foi feita seguindo os seguintes critérios: (i) verificar se as interações proteína-compostos correspondem às preditas pelo modelo farmacofórico (MF-Cristal); (ii) analisar a formação de interações com outros resíduos da cavidade da proteína, priorizando interações com os resíduos *His51B*, *Asp75B*, *Met84A*, *Thr83A* e *Ser135B* e, (iii) verificar se os critérios acima eram satisfeitos em pelo menos três das dez poses geradas. Os compostos capazes de satisfazer todos, ou quase todos, os critérios anteriormente foram selecionados, adquiridos e submetidos a ensaios de inibição enzimática frente a NS2B/NS3pro utilizando protocolos experimentais descritos em detalhes no item 3.6.1.

3.5.4.10. Modelos de busca virtual: Modelo II - DENV

O segundo modelo de busca virtual foi feito aplicando-se, em sequência, três filtros (ancoramento, “grupos funcionais” e inspeção visual) ao banco *ZINC-Curated*, como representado na Figura 3. 6.

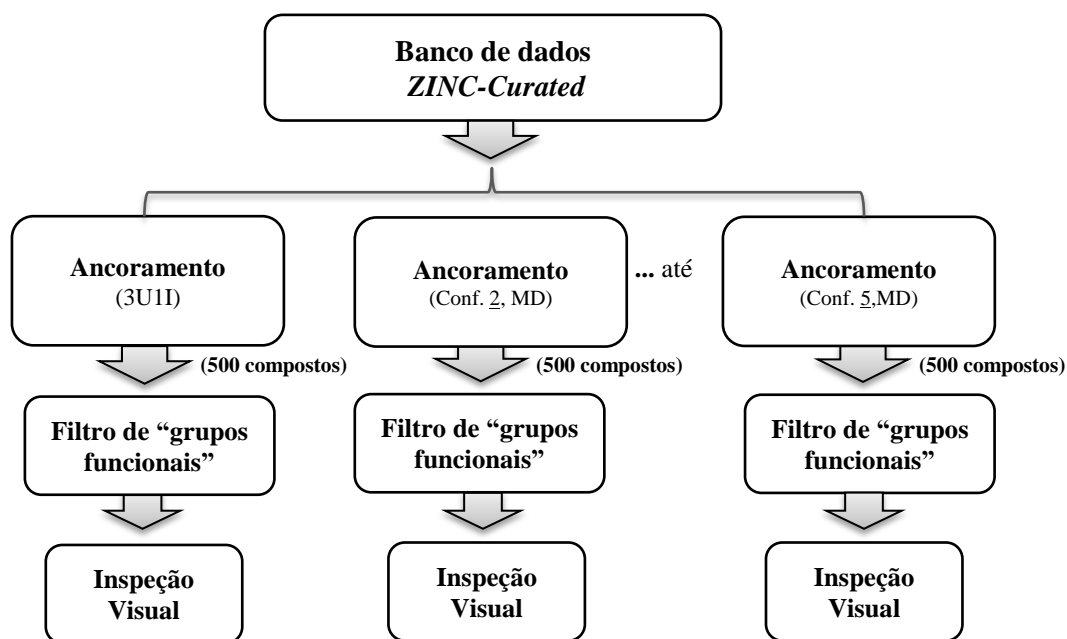


Figura 3. 6. Representação esquemática da construção do segundo modelo de busca virtual: Modelo II – DENV. O Modelo II – DENV foi feito aplicando-se, em sequência, três filtros (ancoramento^(a), “grupos funcionais” e inspeção visual^(b)) ao banco de dados *ZINC-Curated* (preparado como descrito no item 3.5.1.1). O filtro de “grupos funcionais” foi aplicado como descrito no item 3.5.4.5.

^(a) Os quatro ancoramentos foram feitos, separadamente, utilizando o programa *FRED* (MCGANN, 2011), como descrito a seguir. Neste ancoramento foram usadas, separadamente, a estrutura experimental (3U1I) e três conformações (denominadas de Conf. 2, 4 e 5) obtidas nas simulações de DM (como descrito no item 3.5.4.8). Para cada procedimento os 500 compostos com melhor ajuste à cavidade da proteína foram selecionados.

^(b) A inspeção visual foi feita como descrita no item 3.5.4.7, usando os critérios apresentados a seguir.

Neste segundo modelo, todos os procedimentos de ancoramento usando todas as moléculas do banco *ZINC-Curated* foram feitos utilizando o programa *FRED* (MCGANN, 2011). As quatro estruturas (uma experimental - 3U1I e três obtidas nas simulações de DM - Conf. 2, 4 e 5) usadas foram preparadas como descrito no item 3.5.4.1. Para cada estrutura, a caixa contendo os resíduos do sítio ativo foi inicialmente definida a partir da *Ser135B*. A seguir, as caixas foram inspecionadas visualmente e ajustadas de modo acomodar todos os resíduos do sítio ativo. Nos quatro procedimentos de ancoramento foram utilizadas a resolução “*standard*” e a função de ranqueamento *Chemgauss4* (MCGANN, 2012). Para cada um dos procedimentos de ancoramento, a pose com melhor valor de escore de cada uma das moléculas do banco *ZINC-Curated* foi armazenada. Em seguida, os compostos foram

ordenados de acordo com seus valores de escore. Para cada um dos procedimentos de ancoramento, os 500 compostos com menor valor de escore (ou seja, com melhor encaixe no sítio ativo da proteína) foram selecionados e submetidos a um filtro de “grupos funcionais” aplicado exatamente como descrito no item 3.5.4.5. A seguir, os compostos que passaram no filtro de “grupos funcionais” foram submetidos à inspeção visual.

A inspeção visual foi feita como descrito no item 3.5.4.7, avaliando se os compostos selecionados apresentavam as seguintes interações: (i) ligações de hidrogênio com os resíduos *His51B*, *Asp75B*, *Asp81A*, *Asp129B* e *Gly151B* e; (ii) iônicas com os resíduos *Asp75B*, *Asp81A* e *Asp129B* e, (iii) hidrofóbicas com o resíduo *Val155B*. Ainda, avaliou-se a fração da área do composto que está exposta ao solvente. Os compostos capazes de satisfazer grande parte das interações descritas acima e com uma pequena fração de sua área exposta ao solvente foram selecionados, adquiridos e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV, usando protocolos experimentais descritos no item 3.6.1.

3.5.4.11. Modelos de busca virtual: Modelo III - DENV

O terceiro modelo de busca virtual foi feito aplicando-se, em sequência, três filtros (modelo farmacofórico, “grupos funcionais” e inspeção visual) ao banco de dados *NCI*, como representado na Figura 3. 7. Parte do material descrito nos próximos parágrafos faz parte de artigo recentemente publicado no *JCAMD*, do qual sou primeira autora (PICCIRILLO et al., 2016).

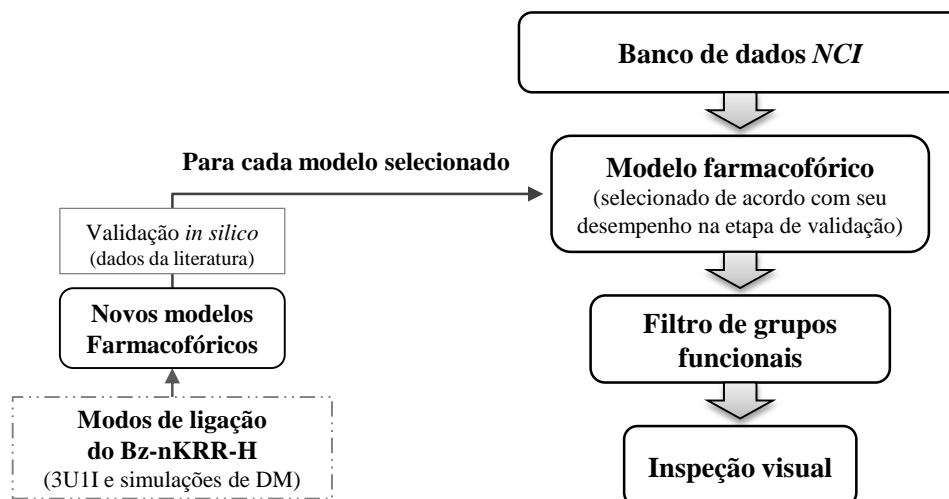


Figura 3. 7. Representação esquemática da construção do terceiro modelo de busca virtual: Modelo III – DENV. O Modelo III – DENV foi feito aplicando-se, em sequência, três filtros (farmacofórico^(a), “grupos funcionais” e inspeção visual^(b)) ao banco de dados *NCI* (preparado como descrito no item 3.5.1.1). O filtro de “grupos funcionais” foi aplicado como descrito no item 3.5.4.5.

^(a) Os outros modelos farmacofóricos foram gerados, como descrito no item 3.5.4.2, a partir das regiões de interações entre a NS2B/NS3pro e o inibidor Bz-nKRR-H observadas na estrutura experimental 3U1I bem como nas simulações de DM (vide item 3.5.4.8). Estes foram submetidos a uma etapa de validação *in silico* (descrita a seguir) e apenas os modelos que apresentaram um bom desempenho, de acordo com Piccirillo e colaboradores (PICCIRILLO et al., 2016), foram selecionados.

^(b) A inspeção visual foi feita como descrita no item 3.5.4.7 usando os critérios apresentados a seguir.

Neste terceiro modelo, os modos de ligação do inibidor Bz-nKRR-H observados nas estruturas representativas, obtidas pelo agrupamento hierárquico (vide item 3.5.4.8), foram utilizados na construção de novos modelos farmacofóricos (classificados, modelos farmacofóricos “Dinâmicos”). Ainda, o modo de ligação do inibidor Bz-nKRR-H observado na estrutura experimental 3U1I foi também usado para construir novos modelos farmacofóricos (classificados, modelos farmacofóricos “Estáticos”). Todos os modelos farmacofóricos foram gerados, como descrito no item 3.5.4.2, considerando todas as possíveis interações entre o inibidor e os resíduos do sítio ativo da NS2B/NS3pro do DENV. O número e o tipo de região de interação variam de acordo com o modelo farmacofórico, incluindo:ceptoras de ligação de hidrogênio; doadoras de ligação de hidrogênio e de interação iônica positiva. Para os modelos farmacofóricos Dinâmicos ainda foram incluídas regiões de interação hidrofóbica. Em seguida, todos os modelos farmacofóricos tiveram duas regiões definidas como obrigatórias, enquanto que as demais regiões foram definidas como opcionais.

Estas duas regiões obrigatórias correspondem a regiões aceptoras e/ou doadoras de ligação de hidrogênio que interagem, necessariamente, com resíduos do sítio ativo que formaram ligações de hidrogênios estáveis com o inibidor Bz-nKRR-H durante as simulações de DM (vide item 3.5.4.8). Ainda, regiões de exclusão de volume foram incluídas em todos os modelos farmacofóricos. O raio de tolerância das regiões de interação aceptoras ou doadoras de ligação de hidrogênio foi ajustado para 1,8 Å, enquanto que o raio das regiões hidrofóbicas e iônica positiva foi ajustado para 1,95 Å.

O desempenho dos modelos farmacofóricos gerados, Dinâmicos e Estáticos, foi avaliado *in silico* usando uma série de compostos peptídeos testados frente à NS2B/NS3pro do DENV retirados da literatura (vide item 3.3.3). Esta série foi preparada como descrito no item 3.5.2.1. Desta série, 28 compostos são ativos, incluindo o próprio inibidor Bz-nKRR-H, e 3 são inativos frente à esta protease. A seguir, todos os modelos foram aplicados à série de peptídeos miméticos, como descrito no item 3.5.4.2, e apenas os compostos capazes de respeitar seis ou mais regiões de interação (duas obrigatórias e quatro opcionais) foram considerados como *hits*. O desempenho de cada um dos novos modelos farmacofóricos, Dinâmicos e Estáticos, foi avaliado verificando a capacidade destes modelos selecionarem como *hits* compostos ativos bem como diferenciarem entres os compostos ativos e inativos. Por fim, apenas os modelos farmacofóricos que apresentaram um desempenho satisfatório na etapa anterior foram selecionados e aplicados ao banco *NCI* como descrito no item 3.5.4.2. A seguir, de modo análogo ao descrito na etapa de avaliação dos modelos, apenas os compostos capazes de respeitar seis ou mais regiões de interação foram considerados como *hits* e selecionados para serem submetidos ao filtro de “grupos funcionais”.

O filtro de “grupos funcionais” foi aplicado como descrito no item 3.5.4.5. A seguir, os compostos que passaram por este filtro foram submetidos à inspeção visual. A inspeção visual foi realizada considerando apenas os compostos selecionados pelo filtro de “grupos

funcionais” que estavam disponíveis para aquisição imediata. Para isto, foi feita uma busca no website do NCI (endereço eletrônico <https://dtp.cancer.gov/dtpstandard/dwindex/index.jsp>) restringindo-se os resultados apenas aos compostos em estoque, em quantidade suficiente (> 5 mg) para a etapa de validação experimental. Em seguida, na inspeção visual (vide item 3.5.4.7) os compostos que apresentaram um melhor ajuste ao modelo farmacofórico com estruturas químicas diversas foram selecionados, adquiridos e testados frente à NS2B/NS3pro do DENV, utilizando protocolos experimentais descritos em detalhes no item 3.6.1.

3.5.4.12. Modelos de busca virtual: Modelo IV - DENV

O quarto modelo de busca virtual foi feito aplicando-se, em sequência, dois filtros (farmacofórico e inspeção visual) ao banco *ZINC-Curated*, como representado na Figura 3. 8.

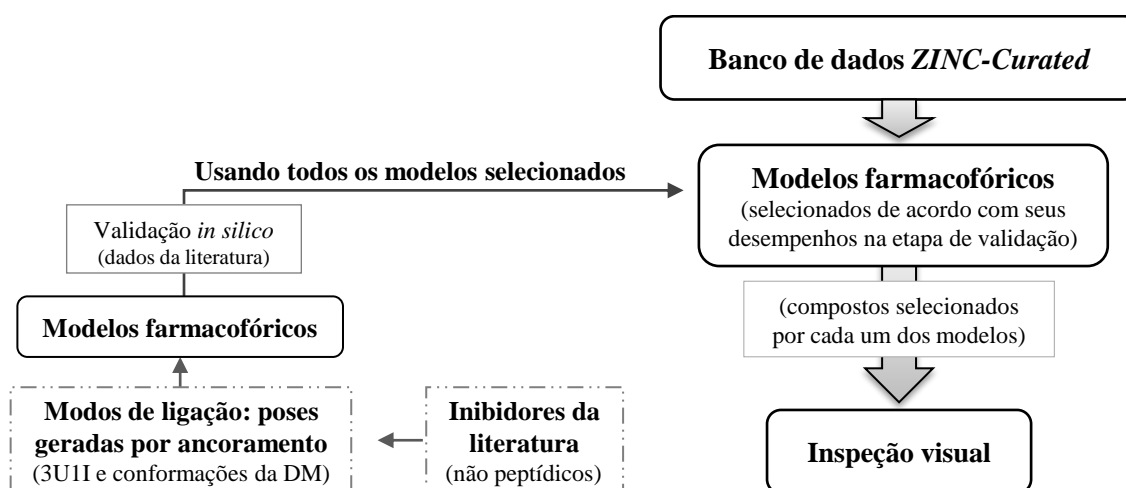


Figura 3. 8. Representação esquemática da construção do quarto modelo de busca virtual: Modelo IV – DENV. O Modelo IV – DENV foi feito aplicando-se, em sequência, dois filtros (farmacofórico^(a) e inspeção visual^(b)) ao banco de dados *ZINC-Curated* (preparado como descrito no item 3.5.1.1).

^(a) Os modelos farmacofóricos foram gerados, como descrito no item 3.5.4.2, a partir das regiões de interações entre a NS2B/NS3pro e cada um dos inibidores da literatura^(c). Estes foram submetidos a uma etapa de validação *in silico* (descrita a seguir) e apenas os modelos que apresentaram um bom desempenho foram selecionados.

^(b) A inspeção visual foi feita como descrita no item 3.5.4.7 usando os critérios apresentados a seguir.

^(c) Os modos de ligação de cada um dos inibidores correspondem a poses geradas por ancoramento. Os procedimentos de ancoramento foram feitos, como descrito no item 3.5.4.6, utilizando mudanças na função de ranqueamento descritas a seguir.

Neste quarto modelo, inibidores da literatura (BODENREIDER et al., 2009; CREGAR-HERNANDEZ et al., 2012; DENG et al., 2012; GANESH et al., 2005; KOH-

STENTA et al., 2015; LAI et al., 2013; LAI; SRIDHAR PRASAD; PADMANABHAN, 2013; LIU et al., 2014; MUELLER et al., 2008; NGUYEN et al., 2013; NITSCHKE; STEUER; KLEIN, 2011; NOBLE et al., 2012; SHIRYAEV et al., 2011; TIEW et al., 2012) com valores de K_i/IC_{50} frente à NS2B/NS3pro do DENV menores que 10 μM foram selecionados para serem usados como referência para criação de outros modelos farmacofóricos.

Para tanto, fez-se necessário obter o modelo de ligação de cada um dos ligantes no sítio ativo da NS2B/NS3pro do DENV através de procedimentos de ancoramento. Os ancoramento foram feitos como descrito no item 3.5.4.6. Neste procedimento, além da conformação da NS2B/NS3pro do DENV observada no cristal 3U1I foram usadas as conformações obtidas pelas simulações de DM. Todas as estruturas da NS2B/NS3pro foram preparadas como descrito no item 3.5.4.1. As estruturas 3D dos inibidores da literatura foram obtidas usando o programa *Sybyl-X* (SYBYL, 2013) e, em seguida, minimizadas (100 ciclos) utilizando as opções padrão do programa *Sybyl-X* (SYBYL, 2013). Nos procedimentos de ancoramento foram introduzidas algumas mudanças na função de ranqueamento. Nesta etapa, os seguintes átomos dos resíduos de aminoácido da NS2B/NS3pro foram indicados como favoráveis para formar ligações de hidrogênio (aceptoras e/ou doadoras) com o inibidor: O-*Asp81A*, OD2-*Asp75B*, OD1-*Asp129B*, OD2-*Asp129B*, HG-*Ser135B*, O-*Gly151B* e HH-*Tyr161B*. Adicionalmente, para os inibidores descritos como covalentes foi “feita” uma ligação covalente entre o centro eletrofílico e o resíduo da *Ser135B*. Todos os inibidores foram, consecutivamente, ancorados 10 vezes sobre a estrutura da enzima. Em seguida, a escolha da pose de melhor ajuste foi feita considerando as interações com os resíduos do sítio ativo bem como a reprodutibilidade das poses com maior valor de *fitness*.

A seguir, para cada um dos inibidores as poses com melhor ajuste foram utilizadas para gerar modelos farmacofóricos. Todos os modelos farmacofóricos foram gerados, como descrito no item 3.5.4.2, considerando todas as possíveis interações entre o inibidor e os

resíduos do sítio ativo da NS2B/NS3pro do DENV. O número e o tipo de região de interação variam de acordo com o modelo farmacofórico, incluindo: aceptoras de ligação de hidrogênio; doadoras de ligação de hidrogênio, de interação iônica positiva, de interação hidrofóbica e de interação aromática. Todas as regiões de interação foram definidas como obrigatórias. Ainda, regiões de exclusão de volume foram incluídas em todos os modelos farmacofóricos. O raio de tolerância das regiões de interação foi ajustado automaticamente pelo programa *LigandScout* (WOLBER; LANGER, 2005).

O desempenho dos modelos farmacofóricos gerados foi avaliado *in silico* utilizados a série de compostos não peptídicos já avaliados experimentalmente quanto a sua atividade biológica frente à NS2B/NS3pro do DENV (vide item 3.3.3) preparada como descrito no item 3.5.2.1. Desta série, 95 compostos são ativos (incluindo os compostos usados para gerar os modelos), 385 são pouco ativos, e 293 são inativos frente a esta protease. Esta validação foi feita usando as extensões *LigandScout* (WOLBER; LANGER, 2005) para o programa *KNIME* (BERTHOLD et al., 2007). Nesta etapa, o número de *hits*, compostos ativos e *decoys* selecionados foram calculados e utilizados para analisar cada um dos modelos gerados. Em seguida, os modelos que selecionaram apenas os compostos classificados como ativos foram selecionados. Por fim, todos os modelos farmacofóricos selecionados na etapa anterior foram aplicados, em sequência, ao banco *ZINC-Curated* como descrito no item 3.5.4.2. A seguir, os compostos selecionados por cada um dos modelos foram submetidos à inspeção visual.

A inspeção visual foi feita como descrito no item 3.5.4.7 usando os seguintes critérios: (i) compostos com bom ajuste ao modelo farmacofórico utilizado para selecioná-los; (ii) compostos com centro quirais para os quais todos os possíveis isômeros ópticos foram selecionados; (iii) compostos com pouca flexibilidade (ex. cadeias carbônicas insaturadas de até 3 átomos); (iv) compostos sem conformações energeticamente desfavoráveis; (v) compostos que não apresentaram grupos $-SO_2$ em regiões doadores de ligação de hidrogênio;

(vi) compostos com poucos átomos polares em regiões hidrofóbicas e/ou expostos ao solvente e (vii) compostos sem grupamentos ésteres susceptíveis a hidrólise. Os compostos que respeitaram todos os critérios descritos acima foram selecionados. Alguns destes compostos foram adquiridos e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV, usando protocolos experimentais descritos no item 3.6.1.

3.5.4.13. Construção e análise de série de análogos ao composto mais ativo dentre os selecionados pelo Modelo IV – DENV

A estrutura do composto que apresentou maior atividade, dentre os selecionados, frente à NS2B/NS3pro, foi utilizado como referência para se encontrar compostos estruturalmente análogos. Esta etapa foi feita utilizando a ferramenta de busca por similaridade do website *Pubchem* (endereço eletrônico <https://pubchem.ncbi.nlm.nih.gov/#>) e aplicando-se um valor de corte de similaridade de 80%. Os compostos análogos selecionados tiveram suas estruturas tridimensionais geradas usando o programa *Corina* (SADOWSKI; GASTEIGER, 1993). Em seguida, para cada um dos compostos análogos, conformações foram geradas utilizando o programa *Omega* (HAWKINS et al., 2010). Esta série de compostos foi submetida, de modo análogo, a sequência de filtros utilizada para selecionar o composto de referência. E, aqueles que passaram por todos os filtros e que estavam disponíveis para aquisição imediata foram comprados e submetidos, de modo análogo, a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV, usando protocolos experimentais descritos no item 3.6.1.

3.5.5. Busca virtual de inibidores da protease do vírus da Febre Aftosa

Nesta tese, foram construídos e utilizados dois diferentes modelos de buscas de inibidores da protease líder do vírus da Febre Aftosa (denominados Modelo I – FA e Modelo II – FA). Como descrito na Introdução e de modo análogo ao feito para NS2B/NS3pro do

DENV, estes modelos de busca virtual foram construídos aplicando-se, em sequência, diferentes filtros de seleção. Inicialmente, será apresentada uma descrição geral dos procedimentos computacionais usados na construção destes modelos. Em seguida, será feita uma descrição detalha destes procedimentos para cada um dos dois modelos de busca virtual, incluindo os tipos e a ordem que estes filtros foram aplicados.

3.5.5.1. Preparação da estrutura da Lb^{pro} do vírus da FA sem ligantes.

A estrutura 3D da sLb^{pro} do vírus da FA (código 2JQG, C51A obtida por RMN) foi preparada usando o programa *Sybyl* (SYBYL, 2009). Esta estrutura corresponde ao mutante Cys51Ala. A estrutura 3D da Lb^{pro} foi obtida por RMN estando disponíveis no arquivo *PDB* cinco modelos de estrutura, numerados, respectivamente de 1 a 5, sendo utilizado o modelo número 1 como sugerido na literatura (CENCIC et al., 2007). Inicialmente, sabendo-se que a *Cys51* é essencial no mecanismo de catálise, fez-se necessário substituir a *Ala51* pela *Cys51* e realizar uma etapa de minimização local dos resíduos vizinhos à *Cys51*, utilizando o campo de força Tripos do programa *Sybyl* (SYBYL, 2009). Em seguida, todas as moléculas de água foram retiradas e todos os hidrogênios (polares e apolares) foram mantidos.

3.5.5.2. Preparação da estrutura da Lb^{pro} do vírus da FA complexada com o inibidor E64-R-P-NH₂

A cadeia A da estrutura 3D da Lb^{pro} do vírus da FA complexada com o inibidor E64-R-P-NH₂ (código 4QBB, resolução 1,60) foi preparada utilizando o programa *Sybyl-X* (SYBYL, 2013). Nesta etapa, todas as moléculas de água, os íons sulfato e o inibidor E64-R-P-NH₂ foram removidos. Em seguida, os átomos ausentes na cadeia lateral da *Gln146* foram adicionados e a posição do anel imidazólico da *His148* catalítica foi invertida.

3.5.5.3. Ancoramento utilizando o programa *GOLD*

Todos os procedimentos de ancoramento foram feitos de modo análogo ao descrito para busca virtual de inibidores da protease do vírus da Dengue (item 3.5.4.6). Nestes procedimentos, os átomos constituintes do sítio ativo da proteína foram definidos considerando-se todos os átomos em um raio de 20 Å a partir do carbono beta da *Cys51*. A função de escore *ChemPLP* (LIEBESCHUETZ; COLE; KORB, 2012) foi utilizada e não foram incluídas mudanças na função de ranqueamento. A seguir, todos os compostos foram, consecutivamente, ancorados 50 vezes sobre a estrutura da Lb^{pro}.

3.5.5.4. Cálculo dos campos de interação molecular e identificação de regiões de mínimo de energia

Inicialmente, os campos de interação molecular foram calculados pelo programa *GRID* (GOODFORD, 1985) utilizando a estrutura 4QBB da sLb^{pro} preparada como descrito no item 3.5.5.2. De modo análogo ao descrito no item 3.5.4.8, a proteína foi colocada em uma caixa com uma grade regularmente espaçada (0,25 Å). As dimensões da caixa foram escolhidas de modo a acomodar todos os resíduos das cavidades S_2 , S_I e S_I' do sítio ativo da sLb^{pro}. Em seguida, os campos de interação molecular foram calculados usando sete diferentes sondas moleculares (posicionadas nos vértices da grade), a saber: “DRY”, “C3”; “N1+”; “O-”; “N1”; “O” e “OH”. Assim, foram identificadas regiões favoráveis às interações de naturezas, respectivamente, hidrofóbica, iônica positiva e negativa, aceptoras de ligação de hidrogênio, doadoras de ligação de hidrogênio e doadoras e aceptoras de ligação de hidrogênio.

A seguir, com a finalidade de identificar os pontos de mínimo de energia de interação foi fixado, para cada sonda, um valor máximo de energia de corte, sendo estes - 1,0 kcal/mol; - 2,0 kcal/mol; - 7,0 kcal/mol, - 7,0 kcal/mol, - 6,0 kcal/mol, - 6,0 kcal/mol e - 6,0 kcal/mol, respectivamente, para sonda “DRY”; “C3”, “N1+”, “O-”, “N1”, “O” e “OH”. E, apenas

aqueles próximos aos resíduos do sítio ativo da sLb^{pro} foram selecionados. As coordenadas destes pontos de mínimo de energia foram convertidas em arquivo *PDB*, podendo ser visualizadas no programa *LigandScout* (WOLBER; LANGER, 2005).

3.5.5.5. Construção e aplicação do modelo farmacofórico

O modelo farmacofórico foi gerado usando a estrutura 4QBB previamente preparada (item 3.5.5.2) e as coordenadas dos pontos de mínimo de energia obtidos pelo programa *GRID* (GOODFORD, 1985). De modo análogo ao descrito para protease do vírus da Dengue (item 3.5.4.2), o modelo farmacofórico foi gerado utilizando o programa *LigandScout* (WOLBER; LANGER, 2005). Nesta etapa, regiões de interação entre os resíduos da proteína e os pontos de mínimo de energia foram identificadas utilizando os valores de distância e de ângulo padrões do programa *LigandScout* (WOLBER; LANGER, 2005). A seleção das regiões de interação a serem mantidas no modelo farmacofórico foi feita considerando as interações observadas para o inibidor E64-R-P-NH₂ descritas (estrutura 4QBB) bem como de dois inibidores previamente identificados, por nós, usando ancoramento descrito no item 3.5.5.3. Ainda, nesta etapa de seleção das regiões que estabeleceram interação(ões) com mais de um resíduo do sítio ativo foram priorizadas. O modelo farmacofórico gerado contém regiões de interação doadoras e aceptoras de ligação de hidrogênio, regiões de interação hidrofóbica e várias regiões de exclusão de volume. Ainda, as regiões de interações de ligação de hidrogênio (direcionais) com mais de três resíduos do sítio ativo foram definidas como não-direcionais. Em seguida, quatro regiões de interação foram definidas como opcionais. Todas as regiões de interação tem um raio de tolerância de 1,5 Å. Por fim, de modo análogo ao descrito para o DENV, o modelo farmacofórico gerado (denominado MF-Lb^{pro}) foi aplicado ao banco *ZINC-Curated* usando o programa *LigandScout* (WOLBER; LANGER, 2005). Como descrito no item 3.5.5.2, todos os compostos capazes de satisfazer as

características mínimas requeridas (ou seja, as regiões de interação obrigatórias) pelo modelo foram selecionados para próxima etapa, independente do seu valor de *Pharmacophore Fit*.

3.5.5.6. Filtro de similaridade

O filtro de similaridade foi feito de modo análogo ao apresentado no item 3.5.4.4. No entanto, para a busca virtual de inibidores da Lb^{pro} do vírus da FA foram utilizados apenas os valores do coeficiente de Tanimoto (**MS1**) calculados como descrito no item 3.5.4.4. Em seguida, foi aplicado um valor de corte de 0,40. Assim, compostos que apresentaram mais de 60% de similaridade entre si foram colocados em um mesmo grupo. E, como descrito no item 3.5.4.4, para cada grupo gerado, o composto com maior valor de *Pharmacophore Fit* foi considerado como representante do respectivo grupo.

3.5.5.7. Modelos de busca virtual: Modelo I – FA

Uma série de 52 compostos comerciais previamente selecionados, por busca virtual (MALVEZZI; DE REZENDE; AMARAL, 2009), como potenciais inibidores covalentes da cruzafina estava disponível em nosso laboratório e foi utilizada tanto para identificar inibidores da Lb^{pro} do vírus da FA bem como para ter informações sobre as interações dentro do sítio ativo. Assim, neste primeiro modelo de busca virtual foi utilizada esta série de 52 compostos aplicando-se filtro de ancoramento e inspeção visual, como apresentado na Figura 3. 9.

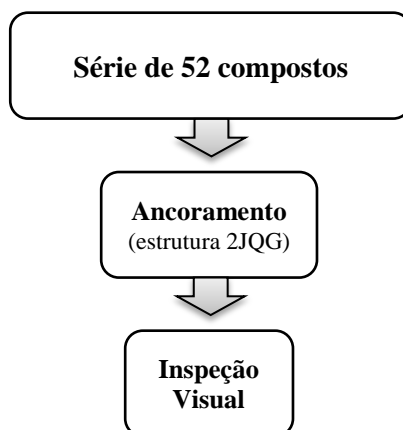


Figura 3. 9. Representação esquemática da construção do primeiro modelo de busca virtual: Modelo I – FA. O Modelo I – FA foi feito aplicando-se, em sequência, dois filtros (ancoramento^(a) e inspeção visual^(b)) a série de 52 compostos previamente selecionados como potenciais inibidores covalentes (preparada como descrito no item 3.5.2.2).

^(a) O ancoramento foi feito, como descrito no item 3.5.5.3, usando a estrutura sem inibidores (2JQG).

^(b) A inspeção visual foi feita como descrita no item 3.5.4.7 usando os critérios descritos a seguir.

O ancoramento foi feito como descrito no item 3.5.5.3, usando a estrutura sem inibidores (2JQG). Sabendo-se que estes compostos foram previamente selecionados como potenciais inibidores covalentes, apenas os compostos cujas poses mostraram o correspondente grupamento eletrofílico próximo à *Cys51* foram selecionados. Estes compostos foram submetidos a ensaios enzimáticos de inibição frente à sLb^{pro} do vírus da FA utilizando protocolos experimentais descritos nos itens 3.6.2.1 e 3.6.2.2.

3.5.5.8. Modelos de busca virtual: Modelo II - FA

Neste segundo modelo, a busca virtual de inibidores Lb^{pro} do vírus da FA foi realizada usando o banco *ZINC-Curated* aplicando-se sequencialmente três diferentes filtros de seleção (farmacofórico; similaridade e inspeção visual), como apresentado na Figura 3. 10.

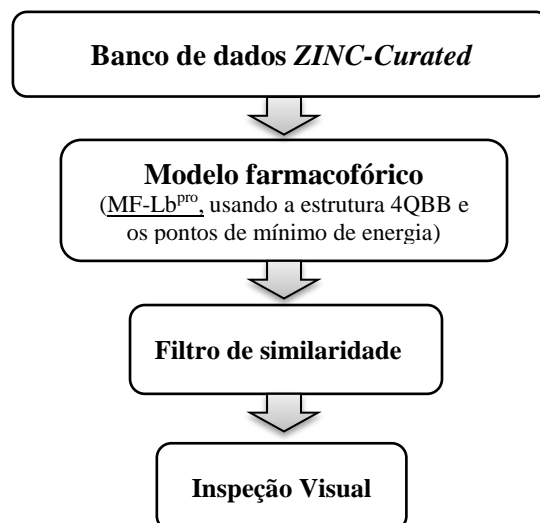


Figura 3. 10. Representação esquemática da construção do segundo modelo de busca virtual: Modelo II – FA. Este foi feito aplicando-se, em sequência, três filtros (farmacofórico^(a), similaridade e inspeção visual^(c)) ao banco *ZINC-Original* (preparado como descrito no item 3.5.1.1). O filtro de similaridade foi aplicado como descrito no item 3.5.5.6.

^(a) O modelo farmacofórico (MF-Lb^{pro}) foi gerado, como descrito no item 3.5.5.5, a partir das regiões de interações entre a sLb^{pro} (estrutura 4QBB) e os pontos de mínimo de energia^(c).

^(b) A inspeção visual foi feita como descrita no item 3.5.4.7 usando os critérios descritos a seguir.

^(c) Os pontos de mínimo de energia foram obtidos a partir dos campos de interação molecular calculados usando o programa *GRID* (GOODFORD, 1985) como descrito no item 3.5.5.4.

O modelo farmacofórico (MF-Lb^{pro}) foi gerado como descrito no item 3.5.5.5 e, em seguida, aplicado ao banco de *ZINC-Curated*. Os compostos capazes de satisfazer as características mínimas requeridas (ou seja, as regiões de interação obrigatórias) pelo modelo foram selecionados para próxima etapa, independente do seu valor de *Pharmacophore Fit*. A seguir, estes compostos foram submetidos a um filtro de similaridade como descrito no item 3.5.5.6. Neste filtro, os compostos foram agrupados e apenas o composto representativo de cada grupo foi analisado e selecionado para a próxima etapa.

Por fim, como descrito no item 3.5.4.7, o ajuste de cada composto ao modelo farmacofórico foi analisado por inspeção visual, utilizando os seguintes critérios: (i) compostos capazes de satisfazer as características do modelo farmacofórico, inclusive suas regiões de interação definidas como opcionais; (ii) compostos que não apresentassem grupamento aromático na região de interação hidrofóbica em S_2 e (iii) compostos apresentando interações com outros resíduos do sítio ativo. Ainda, avaliou-se a fração da área

do composto que está exposta ao solvente. Os compostos capazes de satisfazer todos os critérios descritos acima e com uma pequena fração de sua área exposta ao solvente foram selecionados, adquiridos e submetidos a ensaios de inibição enzimática frente à sLb^{pro} do vírus da FA utilizando protocolos experimentais descritos em detalhes no item 3.6.2.3.

3.6. PROCEDIMENTOS EXPERIMENTAIS

Conforme apresentado e discutido na Introdução, todos os modelos de busca virtual devem ser validados experimentalmente para confirmar sua verdadeira utilidade na seleção de compostos bioativos. Nesta tese, cada um dos modelos de busca virtual foi validado através de testes de inibição enzimática dos compostos selecionados pelos modelos frente à correspondente enzima (protease vírus da Dengue ou protease líder do vírus da Febre Aftosa).

De modo análogo ao apresentado para os procedimentos computacionais, os métodos utilizados nos testes de inibição da protease do vírus da Dengue (NS2B/NS3pro do DENV) e da protease líder do vírus da Febre Aftosa (Lb^{pro} do vírus da FA) serão apresentados sempre na sequência NS2B/NS3pro do DENV e Lb^{pro} do vírus da FA.

3.6.1. Testes de inibição da NS2B/NS3pro do DENV

Todos os compostos selecionados pelos quatro modelos de busca virtual (Modelo I – DENV a Modelo IV – DENV) como potenciais inibidores da NS2B/NS3pro do DENV foram submetidos a uma etapa de validação experimental, através de ensaios de inibição enzimáticas. Os testes de inibição da NS2B/NS3pro do DENV foram realizados nos laboratórios da Prof^a. Wibke Diederich da Philipps-Universität de Marburg, Alemanha.

Nesta etapa de validação, inicialmente, foram determinados os valores de porcentagem de inibição de cada composto (na concentração de 100 µM). Em seguida, os compostos que apresentaram alguma atividade inibitória tiveram os correspondentes valores de IC₅₀

(concentração do composto capaz de inibir 50% da atividade da enzima) determinados. Por fim, para alguns compostos mecanismo de inibição foi caracterizado e o correspondente valor da constante de inibição (K_i) determinado.

3.6.1.1. Determinação dos valores de porcentagem de inibição

Os compostos selecionados pelos quatro modelos de busca virtual (Modelo I – DENV a Modelo IV – DENV) como potenciais inibidores da NS2B/NS3pro do DENV foram avaliados através de ensaios enzimáticos. Inicialmente, estes compostos tiveram seus valores de valores de porcentagem de inibição frente a NS2B/NS3pro do DENV determinados (na concentração de 100 μ M).

Todos os ensaios de inibição foram realizados utilizando o substrato fluorogênico PhAc-KRR-MCA.3TFA (dissolvido em água), seguindo-se protocolo de rotina implementado no laboratório da Prof^a. Wibke Diederich da Philipps-Universität de Marburg, Marburg, Alemanha (WU et al., 2015). As medidas foram feitas utilizando placas de 96 poços (modelo *Nunc*TM da *Thermo Scientific*) contendo solução tampão de ensaio (pH = 9,0 [Tris-HCl (50 mM) com Chaps (1 mM)]), 20 ou 40 μ M do substrato (para DENV2 e DENV3, respectivamente) e 100 μ M do composto com volume final de 200 μ L. A reação foi iniciada adicionando protease do DENV2 (dissolvida em tampão pH = 7,5 [Tris-HCl (20 mM), NaCl (50 mM) e glicerol (5 % v/v)]) ou DENV3 (dissolvida em tampão pH = 8,5 [Tris-HCl (20 mM), NaCl (150 mM) e glicerol (5 % v/v)]) numa concentração final de 50 nM. A variação da intensidade de fluorescência foi seguida em um leitor de placa com agitação e temperatura controlada ($37,0 \pm 0,1^\circ\text{C}$), emitindo radiação em 380 nm e registrando-se fluorescência em 460 nm. As soluções estoque dos compostos foram preparadas em DMSO e logo utilizadas nos testes de inibição. A quantidade de DMSO nos experimentos foi mantida constante, sendo igual a 0,01 % (v/v). Todos os experimentos foram feitos, pelo menos, em triplicata.

O valor de velocidade inicial de formação do produto (v_0) para cada inibidor foi obtido do gráfico de fluorescência em função do tempo, em segundos. A seguir, os valores de porcentagem de inibição foram calculados comparando o valor de v_0 obtido para cada inibidor com o valor obtido em experimento controle, realizado de modo análogo ao descrito anteriormente, no entanto, sem adicionar nenhum composto.

3.6.1.2. Determinação dos valores de IC_{50}

Em seguida, os compostos que apresentaram alguma atividade inibitória frente à protease do DENV, mais de 50% de inibição em 100 μ M, tiveram seus valores de IC_{50} determinados. Os ensaios de determinação dos valores de IC_{50} foram realizados, de modo análogo ao já descrito, variando-se, no entanto, a concentração do inibidor (de 0 a 500 μ M ou de 0 a 1 mM, de acordo com a potência do inibidor).

Procedendo-se como descrito acima, o valor de velocidade inicial de formação do produto (v_0) para cada concentração de inibidor utilizada foi obtido do gráfico de fluorescência em função do tempo, em segundos. A seguir, usando a função LL.2 (COPELAND, 2005) do pacote “*drc*” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013) foi construída a curva dose-resposta, ou seja, gráfico dos valores da fração de enzima ativa (v_i/v_0) em função do logaritmo da concentração do inibidor, sendo v_i a velocidade inicial na presença do inibidor e v_0 a velocidade inicial na ausência do composto. Os valores de IC_{50} foram obtidos a partir da equação de reta de melhor ajuste das curvas dose-resposta (vide Anexo III).

3.6.1.3. Caracterização do mecanismo de inibição e determinação do valor(es) da constante de inibição (K_i)

Por fim, para alguns compostos o mecanismo de inibição foi caracterizado e o correspondente valor(es) da constante de inibição (K_i) determinado.

Procedendo-se como descrito (COPELAND, 2005), a caracterização do mecanismo de inibição foi feita através dos gráficos do Duplo-Recíproco (Equação de Lineweaver- Burk) e de Dixon (gráfico dos inversos da velocidade inicial ($1/v_0$) em função da concentração de inibidor). Para se obter o gráfico Duplo-Recíproco (Equação de Lineweaver- Burk), ensaios de inibição enzimática frente à protease do DENV2 foram feitos de modo análogo ao descrito na etapa de determinação dos valores de inibição utilizando, no entanto, quatro diferentes concentrações do inibidor (variando de 0 a 15 μM) e onze concentrações de substrato (variando de 1,6 a 200 μM). Estes experimentos foram feitos em duplicata. O gráfico de Dixon foi obtido realizando os ensaios de inibição usando-se, no entanto, duas concentrações de substrato (20 e 40 μM) e seis diferentes concentrações do inibidor (0 a 16 μM). Estes experimentos foram feitos em duplicata. Por fim, o valor da constante de inibição (K_i) foi obtido através de dois gráficos, a saber: (i) gráfico da inclinação das retas do gráfico Duplo-Recíproco em função da concentração do inibidor e (ii) gráfico do inverso dos valores de velocidade máxima aparente ($1/V_{\text{max}}^{\text{app}}$) em função da concentração do inibidor, os valores de $V_{\text{max}}^{\text{app}}$ foram calculados a partir das retas de melhor ajuste dos gráficos de Michaelis-Menten obtidas utilizando a função “*nls*” do programa *R* (R CORE TEAM, 2013).

3.6.2. Testes de inibição da sLb^{pro} do vírus da FA

Todos os compostos selecionados pelos dois modelos de busca virtual (Modelo I – FA a Modelo II – FA) como potenciais inibidores da sLb^{pro} do vírus da FA foram submetidos a uma etapa de validação experimental, através de ensaios de inibição enzimática. Os testes de inibição da sLb^{pro} do vírus da FA foram realizados nos laboratórios da Prof^a. Dr^a. Antonia Tavares do Amaral (Departamento de Química Fundamental do Instituto de Química da Universidade de São Paulo – SP) usando a sLb^{pro} do vírus da FA (sorotipo O1k) disponibilizada pelo Prof. Dr. Luiz Juliano Neto (Departamento de Biofísica da UNIFESP).

Nesta etapa de validação, os compostos selecionados pelo Modelo I – FA como potenciais inibidores covalentes da sLb^{pro} do vírus da FA tiveram seus valores de IC₅₀ determinados. Em seguida, para os compostos que apresentaram alguma atividade inibitória frente à sLb^{pro} do vírus da FA, a reversibilidade da inibição foi avaliada. Para os compostos selecionados pelo Modelo II – FA como potenciais inibidores da sLb^{pro}, foram determinados os valores de porcentagem de inibição de cada composto (na concentração de 200 µM).

3.6.2.1. Determinação dos valores de IC₅₀ para os compostos selecionados pelo Modelo I – FA

Diferentemente do descrito para protease do DENV, todos compostos selecionados pelo Modelo I – FA como potenciais inibidores covalentes da sLb^{pro} do vírus da FA tiveram seus valores de IC₅₀ determinados.

Procedendo-se como descrito na literatura (COPELAND, 2005; SANTOS et al., 2009), os ensaios de inibição foram realizados utilizando o substrato fluorogênico KVQRKLKGAGQSSQ-Abz (dissolvido em DMSO) sintetizado e disponibilizado pelo grupo do Prof. Juliano (UNIFESP). As medidas foram feitas utilizando cubeta de quartzo (1,4 mL e caminho óptico de 10 mm) contendo solução tampão de ensaio (pH=7,8 [Borate (50 mM)]), Triton X-100 (0,01 % v/v) e 8,5 pM de sLb^{pro} do vírus da FA com volume final de 1 mL. Para a ativação da enzima, esta solução foi mantida, sob agitação, por 5 min na presença de DTT (2,5 mM) em um banho termostaticado (ajustado para 37 ± 0,1 °C). A reação foi iniciada pela adição do substrato KVQRKLKGAGQSSQ-Abz numa concentração final de 3 µM. A variação da intensidade de fluorescência foi seguida em um fluorímetro equipado com agitador magnético e banho termostaticado (ajustado para 37 ± 0,1 °C), emitindo radiação em 320 nm e registrando-se fluorescência em 420 nm. A seguir, alíquotas 1,0 µL de uma solução estoque do composto foram adicionadas à cubeta e a variação da intensidade de fluorescência

foi acompanhada, até que se fosse observada ou a total inibição (*i.e.* valores de fluorescência mantidos constantes durante o tempo do experimento) ou a perda de solubilidade na cubeta. As soluções estoque dos compostos foram preparadas em DMSO e logo utilizadas nos testes de inibição. O efeito do DMSO na atividade enzimática foi avaliado com um experimento controle feito nas mesmas condições, no entanto, adicionando alíquotas de DMSO. Todos os experimentos foram feitos, pelo menos, em duplicata.

De modo análogo ao descrito pra NS2B/NS3pro do DENV, o valor de velocidade inicial de formação do produto (v_0) para cada concentração de inibidor utilizada foi obtido do gráfico de fluorescência em função do tempo, em segundos. A seguir, utilizando a função LL.2 do pacote “*drc*” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013) foi construída a curva dose-resposta, ou seja, gráfico dos valores da fração de enzima ativa (v_i/v_0) em função do logaritmo da concentração do inibidor, sendo v_i a velocidade inicial na presença do inibidor e v_0 a velocidade inicial na ausência do composto. Os valores de IC_{50} foram obtidos a partir da equação de reta de melhor ajuste das curvas dose-resposta (vide Anexo III).

3.6.2.2. Avaliação da reversibilidade da inibição para os compostos selecionados pelo Modelo I – FA

Em seguida, os compostos selecionados pelo Modelo I - FA que apresentaram alguma atividade inibitória frente à sLb^{pro} do vírus da FA tiveram a reversibilidade da inibição foi avaliada. Estes experimentos foram realizados seguindo protocolo adaptado da literatura (COPELAND, 2005; SANTOS et al., 2009). E, a reversibilidade da inibição foi avaliada medindo-se a recuperação da atividade enzimática, imediatamente, após uma grande diluição (~100 vezes) da solução contendo o complexo enzima-inibidor. Este protocolo foi validado utilizando-se E64 como referência, sendo este descrito como inibidor covalente da sLb^{pro} do vírus do FA ($K_i = 3,4 \pm 0,1 \mu M$ (NOGUEIRA SANTOS et al., 2012)).

Inicialmente, foi preparada uma solução aquosa contendo tampão do ensaio (pH=7,8 [Borato (50 mM)]), Triton X-100 (0,01 % v/v) e 170 nM de sLb^{pro} do vírus da FA com volume final de 100 µL. Para a ativação da sLb^{pro}, esta solução foi mantida, sob agitação, por 5 min na presença de DTT (2,5 mM) em um banho de gelo. A seguir, a esta solução da enzima ativada foi adicionado diferentes alíquotas de uma solução do composto (em DMSO) de acordo com a potencia do mesmo. A solução foi mantida em um banho gelo por 30 min, com agitação a cada 5 min. Em seguida, uma alíquota de 10,0 µL foi retirada desta solução e diluída uma cubeta de quartzo (1,4 mL e caminho óptico de 10 mm) contendo 990,0 µL do tampão de ensaio (pH=7,8 [Borato (50 mM)]). A atividade enzimática foi medida de modo análogo ao descrito na etapa de determinação dos valores de IC₅₀. As soluções estoque dos compostos foram preparadas em DMSO e logo usadas nos testes de inibição.

As curvas de progresso da reação (ou seja, valores de fluorescência em função do tempo, em segundos) foram construídas para cada composto e para o inibidor E64. Por fim, procedendo-se como descrito na literatura (COPELAND, 2005), a reversibilidade da inibição foi avaliada comparando a inclinação destas curvas de progresso da reação com a inclinação da curva do experimento controle, realizado nas mesmas condições, no entanto, sem adicionar nenhum composto/inibidor.

3.6.2.3. Determinação dos valores de porcentagem de inibição para os compostos selecionados pelo Modelo II – FA

Os compostos selecionados pelo Modelo II - FA como potenciais inibidores da Lb^{pro} do vírus da FA foram avaliados através de ensaios enzimáticos. Todos estes compostos tiveram seus valores de valores de porcentagem de inibição frente à sLb^{pro} do vírus da FA determinados (na concentração de 200 µM).

Procedendo-se como descrito na literatura (NOGUEIRA SANTOS et al., 2012), os ensaios de inibição foram realizados utilizando o substrato fluorogênico Z-LR-AMC (dissolvido em DMSO) disponível comercialmente. As medidas foram feitas utilizando cubeta de quartzo (1,4 mL e caminho óptico de 10 mm) contendo solução tampão de ensaio (pH=7,8 [Borato (50 mM)]), Triton X-100 (0,01 % v/v) e 8,4 nM de sLb^{pro} do vírus da FA com volume final de 1 mL. Para a ativação da enzima, esta solução foi mantida, sob agitação, por 5 min na presença de DTT (2,5 mM) em um banho termostatizado (ajustado para $37 \pm 0,1$ °C). A seguir, o composto (200 µM) foi adicionado e a reação foi iniciada adicionando substrato numa concentração final de 100 µM. A variação da intensidade de fluorescência foi seguida em um fluorímetro equipado com agitador magnético e banho termostatizado (ajustado para $37 \pm 0,1$ °C), emitindo radiação em 380 nm e registrando-se fluorescência em 460 nm. As soluções estoque dos compostos foram preparadas em DMSO e logo utilizadas nos testes de inibição. A quantidade de DMSO nos experimentos foi mantida constante, sendo menor ou igual a 0,01 % (v/v). Todos os experimentos foram feitos, pelo menos, em triplicata.

O valor de velocidade inicial de formação do produto (v_0) para cada inibidor foi obtido do gráfico de fluorescência em função do tempo, em segundos. A seguir, os valores de porcentagem de inibição foram calculados comparando o valor de v_0 obtido para cada inibidor com o valor obtido em experimento controle, realizado de modo análogo ao descrito acima, no entanto, sem adicionar nenhum composto.

3.6.3. Avaliação da estabilidade de cada um dos inibidores nas condições do experimento de inibição

Os compostos adquiridos/selecionados que apresentaram atividade inibitória tiveram sua estabilidade nas condições do experimento (ou seja, nas soluções tampão dos ensaios) de inibição avaliada. A estabilidade de cada um dos inibidores foi analisada acompanhando alterações através dos correspondentes espectros de absorção. Nesta análise, para os

compostos adquiridos como potenciais inibidores da NS2B/NS3pro foi utilizada solução tampão de ensaio pH = 9,0 [Tris-HCl (50 mM) e Chaps (1 mM)] e para os compostos adquiridos/selecionados como potenciais inibidores da sLb^{pro} do vírus da FA foi utilizada solução tampão de ensaio pH = 7,8 [Borato (50 mM)]. Para tanto, foi preparada uma solução estoque de cada um dos compostos na solução tampão do correspondente ensaio (pH = 9,0 ou pH = 7,8), a partir de soluções estoque recém-preparadas (em DMSO). Estas soluções foram mantidas à 37 °C com agitação de 500 rpm, utilizando agitador com controle de temperatura. E, a seguir, os espectros foram registrados entre 200 nm – 600 nm, à cada 15 min, durante 1 hora usando um espectrofotômetro Hitachi U-2010 (UV/VIS). Todas as medidas foram feitas em cubeta de quartzo (de 2,0 mL e caminho óptico de 10,0 mm) usando-se como branco 2 mL da correspondente solução tampão do de ensaio (pH = 9,0 ou pH = 7,8). Em seguida, para cada composto, valores de λ^{\max} foram selecionados, a partir dos espectros de absorbância. Variações de absorbâncias nos λ^{\max} selecionados foram utilizadas para avaliar sua estabilidade de cada composto.

3.6.4. Identificação e/ou caracterização dos compostos adquiridos

A identificação e/ou caracterização de alguns compostos adquiridos foi feita registrando-se os respectivos espectros de RMN¹H.

Em um tubo de ressonância foi preparada uma solução de cada composto em DMSO-d₆ 99,9 % (18,8 µM) com TMS (0,1 µL). O espectro de RMN¹H desta solução foi registrado utilizando espectrômetro, operando em transformada de Fourier. O sinal do TMS (0 ppm) foi utilizado como referência para calcular os valores do deslocamento químico dos sinais no espectro de RMN¹H, expressos em ppm. Os compostos selecionados pelo primeiro modelo Modelo I –DENV, ainda tiveram sua estabilidade em DMSO avaliada. Para isto, a solução utilizada para registrar o espectro de RMN¹H foi armazenada a 0 °C durante 3 dias. Após este

período, o espectro de RMN¹H desta solução foi registrado usando as condições descritas acima. Os dois espectros foram comparados considerando os valores do deslocamento químico de cada sinal. Adicionalmente, o espectro do Reagente de Purpald®, que contém parte da estrutura dos compostos selecionados pelo Modelo I –DENV, foi registrado e usado para confirmar os valores de deslocamentos químicos atribuídos aos átomos de hidrogênio ligados aos átomos de nitrogênio.

4. RESULTADOS E DISCUSSÃO

4.1. Introdução

Esta Tese de Doutorado foi desenvolvida no Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos do Instituto de Química da Universidade de São Paulo, liderado pela Profa. Associada Antonia Tavares do Amaral, que vem se dedicando, nos últimos quase 30 anos, ao estudo de diversos aspectos envolvendo compostos bioativos (AMARAL et al., 2017). Entre eles, destacam-se estudos de Relações Quantitativas entre Estrutura Química e Atividade biológica, incluindo-se aqueles com modelagem molecular (QSAR/QSAR-3D) bem como estudos para a determinação de propriedades físico-químicas e/ou estruturais (como lipofilicidade e variação de densidade eletrônica no grupo carbonila) de compostos com atividade biológica. Dentre as classes terapêuticas estudadas pelo grupo, destacam-se: antiarrítmicos e anestésicos locais (CHEFFER et al., 2012; MALVEZZI; DO AMARAL, 2010), antitumorais (RAMINELLI, 2001), antibacterianos (PIRES et al., 2001), antivirais anti-HIV (ISHIKI; GALEMBECK, S. E. AMARAL, 2001); tuberculostáticos (ISHIKI; AMARAL, 2009) e antichagásicos (MALVEZZI, 2008; MALVEZZI et al., 2008; MALVEZZI; DE REZENDE; AMARAL, 2009; SCOTTI et al., 2016). Estes estudos foram planejados visando-se a contribuir para elucidação do mecanismo de ação destes compostos e, quando possível, para prever derivados mais potentes e/ou menos tóxicos. No contexto do planejamento de compostos bioativos, estes estudos consideraram apenas informações do sistema químico, sem considerar a estrutura do alvo biológico, podendo ser classificados como planejamento de fármacos baseado no ligante (do inglês, *Ligand Based Drug Design*). Conforme apresentado por Malvezzi (MALVEZZI, 2008), com a disponibilidade de recursos de hardware e de software para manipulação e tratamento de grande número de dados, foi possível incluir no planejamento de compostos bioativos informações sobre a estrutura do

alvo molecular bem como suas interações com os ligantes. Assim, nos últimos dez anos, além de estratégias de planejamento baseadas no ligante, aquelas com base no receptor (do inglês, *Structure Based Drug Design*) também passaram a ser aplicadas no grupo e para colaboradores (MALVEZZI et al., 2011, 2012; MALVEZZI; DE REZENDE; AMARAL, 2009; MILANI et al., 2014).

O presente trabalho, desenvolvido no âmbito de um projeto de doutoramento no Instituto de Química da Universidade de São Paulo, sob a orientação da Profa. Associada Antonia Tavares do Amaral, foram estudados e incluídos aspectos relacionados à flexibilidade do alvo molecular, até então inéditos ao Laboratório de QSAR e Modelagem Molecular de Compostos Bioativos. Além disso, neste trabalho, foram aplicados diversos filtros físico-químicos e/ou estruturais para a construção de banco de dados de moléculas virtuais, o qual teve sua diversidade estrutural subsequentemente analisada, sendo estes também inéditos no grupo. As estratégias desenvolvidas e/ou implementadas foram: (i) a simulações de Dinâmica Molecular (DM) e análise dos resultados obtidos (ver Introdução item 1.10), (ii) o ancoramento de moléculas orgânicas pequenas usando diferentes conformações da proteína alvo obtidas por simulações de DM (ver Introdução item 1.10), (iii) a construção de modelos farmacofóricos utilizando diferentes modos de ligação obtidos durante a simulação de DM (ver Introdução item 1.7.1), (iv) a aplicação de filtros físico-químicos e/ou estruturais durante a construção do banco de dados de moléculas virtuais (ver Introdução item 1.9), (v) a análise da diversidade química e estrutural do banco de dados construído através de métodos estatísticos e ferramentas quimiométricas. Estas estratégias foram aplicadas, junto com outras metodologias já implementadas no grupo, para a descoberta de novos inibidores de proteases do vírus da Dengue e do vírus da Febre Aftosa.

Os resultados obtidos, descritos nos itens 4.2 e 4.4, que serão e apresentados, a seguir, se referem à: (i) à criação dos bancos de dados de moléculas virtuais (em especial do banco

denominado nesta tese *ZINC-Curated*) bem como à análise da diversidade estrutural do banco de dados *ZINC-Curated*; (ii) ao reconhecimento da estrutura da protease do vírus da Dengue (NS2B/NS3pro do DENV) bem como o estudo da flexibilidade conformacional desta protease; (iii) preparação, construção e validação dos quatro modelos de busca virtual de inibidores da NS2B/NS3pro do DENV, denominados nesta tese Modelo I – DENV ao Modelo IV – DENV; (vi) à construção de uma série de inibidores da NS2B/NS3pro do DENV considerando-se a estrutura do inibidor mais potente obtido pelos modelos de busca virtual; (v) ao reconhecimento da estrutura da protease líder do vírus da Febre Aftosa (Lb^{pro} do vírus FA) e (vi) preparação, construção e validação dos dois modelos de busca virtual de inibidores da Lb^{pro} do vírus da FA, denominados nesta tese Modelo I – FA e Modelo II –FA. Deve-se esclarecer que, os resultados obtidos nesta tese serão apresentados na mesma sequência usada em Materiais e Métodos, ou seja, primeiramente para a NS2B/NS3pro do DENV e depois para a Lb^{pro} do vírus da FA.

4.2. Banco de dados de moléculas virtuais

4.2.1. Construção dos bancos de dados de moléculas virtuais

Uma das etapas fundamentais na busca virtual de compostos bioativos é a preparação dos bancos de dados de moléculas virtuais (BOLOGA; OLAH; OPREA, 2006; SCIOR et al., 2012). Assim, nesta tese, foram preparados e construídos três diferentes bancos de dados de moléculas virtuais, denominados: *ZINC-Original*, *NCI* e *ZINC-Curated*. Procedendo-se como descrito no item 3.5.1.1 (Materiais e Métodos), para os dois primeiros, a redução do número de moléculas foi feita unicamente pela remoção de estruturas duplicadas usando o programa *Open Babel* (O'BOYLE et al., 2011). A Tabela 4.1 apresenta os resultados obtidos: número de moléculas presentes originalmente, no final da construção dos bancos *ZINC-Original* e *NCI* e removidas neste processo, bem como as porcentagens de redução alcançadas (1,44 e

3,75% para os bancos *ZINC-Original* e *NCI*, respectivamente). Por fim, considerando-se que algumas moléculas apresentam certo grau de flexibilidade, fez-se necessário gerar diferentes conformações (até 30 para cada molécula).

Tabela 4.1. Número de moléculas presentes originalmente, no final da construção e removidas neste processo para os bancos de dados *ZINC-Original* e *NCI*. Entre parêntesis são indicadas as correspondentes porcentagens de redução observadas em cada um dos bancos de dados.

Banco de dados	Número de moléculas presentes		
	Originalmente	No final	Removidas ^(a)
<i>ZINC-Original</i>	23.604.544	23.265.387	339.157 (1,4 %)
<i>NCI</i>	280.817	270.569	10.248 (3,6 %)

^(a) Correspondendo a estruturas duplicatas removidas pelo programa Open Babel (O'BOYLE et al., 2011).

Ainda, uma das estratégias para enriquecer bancos de dados com moléculas com características de fármacos é aplicação de alguns filtros de propriedades “*drug-like*” e estruturais. Isto é feito visando melhorar a qualidade dos compostos selecionados pela busca virtual bem como evitar resultados falso-positivos. Estas estratégias foram usadas para criar um banco de dados enriquecido com compostos com características de fármacos, sendo este denominado *ZINC-Curated*. Desta forma, além das etapas aplicadas na construção dos bancos *ZINC-Original* e *NCI* foram incluídos mais dois filtros, a saber: filtro de predição de solubilidade e filtro de propriedades “*drug-like*” e/ou estruturais. Todos os procedimentos usados na construção do banco *ZINC-Curated* estão descritos no item 3.5.1.1.

Existem inúmeras limitações na predição da solubilidade aquosa de um composto. Reconhecendo, no entanto, todas estas limitações e, ao mesmo tempo, sabendo da importância de se considerar a solubilidade nas etapas iniciais da busca virtual, optamos por utilizar um modelo de predição baseado no “*Solubility Forecast Index- SFI*” (HILL; YOUNG, 2010). Resumidamente, este modelo prevê que moléculas com valores de SFI > 5,00 apresentariam grande probabilidade de terem uma baixa solubilidade em água (para mais detalhes vide item 1.6.1). Além deste filtro, optamos por utilizar filtros de propriedades “*drug-like*” e/ou estruturais visando selecionar moléculas com outras características desejáveis para um

fármaco (BOLOGA; OPREA, 2012). Dentre estas, se consideram as seguintes: ter boa biodisponibilidade via oral (“Regra dos 5” de Lipinski (LIPINSKI et al., 1997)); apresentar pelo menos um átomo doador e um átomo aceptor de ligação de hidrogênio (N-H e/ou O-H e O e/ou N, respectivamente); ter mais de nove átomos; ter pelo menos um heteroátomo, ter menos de quatro átomos de halogênios; apresentar menos de dez centros quirais; apresentar menos de dezesseis ligações com livre rotação e não ter grupamentos reativos (descritos no item 1.3.1). Este filtro foi proposto considerando os critérios “*drug*” definidos pelo programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]).

A Tabela 4.2 apresenta o número de moléculas presentes originalmente, no final da construção do banco *ZINC-Curated* e removidas em cada etapa deste processo, acompanhadas das correspondentes porcentagens de redução alcançadas. A remoção de estruturas duplicadas apresentou uma redução menor que 1%. Enquanto que a remoção de moléculas preditas como insolúveis ($SFI > 5,01$), foi a etapa que mais contribuiu para a redução removendo $\sim 8 \times 10^6$ moléculas. Ainda, um número significativo de moléculas ($\sim 3,6 \times 10^6$) foi removido pelo filtro de propriedades “*drug-like*” e/ou estruturais. Considerando todas as etapas de redução envolvidas na construção do banco *ZINC-Curated*, o percentual de moléculas removidas foi igual a 59,9%.

Tabela 4.2. Número de moléculas presentes originalmente, no final e removidas em cada uma das etapas de redução utilizadas na construção do banco de dados *ZINC-Curated*. Entre parêntesis são indicadas as correspondentes porcentagens de redução em cada etapa.

Etapa de redução	Número de moléculas presentes		
	Originalmente	Final	Removidas
Remoção de duplicadas ^(a)	19.607.722	19.450.010	157.712 (0,8 %)
Filtro de solubilidade ^(b)	19.450.010	11.445.533	8.004.477 (41,2 %)
Filtro de propriedades “drug-like” e estruturais ^(c)	11.445.533	7.862.087	3.583.446 (31,3 %)
Todas as etapas	19.607.722	7.862.087	11.745.635 (59,9 %)

^(a) Removidas pelo programa *Open Babel* (O’BOYLE et al., 2011)

^(b) Seguindo o modelo de predição do *SFI* (HILL; YOUNG, 2010)

^(c) Seguindo os critérios presentes no filtro “Drug” do programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]), com algumas modificações.

Adicionalmente, uma das fontes de erros nos procedimentos de busca virtual é à presença de estruturas escritas incorretamente no banco de dados de moléculas virtuais (SCIOR et al., 2012). Assim, na construção do banco *ZINC-Curated* foi incluída uma etapa de correção de estruturas, considerando-se seus correspondentes tautômeros e estados de protonação. No entanto, visando evitar o aumento do número de compostos estruturalmente semelhantes no banco de dados, nesta etapa, foram incluídos apenas o tautômero mais abundante (em água) e o estado de protonação mais favorável (em pH = 7,4). Por fim, de modo análogo ao descrito acima para os bancos *ZINC-Original* e *NCI*, para cada molécula presente no banco *ZINC-Curated* foram geradas até 30 possíveis conformações.

4.2.2. Análise da diversidade das propriedades físico-químicas e/ou estruturais no banco *ZINC-Curated*

Como apresentado e discutido na Introdução, embora a aplicação destes filtros “drug-like” e estruturais seja recomendada na literatura (BAELL; WALTERS, 2014; BOLOGA; OPREA, 2012; LIPINSKI et al., 1997; MEANWELL, 2011; MIGNANI et al., 2016), esta é também bastante criticada (DOAK et al., 2016; LIPINSKI, 2016; SCIOR et al., 2012). Uma

das principais críticas à aplicação destes filtros é a possível diminuição significativa da diversidade estrutural e das correspondentes propriedades físico-químicas dos compostos presentes nestes bancos. Considerando-se todas as limitações bem como as opiniões divergentes ressaltadas na literatura sobre o uso desta estratégia, fizemos uma análise crítica, usando ferramentas estatísticas, sobre o efeito da aplicação destes filtros na diversidade estrutural dos compostos no banco *ZINC-Curated*. Ainda, a diversidade das características físico-químicas e/ou estruturais dos compostos deste banco também foi analisada.

Procedendo-se como descrito no item 3.5.1 (Materiais e Métodos), as propriedades físico-químicas e estruturais dos compostos do banco *ZINC-Curated* foram comparadas com as dos compostos presentes em diferentes bancos de dados de moléculas virtuais dentre os disponíveis gratuitamente, já usados como referência na área de planejamento de compostos bioativos (SCIOR et al., 2012). A escolha destes bancos de dados foi feita com o objetivo de analisar se os critérios e os filtros usados na construção do banco *ZINC-Curated* foram capazes, ou não, de enriquecer este banco com moléculas com características de fármacos sem, no entanto, comprometer significativamente a sua diversidade estrutural. Assim, foram escolhidos bancos de dados que representassem os espaços químicos tanto dos fármacos e dos compostos com alguma atividade biológica bem como representassem o espaço químico com grande diversidade estrutural. Os bancos selecionados estão apresentados a seguir:

- DrugBank (denominado DB), representando o espaço químico dos fármacos;
- Subconjuntos de produtos naturais do banco de dados *ZINC* (denominado ZNP), representando o espaço químico com grande diversidade estrutural;
- ChEMBL (denominado ChEMBL), representando o espaço químico de compostos já avaliados experimentalmente quanto a sua atividade biológica;
- Versão original e atualizada do *ZINC* (denominado *ZINC-2014*), representando a diversidade química e estrutural esperada para bancos de moléculas comerciais sem aplicar filtro algum.

De modo análogo ao descrito acima, na etapa de preparação destes bancos as estruturas duplicadas foram removidas. Para o banco DB, foram também removidos os compostos classificados neste banco como nutracêuticos (e.g. vitaminas e suplementos alimentares). A Tabela 4.3 mostra o número de moléculas presentes originalmente em cada banco, no final e removidas. Nesta tabela, podemos observar que os bancos ZNP e ZINC-2014 apresentavam um número significativo de estruturas duplicadas, correspondendo a 11,8 e 26,2 % do número de moléculas originalmente presentes em cada um dos bancos.

Tabela 4.3. Número de moléculas originalmente presentes, no final da construção, removidas e na amostra representativa de cada um dos bancos de dados.

Banco de dados	Número de moléculas			Na amostra representativa ^(a)
	Originalmente	Final	Removidas (em %)	
DB	1.826	1.607	219 (12,0 %) ^(b)	1.377
ZNP	166.544	146.882	19.662 (11,8 %) ^(c)	9.015
ChEMBL	1.456.020	1.454.442	1.578 (0,1 %) ^(c)	9.541
ZINC-2014	22.724.825	16.782.126	5.942.699(26,2 %) ^(c)	9.599
<i>ZINC-Curated</i>	19.607.722	7.862.087	11.745.635 (59,9 %) ^(d)	9.592

^(a) Calculado considerando intervalo de confiança de 1% e um nível de confiança de 95%.

^(b) Remoção das duplicadas e dos nutracêuticos.

^(c) Remoção apenas as duplicadas.

^(d) Considerando todas as etapas descritas na Tabela 4.2.

Observa-se na Tabela 4.2 que o banco ChEMBL é o único onde o número de moléculas duplicadas é pequeno (0,1 % de moléculas removidas). E ainda, em relação às estruturas presentes originalmente no banco DB, 139 estruturas removidas correspondem aos compostos classificados como nutracêuticos. Ainda, vale lembrar que o número de moléculas presentes no final da construção de cada um dos bancos ChEMBL, ZINC-2014 e *ZINC-Curated* é muito grande ($> 1,5 \times 10^6$). Considerando-se a grandeza destes bancos de dados e visando-se a viabilização das análises apresentadas a seguir (itens 4.2.2.2, 4.2.2.3 e 4.2.2.6), fez-se necessário reduzir o número de compostos a serem analisados, sem que houvesse perda

de informações relevantes. Esta redução foi feita gerando amostras estatisticamente representativas de cada banco de dados.

Assim, como descrito item 3.5.1.2 (Materiais e Métodos), o número de compostos necessários para gerar três amostras representativas de cada um dos cinco bancos de dados analisados foi calculado usando ferramentas estatísticas e está apresentado na Tabela 4.3. Esta estratégia de análise permitiu reduzir significativamente o número de compostos a serem analisados de cerca de 44 milhões para apenas 39 mil, permitindo, assim, o cálculo de diferentes propriedades físico-químicas e/ou estruturais e, principalmente, a análise destas propriedades, apresentadas nos próximos parágrafos.

4.2.2.1. Validação estatística da etapa de amostragem

Considerando-se a necessidade de gerar amostras representativas, antes de se iniciar a análise da diversidade das propriedades físico-químicas e/ou estruturais dos compostos presentes no banco *ZINC-Curated*, fez-se necessário validar esta etapa. Todos os procedimentos envolvidos nesta validação foram realizados usando testes estatísticos e analisando gráficos de caixa, que estão descritos no item 3.5.3 (Materiais e Métodos). Para tanto, foram usadas nove propriedades físico-químicas comumente empregadas na literatura para descrever bancos de dados de moléculas, a saber: massa molecular, XlogP, LogD_{pH7,5} calculados, área de superfície polar (X2d.PSA), número de átomos; número de anéis; número de ligações com rotação livre; número de átomos doadores e de aceptores de ligação de hidrogênio. Nesta tese, átomos ou regiões doadoras e aceptoras de ligação de hidrogênio foram abreviados como *HBD* e *HBA* (do inglês *hydrogen bond donor and acceptor*), respectivamente. Esta etapa de validação foi feita apenas para o banco DB, uma vez que este banco possui apenas 1.607 moléculas sendo, assim, possível calcular e analisar cada uma das nove propriedades físico-químicas descritas para toda a população do banco.

A escolha do teste estatístico a ser usado depende do tipo de distribuição das variáveis (ex. propriedades físico-químicas) que descrevem os bancos de dados (KRZYWINSKI; ALTMAN, 2014). Assim, a análise do tipo de distribuição das variáveis contínuas (massa molecular, XlogP, LogD_{pH7,5} calculado e X2d.PSA) foi feita usando gráficos de probabilidade Normal. Os gráficos de probabilidade Normal gerados para o banco DB mostram que os valores do peso molecular, XlogP e LogD_{pH7,5} calculados e da X2d.PSA não seguem uma distribuição Normal (Figura 4.1). Adicionalmente, o teste estatístico de *Kolmogorov-Smirnov* (bilateral para uma amostra) foi utilizado para confirmar a hipótese que estas variáveis não seguem uma distribuição Normal, mostrando valores de $p < 0,05$ e valores de $D > 0,03$. A partir destes resultados e assumindo-se que as variáveis contínuas não seguem uma distribuição Normal, testes estatísticos não paramétricos foram escolhidos e utilizados para validar a etapa de amostragem.

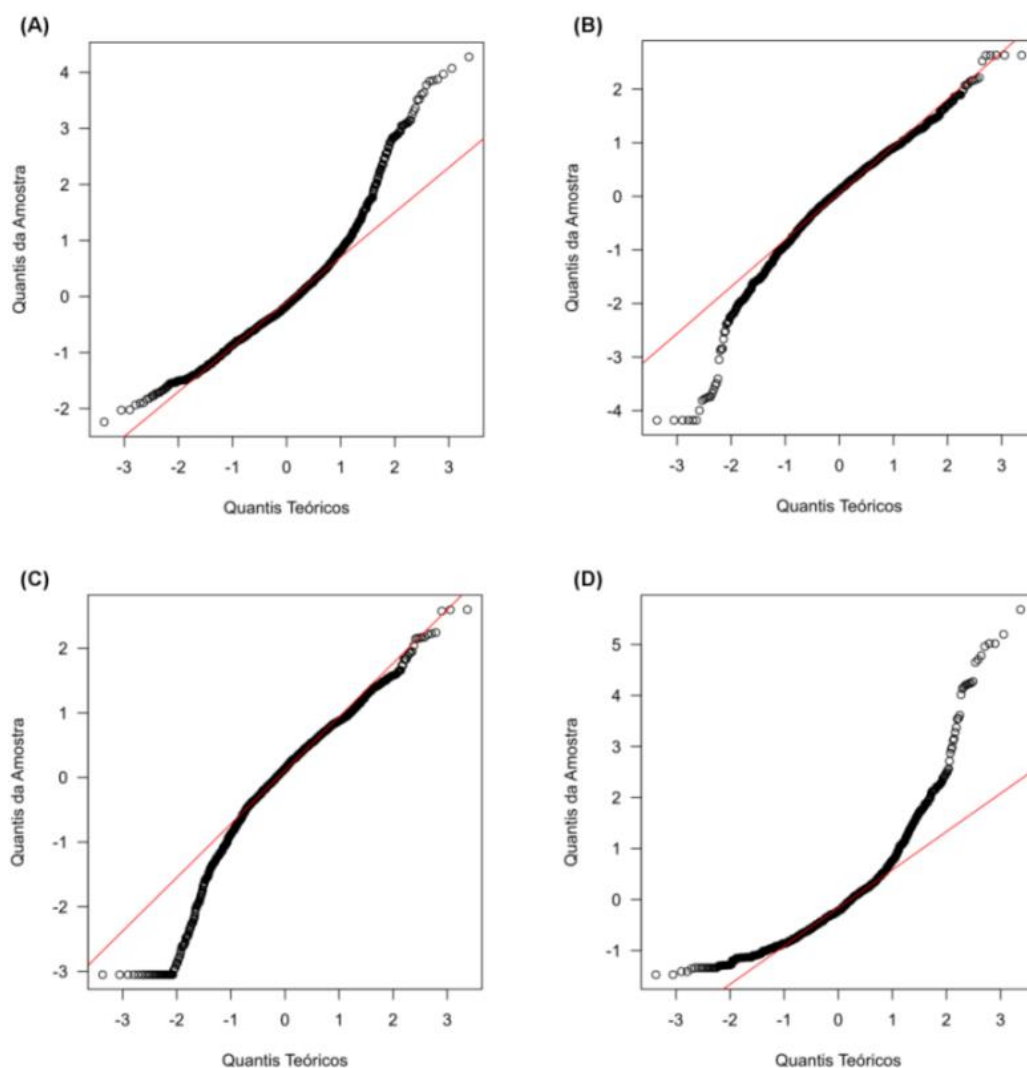


Figura 4.1. Análise do tipo de distribuição das variáveis contínuas para o banco DB. Gráficos de probabilidade Normal^(a) e teste *Kolmogorov-Smirnov* (bilateral para uma amostra)^(b) para os seguintes descritores: (A) peso molecular (valores de $p < 0,05$ e $D = 0,16$); (B) XlogP (valores de $p < 0,05$ e $D = 0,07$); (C) $\log D_{pH7,5}$ (valores de $p < 0,05$ e $D = 0,09$) e (D) Área de superfície polar – X2d.PSA (valores de $p < 0,05$ e $D = 0,10$).

^(a) Gráficos de probabilidade Normal (gráfico dos quantis teóricos obtidos para uma distribuição Normal em função dos quantis da amostra) foram gerados utilizando a função “*qqplot*” do programa R (R CORE TEAM, 2013). Neste gráfico, a linha vermelha representa os valores esperados se os descritores seguissem uma distribuição próxima da Normal.

^(b) O teste *Kolmogorov-Smirnov* (bilateral para uma amostra) foi feito utilizando a função “*ks.test*” do programa R (R CORE TEAM, 2013) e usando $\alpha = 0,05$ e $D_{crit} = 0,03$.

Em seguida, as nove propriedades físico-químicas selecionadas (representando variáveis contínuas e discretas) das três amostras do banco DB foram analisadas, separadamente, usando o teste estatístico não paramétrico de *Kruskal-Wallis* (SPRENT, 1988). Este teste comparou as amostras com base em suas funções de distribuição e verificou

se estas provêm, ou não, de uma mesma população. Para todos os descritores de todos os bancos valores de $p > 0,05$ foram obtidos (variando de 0,077 a 0,986), confirmando a hipótese de que as três amostras provêm da mesma população. A seguir, a distribuição destas propriedades em cada uma das amostras do banco DB (Amostra 1, 2 e 3) foi comparada com a distribuição observada para toda a população do banco DB, usando gráficos de caixa e de porcentagem acumulativa (Figura 4.2 e Figura 4.3, respectivamente).

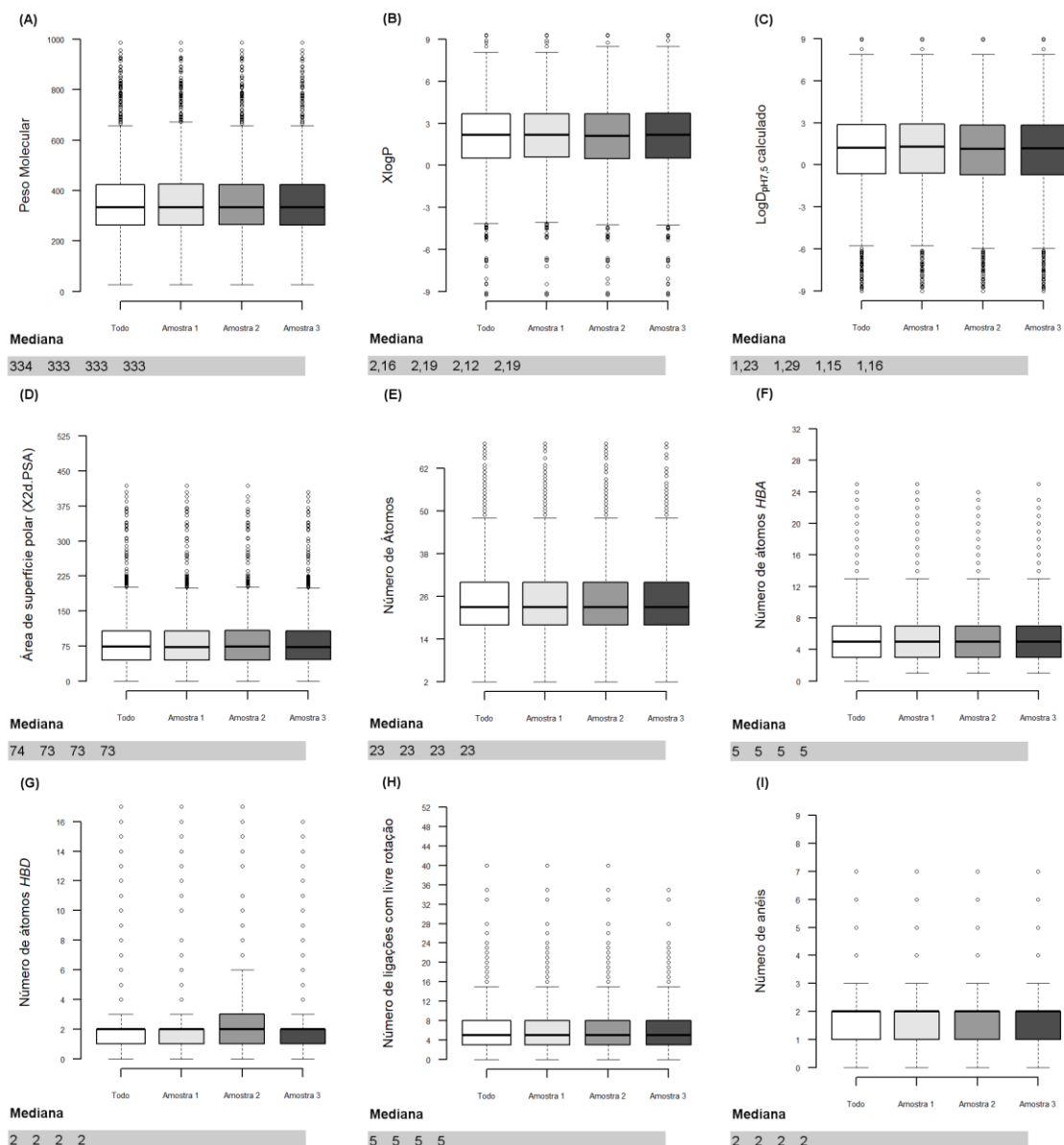


Figura 4.2. Análise comparativa entre as três amostras e toda a população do banco DB através dos gráficos de caixa^(a). Gráficos de caixa e valores de mediana para cada uma das propriedades calculadas^(b) apresentadas, sempre, na sequência, para todo banco (branco), Amostra 1 (cinza claro), Amostra 2 (cinza) e para Amostra 3 (cinza escuro): (A) peso molecular; (B) XlogP; (C) logD_{pH7,5} calculado; (D) Área de superfície polar (X2d.PSA); (E) número de átomos; (F) número de átomos HBA; (G) número de átomos HBD e (H) número de ligações com livre rotação e (I) número de anéis.

(a) Gráficos de caixa tipo *Tukey*, no qual as caixas indicam a distância interquartil ($Q_1 - Q_3$), a linha preta mostra o valor da mediana (Q_2), os *whiskers* (fio de bigode) são segmentos que estendem 1,5 vezes a distancia interquartil e os círculos representam os valores atípicos (*outliers*).

(b) Calculadas (item 3.5.3) usando os programas *Volsurf+* (CRUCIANI et al., 2000) e *Filter* ("FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM", [s.d.]).

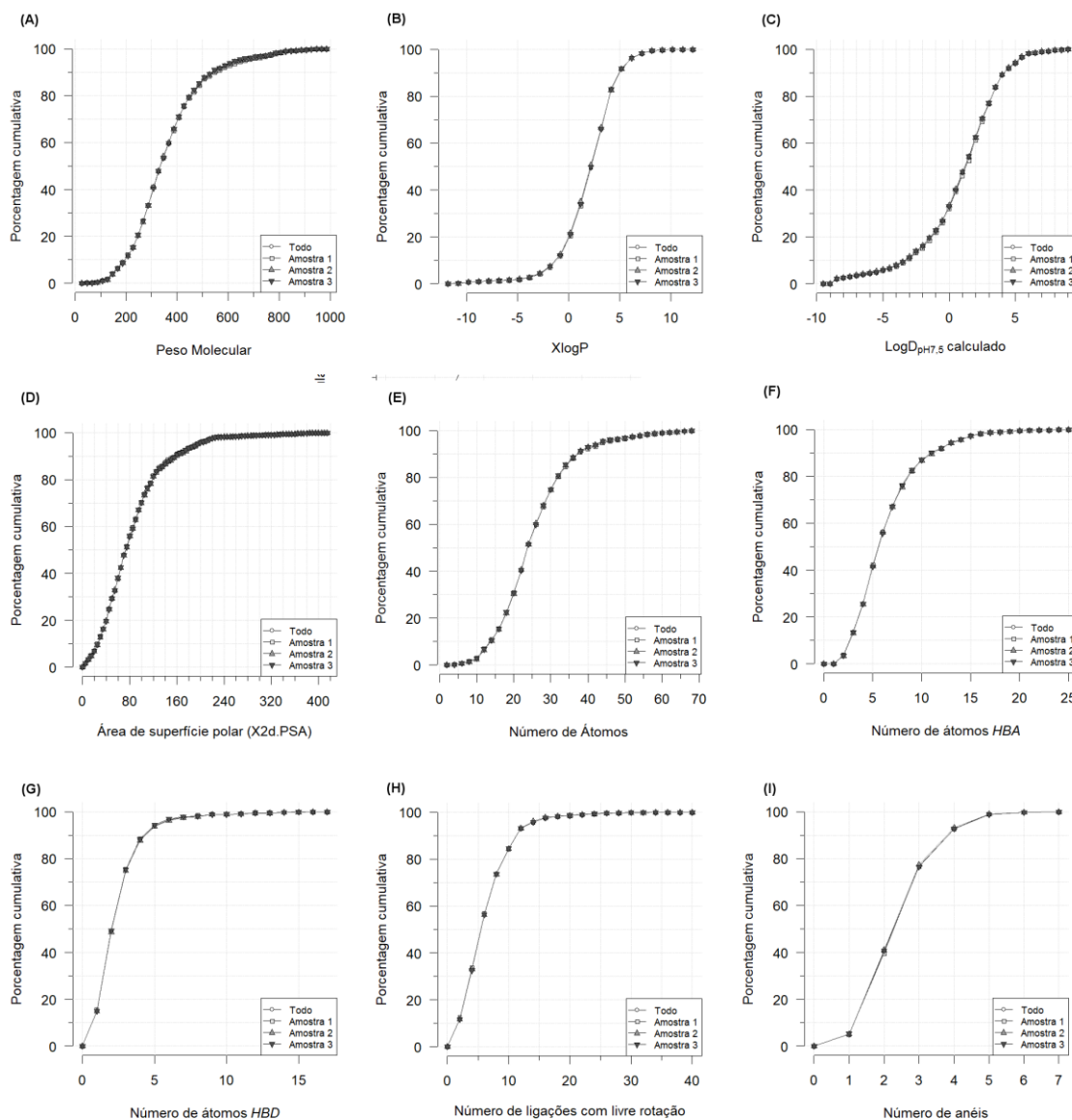


Figura 4.3. Análise comparativa entre das três amostras e toda a população do banco DB através dos gráficos de porcentagem acumulativa^(a). Gráficos da porcentagem cumulativa para cada uma das propriedades físico-químicas calculadas^(b) apresentadas, sempre, na sequência, para todo banco (circulo branco), Amostra 1 (quadrado cinza claro), Amostra 2 (triangulo cinza) e para Amostra 3 (triangulo invertido cinza escuro): (A) peso molecular; (B) XlogP; (C) logD_{pH7.5} calculado; (D) Área de superfície polar (X2d.PSA); (E) número de átomos; (F) número de átomos *HBA*; (G) número de átomos *HBD*; (H) número de ligações com livre rotação e (I) número de anéis.

(a) Gráficos da porcentagem cumulativa, no qual cada ponto representa o número de compostos (em %) com valores abaixo de um determinado valor de corte.

(b) Calculadas (item 3.5.3) usando os programas *Volsurf+* (CRUCIANI et al., 2000) e *Filter* ("FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM", [s.d.]).

Nos gráficos de caixa (Figura 4.2) podemos observar que os valores do primeiro quartil (Q_1), da mediana (Q_2) e do terceiro quartil (Q_3) das propriedades físico-químicas das três amostras são comparáveis com as observadas para toda a população do banco DB, exceto para o terceiro quartil relativo ao número de átomos *HBD* da Amostra 2 (Figura 4.2 (G)).

Contudo, nos gráficos de porcentagem cumulativa nenhuma diferença significativa entre as amostras e toda a população é observada (Figura 4.3) indicando que as amostras são representativas da população. Ainda, visando a entender a diferença observada no gráfico de caixa do número de átomos *HBD*, calculou-se para as três amostras bem como para a população o número de compostos (em %) contendo nenhum (zero) a 17 átomos *HBD* (vide Tabela 4.4).

Tabela 4.4. Número de compostos (em %) contendo nenhum (zero) a 17 átomos *HBD* calculados para a população e para cada uma das amostras.

Número de átomos <i>HBD</i>	Número de compostos (em%)			
	População	Amostra 1	Amostra 2	Amostra 3
0	15,09	15,51	14,85	14,97
1	33,76	33,78	33,66	34,40
2	26,34	26,25	26,34	25,91
3	12,85	12,98	12,76	13,03
4	5,95	5,74	6,04	6,11
5	2,49	2,31	2,46	2,46
6	1,09	1,27	1,27	0,97
7	0,58	0,52	0,52	0,30
8	0,70	0,60	0,82	0,67
9	0,06	0,22	0,07	0,07
10	0,19	0,30	0,22	0,22
11	0,26	0,07	0,30	0,30
12	0,06	0,15	0,30	0,07
13	0,26	0,07	0,15	0,22
14	0,13	0,07	0,07	0,15
15	0,06	0,07	0,07	0,07
16	0,06	0,07	0,07	0,07
17	0,06	0,00	0,00	0,00

Por definição, o primeiro (Q_1) e o terceiro (Q_3) quartis são valores que dividem o conjunto de dados ordenados em dois subconjuntos. Isto é feito de modo que 25% das observações ordenadas estejam abaixo do Q_1 e 75% das observações ordenadas estejam abaixo do Q_3 . Assim, 75% dos compostos ordenados pelo número de átomos *HBD* estão abaixo do Q_3 . Somando-se os valores da porcentagem de compostos contendo 0, 1 e 2 átomos *HBD* (Tabela 4.4) resultou em: 75,19% para a população; 75,54 % para Amostra 1, 74,85% para Amostra 2 e 75,28% para Amostra 3. Considerando-se a definição de Q_1 e Q_3 , podemos entender a razão pela qual na Amostra 2 o valor de Q_3 corresponde a 3 átomos *HBD*. A alteração do valor do terceiro quartil de 2 para 3 modifica tanto a distância interquartil (distância entre Q_1 e Q_3) bem como o tamanho do *whiskers* (1,5 vezes o valor da distância interquartil). Estas modificações alteram significativamente a aparência do gráfico de caixas. No entanto, a diferença entre a porcentagem de composto da Amostra 2 para as demais amostras e para a população é menor do que 0,7 %, portanto, dentro do erro da amostragem que é de $\pm 5\%$. Desta forma, embora o gráfico de caixa do número de átomos *HBD* da Amostra 2 seja diferente das demais amostras e da população, a Amostra 2 foi também considerada como representativa da população.

De acordo com os resultados descritos anteriormente, a etapa de amostragem foi considerada validada. Assim, assumindo-se que as amostras que se originam da mesma população são estatisticamente representativas desta população, qualquer uma das três amostras geradas para um determinado banco poderia ser, então, utilizada para representá-lo. Para evitar um viés decorrente da visualização da distribuição de algumas propriedades das amostras do banco DB, durante a etapa de validação, a escolha da amostra a ser usada foi feita randomicamente, sendo a número 2 escolhida e sempre usada nas análises subsequentes.

4.2.2.2. Distribuição de algumas propriedades físico-químicas e/ou estruturais

Inicialmente, as mesmas propriedades físico-químicas e/ou estruturais utilizadas na etapa de validação da amostragem foram usadas para analisar e comparar os bancos DB, ZNP, ChEMBL, ZINC-2014 e *ZINC-Curated*. Esta comparação foi feita, de modo análogo à etapa de validação da amostragem, através dos correspondentes gráficos de caixa e de porcentagem cumulativa (Figura 4.4 e Figura 4.5, respectivamente).

Comparando os bancos

Nos gráficos de caixa (Figura 4.4) podemos observar que, no geral, as distâncias interquartis dos bancos DB, ZNP e ChEMBL (ou seja, a altura das caixas) são sempre maiores quando comparadas às correspondentes dos bancos ZINC e *ZINC-Curated*. Nos gráficos de caixa, quanto maior a distância interquartil de uma amostra maior é a dispersão dos valores de dada propriedade nesta amostra. Assim, pode-se inferir que os compostos dos bancos DB, ZNP e ChEMBL possuem a maior diversidade nas propriedades físico-químicas e/ou estruturais. Supreendentemente, diferentemente do esperado, embora os valores de mediana observados para o banco ZNP sejam maiores que os do banco DB, a variabilidade do primeiro é igual ou comparável à observada no banco DB. Ainda, as distâncias interquartis do banco *ZINC-Curated* são sempre menores quando comparadas às distâncias dos demais bancos de dados. Estes resultados indicam que o banco *ZINC-Curated* apresenta menor variabilidade destas propriedades quando comparado com os demais, incluindo o banco ZINC-2014. No entanto, os valores de mediana do banco *ZINC-Curated* são idênticos, ou muito próximos, aos valores de mediana do banco DB. Ainda, nestes gráficos quanto maior a sobreposição das caixas mais semelhante é a distribuição de uma determinada propriedade entre os bancos. Interessantemente, comparando a distribuição de cada uma das nove propriedades do banco *ZINC-Curated* com a dos bancos DB e ZINC-2014 podemos observar

que as propriedades dos compostos do banco *ZINC-Curated* estão, no geral, mais próximas daquelas observadas no banco DB do que as observadas no banco ZINC-2014. As únicas exceções são o número de átomos *HBD* (Figura 4.4 (G)) e pelo número de anéis (Figura 4.4 (I)). Como apresentado no item anterior, a distribuição do número de átomos *HBD* da Amostra 2 do banco DB está um pouco diferente da observada na correspondente população. Tendo em vista esta diferença, podemos considerar que a distribuição desta propriedade do banco *ZINC-Curated* também está próxima à observada no banco DB. Pelas mesmas razões, a distribuição das propriedades no banco ZNP se assemelha a distribuição no banco ChEMBL.

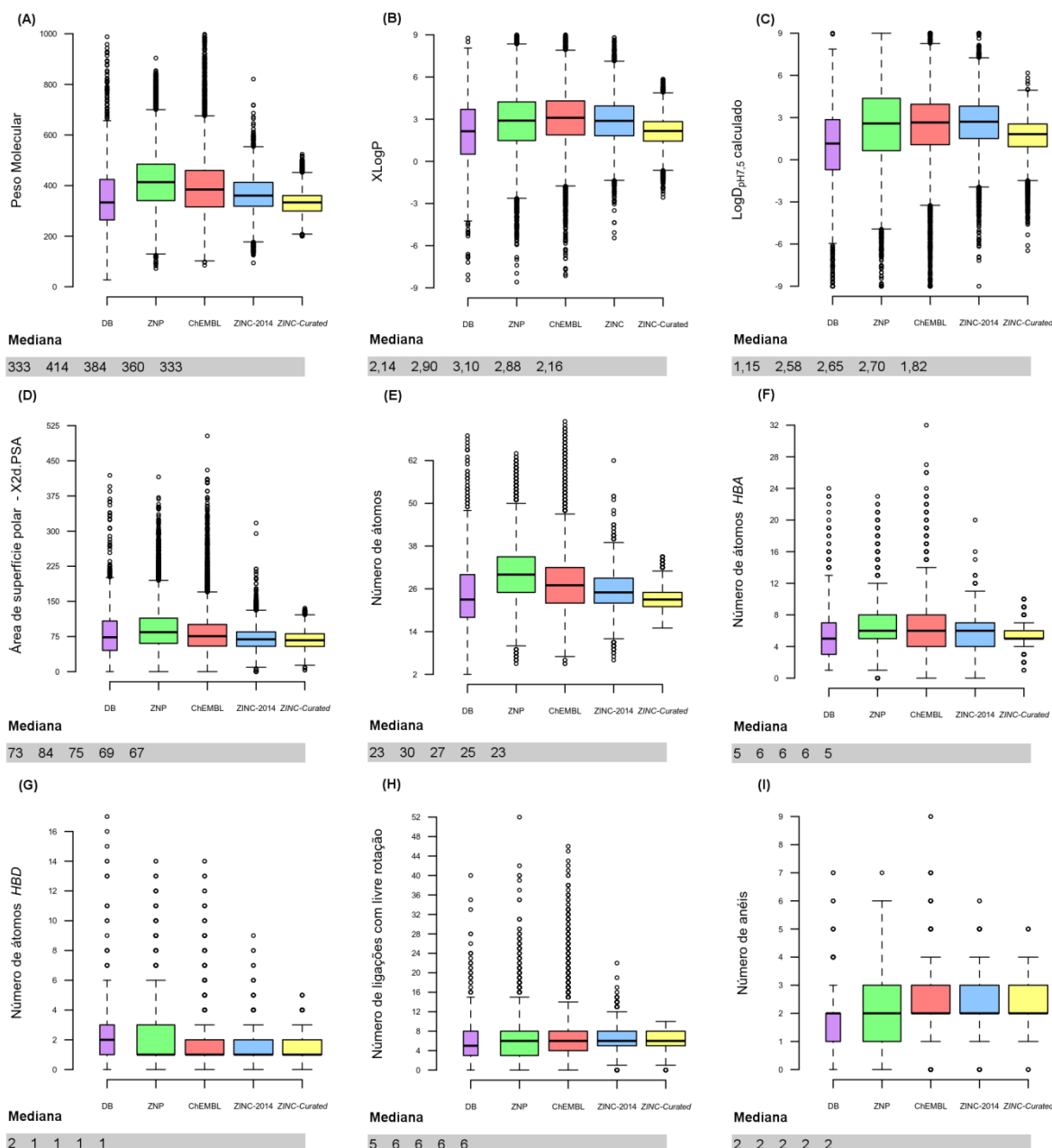


Figura 4.4. Distribuição das propriedades^(a) físico-químicas e/ou estruturais selecionadas para cada um dos bancos de dados. Gráficos de caixa^(b) e valores de mediana para cada uma das propriedades calculadas para cada banco de dados, mostrados sempre, na sequência, DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e ZINC-Curated (amarelo). **(A)** peso molecular; **(B)** XlogP; **(C)** logD_{pH7.5} calculado; **(D)** X2d.PSA; **(E)** número de átomos; **(F)** número de átomos HBA; **(G)** número de átomos HBD; **(H)** número de ligações com livre rotação e **(I)** número de anéis.

(a) Calculadas (item 3.5.3) usando os programas *Volsurf+* (CRUCIANI et al., 2000) e *Filter* ("FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM", [s.d.]).

(b) Gráficos de caixa tipo *Tukey*, para mais detalhes vide a descrição que acompanha a legenda da Figura 4.2.

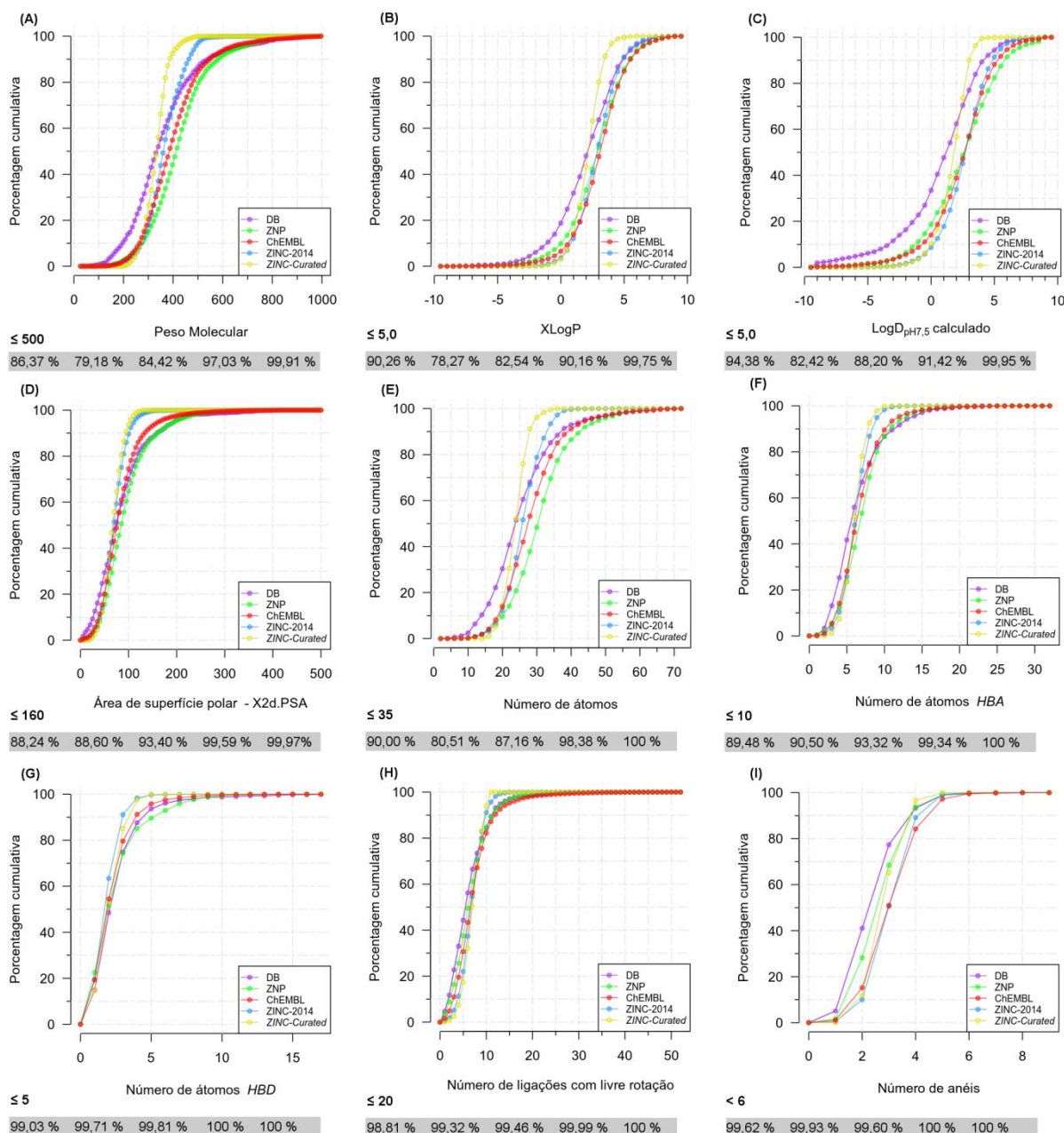


Figura 4.5. Porcentagem acumulativa das propriedades^(a) físico-químicas e/ou estruturais selecionadas para cada um dos bancos de dados. Gráficos da porcentagem cumulativa^(b) e a porcentagem de compostos com valores abaixo dos valores de corte descritos pela “Regra dos 5” de Lipinski e/ou usados no filtro de propriedades aplicado ao banco *ZINC-Curated*, mostrados sempre, na sequência, DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e ZINC-Curated (amarelo). **(A)** peso molecular; **(B)** XlogP; **(C)** logD_{pH7,5} calculado; **(D)** X2d.PSA; **(E)** número de átomos; **(F)** número de átomos HBA; **(G)** número de átomos HBD; **(H)** número de ligações com livre rotação e **(I)** número de anéis.

(a) Calculadas (item 3.5.3) usando os programas *Volsurf+* (CRUCIANI et al., 2000) e *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]).

(b) Gráfico da porcentagem acumulativa, para mais detalhes vide a descrição que acompanha a legenda da Figura 4.3.

Para cada um dos bancos de dados, foram calculados os valores de porcentagem de compostos considerando como corte os mesmos valores aplicados nos filtros usados para construir o banco *ZINC-Curated* (Figura 4.5). Em uma análise preliminar, foi observado que

um número considerável dos compostos apresenta propriedades com valores abaixo daqueles usados como corte, sendo a distribuição de cada uma destas propriedades discutida a seguir. Ainda, a comparação da distribuição de cada propriedade nos bancos foi feita aos pares (dois a dois) através do teste estatístico não paramétrico de *Mann-Whitney U* ($\alpha = 0,05$) aplicando a correção de *Bonferroni*. Este teste assume que os bancos de dados a serem comparados devem ter observações totalmente independentes, ou seja, que um dado composto não pode estar presente em mais de um banco, o que seria pouco provável na nossa análise. Assim, compostos presentes em mais de um banco foram removidos e os parâmetros estatísticos foram recalculados. Porém, semelhantemente ao observado por Wenlock e colaboradores (2003), a remoção dos compostos comuns a mais de um banco (144 compostos, considerando apenas as amostras) não teve efeito significativo sobre as análises feitas. Assim, não consideramos necessário remover os compostos comuns a mais de um banco.

Peso molecular

Os valores de mediana do peso molecular (PM) dos bancos DB e *ZINC-Curated* foram iguais a 333, não sendo observada diferença significativa entre a distribuição desta propriedade nestes dois bancos (teste *Mann-Whitney U*: valor de $p = 0,30$ depois da correção de *Bonferroni*). Já as distribuições do PM dos demais bancos de dados são significativamente diferentes (testes de *Mann-Whitney U*: valores de $p < 0,05$ depois da correção de *Bonferroni*). Estas últimas apresentam valores de mediana maiores do que os observados para os bancos DB e *ZINC-Curated*. É interessante observar que a maioria dos compostos dos bancos DB (86,3%), ZNP (79,2%) e ChEMBL (84,5%) apresentam PM <500 (Figura 4.5(A)), ou seja, compostos não muito grandes. Esta porcentagem é, ainda, maior para os bancos ZINC-2014 (97,1%) e *ZINC-Curated* (99,9%). Além disso, o banco DB é o único banco que possui um número significativo (21,4%) de compostos com baixo PM (<250). Deve-se salientar que o

banco *ZINC-Curated* não possui compostos com $PM < 200$, devido aos valores de corte aplicados no filtro usado quando se construiu este banco ($200 \leq PM \leq 600$). Adicionalmente, foi verificado que o banco ZINC-2014 possui apenas um número muito pequeno (0,4%) de compostos com $PM < 200$. De modo análogo, os bancos DB, ZNP, ChEMBL e ZINC-2014 possuem um número pequeno de compostos com $PM > 600$ (variando entre 1 e 8%). Assim, considerando o uso do PM como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com PM fora da faixa considerada (200 – 600).

XLogP

Os valores de XlogP calculados (no sistema octanol/água, sem incluir correções para ionização) para cada composto foram obtidos como descrito no item 3.5.3, usando o programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]). Considerando os erros associados tanto a determinação experimental de logP bem como a limitação para calcular/prever estes valores, as distribuições dos valores de XlogP serão analisadas nos gráficos de caixa e de porcentagem cumulativa das Figura 4.4 e Figura 4.5, sem considerar seus valores individuais de mediana.

Em geral, os compostos nos bancos DB e *ZINC-Curated* apresentam valores de XlogP menores quando comparados aos dos bancos ZNP, ChEMBL e ZINC-2014, indicando que os compostos dos dois primeiros (DB e *ZINC-Curated*) têm um caráter um pouco menos lipofílico quando comparados com os demais. Nos gráficos de porcentagem cumulativa (Figura 4.5), podemos observar que o banco DB possui 20% de compostos com caráter hidrofílico ($XlogP < 0$), enquanto que para os demais bancos esta é menor que 10%. Nos testes de *Mann-Whitney U* foi observada diferenças estatisticamente significativas nas distribuições dos valores de XlogP para todos os pares de bancos, exceto para DB - *ZINC-Curated* e ZNP - ZINC-2014 (valores de $p = 1$ depois da correção de *Bonferroni*). Ainda, é importante ressaltar

que a grande maioria (entre 92 a 96 %) dos compostos nos bancos DB, ZNP, ChEMBL e ZINC-2014 têm valores de XlogP entre -6 e 6. Assim, considerando o uso do XlogP como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com XlogP fora da faixa considerada ($-6 \leq \text{XlogP} \leq 6$).

LogD_{pH7,5}

Os valores de $\log D_{\text{pH}7,5}$ calculados (no sistema octanol/água, incluído correção(ões) para ionização no pH 7,5) para cada composto foram obtidos como descrito no item 3.5.3, usando o programa *Volsurf+* (CRUCIANI et al., 2000). Como discutido para o XlogP, as distribuições dos valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ serão analisadas graficamente (Figura 4.4 e Figura 4.5), sem considerar seus valores individuais de mediana.

Nos bancos DB e *ZINC-Curated* mais de 75% dos compostos tem valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ menores que 3. A distribuição dos valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ no banco *ZINC-Curated* é a mais próxima da observada para o banco DB. De modo análogo ao descrito para XlogP, o banco de dados DB é o que possui um maior número de compostos com caráter hidrofílico ($\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ calculados < 0). As distribuições dos valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ nos bancos ZNP, ChEMBL e ZINC-2014 são comparáveis, não havendo diferenças estatisticamente significativas entre as distribuições dos bancos ZNP e ChEMBL (teste de *Mann-Whitney U*: valor de $p = 1$ após correção de *Bonferroni*) e entre aquelas dos bancos ZNP e ZINC-2014 (teste de *Mann-Whitney U*: valor de $p = 0,08$ após correção de *Bonferroni*). De modo análogo ao XlogP, é importante ressaltar que a grande maioria (entre 85 e 94 %) dos compostos nos bancos DB, ZNP, ChEMBL e ZINC-2014 têm valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ calculados < 5 . Embora os valores de $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$ não tenham sido diretamente usado como filtro para construir o banco *ZINC-Curated*, estes foram utilizados indiretamente no

filtro de solubilidade¹⁰, na forma de $\log D_{\text{pH}7,4}^{\text{Moka}}$. Isso foi possível, pois $\log D_{\text{pH}7,4}^{\text{Moka}}$ e $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ são calculados de modo análogo sendo, portanto, comparáveis (CRUCIANI et al., 2000; MANNHOLD et al., 2006). O valor de corte no neste filtro de solubilidade foi igual a 5,01. Portanto, apenas compostos com valores de $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ menores do que 5,01 foram mantidos no banco *ZINC-Curated*, independente dos seus números de anéis aromáticos. Assim, considerando o uso indireto de $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ como filtro de solubilidade no banco *ZINC-Curated*, observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ fora do limite considerado (>5,01).

Por fim, de modo geral, para todos os bancos observou-se que os valores de $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ foram menores do que os de XlogP. Além dos erros inerentes, discutidos acima, estas diferenças entre $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ e o XlogP podem ser atribuídas tanto à presença e contribuição de espécies ionizadas para cálculo de $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ quanto às diferenças nas metodologias de cálculo utilizadas pelos dois programas (*Volsurf+* e *Filter*).

Comparação entre os LogP e logD calculados (Volsurf+): contribuição da ionização para partição.

Com o objetivo de verificar se as diferenças observadas $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$ e o XlogP podem, ou não, ser atribuídas à presença e contribuição de espécies ionizadas para cálculo de $\log D_{\text{pH}7,5}^{\text{Volsurf+}}$, valores de logP foram calculados usando o programa *Volsurf+* como descrito no item 3.5.1.2. A Figura 4.6 mostra os gráficos de caixa e de porcentagem cumulativa dos valores de $\log P^{\text{Volsurf+}}$ calculados.

¹⁰ Nesta tese, o *SFI* foi calculado usando os valores de $\log D_{\text{pH}7,4}$ obtidos pelo *Moka* (MILLETTI et al., 2007) e não os obtido pelo *ACDLabs*, como proposto por Hill e Young (2010). Assim, temos que: $SFI = \log D_{\text{pH}7,4}^{\text{Moka}} + \text{número de anéis aromáticos}$.

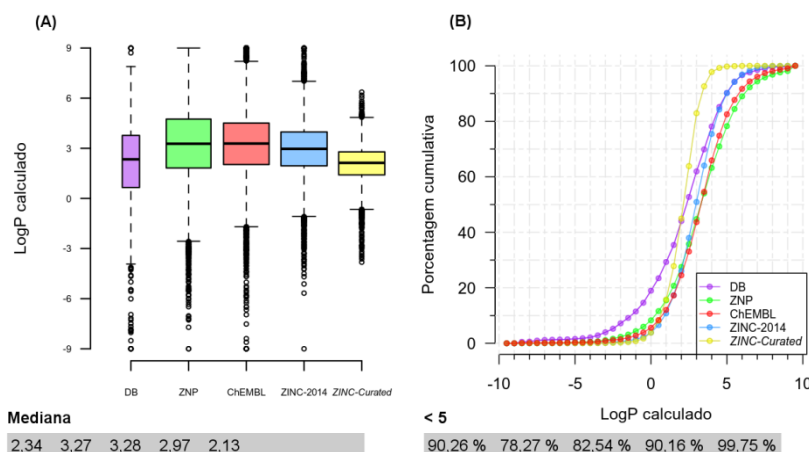


Figura 4.6. Distribuição e porcentagem acumulativa dos valores $\log P^{\text{Volsurf}+}$ para cada um dos bancos de dados. Os valores de $\log P^{\text{Volsurf}+}$ para cada banco de dados estão mostrados sempre na sequência DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e ZINC-Curated (amarelo). (A) Gráfico de caixa^(b) e valores de mediana e (B) gráfico da porcentagem acumulativa^(c) e a porcentagem de compostos com valores $\log P^{\text{Volsurf}+} < 5$.

(a) Calculados (item 3.5.1.2) usando o programa *Volsurf+* (CRUCIANI et al., 2000).

(b) Gráficos de caixa tipo *Tukey*, para mais detalhes vide a descrição que acompanha a legenda da Figura 4.2.

(c) Gráfico da porcentagem acumulativa, para mais detalhes vide a descrição que acompanha a legenda da Figura 4.3.

De modo análogo ao observado para o XlogP, os valores do $\log P^{\text{Volsurf}+}$ calculados são menores que os valores $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$. Sabendo-se ambos foram calculados pelo mesmo programa, as diferenças observadas podem ser atribuídas à presença e contribuição de espécies ionizadas para o cálculo do $\log D_{\text{pH}7,5}^{\text{Volsurf}+}$. Assim, como esperado, podemos concluir que todos os bancos de dados devem possuir compostos com grupamentos ionizáveis que aumentam sua hidrofiliabilidade.

Área de superfície polar (X2d.PSA)

Os valores de área de superfície polar - X2d.PSA calculado para cada composto foram obtidos como descrito no item 3.5.3, usando o programa *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]).

Os bancos ZINC-2014 e ZINC-Curated apresentam os menores valores de mediana de X2d.PSA (69 e 67 Å², respectivamente). Já o banco ZNP é o que apresenta maiores valores de mediana de X2d.PSA (84 Å²) e o maior número de compostos (32%) com valores de

X2d.PSA $>100 \text{ \AA}^2$. Diferentemente do observado para as outras propriedades, o banco ChEMBL ($Q_2 = 75 \text{ \AA}^2$) é o que apresenta a distribuição da X2d.PSA mais próxima à observada para o banco DB ($Q_2 = 73 \text{ \AA}^2$). Não sendo observada diferença estatisticamente significativa entre a distribuição destes dois bancos (teste de *Mann-Whitney U*: valor de $p = 0,35$ após a correção de *Bonferroni*). A maioria (90%) dos compostos dos bancos DB, ZNP, ChEMBL e ZINC-2014 tem valores de X2d.PSA menores do o usado para corte ($X2d.PSA < 160 \text{ \AA}^2$) no filtro aplicado ao banco *ZINC-Curated*. Assim, considerando o uso do X2d.PSA como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com X2d.PSA fora do limite considerado ($>160 \text{ \AA}^2$).

Número de átomos

A distribuição do número de átomos é semelhante à observada para o peso molecular, sendo esta semelhança esperada uma vez que estes descritores são altamente correlacionados. Assim, de modo análogo, não se observou diferença estatisticamente significativa entre as distribuições dos bancos DB e *ZINC-Curated* (teste de *Mann-Whitney U*: valor de $p = 1$ após correção de *Bonferroni*). Por outro lado, as distribuições do peso molecular dos demais bancos de dados são significativamente diferentes (testes de *Mann-Whitney U*: valores de $p < 0,05$ depois da correção de *Bonferroni*). Ainda, o banco DB possui o maior número de compostos (30%) com menos de 20 átomos, enquanto que o banco ZNP é o que possui o maior número de compostos (23%) com mais de 35 átomos. Por fim, em relação ao valor de corte usado como filtro aplicado para construir o banco *ZINC-Curated*, a maioria ($> 77\%$) dos compostos dos bancos DB, ZNP, ChEMBL e ZINC-2014 tem até 35 átomos. De modo análogo, pelo menos 82% dos compostos dos bancos ZNP, ChEMBL e ZINC-2014 têm mais de 20 átomos. Assim, considerando o uso do o número de átomos como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com o número de átomos fora do limite considerado (20 – 35).

Número de átomos aceptores e doadores de ligação de hidrogênio

Os números de átomos *HBA* e *HBD* para cada composto foram obtidos como descrito no item 3.5.3, considerando a definição de Lipinski (LIPINSKI et al., 1997).

Os valores de mediana do número de átomos *HBD* dos cinco bancos são comparáveis (variando entre 5 e 6 átomos). No entanto, todas as distribuições mostraram diferenças estatisticamente significativas no teste de *Mann-Whitney U* (valores de $p < 0,05$ após a correção de *Bonferroni*). Ainda, o número de compostos com menos de 10 átomos *HBD* presentes nos bancos DB, ZNP e ChEMBL é alto e comparável (87, 87 e 89 %, respectivamente). Para o banco ZINC-2014, este número é ainda maior (98%), sendo comparável com o do banco *ZINC-Curated* (99%). Já em relação ao número de átomos *HBA* todos os bancos apresentaram valores de mediana iguais a 1, exceto o banco DB ($Q_2 = 2$). Embora, tenha-se constatado esta diferença no DB, não há diferença estatisticamente significativa entre a distribuição do número de *HBA* deste banco com a distribuição do banco ZNP (teste de *Mann-Whitney U*: valor de $p = 0,07$ após correção de *Bonferroni*). Interessantemente, não há diferença significativa entre a distribuição do número de *HBA* dos bancos *ZINC-Curated* e ChEMBL (teste *Mann-Whitney U*: valor de $p = 1$ após correção de *Bonferroni*) enquanto que aquelas nos bancos *ZINC-Curated* e ZINC-2014 são significativamente diferentes (teste *Mann Whitney U*: valor de $p < 0,05$ após correção de *Bonferroni*). Por fim, pelo menos 89% dos compostos presentes em todos os bancos apresentam número de átomos *HDA* < 10 . Assim, considerando o uso dos números de átomos *HBD* e *HBA* como filtros, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com números de átomos fora dos limites considerados (>5 e >10 , respectivamente).

Número de ligações com livre rotação

O número de ligações com livre rotação para cada composto foi obtido como descrito no item 3.5.3, usando o *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]).

Os valores de mediana do número de ligações com livre rotação de todos os bancos são comparáveis (5 ou 6 ligações). No entanto, todas as distribuições do número de ligações com livre rotação mostraram diferenças estatisticamente significativas no teste de *Mann-Whitney U* (valores de $p < 0,05$ após a correção de *Bonferroni*), exceto para aquelas dos bancos ChEMBL e ZINC-2014 (teste de *Mann-Whitney U*: valor de $p = 0,08$ após a correção de *Bonferroni*). Ainda, quase todos os compostos (de 99 a 100%) dos bancos DB, ZNP, ChEMBL e ZINC-2014 têm até 20 ligações com livre rotação, sendo este o valor de corte usado como filtro aplicado para construir o banco *ZINC-Curated*. Assim, considerando o uso do número de ligações com livre rotação como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com número de ligações com livre rotação fora do limite considerado (>20).

Número de anéis

O número de anéis para cada composto foi obtido como descrito no item 3.5.3, usando o *Filter* (“FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM”, [s.d.]).

Todos os bancos tem valor de mediana relativo ao número de anéis igual a 2. Os compostos do banco DB têm um número menor de anéis (25% com até 1 anel e 50% com 2 anéis). Já no banco ZNP, 50% dos compostos têm de 1 a 3 anéis. Os compostos dos bancos ChEMBL, ZINC-2014 e *ZINC-Curated* por sua vez têm de 2 a 3 anéis. No teste de *Mann-Whitney U*, todas as distribuições do número de anéis mostraram diferenças estatisticamente significativas (valores de $p < 0,05$ após a correção de *Bonferroni*), exceto aquelas nos bancos

ChEMBL e ZINC-2014 (valor de $p = 1$ após a correção de *Bonferroni*). Por fim, menos de 1% dos compostos presentes em todos os bancos possuem mais de 5 anéis, sendo este o valor corte usado como filtro aplicado para construir o banco *ZINC-Curated*. Assim, considerando o uso do número de anéis como filtro, no banco *ZINC-Curated* observou-se uma menor variabilidade, sem que houvesse perda significativa de compostos com número de anéis fora do limite considerado (>5).

Análise de alguns valores de cortes usados para construir o banco ZINC-Curated

Esta análise foi feita determinando-se, inicialmente, para cada banco de dados os valores de algumas propriedades (discutidas acima) abaixo dos quais são encontrados 90% dos compostos em cada banco. Estes foram comparados com aqueles usados como corte nos filtros aplicados para construir o banco *ZINC-Curated* (Tabela 4.5).

Tabela 4.5. Valores das propriedades físico-químicas e/ou estruturais abaixo dos quais são encontrados 90 % dos compostos de cada banco de dados, acompanhados dos valores de cortes usados como filtros para construir o banco *ZINC-Curated*.

Propriedades	Valor de corte	Valor de cada propriedade abaixo do qual são encontrados 90 % dos compostos				
		DB	ZNP	ChEMBL	ZINC-2014	<i>ZINC-Curated</i>
Peso Molecular	600	540	583	548	463	386
XLogP	6	4,87	5,44	5,49	4,93	3,41
LogD _{pH7,5} ^{Volsurf+}		4,11	5,79	5,23	4,83	2,99
Átomos <i>HBD</i>	5	4	5	3	2	3
Átomos <i>HBA</i>	10	11	10	10	8	7
X2d.PSA	160	159	160	133	101	94
Número de átomos	35	36	42*	39	32	27
Ligações com livre rotação	20	11	11	11	9	9
Número de anéis	5	3	3	4	4	3

* Valor acima do valor de corte usado.

No geral, observamos na Tabela 4.5 que para cada um dos filtros usado 90 % dos compostos, em cada um dos bancos, possuem valores próximos, ou abaixo, dos aqueles usados como corte. A única exceção é o número de átomos do banco ZNP, indicado Tabela 4.5 com um asterisco. No entanto, isto não significa, necessariamente, que 90% dos compostos presentes em cada um dos bancos sejam capazes de respeitar, simultaneamente, todos os limites/faixas usados como cortes. Pela análise individual dos valores de cortes usados para construir o banco *ZINC-Curated* pode-se sugerir que se estes fossem aplicados aos demais bancos não haveria uma perda significativa de compostos.

Violação da Regra dos 5 de Lipinski (VR5L)

Além dos filtros descritos acima, outro filtro utilizado foi o número de violações à Regra dos 5 de Lipinski (LIPINSKI et al., 2012), denominado nesta tese *VR5L*. Aqui, consideramos como sendo permitida até duas violações à regra (≤ 2), ao invés de uma única, como proposta originalmente por Lipinski ($VR5L \leq 1$, (LIPINSKI et al., 1997)). Além disso, nesta tese, o XlogP foi usado no lugar do ClogP, proposto originalmente por Lipinski (LIPINSKI et al., 1997).

De modo análogo ao observado para os outros filtros, 90% dos compostos presentes em todos os bancos, inclusive no banco de produtos naturais (ZNP), estão abaixo do valor de corte usado ($VR5L \leq 2$). Assim, considerando o uso do número de *VR5L* como filtro, no banco *ZINC-Curated* observou-se que não houve perda significativa de compostos fora do limite usado (> 2).

Adicionalmente, visando verificar se os compostos de cada um dos bancos de dados seguiam, ou não, a regra original de Lipinski (LIPINSKI et al., 1997), estes compostos foram contados e agrupados considerando nenhuma, uma, duas, três ou quatro violações. Os resultados estão apresentados Tabela 4.6. Nesta tese, todos os compostos de todos os bancos,

inclusive o DB (fármacos), foram considerados e contados independentemente das suas biodisponibilidades orais como classificados por Lipinski (LIPINSKI et al., 1997).

Tabela 4.6. Números de compostos (em %) e de violações à Regra dos 5 de Lipinski (LIPINSKI et al., 1997) encontrados em cada um dos cinco bancos de dados analisados.

Banco de dados	Números de compostos (em %) e de violações encontrados				
	0	1	2	3	4
DB	77,09	11,79	7,76	3,28	0,08
ZNP	67,77	17,09	10,48	4,60	0,06
ChEMBL	72,46	17,08	8,18	2,11	0,17
ZINC-2014	88,86	9,59	1,48	0,07	0,00
<i>ZINC-Curated</i>	99,47	0,52	0,01	0,00	0,00

Como pode ser observado na Tabela 4.6, a maioria ($\geq 68\%$) dos compostos em todos os bancos não viola qualquer um dos critérios da Regra dos 5 de Lipinski (LIPINSKI et al., 2012), ou seja, tem peso molecular ≤ 500 , $XlogP \leq 5$, número de átomos $HBD \leq 5$ e número de átomos $HBA \leq 10$. Ainda, observa-se que mais de 85% dos compostos destes bancos dados seguem a Regra original de Lipinski ($VR5L \leq 1$). Isso significa que a maioria dos compostos analisados nestes bancos – contendo fármacos (DB), produtos naturais (ZNP), compostos avaliados experimentalmente quanto a sua atividade biológica (ChEMBL) e compostos disponíveis comercialmente (ZINC-2014) – estão no “espaço químico de Lipinski”, embora não se tenha feito distinção quanto às suas biodisponibilidades via oral.

Dentre estes resultados, é surpreendente que os produtos naturais do banco ZNP tenham este comportamento, pois estes são considerados, por alguns, exceções da Regra de Lipinski (LIPINSKI, 2016; LIPINSKI et al., 1997; NEWMAN; CRAGG, 2012). Adicionalmente, deve-se lembrar que o banco ZNP, contendo somente 146.882 estruturas, pode não estar, na verdade, representando toda a diversidade esperada para este conjunto de compostos. Assim, os resultados obtidos para o ZNP podem ser decorrentes do uso de um banco de dados não representativo. Na literatura são encontrados relatos corroborando nossos

resultados. Em relação aos fármacos (DB) e aos compostos avaliados experimentalmente quanto a sua atividade biológica (ChEMBL), uma possível explicação para os resultados observados é que a Regra dos 5 de Lipinski, embora por vezes criticada na literatura (DOAK et al., 2016; SCIOR et al., 2012), já foi incorporada para a maioria dos protocolos no desenvolvimento de fármacos. Em relação aos compostos disponíveis comercialmente (ZINC-2014), os resultados observados poderiam ser devido a interesses puramente comerciais ou às limitações sintéticas.

4.2.2.3. Espaço químico

O espaço químico (espaço multidimensional ocupado por todos os compostos químicos) pode ser analisado de várias maneiras. Nesta tese, a comparação do espaço químico foi feita pela Análise de Componentes Principais usando 128 descritores moleculares gerados pelo programa *Volsurf+* (CRUCIANI et al., 2000). Procedendo-se como descrito no item 3.5.3 (Materiais e Métodos), os descritores moleculares foram submetidos a um pré-tratamento. No pré-tratamento, dos 128 descritores gerados, 60 foram selecionados. A seguir, 2 descritores moleculares que não forneceram informações relevantes sobre o sistema (ou seja, descritores com valores constantes) foram excluídos. Esta etapa foi feita visando a diminuir o número de dados a serem tratados e o ruído da amostra (WOLD, 1997). Os 58 descritores (para a lista completa vide Anexo II) restantes foram auto escalonados. Este auto escalonamento removeu qualquer ponderação decorrente, unicamente, das unidades usadas para expressar estes descritores. Após este pré-tratamento, estes 58 descritores moleculares foram usados na Análise de Componentes Principais (ACP) feita como descrito no item 3.5.3. Os resultados foram analisados através do gráfico de escores (Figura 4.7), usando somente a primeira e a segunda componente principal (CP1 e CP2, respectivamente). A CP1 e a CP2 explicam 30 e 23 % da variação dos dados, respectivamente.

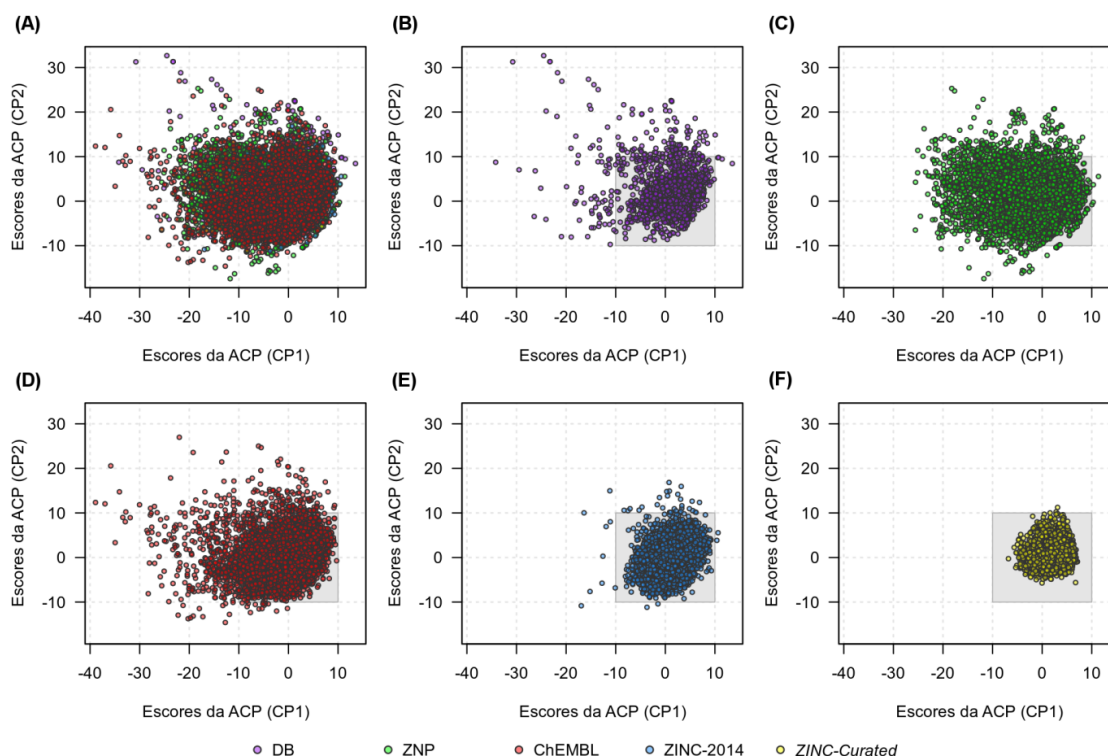


Figura 4.7. Espaço químico dos compostos presentes nos bancos de dados ZINC-Curated, ZINC-2014, DB, ZNP e ChEMBL. O gráfico dos escores^(a) obtidos na ACP^(b) para CP1 e CP2 representam o espaço químico dos compostos presentes nos bancos DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e ZINC-Curated (amarelo). (A) Gráfico dos compostos de todos os bancos. Gráfico dos compostos de cada um dos bancos: (B) DB, (C) ZNP, (D) ChEMBL, (E) ZINC-2014 e (F) ZINC-Curated. A área do retângulo em cinza (CP1 e CP2 entre -10 e 10) mostra a região com maior concentração de compostos, correspondendo a 8 % do DB, 89% do ZNP, 94% do ChEMBL, 99% do ZINC-2014 e 99% do ZINC-Curated.

^(a) Neste gráfico cada ponto representa um composto no eixo de coordenadas de CP1 e CP2.

^(b) Análise de Componentes Principais foi realizada utilizando 58 descritores do programa *Volsurf+* (CRUCIANI et al., 2000) e usando a função *prcomp* do programa *R* (R CORE TEAM, 2013).

Na Figura 4.7(A) podemos observar que há uma sobreposição entre os espaços químicos dos diferentes bancos de dados e que os compostos estão concentrados numa região próxima à origem (CP1 e CP2 entre -10 e 10). Esta região central do espaço químico, comum a todos os bancos, foi demarcada em cinza. A seguir, a porcentagem de compostos de cada banco dentro desta região foi calculada. Os bancos DB, ZNP e ChEMBL apresentam um maior número de compostos fora da “região comum” quando comparados aos bancos ZINC-2014 e ZINC-Curated. Interessantemente, o banco DB é o que possui uma porcentagem maior de compostos fora desta região comum (~ 15%) seguido pelo banco ZNP (~ 11%). Estes resultados indicam que os bancos DB e ZNP possuem uma maior diversidade química, dentre

os bancos analisados. Ainda, pode-se observar que os compostos do banco *ZINC-Curated* estão mais concentrados que os do banco *ZINC-2014*, indicando uma diminuição da diversidade química (Figura 4.8 (A)).

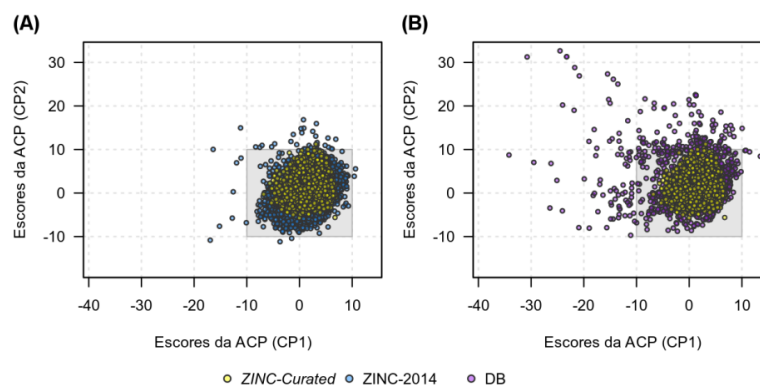


Figura 4.8. Espaço químico dos compostos presentes nos bancos de dados *ZINC-Curated* sobreposto aos bancos *ZINC-2014* e *DB*. O gráfico dos escores^(a) obtidos na ACP^(b) para CP1 e CP2 representam o espaço químico dos compostos presentes nos bancos *ZINC-Curated* (amarelo), *ZINC-2014* (azul) e *DB* (roxo). Gráfico dos compostos do banco *ZINC-Curated* sobrepostos aos compostos do banco *ZINC-2014* (A) e os do banco *DB* (B). A área do retângulo em cinza mostra a região com maior concentração de compostos, correspondendo a 85% do *DB*, 99% do *ZINC-2014* e 99% do *ZINC-Curated*.

^(a) Neste gráfico cada ponto representa um composto no eixo de coordenadas de CP1 e CP2.

^(b) Análise de Componentes Principais foi realizada utilizando 58 descritores do programa *Volsurf+* (CRUCIANI et al., 2000) e usando a função *prcomp* do programa *R* (R CORE TEAM, 2013).

Além do gráfico de escores, a ACP nos fornece o gráfico de pesos que mostra a contribuição de cada descritor para CP1 e CP2 (Figura 4.9). Neste gráfico podemos observar que descritores moleculares relacionados a uma alta hidrofiliicidade contribuem positivamente para CP2, enquanto que descritores relacionados a uma alta hidrofobicidade contribuem negativamente para CP2. Ainda, descritores relacionados a um grande volume e/ou peso molecular contribuem negativamente para CP1. Considerando a contribuição dos descritores, este gráfico foi dividido em quatro quadrantes (denominados Q-I a Q-IV), a saber: Q-I contendo descritores que contribuem positivamente para CP1 e CP2 (ex. MetStab, SOLY, LgS7.5 e DIFF); Q-II contendo descritores que contribuem positivamente para CP1 e negativamente para CP2 (ex. CACO2, LgBB e SKIN); Q-III contendo descritores que contribuem negativamente para CP1 e negativamente para CP2 (ex. HSA, V, S e DRDRDR) e, Q-IV

contendo descritores que contribuem negativamente para CP1 e positivamente para CP2 (ex. PSAA, PSAR e PHSAR).

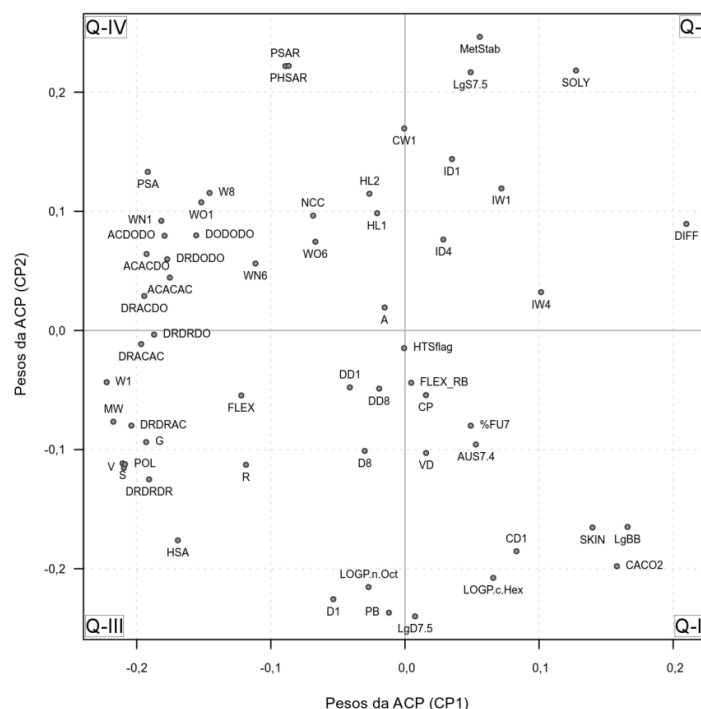


Figura 4.9. Contribuição de cada um dos descritores para as Componentes Principais (CP1 e CP2). O gráfico dos pesos obtidos na ACP^(a) para CP1 e CP2 mostra a contribuição de cada descritor para estas componentes principais. Cada ponto representa um dos 58 descritores calculados pelo programa *Volsurf+* (CRUCIANI et al., 2000). As linhas que cortam os eixos CP1 e CP2 na origem (CP1 = CP2 = 0) dividem o gráfico em quatro quadrantes (denominados Q-I a Q-IV).

(a) Análise de Componentes Principais foi realizada utilizando 58 descritores do programa *Volsurf+* (CRUCIANI et al., 2000) e usando a função *prcomp* do programa *R* (R CORE TEAM, 2013).

De modo análogo, o gráfico de escores foi dividido em quatro quadrantes (Q-I a Q-IV) considerando os valores de escore de CP1 e de CP2 (Q-I: CP1 e CP2 >0; Q-II: CP1 >0 e CP2 <0; Q-III: CP1 e CP2 <0 e Q-IV: CP1 <0 e CP2 >0). Considerando a distribuição dos descritores moleculares no gráfico de pesos da ACP (Figura 4.9), podemos inferir sobre algumas propriedades físico-químicas (calculadas) e/ou estruturais dos compostos presentes nos bancos, de acordo com o quadrante que estes se encontram. Assim, compostos presentes nos quadrantes Q-I e Q-IV são mais hidrofílicos, enquanto que compostos presentes nos quadrantes Q-II e Q-III são mais lipofílicos. Adicionalmente, compostos presentes nos quadrantes Q-III e Q-IV são grandes, ou seja, volumosos e/ou com alto peso molecular. A

seguir, o número de compostos (em %) de cada banco de dados nos quatro quadrantes foi calculado e está mostrado na Tabela 4.7.

Tabela 4.7. Número de compostos (em %) localizados em cada um dos quatro quadrantes (Q-I, Q-II, Q-III e Q-IV) do gráfico dos escores obtidos na ACP para cada um dos bancos de dados.

Banco de Dados	Número de compostos (em %) em cada quadrante			
	Q-I	Q-II	Q-III	Q-IV
DB	42,02	20,37	15,21	22,40
ZNP	18,00	21,34	32,11	28,55
ChEMBL	24,42	26,47	29,01	16,09
ZINC-2014	28,69	36,01	26,28	9,02
<i>ZINC-Curated</i>	42,58	38,68	8,68	10,06

É interessante observar que a maior parte dos compostos do banco DB (42 %) se encontra no quadrante Q-I (compostos hidrofílicos pouco volumosos e/ou com baixo peso molecular). Um perfil parecido pode ser observado para o banco *ZINC-Curated*, contendo 42 % dos compostos no quadrante Q-I. Estes resultados corroboram com os já observados nas análises apresentadas acima, usando o peso molecular e os valores de XlogP. Os demais compostos do banco DB se distribuem, principalmente, entre os quadrantes Q-II (compostos lipofílicos pequenos, ou seja, pouco volumosos e/ou com baixo peso molecular) e Q-IV (compostos hidrofílicos grandes). Já no banco *ZINC-Curated* os demais compostos se encontram, principalmente, no quadrante Q-II (38,68 %). Comparando a distribuição dos compostos do banco *ZINC-Curated* com a distribuição dos compostos do banco ZINC-2014 podemos observar que este último apresenta compostos mais lipofílicos concentrados, principalmente, no quadrante Q-II (36 %). Desta forma, o banco *ZINC-Curated* possui compostos com características mais próximas às observadas para os compostos do banco DB do que as observadas para os compostos do banco ZINC-2014. Esta observação pode ser confirmada analisando-se a sobreposição do espaço químico ocupado pelos compostos presentes nos bancos DB e *ZINC-Curated* (Figura 4.8 (B)). Adicionalmente, os compostos do

banco ZNP são grandes, apresentando tanto caráter lipofílico (32,11 % no Q-III) bem como hidrofílico (28,55 % no Q-IV). Por fim, no banco ChEMBL estão presentes, no geral, compostos lipofílicos (grandes e pequenos) bem com compostos hidrofílicos pequenos.

Em geral, os resultados mostram que embora a aplicação dos filtros ao banco de dados *ZINC-Curated* tenha acarretado numa diminuição da diversidade química, este banco foi enriquecido com compostos com características próximas às encontradas em fármacos, representados, nesta análise, pelo banco DB. Dentre estas, cita-se uma maior solubilidade em água bem como uma maior estabilidade metabólica, sendo ambas as características representadas por descritores presentes no quadrante Q-I (SOLY e MetStab, respectivamente). Considerando a importância destas características, gráficos de caixa foram gerados para estes dois descritores e estão apresentados na Figura 4.10. Sendo cada um destes descritores apresentados e analisados separadamente nos itens 4.2.2.4 e 4.2.2.5, respectivamente.

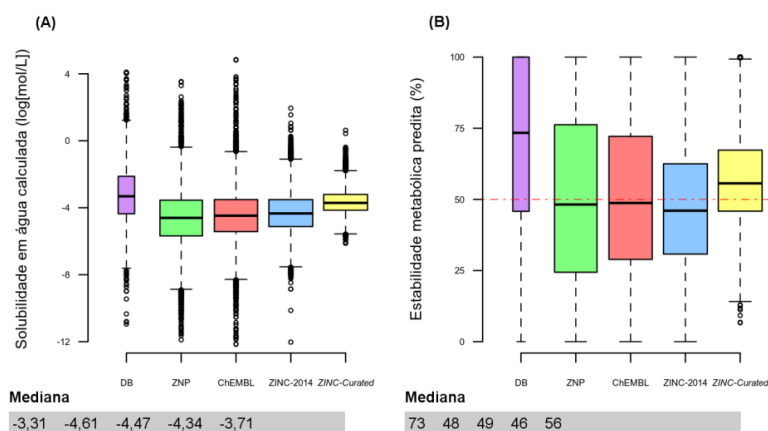


Figura 4.10. Distribuição da solubilidade em água calculada e da estabilidade metabólica predita para cada um dos bancos de dados. Gráficos de caixa^(a) e valores de mediana das duas propriedades calculadas para cada um dos bancos de dados, mostrados sempre na sequência DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e *ZINC-Curated* (amarelo): (A) solubilidade em água calculada (log[mol/L], a 25 °C)^(b) e (B) estabilidade metabólica frente à CYP3A4 humana predita (%)^(c).

(a) Gráficos de caixa tipo *Tukey*, para mais detalhes vide a descrição que acompanha a legenda da Figura 4.2.

(b) Calculada pelo programa *Volsurf+* (CRUCIANI et al., 2000) usando um modelo de *PLS* (3 variáveis latentes, $R^2 = 0,74$; $q^2 = 0,73$; SDEC = 0,89 e $n = 1028$) com um erro de predição de 0,7 log (MANNHOLD et al., 2006).

(c) Predita pelo programa *Volsurf+* (CRUCIANI et al., 2000) usando um modelo qualitativo (2 variáveis latentes, $R^2 = 0,44$; $q^2 = 0,43$; SDEC = 0,74 e $n = 1507$) no qual compostos com valores de MetStab % >50 são preditos como estáveis e compostos com valores de MetStab % <50 como instáveis (MANNHOLD et al., 2006).

4.2.2.4. Análise da solubilidade em água predita

A predição da solubilidade em água é extremamente difícil e não é, por nós, conhecido um modelo universal capaz de prever de forma eficiente a solubilidade de compostos quimicamente diversos. Para a solubilidade calculada usando o descritor do *Volsurf+*, o valor do erro de predição descrito por Mannhold e colaboradores é igual 0,7 log (MANNHOLD et al., 2006). Contudo, estes autores defendem o uso desse modelo para classificar um conjunto de compostos desde que estes compostos não sejam muito similares. Tendo em vista que os compostos presentes nos bancos não são muito similares (conforme será apresentado no próximo item), o modelo de solubilidade do *Volsurf+* pode ser usado para classificá-los. Desta forma, apenas a distribuição dos valores preditos de solubilidade mostrados na Figura 4.10 serão analisados. Nestes gráficos de caixa, podemos observar que o banco DB é o que possui um maior número de compostos com valores de solubilidade >-4 log (mol/L), indicando que este é o banco com o maior número de compostos solúveis. A distribuição dos valores de solubilidade preditos do banco *ZINC-Curated* é a que mais se aproxima da observada para o banco DB. Ainda, o fato dos valores de solubilidade preditos serem menores no banco ZINC-2014 quando comparados aos do banco *ZINC-Curated*, sugere que o filtro de solubilidade (*SFI*) aplicado ao último o enriqueceu com compostos solúveis.

Neste contexto, optamos por analisar a porcentagem de compostos comprovadamente insolúveis, classificados corretamente como tendo uma baixa solubilidade usando o *SFI* (HILL; YOUNG, 2010). Com este objetivo, calculamos o valor do *SFI* de cada composto do banco DB de modo análogo ao descrito na construção do banco *ZINC-Curated* (vide item 3.5.3). Dos compostos presentes neste banco, 21,44% apresentaram valores de *SFI* $>5,01$. Para cada um destes compostos (*SFI* $>5,01$), os valores de solubilidade experimentalmente determinados (em água) foram, quando disponíveis, retirados da literatura (KIM et al., 2016; O'NEIL, 2006; "U.S. Food and Drug Administration", 2017, "USP - United States

Pharmacopeial Convention”, 2017; WISHART, 2006). Ainda, para os compostos cujos valores de solubilidade experimental não estavam disponíveis, estes foram “estimados” considerando os valores calculados pelo programa *ALOGPS* (KIM et al., 2016), uma vez que os valores estavam disponíveis no website do banco DrugBank (WISHART, 2006). Em seguida, estes valores foram usados para dividir os compostos em três categorias, seguindo a classificação apresentada por Hill e Young (HILL; YOUNG, 2010). Sendo estas: compostos com baixa solubilidade ($<30 \mu\text{M}$), compostos com solubilidade intermediária ($30 - 200 \mu\text{M}$) e compostos com alta solubilidade ($>200 \mu\text{M}$). O gráfico de setores apresentado na Figura 4.11 mostra as porcentagens de compostos com baixa solubilidade, solubilidade intermediária e alta solubilidade diferenciando, os valores experimentais retirados da literatura dos preditos usando o programa *ALOGPS* (KIM et al., 2016).

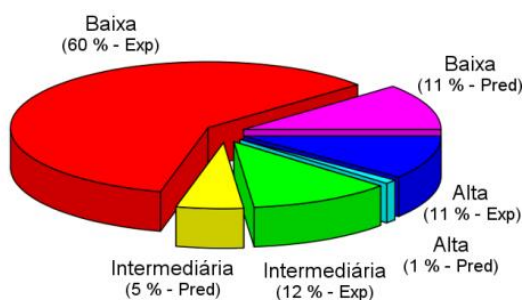


Figura 4.11. Solubilidade em água experimentalmente determinada e predita para os compostos do banco DB que apresentaram valores de $SFI^{(a)}$ $> 5,01$. Gráficos de setor com os correspondentes valores de porcentagem seguindo-se a classificação apresentada por Hill e Young^(b), mostrando a porcentagem de compostos com: baixa solubilidade, experimental^(c) (vermelho) e predita^(d) (magenta); alta solubilidade, experimental (azul) e predita (ciano), e solubilidade intermediária, experimental (verde) e predita (amarelo).

(a) Valores de SFI (HILL; YOUNG, 2010) foram calculados como descrito no item 3.5.3.

(b) Segundo a classificação de Hill e Young (2010), os compostos foram divididos em três categorias de acordo com sua solubilidade: alta ($> 200 \mu\text{M}$), intermediária ($30 - 200 \mu\text{M}$) e baixa ($< 30 \mu\text{M}$).

(c) Solubilidade experimentalmente determinada obtida da literatura (KIM et al., 2016; O'NEIL, 2006; "U.S. Food and Drug Administration", 2017, "USP - United States Pharmacopeial Convention", 2017; WISHART, 2006).

(d) Solubilidade predita usando os valores do programa *ALOGPS* (KIM et al., 2016), disponíveis no website do banco DrugBank (WISHART, 2006).

Neste gráfico podemos observar que a maioria (71%) dos compostos realmente apresentam uma baixa solubilidade, como predito. Além disso, a porcentagem de compostos

com alta solubilidade, preditos como insolúveis (falsos negativos), é consideravelmente pequena (12%). Estas porcentagens são comparáveis com as, por nós, estimadas considerando os dados apresentados na Figura 9 do artigo de Hill e Young (2010). Para estes dados, estimamos que 54, 29 e 17% dos 36.610 compostos com valores de *SFI* >5 teriam solubilidade baixa, intermediária e alta, respectivamente. Assim, a porcentagem esperada de compostos preditos corretamente como insolúveis, usando o modelo de *SFI*, está entre 54 e 71%. Baseado nisto, pode-se considerar que o aumento do número de compostos com valores mais altos de solubilidade, segundo o modelo do *Volsurf+* (Figura 4.10), realmente indica que o banco de dados *ZINC-Curated* foi enriquecido com compostos mais solúveis.

4.2.2.5. Análise da estabilidade metabólica predita

Apresentar uma certa estabilidade metabólica é uma característica desejável para a grande maioria dos fármacos (KIRCHMAIR et al., 2015). Nesta análise, a estabilidade frente à CYP3A4 foi inferida usando o descritor MetStab (MANNHOLD et al., 2006) calculado pelo programa *Volsurf+* (CRUCIANI et al., 2000). Segundo os autores, este modelo deve ser considerado de forma qualitativa. Assim, compostos com valores acima de 50 são considerados estáveis, enquanto que, compostos com valores abaixo de 50 são tidos como instáveis (MANNHOLD et al., 2006). No gráfico (B) da Figura 4.10, é interessante observar que apenas 29% dos compostos presentes no banco DB são preditos como instáveis frente à CYP3A4 (ou seja, valores menores que 50). Uma porcentagem comparável de compostos com valores menores que 50 foi observada para o banco *ZINC-Curated* (35%). Além disso, esta porcentagem é muito menor do que a observada para o banco *ZINC-2014* (56%), indicando que a remoção de grupos funcionais reativos enriqueceu o banco *ZINC-Curated* com compostos mais estáveis.

4.2.2.6. Diversidade estrutural

Os valores do coeficiente de Tanimoto (T_c), calculados a partir de *fingerprints* moleculares (WILLETT, 2011), podem ser usados como filtros de seleção na busca por similaridade 2D. Assim, nesta tese, a diversidade estrutural (dentro de um mesmo banco) e a similaridade (entre diferentes bancos) foram analisadas usando valores de T_c , calculados como descrito no item 3.5.3. A diversidade estrutural de cada banco foi analisada considerando-se, inicialmente, valores de T_c calculados para todos os possíveis pares de compostos presentes dentro do correspondente banco (gráficos (A) e (B) da Figura 4.12).

Nos gráficos A e B (Figura 4.12), é interessante observar que os compostos presentes no banco DB são os mais dissimilares entre si (69% dos pares de compostos apresentam valores $T_c < 0,2$). Os compostos disponíveis no banco ChEMBL são um pouco mais dissimilares entre si (58% dos pares de compostos tem valores $T_c < 0,2$) que os compostos presentes no banco ZNP (53% dos pares de compostos têm valores de $T_c < 0,2$). Estes resultados não seriam esperados, uma vez que os produtos naturais são reconhecidos na literatura como um conjunto de compostos com grande diversidade química/estrutural (HARVEY; EDRADA-EBEL; QUINN, 2015). Dentre as possíveis explicações para nossos achados, citam-se que: (i) o *fingerprint* molecular usado não ser capaz de distinguir isômeros enantioméricos (ou seja, isômeros R/S, bastante frequente em produtos naturais (FEHER; SCHMIDT, 2003)) e (ii) para alguns pares de compostos com valor de coeficiente de $T_c = 1$, um dos compostos pode ser uma subestrutura do outro, pois cada fragmento é contado uma única vez (ou seja, dois compostos de diferentes tamanhos com os mesmos fragmentos terão $T_c = 1$, mesmo que um deles tenha o dobro de fragmentos que o outro). Adicionalmente, os compostos do banco *ZINC-Curated* são um pouco menos dissimilares que os do banco ZINC-2014, indicando que a remoção de grupos funcionais acarretou numa certa perda de diversidade estrutural.

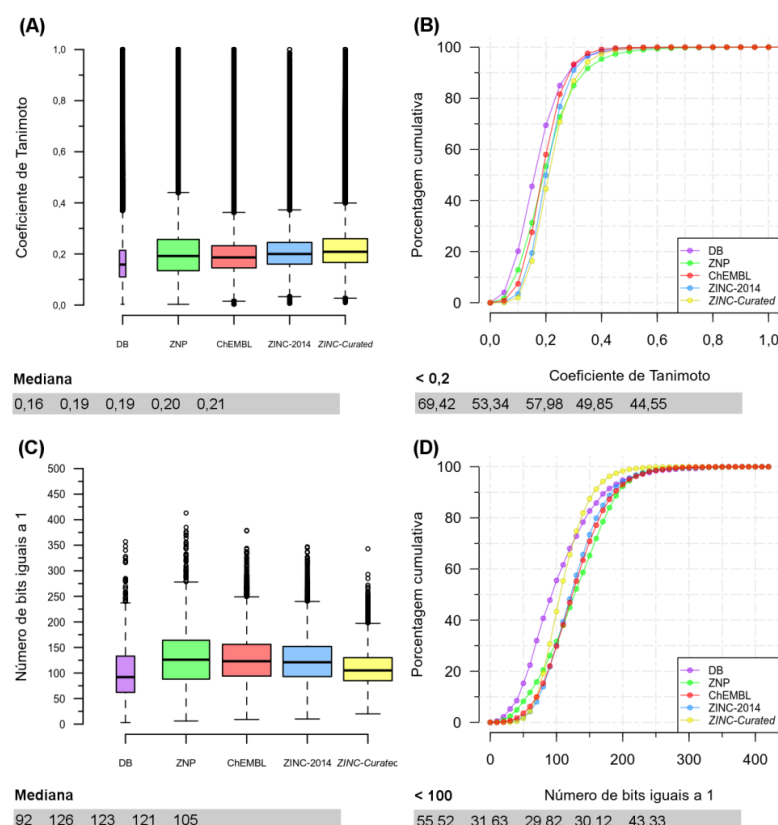


Figura 4.12. Distribuição dos valores do $T_c^{(a)}$ e do número de bits para cada banco de dados. Gráficos de caixa^(b) e valores de mediana (esquerda) e gráficos de porcentagem cumulativa^(b) (direita) para cada um dos bancos, mostrados sempre na sequência: DB (roxo), ZNP (verde), ChEMBL (vermelho), ZINC-2014 (azul) e ZINC-Curated (amarelo): (A) e (B) distribuição dos valores do $T_c^{(a)}$ para todos os pares de compostos presentes dentro de um mesmo banco e (C) e (D) número de *bits* iguais a 1 em cada um dos compostos.

(a) Valores do coeficiente de Tanimoto (T_c) calculados utilizando a função *fpSim* do pacote *ChemmineR* (CAO et al., 2008) usando os *fingerprints* tipo FP2 calculados pelo programa *Open Babel* (O'BOYLE et al., 2011).

(b) Para mais detalhes vide legenda da Figura 4.2.

Além da análise da diversidade estrutural, podemos analisar a complexidade¹¹ estrutural, dos compostos presentes em cada banco, contando o número de fragmentos diferentes em uma mesma molécula, ou seja, número de *bits* iguais a 1. Esta contagem está apresentada nos gráficos (C) e (D) da Figura 4.12. Nestes gráficos, podemos observar que embora os compostos presentes no banco DB sejam os mais dissimilares, entre si, estes compostos são estruturalmente mais simples. Neste contexto, o banco ZINC-Curated é o que mais se assemelha ao banco DB. Como esperado, o banco ZNP apresenta os compostos com a

¹¹ Nesta tese, complexidade estrutural se refere à presença de diferentes grupos funcionais numa mesma molécula.

maior complexidade estrutural. Para exemplificar esta simplicidade/complexidade estrutural, selecionamos o composto com o maior, e com o menor, número de *bits* iguais a 1, presentes em cada um dos bancos (Figura 4.13).

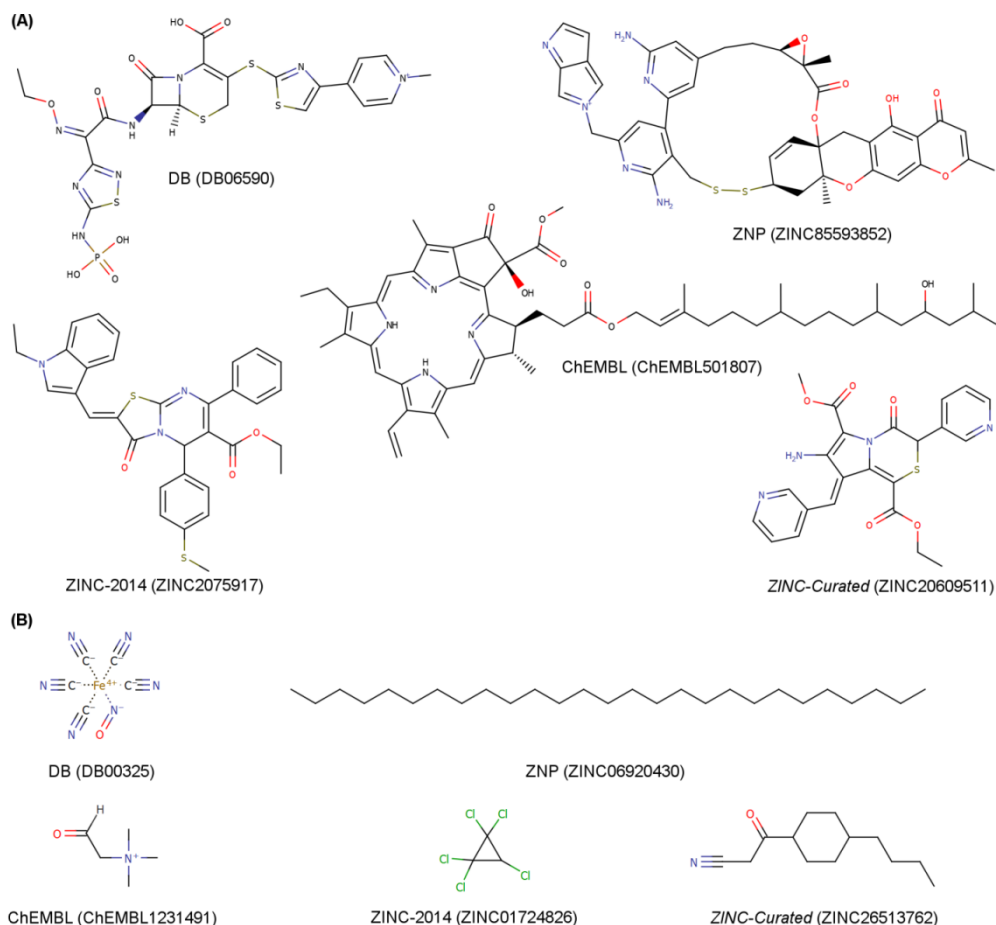


Figura 4.13. Representação em 2D das estruturas dos compostos com o maior e com o menor número de fragmentos^(a) (bits iguais a 1) para cada um dos bancos de dados. (A) Estrutura do composto com maior número de fragmentos nos bancos DB (DB06590, com 340 *bits* =1), ZNP (ZINC85593852, com 413 *bits* =1), ChEMBL (ChEMBL501807, com 378 *bits* =1), ZINC-2014 (ZINC2075917, com 347 *bits* =1) e ZINC-Curated (ZINC20609511, com 340 *bits* =1). (B) Estrutura do composto com menor número de fragmentos nos bancos DB (DB00325, com 3 *bits* =1), ZNP (ZINC06920430, com 6 *bits* =1), ChEMBL (ChEMBL1231491, com *bits* =1), ZINC-2014 (ZINC01724826, com 10 *bits* =1) e ZINC-Curated (ZINC26513762, com 20 *bits* =1).

^(a) Fragmentos gerados usando os *fingerprints* calculados pelo *Open Babel* (O'BOYLE et al., 2011), correspondendo aos diferentes grupos funcionais presentes numa mesma molécula.

Em seguida, para analisar a similaridade entre diferentes bancos, os valores Tc foram calculados para todos os possíveis pares de compostos de dois bancos diferentes, considerando-se todas as possíveis combinações. E os resultados foram, inicialmente, analisados calculando-se o número de compostos similares entre o banco DB e cada um dos

outros bancos de dados, usando diferentes valores de corte de Tc que variou (variando de 0,5 a 1,0). A Tabela 4.8 mostra o número de compostos similares entre o banco DB e cada um dos bancos de dados.

Tabela 4.8. Número de compostos similares (em %) entre o banco DB e cada um dos bancos considerando seis diferentes valores de corte do Tc, variando de 0,5 a 1,0.

Banco de dados	Número de compostos similares ao banco DB (em %)					
	Tc $\geq 0,5$	Tc $\geq 0,6$	Tc $\geq 0,7$	Tc $\geq 0,8$	Tc $\geq 0,9$	Tc = 1,0
ZNP	51,75	24,10	9,23	3,05	1,18	0,27
ChEMBL	37,63	17,33	8,25	3,46	1,17	0,21
ZINC-2014	27,97	7,92	1,63	0,33	0,13	0,00
<i>ZINC-Curated</i>	36,92	11,53	2,25	0,24	0,07	0,00

Nesta tabela, podemos observar que o banco ZNP é que possui maior número de compostos similares ao banco DB. Esta semelhança pode ser considerada esperada, pois dos 1.355 novos fármacos aprovados entre 1981 e 2010, 26 % são produtos naturais ou derivados sintéticos de produtos naturais e 24% são produtos sintéticos baseados em produtos naturais (NEWMAN; CRAGG, 2012). O banco ChEMBL, é segundo banco com maior número de compostos similares ao banco DB. É interessante observar que o banco *ZINC-Curated* possui mais compostos estruturalmente semelhantes ao banco DB do que ao banco ZINC-2014. Estes resultados corroboram com os resultados apresentados anteriormente nesta tese, nos quais se observou que o banco *ZINC-Curated* foi enriquecido com compostos com características próximas as observadas no banco DB. Uma limitação desta análise é a necessidade de se considerar um banco como referência. A seguir, visando-se a analisar a similaridade entre todos os pares de bancos, sem ter de considerar um como referência, foi usado o índice de Similaridade *Hausdorff-Like* - *HLS* (do inglês, *Hausdorff-like Similarity*) (MAURI et al., 2016). O *HLS* foi, recentemente, proposto por Todeschini como um método para comparar dois sistemas químicos (MAURI et al., 2016). Para isso, são calculados valores de similaridade (variando de 0 a 1) entre todos os pares de elementos destes dois sistemas.

Nesta tese, os bancos de dados, os compostos e a valores de Tc correspondem aos sistemas químicos, aos seus correspondentes elementos e a medida de similaridade variando de 0 a 1, respectivamente. Os valores dos índices de *HLS* calculados para todos os pares de bancos estão mostrados na Tabela 4.9.

Tabela 4.9. Valores dos índices de Similaridade Hausdorff-Like (*HLS*) (MAURI et al., 2016) entre dois bancos calculados para todos os possíveis pares de bancos.

Banco de dados	Valores dos Índices de <i>HLS</i>				
	DB	ZNP	ChEMBL	ZINC-2014	ZINC-Curated
DB	1,00				
ZNP	0,53	1,00			
ChEMBL	0,46	0,53	1,00		
ZINC-2014	0,48	0,57	0,58	1,00	
ZINC-Curated	0,46	0,44	0,57	0,64	1,00

O índice de *HLS* também mostra que o banco ZNP possui compostos estruturalmente mais similares aos compostos do banco DB ($HLS = 0,53$). No entanto, diferentemente da análise anterior, a semelhança estrutural entre o banco DB e os bancos ChEMBL, ZINC-2014 e ZINC-Curated é comparável. Como esperado, os bancos ZINC-2014 e ZINC-Curated são os mais parecidos entre si ($HLS = 0,64$). Ainda, podemos observar que embora o banco ZINC-2014 seja o que apresenta a maior similaridade com o banco ZNP ($HLS = 0,57$), este resultado é esperado, uma vez que o ZNP é um subconjunto de produtos naturais do ZINC. No entanto, o ZINC-Curated é o que apresenta a menor similaridade com o banco ZNP ($HLS = 0,44$). Esta diferença se deve ao fato dos compostos do banco ZINC-Curated serem estruturalmente mais simples do que os dos bancos ZINC-2014 e ZNP. Adicionalmente, a semelhança entre ChEMBL e os bancos ZINC-2014 e ZINC-Curated é comparável.

No geral, a análise da similaridade estrutural mostra que a aplicação dos filtros acarretou numa certa diminuição da diversidade estrutural do banco ZINC-Curated quando

comparada a do ZINC-2014. Embora estes resultados não fossem desejados, eles podem ser considerados esperados. No entanto, é importante ressaltar que parte da diversidade estrutural perdida pode ser devido à remoção de grupos funcionais reativos, sendo alguns deles relacionados à ocorrência de resultados falso-positivos e/ou associados à falta de estabilidade e toxicidade dos compostos.

Por fim, considerando todas as análises feitas, embora com uma certa diminuição na diversidade das propriedades físico-químicas e/ou estruturais tenha sido observada, o banco *ZINC-Curated* foi considerado como enriquecido em compostos com características de fármaco podendo ser usado nas buscas virtuais.

4.3. Busca virtual de inibidores da protease do vírus da Dengue

Nesta tese, foram construídos e utilizados quatro diferentes modelos de buscas de inibidores da protease do vírus da Dengue (denominados Modelo I – DENV a Modelo IV - DENV). Como apresentado e descrito na Introdução, estes modelos de busca virtual foram feitos aplicando-se, em sequência, diferentes filtros de seleção. Inicialmente, será apresentado o item referente ao reconhecimento do alvo. Em seguida, serão apresentados os resultados de cada um dos quatro modelos, incluindo os compostos selecionados bem como a etapa de validação experimental. A ordem apresentada dos modelos corresponde à ordem cronológica em que foram construídos.

4.3.1. Reconhecimento da estrutura e características do sítio ativo da protease do vírus da Dengue para escolha da estrutura 3D

Conforme apresentado e discutido na Introdução, para o sucesso de estratégias baseadas na estrutura do receptor, além da disponibilidade da estrutura tridimensional do receptor, outros aspectos relacionados à interação receptor-ligante precisam ser considerados. Dentre estes, citam-se: (i) a identificação da função do receptor estudado; (ii) a identificação

do seu sítio de ligação; (iii) as características físico-químicas deste sítio de ligação; (iv) a flexibilidade conformacional, quando relevante, e (v) a qualidade da estrutura tridimensional disponível.

Nesta tese, o primeiro alvo a ser apresentado e discutido será a protease do vírus da Dengue (DENV). A estrutura tridimensional desta protease foi elucidada pela primeira vez por Erbel e colaboradores, em 2006 (ERBEL et al., 2006) utilizando cristalografia de Raios-X. Atualmente, existem pelo menos quinze estruturas da NS2B/NS3pro do DENV depositadas no *PDB*, sendo três estruturas com os dois domínios (helicase e protease) da proteína NS3 junto com a região da proteína NS2B (códigos PDB: 2VBC; 2WZQ E 2WHX) (LUO et al., 2008) e doze contendo o domínio protease da proteína NS3 junto com a região da proteína NS2B (códigos PDB: 2M9P; 2M9Q; 4M9F; 4M9I; 4M9K; 4M9M; 4M9T; 3U1I; 1UIJ; 3LP6; 3LKW, 2FOM) (CHANDRAMOULI et al., 2010; ERBEL et al., 2006; GIBBS; STEELE; TOUNGE, 2014, 2014; NOBLE et al., 2012; YILDIZ et al., 2013). A Tabela 4.10 apresenta os códigos PDB de cada uma delas, os correspondentes métodos de obtenção, a resolução (quando obtida por cristalografia de Raios-X), o sorotipo do DENV usado, a informação da presença, ou não, de ligantes, os domínios da proteína NS3 e a conformação da região NS2B. A representação das estruturas do(s) ligante(s) presentes nas estruturas 3D da NS2B/NS3pro do DENV disponíveis no *PDB* estão apresentadas na Figura 4.14.

Tabela 4.10. Códigos PDB das quinze estruturas da proteína NS2B/NS3pro do DENV disponíveis no PDB, sorotipo do DENV, método de obtenção, valores de resolução (Å), ligante, domínio da proteína NS3 e a conformação da região NS2B.

Código PDB	Sorotipo do DENV	Método de obtenção	Resolução (Å)	Ligante	Domínios da proteína NS3	Conformação da região NS2B
2FOM	2	Raios-X	1,5	Não	Protease	Parcialmente resolvida
2VBC	4	Raios-X	3,2	Não	Ambos	Parcialmente resolvida
2WHX	4	Raios-X	2,2	Não	Ambos	Parcialmente resolvida
2WZQ	4	Raios-X	2,8	Não	Ambos	Parcialmente resolvida
3L6P	1	Raios-X	2,2	Não	Protease	β -harpin periférico
3LKW	1	Raios-X	2,0	Não	Protease	β -harpin periférico
3U1I	3	Raios-X	2,3	Bz-nKRR-H ^(a)	Protease	β -harpin formando parte do sítio ativo
3U1J	3	Raios-X	1,8	Aprotinina ^(b)	Protease	Parcialmente resolvida
4M9K	2	Raios-X	1,5	Não	Protease	Parcialmente resolvida
4M9M	2	Raios-X	1,5	Não	Protease	Parcialmente resolvida
4M9I	2	Raios-X	2,4	Não	Protease	Parcialmente resolvida
4M9F	2	Raios-X	2,7	Não	Protease	Parcialmente resolvida
4M9T	2	Raios-X	1,7	Não	Protease	Parcialmente resolvida
2M9P	2	RMN	-	Bz-nKRR-CF ₃ ^(c)	Protease	β -harpin formando parte do sítio ativo
2M9Q	2	RMN	-	Bz-nKRR-CF ₃ ^(c)	Protease	β -harpin formando parte do sítio ativo

^(a) Benzoyl-norleucina-lisina-arginina-arginina-aldeído, estrutura apresentada na Figura 4.14(A).

^(b) Estrutura apresentada na Figura 4.14(B).

^(c) Benzoyl-norleucina-lisina-arginina-arginina-trifluormetil, estrutura apresentada na Figura 4.14(C).

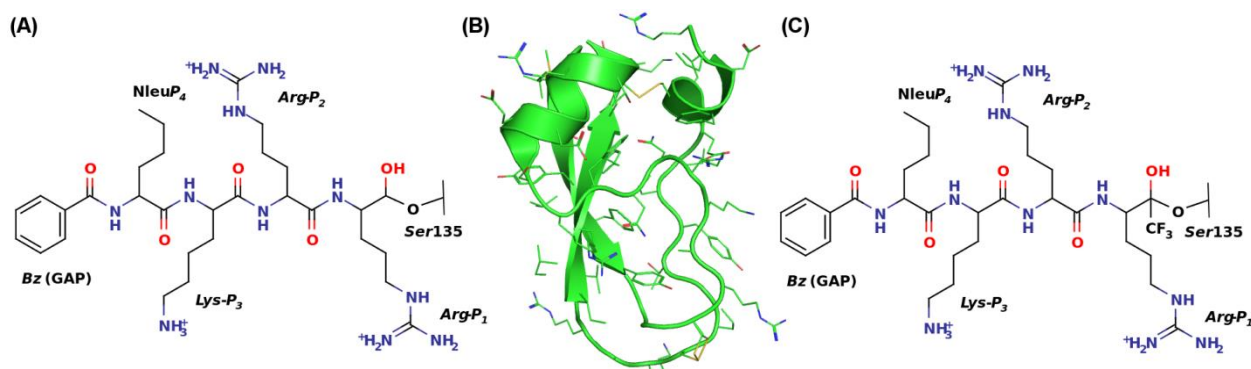


Figura 4.14. Representação das estruturas dos ligantes presentes nas estruturas tridimensionais da NS2B/NS3pro do DENV disponíveis no *PDB*. (A) Estrutura 2D do Bz-nKRR-H, inibidor peptídico mimético covalentemente ligado a *Ser135* ($K_i = 5,8 \mu\text{M}$) como encontrado no PDB 3U1I (resolução 2,30 Å, (NOBLE et al., 2012)). (B) Estrutura 3D da aprotinase, inibidor polipeptídico ($K_i = 26\text{nM}$) como encontrado no PDB 3U1J (resolução 1,80 Å, (NOBLE et al., 2012)). (C) Estrutura 2D do Bz-nKRR-CF₃, inibidor peptídico mimético covalentemente ligado a *Ser135* ($K_i = 0,85 \mu\text{M}$) como encontrado nos *PDBs* 2M9P e 2M9Q (obtidos por RMN, (GIBBS; STEELE; TOUNGE, 2014, 2014)).

Conforme apresentado na Introdução, embora a tríade catalítica (*His51*, *Asp75* e *Ser135*) se encontre no domínio protease da proteína NS3 (denominado NS3pro), a presença de alguns resíduos da proteína NS2B (~40 aminoácidos) é essencial para atividade catalítica desta protease (denominada NS2B/NS3pro). Assim, esta região da proteína NS2B seria responsável por formar parte do sítio ativo da protease (NOBLE et al., 2012). No entanto, nas estruturas da NS2B/NS3pro obtidas sem ligante (*PDBs* 4M9F; 4M9I; 4M9K; 4M9M; 4M9T; 3LP6; 3LKW; 2FOM 2VBC; 2WZQ, 2WHX), a região NS2B apresenta diferentes conformações e não interage com a porção central do domínio NS3pro (vide região colorida de amarelo na Figura 4.15 (A) e (B)), com exceção de uma única folha β . Esta prolonga o núcleo de folhas β do domínio NS3pro (vide folha β colorida de azul escuro na Figura 4.15).

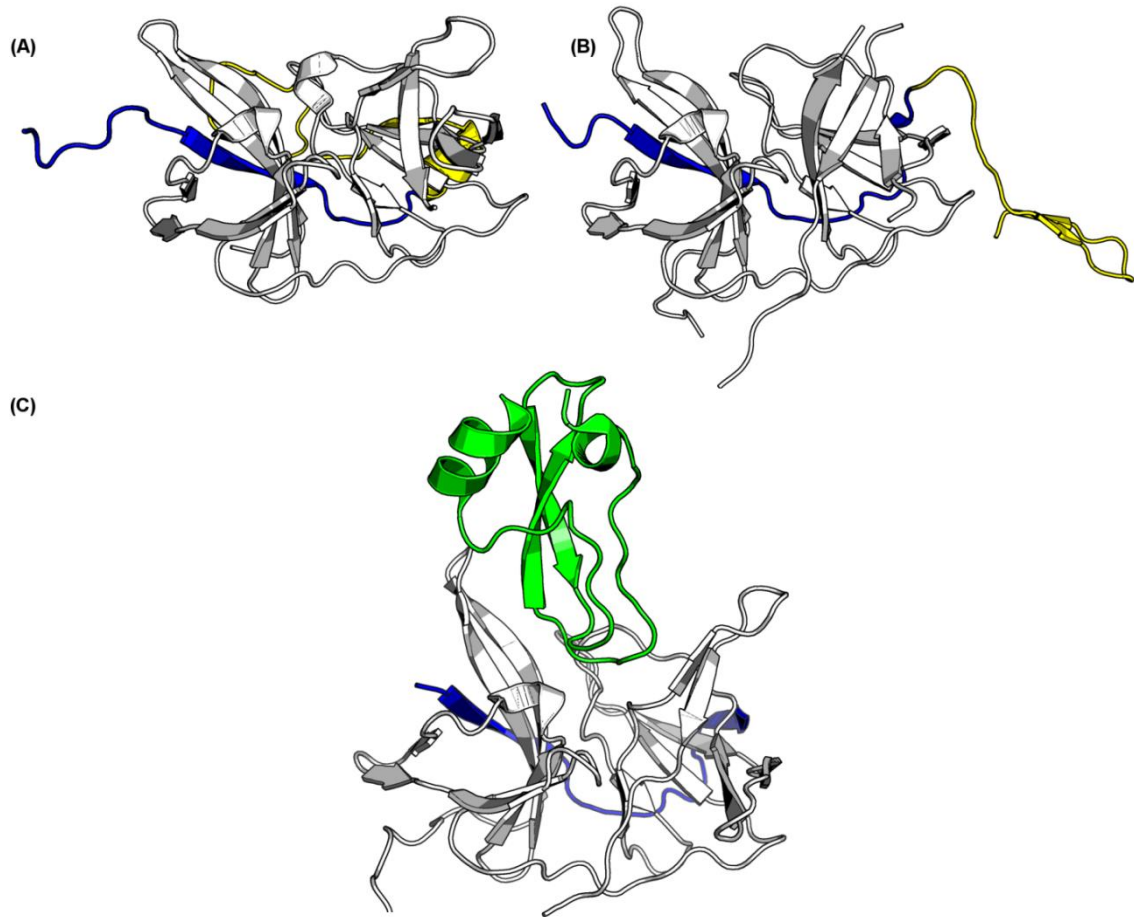


Figura 4.15. Diferentes conformações da região NS2B observadas nas estruturas tridimensionais da NS2B/NS3pro do DENV cristalizadas na ausência de ligante e na presença da aprotinina: **(A)** NS2B/NS3pro do DENV2 com a região NS2B parcialmente resolvida (amarelo), PDB 2FOM (resolução 1,5 Å, (ERBEL et al., 2006)); **(B)** NS2B/NS3pro do DENV1 com a região NS2B formando um β -harpin periférico (amarelo), PDB 3L6P (resolução 1,2 Å, (CHANDRAMOULI et al., 2010)) e, **(C)** NS2B/NS3pro do DENV3 em complexo com aprotinina (verde), PDB 3U1J (resolução 1,8 Å, (NOBLE et al., 2012)). As cadeias principais das proteínas estão representadas em *cartoon*. O domínio NS3pro está mostrado em cinza e a região NS2B formando uma folha β presentes em todas as estruturas está mostrada em azul escuro. Figura preparada com o programa *PyMOL* (SCHRÖDINGER, 2013).

De modo análogo, no complexo NS2B/NS3pro com aprotinina (inibidor polipeptídico de 58 aminoácidos e $K_i = 26$ nM, Figura 4.14 (B)), a região NS2B não interage com a região NS3pro, uma vez que a cavidade de ligação desta região é completamente ocupada pela aprotinina (Figura 4.15 (C)). No entanto, no complexo NS2B/NS3pro com o inibidor Bz-nKRR-H covalentemente ligado à *Ser135* (inibidor peptídico mimético com $K_i = 5,8$ μ M (YIN et al., 2006), Figura 4.14 (A)), a região NS2B envolve a NS3pro formando um β -harpin (duas folhas betas antiparalelas) que compõem parte do sítio ativo da protease (vide β -harpin colorido de amarelo na Figura 4.16). Analogamente, esta estrutura de β -harpin já havia sido

observada para complexos da NS2B/NS3pro do vírus da Febre do Nilo (WNV, *West Nile virus*) com outros inibidores peptídicos, sendo que esta apresenta 56% de identidade em relação a protease do DENV (NOBLE et al., 2012; WICHAPONG et al., 2010). Em 2014, duas estruturas da NS2B/NS3pro do DENV complexadas com o inibidor Bz-nKRR-CF₃ covalentemente ligado à *Ser135* (inibidor peptídeo mimético com $K_i = 0,85 \mu\text{M}$ (YIN et al., 2006), Figura 4.14 (C)) foram obtidas por RMN (2M9P e 2M9Q, (GIBBS; STEELE; TOUNGE, [s.d.]; GIBBS; TOUNGE; STEELE, [s.d.]b)). Em ambos os complexos, a região NS2B envolve o domínio NS3pro formando um β -harpin. Cabe ressaltar ainda que nas três estruturas contendo os dois domínios da proteína NS3 – 2VBC; 2WZQ e 2WHX (LUO et al., 2008) – observa-se uma clara separação entre o domínio protease e helicase da proteína NS3 sendo, assim, possível considerar apenas o domínio NS3pro tanto para estudos *in vitro* bem como *in silico*.

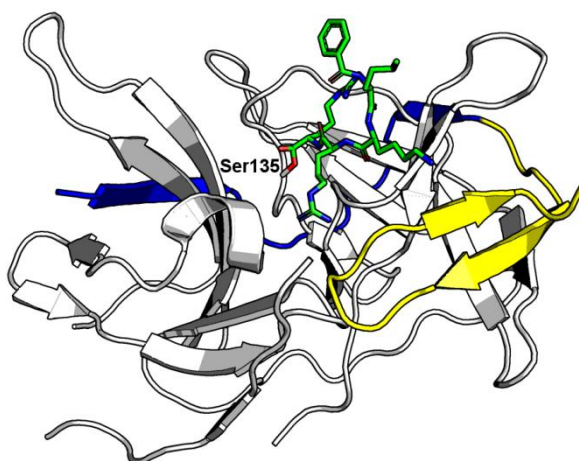


Figura 4.16. Região NS2B formando um β -harpin que compõem parte do sítio ativo da NS2B/NS3pro do DENV. NS2B/NS3pro do DENV3 co-cristalizada com o inibidor Bz-nKRR-H (verde) covalentemente ligado à *Ser135*, PDB 3U1I (resolução 2,3 Å, (NOBLE et al., 2012)), na qual a NS2B forma um β -harpin que compõem parte do sítio ativo (amarelo). As cadeias principais das proteínas estão representadas em *cartoon*. O domínio NS3pro está mostrado em cinza e a região NS2B formando uma folha β presente nesta estrutura está mostrada em azul escuro. O inibidor Bz-nKRR-H e a cadeia lateral da *Ser135* estão representados em *sticks*. Figura preparada com o programa PyMOL (SCHRÖDINGER, 2013).

Desta forma, considerando-se que a ótima atividade catalítica depende tanto da presença da região NS2B bem como desta estar formando parte do sítio ativo da NS2B/NS3pro, a estrutura 3D da NS2B/NS3pro co-cristalizada com o inibidor Bz-nKRR-H

(3U1I, resolução 2,3 Å, (NOBLE et al., 2012)) foi escolhida como referência para se iniciar os estudos de modelagem molecular, embora esta não seja a estrutura com melhor resolução disponível no *PDB*. Vale salientar que embora as estruturas 2M9P e 2M9Q também apresentem as regiões NS2B interagindo com o domínio NS3pro, detalhes destas estruturas não estão disponíveis no *PDB*, até a elaboração desta tese, impossibilitando que estas fossem utilizadas nos nossos estudos de modelagem molecular.

4.3.1.1. Reconhecimento da estrutura da proteína por inspeção visual

O reconhecimento da estrutura da NS2B/NS3pro foi realizado, inicialmente, por inspeção visual utilizando a estrutura 3U1I (2,3 Å). Esta estrutura apresenta duas macromoléculas na unidade assimétrica formadas respectivamente pelas cadeias A (NS2B) e B (NS3pro) e C (NS2B) e D (NS3pro), respectivamente. Na primeira macromolécula, o inibidor Bz-nKRR-H se encontra no sítio ativo da NS2B/NS3pro. Já na segunda macromolécula, Noble e colaboradores observaram a presença de um ligante menor (modelado como sendo *Arg-Arg*). Segundo estes autores, esta diferença se deve a efeitos de empacotamento cristalino. Sendo assim, em todos os estudos de modelagem apresentados nesta tese apenas as cadeias A e B da estrutura 3U1I foram usadas. Conforme descrito por Noble e colaboradores (NOBLE et al., 2012), o carbono carbonílico do resíduo *Arg(P_I)* do inibidor Bz-nKRR-H se encontra a uma distância de 1,3 Å do átomo de oxigênio da cadeia lateral da *Ser135*. Esta distância é concordante com a formação de uma ligação covalente dando origem ao hemiacetal observado nesta estrutura (Figura 3.17 (A)). Ainda, pode-se observar densidade eletrônica entre estes dois átomos (NOBLE et al., 2012).

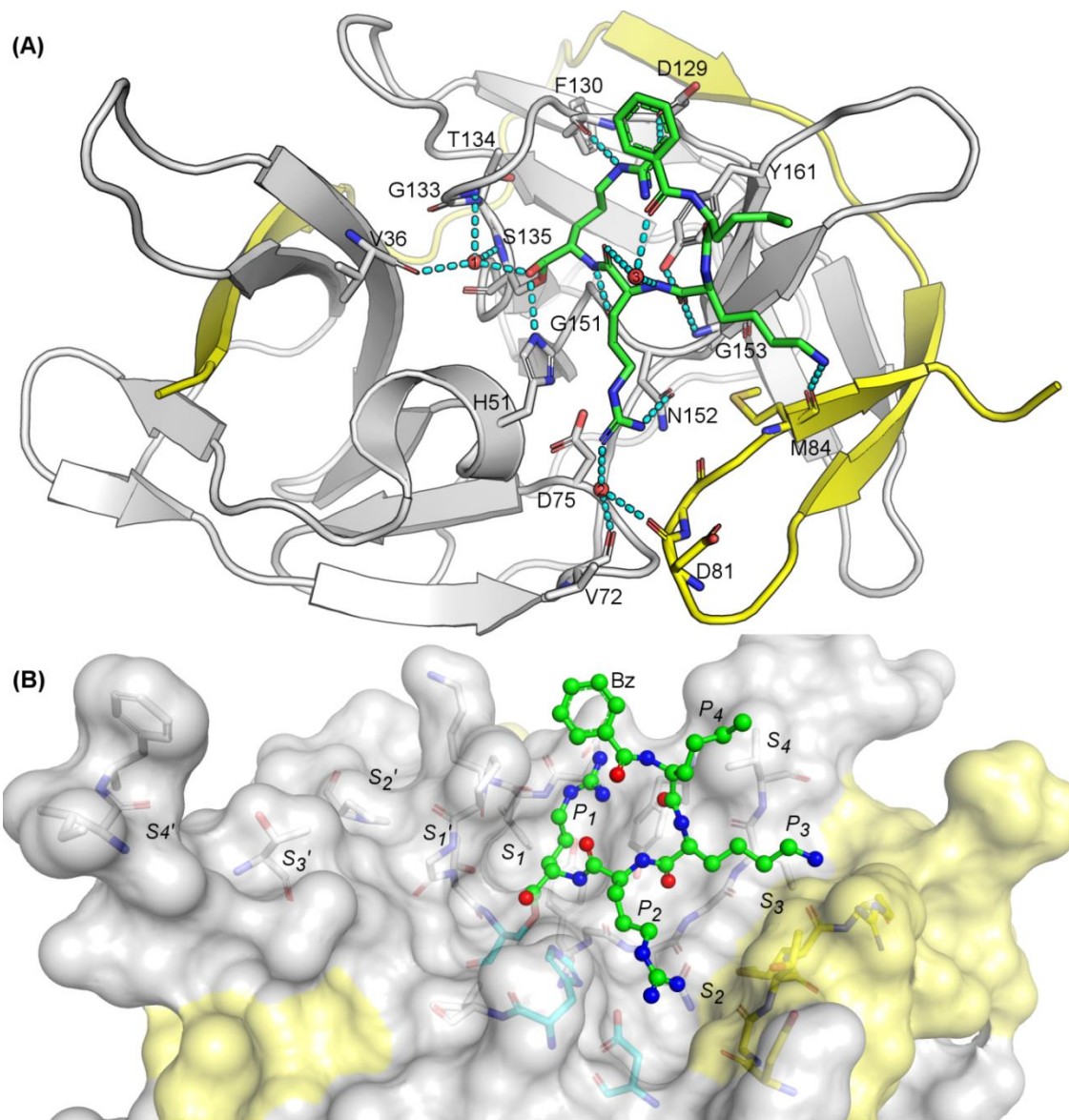


Figura 4.17. Representação das interações proteína - Bz-nKRR-H e das cavidades S_1 a S_4' da NS2B/NS3pro do DENV^(a). (A) Interações entre os resíduos do sítio ativo e o inibidor Bz-nKRR-H covalentemente ligado a *Ser135*, intermediadas, ou não, por moléculas de água. O inibidor Bz-nKRR-H e os resíduos do sítio ativo, a proteína e as moléculas de água estão mostrados como *sticks*, *cartoon* e esferas, respectivamente. Os átomos de carbono do inibidor e dos resíduos do domínio NS3pro e da região NS2B estão mostrados em verde, cinza e amarelo, respectivamente. Ligações de hidrogênio estão mostradas com linhas pontilhadas. As moléculas de água números 1, 2 e 3, correspondem às moléculas 206, 225 e 202 do PDB 3U1I. (B) Cavidades S_1 - S_4' da NS2B/NS3pro do DENV. As cavidades S_1 a S_4 foram definidas com base nos resíduos do inibidor Bz-nKRR-H (P_1 a P_4) e as cavidades S_1' a S_4' foram definidas conforme descrito na literatura (NOBLE et al., 2012). A proteína, os resíduos das cavidades e o inibidor estão mostrados como superfície, *sticks* e *ball and sticks*, respectivamente. Os átomos de carbono dos resíduos do domínio NS3pro, da região NS2B e da tríade catalítica e do inibidor estão mostrados em cinza claro, amarelo, ciano e verde, respectivamente. Figura preparada com o programa *PyMOL* (SCHRÖDINGER, 2013).

^(a) A estrutura NS2B/NS3pro do DENV3 corresponde ao PDB 3U1I (resolução 2,3 Å (NOBLE et al., 2012)).

Os resíduos que compõem a tríade catalítica desta estrutura (*Asp75*, *His51* e *Ser135*) estão localizados entre as cavidades do sítio ativo S_1 e S_2 . Dentre os resíduos que formam a

cavidade S_1 destacam-se a *His51*, o *Asp129*, a *Phe130*, a *Ser135*, a *Tyr150* e a *Tyr160*. O resíduo *His51* faz uma ligação de hidrogênio com o oxigênio do hemiacetal formado. É interessante observar que não há interação direta entre os resíduos da cavidade oxânion (*Gly133*, *Thr134* e *Ser135* do domínio NS3pro) e o hemiacetal formado, uma vez que esta cavidade é ocupada por uma molécula de água (HOH206 do PDB 3U1I) e um íon sulfato (NOBLE et al., 2012). Contudo, esta molécula de água (denomina nesta tese água número 1) intermedeia as ligações de hidrogênio entre o hemiacetal e os resíduos *Val36*, *Gly133* e *Ser135* da protease (Figura 4.17(A)). Os resíduos *Asp129* e *Phe130* formam ligações de hidrogênio com o grupo guanidino da *Arg(P₁)* do inibidor Bz-nKRR-H. Além destas interações, este grupo guanidino forma uma ligação iônica e uma interação cátion- π com os resíduos *Asp129* e *Tyr160*, respectivamente. A cavidade S_2 é composta pelos resíduos *His51*, *Val72*, *Asp75*, *Gly151*, *Asn152* do domínio NS3pro e pelos resíduos *Asp81* e *Gly82* da região NS2B. Os resíduos *Gly151* e *Asn152* formam ligações de hidrogênio diretamente com a *Arg(P₂)* do inibidor Bz-nKRR-H, já para os resíduos *Val72* e *Asp81* (ambos da região NS2B) as ligações de hidrogênio com a *Arg(P₂)* são intermediadas por uma molécula de água (denomina nesta tese água número 2, correspondendo a HOH225 do PDB 3U1I). De modo análogo, a cavidade S_3 é composta por resíduos do domínio NS3pro (*Gly153*, *Val154* e *Tyr160*) e da região NS2B (*Thr83*, *Met84* e *Arg85*). A *Lys(P₃)* do inibidor Bz-nKRR-H forma ligações de hidrogênio com os resíduos *Gly153*, *Tyr160* e *Met84* (região NS2B). Por fim, a norleucina – n(P_4) e o grupo benzoil (Bz) parecem não formar interações com os resíduos do sítio ativo ficando exposto ao solvente. Além das interações descritas acima, este inibidor estabelece três ligações de hidrogênio intramoleculares intermediadas por uma única molécula de água (denominada nesta tese água número 3, correspondendo a HOH202 do PDB 3U1I). Embora nesta estrutura não se observa qualquer interação entre o inibidor Bz-nKRR-H e os resíduos das cavidades S_1' a S_4' , a posição destas cavidades foi inferida de dados de literatura

(NOBLE et al., 2012) e mostrada na Figura 4.17 (B). Vale ainda ressaltar que as cavidades que compõem o sítio ativo da NS2B/NS3pro são rasas e expostas ao solvente.

4.3.1.2. Reconhecimento da estrutura da proteína auxiliado por computador

Numa segunda etapa, a análise da superfície e da natureza das interações possíveis na NS2B/NS3pro do DENV (PDB 3U1I) foi feita através dos potenciais eletrostáticos e das densidades doadora e/ou aceptora de ligação de hidrogênio calculados pela ferramenta *MOLCAD* do programa *SYBYL-X* (SYBYL, 2013). Estas três propriedades da superfície estão apresentadas na Figura 4.18. O potencial eletrostático calculado para superfície da NS2B/NS3pro do DENV variou de 229 até -316 kcal/mol elétron. Conforme descrito na literatura (NITSCHKE et al., 2014), podemos observar o caráter polar do sítio ativo da NS2B/NS3pro do DENV, principalmente nas cavidades S_1 , S_2 e S_3 . Adicionalmente, as cavidades S_1 , S_2 e S_3 tem um potencial eletrostático negativo sendo, portanto, favoráveis às interações com ligantes contendo grupos carregados positivamente. Esta preferência por grupos carregados positivamente está de acordo com as observações experimentais de que os substratos naturais (porções da poliproteína viral) da NS2B/NS3pro possuem, geralmente, resíduos de aminoácidos básicos (*Lys-Arg*, *Arg-Arg* ou *Arg-Lys*) nas posições P_2 - P_1 . Além disso, estes substratos possuem resíduos com caráter alifático ou polar neutro em P_1' – *Gly/Ser/Ala* (ARAVAPALLI et al., 2012; NOBLE et al., 2012) – concordando também com o potencial eletrostático observado para a cavidade S_1' (nem positivo e nem negativo).

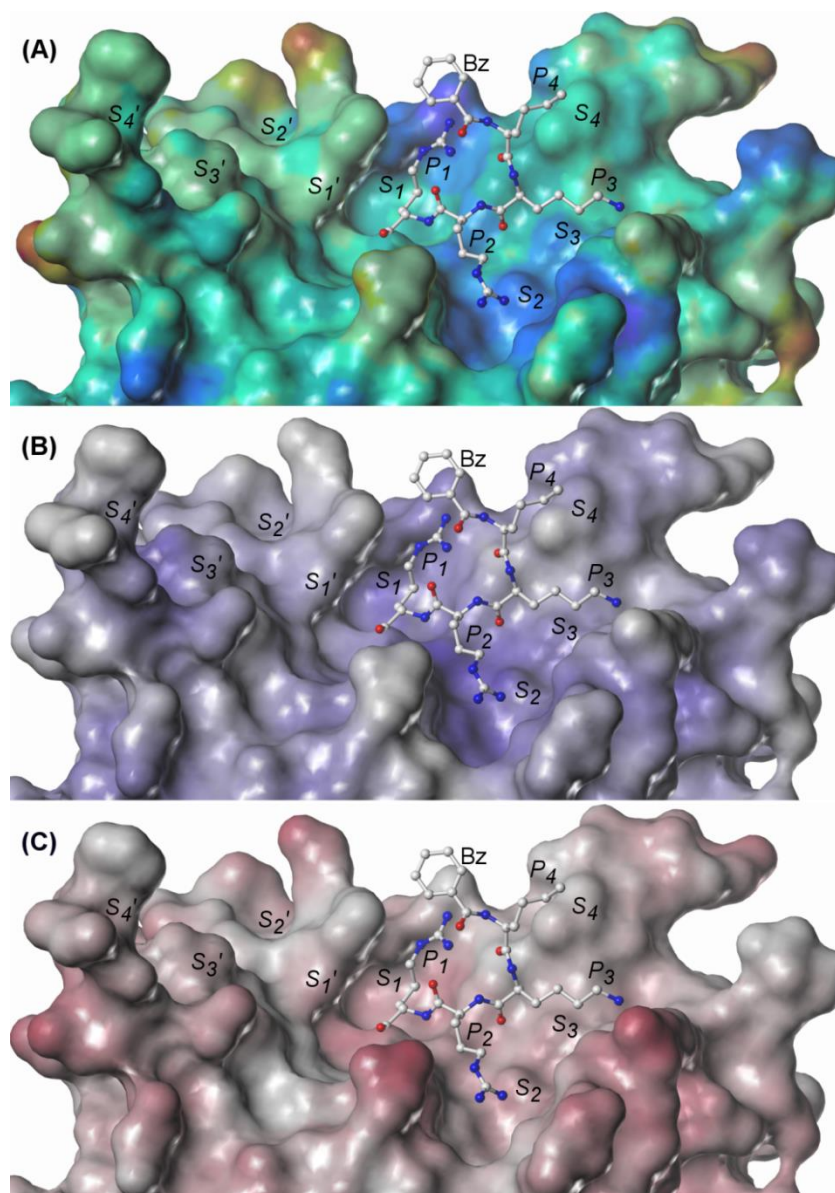


Figura 4.18. Representações da superfície^(a) da NS2B/NS3pro do DENV^(b) colorida de acordo com algumas propriedades^(c): (A) potencial eletrostático variando de 229 até -316 kcal/mol elétron (vermelho e roxo, respectivamente); (B) a densidade *HBA* variando de 0.000 até 0.085 Å⁻² (cinza e azul, respectivamente) e (C) a densidade de *HBD* variando de 0.000 até 0.172 Å⁻² (cinza e vermelho, respectivamente). As cavidades *S₄* a *S₁'* foram definidas com descrito na Figura 4.17. A proteína e o inibidor estão mostrados como superfície e *ball and sticks*, respectivamente. Figura preparada com o programa *SYBYL-X* (SYBYL, 2013).

^(a) Superfície acessível ao solvente gerada usando o algoritmo de *Connolly* (CONNOLLY, 1983) do programa *SYBYL-X* (SYBYL, 2013).

^(b) A estrutura NS2B/NS3pro do DENV3 corresponde ao PDB 3U1I (resolução 2,3 Å, (NOBLE et al., 2012)).

^(c) Propriedades calculadas utilizando a ferramenta *MOLCAD* do programa *SYBYL-X* (SYBYL, 2013) após adicionar todos os átomos de hidrogênio e calcular as cargas parciais tipo *Gasteiger-Hueckel* (PURCELL; SINGER, 1967).

Ainda, vale ressaltar o caráter eletrostático positivo próximo a *His51* (entre *S₁* e *S₂*), uma vez que nesta região se forma o intermediário tetraédrico carregado negativamente que deve ser estabilizado pelos resíduos da cavidade oxânion. Em relação à densidade doadora e

aceptora de ligação de hidrogênio, se observa uma preferência por grupos doadores de ligação de hidrogênio nas cavidades S_1 e S_2 , ou seja, região com alta densidade de aceptores de ligação de hidrogênio (azul). É interessante observar que o valor da densidade de aceptores de ligação de hidrogênio na cavidade S_2 é maior quando comparado ao da cavidade S_1 . Na estrutura 3U1I, esta região é ocupada pela molécula de água número 2 (Figura 4.17). Já grupos aceptores de ligação de hidrogênio seriam mais favoráveis na região formada pelas cavidades S_1' , S_1 e S_2 próxima à *His51* e à cavidade oxânion. Esta preferência é concordante com o mecanismo de catálise desta enzima, no qual ocorre a formação de um intermediário tetraédrico entre as cavidades S_1' , S_1 e S_2 que é estabilizado por ligações de hidrogênio.

No reconhecimento da NS2B/NS3pro do DENV por inspeção visual foi observado que seu sítio ativo é raso e achatado. Esta observação foi comprovada calculando-se a profundidade desta cavidade utilizando a ferramenta *MOLCAD* do programa *SYBYL-X* (SYBYL, 2013). Nesta representação, cavidades profundas são coloridas em vermelho e cavidades rasas em azul como mostrado na Figura 4.19. E, assim com esperado, a coloração das cavidades do sítio ativo da NS2B/NS3pro do DENV variam de azul escuro a um amarelo claro, indicando um sítio raso com porções expostas ao solvente. A profundidade do sítio ativo é uma das características utilizadas para medir sua *druggability*¹² (VOLKAMER et al., 2012). Assim, sítios ativos rasos podem ser um indicativo de sítios não *druggables*.

¹² Probabilidade de um alvo biológico de interesse ser modulado por uma molécula pequena.

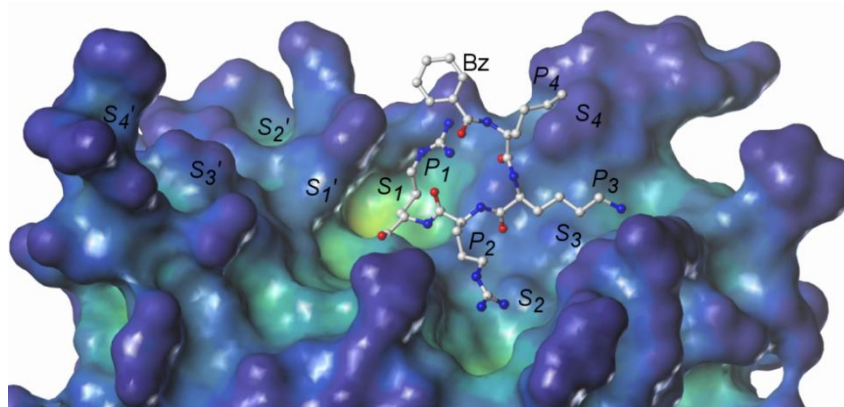


Figura 4.19. Representação da profundidade^(a) da superfície^(b) da NS2B/NS3pro do DENV. Superfície da NS2B/NS3pro do DENV colorida de acordo com a profundidade de cada região variando de -0,3 Å (azul) a 11 Å (vermelho). As cavidades S_4 e S_1' foram definidas com descrito na Figura 4.17. A proteína e o inibidor estão mostrados como superfície e *ball and sticks*, respectivamente.

^(a) A profundidade foi calculada usando a ferramenta *MOLCAD* do programa *SYBYL-X* (SYBYL, 2013).

^(b) Superfície acessível ao solvente gerada como descrito na Figura 4.18.

A análise da *druggability* deste sítio ativo foi feita com auxílio da ferramenta *DoGSiteScorer* (VOLKAMER et al., 2012). Nove sítios foram identificados e classificados de acordo com seu valor de *drugScore* (denominados **1** a **9**). Estes sítios estão mostrados na Figura 4.20 e seus respectivos valores de *drugScore*, de profundidade e da razão superfície/volume estão apresentados na Tabela 4.11.

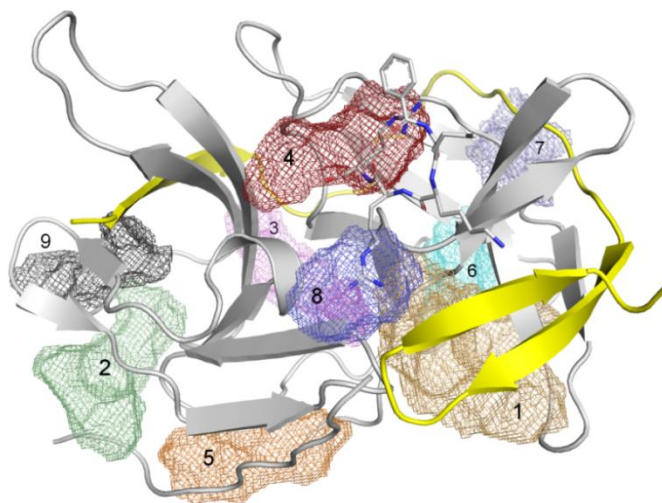


Figura 4.20. Sítios de ligação identificados^(a) para a NS2B/NS3pro do DENV. Os sítios de ligação **1** a **9** estão mostrados como superfície de malha (bege, verde, rosa, vermelho, laranja, ciano, azul claro, azul escuro e preto, respectivamente). A proteína e o inibidor estão mostrados como *cartoon* e *sticks*, respectivamente. A região NS2B e o domínio NS3pro estão em amarelo e cinza, respectivamente. Figura preparada com o programa *PyMOL* (SCHRÖDINGER, 2013).

^(a) Identificados para estrutura 3U1I (resolução 2,3 Å, (NOBLE et al., 2012)), usando a ferramenta *DoGSiteScorer* (VOLKAMER et al., 2012) e classificados de acordo com seus valores de *drugScore* (variando de 0,80 a 0,23).

Tabela 4.11. Valores de *DrugScore*, da profundidade e da razão superfície/volume de cada um dos nove sítios de ligação identificados para a NS2B/NS3pro do DENV (estrutura 3U1I).

Sítio de ligação	<i>DrugScore</i>	Profundidade	Superfície/volume
1	0,80	18,39	1,38
2	0,57	13,89	1,90
3	0,58	14,20	1,53
4	0,39	10,72	1,42
5	0,51	13,15	2,10
6	0,32	9,14	2,14
7	0,52	14,51	2,35
8	0,23	8,14	1,30
9	0,46	13,45	2,13

Dos nove sítios de ligação identificados, dois sítios (**4** e **8**) se encontram na região do sítio ativo da NS2B/NS3pro do DENV. Estes sítios de ligação estão localizados nas cavidades S_1 e S_2 (Figura 4.21 (A)). Os baixos valores de *drugScore* (0,39 e 0,23 para **4** e **8**, respectivamente) obtidos se devem, principalmente, ao fato deles serem rasos. Contudo, estes dois sítios juntos cobrem 44 % da área total do inibidor Bz-nKRR-H e quase toda a área não exposta ao solvente deste inibidor. Os sítios de ligação com maior valor de *drugScore* (**1** e **2**) se encontram abaixo do sítio ativo da NS2B/NS3pro do DENV, sendo o primeiro formado tanto pela região NS2B quanto pelo domínio NS3pro (Figura 4.21 (B)). O sítio de ligação **1** já foi sugerido na literatura como um possível sítio alostérico (WU et al., 2015).

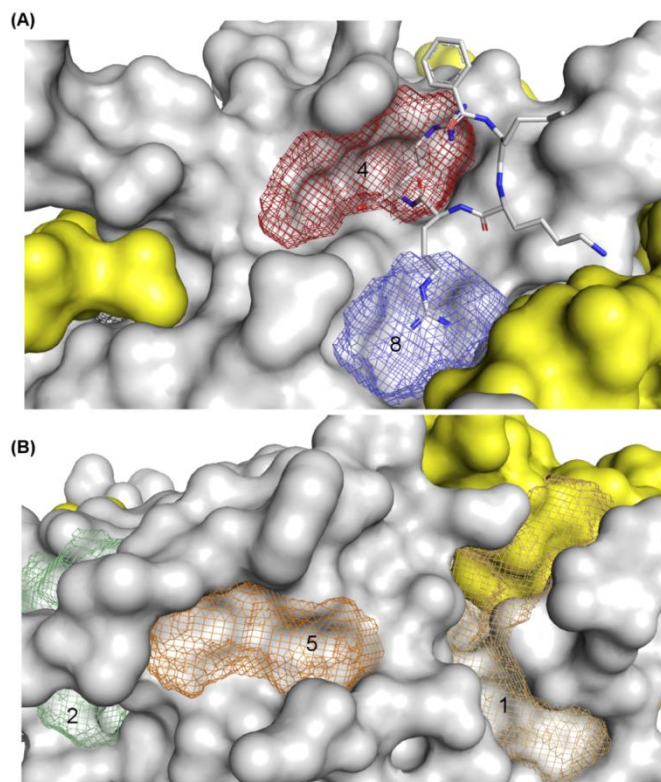


Figura 4.21. Representação dos sítios de ligação identificados^(a) na região do sítio ativo da NS2B/NS3pro do DENV e do sítio de ligação com maior valor de *drugScore*. (A) Sítios de ligação, denominados **4** e **8**, localizados nas cavidades S_1 e S_2 do sítio ativo da NS2B/NS3pro do DENV. (B) Sítios de ligação, denominados **1** e **2**, com os melhores valores de *drugScore* (0,80 e 0,57, respectivamente) localizado a abaixo do sítio ativo da NS2B/NS3pro (rotação de 270° entorno do eixo x) e sítio de ligação **5** (*drugScore* = 0,51). Os sítios de ligação 1, 2, 4, 5 e 8 estão mostrados como superfície de malha (bege, verde, vermelho, laranja e azul escuro, respectivamente). A proteína e o inibidor estão mostrados como superfície e *sticks*, respectivamente. A região NS2B e o domínio NS3pro estão representados em amarelo e cinza, respectivamente. Figura preparada usando o *PyMOL* (SCHRÖDINGER, 2013).

^(a) Identificados para estrutura 3U1I (resolução 2,3 Å (NOBLE et al., 2012)), usando a ferramenta *DoGSiteScorer* (VOLKAMER et al., 2012) e classificados de acordo com seus valores de *drugScore* (variando de 0,80 a 0,23)

Embora o sítio ativo da NS2B/NS3pro do DENV seja superficial e achatado, ainda foi possível identificar sítios de ligação próximos às cavidades S_1 e S_2 . Ainda, estas cavidades apresentam um alto potencial eletrostático e uma grande densidade de doadores e de aceptores de ligação de hidrogênio. Sendo assim, as cavidades S_1 e S_2 do sítio ativo da NS2B/NS3pro do DENV foram escolhidas como cavidades alvo para os procedimentos de busca virtual.

4.3.1.3. Escolha da conformação do receptor e considerações sobre sua flexibilidade conformacional

Conforme apresentado na Introdução, uma das limitações nos estudos baseados na estrutura do receptor é considerar o receptor como uma estrutura rígida, sendo este ainda um

dos aspectos pouco explorados nas buscas virtuais (SCIOR et al., 2012). Neste contexto, torna-se importante não só escolher uma conformação relevante do receptor, mas também considerar, dentro do possível, aspectos relacionados à sua flexibilidade conformacional. Como apresentado no item 4.2.1.1, nas estruturas cristalográficas obtidas na ausência de ligante se observam diferenças significativas na posição e no enovelamento da região NS2B. Na literatura as diferentes formas de enovelamento da região NS2B da protease do DENV são, às vezes, denominadas de “aberta” e “fechada”, sendo esta última a conformação observada na estrutura 3U1I (NITSCHKE et al., 2014). Ainda, a formação da conformação “fechada” é descrita como consequência da ligação do substrato/ligante ao sítio ativo, que levaria ao ajuste induzido da região NS2B (NITSCHKE et al., 2014). Entretanto, mais recentemente, estudos de RMN e de dinâmica molecular (DM) evidenciaram que a conformação “fechada” é a conformação predominante independentemente da presença, ou ausência, do ligante (DE LA CRUZ et al., 2011; PICCIRILLO et al., 2016). Outro estudo feito por RMN (DE LA CRUZ et al., 2014) mostrou que a conversão entre a conformação “fechada” e “aberta” aumenta significativamente com os aumentos do pH e da concentração de sal, indicando a importância das interações eletrostáticas para a associação entre a região NS2B e o domínio NS3pro, também observadas por estudos de DM (PICCIRILLO et al., 2016). Por fim, nestes estudos as conformações “abertas” observadas nas estruturas cristalográficas são explicadas tanto pela presença de artefatos do empacotamento cristalino bem como por uma possível degradação desta protease (CHEN et al., 2014). Contudo, embora a conformação “fechada” seja a predominante em solução, todos estes estudos observaram uma alta flexibilidade na região NS2B. Baseado na literatura (NITSCHKE et al., 2014) e no apresentado acima, a conformação “fechada” é considerada como a única conformação relevante da NS2B/NS3pro do DENV. Nesta tese, inicialmente, a busca virtual foi realizada sem considerar a flexibilidade conformacional já descrita para esta protease (Modelo I –

DENV). Posteriormente, alguns aspectos relacionados a esta flexibilidade foram, então, incluídos na busca virtual, resultando em mais três modelos (Modelo II – DENV a Modelo IV – DENV).

4.3.2. Modelos de busca virtual: Modelo I - DENV

Uma das alternativas descritas na literatura para superar as dificuldades associadas à proposição de modelos de busca virtual válidos é a de aplicação diferentes filtros de seleção de modo sequencial (KLEBE, 2006). Assim, procedendo-se como descrito item 3.5.4.9 (Materiais e Métodos) e partindo-se do banco de dados *ZINC-Original*, foram aplicados filtros de diferentes naturezas para selecionar inibidores da NS2B/NS3pro do DENV. Os filtros de seleção aplicados foram farmacofórico, “*drug-like*”, similaridade, ancoragem e inspeção visual. Os compostos selecionados por este primeiro modelo de busca virtual (denominado Modelo I - DENV) foram testados como inibidores da NS2B/NS3pro do DENV. A Figura 4.22 mostra uma representação esquemática da sequência de aplicação de cada filtro bem como os valores da porcentagem de redução do número de compostos em cada etapa.

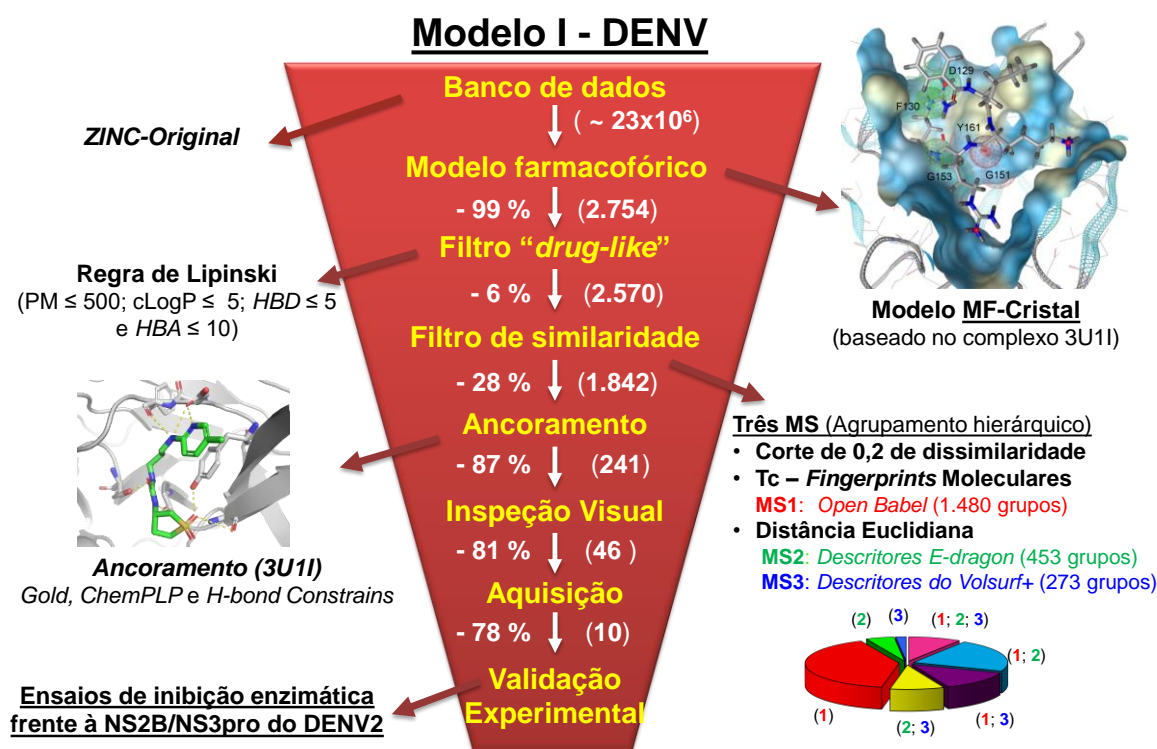


Figura 4.22. Representação esquemática da sequência de filtros de seleção aplicados ao banco *ZINC-Original* usada no primeiro modelo de busca virtual (**Modelo I – DENV**). Entre parênteses são indicados os números de compostos selecionados em cada etapa e ao lado a porcentagem de redução por cada filtro.

4.3.2.1. Filtro farmacofórico

A primeira etapa do procedimento de busca virtual foi a aplicação de um filtro farmacofórico ao banco de dados *ZINC-Original*, como descritos no item 3.5.4.9. O modelo farmacofórico foi construído usando apenas as regiões de interação entre o inibidor Bz-nKRR e a NS2B/NS3pro do DENV presentes na estrutura 3U1I. O reconhecimento de todas as regiões de interação entre este inibidor e a NS2B/NS3pro do DENV foi feito pelo programa *LigandScout* (WOLBER; LANGER, 2005). Nesta etapa, foi possível identificar todas as 11 interações reconhecidas anteriormente por inspeção visual (item 4.2.1), sendo estas: 6 regiões HBD envolvendo os resíduos *His51*, *Gly151*, *Asp129*, *Phe130* e *Met84* (NS2B); 2 regiões HBA envolvendo os resíduos *Gly153* e *Tyr161*; 2 regiões de interação iônica com carga positiva com os resíduos *Asp75* e *Asp129* e, 1 região de interação cátion- π com a *Tyr161*. Estas regiões de interação estão apresentadas na Figura 4.23 (A) e (B).

Sabendo-se que para séries de inibidores peptídeo miméticos, modificações nos resíduos P_1 tem uma grande influência na atividade inibitória e que a cavidade S_1 foi predita como mais *druggable* que a cavidade S_2 (vide item 4.2.1.2), neste primeiro modelo farmacofórico apenas as regiões de interação com resíduos próximos à cavidade S_1 foram consideradas, como descritos no item 3.5.4.9. Ainda, dentre estas interações optou-se por manter apenas as ligações de hidrogênio com os resíduos *Asp129*, *Phe130*, *Gly151*, *Gly153* e *Tyr161*. Estas interações geraram o modelo farmacofórico MF-Cristal que contém seis regiões de interação, a saber: quatro regiões *HBD* e duas regiões *HBA* (Figura 4.23 (C)).

O modelo farmacofórico MF-Cristal aplicado ao banco *ZINC-Original* selecionou 2.756 compostos (*Pharmacophore Fit* entre 61,30 e 68,10) capazes de satisfazer todas as características do modelo. Desta forma, a aplicação deste filtro representou uma redução de 99,99% em relação ao número de compostos no banco *ZINC-Original* ($\sim 23 \times 10^6$).

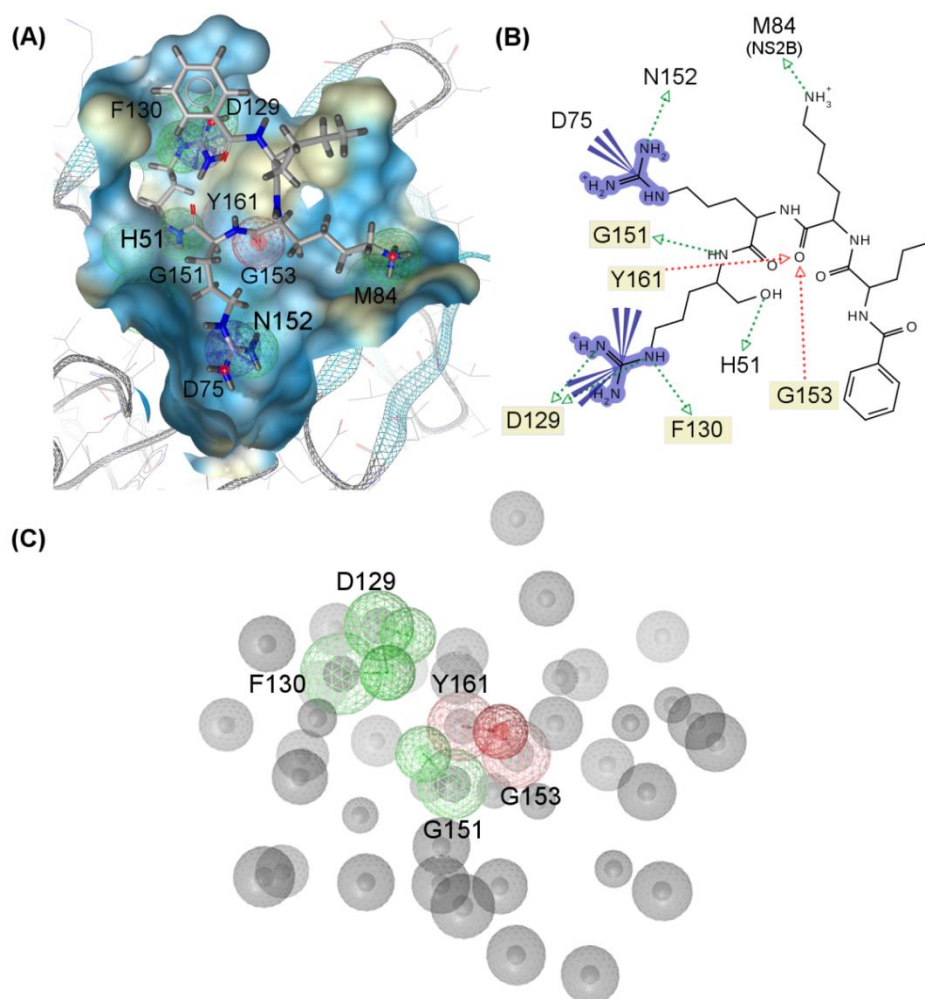


Figura 4.23. Representação esquemática das 11 regiões de interação identificadas entre o inibidor Bz-nKRR-H e a NS2B/NS3pro do DENV e modelo farmacofórico MF-Cristal^(a) gerado a partir destas interações. (A) Sítio de ligação do inibidor Bz-nKRR-H mostrando as 11 regiões de interações identificadas. (B) Representação 2D das 11 regiões de interação identificadas, destacando as interações utilizadas para gerar o modelo MF-Cristal. (C) Modelo farmacofórico MF-Cristal com 6 regiões de interação^(b) obrigatórias (2 *HBD* e 4 *HBA*) e várias regiões de exclusão de volume^(c). Regiões de interação *HBD*, *HBA*, iônica com carga positiva, cátion- π e regiões de exclusão de volume estão representadas em verde, vermelho, azul, azul e cinza, respectivamente.

^(a) Gerado com o programa *LigandScout* (WOLBER; LANGER, 2005) a partir da estrutura 3U1I (vide 3.5.4.9).

^(b) Todas as regiões de interação *HBA* e *HBD* possuem um raio de 1,8 Å.

^(c) Definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio de ligação da NS2B/NS3pro.

4.3.2.2. Filtro “drug-like”

A segunda etapa do procedimento de busca virtual foi a aplicação de um filtro físico-químico aos 2.756 compostos selecionados pelo modelo farmacofórico visando selecionar compostos com características de fármacos (*e.g.* boa permeabilidade e biodisponibilidade via oral e baixa toxicidade). Ao mesmo tempo, reconhecemos que a aplicação destes filtros embora extensivamente recomendada (BOLOGA; OLAH; OPREA, 2006; LIPINSKI et al.,

1997) é também bastante criticada na literatura (SCIOR et al., 2012). Procedendo-se como descrito no item 3.5.4.9, as propriedades físico-químicas utilizadas neste filtro foram baseadas na “Regra dos 5” de Lipinski (LIPINSKI et al., 1997). Os valores de corte das quatro propriedades da “Regra dos 5” bem como os correspondentes intervalos de variação (máximo e mínimo) calculados para os 2.756 compostos estão apresentados na Tabela 4.12.

Tabela 4.12. Valores das propriedades físico-químicas (peso molecular, ClogP e número de átomos *HBD* e *HDA*) sugeridos por Lipinski (LIPINSKI et al., 1997) e dos correspondentes intervalos de variação calculados^(a) para os 2.756 compostos selecionados pelo modelo MF-Cristal.

Propriedade físico-química	Valores	
	Pela Regra dos 5	Intervalo de variação
Peso Molecular	500	121 – 528
logP calculado	5	-7,64 – 5,7
Número de átomos <i>HBD</i>	5	3 – 10
Número de átomos <i>HBA</i>	10	1 – 17

^(a) Propriedades calculadas utilizando o programa *LigandScout* (WOLBER; LANGER, 2005).

Dos 2756 compostos selecionados, 186 violaram pelo menos um dos quatro critérios da “Regra dos 5”, a saber peso molecular > 500 (19 compostos), LogP calculado > 5 (10 compostos), número de átomos *HBD* > 5 (156 compostos) e número de átomos *HBA* > 10 (33 compostos). De acordo com a literatura (LIPINSKI et al., 2012) os critérios mais comumente violados são peso molecular e logP calculado. No Modelo I - DENV, no entanto, os critérios mais violados foram o número de átomos doadores/aceptores de ligação de hidrogênio. Esta observação pode ser explicada considerando o tamanho (pequeno) e a natureza polar das regiões de interação do modelo MF-Cristal (Figura 4.23 (C)). De acordo com a “Regra dos 5” compostos que violem dois ou mais dos critérios sugeridos por ela, deveriam provavelmente apresentar baixa biodisponibilidade via oral. No entanto, neste primeiro modelo de busca virtual, na tentativa de ser mais rigorosos, optamos por excluir todos os compostos (186) que violaram pelo menos um destes critérios, restando 2.570 compostos para próxima etapa.

Ainda, conforme apresentado na Introdução, compostos com características de fármacos (“*drug-like*”) além de apresentarem uma boa permeabilidade/biodisponibilidade

devem apresentar uma baixa toxicidade (LIPINSKI et al., 2012; OPREA et al., 2001). A toxicidade de um composto pode ser devido, dentre outros fatores, tanto a uma baixa seletividade bem como a presença de grupos funcionais com toxicidade conhecida (MEANWELL, 2011). Com este objetivo, neste primeiro modelo foram retirados derivados de monossacarídeos (hexoses e pentoses) - estruturas com grande número de grupos hidroxilas, que através das muitas ligações de hidrogênio, poderiam acarretar uma baixa especificidade. Assim, com este filtro, 2.433 compostos (redução de 11,72 %) seguiram para próxima etapa.

4.3.2.3. Filtro de Similaridade

Uma primeira análise visual das estruturas dos 2.433 compostos selecionados permitiu observar que muitos compostos apresentavam estruturas químicas semelhantes, apresentando diferenças em regiões que não se sobrepunham ao modelo farmacofórico e que, geralmente, se encontravam voltadas para fora do sítio ativo. Considerando o custo computacional envolvido na etapa de ancoramento e que compostos similares podem apresentar atividades biológicas comparáveis, a terceira etapa da busca virtual foi a aplicação de um filtro de similaridade visando-se excluir compostos similares mantendo, no entanto, a diversidade estrutural (LARSON, 2012). Procedendo-se como descrito no item 3.5.4.4, os compostos similares foram agrupados usando três diferentes medidas de similaridade (**MS1**, **MS2** e **MS3**). O número de grupos gerados usando cada uma das medidas de similaridade está apresentado na Tabela 4.13.

Tabela 4.13. Número de grupos gerados^(a) usando cada uma das três medidas de similaridade: **MS1, **MS2** e **MS3**.**

Medida de similaridade	Número de grupos
MS1 (<i>fingerprints</i> moleculares)	1480
MS2 (descritores moleculares - <i>E-dragon</i>)	453
MS3 (descritores moleculares – <i>Volsurf+</i>)	273

^(a) Aplicando-se um valor de corte de 0,20 de dissimilaridade.

Analisando-se os resultados apresentados na Tabela 4.13, podemos observar que o agrupamento feito com a medida de similaridade **MS1** (*fingerprints* moleculares) gerou o maior número de grupos, dentre eles. Já o agrupamento feito com a medida de similaridade **MS3** (descritores moleculares – *Volsurf+*) gerou o menor número de grupos. Na medida de similaridade **MS1** a diferença estrutural entre dois compostos se baseia unicamente na presença/ausência de diferentes fragmentos, já nas medidas de similaridade **MS2** e **MS3** a diferença entre os compostos depende da diferença entre suas características físico-químicas representadas pelos diferentes descritores moleculares. Nas medidas de similaridade baseadas em descritores moleculares (**MS2** e **MS3**) compostos estruturalmente diferentes, mas com propriedades físico-químicas semelhantes, são colocados em um mesmo grupo. Assim, seria esperado que estas duas medidas de similaridade gerassem um número menor de grupos quando comparados aos da **MS1** (usando *fingerprints* moleculares). Ainda, quando comparamos o número de grupos utilizando os descritores calculados pelos programas *E-dragon* (**MS2**) e *Volsurf+* (**MS3**), a diferença observada pode ser uma consequência tanto da natureza dos descritores quanto ao número diferente de descritores (1.600 e 128, respectivamente) gerados em cada uma das medidas. Desta forma, o maior número de descritores moleculares calculados pelo programa *E-dragon* permite diferenciar mais os compostos refletindo, portanto, em um maior número de grupos. A seguir, um representante de cada grupo foi selecionado.

A análise dos compostos selecionados usando cada uma das três medidas de similaridade permitiu observar que 164 compostos (~9%) foram selecionados pelas três medidas; 810 compostos (~44%) foram selecionados por duas e 868 (~47%) foram selecionados por apenas uma (Figura 4.24). Em seguida, visando analisar se os compostos selecionados por cada uma das medidas de similaridade apresentavam diferentes características físico-químicas, optamos por avaliar as quatro propriedades já calculadas no

filtro “*drug-like*”. A Figura 4.25 mostra a distribuição dos valores de peso molecular, logP calculado, número de átomos *HBD* e *HBA* considerando os compostos selecionados usando as medidas MS1, MS2 e MS3.

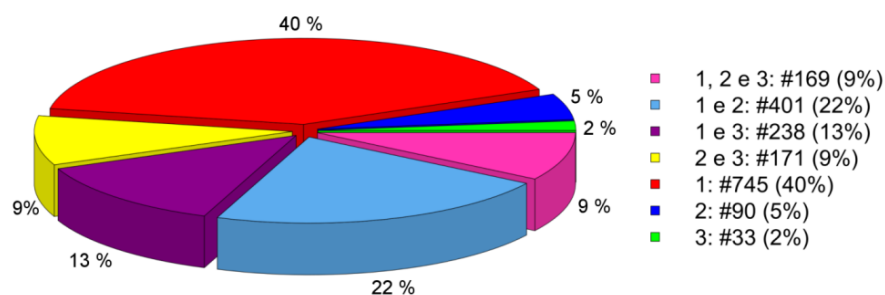


Figura 4.24. Análise dos compostos selecionados utilizando cada uma das medidas de similaridade. Gráfico de setor com os correspondentes valores de porcentagem de compostos selecionados^(a) por uma, duas ou três medidas de similaridade, sendo 1, 2 e 3 as medidas MS1, MS2 e MS3, respectivamente.

(a) Correspondendo as estruturas representativas de cada um dos grupos gerados, com descrito no item 3.5.4.4.

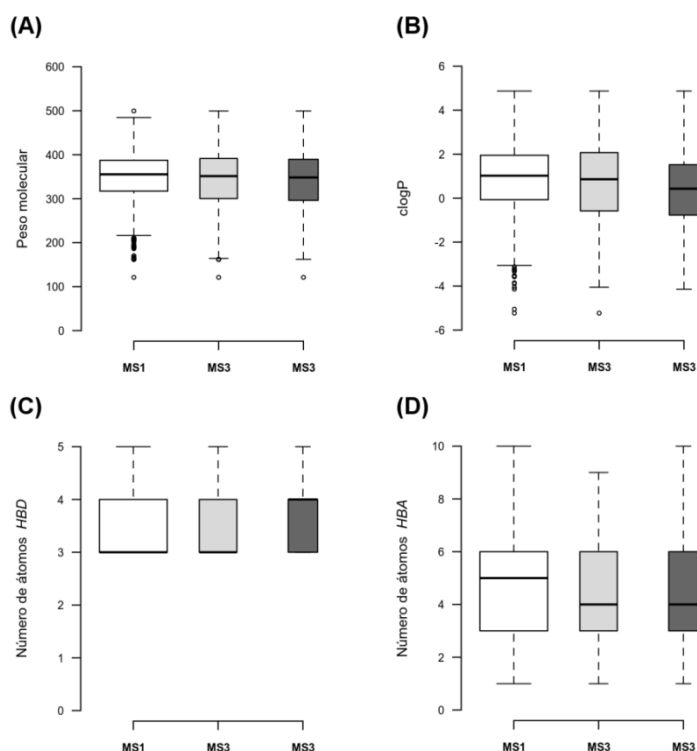


Figura 4.25. Análise comparativa entre os compostos selecionados usando as três medidas de similaridade. Gráficos de caixa^(a) para cada uma das propriedades^(b) físico-químicas usadas no filtro *drug-like* apresentados sempre na sequência MS1 (branco) MS2 (cinza claro) MS3 (cinza escuro): (A) peso molecular; (B) logP calculado; (C) número de átomos *HBD* e (D) número de átomos *HBA*.

(a) Para mais detalhes vide a descrição que acompanha a legenda da Figura 4.2.

(b) Calculadas usando o programa *LigandScout* (WOLBER; LANGER, 2005), vide item 3.5.4.3.

A análise dos gráficos de caixa apresentados na Figura 4.25 permite observar que não há diferença significativa entre a distribuição das propriedades físico-químicas. Assim, os

compostos que foram selecionados usando qualquer uma das medidas de similaridade foram mantidos para próxima etapa. Desta forma, foram selecionados 1.842 compostos (redução de ~24 %) para realizar procedimentos de ancoramento.

4.3.2.4. Ancoramento

Uma discussão detalhada sobre filtros de ancoramento aplicados para seleção de compostos bioativos está apresentada no item 1.8 da Introdução desta tese.

As moléculas selecionadas pelo filtro de similaridade foram, a seguir, ancoradas no sítio ativo da NS2B/NS3pro do DENV. Procedendo-se como descrito no item 3.5.4.9 (Materiais e Métodos), foi realizada uma etapa de validação do ancoramento (“*re-docking*”) usando o inibidor Bz-nKRR-H, a estrutura 3U1I e, inicialmente, as opções padrão da função de ranqueamento. No entanto, a pose de melhor *fitness* obtida para este inibidor não apresentou o mesmo modo de ligação observado na estrutura 3U1I (Figura 4.26(A)).

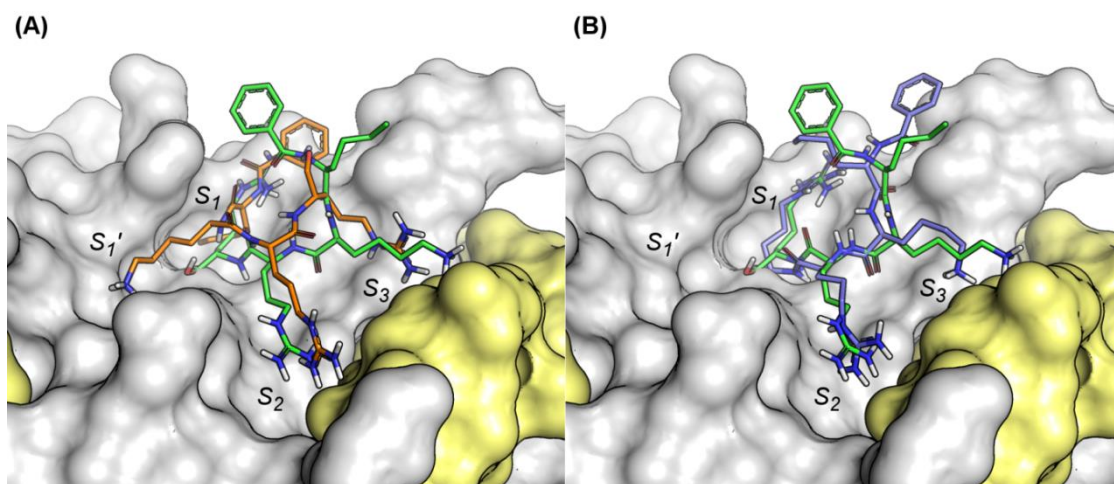


Figura 4.26. Validação do procedimento de ancoramento sobre a cavidade da NS2B/NS3pro do DENV. (A) Pose de melhor *fitness*^(a) (laranja) obtida sobreposta ao modo de ligação do inibidor Bz-nKRR-H observado na estrutura 3U1I (verde). (B) Pose de melhor *fitness*^(b) (azul) obtida após as modificações na função de ranqueamento sobreposta ao modo de ligação do inibidor Bz-nKRR-H observado na estrutura 3U1I (verde). A proteína e o inibidor Bz-nKRR-H estão representados em superfície e *sticks*, respectivamente. A região NS2B em amarelo e o domínio NS3pro em cinza. Figura preparada com o programa *PymOL* (SCHRÖDINGER, 2013).

^(a) Obtida usando a função de ranqueamento *ChemPLP* e o programa *GOLD* (VERDONK et al., 2003)

^(b) Obtida usando a função de ranqueamento *ChemPLP*, com modificações, e o programa *GOLD* (VERDONK et al., 2003)

Nesta pose, o grupo benzoil ocupa o sítio S_1 (posição que deveria ser ocupada pela *Arg-P₁*) e a *Arg-P₁* ocupa a cavidade S_3 (que deveria ser ocupado pela *Lys-P₃*). Ainda, o valor

do *RMSD* entre a pose do inibidor Bz-nKRR-H obtida por ancoramento e observada na estrutura 3U1I foi igual a 7,7 Å. Com base nestes resultados, este ancoramento não foi considerado validado. Estes resultados podem ser devido, entre outros fatores, a menor taxa de acerto dos programas de ancoramento em posicionar corretamente ligantes com alta flexibilidade (ou seja, com grande número de ligações com livre rotação) (BOSTRÖM; NORRBY; LILJEFORS, 1998; LARSON, 2012). Assim, visando favorecer o posicionamento correto deste inibidor no sítio ativo da NS2B/NS3pro do DENV foi necessário introduzir algumas mudanças na função de ranqueamento e incluir uma ligação covalente entre ele e a proteína. Estas mudanças na função de ranqueamento consistiram em indicar alguns átomos dos resíduos de aminoácido do sítio ativo como favoráveis para formar ligações de hidrogênio (aceptoras e/ou doadoras) com o inibidor (para uma descrição completa destas mudanças vide item 3.5.4.9). A inclusão das mudanças na função de ranqueamento e a formação da ligação covalente produziram resultados significativamente melhores que os obtidos, anteriormente, usando os parâmetros padrões do programa. Na pose de melhor *fitness* obtida para o inibidor Bz-nKRR-H, os resíduos *Arg-P₁*, *Arg-P₂* e *Lys-P₃* foram posicionados de modo semelhante ao observado na estrutura 3U1I, ocupando as correspondentes cavidades *S₁*, *S₂* e *S₃*. Ainda, o valor de *RMSD* obtido foi igual a 3,18 Å. O resíduo norleucina (*n-P4*) e o grupo benzoil (Bz) do inibidor, entretanto, adotam uma orientação invertida em relação à observada na estrutura 3U1I (Figura 4.26 (B)). Esta diferença pode explicar o alto valor de *RMSD* obtido. A inspeção visual da estrutura 3U1I mostrou que o resíduo *n-P4* e o grupo Bz estão expostos ao solvente formando contatos com uma segunda molécula da célula unitária, devido ao efeito do empacotamento cristalino (NOBLE et al., 2012). Considerando que estas posições experimentais podem ser um artefato da cristalização ausente no ancoramento, parece razoável propor que duas regiões do inibidor expostas ao solvente e de mesma natureza — ambas hidrofóbicas — poderiam adotar orientações invertidas às observadas no cristal. Com

base no descrito acima, o valor do *RMSD* foi calculado, agora, apenas para os resíduos P_1 , P_2 e P_3 do inibidor sendo igual a 1,32 Å. Assim, o ancoramento foi considerado validado e, então, aplicado aos 1.842 compostos previamente selecionados pelo filtro de similaridade. Cabe ressaltar que nestes ancoramentos todas as mudanças na função de ranqueamento foram mantidas, exceto pela formação da ligação covalente.

Tendo em vista que o primeiro filtro aplicado foi o modelo farmacofórico MF-Cristal (Figura 4.23) e que a construção deste modelo foi baseada em dados experimentais (estrutura 3U1I), neste filtro de ancoramento foram selecionados apenas os compostos cujas poses geradas satisfizessem todas as interações preditas pelo modelo MF-Cristal. Procedendo-se como descrito no item 3.5.4.9 (Materiais e Métodos), a análise dos descritores de ligação de hidrogênio mostrou que 241 compostos tiveram pelo menos uma pose capaz de satisfazer todas as interações preditas pelo modelo farmacofórico. Estes 241 compostos (redução de 86,92 %) foram selecionados e, em seguida, submetidos a uma análise por inspeção visual.

4.3.2.5. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV

Uma discussão detalhada sobre a necessidade de se realizar a inspeção visual dos resultados obtidos pelo ancoramento está apresentada no item 1.8 da Introdução desta tese.

Procedendo-se como descrito no item 3.5.4.9 (Materiais e Métodos), os 241 compostos selecionados no filtro de ancoramento foram analisados, por inspeção visual, para verificar seus ajustes no sítio ativo da NS2B/NS3pro do DENV e se suas poses se mostravam reproduzíveis. Nesta inspeção visual, foi possível observar que além dos resíduos utilizados para construir o modelo MF-Cristal (*Asp129*, *Phe130*, *Gly151*, *Gly153* e *Try161*) os compostos interagiram, também, com outros resíduos do sítio ativo, a saber: (i) *His51*, *Asp75* e *Ser135* (tríade catalítica); (ii) *Val72*, *Lys131*, *Gly133*, *Thr134*, *Try150*, *Asn152*, *Val154* e *Val155* (domínio NS3pro) e (iii) *Arg54*, *Asp81*, *Thr83*, *Met84*, *Arg85* e *Ile86* (região NS2B).

Os compostos com pelo menos três poses capazes de interagir com os resíduos da tríade catalítica e/ou com os resíduos *Met84* e *Thr83* da região NS2B – sugeridos pela literatura como responsáveis pelas diferenças de especificidade entre as N2SB/NS3pro do DENV e do WNV (NOBLE et al., 2012) – foram priorizados nesta seleção visual.

Nesta etapa, 46 compostos foram selecionados como potenciais inibidores da NS2B/NS3pro do DENV, representando uma redução de 80,09 % em relação aos 241 compostos selecionados na etapa anterior. Destes 46 potenciais inibidores, 10 (21,74 %) estavam disponíveis e foram adquiridos para validação experimental do Modelo I - DENV. Por fim, a aplicação de todos os filtros de seleção representou uma redução total de mais de 99 % em relação ao banco *ZINC-Original* ($\sim 23 \times 10^6$ compostos). A Tabela 4.14 mostra a representação das estruturas dos dez compostos (denominados **DV-01** a **DV-10**) selecionados pelo Modelo I – DENV que foram adquiridos e submetidos a testes de inibição enzimática frente à NS2B/NS3pro do DENV.

Tabela 4.14. Representação da estrutura 2D dos 10 compostos (DV-01 a DV-10) selecionados pelo Modelo I – DENV como potenciais de inibidores da NS2B/NS3pro do DENV e que foram adquiridos e submetidos ao teste de inibição enzimática. Estas estão acompanhadas dos seus correspondentes códigos ZINC, valores de peso molecular e empresa fornecedora.

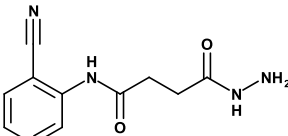
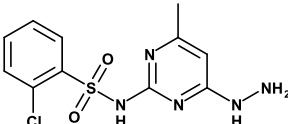
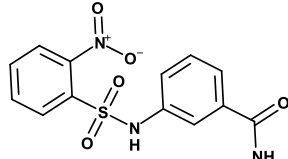
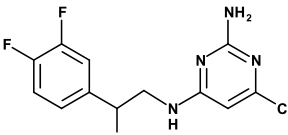
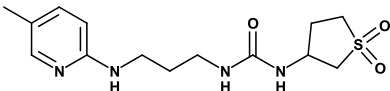
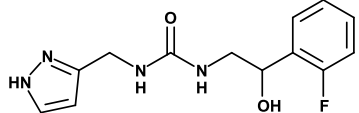
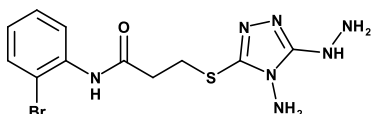
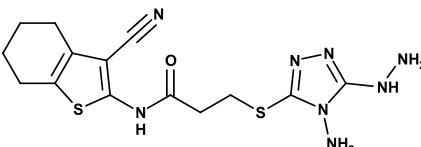
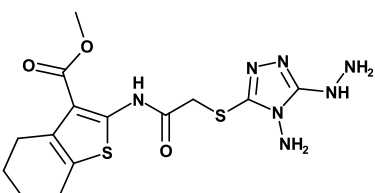
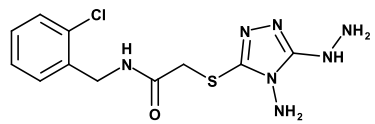
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-01		ZINC01654187	232,34	NCI
DV-02		ZINC00270493	313,76	Vitas M Labs
DV-03		ZINC6214069	322,32	Vitas M Labs
DV-04		ZINC29588623	301,70	Enamine

Tabela 4.14. (Continua)

#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-05		ZINC44966262	327,42	Enamine
DV-06		ZINC48260384	278,28	Enamine
DV-07		ZINC20922978	373,25	Scientific Exchange
DV-08		ZINC20922301	379,48	Scientific Exchange
DV-09		ZINC20922428	398,48	Scientific Exchange
DV-10		ZINC21035463	328,80	Scientific Exchange

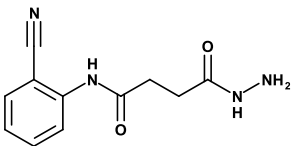
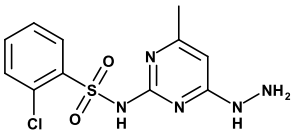
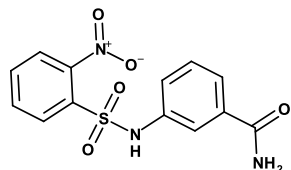
4.3.2.6. Validação experimental

Como apresentado e discutido no item 1.4 da Introdução desta tese, por melhor e mais adequados que tenham sido os filtros usados para construir o modelo de busca virtual, a verdadeira utilidade deste modelo na busca de compostos bioativos só pode ser avaliada e/ou comprovada após a sua validação experimental, através, por exemplo, de ensaios enzimáticos. Assim, os 10 compostos selecionados pelo Modelo I - DENV foram adquiridos e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV2. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Wibke Diederich (Philipps-Universität de Marburg, Marburg, Alemanha) e estão descritos no item 3.6.1.

Ensaio de inibição enzimática frente à NS2B/NS3pro do DENV

Como proposto no item 3.6.1, somente compostos com valores de atividade inibitória maiores, ou próximos a 50%, foram considerados como tendo atividade inibitória significativa. Assim, dos 10 compostos adquiridos, dois mostraram atividade inibitória significativa frente à NS2B/NS3pro do DENV2 na concentração utilizada (100 μ M), sendo eles: compostos **DV-08** (ZINC20922301) e **DV-09** (ZINC20922428). Os valores de porcentagem de inibição dos compostos **DV-08** e **DV-09** (em 100 μ M) foram iguais a 44 ± 8 e 53 ± 11 , respectivamente. Por outro lado, oito deles não foram considerados ativos na concentração usada, sendo estes: compostos **DV-01** (ZINC01654187), **DV-02** (ZINC00270493), **DV-03** (ZINC6214069), **DV-04** (ZINC29588623), **DV-05** (ZINC44966262), **DV-06** (ZINC48260384), **DV-07** (ZINC20922978) e **DV-10** (ZINC21035463). A Tabela 4.15 mostra os valores de porcentagem de inibição (em 100 μ M) obtidos para cada um dos 10 compostos testados e os correspondentes desvios padrões.

Tabela 4.15. Valores de porcentagem de inibição (% , em 100 μ M)^(a) e dos correspondentes desvios padrão^(b) de cada um dos compostos selecionados pelo Modelo I – DENV.

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M)
DV-01		14 ± 6
DV-02		15 ± 8
DV-03		2 ± 2

^(a) Em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a. Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

Tabela 3.15. (Continua).

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M)
DV-04		15 ± 3
DV-05		10 ± 3
DV-06		10 ± 9
DV-07		26 ± 6
DV-08 ^(c)		44 ± 8
DV-09 ^(c)		53 ± 11
DV-10		20 ± 11

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Para estes inibidores **DV-08** e **DV-09**, os valores de IC₅₀ foram determinados a partir de suas curvas dose-resposta, construídas como descrito no item 3.6.1.2 e apresentadas na Figura 4.27. Estes foram iguais a 266 ± 13 e a 195 ± 36 μ M para os inibidores **DV-08** e **DV-09**, respectivamente.

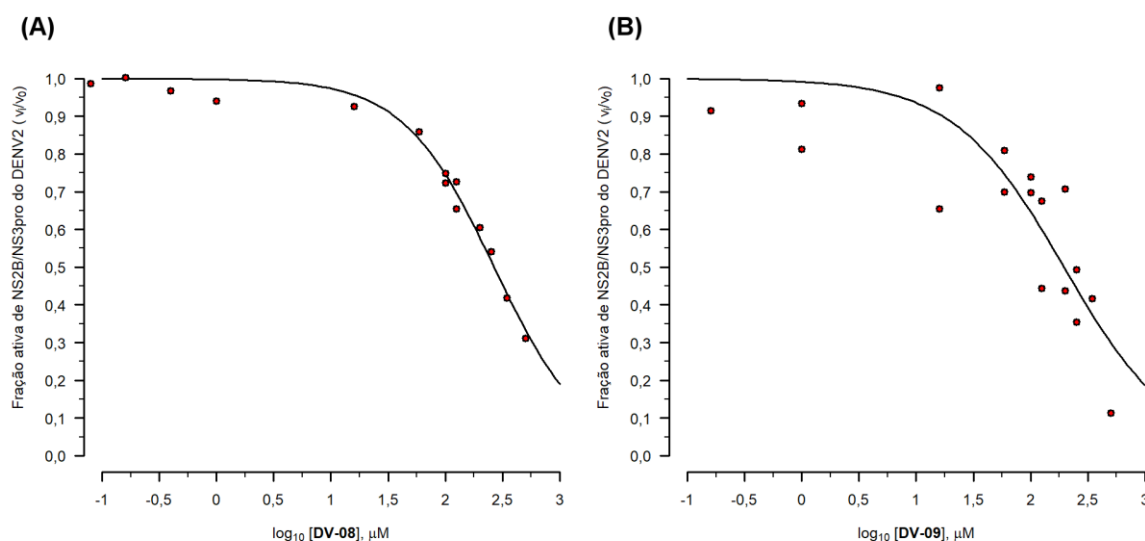


Figura 4.27. Curvas dose-resposta^(a) para os compostos DV-08 e DV-09. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: **DV-08 (A)** e **DV-09 (B)**. Experimentos realizados uma única vez ou em duplicata

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa *R* (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: **DV-08** ou **DV-09**.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM.
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a. Diederich) e PhAc-KRR-MCA (20 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{ nm}$.

Embora os compostos **DV-08** e **DV-09** tenham apresentado uma certa atividade inibitória frente à NS2B/NS3pro do DENV2, seus valores de IC₅₀ foram consideravelmente altos (> 100 μM). Adicionalmente, os graus de pureza destes inibidores **DV-08** e **DV-09** não foram avaliados devido a sua baixa atividade inibitória.

Avaliação da estabilidade dos inibidores nas condições do experimento de inibição frente à da NS2B/NS3pro do DENV2

O inibidor **DV-09** se mostrou estável durante os experimentos de inibição. Por outro lado, foi observada uma mudança de cor significativa, de marrom para rosa, nas soluções estoque do inibidor **DV-08** (em DMSO). Assim, fez-se necessário analisar a estabilidade deste inibidor nas condições do experimento de inibição (solução tampão pH 9,0 [Tris Base (50 mM) e CHAPS (1 mM)]). Nesta tese, procedendo-se como descrito no item 3.5.6.1, a análise

da estabilidade foi feita acompanhando alterações nos correspondentes espectros de absorção, durante 5 horas.

Assim, a partir do espectro de absorbância da solução do inibidor **DV-08** (197 μM em tampão pH =9,0), dois comprimentos de onda máximos ($\lambda_1^{\text{max}} = 313 \text{ nm}$ e $\lambda_2^{\text{max}} = 380 \text{ nm}$) foram selecionados e usados para analisar as estabilidades. A Tabela 4.16 mostra os valores de absorbância do inibidor **DV-08** registrados a cada hora, durante 5 horas, para cada um dos comprimentos de onda selecionados.

Tabela 4.16. Valores de absorbância observados para a solução de DV-08 (197 μM em tampão^(a) pH = 9,0), registrados a cada hora durante 5 horas, em $\lambda_1^{\text{max}} = 313 \text{ nm}$ e $\lambda_2^{\text{max}} = 380 \text{ nm}$.

Tempo (horas)	Valores de absorbância	
	λ_1^{max} (313 nm)	λ_2^{max} (380 nm)
0	1,714	0,074
1	1,703	0,103
2	1,678	0,128
3	1,661	0,152
4	1,651	0,167
5	1,631	0,164

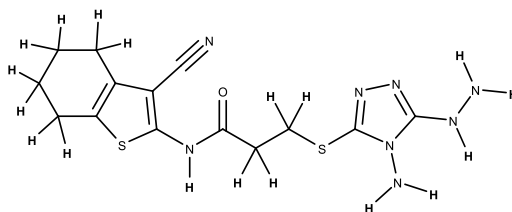
^(a) Solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)].

Considerando-se o erro sistemático do equipamento utilizado ($\pm 0,004$ e $\pm 0,002$ para valores de absorbância $> 0,5$ e $< 0,5$, respectivamente), as variações observadas nos valores de absorbância não foram significativas ($\Delta = 0,011$ para λ_1^{max}) para tempos até 1 hora (tempo máximo utilizado nos experimentos de inibição). No entanto, após 1 hora de experimento variações significativas ($\Delta = 0,036$ a $0,083$ e $\Delta = 0,054$ a $0,093$ para λ_1^{max} e λ_2^{max} , respectivamente) foram observadas. De acordo com este critério de análise, pode-se considerar que o inibidor **DV-08** mostrou-se estável apenas durante o tempo do experimento (≤ 1 hora), o que não ocorria em tempos maiores.

Para os tempos maiores, considerando a baixa estabilidade observada para **DV-08**, fez-se necessário verificar se a mudança de cor estava associada a qualquer mudança estrutural. Assim, procedendo-se como descrito no item 3.5.6.2 (Materiais e Métodos), a identificação e/ou caracterização do inibidor **DV-08** bem como dos seus possíveis produtos de degradação foi feita pela análise dos espectros de RMN¹H de uma solução do inibidor **DV-08** registrados nos diferentes intervalos de tempo (t = 0 e t = 3 dias) onde se observava mudança de cor. Os valores de deslocamento químico (δ , ppm) bem como os valores das constantes de acoplamento J (Hz) observados nestes espectros estão mostrados na Tabela 4.17.

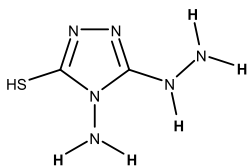
As atribuições dos sinais NH do **DV-08** foram confirmadas comparando-as com aquelas observadas para o reagente de Purpald®, que contém parte da estrutura do inibidor **DV-08**. Os valores de deslocamento químico (δ) observados nos espectros de RMN¹H, registrados nas mesmas condições, para este reagente estão apresentados na Tabela 4.18.

Tabela 4.17. Valores dos deslocamentos químicos (δ , ppm) relativos ao TMS ($\delta = 0$ ppm) e das constantes de acoplamento J (Hz) obtidos nos espectros de RMN¹H (DMSO-d₆ 99,9 %) para o inibidor **DV-08** (3-[(4-amino-5-hidrazino-1,2,4-triazol-3-il)sulfanil]-N-(3-ciano-4,5,6,7-tetrahidrobenzo-tiofen-2-il) nos diferentes tempos onde se observava mudança de cor (t = 0 e t = 3 dias).



Espectro de RMN ¹ H	δ (ppm) e J (Hz)
t = 0 (cor marrom)	12,26 (s, 1H, CONH); 7,15 (s, 1H, (NH)); 5,27 (s, 2H, NH-NH ₂); 4,09 (s, 2H, NNH ₂); 4,03 (d, 2H, SCH ₂ , J=4,64); 3,27 (d, 2H, COCH ₂ CH ₂ , J=4,64); 2,61 – 2,58 (m, 2H, SCCH ₂); 2,47 – 2,45 (m, 2H, CNCCCCH ₂); 1,76 -1,74 (m, 4H, (CH ₂ CH ₂) ₂). Possíveis Impurezas: 2,89 (s, 0,16H); 2,73 (s, 0,21H);
t = 3 dias (cor rosa)	12,68 (s, 1H, CONH); 7,15 (s, 1H, (NH)); 5,27 (s, 2H, NH-NH ₂); 4,09 (s, 2H, NNH ₂); 4,03 (d, 2H, SCH ₂ , J=4,73); 3,26 (d, 2H, COCH ₂ CH ₂ , J=4,73); 2,61 – 2,57 (m, 2H, SCCH ₂); 2,47 – 2,45 (m, 2H, CNCCCCH ₂); 1,76 -1,74 (m, 4H, (CH ₂ CH ₂) ₂). Possíveis Impurezas: 8,44 (s, 0,25H); 5,68 (s, 0,41H); 2,89 (s, 0,16H); 2,73 (s, 0,21H)

Tabela 4.18. Valores de deslocamentos químicos (δ , ppm) relativos ao TMS ($\delta = 0$ ppm) obtidos no espectro de RMN¹H (DMSO-d₆ 99,9 %) para o reagente de Purpald® (4-Amino-3-hidrazino-5-mercaptop-1,2,4-triazol).



Composto	δ (ppm)
Purpald®	7,15 (s, 1H, NH); 5,27 (s, 2H, NHNH ₂); 4,09 (s, 2H, NNH ₂)

Deste modo, pode-se confirmar que os sinais NH para o **DV-08** foram atribuídos de forma correta. Ainda mais, pela Tabela 4.17, observa-se que os espectros de RMN¹H do inibidor **DV-08** registrados em tempos diferentes ($t = 0$ e $t = 3$ dias) se mostraram concordantes com a estrutura proposta, entretanto, alguns sinais referentes às possíveis impurezas foram observados. Para o espectro registrado após 3 dias, foi observado o aparecimento de dois novos sinais referentes à possíveis impurezas que não haviam sido observadas de início, sugerindo que estes sejam responsáveis pela mudança de cor observada para o inibidor **DV-08**. Nenhuma das possíveis impurezas observadas nestes espectros pôde ser identificada.

Pelas análises feitas, conclui-se que o inibidor **DV-08** apresenta estabilidade nas condições do experimento até uma hora não sendo possível, no entanto, identificar a(s) porção estrutural(is) do inibidor **DV-08** responsável(is) pela mudança de cor.

Tanto à baixa atividade inibitória dos compostos **DV-08** e **DV-09** bem como a falta de estabilidade do inibidor **DV-08** sugerem que o modelo de busca virtual (Modelo I - DENV) precisava ser melhorado. Assim, visando melhorar o modelo de busca virtual duas alterações foram propostas e realizadas: (1) criação de outro banco de dados de compostos (“*ZINC-Curated*”), removendo os compostos que apresentaram propriedades e/ou estruturas indesejadas para um composto com característica de fármaco e (2) inclusão da flexibilidade

do sítio ativo da NS2B/NS3pro do DENV, em especial dos resíduos da região NS2B. Estes aspectos foram considerados e serão apresentados a seguir, no decorrer desta tese (itens 4.3.3 a 4.3.6).

4.3.3. Estudos da flexibilidade conformacional da NS2B/NS3pro do DENV por simulações de Dinâmica Molecular (DM)

Considerando-se que a região NS2B da NS2B/NS3pro do DENV apresenta flexibilidade (vide item 4.3.1.3), somente a partir do segundo modelo de busca virtual (apresentado no item 4.3.4), tentou-se incluir a flexibilidade conformacional desta protease, considerando, em especial, os resíduos do sítio ativo e da região NS2B. Uma das ferramentas para se estudar a flexibilidade conformacional de proteínas são as simulações de Dinâmica Molecular (DM), para mais detalhes vide item 1.10. Assim, simulações de DM foram feitas para explorar a flexibilidade conformacional dos resíduos do sítio ativo, em especial, os da região NS2B. Todas as simulações de DM foram realizadas durante uma visita de três meses aos laboratórios da Prof. Christoph A. Sotriffer (Universidade de Würzburg, Alemanha) e estão descritos no item 3.5.4.8 (Materiais e Métodos). Nestas simulações foram utilizadas três diferentes estruturas iniciais, a saber: a primeira corresponde à estrutura 3U1I inalterada, contendo a *Ser135* covalentemente ligada ao inibidor Bz-nKRR-H (denominada “estrutura covalente”, Figura 4.28 (A)); a segunda foi obtida removendo completamente o inibidor Bz-nKRR-H (denominada “estrutura livre”) e, a terceira foi obtida mantendo o inibidor Bz-nKRR-H no sítio ativo, mas clivando a ligação covalente entre este e a *Ser135* (denominada “estrutura não-covalente”, Figura 4.28 (B)). Parte do material descrito nos próximos parágrafos faz parte de artigo recentemente publicado no *JCAMD*, do qual sou primeira autora (PICCIRILLO et al., 2016), colocado como anexo.

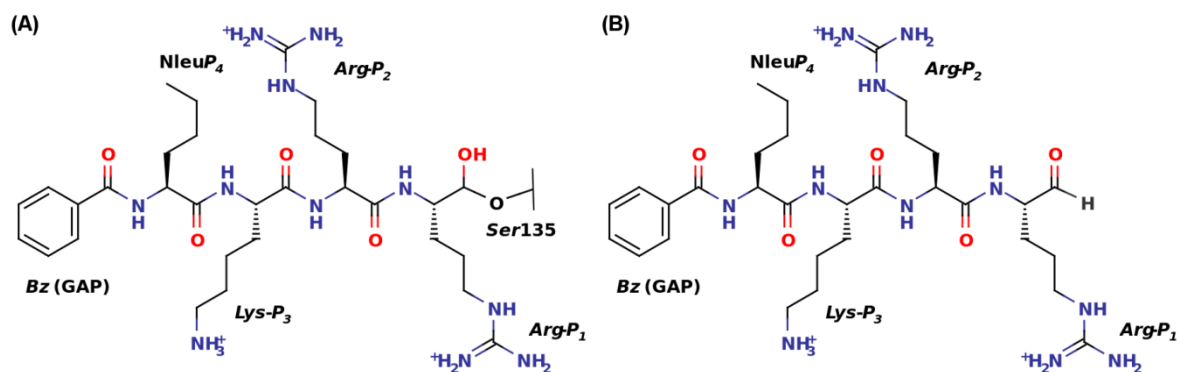


Figura 4.28. Representação 2D da estrutura do inibidor Bz-nKRR-H: (A) inibidor covalentemente ligado à *Ser135* (simulação de DM da estrutura “covalente”) e (B) inibidor livre (simulação de DM da estrutura “não-covalente”). Os resíduos do inibidor foram denominados P_1 a P_4 , no sentido N-terminal, seguindo a nomenclatura usual dos substratos de protease.

Visando analisar a flexibilidade dos resíduos do sítio ativo da NS2B/NS3pro do DENV, fez-se necessário definir os resíduos que fazem parte deste sítio. Seguindo o critério descrito no item 3.5.4.8, o sítio ativo da NS2B/NS3pro do DENV contém 17 resíduos sendo estes: *Asp81*, *Gly82*, *Thr83*, *Met84* e *Arg85* da região NS2B e *His51*, *Asp75*, *Asp129*, *Phe130*, *Ser135*, *Tyr150*, *Gly151*, *Asn152*, *Gly153*, *Val154*, *Val155* e *Tyr161* do domínio NS3pro (Figura 4.29).

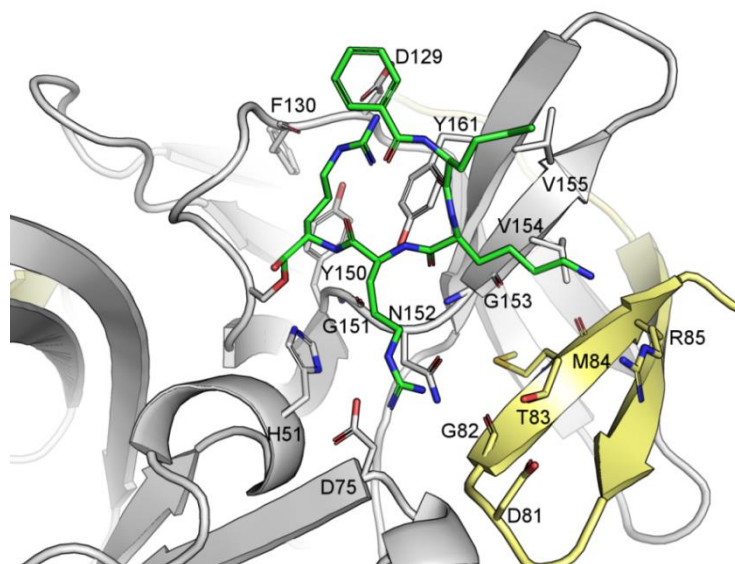


Figura 4.29. Estrutura 3U1I (NS2B/NS3pro do DENV3) complexada com o inibidor Bz-nKRR-H mostrando os resíduos do sítio ativo^(a). A cadeia principal está mostrada na forma de “cartoon” e tanto a cadeia lateral dos resíduos do sítio ativo (átomos de carbono em cinza e em amarelo diferenciando os resíduos da região NS2B e dos resíduos do domínio NS3pro) bem como do inibidor Bz-nKRR-H (átomos de carbono em verde) estão mostrados na forma de “sticks”. Figura retirada de Piccirillo, *et. al* 2016.

^(a) Definidos nesta tese como sendo todos os resíduos da estrutura 3U1I que apresentaram pelo menos três átomos distantes 4 Å do inibidor Bz-nKRR-H.

4.3.3.1. Agrupamento hierárquico dos *snapshots* das simulações de DM e abordagem *GRID/PCA*

Como descrito na literatura (SCHAEFER; KISKER; SOTRIFER, 2011), a flexibilidade dos resíduos do sítio ativo foi analisada através do agrupamento hierárquico baseado nos valores de *RMSD*. O número de grupos foi gerado considerando-se um valor de corte de *RMSD* igual a 3 Å, que resultou em cinco grupos (G1 a G5). O grupo G1 contém estados conformacionais obtidos nas simulações das estruturas “*covalente*”, “*livre*” e “*não-covalente*” (10%, 17% e 35%, respectivamente). O grupo G2 contém a maioria dos estados obtidos nas simulações das estruturas “*covalente*” e “*não-covalente*” (84 e 55%, respectivamente), além de alguns estados obtidos na simulação da estrutura “*livre*” (18%). De modo análogo, o grupo G3 contém a maioria dos estados obtidos na simulação da estrutura modificada (47%) e, alguns poucos estados obtidos nas simulações das estruturas “*covalente*” e “*não-covalente*” (6 e 10%, respectivamente). Já os grupos G4 e G5 contêm apenas os estados obtidos na simulação da estrutura “*livre*” (9% cada). Para cada grupo, o *snapshot* central (Conf-1 a Conf-5) foi escolhido como estrutura representativa do correspondente grupo e, utilizado nas análises subsequentes bem como na inspeção visual. Na análise das estruturas representativas dos grupos (Conf-1 a Conf-5) foi observado que os resíduos da região NS2B (*Asp81*, *Thr83*, *Met84* e *Arg85*) adotam conformações diferentes das observadas na estrutura 3U1I (Figura 4.30). Já os resíduos do domínio NS3pro (*His51*, *Asp75*, *Asp129*, *Phe130*, *Ser135*, *Tyr150*, *Gly151*, *Asn152*, *Gly153*, *Val154*, *Val155* e *Tyr161*) adotam conformações próximas às observadas na estrutura 3U1I, sendo os resíduos *Asp129*, *Val155* e *Tyr161* os mais flexíveis. A flexibilidade observada para os resíduos da região NS2B envolve tanto a cadeia lateral bem como a cadeia principal, enquanto que a flexibilidade dos resíduos *Asp129*, *Val155* e *Tyr161* do domínio NS3pro envolve apenas a cadeia lateral. Cabe ressaltar que embora a região NS2B apresente flexibilidade, nenhuma das estruturas representativas

esta região deixou de interagir com a parte central do domínio NS3pro. De modo análogo ao descrito na literatura (DE LA CRUZ et al., 2014), as mudanças conformacionais observadas durante as simulações de DM correspondem às mudanças na conformação “fechada” e não ao equilíbrio entre as conformações “aberta” e “fechada”.

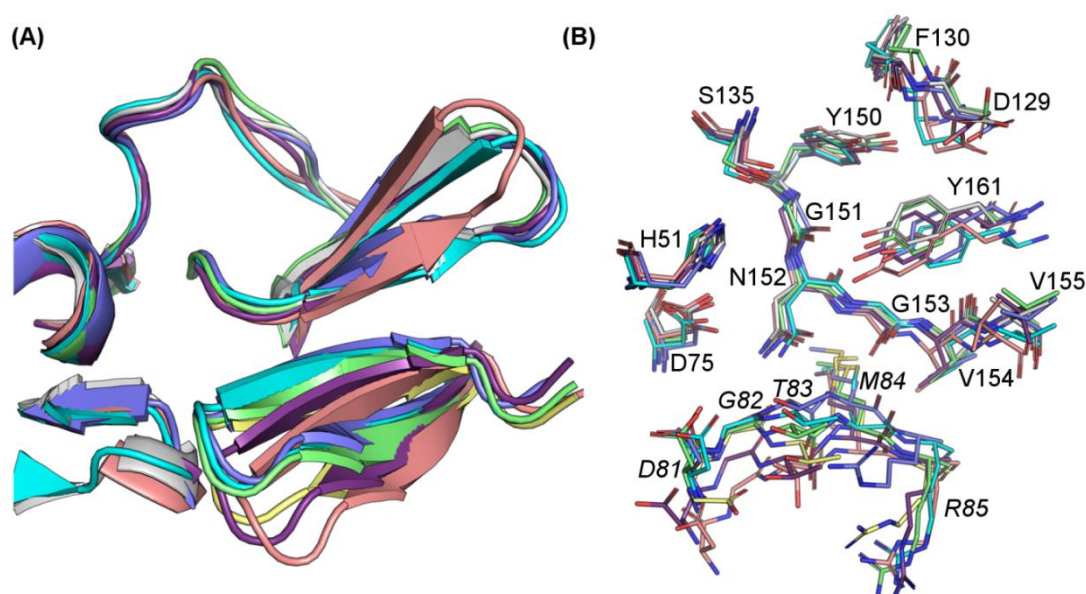


Figura 4.30. Estrutura 3U1I (NS2B/NS3pro do DENV3) sobreposta à estrutura representativa^(a) (Conf-1 a Com-5) de cada grupo^(b) (G1 a G5). (A) Cadeia principal da estrutura 3U1I (região NS2B e domínio NS3pro em amarelo e cinza, respectivamente) e das estruturas representativas (Conf-1 a Conf-5 em ciano, verde, azul, roxo e rosa, respectivamente) mostradas na forma de “cartoon”. (B) Resíduos do sítio ativo^(c) da estrutura 3U1I (região NS2B e domínio NS3pro átomos de carbono em amarelo e cinza, respectivamente) e das estruturas representativas (Conf-1 a Conf-5 átomos de carbonos em ciano, verde, azul, roxo e rosa, respectivamente) mostrados na forma de “sticks”. Figura retirada de Piccirillo, *et. al* 2016.

^(a) Snapshot central do correspondente grupo.

^(b) Grupos obtidos através do agrupamento hierárquico usando os valores de *RMSD* e um valor de corte de 3 Å.

^(c) Definidos nesta tese como sendo todos os resíduos da estrutura 3U1I que apresentaram pelo menos três átomos distantes 4 Å do inibidor Bz-nKRR-H.

Visando-se analisar se todas estas cinco estruturas representativas precisariam ser incluídas em nossos modelos de busca virtual, as diferenças estruturais entre a estrutura 3U1I (cristal) e cada uma das estruturas representativas (Conf-1 a Conf-5) foram analisadas. Esta análise foi realizada através da abordagem *GRID/PCA* (PASTOR; CRUCIANI, 1995) como descrito no item 3.5.4.8 (Materiais e Métodos). As Componentes Principais (CP) 1, 2 e 3 explicaram respectivamente 24, 16 e 13% da variância dos dados, respectivamente. No

gráfico de escore das CP1, CP2 e CP3 (Figura 4.31), a CP1 ranqueou as energias de interação das sondas moleculares enquanto que as CP2 e CP3 separaram/distinguiram as seis estruturas.

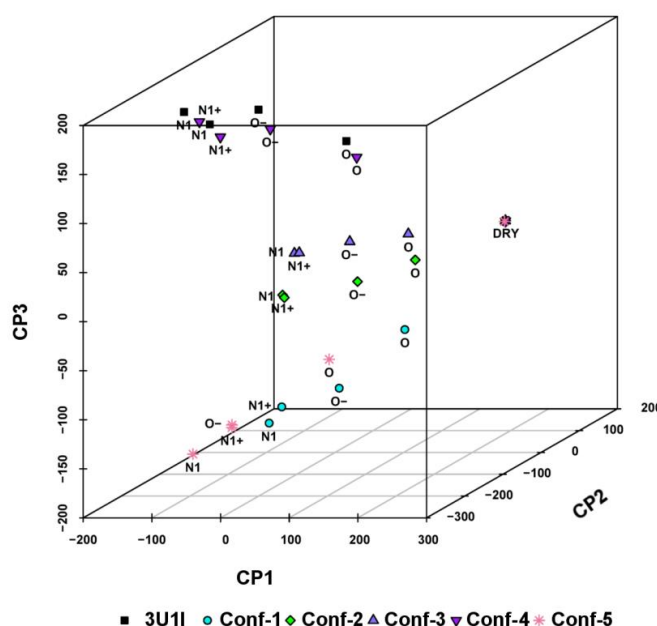


Figura 4.31. Gráfico de escore^(a) das CPs 1, 2 e 3 para as energias dos campos de interação calculados para as seis estruturas (3UII, Conf-1, Conf-2, Conf-3, Conf-4 e Conf-5). Cada ponto representa a interação entre a estrutura e a sonda molecular. Os quadrados em preto, círculos em ciano, losangos em verde, triângulos em azul marinho, triângulos em roxo e asteriscos em rosa representam as interações com as estruturas 3UII, Conf-1, Conf-2, Conf-3, Conf-4 e Conf-5, respectivamente. Figura retirada de Piccirillo, *et. al* 2016.

(a) Obtidos pela ACP das energias dos campos de interação molecular calculados para cada estrutura usando o programa *GRID* (GOODFORD, 1985) e cinco diferentes sondas moleculares ("DRY", "N1+", "N1", "O" e "O-").

Neste gráfico, as sondas moleculares “N1” e “N1+” apresentaram os menores valores de CP1, indicando uma forte interação entre estas sondas e os resíduos do sítio ativo. Já a sonda “DRY”, representando interações hidrofóbicas, apresenta os maiores valores de CP1. Considerando a grande densidade de átomos *HBA* no sítio ativo da NS2B/NS3pro do DENV (vide Figura 4.18), era esperado que sondas representando grupos *HBD* (“N1” e “N1+”) apresentassem uma forte interação com os resíduos do sítio ativo da NS2B/NS3pro do DENV. Além disso, a sonda “N1” apresentou valores de CP1 similares ou menores do que os valores observados para a sonda “N1+”. A princípio, estes resultados não eram esperados, uma vez que a NS2B/NS3pro do DENV apenas reconhece substratos com resíduos básicos nas posições P_1 e P_2 bem como pelo fato desta protease ter resíduos ácidos nas cavidades S_1 e S_2 .

(*Asp129* na cavidade S_1 e *Asp75* e *Asp81*-NS2B na cavidade S_2). No entanto, cabe relembrar que as cavidades do sítio ativo desta protease são rasas e que o resíduo *Asp129* está exposto ao solvente, o que pode fazer com que os valores de energia da interação iônica com os resíduos ácidos do sítio ativo desta protease sejam menores. Ainda, os valores de CP2 e CP3 para as estruturas representativas Conf-2 e Conf-3 foram parecidos. Estes resultados indicam que embora estas estruturas pertençam a grupos diferentes (G2 e G3), os campos de interação moleculares calculados para ambas as estruturas são semelhantes. A estrutura Conf-2 foi escolhida como representativa dos grupos G2 e G3, porque esta representa o grupo G2 que contém o maior número de estados conformacionais. De modo análogo, a estrutura representativa Conf-4 apresentou valores de CP2 e CP3 comparáveis com os observados para a estrutura 3U1I. No entanto, os valores de CP2 e CP3 das estruturas representativas Conf-1 e Conf-5 foram significativamente diferentes entre si e entre os valores observados para as demais estruturas, exceto pela sonda “DRY”. Adicionalmente, os campos de interação molecular calculados para a estrutura Conf-2 são claramente diferentes dos calculados para a estrutura 3U1I, mostrando diferenças que podem ser importantes para a interação ligante-proteína. Entre estas diferenças, a inspeção visual dos campos de interação destas duas estruturas (3U1I e Conf-2) indicou que cavidade S_2 da estrutura Conf-2 é mais favorável às interações de ligação de hidrogênio quando comparada a da estrutura 3U1I. Esta diferença se deve ao fato de que nas estruturas Conf-2 a cadeia lateral do *Asp81* da região NS2B estava voltada para cavidade S_2 , já na estrutura 3U1I a cadeia lateral deste resíduo está exposta ao solvente. Na estrutura Conf-4 a cadeia lateral do *Asp81* da região NS2B está um pouco distante da porção central do domínio NS3pro, adotando uma orientação próxima à observada na estrutura 3U1I. Desta forma, os campos de interação molecular calculados para estas duas estruturas são semelhantes. Visando incluir estas diferenças estruturais observadas nos

modelos de busca virtual, quatro estados conformacionais diferentes (representados pelas estruturas 3U1I, Conf-1, Conf-2 e Conf-5) devem ser usados.

4.3.3.2. Identificação das interações moleculares conservadas durante as simulações de DM

Adicionalmente, todas as ligações de hidrogênio, interações iônicas e interações $\text{cation-}\pi$ observadas entre o inibidor Bz-nKRR-H (covalentemente ligado, ou não) e os resíduos do sítio ativo da NS2B/NS3pro foram examinadas ao longo das simulações de DM. Os resultados obtidos estão mostrados na Figura 4.32. Seguindo os critérios definidos no item 3.5.4.8, foi possível identificar quais interações se mantiveram estáveis durante as simulações de DM. Assim, ligações de hidrogênio envolvendo os resíduos *Asp81* da região NS2B e os resíduos *Asp129* e *Tyr161* do domínio NS3pro se mostraram estáveis durante as simulações de DM das estruturas “covalente” e “não-covalente” (valores de ocupância maiores que 30 %). Enquanto que, ligações de hidrogênio com os resíduos *His51* e *Gly151* do domínio NS3pro se mantiveram estáveis apenas durante a simulação da estrutura “covalente” (valores de ocupância iguais a 32 e a 51 %, respectivamente). As ligações de hidrogênio com os resíduos *Phe130*, *Asn152* e *Gly153* do domínio NS3pro observadas na estrutura 3U1I (cristal), não se mostraram estáveis durante nenhuma das simulações de DM (valores de ocupância menores que 5%). A ligação de hidrogênio com o resíduo *Met84* da região NS2B, também observada na estrutura 3U1I, apresentou uma estabilidade média ao longo das simulações de DM das estruturas “covalente” e “não-covalente” (valores de ocupância iguais a 24% e a 12% respectivamente). De modo análogo, ligações de hidrogênio com o resíduo *Asp75* do domínio NS3pro, presentes na estrutura 3U1I, mostraram uma estabilidade média durante as duas simulações de DM (valores de ocupância iguais a 10% e a 20% respectivamente para estrutura “covalente” e “não-covalente”). Adicionalmente, as ligações de hidrogênio envolvendo o resíduo *Asp81* da região NS2B, conservadas nas duas simulações, não estão

presentes na estrutura 3U1I. Esta diferença se deve ao fato de que na estrutura 3U1I (cristal) o resíduo *Asp81* está exposto ao solvente.

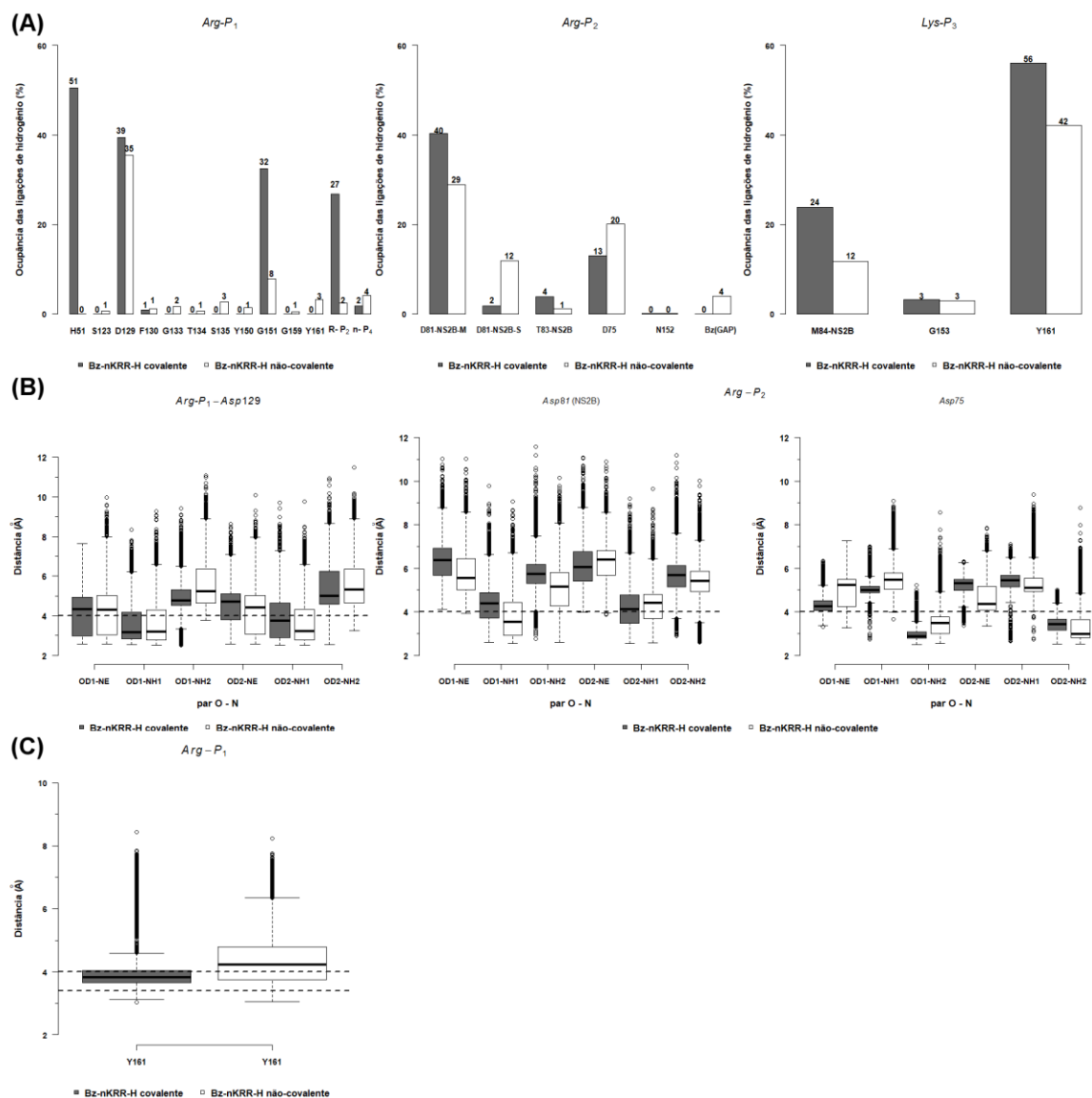


Figura 4.32. Interações envolvendo o Bz-nKRR-H e os resíduos do sítio ativo observadas ao longo das simulações de dinâmica molecular das estruturas “covalente” (cinza claro) e “não-covalente” (cinza escuro). **(A)** Gráfico de barra dos valores de ocupância das ligações de hidrogênio (%) das interações envolvendo os resíduos *Arg(P₁)*, *Arg(P₂)* e *Lys(P₃)*. **(B)** Gráficos de caixa^(a) dos valores de distância de interação (Å) para as interações iônicas envolvendo os resíduos *Arg(P₁)* e *Arg(P₂)*. **(C)** Gráficos de caixa^(a) dos valores de distância (Å) para interação cation- π entre a *Tyr161* e a *Arg(P₁)*. As linhas tracejadas em preto mostram os valores de corte (4 Å e 3,4 - 4 Å) usados para definir a estabilidade destas interações (BARLOW; THORNTON, 1983; BISSANTZ; KUHN; STAHL, 2010)(BISSANTZ; KUHN; STAHL, 2010). Figura retirada de Piccirillo, *et. al* 2016.

^(a) Para mais detalhes vide a descrição que acompanha a legenda da Figura 4.2.

Ainda, interações iônicas estáveis com os resíduos *Asp75* e *Asp129* do domínio NS3pro foram observadas em ambas as simulação de DM. Enquanto que as interações iônicas

com o resíduo *Asp81* da região NS2B apresentaram uma estabilidade média. Conforme apresentado na etapa de reconhecimento da estrutura da NS2B/NS3pro do DENV (item 4.3.1.1), o resíduo *Arg-P_I* do inibidor forma interações cátion- π com a *Tyr161* do domínio NS3pro. Interações envolvendo sistemas- π não são diretamente consideradas pelo campo de força ff99SB utilizado nas nossas simulações de DM. Contudo, Cerutti e colaboradores (CERUTTI et al., 2010) observaram em suas simulações de DM, realizadas utilizando este mesmo campo força, que alguns resíduos do sítio ativo mantiveram as orientações observadas em sua estrutura cristalográfica, nas quais se observava a formação de interações de tipo π -stacking. Baseado nisto, nos propomos (PICCIRILLO et al., 2016) considerar a interação cátion- π como estável se a distância entre o grupo guanidino do resíduo *Arg-P_I* do Bz-nKRR-H e o anel aromático da *Tyr161* se mantiver, ao longo de toda simulação de DM, compatível com a esperada para este tipo de interação (entre 3,4 a 4 Å (BISSANTZ; KUHN; STAHL, 2010)). Desta forma, uma interação cátion- π estável envolvendo a *Tyr161* foi observada apenas na simulação “covalente”.

A análise da estabilidade destas interações ao longo das simulações de DM mostrou que as interações envolvendo o resíduo *Met84* da região NS2B e os resíduos *Phe130*, *Asn152* e *Gly153* do domínio NS3pro não são estáveis. Cabe lembrar que ligações de hidrogênio com os resíduos *Phe130* e *Gly153* do domínio NS3pro foram incluídas no primeiro modelo farmacofórico gerado, o que poderia explicar a baixa atividade dos compostos selecionados por este modelo. Desta forma as interações com os resíduos *Met84* (NS2B), *Phe130*, *Asn152* e *Gly153* deverão ser incluídas com cuidado nos próximos modelos farmacofóricos gerados. Adicionalmente, interações tanto com *Asp81* da região NS2B bem com *His51*, *Asp129* e *Gly151* do domínio NS3pro serão priorizadas na construção dos próximos modelos farmacofóricos.

4.3.3.3. Identificação dos modos de ligação do inibidor Bz-nKRR-H

De modo análogo ao descrito acima, o espaço conformacional do inibidor Bz-nKRR-H foi analisado através do agrupamento hierárquico baseado nos valores de *RMSD*. Como descrito no item 3.5.4.8 (Materiais e Métodos), a análise das simulações de DM das estruturas “*covalente*” e “*não-covalente*” foi realizada em separado tendo em vista que o número de átomos do inibidor Bz-nKRR-H varia de acordo com estrutura utilizada (presença, ou ausência da ligação covalente com a *Ser135*). O número de grupos foi gerado considerando-se um valor de corte de *RMSD* igual a 5 Å, resultando em seis (denominados Cov-1 a Cov-6) e sete (denominados Nonc-1 a Nonc-7) grupos para as simulações das estruturas “*covalente*” e “*não-covalente*”, respectivamente. O grupo Cov-6 contém 64% dos estados conformacionais do inibidor Bz-nKRR-H observados durante a trajetória da estrutura “*covalente*”, enquanto que os demais estados conformacionais desta trajetória estão distribuídos entre os grupos Cov-1, Cov-2, Cov-3, Cov-4 e Cov-5 (2, 14, 3, 3 e 14%, respectivamente). Já 64% dos estados conformacionais do inibidor Bz-nKRR-H obtidos na simulação “*não-covalente*” estão igualmente distribuídos entre os grupos Nonc-2, Nonc-3 e Nonc-7 (23, 20 e 21% dos estados conformacionais, respectivamente). Os grupos Nonc-1, Nonc-4, Nonc-5 e Nonc-6 contêm 16, 6, 13 e 1% dos estados conformacionais do inibidor Bz-nKRR-H obtidos na simulação “*não-covalente*”. Estes resultados mostram que durante a simulação da estrutura “*covalente*” o inibidor Bz-nKRR-H adota um modo de ligação preferencial. Por outro lado, durante a simulação da estrutura “*não-covalente*”, pelo menos três modos de ligação, igualmente favoráveis, são observados. Para os grupos mais populosos (Cov-6, Nonc-2, Nonc-3 e Nonc-7), o *snapshot* central foi escolhido como estrutura representativa do correspondente grupo (Conf-Cov6, Conf-Nonc2, Conf-Nonc3 e Conf-Nonc7) e utilizado nas análises subsequentes bem como na inspeção visual. Analisando estas estruturas representativas, foi observado que

Arg-P₂ é o único resíduo que fica próximo à estrutura 3U1I (cristalográfica) em todas as estruturas representativas, como apresentado na Figura 4.33.

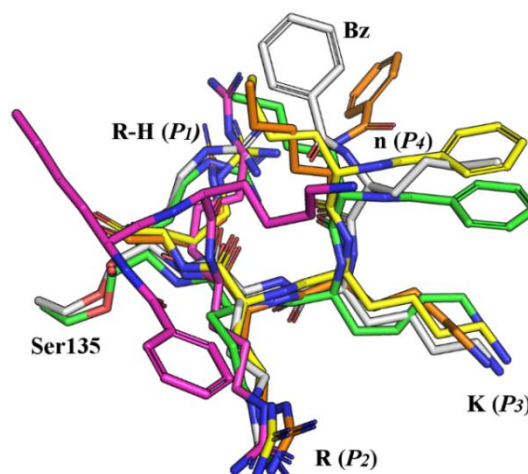


Figura 4.33. Representação do inibidor Bz-nKRR-H da estrutura 3U1I (NS2B/NS3pro do DENV3) sobreposto aos inibidores das estruturas representativas^(a) (Conf-Cov6, Conf-Nonc2, Conf-Nonc3 e Conf-Nonc7) dos grupos^(b) mais populosos. Todos os átomos de oxigênio e de nitrogênio estão mostrados em azul e em vermelho, respectivamente. Os átomos de carbono da estrutura 3U1I estão mostrados em cinza e os átomos de carbono das estruturas representativas Conf-Cov6, Conf-Nonc2, Conf-Nonc3 e Conf-Nonc7 estão mostrados em verde, em laranja, em amarelo e em rosa, respectivamente. Para melhor visualização, os átomos de hidrogênio foram excluídos e os demais mostrados na forma de “sticks”. Figura retirada de Piccirillo, *et. al* 2016.

(a) *Snapshot* central do grupo.

(b) Grupos obtidos através do agrupamento hierárquico usando os valores de *RMSD* e um valor de corte de 5 Å.

Nas estruturas Conf-Cov6, Conf-Nonc2 e Conf-Nonc3 o resíduo *nLeu-P₄* e o grupo *Bz* do inibidor adotam uma orientação invertida em relação à observada na estrutura 3U1I. É interessante ressaltar que esta mesma orientação invertida havia sido observada na etapa de validação do ancoramento, utilizando o programa *GOLD* (JONES et al., 1997), descrita no item 4.2.2.4. De modo análogo ao apresentado na etapa de validação do procedimento de ancoramento, estas posições experimentais foram consideradas como um artefato da cristalização sendo este ausente nas simulações de DM. Sendo assim, parece razoável propor que duas regiões do inibidor expostas ao solvente e de mesma natureza — ambas hidrofóbicas — poderiam adotar orientações invertidas às observadas no cristal. Ainda, nas estruturas Conf-Cov6, Conf-Nonc2 e Conf-Nonc3 os resíduos *Arg-P₁*, *Arg-P₂* e *Lys-P₃* do inibidor Bz-

nKRR-H apresentam orientações próximas às observadas na estrutura 3U1I. A estabilidade das posições dos resíduos *Arg-P₁* e *Arg-P₂* era esperada, considerando sua importância no reconhecimento dos substratos naturais da NS2B/NS3pro do DENV (GOUVEA et al., 2007). Por fim, na estrutura Conf-Nonc7 as regiões do inibidor Bz-nKRR-H estão expostas ao solvente, exceto pelo resíduo *Arg-P₂* que mantém uma orientação próxima à observada no cristal. Ainda, estes quatro modos de ligação do inibidor Bz-nKRR-H identificados durante as simulações de DM, representados pelas estruturas Conf-Cov6, Conf-Nonc2, Conf-Nonc3 e Conf-Nonc7, serão utilizados na construção de outros modelos farmacofóricos.

4.3.4. Modelos de busca virtual: Modelo II - DENV

Conforme descrito no item 4.2.2, os resultados obtidos na etapa de validação experimental do Modelo I – DENV sugerem que este primeiro modelo de busca virtual precisava ser melhorado. Assim, a partir dos conhecimentos e experiências adquiridos na preparação do primeiro modelo de busca virtual bem como no estudo da flexibilidade conformacional da NS2B/NS3pro do DENV através de simulações de DM, foi proposto um segundo modelo de busca virtual de inibidores da NS2B/NS3pro do DENV, denominado Modelo II – DENV. Neste segundo modelo foram introduzidas algumas modificações, a saber: (i) no banco de dados de moléculas virtuais usado para fazer a busca virtual; (ii) nos tipos de filtro de seleção aplicados ao banco de dados; (iii) no programa de ancoramento utilizado; (iv) nos critérios de cada um destes filtros de seleção e (v) na sequência de aplicação dos diferentes filtros de seleção.

A primeira mudança em relação ao Modelo I – DENV aplicada na construção deste Modelo II – DENV foi a construção de outro banco de dados de moléculas virtuais para ser utilizado na busca virtual, denominado nesta tese *ZINC-Curated*. Dentre as principais etapas envolvidas na construção deste banco, cita-se a aplicação de um filtro de solubilidade e, em seguida, de um filtro de propriedades físico-químicas e/ou estruturais. Este segundo filtro foi

proposto visando, em especial, remover moléculas com grupos funcionais reativos evitando, assim, a seleção de compostos com baixa estabilidade, que já havia sido observada por nós no Modelo I - DENV. Uma análise detalhada dos resultados de cada etapa envolvida na criação do banco *ZINC-Curated* bem como da análise da aplicação destes filtros na diversidade das propriedades físico-químicas e/ou estruturais deste banco já foram apresentados e discutidos em detalhes no item 4.2 desta tese.

Assim, partindo-se do banco de dados *ZINC-Curated*, foram aplicados três filtros de seleção na seguinte sequência: ancoramento, “grupos funcionais” e inspeção visual. O número menor de moléculas presentes no banco *ZINC-Curated* ($7,8 \times 10^6$ moléculas) quando comparado ao número no banco *ZINC-Original* ($23,3 \times 10^6$ moléculas) bem como a utilização do programa de ancoramento *FRED* (MCGANN, 2011) permitiu aplicar o filtro de ancoramento a todo banco de moléculas virtuais. Além da estrutura 3U1I (retirada do *PDB*), neste filtro foram incluídas as diferentes conformações do sítio ativo identificadas durante as simulações de DM. Ainda, foi, por nós, proposto e aplicado um filtro de “grupos funcionais”. E, de modo análogo ao primeiro modelo de busca virtual, o último filtro foi a análise por inspeção visual usando, agora, outros critérios de seleção, descritos no item 3.5.4.10 (Materiais e Métodos). Por fim, os compostos selecionados pelo Modelo II – DENV foram testados como inibidores da NS2B/NS3pro do DENV. A Figura 4.34 mostra uma representação esquemática da sequência de aplicação de cada filtro bem como os valores da porcentagem de redução do número de compostos em cada etapa.

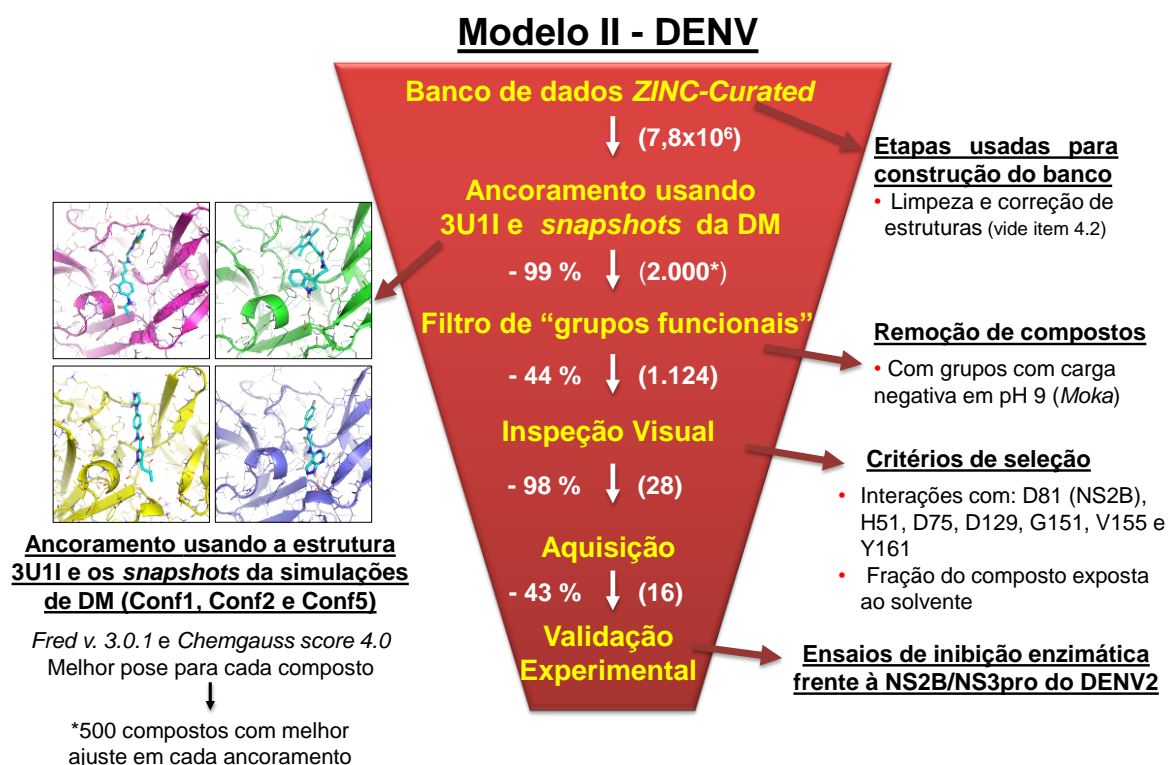


Figura 4.34. Representação esquemática da sequência de filtros de seleção aplicados ao banco *ZINC-Curated* representando usada no segundo modelo de busca virtual (**Modelo II – DENV**). Entre parênteses são indicados os números de compostos selecionados em cada etapa e ao lado a porcentagem de redução por cada filtro. *Considerando-se os 500 compostos com melhor encaixe no sítio ativo da NS2B/NS3pro do DENV obtidos em cada um dos quatro procedimentos de ancoramento.

4.3.4.1. Ancoramento

A primeira etapa do procedimento de busca virtual foi a aplicação de um filtro ancoramento a todas as moléculas presentes no banco de dados *ZINC-Curated*. Todos os procedimentos envolvidos nesta etapa foram feitos usando o programa *FRED* (MCGANN, 2011) e estão descritos no item 3.5.4.10 (Materiais e Métodos). Considerando o tamanho do banco de *ZINC-Curated*, a escolha do programa de ancoramento *FRED* (MCGANN, 2011) ao invés do programa *GOLD* (JONES; WILLETT; GLEN, 1995; VERDONK et al., 2003) se justifica devido: (i) ao menor custo computacional do ancoramento, pois o primeiro usa um algoritmo de complementaridade de formato ao invés de algoritmo genético (vide item 1.8); (ii) à disponibilidade de distribuir o cálculo em diferentes estações de trabalho, o que não é

possível com o programa *GOLD* devido ao tipo de licença que temos no nosso grupo e (iii) aos exemplos de sucesso utilizando o programa *FRED* (MCGANN, 2011) na busca virtual descritos na literatura (BRUS et al., 2014; MCGANN, 2011, 2012).

Visando considerar parte da flexibilidade dos resíduos do sítio ativo da NS2B/NS3pro do DENV, além da estrutura 3U1I (cristalográfica), três diferentes estados conformacionais identificados nas simulações de DM (representados pelas estruturas Conf-2, Conf-4 e Conf-5) foram usados nestes ancoramentos. Cabe esclarecer que esta etapa foi realizada concomitantemente com a análise dos resultados das simulações de DM (item 4.2.3.1). Desta forma, ainda não havia sido observado que: os campos de interação moleculares da estrutura Conf-4 eram similares aos observados para a estrutura 3U1I e que os campos calculados para a estrutura Conf-1 eram diferentes dos campos calculados para as demais estruturas. Ainda, em uma primeira análise, por inspeção visual, as estruturas Conf-1, Conf-3 e Conf-2 pareciam ser semelhantes entre si (Figura 4.30), enquanto, que as estruturas Conf-4, Conf-5 e 3U1I pareciam ser diferentes umas das outras (Figura 4.30). Baseado nesta primeira análise, a estrutura Conf-4 foi escolhida para ser usada nos ancoramentos ao invés da estrutura Conf-1, sendo esta última sugerida ao final da análise das simulações de DM como relevante para ser incluída na busca virtual. Assim, todas as moléculas presentes no banco de dados *ZINC-Curated* foram submetidas aos ancoramentos no sítio ativo da NS2B/NS3pro do DENV usando, separadamente, as estruturas 3U1I, Conf-2, Conf-4 e Conf-5. Os resultados obtidos em cada ancoramento foram, inicialmente, analisados separadamente e os 500 compostos com o menor valor de escore (ou seja, com o melhor encaixe no sítio ativo da proteína) obtidos em cada ancoramento foram selecionados para próxima etapa. Cabe ressaltar que embora este critério de seleção tenha suas limitações, devido aos erros inerentes das funções de ranqueamento, neste primeiro filtro, o grande número de compostos (7,8 milhões) inviabilizou que a seleção dos compostos fosse realizada por inspeção visual. A seguir, os compostos

selecionados em cada ancoramento foram agrupados, resultando na seleção de 2.000 compostos (redução de 99,99%) que foram submetidos a um filtro de “grupos funcionais”.

4.3.4.2. Filtro de “grupos funcionais”

A segunda etapa da busca virtual foi a aplicação de um filtro de “grupos funcionais” aos 2.000 compostos selecionados pelo ancoramento. Conforme apresentado e discutido em detalhes no item 4.2, durante a etapa de criação do banco *ZINC-Curated*, filtros de propriedades físico-químicas e/ou estruturais foram aplicados para enriquecer o banco de dados com compostos com características de fármacos. Cabe ressaltar, no entanto, que além das propriedades já consideradas, outras propriedades podem ser incluídas dependendo do receptor alvo estudado. Neste contexto, no item 4.3.1 desta tese estão apresentadas e discutidas algumas características do sítio ativo da NS2B/NS3pro do DENV. Dentre estas, cita-se sua preferência por substratos contendo resíduos de aminoácidos básicos em P_1 e P_2 , que é justificada, dentre outros fatores, pela presença de três resíduos aminoácidos ácidos nas cavidades S_1 e S_2 (*Asp75* e *Asp129* do domínio NS3pro e *Asp81* da região NS2B). Vale lembrar que estes três resíduos apresentaram interações estáveis ao longo das simulações de DM (item 4.3.3.2).

Diante disso, foi por nós proposto um filtro de grupos funcionais objetivando remover os compostos com cargas negativas e com baixa estabilidade nas condições dos experimentos de inibição (pH 9,0). Procedendo como descrito no item 3.5.4.5 (Materiais e Métodos), inicialmente, foram removidos compostos contendo os seguintes grupos funcionais: dicarbonílicos, ácidos carboxílicos, grupos contendo átomos ionizáveis com cargas negativas (em pH 7,4), grupos protetores (ex. acetal); grupos contendo $N_{sp3}-N_{sp3}$ e ésteres com bons grupos de partida que poderiam favorecer reações de hidrólise levando à formação de ácidos carboxílicos. O número de compostos apresentando cada um dos grupos funcionais utilizados neste filtro está apresentado na Tabela 4.19. Todos compostos eliminados nesta etapa foram

submetidos, na sequência, a uma inspeção visual, na qual se verificou, em especial, se os ésteres de ácidos carboxílicos pareciam ser suscetíveis à hidrólise.

Tabela 4.19. Número de compostos removidos para cada um dos grupos funcionais considerados neste filtro.

Grupo funcional considerado	Número de compostos removidos
Dicarbonílicos	87
Ácidos carboxílicos	79
Grupos ionizáveis com cargas negativas ^(a) (em pH 7,4)	31
Grupos protetores	3
Grupo contendo N _{sp3} -N _{sp3}	66
Ésteres ^(b)	26

^(a) Calculadas durante a criação do banco *ZINC-Curated*.

^(b) Contendo grupos de partida capazes de favorecer reações de esterificação.

Em uma segunda etapa, 584 compostos cujas espécies, ainda, apresentaram carga negativa em pH 9,0 foram eliminados. A aplicação deste filtro resultou 1.124 compostos (redução de 43,80%) que foram, a seguir, submetidos a uma análise por inspeção visual.

4.3.4.3. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV

De modo análogo ao primeiro modelo de busca virtual, os 1.124 compostos selecionados no filtro de “grupos funcionais” foram analisados por inspeção visual para verificar seus ajustes no sítio ativo da NS2B/NS3pro do DENV. Esta análise foi feita procedendo-se como descrito nos descritos nos itens 3.5.4.10 (Materiais e Métodos). Neste segundo modelo, diferentemente do descrito para o Modelo I – DENV, os critérios usados nesta etapa foram definidos visando selecionar compostos capazes de estabelecer interações com resíduos do sítio ativo que formaram interações estáveis ao longo das simulações de DM,

como: *His51*, *Asp75*, *Asp129*, *Gly151* e *Asp81* (NS2B). Todos os critérios utilizados nesta inspeção estão descritos no item 3.5.4.10.

Nesta etapa, 28 compostos foram selecionados como potenciais inibidores da NS2B/NS3pro do DENV, representando uma redução de 97,51 % em relação aos 1.124 compostos selecionados na etapa anterior. Destes 28 potenciais inibidores, 16 (57,14%) estavam disponíveis e foram adquiridos para validação experimental do Modelo I - DENV. Por fim, a aplicação de todos os filtros de seleção representou uma redução total de mais de 99 % em relação ao banco *ZINC-Curated* ($\sim 7,8 \times 10^6$ compostos). A Tabela 4.20 mostra a representação das estruturas dos dezesseis compostos (denominados **DV-11** a **DV-26**) selecionados pelo Modelo II – DENV que foram adquiridos e submetidos a testes de inibição enzimática frente à NS2B/NS3pro do DENV.

Tabela 4.20. Representação da estrutura 2D dos 16 compostos (DV-11 a DV-26) selecionados pelo Modelo II – DENV como potenciais de inibidores da NS2B/NS3pro do DENV e que foram adquiridos e submetidos ao teste de inibição enzimática. Estas estão acompanhadas dos seus correspondentes códigos ZINC, valores de peso molecular e empresa fornecedora.

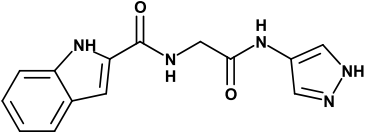
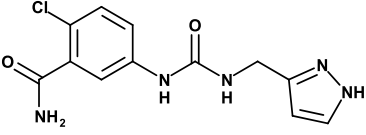
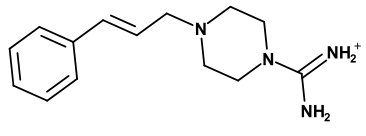
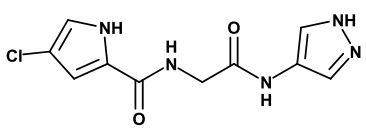
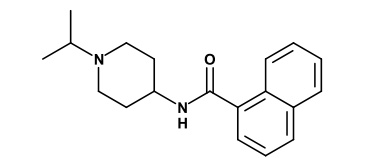
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-11		ZINC69421108	283,29	Enamine
DV-12		ZINC40540323	293,71	Enamine
DV-13		ZINC25605802	244,34	UORSY
DV-14		ZINC58246768	267,67	Enamine
DV-15		ZINC12382552	296,41	Vitas M Labs

Tabela. 4.20 (Continua).

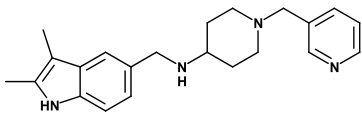
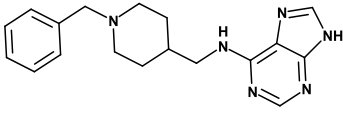
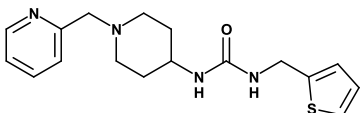
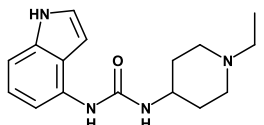
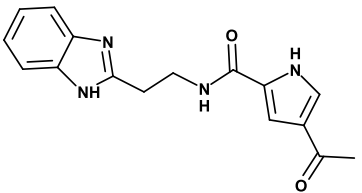
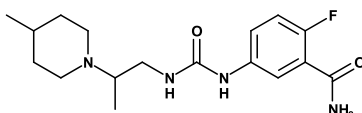
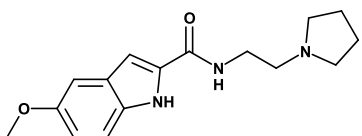
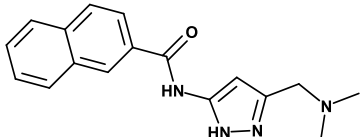
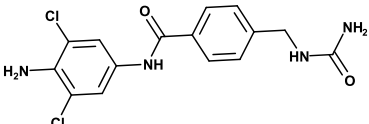
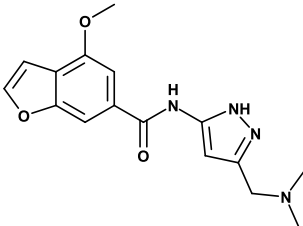
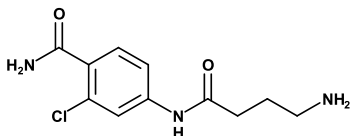
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-16		ZINC12449348	348,48	ChemBridge
DV-17		ZINC58145695	322,41	Enamine
DV-18		ZINC40549171	330,45	Enamine
DV-19		ZINC71803380	286,37	Enamine
DV-20		ZINC69433258	296,32	Enamine
DV-21		ZINC65585171	336,40	Enamine
DV-22		ZINC09581802	287,36	Vitas M Lab
DV-23		ZINC72258613	294,35	Princeton BioMolecular
DV-24		ZINC08935785	353,20	Enamine

Tabela. 4.20 (Continua).

#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-25		ZINC72258577	314,34	Princeton BioMolecular
DV-26		ZINC37046989	255,70	UORSY

4.3.4.4. Validação experimental

De modo análogo ao Modelo I - DENV, a verdadeira utilidade do modelo Modelo II – DENV na busca de compostos bioativos foi comprovada com sua validação experimental, através de ensaios enzimáticos. Assim, os 16 compostos selecionados pelo Modelo II - DENV foram submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV2. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Wibke Diederich (Philipps-Universität de Marburg, Alemanha) e estão descritos no item 3.6.1.

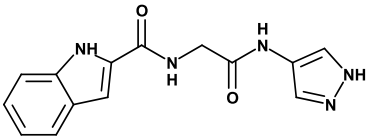
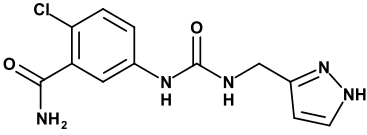
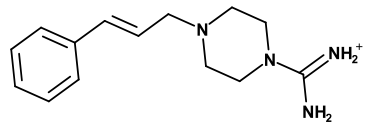
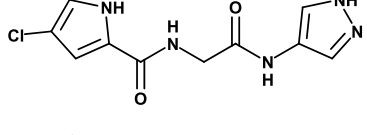
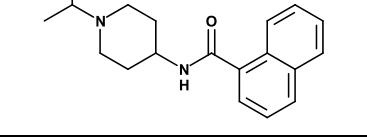
Ensaio de inibição enzimática frente à NS2B/NS3pro do DENV

Procedendo-se como descrito 3.6.1.1 (Materiais e Métodos), inicialmente, foram determinados os valores de porcentagem de inibição da NS2B/NS3pro do DENV2 de cada composto (em 100 μ M), sendo apresentados na Tabela 4.21. Ainda, como proposto no item 3.6.1.1 somente compostos com valores de atividade inibitória maiores, ou próximos a 50%, foram considerados como tendo atividade inibitória significativa.

Assim, dos 16 compostos adquiridos, o composto **DV-14** (ZINC58246768) foi o que apresentou a maior atividade inibitória (inibição de 39 ± 3 %, em 100 μ M), no entanto, este valor é menor do que o observado para os compostos **DV-08** e **DV-09** selecionados pelo

primeiro modelo de busca virtual. Os demais compostos não apresentaram atividade inibitória significativa na concentração usada, sendo eles: compostos **DV-11** (ZINC69421108), **DV-12** (ZINC40540323), **DV-13** (ZINC25605802), **DV-15** (ZINC12382552), **DV-16** (ZINC12449348), **DV-17** (ZINC58145695), **DV-18** (ZINC40549171), **DV-19** (ZINC71803380), **DV-20** (ZINC69433258), **DV-21** (ZINC65585171), **DV-22** (ZINC09581802), **DV-23** (ZINC72258613), **DV-24** (ZINC08935785), **DV-25** (ZINC72258577) e **DV-26** (ZINC37046989). A Tabela 4.21 mostra os valores de porcentagem de inibição (em 100 μ M) obtidos para cada um dos 16 compostos testados bem como os correspondentes desvios padrões. Considerando a baixa atividade inibitória destes compostos seus correspondentes valores de IC_{50} não foram determinados.

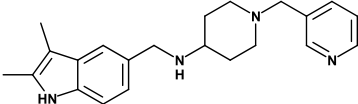
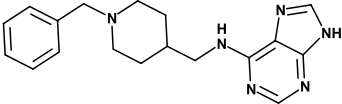
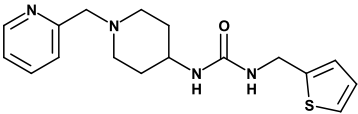
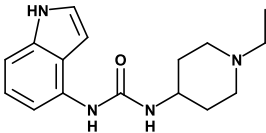
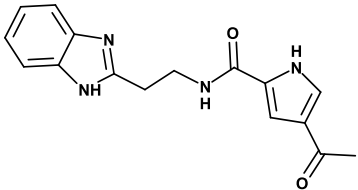
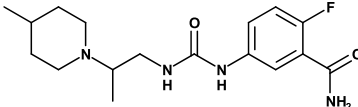
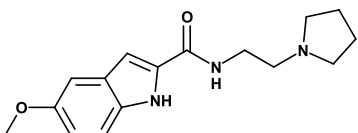
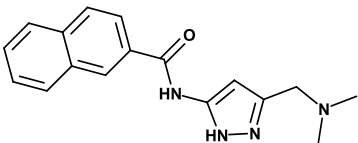
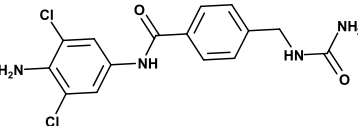
Tabela 4.21. Valores de porcentagem de inibição (em 100 μ M)^(a) e dos correspondentes desvios padrão^(b) de cada um dos compostos (DV-11 a DV-26) selecionados pelo Modelo II – DENV.

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M)
DV-11		7 ± 6
DV-12		9 ± 12
DV-13		13 ± 7
DV-14		39 ± 3
DV-15		7 ± 1

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

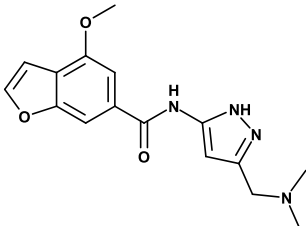
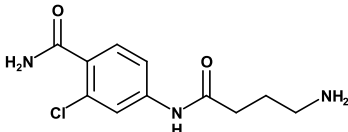
Tabela 4.21. (Continua).

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M)
DV-16		19 ± 6
DV-17		12 ± 5
DV-18		12 ± 4
DV-19		25 ± 11
DV-20		13 ± 6
DV-21		14 ± 1
DV-22		20 ± 9
DV-23		13 ± 1
DV-24		15 ± 12

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^ª Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

Tabela 4.21. (Continua).

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M)
DV-25		12 ± 6
DV-26		13 ± 4

(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a. Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

(b) Considerando pelo menos três experimentos independentes.

Avaliação da estabilidade dos compostos nas condições do experimento de inibição frente à da NS2B/NS3pro do DENV2

Como planejamos com relação à estabilidade, todos os 16 compostos adquiridos (**DV-11** a **DV-26**) se mostraram estáveis nas condições do experimento de inibição (tampão pH 9,0). Desta forma, a baixa atividade inibitória observada para estes compostos não pode ser atribuída a uma falta de estabilidade nas condições dos experimentos de inibição. Ainda, considerando que nenhum dos 16 compostos selecionados pelo Modelo II – DENV apresentou atividade inibitória frente à NS2B/NS3pro do DENV, optamos por não realizar as etapas de identificação, caracterização e determinação dos graus de pureza destes compostos.

Assim, considerando que os resultados obtidos, embora tenha selecionado compostos com maior estabilidade que o Modelo I - DENV, o Modelo II –DENV não foi capaz de selecionar qualquer composto com atividade inibitória significativa frente à NS2B/NS3pro do DENV. Dentre as possíveis justificativas para a falta de sucesso desse modelo, quanto à atividade inibitória, podemos considerar:

- a) O fato de a seleção dos compostos ter sido, inicialmente, baseada nos valores de escore da melhor pose, sem a análise concomitante da reprodutibilidade de suas correspondentes poses de ancoramento.
- b) A não inclusão de uma etapa de validação do procedimento de ancoramento que poderia ter demonstrado alguma limitação do programa *FRED* (MCGANN, 2011)
- c) O fato de ter sido utilizado apenas o filtro de ancoramento. Este pode não ter sido ideal pra sítios rasos e achatados, como o sítio ativo da NS2B/NS3pro do DENV, pois para estes espera-se uma menor taxa de sucesso para proteases (IRWIN; SHOICHET, 2016).

Independentemente da razão do insucesso do Modelo II – DENV, os resultados obtidos até este modelo não foram satisfatórios. Assim, um novo modelo de busca virtual foi proposto e construído.

4.3.5. Modelos de busca virtual: Modelo III - DENV

Os resultados obtidos na etapa de validação experimental do Modelo I – DENV sugere que este modelo de busca virtual precisaria ser melhorado (vide item 4.3.2.6). Assim, foi proposto um terceiro modelo de busca virtual de inibidores da NS2B/NS3pro do DENV, denominado Modelo III – DENV. Cabe esclarecer que este terceiro modelo de busca virtual foi proposto e construído concomitantemente com o segundo modelo de busca virtual (Modelo II – DENV), incluindo alguns outros aspectos relacionados à flexibilidade conformacional da NS2B/NS3pro do DENV. Neste terceiro modelo foram introduzidas algumas modificações em relação ao segundo modelo de busca virtual, especificamente para: (i) o banco de moléculas virtuais usado para fazer a busca virtual; (ii) o primeiro filtro de seleção aplicado ao banco de moléculas virtuais; (iii) a inclusão de uma etapa de análise do desempenho dos modelos farmacofóricos, usando uma série de inibidores peptidomiméticos da literatura e (iv) os critérios usados em cada um dos filtros de seleção.

A primeira mudança em relação ao primeiro e segundo modelos de busca virtual na construção do Modelo III – DENV foi a escolha do banco de moléculas virtuais *NCI* para ser

usado na busca virtual. Esta escolha foi feita considerando que *NCI* contém vários compostos com atividade biológica já descrita, além dos seus compostos serem enviados gratuitamente e, em geral, mais rapidamente do que para os comerciais. Assim, partindo-se do banco *NCI*, foram aplicados três filtros de seleção na seguinte sequência: farmacofórico, “grupos funcionais” e inspeção visual. Na construção destes modelos farmacofóricos, além do modo de ligação do inibidor Bz-nKRR-H observado na estrutura 3U1I foram incluídos os diferentes modos de ligação deste inibidor identificados durante as simulações de DM. Ainda, o desempenho destes modelos farmacofóricos foi avaliado, *in silico*, utilizando uma série de peptidomiméticos da literatura (YIN et al., 2006). Esta etapa de validação foi incluída visando remover os modelos que não fossem capazes de diferenciar os compostos ativos dos inativos. De modo análogo ao Modelo II – DENV, um filtro de “grupos funcionais” foi aplicado seguido de inspeção visual usando, agora, outros critérios de seleção (descritos no item 3.5.4.12, Materiais e Métodos). Por fim, os compostos selecionados pelo Modelo III – DENV foram testados como inibidores da NS2B/NS3pro do DENV. A Figura 4.36 mostra uma representação esquemática da sequência de aplicação de cada filtro bem como os valores da porcentagem de redução do número de compostos em cada etapa. Parte do material descrito nos próximos parágrafos faz parte de artigo recentemente publicado no *JCAMD*, do qual sou primeira autora (PICCIRILLO et al., 2016), colocado como anexo.

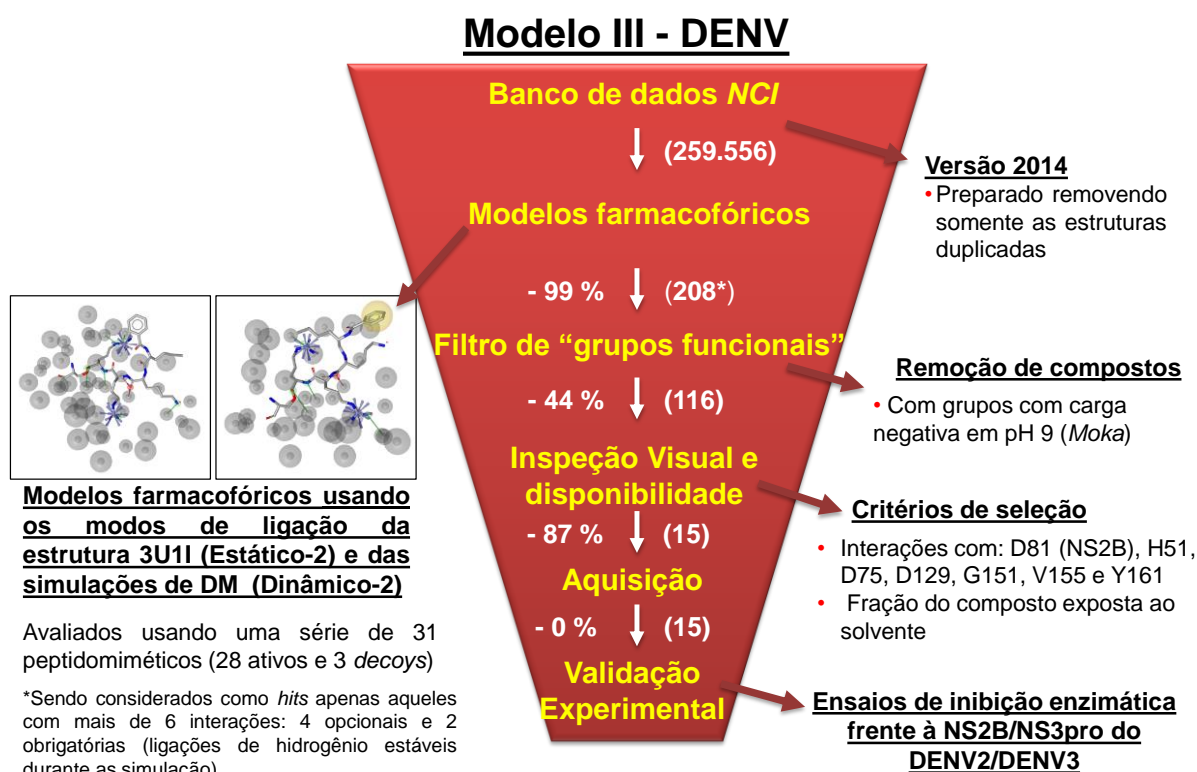


Figura 4.35. Representação esquemática da sequência de filtros de seleção aplicados ao banco NCI usada no terceiro modelo de busca virtual (Modelo III – DENV). Entre parênteses são indicados os números de compostos selecionados em cada etapa e ao lado a porcentagem de redução por cada filtro.*Compostos selecionados utilizando dois diferentes modelos farmacofóricos previamente validados *in silico* aplicando-se os critérios definidos na etapa de validação.

4.3.5.1. Filtro farmacofórico

De modo análogo ao primeiro modelo de busca virtual, a primeira etapa do procedimento de busca virtual foi a aplicação de um filtro farmacofórico ao banco NCI (preparado com descrito no item 3.5.1.1). Visando considerar a flexibilidade do inibidor Bz-nKRR-H dentro do sítio ativo da NS2B/NS3pro do DENV, além do modo de ligação observado para este inibidor na estrutura 3U1I (cristalográfica), os modos de ligação, posição e orientação do inibidor Bz-nKRR-H dentro do sítio ativo das estruturas representativas Conf-Cov6, Conf-Nonc2 e Conf-Nonc3 foram usados para gerar outros modelos farmacofóricos. O modo de ligação do inibidor Bz-nKRR-H observado na estrutura Conf-Nonc7 mostra apenas o resíduo *Arg-P₂* do inibidor interagindo com os resíduos do sítio ativo e, por isto, não foi usado para gerar outros modelos farmacofóricos. A obtenção das estruturas representativas

Conf-Cov6, Conf-Nonc2 e Conf-Nonc3, usando simulações de DM, está descrita em detalhes no item 4.3.3.

Construção de outros modelos farmacofóricos — utilizando os modos de ligação do Bz-nKRR-H obtidos durante as simulações de DM (200 ns)

Todos os modelos farmacofóricos gerados com base nas estruturas representativas (Conf-Cov6, Conf-Nonc2 e Conf-Nonc3) e na estrutura cristalográfica foram classificados, nesta tese, como modelos farmacofóricos “dinâmicos” e “estáticos”, respectivamente. Procedendo-se como descrito no item 3.5.4.12 (Materiais e Métodos), estes modelos farmacofóricos foram gerados considerando todas as interações possíveis entre o inibidor Bz-nKRR-H e os resíduos do sítio ativo da correspondente estrutura usada (Figura 4.36).

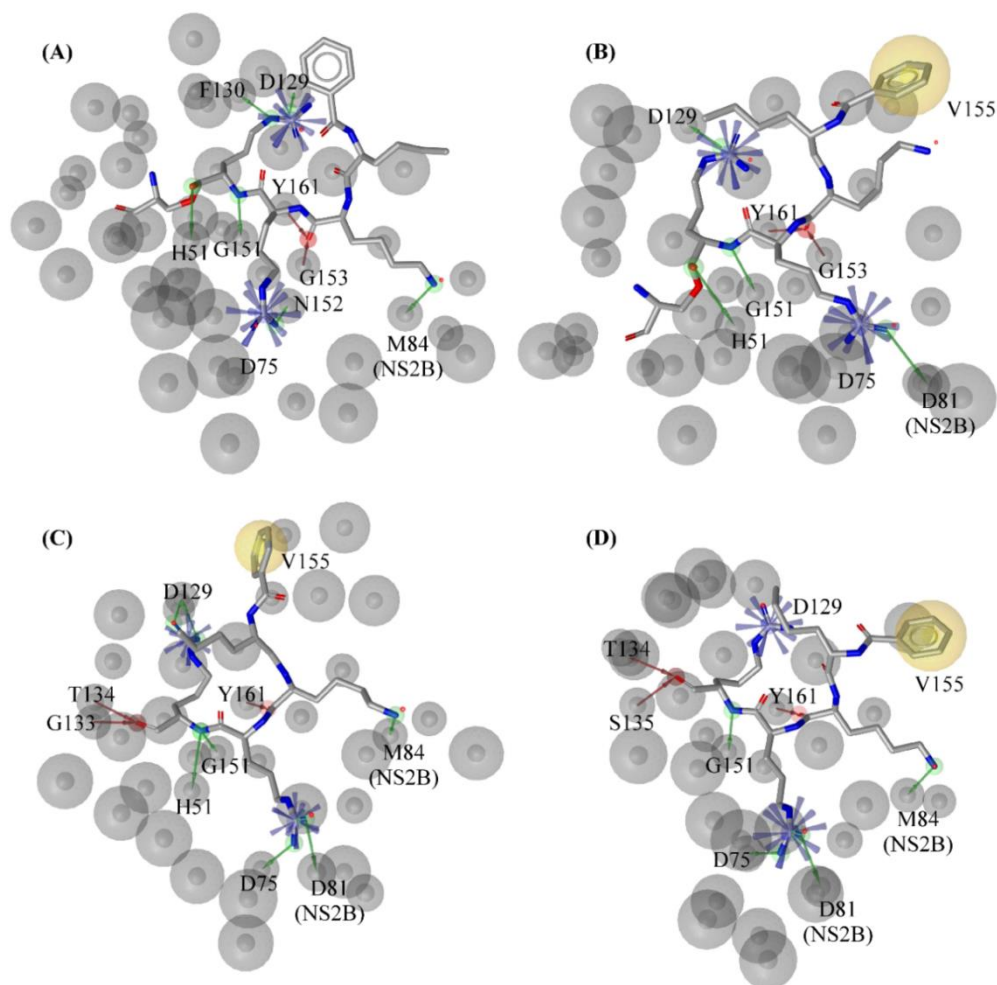


Figura 4.36. Representação esquemática dos modelos farmacofóricos^(a) gerados para as estruturas^(b) 3U1I, Conf-Cov6, Conf-Nonc2 e Conf-Nonc3 da NS2B/NS3pro do DENV3, classificados como: “estáticos” e “dinâmicos”. (A) Modelo farmacofórico “estático” gerado para estrutura 3U1I com 10 regiões de interação^(c). (B) Modelo farmacofórico “dinâmico” gerado para estrutura Conf-Cov6 com 9 regiões de interação^(c). (C) Modelo farmacofórico “dinâmico” gerado para estrutura Conf-Nonc2 com 13 regiões de interação^(c). (D) Modelo farmacofórico “dinâmico” gerado para estrutura Conf-Nonc3 com 10 regiões de interação^(c). Todos os modelos contêm regiões de exclusão de volume^(d) mostradas em cinza. As regiões *HBA*, *HBD*, hidrofóbicas e iônicas positivas estão representadas em verde, vermelho, azul e amarelo, respectivamente. As setas indicam regiões de interação direcionais. Figura retirada de Piccirillo, *et. al* 2016.

- (a) Gerados com o programa *LigandScout* (WOLBER; LANGER, 2005) a partir das interações identificadas entre o inibidor Bz-nKRR-H e os resíduos do sítio ativo, utilizando os valores de corte padrão do programa.
- (b) A estrutura 3U1I corresponde à estrutura cristalográfica obtida por raios-X e às estruturas Conf-Cov6, Conf-Nonc2 e Conf-Nonc3 correspondem às estruturas representativas dos grupos Cov6, Nonc2 e Nonc3 obtidos pelo agrupamento hierárquico dos *snapshots* das simulações de DM (item 4.3.3.3).
- (c) Todas as regiões de interação *HBA* e *HBD* possuem um raio de 1,8 Å e todas as regiões de interação iônica positiva ou hidrofóbica possuem um raio de 1,95 Å.
- (d) Definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio ativo de cada estrutura.

O número e natureza da região de interação variam de acordo com o modelo farmacofórico, como apresentado na Figura 4.36, a saber: (i) modelo farmacofórico “estático” gerado com base na estrutura cristalográfica 3U1I, incluindo seis regiões *HBA*, duas regiões

HBD e duas regiões interação iônica positiva; (ii) modelo farmacofórico “*dinâmico*” gerado com base na estrutura representativa Conf-Cov6, incluindo quatro regiões *HBA*, duas regiões *HBD*, duas regiões interação iônica positiva e uma região de interação hidrofóbica; (iii) modelo farmacofórico “*dinâmico*” gerado com base na estrutura representativa Conf-Nonc2, incluindo sete regiões *HBA*, três regiões *HBD*, duas regiões interação iônica positiva e uma região de interação hidrofóbica e (iv) modelo farmacofórico “*dinâmico*” gerado com base na estrutura representativa Conf-Nonc3, incluindo quatro regiões *HBA*, três regiões *HBD*, duas regiões interação iônica positiva e uma região de interação hidrofóbica.

Como descrito no item 4.3.3.2, interações envolvendo ligações de hidrogênio estáveis entre o inibidor Bz-nKRR-H e os resíduos *Asp129*, *Tyr161* e *Asp81* (NS2B) foram identificadas ao longo da simulação com a estrutura “*covalente*”. De modo análogo, interações envolvendo ligações de hidrogênio estáveis entre o inibidor Bz-nKRR-H e os resíduos *His51*, *Gly151*, *Tyr161* e *Asp81* (NS2B) foram também identificadas ao longo da simulação com a estrutura “*não-covalente*”. Todas estas interações estáveis foram utilizadas para definir duas regiões de interação de ligação de hidrogênio (aceptoras e/ou doadoras) como obrigatórias para cada um dos quatro novos modelos farmacofóricos gerados, sendo as demais regiões definidas como opcionais. Considerando que mais de duas interações de ligação de hidrogênio estáveis foram identificadas nas simulações de DM com as estruturas “*covalente*” e “*não-covalente*” (200 ns, cada), optou-se por usar todas as combinações, dois a dois, possíveis. Assim, a partir destas combinações foram gerados dezesseis modelos farmacofóricos, a saber: seis modelos farmacofóricos a partir das combinações do modelo *estático* (denominados Estático-1 a Estático-6); seis modelos farmacofóricos a partir das combinações do modelo *dinâmico* baseado na estrutura Conf-Cov6 (denominados Dinâmico-1 a Dinâmico-6); três modelos farmacofóricos a partir das combinações do modelo *dinâmico* baseado na estrutura Conf-Nonc2 (denominados Dinâmico-7 a Dinâmico -9) e, um modelo

farmacofórico a partir das combinações do modelo *dinâmico* baseado na estrutura Conf-Nonc3 (denominado Dinâmico-10). Os resíduos do sítio ativo associados com as duas regiões obrigatórias, representando ligações de hidrogênio estáveis, de cada um dos dezesseis modelos farmacofóricos estão apresentados na Tabela 4.22.

Tabela 4.22. Resíduos do sítio ativo associados com as regiões obrigatórias (representando ligações de hidrogênio estáveis) para cada um dos 16 modelos farmacofóricos gerados.

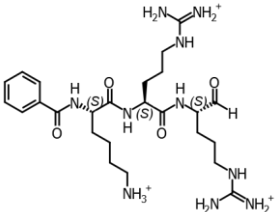
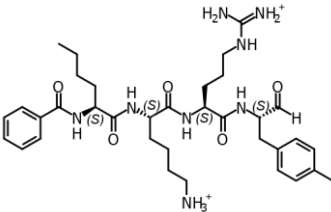
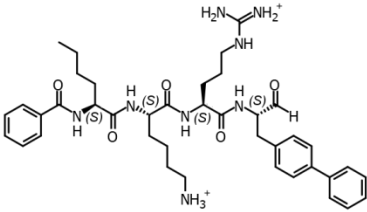
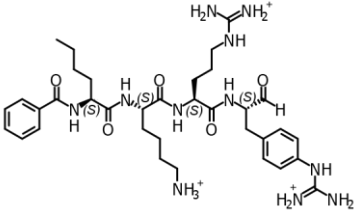
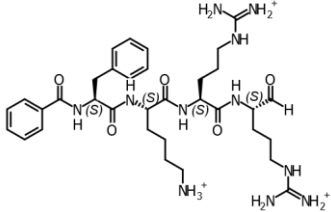
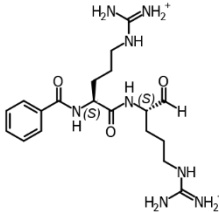
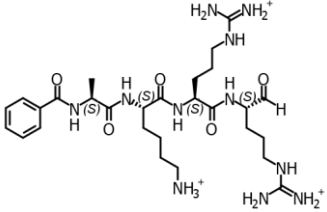
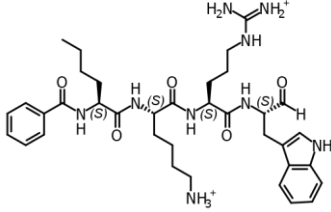
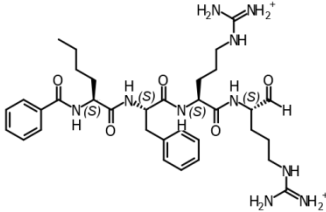
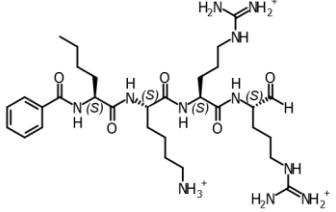
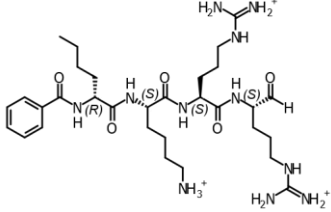
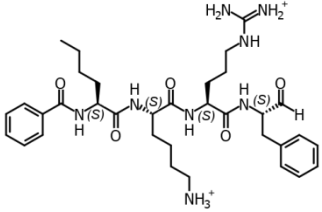
Modelo farmacofórico	Resíduos do sítio ativo associados com as regiões obrigatórias
Estático-1 (3U1I)	<i>Gly151 e Tyr161</i>
Estático-2 (3U1I)	<i>His51 e Tyr161</i>
Estático-3 (3U1I)	<i>Asp129 e Tyr161</i>
Estático-4 (3U1I)	<i>Gly151 e Asp129</i>
Estático-5 (3U1I)	<i>His51 e Asp129</i>
Estático-6 (3U1I)	<i>His51 e Gly151</i>
Dinâmico-1 (Conf-Cov6)	<i>His51 e Asp129</i>
Dinâmico-2 (Conf-Cov6)	<i>Gly151 e Asp129</i>
Dinâmico-3 (Conf-Cov6)	<i>His51 e Gly151</i>
Dinâmico-4 (Conf-Cov6)	<i>Asp81 (NS2B) e His51</i>
Dinâmico-5 (Conf-Cov6)	<i>Asp81 (NS2B) e Gly151</i>
Dinâmico-6 (Conf-Cov6)	<i>Asp81 (NS2B) e Asp129</i>
Dinâmico-7 (Conf-Nonc2)	<i>Asp129 e Tyr161</i>
Dinâmico-8 (Conf-Nonc2)	<i>Asp129 e Asp81 (NS2B)</i>
Dinâmico-9 (Conf-Nonc2)	<i>Tyr161 e Asp81 (NS2B)</i>
Dinâmico-10 (Conf-Nonc3)	<i>Asp81 (NS2B) e Tyr161</i>

Avaliação do desempenho dos 16 modelos farmacofóricos usando série de peptidomiméticos da literatura

A capacidade destes 16 modelos farmacofóricos selecionarem inibidores conhecidos e diferenciarem compostos ativos de inativos foi avaliada usando uma série de 31 peptidomiméticos da literatura (YIN et al., 2006) já testados frente à NS2B/NS3pro do

DENV. Esta classe de compostos foi escolhida, pois é a mesma do inibidor presente na estrutura usada como ponto de partida nas simulações de DM (item 4.3.3). Esta série foi preparada procedendo-se como descrito no item 3.5.2.1 (Materiais e Métodos) e contém 28 compostos ativos, incluindo o próprio inibidor Bz-nKRR-H, e ainda 3 inativos. As estruturas e os valores das constantes de inibição (K_i) destes peptídomiméticos frente a esta protease (YIN et al., 2006) estão apresentados na Tabela 4.23.

Tabela 4.23. Representação esquemática das estruturas dos 31 peptidomiméticos retirados da literatura e os respectivos valores da constante de inibição (K_i, μM) frente à NS2B/NS3pro do DENV (YIN et al., 2006). Valores dos erros/desvios não relatados.

#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)
1		1,5	5		6,0	9		11,6
2		2,8	6		6,8	10		12,0
3		5,3	7		7,5	11		15,8
4*		5,8	8		9,4	12		15,9

*Bz-nKRR-H

Tabela 4.23. (Continua).

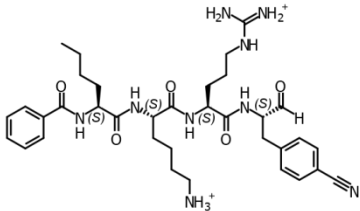
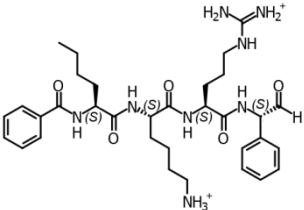
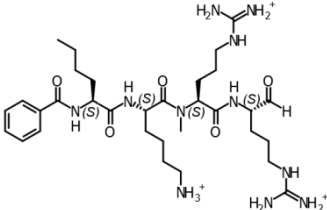
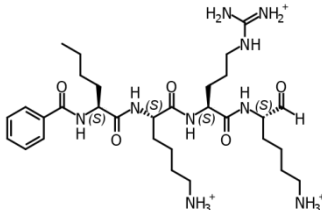
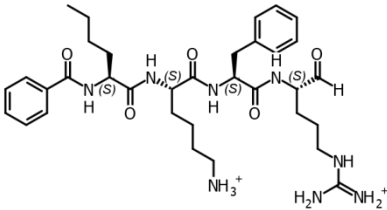
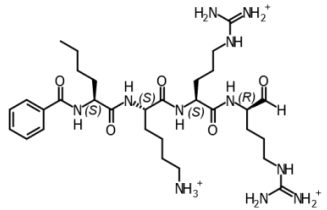
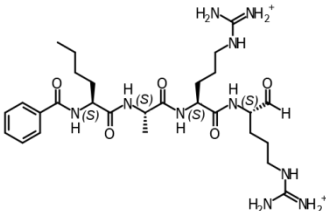
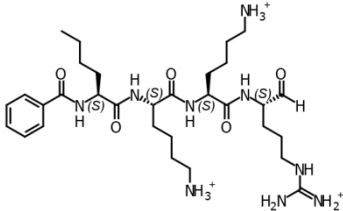
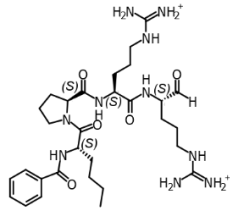
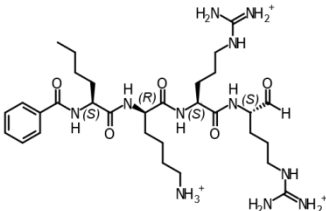
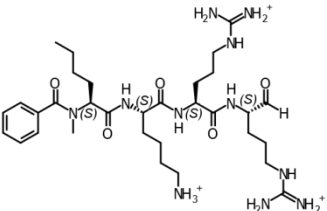
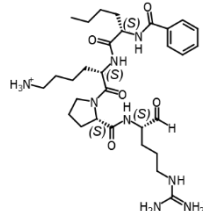
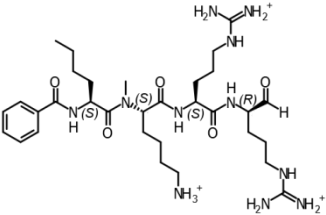
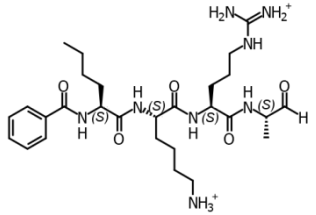
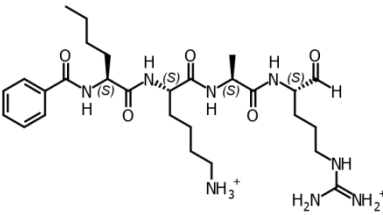
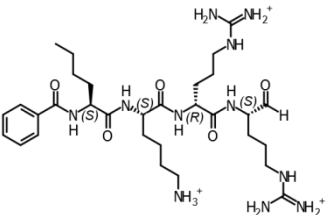
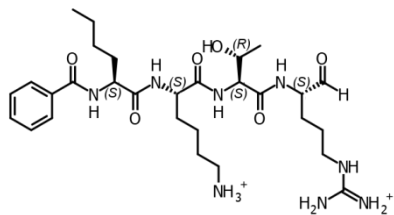
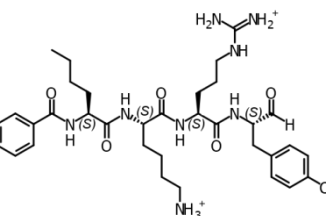
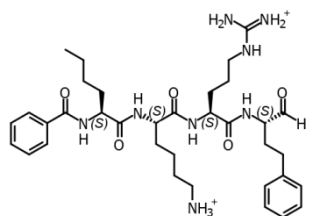
#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)
13		18,6	17		33,0	21		47,4
14		20,5	18		40,7	22		51,0
15		22,1	19		41,3	23		61,4
16		28,6	20		43,7	24		109,0

Tabela 4.23. (Continua).

#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)	#	Peptidomimético	K _i (μM)
25		113,3	28		178,0	31		Inativo > 500
26		115,0	29		Inativo > 500			
27		138,0	30		Inativo > 500			

Todos os 16 modelos farmacofóricos foram aplicados à série de peptidomiméticos e, apenas compostos capazes de respeitar seis ou mais regiões de interação (duas obrigatórias e quatro opcionais) foram considerados como *hit*¹³. Os *hits* e o número de *hits* (total, ativos e *decoys*¹⁴) selecionados para cada um dos modelos estão mostrados na Tabela 4.24.

Tabela 4.24. Hits e número de hits total, ativos e inativos selecionados para cada um dos 16 modelos farmacofóricos gerados.

Modelo farmacofórico	Hits selecionados	Número de hits		
		Total	Ativos	Decoys
Estático-1	1, 3, 4 ^a , 10 - 13, 15, 17, 18, 20, 24, 28, 29 ^b , 30 ^b , 31 ^b	16	13 (81%)	3 (19%)
Estático-2	1, 4 ^a , 10, 17, 18, 20, 24, 26, 28	9	9 (100%)	0
Estático-3	1, 2, 3, 4 ^a , 5, 7, 9, 10, 12, 14, 16, 17, 18, 20, 22, 23, 24, 26, 27, 29 ^b , 30 ^b	21	19 (90%)	2 (10%)
Estático-4	1, 4 ^a , 10, 18, 20, 21, 24, 31 ^b	8	7 (88%)	1 (12%)
Estático-5	1, 18, 20, 26, 31 ⁱ	5	4 (80%)	1 (20%)
Estático-6	1, 10, 18, 20, 24, 31 ⁱ	6	5 (83%)	1 (17%)
Dinâmico-1	4 ^a , 22, 28, 30 ^b	4	3 (75%)	1 (25%)
Dinâmico-2	4 ^a , 9, 10, 14, 19, 22	6	6 (100%)	0
Dinâmico-3	4 ^a , 10, 21, 22, 29 ^b , 31 ^b	6	4 (67%)	2 (33%)
Dinâmico-4	4 ^a , 10, 21, 29 ⁱ , 30 ⁱ	5	3 (60%)	2 (40%)
Dinâmico-5	1, 3, 4 ^a , 6, 8, 9, 10, 11, 14, 15, 19, 21, 29 ^b	13	12 (92%)	1 (8%)
Dinâmico-6	1, 3, 4 ^a , 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 23, 29 ^b , 30 ^b	20	18 (90%)	2 (10%)
Dinâmico-7	3, 4 ^a , 7, 8, 10 - 12, 15, 16, 22, 25, 29 ^b	12	11 (92%)	1 (8%)

^(a) Inibidor Bz-nKRR-H.

^(b) Composto inativo ($K_i > 500 \mu\text{M}$).

¹³ *Hit* é molécula com estrutura identificada/conhecida que possui uma atividade em um ensaio biológico abaixo de determinado valor corte (BUCKLE et al., 2014), correspondendo aos compostos selecionados na busca virtual.

¹⁴ Falsos positivos.

Tabela 4.24. (Continua).

Modelo farmacofórico	<i>Hits</i> selecionados	Número de <i>hits</i>		
		Total	Ativos	<i>Decoys</i>
Dinâmico-8	1, 2, 6, 8, 10, 11, 14, 15, 16, 19, 21, 22, 24, 25, 29 ^b	15	14 (93%)	1 (7%)
Dinâmico-9	1, 3, 4 ^a , 6, 8, 10, 11, 14, 15, 22, 25, 27, 28, 29 ^b	14	13 (93%)	1 (7%)
Dinâmico-10	1, 3, 4 ^a , 5, 6, 7, 8, 10, 11, 12, 13, 14, 15, 16, 17, 22, 24, 25, 26, 27, 28, 29 ^b , 31 ^b	23	21 (91%)	2 (9%)

^(a) Inibidor Bz-nKRR-H.^(b) Composto inativo ($K_i > 500 \mu\text{M}$).

Analisando os resultados da Tabela 4.24 observa-se que nenhum dos modelos foi capaz de selecionar todos os compostos ativos da série de peptidomiméticos. Como esperado, a maioria dos modelos farmacofóricos foi capaz de selecionar o inibidor Bz-nKRR-H. Embora os modelos Estático-1, Estático-3, Dinâmico-3, Dinâmico-4, Dinâmico-6 e Dinâmico-10 tenham selecionado grande parte dos compostos ativos, estes modelos também selecionaram compostos inativos. O fato de tanto compostos ativos como inativos terem sido selecionados sugere que estes modelos farmacofóricos seriam pouco seletivos. Em contrapartida, os modelos farmacofóricos Estático-2 e Dinâmico-2 não selecionaram qualquer composto inativo, sendo ainda capazes de selecionar 32 % e 21 % dos compostos ativos, respectivamente. É interessante observar que estes dois modelos selecionaram *hits* diferentes, exceto para os *hits* número **4** (inibidor Bz-nKRR-H) e número **10**. A aplicação, concomitante, dos modelos Estático-2 e Dinâmico-2 permite selecionar cerca de 46 % dos compostos ativos. Estes resultados sugerem que os modelos Estático-2 (estrutura 3U1I) e Dinâmico-2 (simulações de DM) se complementam permitindo identificar um número maior de inibidores da NS2B/NS3pro do DENV.

Aplicação dos modelos farmacofóricos Estático-2 e Dinâmico-2 ao banco ZINC-Curated

Considerando os resultados obtidos na etapa de avaliação do desempenho dos 16 modelos farmacofóricos gerados, apenas os modelos Estático-2 e Dinâmico-2 foram utilizados como filtro neste terceiro modelo de busca virtual. Procedendo-se como descrito no item 3.5.4.1, os dois modelos foram aplicados ao banco *NCI* usando o programa *LigandScout* (WOLBER; LANGER, 2005). Durante esta etapa, o programa selecionou, respectivamente, 21.701 (redução de 91,64%) e 17.005 (redução de 93,45%) compostos capazes de satisfazer as características mínimas (ou seja, as duas regiões de interação obrigatórias) requeridas pelos modelos farmacofóricos Estático-2 e Dinâmico-2. Dos 21.701 compostos selecionados pelo modelo farmacofórico Estático-2, apenas 156 (redução de 99,94 %) foram considerados *hits* (compostos capazes de respeitar seis ou mais regiões de interação). De modo análogo, dos 17.005 compostos selecionados pelo modelo farmacofórico Dinâmico-2, apenas 52 (redução de 99,69 %) foram considerados *hits*.

4.3.5.2. Filtro de “grupos funcionais”

Os 208 *hits* selecionados foram submetidos a um filtro de “grupos funcionais”, de modo análogo ao descrito para o segundo modelo de busca virtual (item 4.3.4.2), visando excluir tanto compostos com grupos com carga negativa em pH 9,0 (condições do experimento de inibição) bem como compostos contendo grupos pouco estáveis nas condições do experimento. De modo análogo ao descrito no item 4.3.4.2, numa primeira etapa, foram removidos os compostos contendo os seguintes grupos funcionais: dicarbonílicos; ácidos carboxílicos; grupos protetores (ex. acetal); grupos contendo N_{sp^3} - N_{sp^3} e ésteres de ácidos carboxílicos suscetíveis à hidrólise que poderiam formar os ácidos correspondentes. O número de compostos eliminados por cada um dos grupos funcionais

utilizados neste filtro está apresentado na Tabela 4.25. Todos compostos eliminados nesta etapa foram submetidos à inspeção visual, na qual se verificou, em especial, se os ésteres de ácidos carboxílicos pareciam ser suscetíveis à hidrólise.

Tabela 4.25. Número de compostos removidos por cada um dos grupos funcionais usados neste filtro.

Grupo funcional usado	Número de compostos removidos
Dicarbonílicos	2
Ácidos carboxílicos	39
Grupos protetores	1
Grupo contendo N_{sp3} - N_{sp3}	2
Ésteres de ácidos carboxílicos ^(a)	11

^(a) Suscetíveis à hidrólises.

Em uma segunda etapa, 37 compostos cujas espécies, ainda, apresentavam carga negativa em pH 9,0 foram eliminados. A aplicação deste filtro resultou em 116 compostos (redução de 44,23%) que foram, a seguir, submetidos a uma análise por inspeção visual.

4.3.5.3. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV

Como discutido na Introdução e de modo análogo aos demais modelos de busca virtual, os 116 compostos selecionados no filtro de “grupos funcionais” foram analisados, individualmente, por inspeção visual para verificar seus ajustes no sítio ativo da NS2B/NS3pro do DENV. Embora os compostos do banco *NCI* estejam disponíveis gratuitamente, sua remessa depende do composto estar, ou não, disponível no momento da solicitação. Assim, como descrito no item 3.5.4.11 (Materiais e Métodos), os compostos foram analisados quanto a sua disponibilidade para remessa. Dos 116 compostos selecionados, apenas 44 (37,93 %) estavam disponíveis em quantidade suficiente (> 5 mg). A seguir, estes 44 compostos foram analisados por inspeção visual para verificar seus ajustes

aos correspondentes modelos farmacofóricos (Estático-2 ou Dinâmico-2) utilizados na etapa de seleção de cada composto.

Nesta etapa, 15 compostos foram selecionados como potenciais inibidores da NS2B/NS3pro do DENV, representando uma redução de 87,07 % em relação aos 116 compostos selecionados na etapa anterior. Todos estes 15 potenciais inibidores foram adquiridos para validação experimental do Modelo III - DENV. Por fim, a aplicação de todos os filtros de seleção, descritos acima, representou uma redução total de mais de 99 % em relação ao banco *NCI* ($\sim 259 \times 10^3$ compostos). A Tabela 4.26 mostra a representação das estruturas dos quinze compostos (denominados **DV-27** a **DV-41**) selecionados pelo Modelo III – DENV que foram adquiridos e submetidos a testes de inibição enzimática frente à NS2B/NS3pro do DENV.

Tabela 4.26. Representação da estrutura 2D dos 15 compostos (DV-27 a DV-41) selecionados pelo Modelo III – DENV como potenciais inibidores da NS2B/NS3pro do DENV e que foram adquiridos e submetidos ao teste de inibição enzimática frente a esta protease. Estas estão acompanhadas dos seus correspondentes códigos *NCI* e valores de peso molecular.

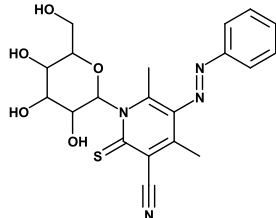
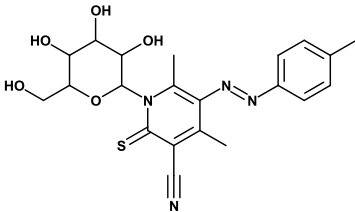
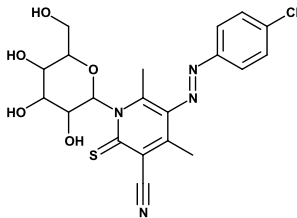
#Ref	Estrutura	Código NCI	Peso Molecular
DV-27		NSC667731	430,48
DV-28		NSC667732	444,50
DV-29		NSC667734	464,92

Tabela 4.26. (Continua).

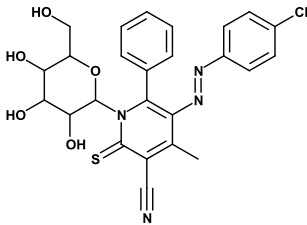
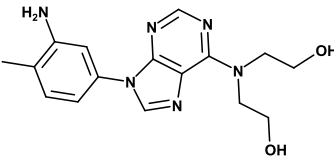
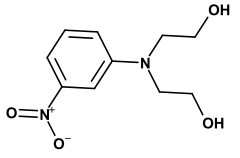
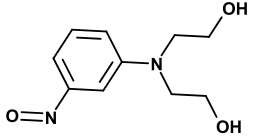
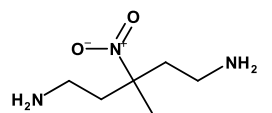
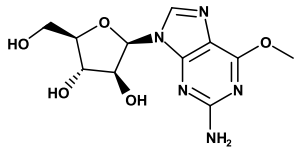
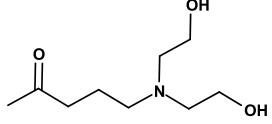
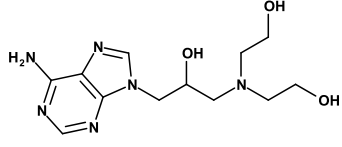
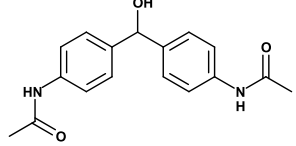
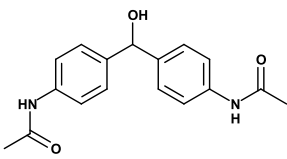
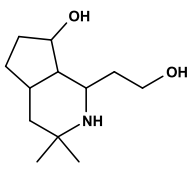
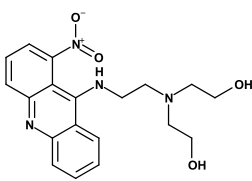
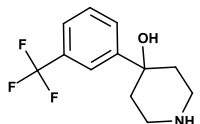
#Ref	Estrutura	Código NCI	Peso Molecular
DV-30		NSC667738	526,99
DV-31		NSC100107	328,37
DV-32		NSC103358	226,23
DV-33		NCS04966	210,23
DV-34		NCS116805	161,20
DV-35		NCS132361	297,27
DV-36		NCS140276	189,27
DV-37		NCS371365	296,32
DV-38		NCS4288	298,34

Tabela 4.26. (Continua).

#Ref	Estrutura	Código NCI	Peso Molecular
DV-38		NCS4288	298,34
DV-39		NCS54839	213,32
DV-40		NCS645806	370,41
DV-41		NCS83235	281,71

4.3.5.4. Validação experimental

De modo análogo aos demais modelos, a verdadeira utilidade do modelo Modelo III – DENV foi comprovada com sua validação experimental, através de ensaios enzimáticos. Assim, os 15 compostos selecionados por este modelos nos foram disponibilizados pelo *Drug Synthesis and Chemistry Branch* – que é parte do *Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute (NCI)* – e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV2. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Wibke Diederich (Philipps-Universität de Marburg, Alemanha) e estão descritos no item 3.6.1.

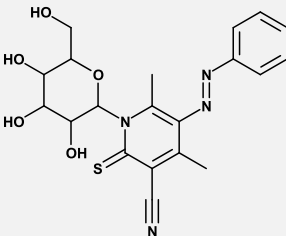
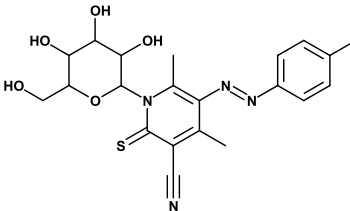
Ensaio de inibição enzimática frente à NS2B/NS3pro do DENV

Procedendo-se como descrito 3.6.1.1 (Materiais e Métodos), inicialmente, foram determinados os valores de porcentagem de inibição da NS2B/NS3pro do DENV2 de cada composto (em 100 μ M), sendo apresentados na Tabela 4.27. Ainda, como proposto no item

3.6.1.1 somente compostos com valores de atividade inibitória maiores, ou próximos a 50%, foram considerados como tendo atividade inibitória significativa.

Assim, dos 15 compostos adquiridos, quatro (26,67%) mostraram atividade inibitória significativa frente à NS2B/NS3pro do DENV2 na concentração utilizada (100 μ M), sendo eles: compostos **DV-27** (NCS667731), **DV-29** (NCS667734), **DV-30** (NCS667738) e **DV-33** (NCS103358). Os valores de porcentagem de inibição dos compostos **DV-27**, **DV-29**, **DV-30**, e **DV-33** (em 100 μ M) foram iguais a 52 ± 12 ; 67 ± 2 ; 66 ± 3 e 74 ± 3 , respectivamente. Por outro lado, onze deles não foram considerados significativamente ativos, sendo eles: compostos **DV-28** (NCS667732), **DV-31** (NCS100107), **DV-32** (NCS04966), **DV-34** (NCS116805), **DV-35** (NCS132361), **DV-36** (NCS4288), **DV-37** (NCS140276), **DV-38** (NCS371365), **DV-39** (NCS54839), **DV-40** (NCS645806) e **DV-41** (NCS83235).

Tabela 4.27. Valores de porcentagem de inibição^(a) (em 100 μ M) e dos correspondentes desvios padrão^(b) de cada um dos compostos (DV-27 a DV-41) selecionados pelo Modelo III – DENV.

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M) ^(a,b)
DV-27^(c)		52 ± 12
DV-28		42 ± 12

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^{da} Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Tabela 4.27. (Continua).

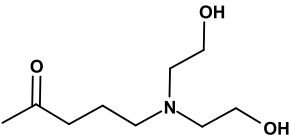
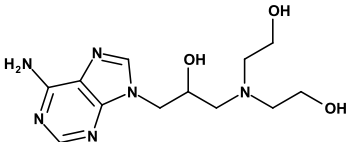
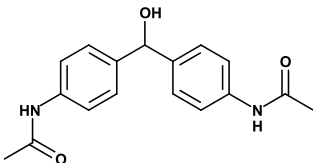
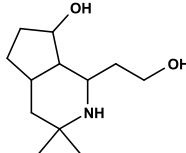
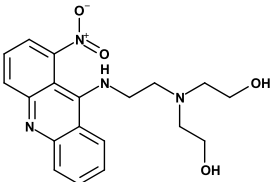
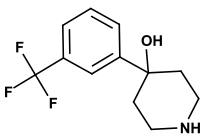
#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M) ^(a,b)
DV-29 ^(c)		67 \pm 2
DV-30 ^(c)		66 \pm 3
DV-31		0
DV-32		15 \pm 6
DV-33 ^(c)		74 \pm 3
DV-34		10 \pm 5
DV-35		13 \pm 6

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a. Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Tabela 4.27. (Continua).

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M) ^(a,b)
DV-36		0
DV-37		10 \pm 5
DV-38		20 \pm 6
DV-39		0
DV-40		30 \pm 6
DV-41		10 \pm 2

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a. Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

Em seguida, para cada um dos quatro compostos ativos (**DV-27**, **DV-29**, **DV-30** e **DV-33**), o valor de IC₅₀ foi determinado a partir de sua curva dose-resposta construída como descrito no item 3.6.1.2 (Materiais e Métodos) e apresentada na Figura 4.37. Estes valores estão apresentados na Tabela 4.28.

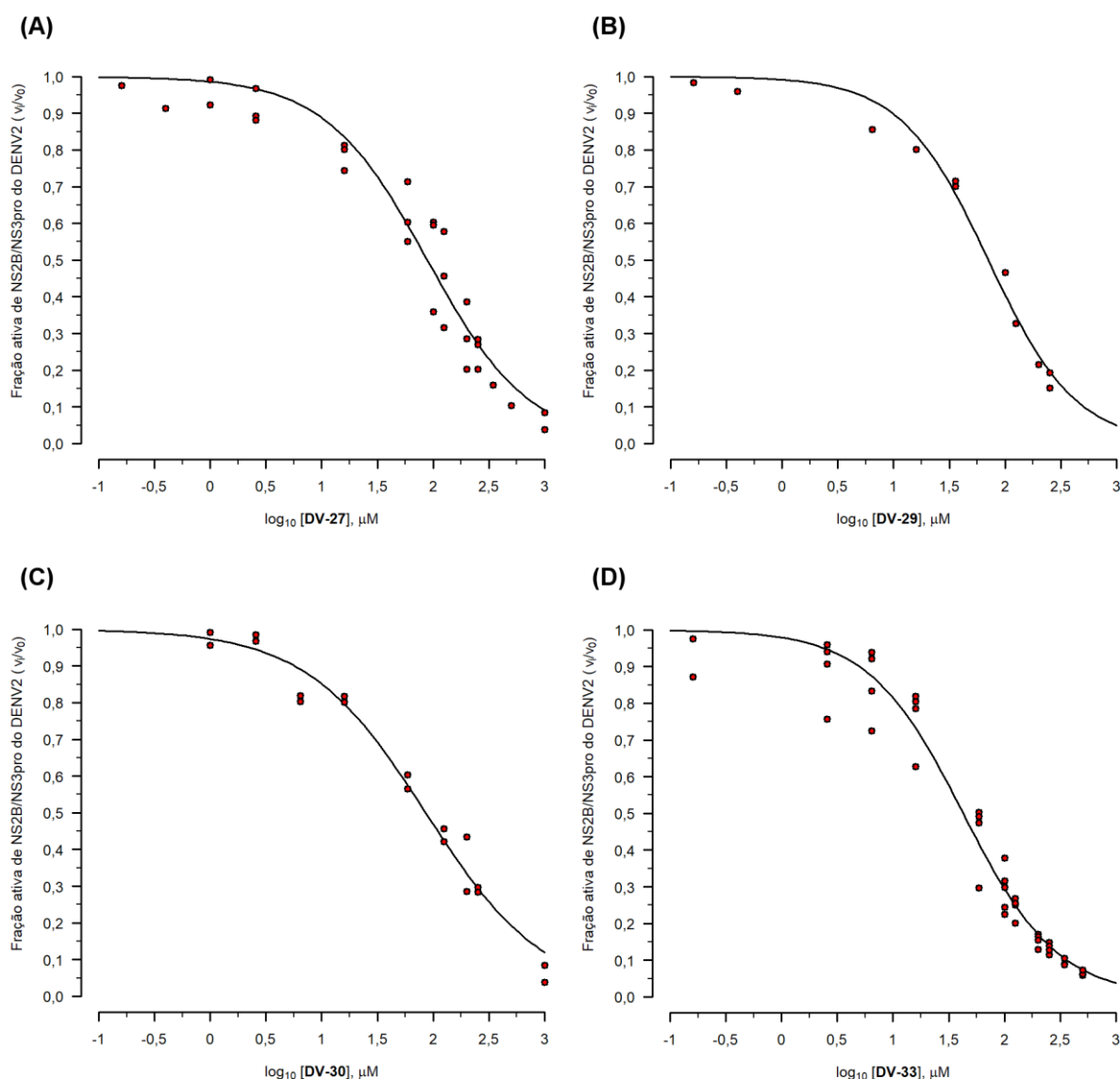
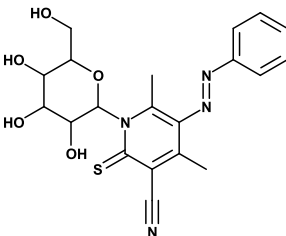
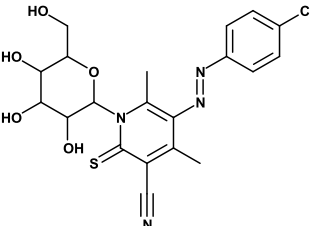
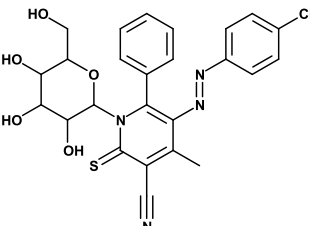
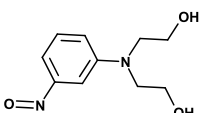


Figura 4.37. Curvas dose-resposta^(a) dos compostos DV-27, DV-29, DV-30 e DV-33. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-27 (A), DV-29 (B), DV-30 (C) e DV-33 (D). Experimentos realizados, pelo menos, em duplicata, exceto para DV-29.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: DV-27, DV-29, DV-30 e DV-33.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 1000,00 μM.
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (20 μM), à 37 ± 0,1°C e utilizando-se $\lambda_{em} = 380\text{nm}$ e $\lambda_{ex} = 460\text{ nm}$.

Tabela 4.28. Valores de IC_{50} frente à NS2B/NS3pro do DENV2 (μM) e dos correspondentes erros padrão de ajuste^(a) obtidos para os compostos DV-27, DV-29, DV-30 e DV-33.

#Ref	Estrutura	IC_{50} (μM)
DV-27		88 ± 8
DV-29		71 ± 5
DV-30		86 ± 7
DV-33		42 ± 3

^(a) Calculado pelo pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) do programa R (R CORE TEAM, 2013).

Analisando a Tabela 4.28 podemos observar que os inibidores **DV-27**, **DV-29** e **DV-30** apresentaram valores de IC_{50} comparáveis considerando-se o erro padrão de ajuste ($\sim 80 \mu M$). Considerando que estes três inibidores são estruturalmente semelhantes, os resultados obtidos indicam que as substituições tanto de um grupo metila pela fenila na porção central da molécula bem como de um átomo de hidrogênio pelo cloro na posição 4 do anel benzênico (porção lateral da molécula) têm pouco efeito sobre a atividade destes compostos, dentro do erros de ajuste. É interessante observar que o inibidor **DV-33** foi o mais ativo ($IC_{50} = 42 \pm 3 \mu M$ frente à NS2B/NS3pro do DENV2), embora seja bem menor que os inibidores **DV-27**, **DV-29** e **DV-30**.

Adicionalmente, para o **DV-33**, o valor de IC_{50} frente à NS2B/NS3pro do DENV3 foi determinado, de modo análogo, a partir da curva dose-resposta apresentada na Figura 4.38, sendo este igual a $40 \pm 2 \mu\text{M}$. Ainda, vale observar que embora o composto **DV-32** seja, estruturalmente, análogo ao composto **DV-33** – contendo um grupo nitro no lugar de um nitroso – o primeiro (**DV-32**) não mostrou atividade significativa frente à NS2B/NS3pro do DENV2. Estes resultados sugerem que a inibição apresentada pelo **DV-33** pode estar relacionada a uma maior reatividade do grupo nitroso, nas condições do experimento.

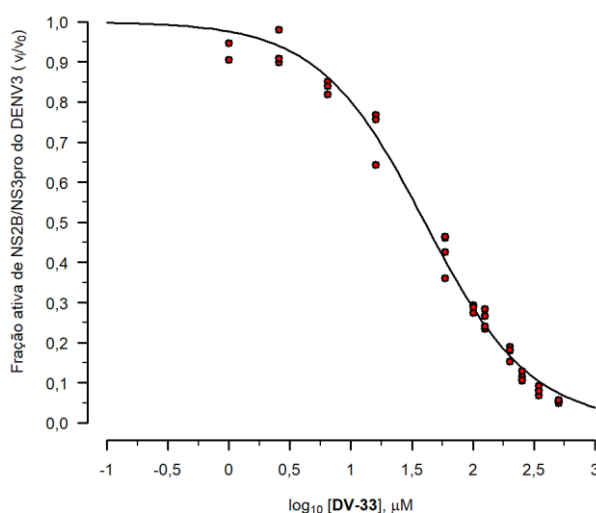


Figura 4.38. Curva dose-resposta^(a) do inibidor **DV-33** frente à NS2B/NS3pro do DENV3. Fração ativa^(b) de NS2B/NS3pro do DENV3 em função dos valores do logaritmo da concentração^(c) (em μM) do composto **DV-33**. Experimentos realizados, pelo menos, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa *R* (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), **DV-33**, determinado a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s).
- (c) Os valores da concentração de **DV-33** utilizados no teste variaram de 0,00 a 500,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV3 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (40 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{ nm}$.

Avaliação da estabilidade dos inibidores nas condições do experimento de inibição frente à da NS2B/NS3pro do DENV2/DENV3

Dos quatro inibidores selecionados pelo Modelo III – DENV, os inibidores **DV-27**, **DV-29** e **DV-30** são sensíveis à luz, constando da descrição disponibilizada pelo *Drug Synthesis and Chemistry Branch* – que é parte do *Developmental Therapeutics Program, Division of Cancer Treatment and Diagnosis, National Cancer Institute (NCI)*. Assim, estes compostos bem como suas soluções estoques (em DMSO) foram mantidas protegidas da luz. Cabe ressaltar que na construção do banco *NCI* não foram incluídos filtros físico-químicos e/ou estruturais visando remover compostos com grupos funcionais reativos (ex. N=N), o que poderia justificar a falta de estabilidade observada. Ainda, embora reconhecendo a importância de analisar estes quatro compostos do *NCI*, quanto a sua estabilidade e grau de pureza, estas não foram realizadas devido à falta de material.

Embora três dos quatro inibidores identificados pelo Modelo III – DENV possam ter problemas com a estabilidade (sensíveis à luz), os inibidores por ele identificados são mais ativos do que os encontrados pelo Modelo I – DENV. Assim, ao se considerar a flexibilidade conformacional do inibidor Bz-nKRR-H durante a criação dos modelos farmacofóricos, em conjunto com a escolha de modelos baseando-se numa avaliação de desempenho (validação *in silico*), permitiu construir um modelo de busca virtual bem melhor do que os anteriores.

4.3.6. Modelos de busca virtual: Modelo IV - DENV

Considerando os resultados preliminares não promissores obtidos para alguns compostos antes da visita aos laboratórios da Prof^a. W. Diederich, Philipps-Universität de Marburg, Alemanha (de 30/01/2017 a 24/02/2017), foi construído um quarto modelo de busca virtual (Modelo IV – DENV). Procedendo-se como descrito no item 3.5.4.12 (Materiais e Métodos), neste quarto modelo foram introduzidas algumas modificações, a saber: (i) nos

tipos de filtros de seleção aplicados ao banco de moléculas virtuais; (ii) na classe de compostos usada para criar os modelos farmacofóricos; (iii) na série de compostos usados na análise do desempenho dos mesmos e (iv) nos critérios de cada um destes filtros de seleção.

Assim, partindo-se do banco de dados *ZINC-Curated*, foram aplicados dois filtros virtuais para selecionar inibidores de NS2B/NS3pro do DENV, a saber: farmacofórico (previamente validados *in silico*) e inspeção visual. Diferentemente dos demais modelos de busca virtual, este modelo foi criado e avaliado quanto ao desempenho utilizando inibidores não peptídicos da literatura. Os compostos selecionados pelo Modelo IV - DENV foram testados como inibidores da NS2B/NS3pro do DENV. A Figura 4.39 mostra uma representação esquemática da sequência de aplicação de cada filtro bem como os valores da porcentagem de redução do número de compostos em cada etapa.

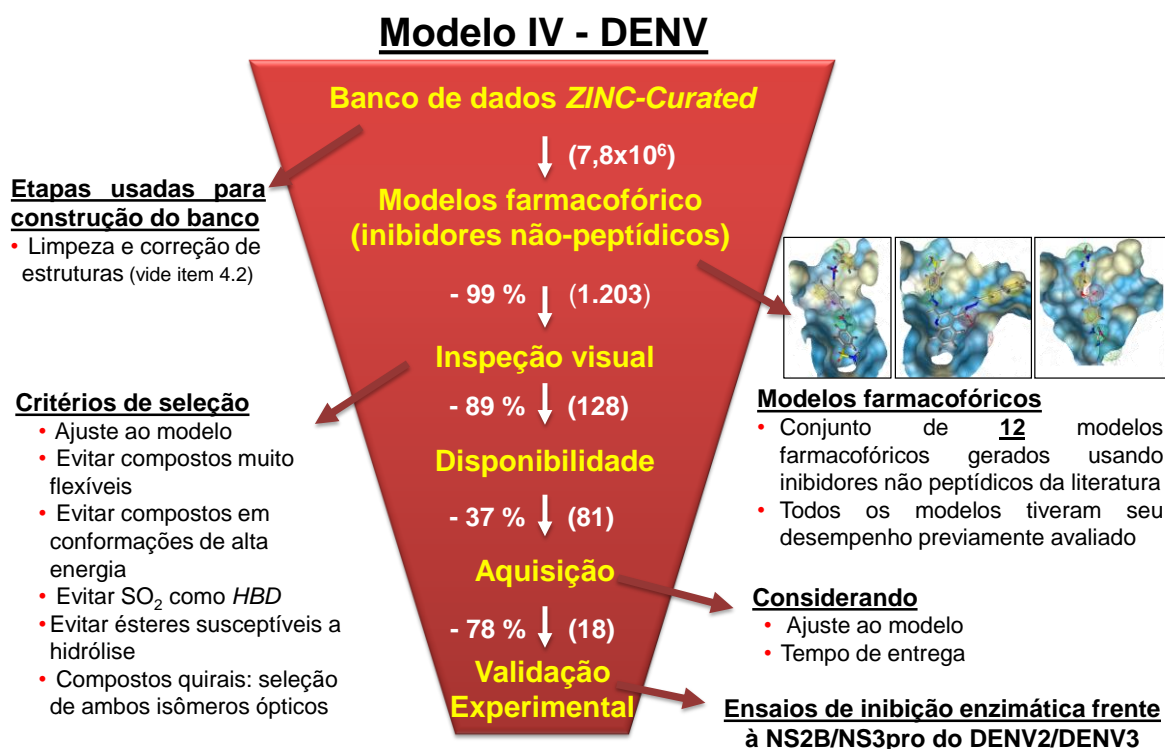


Figura 4.39. Representação esquemática da sequência de filtros de seleção aplicados ao banco *ZINC-Curated* usada no quarto modelo de busca virtual (Modelo IV – DENV). Entre parênteses são indicados os números de compostos selecionados em cada etapa e ao lado a porcentagem de redução por cada filtro.

4.3.6.1. Filtro farmacofórico

De modo análogo ao Modelo III – DENV, a primeira etapa do procedimento de busca virtual foi a aplicação de um filtro farmacofórico ao banco *ZINC-Curated*. Para tanto, procedendo-se como descrito no item 3.5.2., uma série de compostos não peptídicos, já avaliados experimentalmente quanto a sua atividade inibitória frente à NS2B/NS3pro do DENV, foi construída usando dados de literatura (obtidos diretamente de artigos científicos ou do banco ChEMBL (BENTO et al., 2014)). Estes compostos foram classificados de acordo com sua atividade em: ativos ($K_i/IC_{50} < 50 \mu M$), pouco ativos (K_i/IC_{50} entre 50 e 300 μM), inativos ($K_i/IC_{50} > 300 \mu M$) e não-competitivos. Adicionalmente, visando considerar os resultados negativos obtidos nos dois primeiro modelos (Modelo I – DENV e Modelo II – DENV), os compostos **DV-01** a **DV-26** foram incluídos nesta série, agora, como inativos. A Tabela 4.29 resume a composição final desta série, indicando, entre parênteses, o número de compostos obtidos diretamente de artigos da literatura.

Tabela 4.29. Composição da série de compostos não peptídicos, classificados como: ativos ($K_i/IC_{50} < 50 \mu M$); pouco ativos (K_i/IC_{50} entre 50 a 300 μM); inativos ($K_i/IC_{50} > 300 \mu M$) e, não competitivos frente à NS2B/NS3pro do DENV. O número entre parênteses corresponde ao número de compostos obtidos diretamente de artigos da literatura.

Série de compostos	Classificados segundo sua atividade como			
	Ativos	Pouco Ativos	Inativos	Não competitivos
Não peptídicos	650 (95)	383 (174)	293 (167)	9 (9)

Considerando-se que os compostos disponíveis no banco ChEMBL (BENTO et al., 2014) nem sempre possuem uma descrição detalhada dos procedimentos experimentais usados, optamos por considerar como ativos apenas os selecionados diretamente de artigos da literatura, resultando 95 compostos. Em seguida, dos 95 compostos classificados como ativos, quinze ($K_i/IC_{50} < 10 \mu M$) foram selecionados para serem usados como referência para construção dos outros modelos farmacofóricos. A representação das estruturas 2D destes 15

inibidores (denominados **LIT-01** a **LIT-15**), seus valores de atividade frente à NS2B/NS3pro do DENV bem como a correspondente literatura estão apresentados na Tabela 4.30.

Tabela 4.30. Representação da estrutura 2D, valor da atividade inibitória frente à NS2B/NS3pro do DENV de cada um dos 15 inibidores (LIT-01 a LIT-15) selecionados da literatura usados na construção dos modelos farmacofóricos do Modelo IV- DENV.

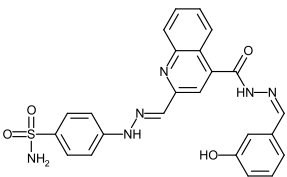
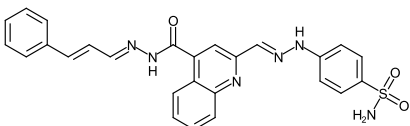
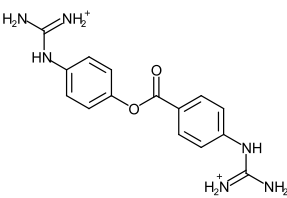
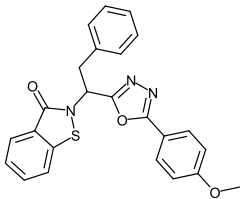
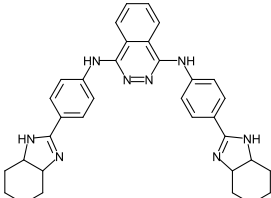
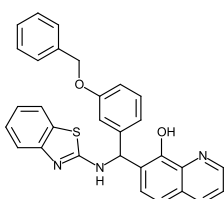
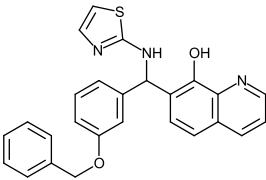
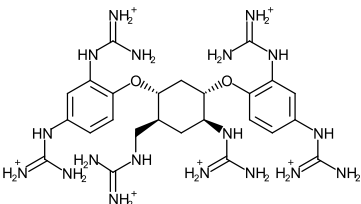
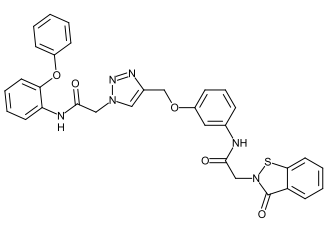
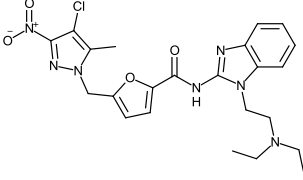
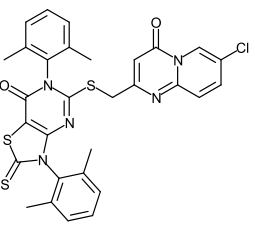
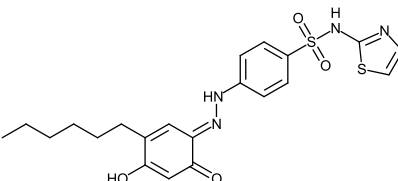
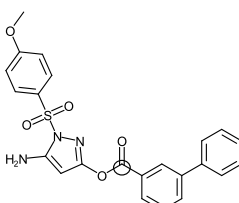
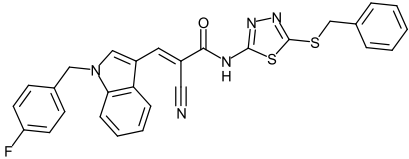
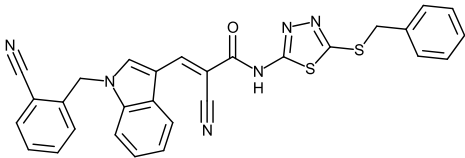
Inibidor número	Estrutura	Atividade frente à NS2B/NS3pro do DENV
LIT-01		IC ₅₀ = 7,8 µM (DENG et al., 2012)
LIT-02		IC ₅₀ = 7,5 µM (DENG et al., 2012)
LIT-03		K _i = 2,0 µM (KNEHANS et al., 2011)
LIT-04		80 % (em 25 µM) (LAI et al., 2013)
LIT-05		IC ₅₀ = 1,1 µM (BODENREIDER et al., 2009)
LIT-06		IC ₅₀ = 1,3 µM (LAI; SRIDHAR PRASAD; PADMANABHAN, 2013)

Tabela 4.30. (Continua).

Inibidor número	Estrutura	Atividade frente à NS2B/NS3pro do DENV
LIT-07		IC ₅₀ = 2,0 µM (LAI; SRIDHAR PRASAD; PADMANABHAN, 2013)
LIT-08		IC ₅₀ = 5,0 µM (CREGAR-HERNANDEZ et al., 2012)
LIT-09		IC ₅₀ = 3,5 µM (TIEW et al., 2012)
LIT-10		K _i = 4,0 µM (NGUYEN et al., 2013)
LIT-11		K _i = 3,4 µM (NGUYEN et al., 2013)
LIT-12		K _i = 2,0 µM (SHIRYAEV et al., 2011)
LIT-13 ^(a)		IC ₅₀ = 0,5 µM* (KOH-STENTA et al., 2015)

^(a) Inibidor covalente, a esfera indica o centro eletrofílico que é atacado pelo oxigênio da *Ser135*.

Tabela 4.30. (Continua).

Inibidor número	Estrutura	Atividade frente à NS2B/NS3pro do DENV
LIT-14		IC ₅₀ = 2,2 µM (LIU et al., 2014)
LIT-15		IC ₅₀ = 2,9 µM (LIU et al., 2014)

Construção e avaliação do desempenho dos modelos farmacofóricos utilizando inibidores não peptídicos da literatura

Como apresentado no item 4.2.1, não existem disponíveis, estruturas 3D experimentais da NS2B/NS3pro do DENV ligada a qualquer um dos 15 inibidores referidos acima (Tabela 4.30). Assim, como descrito no item 3.5.4.12 (Materiais e Métodos), o modo de ligação de cada um destes inibidores no sítio ativo da NS2B/NS3pro do DENV foi obtido por ancoramento. Inicialmente, todos os inibidores foram ancorados no sítio ativo da estrutura 3U1I e as poses de melhor ajuste foram utilizadas para gerar vários modelos farmacofóricos. De modo análogo ao Modelo III – DENV e procedendo-se como descrito no item 3.5.4.12, o desempenho de cada modelo foi avaliado *in silico* usando a série de compostos não peptídicos (Tabela 4.29). Nesta avaliação, os compostos classificados como pouco ativos, inativos e não competitivos foram considerados como *decoys*, resultando em 685 *decoys* e 95 compostos ativos. A Tabela 4.31 apresenta o número de compostos selecionados – total, ativos e *decoys* – para os modelos farmacofóricos de melhor desempenho gerados para cada um dos 15 inibidores usando a estrutura 3U1I. Estes modelos foram denominados MF-LIT01-3U1I a MF-LIT16-3U1I, de acordo com o inibidor e a estrutura usada no ancoramento.

Tabela 4.31. Número de compostos selecionados (total, ativos e *decoys*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 15 inibidores, usando as poses de ancoramento no sítio ativo da estrutura 3U1I.

Inibidor número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-01^(a)	MF-LIT01-3U1I	1	1	0
LIT-02^(a)	MF-LIT02-3U1I	1	1	0
LIT-03	MF-LIT03-3U1I	0	0	0
LIT-04	MF-LIT04-3U1I	78	24	54
LIT-05	MF-LIT05-3U1I	7	4	3
LIT-06^(a)	MF-LIT06-3U1I	6	6	0
LIT-07^(a)	MF-LIT07-3U1I	2	2	0
LIT-08	MF-LIT08-3U1I	3	1	2
LIT-09	MF-LIT09-3U1I	13	8	5
LIT-10	MF-LIT10-3U1I	7	5	2
LIT-11	MF-LIT11-3U1I	47	7	40
LIT-12	MF-LIT12-3U1I	24	1	23
LIT-13	MF-LIT13-3U1I	0	0	0
LIT-14	MF-LIT14-3U1I	5	1	4
LIT-15	MF-LIT15-3U1I	10	8	2

^(a) Modelos considerados com um desempenho razoável, destacados em cinza na tabela.

Dos 15 modelos farmacofóricos construídos, apenas quatro (MF-LIT01-3U1I, MF-LIT02-3U1I, MF-LIT06-3U1I e MF-LIT07-3U1I) não selecionaram qualquer *decoy* sendo, ainda, capazes de selecionar pelo menos um composto ativo (sendo este o próprio inibidor usado para construir o modelo). Embora estes quatro modelos não tenham sido capazes de selecionar um grande número de ativos, estes os modelos foram considerados como tendo um desempenho razoável, devido à diversidade estrutural na série de compostos. Sendo assim,

estes foram usados nas próximas etapas. Adicionalmente, os modelos MF-LIT03-3U1I e MF-LIT13-3U1I não selecionaram qualquer composto. Já os demais modelos que selecionaram um, ou mais, *decoys* foram considerados como tendo desempenho insatisfatório.

Assim, os compostos usados na construção destes últimos modelos foram novamente ancorados utilizando-se, agora, a estrutura Conf-1 obtida nas simulações de DM (descritas em detalhes no item 4.3.3.1). De modo análogo ao ancoramento usando a estrutura 3U1I, as poses de melhor ajuste foram utilizadas para gerar outros modelos farmacofóricos. O desempenho destes modelos foi subsequentemente analisado e os resultados dos melhores modelos (denominados de modo análogo aos anteriores) estão apresentados na Tabela 4.32.

Tabela 4.32. Número de compostos selecionados (total, ativos e *decoys*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 11 inibidores, usando as poses de ancoramento no sítio ativo da estrutura Conf-1 (obtida por simulações de DM).

Inibidor número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-03^(a)	MF-LIT03-Conf1	1	1	0
LIT-04	MF-LIT04-Conf1	10	7	3
LIT-05	MF-LIT05-Conf1	7	4	3
LIT-08^(a)	MF-LIT08-Conf1	1	1	0
LIT-09	MF-LIT09-Conf1	-	-	-
LIT-10^(a)	MF-LIT10-Conf1	1	1	0
LIT-11	MF-LIT11-Conf1	36	2	34
LIT-12	MF-LIT12-Conf1	0	0	0
LIT-13	MF-LIT13-Conf1	0	0	0
LIT-14	MF-LIT14-Conf1	0	0	0
LIT-15	MF-LIT15-Conf1	4	0	4

^(a) Modelos considerados com um desempenho razoável, destacados em cinza na tabela.

Três modelos farmacofóricos (MF-LIT03-Conf1, MF-LIT08-Conf1 e MF-LIT10-Conf1) não selecionaram qualquer *decoy* sendo, ainda, capazes de selecionar pelo menos um composto ativo (sendo este o próprio inibidor usado para construir o modelo). De modo análogo ao descrito acima, estes foram considerados com tendo um desempenho razoável podendo seguir para as próximas etapas. Já os demais modelos foram considerados como tendo desempenho insatisfatório, por não selecionarem qualquer ativo ou pelo menos um *decoy*. Assim, os compostos usados na construção destes últimos modelos foram novamente ancorados no sítio ativo da estrutura Conf-2 obtida nas simulações de DM (descritas em detalhes no item 4.3.3.1). De modo análogo ao descrito acima, as poses de melhor ajuste foram usadas para gerar outros modelos farmacofóricos. O desempenho destes modelos foi subsequentemente analisado e os resultados dos melhores modelos (denominados de modo análogo aos anteriores) estão apresentados na Tabela 4.33.

Tabela 4.33. Número de compostos selecionados (total, ativos e *decoys*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 8 inibidores, usando as poses de ancoramento no sítio ativo da estrutura Conf-2 (obtida por simulações de DM).

Inibidor número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-04	MF-LIT04-Conf2	14	8	6
LIT-05	MF-LIT05-Conf2	5	4	1
LIT-09	MF-LIT09-Conf2	0	0	0
LIT-11	MF-LIT11-Conf2	1	0	1
LIT-12	MF-LIT12-Conf2	0	0	0
LIT-13^(a)	MF-LIT13-Conf2	1	1	0
LIT-14	MF-LIT14-Conf2	0	0	0
LIT-15	MF-LIT15-Conf2	0	0	0

^(a) Modelo considerado com um desempenho razoável, destacado em cinza na tabela.

Apenas um modelo farmacofórico (MF-LIT13-Conf2) apresentou desempenho considerado como razoável, de acordo com os critérios definidos acima, podendo seguir para as próximas etapas. Os compostos usados para construir os demais modelos foram ancorados no sítio ativo da estrutura Conf-5 obtida nas simulações de DM (descritas em detalhes no item 4.3.3.1). Em seguida, as poses de melhor ajuste foram usadas para gerar outros modelos farmacofóricos. O desempenho destes modelos foi, subsequentemente, analisado e os resultados dos melhores modelos estão apresentados na Tabela 4.34.

Tabela 4.34. Número de compostos selecionados (total, ativos e *decoys*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 7 inibidores, utilizando as poses de ancoramento no sítio ativo da estrutura Conf-5 (obtida por simulações de DM).

Inibidores número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-04	MF-LIT04-Conf5	10	7	2
LIT-05	MF-LIT05-Conf5	1	0	1
LIT-09	MF-LIT09-Conf5	0	0	0
LIT-11	MF-LIT11-Conf5	0	0	0
LIT-12	MF-LIT12-Conf5	0	0	0
LIT-14^(a)	MF-LIT14-Conf5	1	1	0
LIT-15	MF-LIT15-Conf5	0	0	0

^(a) Modelo considerado com um desempenho razoável, destacado em cinza na tabela.

De modo análogo ao observado acima, apenas um modelo farmacofórico (MF-LIT14-Conf5) apresentou desempenho considerado como razoável, podendo seguir para as próximas etapas. Os compostos usados para construir os demais modelos foram ancorados no sítio ativo da estrutura DV1 obtida nas simulações de DM. Esta estrutura (DV1) corresponde a uma das conformações majoritárias obtidas na simulação DM utilizando a NS2B/NS3pro do DENV1 e descrita em detalhes em nosso artigo colocado em anexo (PICCIRILLO et al., 2016). De modo análogo ao ancoramento usando a estrutura 3U1I, as poses de melhor ajuste foram

utilizadas para gerar outros modelos farmacofóricos. O desempenho destes modelos foi, em seguida, analisado e os resultados dos melhores modelos estão apresentados na Tabela 4.35.

Tabela 4.35. Número de compostos selecionados (total, ativos e *decoys*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 6 inibidores, usando as poses de ancoramento no sítio ativo da estrutura DV-1 (obtida por simulações de DM).

Inibidor número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-04	MF-LIT04-DV1	9	7	2
LIT-05	MF-LIT05-DV1	2	0	2
LIT-09	MF-LIT09-DV1	0	0	0
LIT-11	MF-LIT11-DV1	0	0	0
LIT-12	MF-LIT12-DV1	0	0	0
LIT-15	MF-LIT15-DV1	0	0	0

Nenhum dos seis modelos farmacofóricos gerados foram capazes de selecionar apenas compostos ativos. Assim, o desempenho destes modelos foi considerado insatisfatório e estes seis compostos foram ancorados no sítio ativo da estrutura DV2 obtida nas simulações de DM. De modo análogo ao descrito para estrutura DV1, a estrutura DV2 corresponde à conformação majoritária obtidas na simulação de DM usando a NS2B/NS3pro do DENV2, descrita em detalhes em nosso artigo colocado como anexo (PICCIRILLO et al., 2016). De modo análogo ao descrito acima, as poses de melhor ajuste foram utilizadas para gerar outros modelos farmacofóricos. O desempenho destes modelos foi subsequentemente analisado e os resultados dos melhores modelos estão apresentados na Tabela 4.36.

Tabela 4.36. Número de compostos selecionados (total, ativos e *decoy*) pelos modelos farmacofóricos de melhor desempenho gerados para cada um dos 6 inibidores, usando as poses de ancoramento no sítio ativo da estrutura DV-2 (obtida por simulações de DM).

Inibidor número	Modelo farmacofórico	Número de compostos selecionados		
		Total	Ativos	<i>Decoys</i>
LIT-04	MF-LIT04-DV2	16	1	15
LIT-05	MF-LIT05-DV2	4	2	2
LIT-09	MF-LIT09-DV2	2	1	1
LIT-11	MF-LIT11-DV2	0	0	0
LIT-12	MF-LIT12-DV2	0	0	0
LIT-15	MF-LIT15-DV2	4	3	1

Infelizmente, os modelos gerados para **LIT-04**, **LIT-05**, **LIT-09**, **LIT-11**, **LIT-12** e **LIT-15** também não foram capazes de selecionar apenas compostos ativos. Considerando-se que todas as estruturas identificadas como relevantes em nossos estudos de DM já foram usadas no ancoramento, todos os modelos gerados para estes seis inibidores foram, novamente, analisados por inspeção visual. Nesta análise tentou-se identificar alguma diferença entre os ativos e os *decoys* selecionados pelo modelo que pudesse ser usada como critério para diferenciá-los. Para o modelo MF-LIT04-DV1, foi observada uma diferença entre os valores de *Pharmacophore Fit* calculados para os 7 compostos ativos (valores > 66) quando comparados aos calculados para os 2 *decoys* (valores < 66). Com base nesta diferença, foi considerado que este último (MF-LIT04-DV1) apresenta um desempenho razoável desde que os compostos selecionados tenham valores de *Pharmacophore Fit* > 66. Para o modelo MF-LIT05-Conf2, foi observado que o *decoy* selecionado é uma estrutura análoga à do inibidor **LIT-05** com $IC_{50} = 55 \mu M$ frente à NS2B/NS3pro do DENV2. Considerando que este valor de IC_{50} é próximo ao valor de corte utilizado para classificar/separar os compostos ativos dos pouco ativos, o modelo MF-LIT05-Conf2 foi considerado com tendo um desempenho razoável. Durante esta inspeção visual, não se encontrou qualquer diferença entre

os ativos e os *decoys* selecionados pelos modelos gerados usando os inibidores **LIT-09**, **LIT-11** e **LIT-12**. Assim, o desempenho de todos dos modelos gerados, com base nestes três inibidores, foi novamente considerado como sendo insatisfatório. Adicionalmente, na inspeção visual dos compostos selecionados pelo MF-LIT15-DV2 foi observado que o único *decoy* selecionado apresentava um átomo de oxigênio do grupo sulfonil ($-\text{SO}_2$) como acceptor de ligação de hidrogênio. No entanto, na literatura (BISSANTZ; KUHN; STAHL, 2010) é descrito que a maioria (74 a 80%) dos oxigênios do grupo sulfonil são encontrados próximos a grupos alifáticos. Isto sugere que estes têm um maior caráter hidrofóbico do que acceptor de ligação de hidrogênio. Com base nesta descrição (BISSANTZ; KUHN; STAHL, 2010), o *decoy* selecionado pelo MF-LIT15-DV2 foi desconsiderado e o desempenho deste modelo também foi considerado razoável, pois sem este *decoy* este modelo selecionou três compostos sendo todos ativos.

Por fim, considerando todos os modelos farmacofóricos gerados usando os 15 inibidores selecionados da literatura (Tabela 4.30), 12 modelos foram avaliados com um desempenho razoável. O número e a natureza da região de interação variam de acordo com o modelo farmacofórico, como apresentado nas Figura 4.40 e Figura 4.41, a saber:

- MF-LIT01-3U1I contém 6 regiões de interação obrigatórias, sendo 3 *HBD*, 1 *HBA* e 2 hidrofóbicas (Figura 4.40 (A));
- MF-LIT02-3U1I contém 7 regiões de interação obrigatórias, sendo 2 *HBD*, 1 *HBA*, 3 hidrofóbicas e 1 de interação com anéis aromáticos (Figura 4.40 (B));
- MF-LIT03-Conf1 contém 6 regiões de interação obrigatórias, sendo 2 *HBD*, 1 *HBA*, 2 hidrofóbicas e 1 de interação com anéis aromáticos (Figura 4.40 (C));
- MF-LIT04-DV1 contém 6 regiões de interação obrigatórias, sendo 1 *HBA*, 3 hidrofóbicas e 2 de interação com anéis aromáticos (Figura 4.40 (D));
- MF-LIT05-Conf2 contém 5 regiões de interação obrigatórias, sendo 2 *HBD*, 2 hidrofóbicas e 1 de interação iônica positiva (Figura 4.40 (E));

- MF-LIT06-3U1I contém 8 regiões de interação obrigatórias, sendo 2 *HBD*, 2 *HBA* e 4 hidrofóbicas (Figura 4.40 (F));
- MF-LIT07-3U1I contém 6 regiões de interação obrigatórias e 1 opcional, sendo 2 *HBD*, 2 *HBA*, 2 hidrofóbicas (desta uma é opcional) e 1 de interação com anéis aromáticos (Figura 4.41(A));
- MF-LIT08-Conf1 contém 8 regiões de interação, sendo 5 *HBD*, 1 hidrofóbica e 2 de interação iônica positiva (Figura 4.41(B));
- MF-LIT10-Conf1 contém 6 regiões de interação, sendo 1 *HBA*, 3 hidrofóbicas e 2 de interação com anéis aromáticos (Figura 4.41(C));
- MF-LIT13-Conf2 contém 6 regiões de interação, sendo 1 *HBD*, 1 *HBA*, 4 hidrofóbicas e 1 de interação com anéis aromáticos (Figura 4.41(D));
- MF-LIT14-Conf5; contém 5 regiões de interação, sendo 2 *HBA* e 3 hidrofóbicas (Figura 4.41(E));
- MF-LIT15-DV2 contém 6 regiões de interação, sendo 4 *HBA* e 3 hidrofóbicas (Figura 4.41(F)).

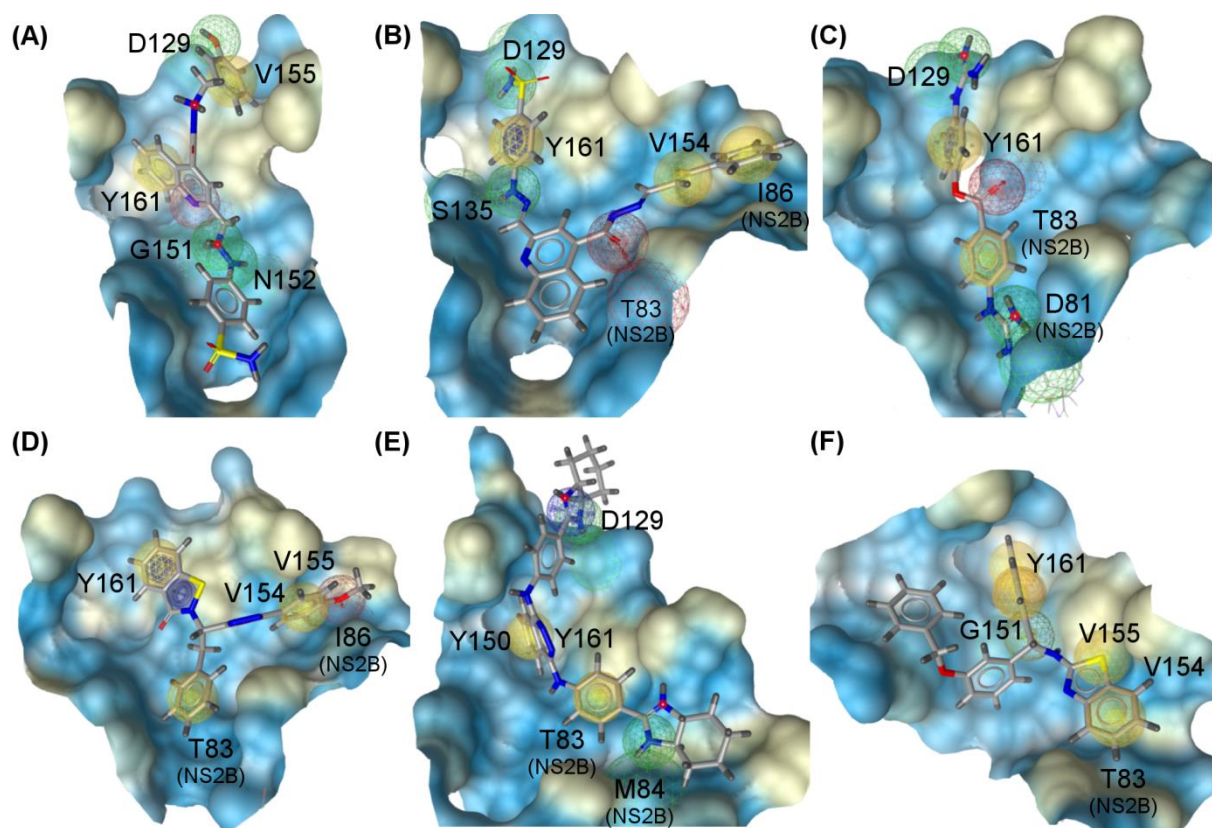


Figura 4.40. Representação esquemática dos modelos farmacofóricos^(a) de melhor desempenho gerados usando as poses de ancoramento^(b) dos inibidores^(c) LIT-01, LIT-02, LIT-03, LIT-04, LIT-05 e LIT-06. (A) MF-LIT01-3U1I com 6 regiões de interação^(d). (B) MF-LIT02-3U1I com 7 regiões de interação^(d). (C) MF-LIT03-Conf1 com 6 regiões de interação^(d). (D) MF-LIT04-DV1 com 6 regiões de interação^(d). (E) MF-LIT05-Conf2 com 5 regiões de interação^(d). (F) MF-LIT06-3U1I com 8 regiões de interação^(d). Todos os modelos contêm regiões de exclusão de volume^(e) que não estão sendo mostradas para facilitar a visualização dos modelos. As regiões HBD, HBA, hidrofóbicas, de interação aromática e iônica positiva estão representadas em verde, vermelho, amarelo, azul e azul, respectivamente. As setas e os círculos azuis indicam regiões de interação direcionais.

^(a) Gerados com o programa *LigandScout* (WOLBER; LANGER, 2005) a partir das interações identificadas entre cada inibidor e os resíduos do sítio ativo, utilizando os valores de corte padrão do programa.

^(b) Ancoramento realizado com o programa *GOLD* (JONES et al., 1997) no sítio ativo da protease usando as estruturas: 3U1I (cristalográfica) e Conf-1 Conf-2 e DV1, obtidas por simulações de DM (vide item 3.3.3.1).

^(c) Inibidores obtidos de artigos da literatura (vide Tabela 4.30).

^(d) Todas as regiões de interação possuem um raio de 1,5 Å.

^(e) Definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio ativo de cada estrutura.

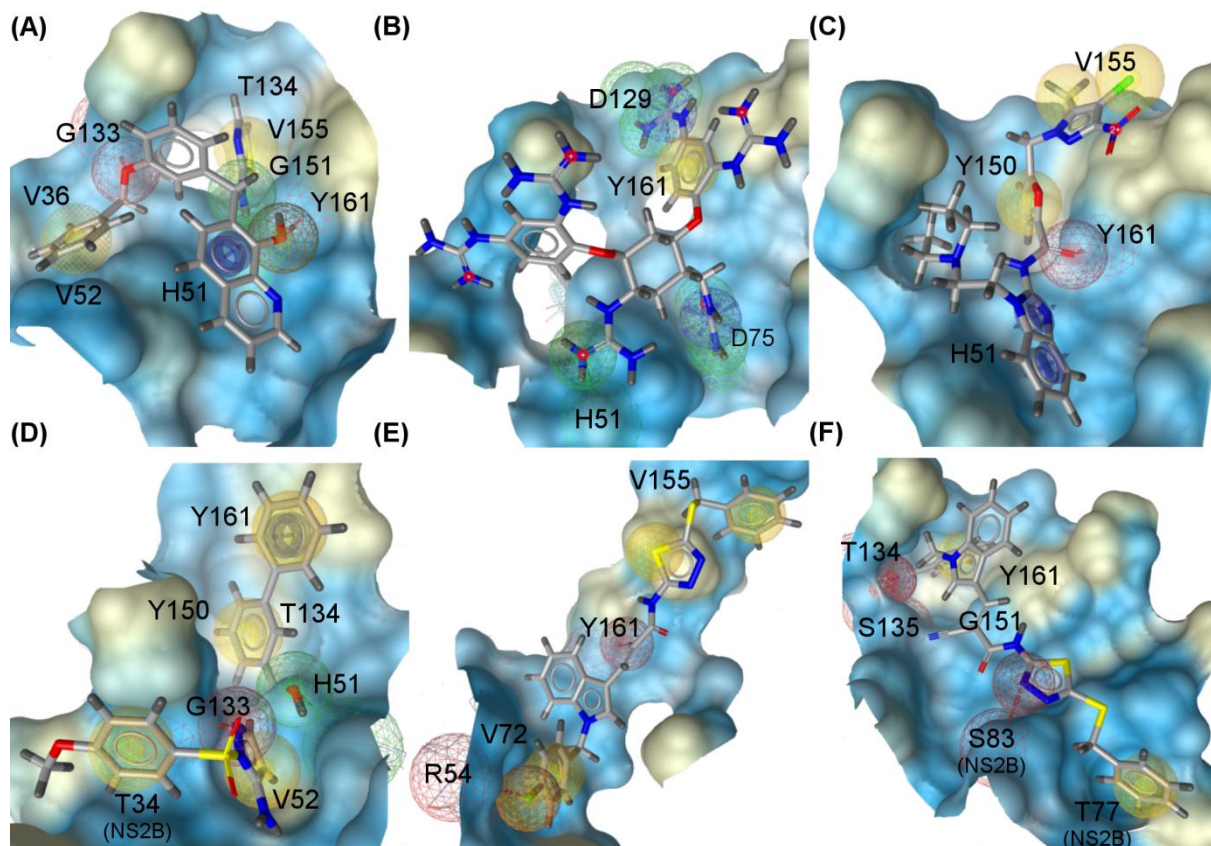


Figura 4.41. Representação esquemática dos novos modelos farmacofóricos^(a) de melhor desempenho gerados usando as poses de ancoramento^(b) dos inibidores^(c) LIT-07, LIT-08, LIT-10, LIT-13, LIT-14 e LIT-15. (A) MF-LIT07-3U1I com 6 regiões de interação^(d). (B) MF-LIT08-Conf1 com 8 regiões de interação^(d). (C) MF-LIT10-Conf1 com 6 regiões de interação^(d). (D) MF-LIT13-Conf2 com 6 regiões de interação^(d). (E) MF-LIT14-Conf5 com 5 regiões de interação^(d). (F) MF-LIT15-DV2 com 6 regiões de interação^(d). Todos os modelos contêm regiões de exclusão de volume^(e) que não estão sendo mostradas para facilitar a visualização dos modelos. As regiões *HBD*, *HBA*, hidrofóbicas, de interação aromática e iônica positiva estão representadas em verde, vermelho, amarelo, anéis azul e azul, respectivamente. As setas e os círculos azuis indicam regiões de interação direcionais.

- (a) Gerados com o programa *LigandScout* (WOLBER; LANGER, 2005) a partir das interações identificadas entre cada inibidor e os resíduos do sítio ativo, utilizando os valores de corte padrão do programa.
- (b) Ancoramento realizado com o programa *GOLD* (JONES et al., 1997) no sítio ativo da protease usando as estruturas: 3U1I (cristalográfica) e Conf-1, Conf-2, Conf-5 e DV2, obtidas por simulações de DM (vide item 4.3.3.1).
- (c) Inibidores obtidos de artigos da literatura (vide Tabela 4.30).
- (d) Todas as regiões de interação possuem um raio de 1,5 Å.
- (e) Definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio ativo de cada estrutura.

Aplicação dos modelos farmacofóricos gerados ao banco ZINC-Curated

Considerando os resultados obtidos na etapa de avaliação do desempenho dos modelos farmacofóricos, 12 modelos farmacofóricos (MF-LIT01-3U1I, MF-LIT02-3U1I, MF-LIT03-Conf1, MF-LIT04-DV1, MF-LIT05-Conf2, MF-LIT06-3U1I, MF-LIT07-3U1I, MF-LIT08-Conf1, MF-LIT10-Conf1, MF-LIT13-Conf2, MF-LIT14-Conf5 e MF-LIT15-DV2) foram

usados como filtro nesta busca virtual de inibidores da NS2B/NS3pro do DENV. Procedendo-se como descrito no item 3.5.4.12 (Materiais e Métodos), estes modelos foram aplicados, em sequência, ao banco *ZINC-Curated* utilizando o programa *LigandScout* (WOLBER; LANGER, 2005). Durante esta etapa, o programa selecionou 1.325 compostos capazes de satisfazer as características mínimas (ou seja, regiões de interação obrigatórias) requeridas por pelo menos um dos 12 modelos farmacofóricos. Destes 1.325 compostos, 122 (9,21%) foram eliminados considerando os critérios adicionais definidos para MF-LIT04-DV1 e MF-LIT15-DV2 na inspeção (vide página 292). Por fim, os 1.203 compostos (redução de 99,98%) restantes foram submetidos a um filtro de inspeção visual.

4.3.6.2. Inspeção visual e seleção dos compostos para testes de inibição frente à NS2B/NS3pro do DENV

De modo análogo aos três modelos de busca virtual, os 1.203 compostos selecionados no filtro farmacofórico foram analisados por inspeção visual para verificar seus ajustes no sítio ativo da NS2B/NS3pro do DENV. Esta análise foi feita como descrito no item 3.5.4.12 (Materiais e Métodos). Neste quarto modelo, diferentemente do descrito para os demais modelos, além do ajuste no sítio ativo, foram definidos outros critérios considerando-se: (i) a flexibilidade dos compostos; (ii) a seleção de apenas um estereoisômero (para compostos com centros quirais) e/ou de confôrmeros de alta energia e (iii) a presença de grupos sufonil como aceptores de ligação de hidrogênio e/ou de ésteres de ácidos carboxílicos susceptíveis à hidrólise. Uma descrição detalhada destes critérios está apresentada no item 3.5.4.12.

Nesta etapa, 129 compostos foram selecionados como potenciais inibidores da NS2B/NS3pro do DENV, representando uma redução de 89,28% em relação aos 1.203 compostos selecionados na etapa anterior. Dos 129 potenciais inibidores, 81 (62,79%) estavam disponíveis para aquisição imediata. Destes, 18 compostos (22,22) foram adquiridos para validação experimental. Por fim, a aplicação de todos os filtros de seleção representou

uma redução total de mais de 99 % em relação ao banco *ZINC-Curated* ($\sim 7,8 \times 10^6$ compostos). A Tabela 4.37 mostra a representação das estruturas dos dezoito compostos (denominados **DV-42** a **DV-59**) selecionados pelo Modelo IV – DENV que foram adquiridos e submetidos a testes de inibição enzimática frente à NS2B/NS3pro do DENV.

Tabela 4.37. Representação da estrutura 2D dos 18 compostos (DV-42 a DV-59) selecionados pelo Modelo IV – DENV como potenciais inibidores da NS2B/NS3pro do DENV que foram adquiridos e submetidos ao teste de inibição enzimática frente a esta protease. Estas estão acompanhadas dos seus correspondentes códigos ZINC, valores de peso molecular e empresa fornecedora.

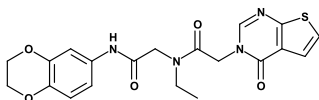
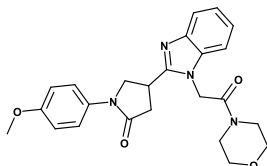
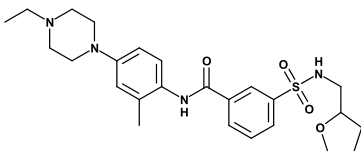
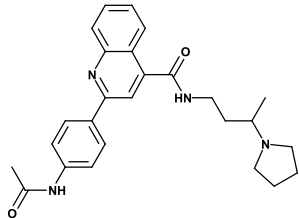
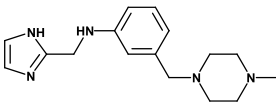
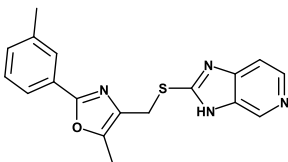
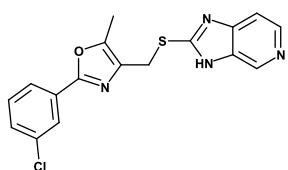
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-42		ZINC9591679	428,46	Enamine
DV-43		ZINC15922057	434,50	Enamine
DV-44		ZINC28731880	486,63	Enamine
DV-45		ZINC21662916	430,55	ChemDiv
DV-46		ZINC80196225	285,40	UORSY
DV-47		ZINC20865739	336,41	ChemDiv
DV-48		ZINC20865752	356,83	ChemDiv

Tabela 4.37. (Continua).

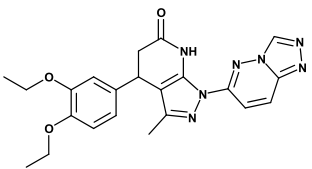
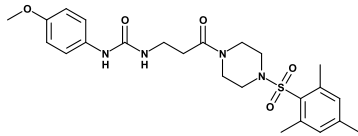
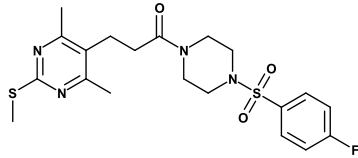
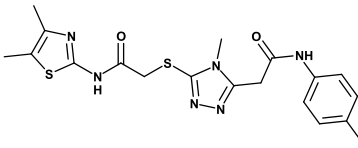
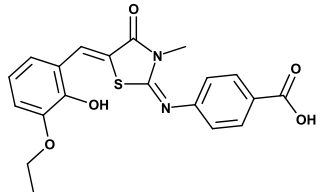
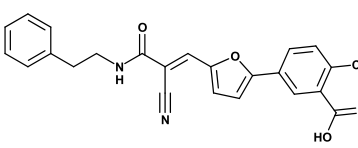
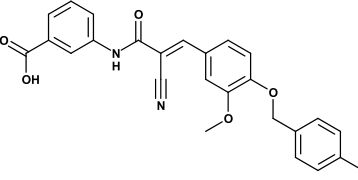
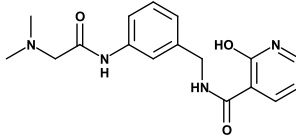
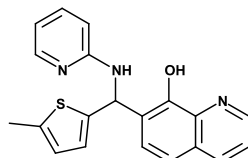
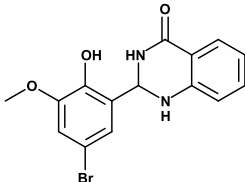
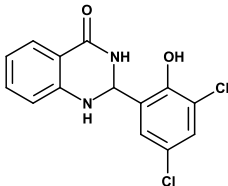
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-49		ZINC33088092	433,47	ChemDiv
DV-50		ZINC14109553	488,60	Enamine
DV-51		ZINC14070608	452,56	Enamine
DV-52		ZINC9302446	434,51	Vitas M Lab
DV-53		ZINC5032850	398,43	Vitas M Lab
DV-54		ZINC13424134	420,85	Vitas M Lab
DV-55		ZINC1438749	446,434	Vitas M Lab
DV-56		ZINC58166454	328,37	Enamine
DV-57		ZINC4123995	347,44	Vitas M Lab

Tabela 4.37. (Continua).

#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-58		ZINC9512483	349,18	Enamine
DV-59		ZINC9609087	309,15	Enamine

4.3.6.3. Validação experimental

De modo análogo aos demais modelos, a verdadeira utilidade do modelo Modelo IV – DENV na busca de compostos bioativos deve ser comprovada com sua validação experimental, através de ensaios enzimáticos. Assim, os 18 compostos selecionados pelo Modelo IV - DENV de busca virtual foram adquiridos e submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV2. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Wibke Diederich (Philipps-Universität de Marburg, Marburg, Alemanha) e estão descritos no item 3.6.1 (Materiais e Métodos).

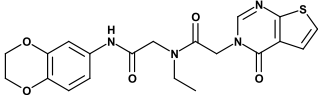
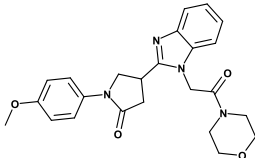
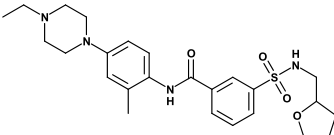
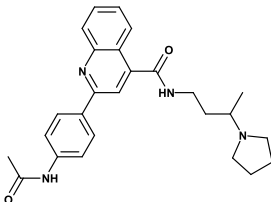
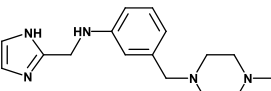
Ensaio de inibição enzimática frente à NS2B/NS3pro do DENV

Procedendo-se como descrito 3.6.1.1 (Materiais e Métodos), inicialmente, foram determinados os valores de porcentagem de inibição da NS2B/NS3pro do DENV2 de cada composto (em 100 μ M), sendo apresentados na Tabela 4.38. Ainda, como proposto no item 3.6.1.1, somente compostos com valores de atividade inibitória maiores, ou próximos a 50%, foram considerados como tendo atividade inibitória significativa.

Assim, dos 18 compostos adquiridos, quatro (22,22%) mostraram atividade inibitória significativa frente à NS2B/NS3pro do DENV2 na concentração utilizada (100 μ M), sendo eles: compostos **DV-53** (ZINC5032850), **DV-54** (ZINC13424134), **DV-55** (ZINC1438749) e

DV-57 (ZINC4123995). Os valores de porcentagem de inibição dos compostos **DV-53**, **DV-54**, **DV-55** e **DV-57** foram iguais a 62 ± 3 ; 82 ± 3 ; 89 ± 7 e 86 ± 5 , respectivamente. Por outro lado, quatorze deles não mostraram atividade inibitória significativa na concentração usada, sendo eles: compostos **DV-42** (ZINC9591679), **DV-43** (ZINC15922057), **DV-44** (ZINC28731880), **DV-45** (ZINC21662916), **DV-46** (ZINC80196225), **DV-47** (ZINC20865739), **DV-48** (ZINC20865752), **DV-49** (ZINC33088092), **DV-50** (ZINC14109553), **DV-51** (ZINC14070608), **DV-52** (ZINC9302446), **DV-56** (ZINC58166454), **DV-58** (ZINC9512483) e **DV-59** (ZINC9609087).

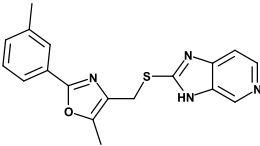
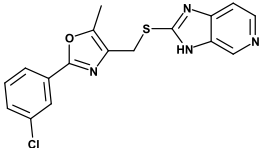
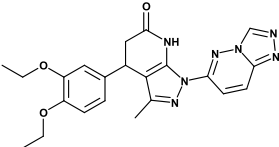
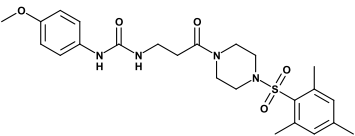
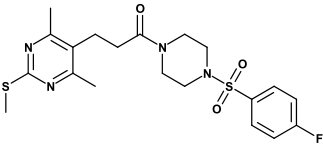
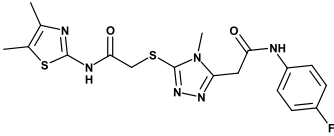
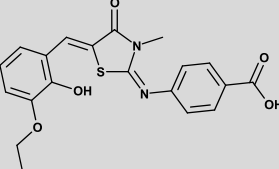
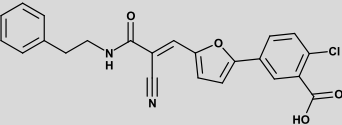
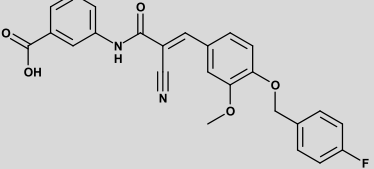
Tabela 4.38. Valores de porcentagem de inibição (em 100 μM)^(a) e dos correspondentes desvios padrão^(b) de cada um dos compostos (**DV-42** a **DV-59**) selecionados pelo Modelo IV – DENV.

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% em 100 μM) ^(a,b)
DV-42		31 ± 9
DV-43		6 ± 5
DV-44		3 ± 2
DV-45		10 ± 5
DV-46		0

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^{ra} Diederich), PhAc-KRR-MCA (20 μM) e cada um dos compostos (100 μM).

^(b) Considerando pelo menos três experimentos independentes.

Tabela 4.38. (Continua).

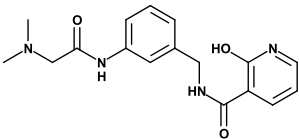
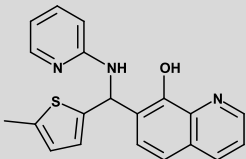
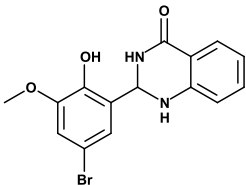
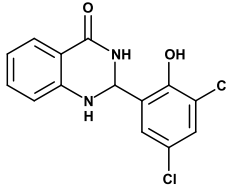
#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μ M) ^(a, b)
DV-47		0
DV-48		36 \pm 4
DV-49		32 \pm 7
DV-50		23 \pm 9
DV-51		20 \pm 12
DV-52		39 \pm 9
DV-53 ^(c)		62 \pm 3
DV-54 ^(c)		82 \pm 3
DV-55 ^(c)		89 \pm 7

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Tabela 4.38. (Continua)

#Ref	Estrutura	Inibição da NS2B/NS3pro do DENV2 (% em 100 μ M) ^(a, b)
DV-56		2 ± 1
DV-57 ^(c)		86 ± 5
DV-58		8 ± 6
DV-59		39 ± 8

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 μ M) e cada um dos compostos (100 μ M).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Em seguida, para cada um dos quatro compostos ativos (**DV-53**, **DV-54**, **DV-55** e **DV-57**), os valores de IC₅₀ foram determinados a partir de suas curvas dose-resposta construídas como descrito no item 3.6.1.2 (Materiais e Métodos) e apresentadas na Figura 4.42. Estes valores estão apresentados na Tabela 4.39.

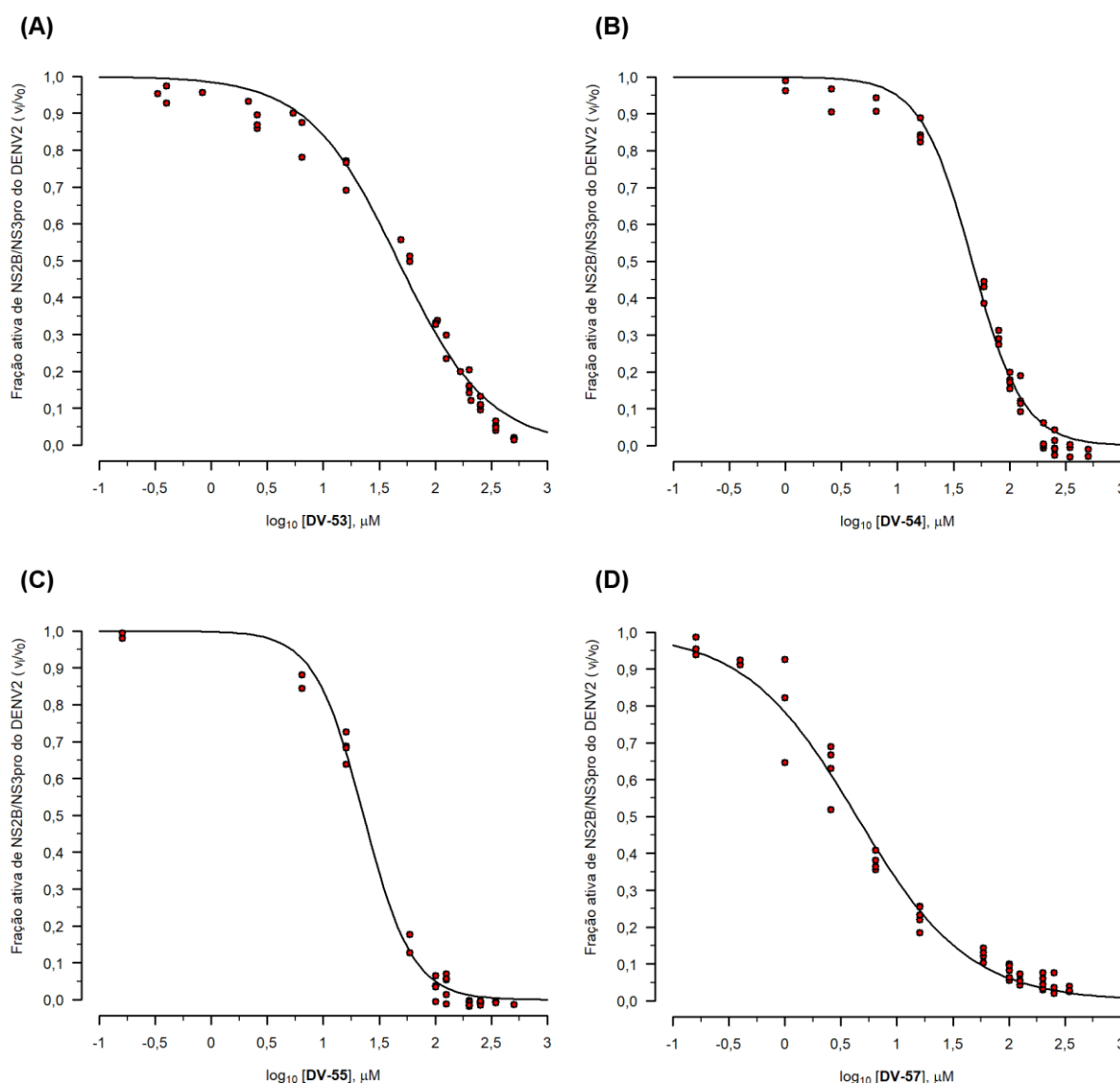


Figura 4.42. Curvas dose-resposta^(a) dos compostos DV-53, DV-54, DV-55 e DV-57. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-53 (A), DV-54 (B), DV-55 (C) e DV-57 (D). Experimentos realizados, pelo menos, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: DV-53, DV-54, DV-55 e DV-57.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (20 μM), à 37 $\pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{ nm}$.

Adicionalmente, procedendo-se de modo análogo, os valores de IC_{50} frente à NS2B/NS3pro do DENV3 destes quatro inibidores também foram determinados a partir da curva dose-resposta construídas como descrito no item 3.6.1.2 e apresentadas na Figura 4.43. Estes valores estão apresentados na Tabela 4.39.

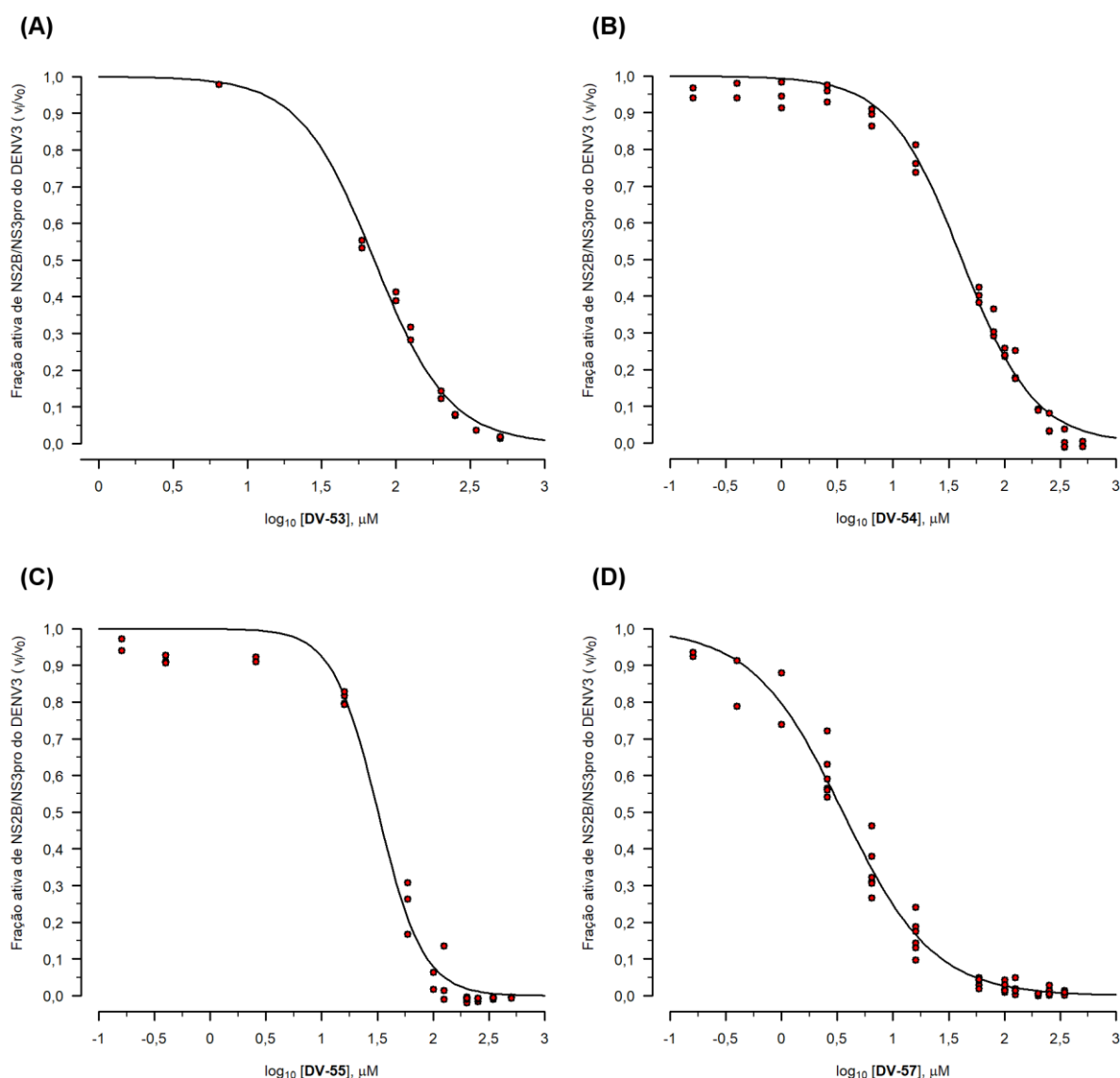
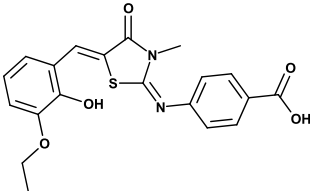
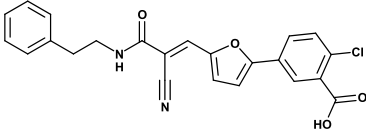
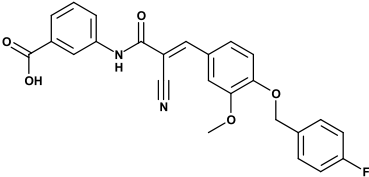
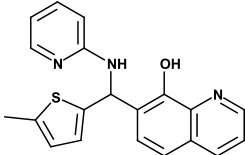


Figura 4.43. Curvas dose-resposta^(a) dos compostos DV-53, DV-54, DV-55 e DV-57. Fração ativa^(b) de NS2B/NS3pro do DENV3 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-53 (A), DV-54 (B), DV-55 (C) e DV-57 (D). Experimentos realizados, pelo menos, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: DV-53, DV-54, DV-55 e DV-57.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM.
- (d) Valores de fluorescência obtido em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV3 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (40 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{nm}$.

Tabela 4.39. Valores de IC_{50} (μM) frente à NS2B/NS3pro do DENV2 e do DENV3 e dos correspondentes erros padrão de ajuste^(a) obtidos para os compostos DV-53, DV-54, DV-55 e DV-57.

#Ref	Estrutura	IC_{50} (μM)	
		DENV2	DENV3
DV-53		$80,2 \pm 7,5$	$71,2 \pm 2,6$
DV-54		$46,2 \pm 2,1$	$41,3 \pm 2,0$
DV-55		$22,8 \pm 0,7$	$32,0 \pm 1,6$
DV-57		$4,7 \pm 0,3$	$3,5 \pm 0,2$

^(a) Calculado pelo pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) do programa R (R CORE TEAM, 2013).

Analisando a Tabela 4.39 podemos observar que os valores de IC_{50} dos compostos DV-53, DV-54, DV-55 e DV-57 determinados frente à NS2B/NS3pro do DENV2 foram próximos aos para a protease do DENV3. Estes valores coincidentes, dentro dos erros de ajuste, são considerados interessantes, pois se espera que um bom inibidor desta protease seja ativo contra todos os sorotipos do DENV. Considerando estes quatro inibidores selecionados pelo Modelo IV –DENV, a taxa de acerto deste modelo foi ~ 22%. Ainda, o composto DV-57 foi o inibidor mais ativo dentre todos ($IC_{50} = 4,7 \pm 0,3$ e $3,5 \pm 0,2$ para a protease do DENV2 e DENV3, respectivamente).

Por fim, procedendo-se como descrito no item 3.6.1.3, o inibidor **DV-57** teve seu tipo de mecanismo de inibição frente à NS2B/NS3pro do DENV2 determinado através dos gráficos Duplo-recíproco e Dixon (Figura 4.44 e Figura 4.45, respectivamente).

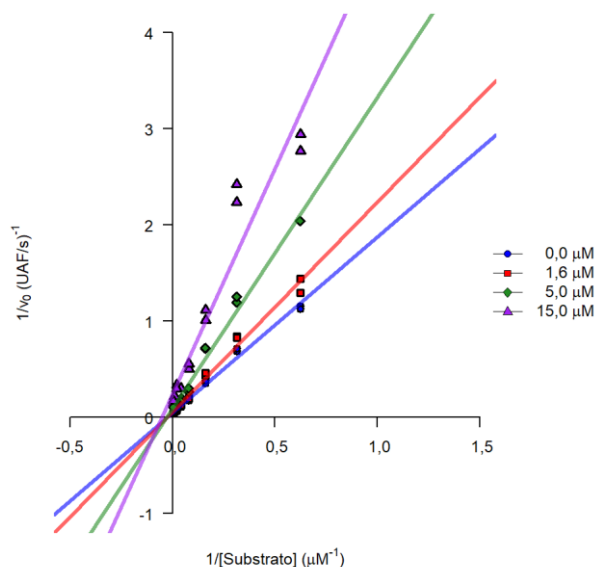


Figura 4.44. Gráfico Duplo-recíproco^(a) para o inibidor DV-57. Valores do inverso da velocidade inicial (v_0)^(b) em função dos valores do inverso da concentração^(c) do substrato PhAc-KRR-MCA (μM^{-1}) obtidos utilizando quatro diferentes concentrações do inibidor **DV-57** (0,0; 1,6, 5,0 e 15,0 μM). Experimentos realizados em duplicata.

- (a) Construída utilizando-se o programa *R* (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto obtidos utilizando diferentes concentrações^(c) do substrato e do inibidor **DV-57** (variando de 0,0 a 15,0 μM), determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s).
- (c) Os valores da concentração do substrato PhAc-KRR-MCA variaram de 1,6 a 200,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (variando de 1,6 a 200,00 μM) e **DV-57** (variando de 0 a 15 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{ nm}$.

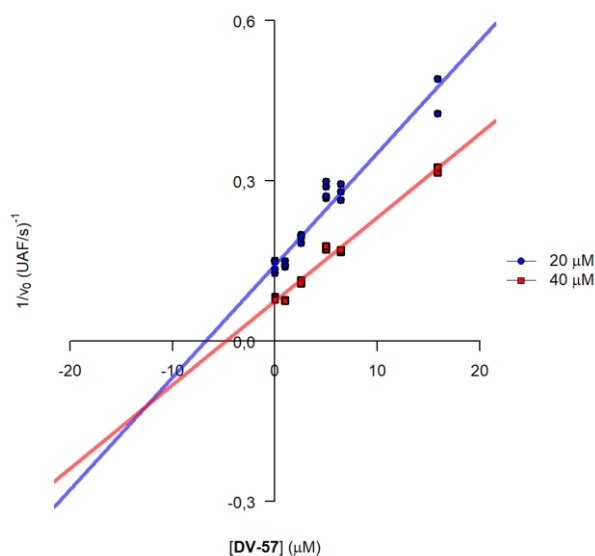


Figura 4.45. Gráfico de Dixon^(a) para o inibidor DV-57. Valores do inverso da velocidade inicial (v_0)^(b) em função em função dos valores da concentração^(c) do substrato PhAc-KRR-MCA (μM) obtidos utilizando duas diferentes concentrações do inibidor **DV-57** (20 e 40 μM). Experimentos realizados, pelo menos, em duplicata.

- (a) Construída utilizando-se o programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto obtidos utilizando diferentes concentrações do substrato^(c) e do inibidor **DV-57** (variando de 0 a 15 μM), determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s).
- (c) Os valores da concentração do substrato PhAc-KRR-MCA variaram de 0 a 15,87 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Profª Diederich), PhAc-KRR-MCA (variando de 0 a 15,87 μM) e **DV-57** (20 ou 40 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{nm}$.

Os dois gráficos mostram que o inibidor **DV-57** apresenta um mecanismo de inibição reversível do tipo misto, ou seja, não competitivo. Conhecendo-se o tipo de mecanismo de inibição de **DV-57** e procedendo-se como descrito no item 3.6.1.3, foi possível calcular seus valores de K_i e αK_i , sendo estes iguais a 10,4 e 6,1 μM , respectivamente. O fato do valor do αK_i ser menor do que o do K_i indica que a afinidade do inibidor pela enzima é maior quando esta está ligada ao substrato.

O inibidor **DV-57** foi selecionado usando o modelo farmacofórico MF-LIT06-3U1I (Figura 4.40 (F)) e seus valores de IC_{50} , K_i e αK_i estão na mesma ordem de grandeza que o do **LIT-06** ($\text{IC}_{50} = 1,3\ \mu\text{M}$, (LAI; SRIDHAR PRASAD; PADMANABHAN, 2013)) usado para construir MF-LIT06-3U1I. No entanto, o mecanismo de inibição misto determinado para

inibidor **DV-57** indica que este não se liga no sítio ativo da enzima, como proposto pelo modelo MF-LIT06-3U1I (Figura 4.40 (F)). Comparando a estrutura dos inibidores **DV-57** e **LIT-06** podemos observar que ambos apresentam o grupo 8-hidroxiquinolina ligado a um grupo amina através de um carbono quiral (Figura 4.46). Entretanto, o inibidor **DV-57** é consideravelmente menor que o **LIT-06**. Esta diferença de tamanho pode ser considerada esperada tendo em vista as características do modelo MF-LIT06-3U1I (Figura 4.46).

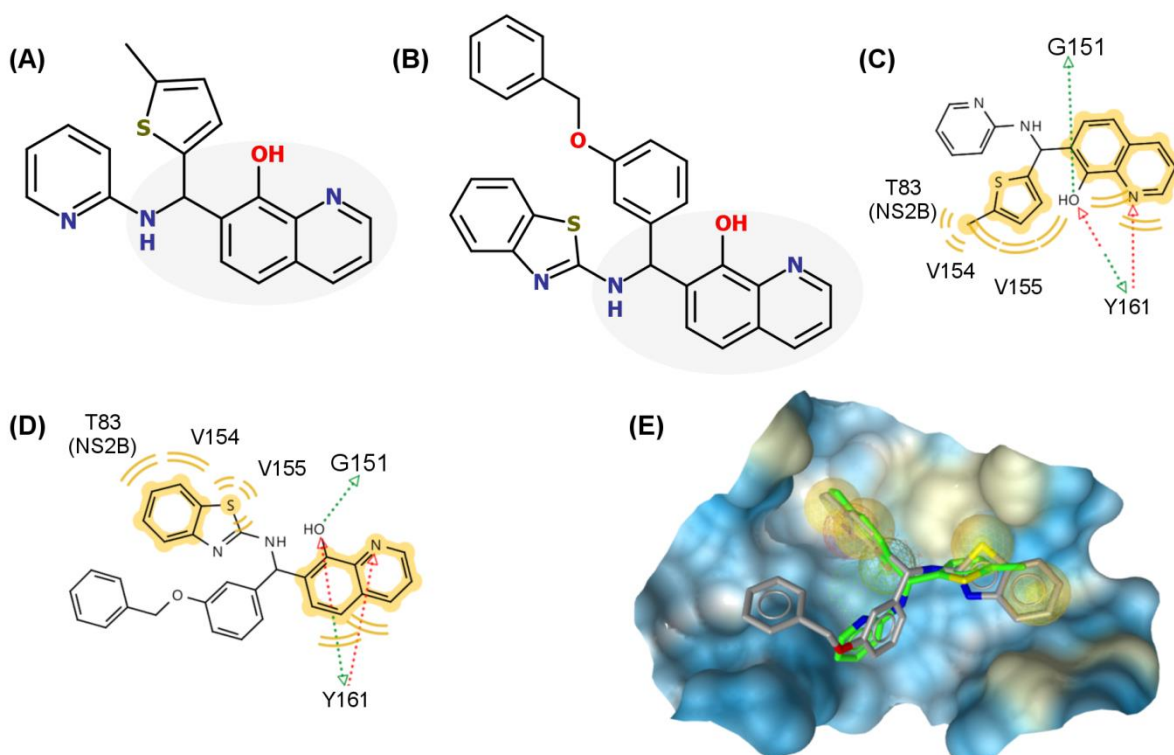


Figura 4.46. Comparação entre os inibidores **DV-57** ($IC_{50} = 4,7 \mu M$, selecionado pelo Modelo IV- DENV^(a)) e **LIT-06** ($IC_{50} = 1,3 \mu M$, selecionado da literatura (LAI; SRIDHAR PRASAD; PADMANABHAN, 2013)). Representação das estruturas 2D de **DV-57** (A) e de **LIT-06** (B), mostrando em cinza a subestrutura comum. Representação 2D da sobreposição de **DV-57** (C) e de **LIT-06** (D) ao Modelo IV – DENV. (E) Representação 3D da sobreposição dos inibidores **DV-57** (verde) e **LIT-06** (cinza) ao Modelo IV – DENV no sítio ativo da NS2B/NS3pro do DENV^(b). Em verde, vermelho e amarelo estão mostradas as interações doadoras e receptoras de ligação de hidrogênio e hidrofóbicas, respectivamente.

^(a) Gerados com o programa *LigandScout* (WOLBER; LANGER, 2005) a partir das interações identificadas entre o inibidor **LIT-06** e os resíduos do sítio ativo, utilizando os valores de corte padrão do programa.

^(b) A NS2B/NS3pro do DENV corresponde a estrutura 3U1I.

Adicionalmente, conforme apresentado e discutido na Introdução desta tese (item 1.5.5), na fase de otimização de um *hit* tem sido sugerido pela literatura considerar não apenas a potência dos compostos, mas também suas eficiências de ligação. Dentre as métricas já

apresentadas para descrever a eficiência de ligação, cita-se a *LE* (do inglês *ligand efficiency*). Os inibidores **DV-57** e **LIT-06** têm valores de *LE* iguais a 0,29 e 0,23 kcal/mol, respectivamente. Embora a comparação entre valores de *LE* seja mais indicada para atividades medidas nas mesmas condições experimentais (HOPKINS et al., 2014), o fato do valor de *LE* do **DV-57** ser um pouco menor do que o do **LIT-06** sugere que o primeiro seria um *hit* melhor. Além disso, **DV-57** apresenta um valor próximo ao mínimo sugerido pela literatura ($LE > 0,30$ kcal/mol (MIGNANI et al., 2016)). Por fim, considerando os sítios identificados no item 4.3.1.2 (Figura 4.20 e Figura 4.21), um possível sítio de ligação de **DV-57** seria o localizado abaixo do sítio ativo (marcado nas figuras como sítio número **1**), já sugerido como um potencial sítio alostérico. Esta hipótese será, futuramente, analisada através de procedimentos de ancoramento neste sítio.

Cabe esclarecer que, em virtude do curto tempo de duração da visita aos laboratórios da Prof^a. Diederich, não houve tempo para determinar o mecanismo de inibição para os demais inibidores selecionados por este modelo (**DV-53**, **DV-54** e **DV-55**).

Avaliação da estabilidade dos inibidores nas condições do experimento de inibição frente à da NS2B/NS3pro do DENV2/DENV3

Como planejamos com relação à estabilidade, todos os 18 compostos adquiridos (**DV-42** a **DV-59**) se mostraram estáveis nas condições do experimento de inibição (tampão pH 9,0). Ainda, a identificação e/ou caracterização dos quatros compostos com atividade inibitória (**DV-53**, **DV-54**, **DV-55** e **DV-57**) está em andamento.

Por fim, em resumo, o Modelo IV – DENV foi capaz de selecionar compostos com boa estabilidade nos experimentos de inibição. Adicionalmente, quatro compostos (22,22%) selecionados pelo Modelo IV – DENV apresentaram atividade inibitória significativa frente à NS2B/NS3pro do DENV2 e à do DENV3. Diferentemente do esperado, o mecanismo de inibição determinado para o inibidor mais ativo (**DV-57**) não segue o proposto pelo modelo

farmacofórico usado para selecioná-lo. Contudo, o modelo Modelo IV – DENV se mostrou capaz de encontrar compostos ativos e com boa estabilidade. Assim, as alterações incluídas neste modelo conseguiram melhorar, consideravelmente, os resultados da busca virtual.

4.3.7. Série de análogos ao composto mais ativo dentre os selecionados pelo quarto modelo de busca virtual (Modelo IV – DENV)

O objetivo da busca virtual é identificar o maior número compostos, com a maior diversidade estrutural possível, que sejam ativos frente a determinado alvo de interesse (ou seja, *hits*). No entanto, embora seja desejado, nem sempre os *hits* identificados apresentam a potência esperada (SCIOR et al., 2012). Sendo bastante comum que estes apresentem atividades na ordem de poucos μM (SCIOR et al., 2012). Estes *hits* podem, no entanto, ser, considerados como estruturas de partida que devem ser modificadas e otimizadas, com o objetivo de se obter um composto mais potente (atividade na ordem de pM a nM). Sendo assim, **DV-57** foi considerado como uma estrutura de partida. Considerando que a síntese de compostos está fora dos objetivos desta tese, propusemos tentativamente melhorar a atividade de **DV-57** através de abordagens de baseadas na estrutura de ligante (*LDBB*) partindo de compostos análogos estruturalmente similares a DV-57 e, ainda disponíveis para compra.

4.3.7.1. Construção da série de análogos

Estes estudos foram propostos antes de conhecer o mecanismo de inibição de **DV-57** (inibidor do tipo misto, vide pg. 308) e realizados como descrito no item 3.5.4.13 (Materiais e Métodos). A busca por similaridade 2D, usando **DV-57** como referência, resultou em 240 compostos com similaridade estrutural de pelo menos 80 % em relação a ele. Em seguida, estes compostos foram submetidos aos mesmos filtros usados para selecionar **DV-57** (modelo farmacofórico MF-LIT06-3U1I seguido de inspeção visual, vide itens 4.3.6.1 e 4.3.6.2), resultando na seleção de 123 (51,25%) compostos. Destes, 46 (37,40%) estavam disponíveis

para serem comprados e 13 deles (28,26%) foram adquiridos considerando-se: (i) a possibilidade de fazer outras interações com o sítio ativo da protease do DENV, além das descritas pelo modelo MF-LIT06-3U1I e (ii) o tempo de entrega, pois esta etapa foi realizada durante a visita aos laboratórios da Prof^a. Diederich (de 30/01/2017 a 24/02/2017). A Tabela 4.40 mostra a representação das estruturas dos treze compostos (denominados **DV-60** a **DV-72**) análogos estruturais do **DV-57** que foram adquiridos e também submetidos a testes de inibição enzimática frente à NS2B/NS3pro do DENV.

Tabela 4.40. Representação da estrutura 2D dos 13 compostos (**DV-60** a **DV-72**) da série de análogos ao **DV-57** que foram adquiridos e submetidos ao teste de inibição enzimática frente a esta protease. Estas estão acompanhadas dos seus correspondentes códigos ZINC, valores de peso molecular e empresa fornecedora.

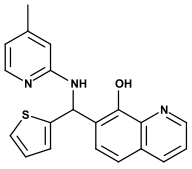
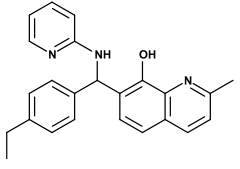
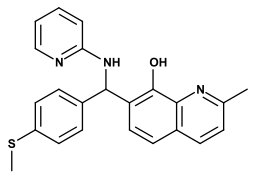
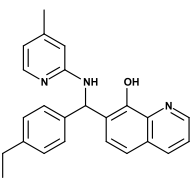
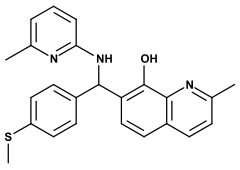
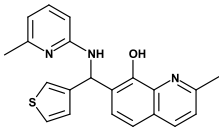
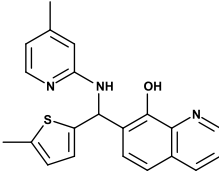
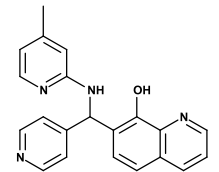
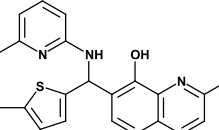
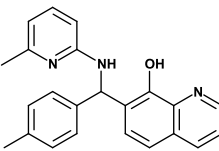
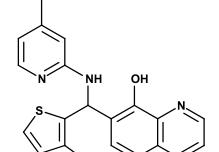
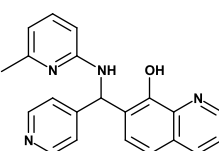
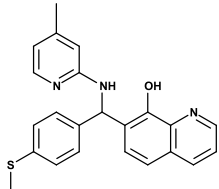
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-60		ZINC13062711	347,44	Enamine
DV-61		ZINC02073889	369,47	Vitas M Lab
DV-62		ZINC02470437	387,50	Vitas M Lab
DV-63		ZINC02874296	369,47	Vitas M Lab
DV-64		ZINC04755363	401,53	Vitas M Lab

Tabela. 4.40. (Continua).

#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
DV-65		ZINC05900688	361,46	Vitas M Lab
DV-66		ZINC03656379	361,46	Vitas M Lab
DV-67		ZINC00086532	342,40	Vitas M Lab
DV-68		ZINC01132808	375,49	Vitas M Lab
DV-69		ZINC18173640	355,44	Vitas M Lab
DV-70		ZINC18173644	361,46	Vitas M Lab
DV-71		ZINC19799769	342,40	Vitas M Lab
DV-72		ZINC01171678	387,50	Vitas M Lab

4.3.7.2. Validação experimental

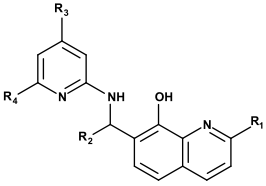
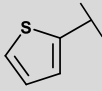
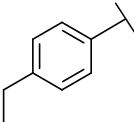
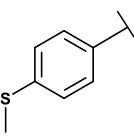
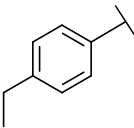
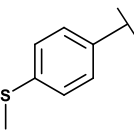
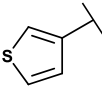
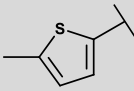
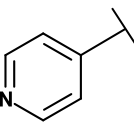
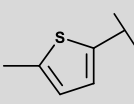
De modo análogo ao realizado para o Modelo IV - DENV, a validação experimental foi feita através de ensaios enzimáticos. Assim, os 13 compostos, análogos ao **DV-57**, adquiridos foram submetidos a ensaios de inibição enzimática frente à NS2B/NS3pro do DENV2. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Wibke Diederich (Philipps-Universität de Marburg, Marburg, Alemanha) e estão descritos no item 3.6.1 (Materiais e Métodos).

Ensaio de inibição enzimática frente à NS2B/NS3pro do DENV

Procedendo-se como descrito 3.6.1.1 (Materiais e Métodos), inicialmente, foram determinados os valores de porcentagem de inibição da NS2B/NS3pro do DENV2 de cada composto (em 100 μ M), sendo estes apresentados na Tabela 4.41. Ainda, de modo análogo ao proposto para o Modelo IV – DENV somente compostos com valores de atividade inibitória maiores, ou próximos a 50%, foram considerados como tendo atividade inibitória significativa.

Assim, dos 13 compostos adquiridos, quatro (30,77%) mostraram atividade inibitória significativa frente à NS2B/NS3pro do DENV2 na concentração utilizada (100 μ M), sendo eles: compostos **DV-60** (ZINC13062711), **DV-66** (ZINC03656379), **DV-68** (ZINC01132808) e **DV-70** (ZINC18173644). Os valores de porcentagem de inibição dos compostos **DV-60**, **DV-66**, **DV-68** e **DV-70** foram iguais a 79 ± 4 ; 80 ± 1 , 60 ± 6 e 80 ± 1 , respectivamente. Por outro lado, nove deles não mostraram atividade inibitória significativa, sendo eles: compostos **DV-61** (ZINC02073889), **DV-62** (ZINC02470437), **DV-63** (ZINC02874296), **DV-64** (ZINC04755363), **DV-65** (ZINC05900688), **DV-67** (ZINC00086532), **DV-69** (ZINC18173640), **DV-71** (ZINC19799769) e **DV-72** (ZINC01171678).

Tabela 4.41. Valores de porcentagem de inibição (em 100 μM)^(a) e dos correspondentes desvios padrão^(b) de cada um dos compostos (DV-60 a DV-72) da série de análogos ao DV-57.

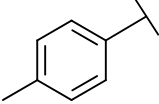

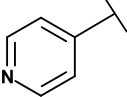
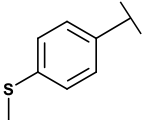
<div> <div>Estrutura geral</div>  </div>					
#Ref	R ₁	R ₂	R ₃	R ₄	Inibição da NS2B/NS3pro do DENV2 (% , em 100 μM)
DV-60 ^(c)	-H		-H	-CH ₃	79 \pm 4
DV-61	-CH ₃		-H	-H	4 \pm 1
DV-62	-CH ₃		-H	-H	17 \pm 6
DV-63	-H		-H	-CH ₃	13 \pm 2
DV-64	-CH ₃		-CH ₃	-H	16 \pm 12
DV-65	-CH ₃		-CH ₃	-H	42 \pm 8
DV-66 ^(c)	-H		-H	-CH ₃	80 \pm 1
DV-67	-H		-H	-CH ₃	35 \pm 8
DV-68 ^(c)	-CH ₃		-CH ₃	-H	60 \pm 6

^(a) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 μM) e cada um dos compostos (100 μM).

^(b) Considerando pelo menos três experimentos independentes.

^(c) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Tabela 4.41.(Continua)

#Ref	R ₁	R ₂	R ₃	R ₄	Inibição da NS2B/NS3pro do DENV2 (% , em 100 µM)
DV-69	-H		-CH ₃	-H	17 ± 6
DV-70 ^(c)	-H		-H	-CH ₃	76 ± 1
DV-71	-H		-CH ₃	-H	31 ± 1
DV-72	-H		-H	-CH ₃	27 ± 1

^(d) Determinados em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich), PhAc-KRR-MCA (20 µM) e cada um dos compostos (100 µM).

^(e) Considerando pelo menos três experimentos independentes.

^(f) Compostos considerados significativamente ativos, destacados em cinza na tabela.

Diferentemente dos demais modelos, todos os compostos, exceto o **DV-69**, tiveram seus valores de IC₅₀ frente à NS2B/NS3pro do DENV2 determinados a partir de suas curvas dose-resposta, construídas como descrito no item 3.6.1.2 (Materiais e Métodos) e apresentadas nas Figura 4.47, Figura 4.48 e Figura 4.49. Estes valores estão apresentados na Tabela 4.42.

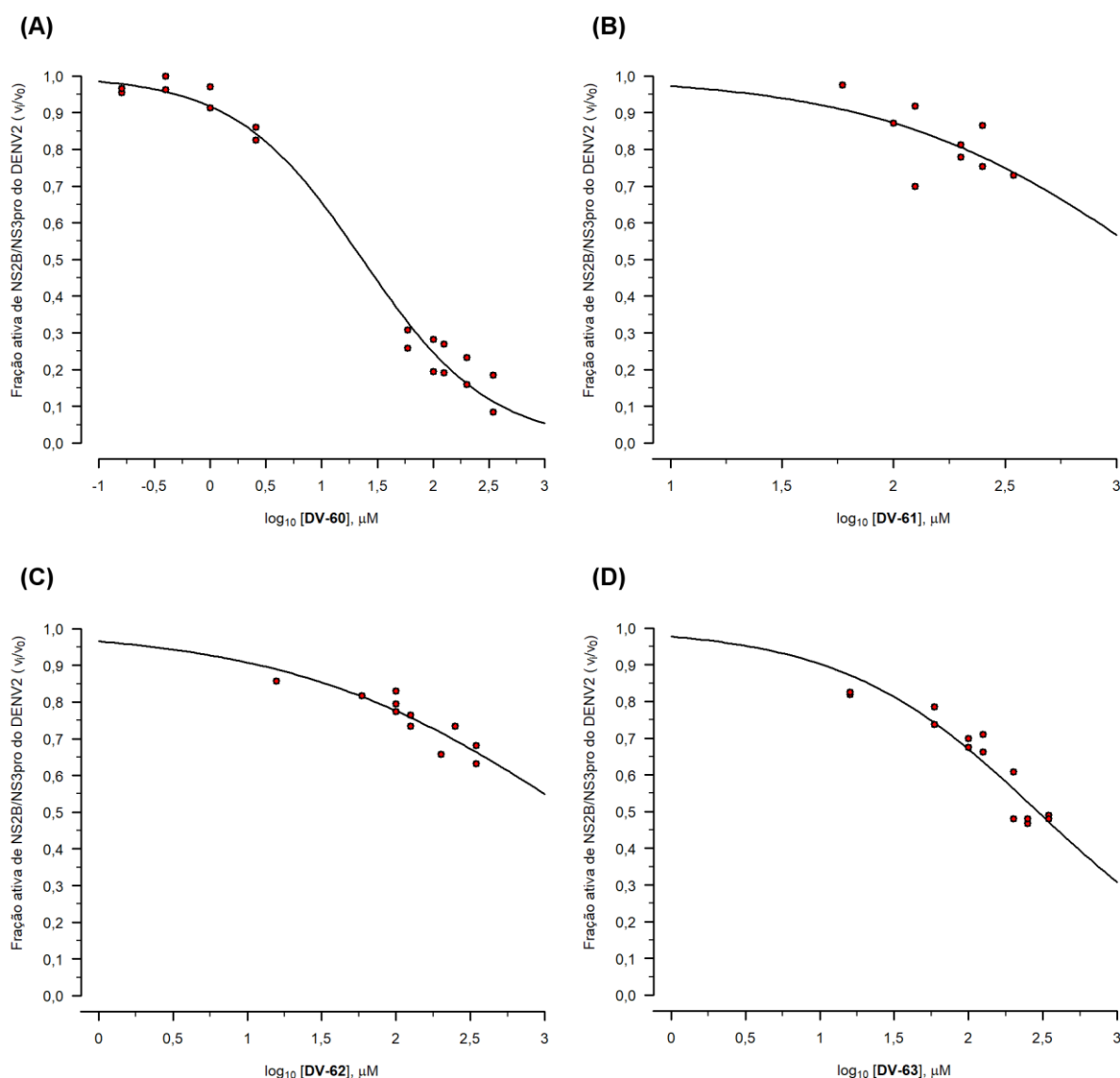


Figura 4.47. Curvas dose-resposta^(a) dos compostos DV-60, DV-61, DV-62 e DV-63. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-60 (A), DV-61 (B), DV-62 (C) e DV-63 (D). Experimentos realizados, dentro do possível, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (INDEX; DEVELOPMENT; TEAM, 2009).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: DV-61, DV-62, DV-63 e DV-64.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (20 μM), à $37 \pm 0,1^\circ\text{C}$ e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{nm}$.

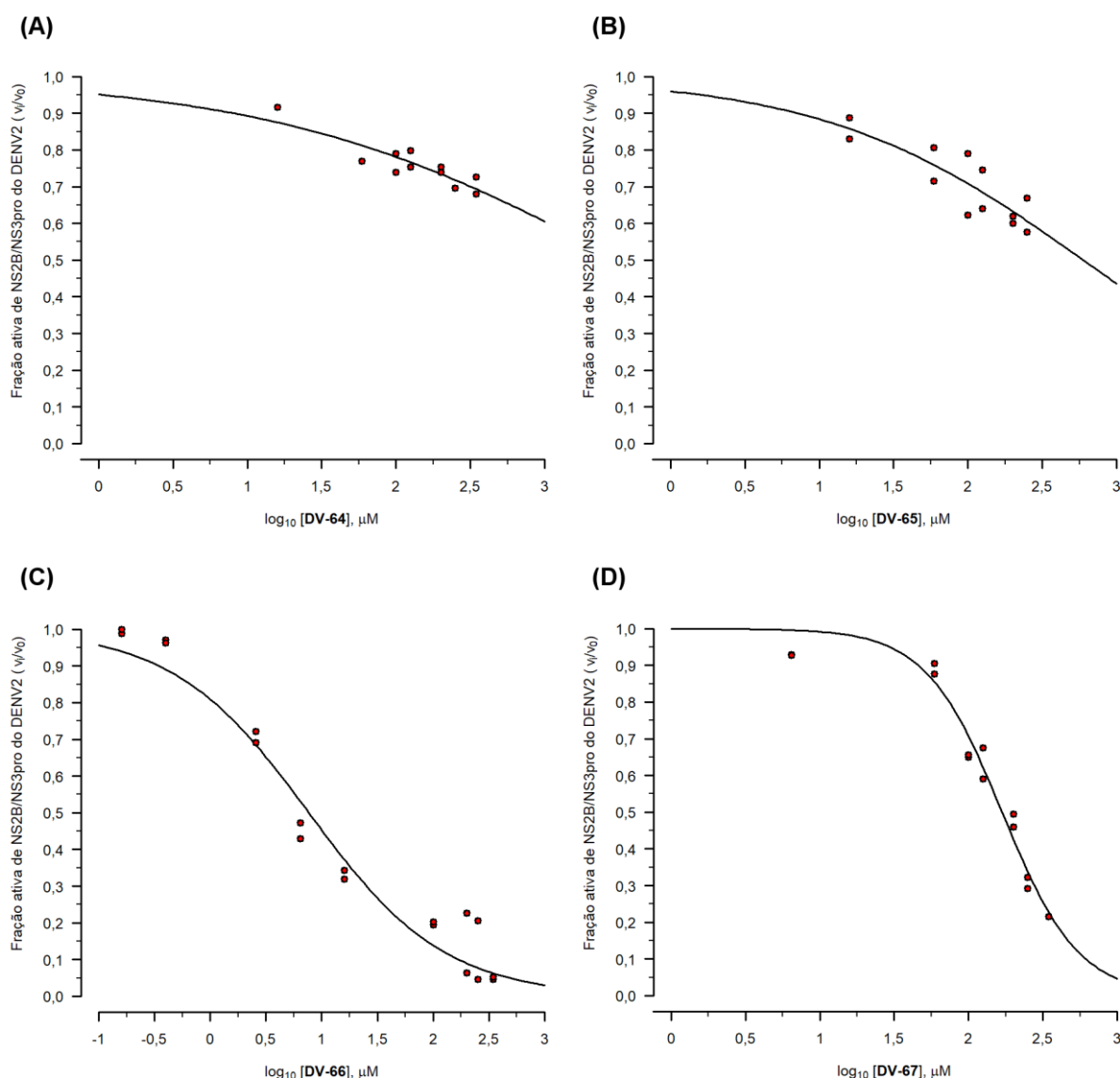


Figura 4.48. Curvas dose-resposta^(a) dos compostos DV-64, DV-65, DV-66 e DV-67. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-64 (A), DV-65 (B), DV-66 (C) e DV-67 (D). Experimentos realizados, dentro do possível, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: **DV-64, DV-65, DV-66 e DV-67**.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (20 μM), à 37 \pm 0,1°C e utilizando-se $\lambda_{\text{em}} = 380\text{nm}$ e $\lambda_{\text{ex}} = 460\text{ nm}$.

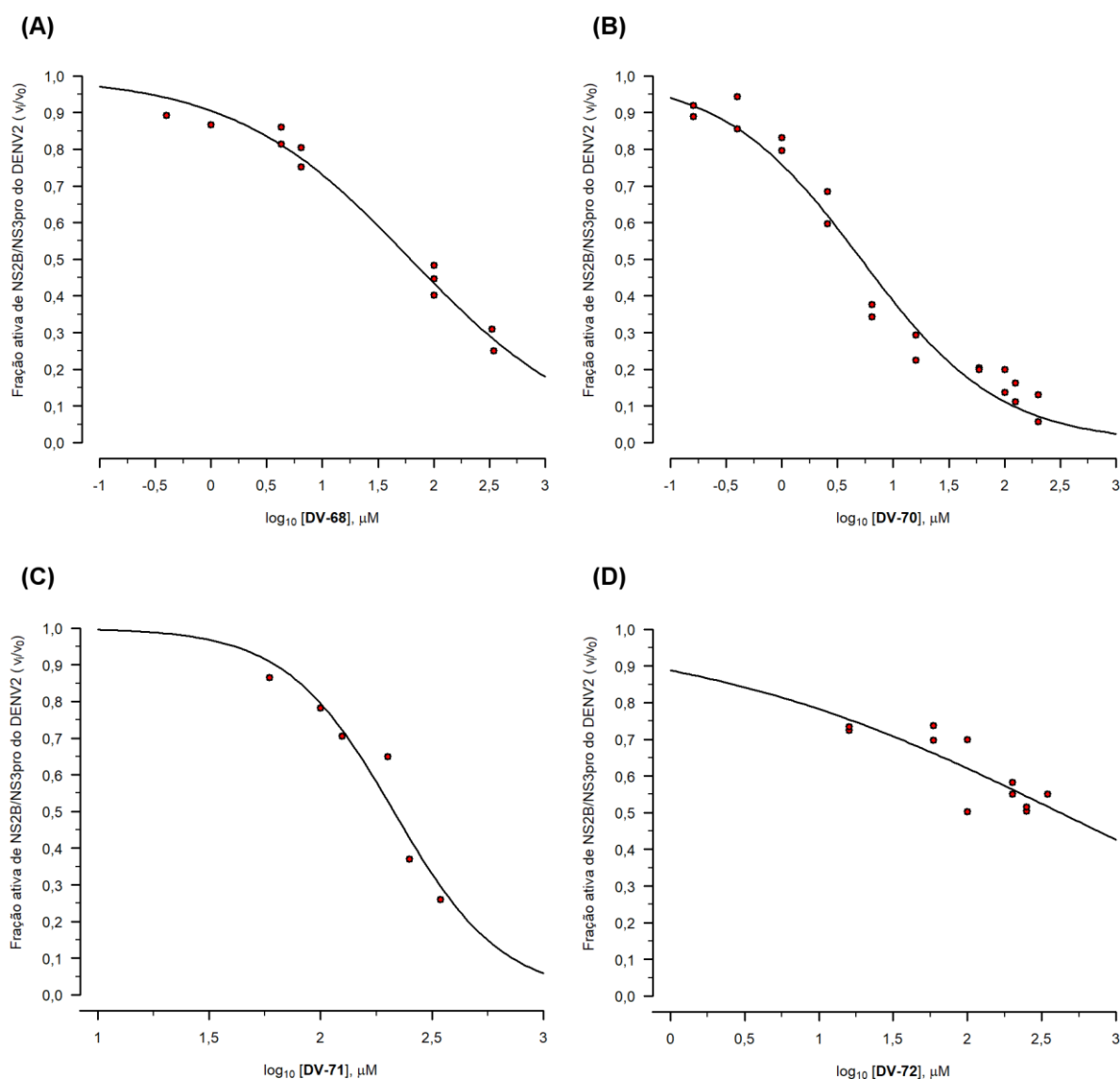


Figura 4.49. Curvas dose-resposta^(a) dos compostos DV-68, DV-70, DV-71 e DV-72. Fração ativa^(b) de NS2B/NS3pro do DENV2 em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: DV-68 (A), DV-70 (B), DV-71 (C) e DV-72 (D). Experimentos realizados, dentro do possível, em duplicata.

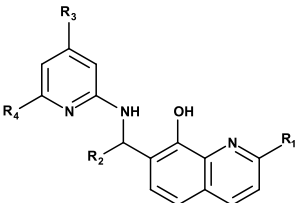
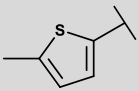
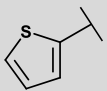
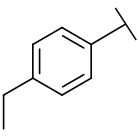
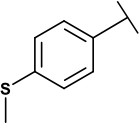
- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v₀), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: DV-68, DV-70, DV-71 e DV-72.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 500,00 μM.
- (d) Valores de fluorescência obtidos em solução tampão pH=9,0 [Tris Base (50 mM) e CHAPS (1 mM)] contendo NS2B/NS3pro do DENV2 (50 nM, disponível no grupo da Prof^a Diederich) e PhAc-KRR-MCA (20 μM), à 37 ± 0,1°C e utilizando-se λ_{em} = 380nm e λ_{ex} = 460 nm.

Nas curvas dose-resposta dos compostos DV-60, DV-66, DV-67, DV-68 e DV-70

(Figura 4.47, Figura 4.48 e Figura 4.49) se observa inibição próxima a 90 % para a maior concentração utilizada de cada um destes compostos. Sendo assim, seus valores de IC₅₀ foram

obtidos com baixa incerteza acompanhados dos correspondentes valores pequenos de erro de ajuste. Por outro lado, nas curvas dose-resposta dos demais compostos desta série (Figura 4.47, Figura 4.48 e Figura 4.49) não se observa inibição próxima a 90 % para a maior concentração utilizada de cada um destes compostos. Desta forma, seus valores de IC₅₀ foram obtidos com altas incertezas, acompanhados dos correspondentes valores grandes de erro de ajuste. Para estes compostos não foi possível aumentar o valor da concentração usado devido a limitações quanto à solubilidade nas condições do experimento (solução tampão pH 9,0). Assim, os valores de IC₅₀ determinados para estes últimos foram considerados somente como estimativas.

Tabela 4.42. Valores estimados de IC₅₀ (μM) frente à NS2B/NS3pro do DENV2 e dos correspondentes erros padrão do ajuste^(a) obtidos para 13 compostos da série de análogos ao inibidor DV-57.

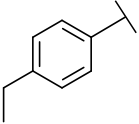
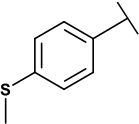
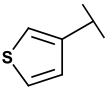
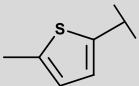
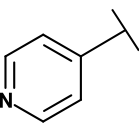
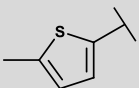
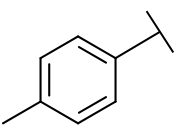
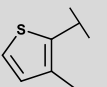
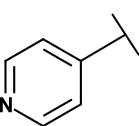
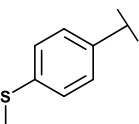
<div> <div>Estrutura geral</div>  </div>					
#Ref	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)
DV-57 ^(c)	-H		-H	-H	4,7 ± 0,3
DV-60 ^(c)	-H		-H	-CH ₃	23,2 ± 2,5
DV-61	-CH ₃		-H	-H	> 1000 ^(b)
DV-62	-CH ₃		-H	-H	> 1000 ^(b)

^(a) Calculado pelo pacote “drc”(RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) do programa R (R CORE TEAM, 2013).

^(b) Curvas dose resposta com poucos pontos devido a limitações experimentais.

^(c) Compostos considerados significativamente ativos.

Tabela 4.42. (Continua)

#Ref	R ₁	R ₂	R ₃	R ₄	IC ₅₀ (μM)
DV-63	-H		-H	-CH ₃	292,5 ± 36,4 ^(b)
DV-64	-CH ₃		-CH ₃	-H	> 1000 ^(b)
DV-65	-CH ₃		-CH ₃	-H	> 500 ^(b)
DV-66 ^(c)	-H		-H	-CH ₃	7,6 ± 1,2
DV-67	-H		-H	-CH ₃	167,9 ± 7,7
DV-68 ^(c)	-CH ₃		-CH ₃	-H	62,1 ± 6,3
DV-69	-H		-CH ₃	-H	ND ^(d)
DV-70 ^(c)	-H		-H	-CH ₃	5,1 ± 0,7
DV-71	-H		-CH ₃	-H	212,9 ± 18,2 ^(b)
DV-72	-H		-H	-CH ₃	420,0 ± 164,2 ^(b)

^(a) Calculado pelo pacote “drc”(RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) do programa R (R CORE TEAM, 2013).

^(b) Curvas dose resposta com poucos pontos devido a limitações experimentais.

^(c) Compostos considerados significativamente ativos.

^(d) Não determinado (ND) por falta de tempo.

Ainda, desta série de análogos, os mais ativos frente à NS2B/NS3pro do DENV são os compostos **DV-66** e **DV-70** com valores de IC₅₀ iguais a 7,6 ± 1,2 e 5,1 ± 0,7 μM, respectivamente. Infelizmente, diferentemente do planejado, nenhum composto desta série se

mostrou mais ativo que **DV-57**. Como já relatado, a construção desta série foi proposta, concomitantemente, com a determinação do mecanismo de inibição do **DV-57** e, ainda, não se havia observado seu mecanismo de inibição misto. O fato dos 13 compostos desta série serem estruturalmente mais similares ao inibidor **DV-57** do que ao inibidor **LIT-06** (descrito como competitivo) sugere que estes também possam atuar por um mecanismo de inibição misto. Portanto, embora os compostos desta série pudessem fazer outras interações, além das preditas pelo modelo MF-LIT06-3U1I, o fato destes não se ligarem no sítio ativo impediu que as interações consideradas pudessem resultar em melhores inibidores do que **DV-57**.

4.3.7.3. Análise preliminar da Relação Estrutura Atividade

Adicionalmente, fizemos uma análise preliminar de Relação Estrutura Atividade Biológica (SAR, do inglês *Structure Activity Relationship*) desta série de compostos sem gerar qualquer modelo matemático. Nesta análise, foi observado que a adição de uma metila no anel 8-hidroxiquinolina parece diminuir a atividade inibitória destes compostos. A adição de uma metila na posição 4 do anel piridínico parece ser tolerada, embora não se observe um aumento significativo na atividade inibitória (ex. **DV-57** quando comparado a **DV-66**). Ainda, esta adição da metila na posição 4 é preferível a adição na posição 6 (ex. **DV-67** quando comparado ao **DV-71**). Além disso, a presença do grupo tiofeno é bastante importante para a atividade inibitória. Uma análise mais completa desta relação estrutura atividade inibitória será, futuramente, realizada através, por exemplo, de modelos Quantitativos de Relação Estrutura Atividade (QSAR).

4.4. Busca virtual de inibidores da protease do vírus da Febre Aftosa

Nesta tese, foram construídos e utilizados dois diferentes modelos de buscas de inibidores da protease líder do vírus da Febre Aftosa (denominados Modelo I – FMDV e Modelo II – FMDV). De modo análogo a busca virtual da NS2B/NS3pro do DENV, estes

modelos de busca virtual foram feitos aplicando-se, em sequência, diferentes filtros de seleção. Inicialmente, será apresentado o item referente ao reconhecimento do alvo. Em seguida, serão apresentados os resultados de cada um dos modelos, incluindo os compostos selecionados bem como a etapa de validação experimental. A ordem apresentada dos modelos corresponde à ordem cronológica em que foram construídos.

4.4.1. Reconhecimento da estrutura e características do sítio ativo da protease líder do vírus da Febre Aftosa para escolha da estrutura 3D

Conforme apresentado na busca virtual da NS2B/NS3pro do DENV, para o sucesso de estratégias baseadas na estrutura do receptor, além da disponibilidade da estrutura 3D do receptor, outros aspectos relacionados à interação receptor-ligante precisam ser considerados.

Nesta tese, o segundo alvo a ser apresentado e discutido será a protease líder do vírus da Febre Aftosa (L^{pro} do vírus FA). Como apresentado na Introdução (item 1.11.2), existem duas isoformas da L^{pro} , denominadas Lab^{pro} e Lb^{pro} . Estas isoformas tem a mesma atividade, no entanto, a Lb^{pro} sugerida como a isoforma biologicamente ativa da L^{pro} (NOGUEIRA SANTOS et al., 2012; SANTOS et al., 2009). Além disso, uma forma menor desta última isoforma também é encontrada, sendo denominada sLb^{pro} . A única diferença entre as formas Lb^{pro} e sLb^{pro} é o número de resíduos em suas extensões C-terminais (CTE), que diferem por 6 ou 7 resíduos. Baseado no exposto, nesta tese, as duas formas da Lb^{pro} (Lb^{pro} e sLb^{pro}) foram usadas, indiscriminadamente, como alvo para a busca virtual. A estrutura tridimensional da Lb^{pro} foi descrita pela primeira vez por Guarné e colaboradores em 1996 (GUARNÉ et al., 1996). No entanto, a elucidação completa desta estrutura por cristalografia de Raios-X, só foi possível em 1998 (1QOL (GUARNE et al., 1998)). Atualmente, existem pelo menos cinco estruturas da Lb^{pro} do vírus da FA depositadas no *PDB*, sendo três obtidas por cristalografia de Raio-X (1QOL, 1QMY e 4QBB (GUARNE et al., 1998; GUARNÉ et al., 2000; STEINBERGER et al., 2014)) e duas por RMN (2JQG e 2JQF (CENCIC et al., 2007)). A

Tabela 4.43 apresenta os códigos PDB de cada uma delas, os correspondentes métodos de obtenção, a resolução (quando obtida por cristalografia de Raios-X), a forma da Lb^{pro} usada, a informação da presença, ou não, de ligantes e de mutação na cisteína catalítica (*Cys51*). A representação da estrutura do único ligante complexado com a Lb^{pro} do vírus da FA, considerando as estruturas 3D disponíveis no *PDB*, está apresentada na Figura 4.50.

Tabela 4.43. Códigos PDB das quinze estruturas da proteína Lb^{pro} do vírus da FA disponíveis no PDB, forma da Lb^{pro}, método de obtenção, valores de resolução (Å), ligante e mutações da *Cys51* catalítica.

Código PDB	Forma da Lb ^{pro}	Método de Obtenção	Resolução (Å)	Ligante	<i>Cys51</i> catalítica
1QOL	Lb ^{pro}	Raios-X	3,0	Não	C51A
1QMY	sLb ^{pro}	Raios-X	1,9	Não	C51A
2JQG	sLb ^{pr}	RMN	-	Não	C51A
2JQF	Lb ^{pro}	RMN	-	Não	C51A
4QBB	sLb ^{pro}	Raios-X	1,6	E64-R-P-NH ₂ ^(a)	Covalentemente ligada ao E64-R-P-NH ₂

^(a) N-(trans-epoxisuccinil)-L-leucina-4-guanidinobutilamida-arginina-prolina-amina, estrutura apresentada na Figura 4.50.

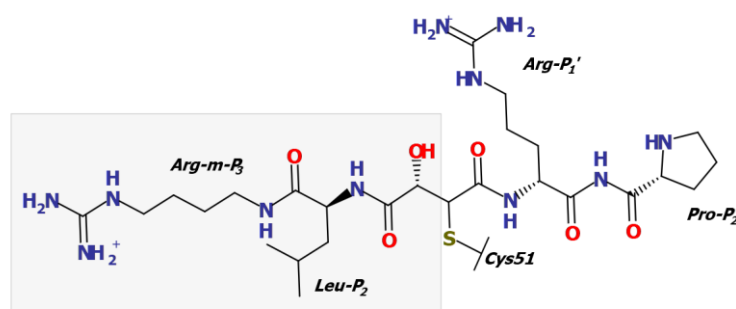


Figura 4.50. Representação da estrutura do ligante presente na estrutura tridimensional da Lb^{pro} do vírus da FA disponível no PDB. Estrutura 2D do E64-R-P-NH₂, inibidor peptídeo mimético covalentemente ligado à *Cys51* ($K_i = 30$ nM) como encontrado no PDB 4BBQ (resolução 1,6 Å, (STEINBERGER et al., 2014)). O retângulo cinza, mostra a região do inibidor que corresponde ao E64 (trans-Epoxisuccinil-L-leucilamido(4-guanidino)butano) após a reação entre o grupo epóxido e o enxofre da *Cys51* da Lb^{pro}.

Diferentemente da NS2B/NS3pro do DENV, todas as estruturas 3D apresentadas na Tabela 4.43 correspondem a sequência da Lb^{pro}/sLb^{pro} obtida para o mesmo sorotipo do vírus

da FA (sorotipo O1k). Ainda, como apresentado na Introdução (item 1.11.2), a tríade catalítica da Lb^{pro} é composta pelos resíduos *Cys51*, *His148* e *Asp163* e um dos seus substratos naturais é a porção C-terminal da própria proteína. Desta forma, as estabilidades das Lb^{pro}/sLb^{pro}, na ausência de ligantes, durante os experimentos de cristalografia de raios-X e de RMN foram garantidas substituindo-se a cisteína catalítica por uma alanina (mutante C51A). Esta mutação inativa a Lb^{pro} impedindo sua auto proteólise (GUARNÉ et al., 1996).

Nas estruturas obtidas utilizando a Lb^{pro} (2JQF e 1QOL), se observa a formação de dímeros simétricos e tetrâmeros assimétricos, respectivamente. A formação deste dímeros/tetrâmeros se deve às interações dos últimos resíduos de aminoácidos da região CTE de uma determinada proteína com parte do sítio ativo de outra proteína adjacente (Figura 4.51(A)). A literatura (SANTOS et al., 2009) sugere que a formação do dímero deva ocorrer para soluções com concentrações de Lb^{pro} superiores a 2 µM. Embora o dímero da Lb^{pro} seja estável em solução, este precisa ser desfeito para que os outros substrato naturais desta enzima possam se ligar de forma eficiente ao sítio ativo (CENCIC et al., 2007). Diferentemente do observado para Lb^{pro}, as estruturas determinadas usando a sLb^{pro} (1QMY e 2JQG) são monoméricas (Figura 4.51(B)). Ainda, suas regiões CTE não apresentam conformações definidas devido a alta flexibilidade destas regiões (CENCIC et al., 2007; GUARNÉ et al., 2000). De modo análogo, a estrutura da sLb^{pro} complexada com o inibidor E64-R-P-NH₂, covalentemente ligado a *Cys51*, se encontra na forma monomérica (Figura 4.51(C)).

Com base no exposto, as formas monoméricas obtidas usando a sLb^{pro} (2JQG, 1QMY e 4QBB) foram escolhidas para serem utilizadas na busca virtual de inibidores desta protease. A estrutura 3D da sLb^{pro} complexada com o inibidor E64-R-P-NH₂ (4QBB, resolução 1,60 Å (STEINBERGER et al., 2014)) só foi disponibilizada no *PDB* em 11/2014. Assim, para o primeiro modelo de busca virtual (Modelo I - FA) não havia sido descrita, qualquer estrutura da 3D da sLb^{pro} com ligantes, justificando o uso das estruturas sem ligantes (2JQG e 1QMY).

Dentre estas, a estrutura 2JQG da sLb^{pro} foi escolhida como referência, pois esta foi obtida mais recentemente e, é considerada de melhor qualidade (CENCIC et al., 2007). Já para o segundo modelo de busca virtual (Modelo II – FA), a estrutura com ligantes foi usada.

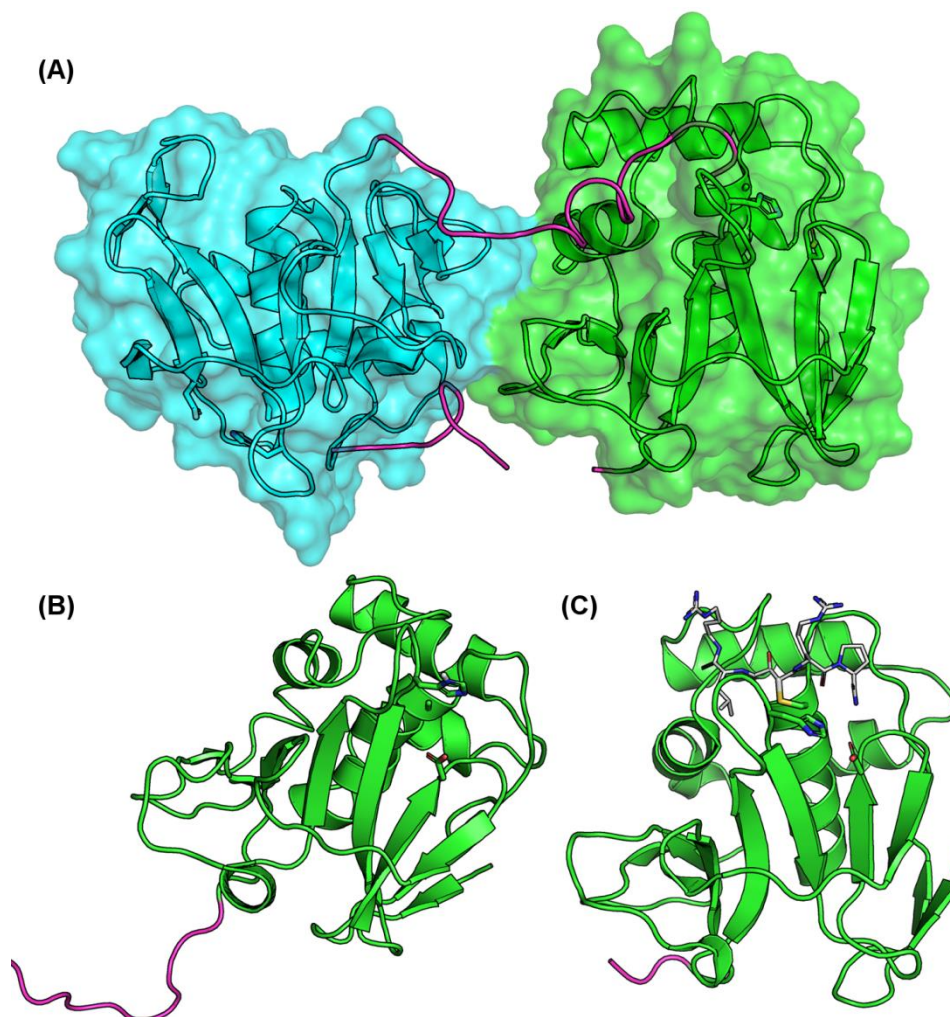


Figura 4.51. Estruturas 3D da Lb^{pro} e da sLb^{pro} do vírus da FA obtidas na forma de dímero e de monômero, respectivamente. (A) Lb^{pro} obtida como dímero devido à interação da porção CTE (magenta) de uma determinada proteína (cadeia S, ciano) com o sítio ativo de outra proteína adjacente (cadeia R, verde) (PDB 2JQF (CENCIC et al., 2007)). (B) sLb^{pro} (verde) obtida como monômero e com sua região CTE (magenta) sem conformação definida (PDB 2JQG (CENCIC et al., 2007)). (C) sLb^{pro} (verde) complexada com o inibidor covalente E64-R-P-NH₂ (cinza) obtida como monômero com apenas três resíduos de sua porção CTE (magenta) (cadeia A do PDB 4QBB, resolução 1,6 Å (STEINBERGER et al., 2014)). As cadeias principais da Lb^{pro}/sLb^{pro} e as cadeias laterais dos resíduos correspondentes à tríade catalítica (“Ala51”/Cys51, His148 e Asp163) estão mostradas em “cartoon” e em “sticks”, respectivamente. Figura preparada com o programa PyMOL (SCHRÖDINGER, 2013).

4.4.1.1. Reconhecimento da estrutura da proteína por inspeção visual

De modo análogo ao apresentado para protease do DENV (item 4.3.1.1), o reconhecimento da estrutura da Lb^{pro}/sLb^{pro} foi realizado, inicialmente, por inspeção visual.

Esta etapa foi feita usando, separadamente, as estruturas 1QOL, 2JQF e 4QBB. No entanto, considerando a similaridade estrutural entre elas, apenas a da estrutura 4BQQ (resolução 1,60 Å (STEINBERGER et al., 2014)) será apresentada. A estrutura 4QBB foi cristalizada com três moléculas na unidade assimétrica (correspondendo as cadeias A, B e C). Nesta estrutura, não se observa a densidade eletrônica completa nos dois resíduos de arginina do inibidor E64-R-P-NH₂, devido às suas altas flexibilidades. Steinberger e colaboradores escolheram a cadeia B como a de referência por apresentar uma melhor resolução do inibidor, no entanto, esta se deve a contatos de empacotamento cristalino (STEINBERGER et al., 2014). Ainda, o valor do RMSD, considerando todos os átomos, entre as três cadeias é menor que 1,0 Å. Nesta tese, escolheu-se utilizar cadeia A que embora não apresente a melhor resolução do inibidor, esta sofre menos contatos cristalinos.

Nesta inspeção visual podemos observar que o inibidor E64-R-P-NH₂ ocupa as cavidades S_3 a S_2' e que há uma ligação covalente entre o inibidor e a Cys51 catalítica (Figura 4.50 e Figura 4.52(A)). A tríade catalítica desta estrutura é formada pelos resíduos de aminoácidos Cys51, His148 e Asp163, localizados entre as cavidades do sítio ativo S_1 e S_2' (Figura 4.52(B)). O átomo de nitrogênio terminal do resíduo Pro(P_2') do E64-R-P-NH₂ faz uma ligação de hidrogênio com a cadeia lateral Asp164 ocupando a cavidade S_2' (Asn46, His148, Asp163 e Asp164). A cavidade S_1' é composta por três resíduos ácidos (Asp49, Glu96 e Glu147) e pela His148 (Figura 4.52(B)). Na estrutura 4QBB, esta cavidade é ocupada pela Arg(P_1') do E64-R-P-NH₂. Embora a posição do grupo guanidino deste resíduo não tenha sido determinada, parece razoável propor que este grupo seria capaz de formar interações iônicas com os resíduos Asp49, Glu96 e Glu147 (Figura 4.52(A)). A cavidade S_1 da Lb^{pro} do vírus do FA é estreita diferindo de outras proteases da família da papaína, as quais apresentam cavidades S_1 grandes com pouca influência sobre a especificidade dos substratos (GUARNE et al., 1998). Dentre os resíduos que formam a cavidade S_1 se destacam: Asn46, Asp49, Cys51,

Glu96, *Gly97* e *Glu147*. No inibidor E64-R-P-NH₂ a cadeia lateral do resíduo *P₁* corresponde a um grupo epóxido que é atacado pela *Cys51*. Além da ligação covalente com a *Cys51*, o oxigênio carbonílico deste resíduo faz ligações de hidrogênio com *Asn46* (cadeia lateral) e *Cys51* (cadeia principal). Na estrutura 1QOL, a posição *P₁* é ocupada pela *Lys201* da CTE (GUARNE, 1998) que forma interações iônicas com os resíduos ácidos. A presença de resíduos ácidos tanto em *S₁'* bem como *S₁* também difere a Lb^{pro} das demais papaínas celulares (STEINBERGER et al., 2014). A cadeia principal do resíduo Leu(*P₂*) do E64-R-P-NH₂ faz duas ligações de hidrogênio com as cadeias principais dos resíduos *Trp52* e *Gly98* da cavidade *S₂* (Figura 4.52(A)). Ainda, sua cadeia lateral faz interações hidrofóbicas com os resíduos *Leu143* e *Ala149* presentes nesta cavidade. Na estrutura 1QOL, interações hidrofóbicas com resíduos da cavidade *S₂* (*Trp52*, *Gly97-Pro100*, *Leu143*, *Glu147-Ala149* e *Leu178*) também são observadas para a Leu200(*P₂*) da porção CTE. Como já descrito, não foi possível determinar a posição do grupo guanidino da Arg-m(*P₃*) do E64-R-P-NH₂. Ainda sim, parece razoável propor que este grupo seria capaz de formar interações iônicas com os resíduos *Glu93* e *Glu147* desta cavidade. Na estrutura 1QOL, a *Lys199*(*P₃*) da CTE faz uma ligação de hidrogênio com a cadeia principal do *Glu93*. Ainda, nesta estrutura, pelas posições dos resíduos Arg198(*P₄*), Gln197(*P₅*) e Val196(*P₆*) podemos determinar alguns resíduos que compõem as cavidades *S₄*, *S₅* e *S₆*, sendo estes: *Gly98*, *Pro99*, *Leu143* e *Gln146* em *S₄*; *Pro99* em *S₅* e *Pro99*, *Ala101*, *Val127* e *Leu178* em *S₆* (Figura 4.52(B)). Por fim, vale observar que de modo análogo ao Bz-nKRR-H (DENV), as cadeias laterais de alguns resíduos do inibidor E64-R-P-NH₂ estão expostas ao solvente (vide Figura 4.52(B)), sendo estes os resíduos Pro(*P₂'*), Arg(*P₁'*) e Arg-m(*P₃*).

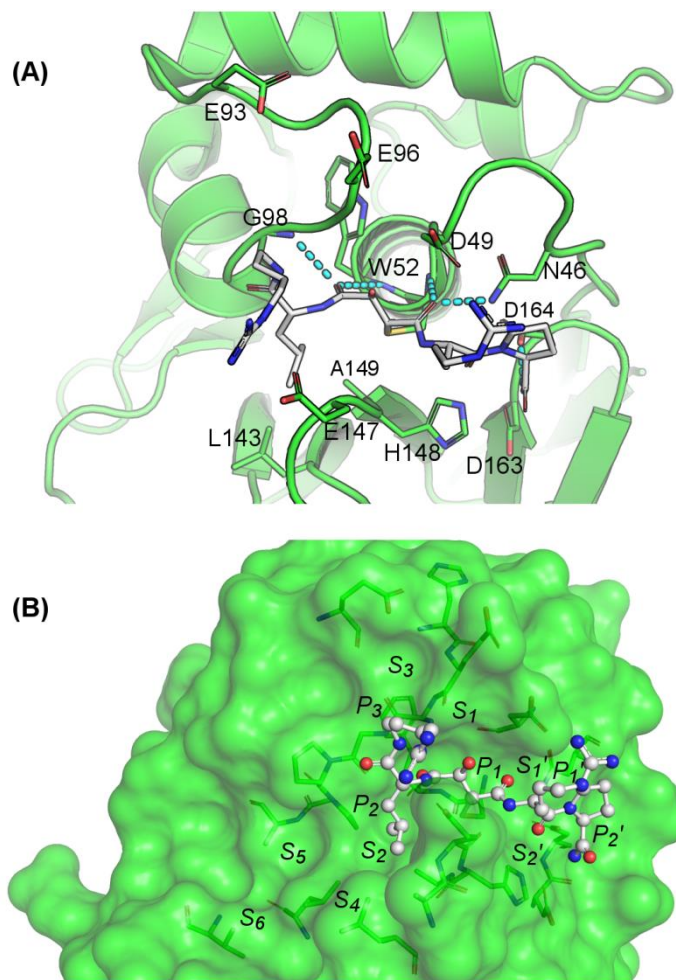


Figura 4.52. Representação das interações proteína – E64-R-P-NH₂ e das cavidades S_{2'} a S₆ da Lb^{pro} do vírus da FA^(a). (A) Interações entre os resíduos do sítio ativo e o inibidor E64-R-P-NH₂ covalentemente ligado a Cys51. O inibidor E64-R-P-NH₂ (cinza) e os resíduos do sítio ativo (verdes) estão mostrados como *sticks*. A cadeia principal da Lb^{pro} está mostrada como *cartoon*. (B) Cavidades S_{2'} – S₆ da Lb^{pro} do vírus da FA. As cavidades S_{2'} a S₃ foram definidas com base nos resíduos do inibidor E64-R-P-NH₂ (P_{2'} a P₃), enquanto que as cavidades S₄ a S₆ foram definidas considerando a posição dos resíduos da Arg198(P₄), Gln197(P₅) e Val196(P₆) da região CTE como observado na estrutura 1QOL (GUARNE, 1998). A proteína, os resíduos das cavidades e o inibidor estão mostrados como superfície, *sticks* e *ball and sticks*, respectivamente. Os átomos de carbono da Lb^{pro} e do inibidor estão mostrados em verde e cinza, respectivamente. Figura preparada com o programa PyMOL (SCHRÖDINGER, 2013).

^(a) A estrutura Lb^{pro} do vírus da FA corresponde ao PDB 4QBB (resolução 1,6 Å (STEINBERGER et al., 2014)).

4.4.1.2. Reconhecimento da estrutura da proteína auxiliado por computador

De modo análogo a NS2B/NS3pro do DENV (item 4.3.1.2), numa segunda etapa, a análise da superfície e da natureza das interações possíveis na Lb^{pro}/sLb^{pro} do vírus da FA (PDB 4QBB) foi feita através dos potenciais eletrostáticos e das densidades doadora e/ou aceptora de ligação de hidrogênio. Estas três propriedades da superfície estão apresentadas na Figura 4.53. O potencial eletrostático calculado para superfície da sLb^{pro} do vírus da FA

variou de -626 até -57 kcal/mol elétron (regiões vermelhas e roxas, respectivamente). Conforme descrito na literatura (GUARNE, 1998), podemos observar o caráter polar das cavidades S_1 , S_1' e S_2' do sítio ativo da Lb^{pro} do vírus da FA. Ainda, podemos observar o caráter mais apolar das cavidades S_2 , S_4 , S_5 e S_6 quando comparada com as demais. Diferentemente, do observado para protease do DENV, não existem regiões com potencial eletrostático positivo na superfície da Lb^{pro} do vírus da FA. Adicionalmente, as cavidades S_1 , S_1' e S_2' tem um potencial eletrostático bastante negativo sendo, portanto, favoráveis às interações com ligantes contendo grupos carregados positivamente. Esta preferência por grupos carregados positivamente está de acordo com estudo da especificidade da Lb^{pro} frente a diferentes substratos peptídicos (SANTOS et al., 2009), no qual se observou uma preferência por substratos contendo resíduos básicos na posição P_1 , em especial por lisinas. Neste estudo ainda se observou uma preferência por substratos contendo resíduos alifáticos na porção P_2 com preferência por resíduos alifáticos não aromáticos, o que concorda com o caráter apolar observado para cavidade S_2 .

Em relação à densidade doadora e aceptora de ligação de hidrogênio, se observa uma preferência por grupos doadores de ligação de hidrogênio nas cavidades S_1' e S_2' . Já na cavidade S_1 se observa uma preferência por grupos aceptores de ligação de hidrogênio. Na estrutura 4QBB, esta região é ocupada pela cadeia principal do resíduo P_1 , contendo um oxigênio carbonílico que é um doador de ligação de hidrogênio (Figura 4.52 (A)). Por fim, diferentemente da NS2B/NS3pro, na Lb^{pro} a cavidade oxânion – responsável por estabilizar o intermediário tetraédrico carregado negativamente que é formado durante a catálise – é formada pelos resíduos *Cys51* e *Asn46*. Assim, o caráter eletrostático um pouco menos negativo e a densidade de doadores de ligação de hidrogênio próximo aos resíduos *Cys51* e *Asn46* (entre S_1 e S_1') estão de acordo com o mecanismo de catálise da Lb^{pro}.

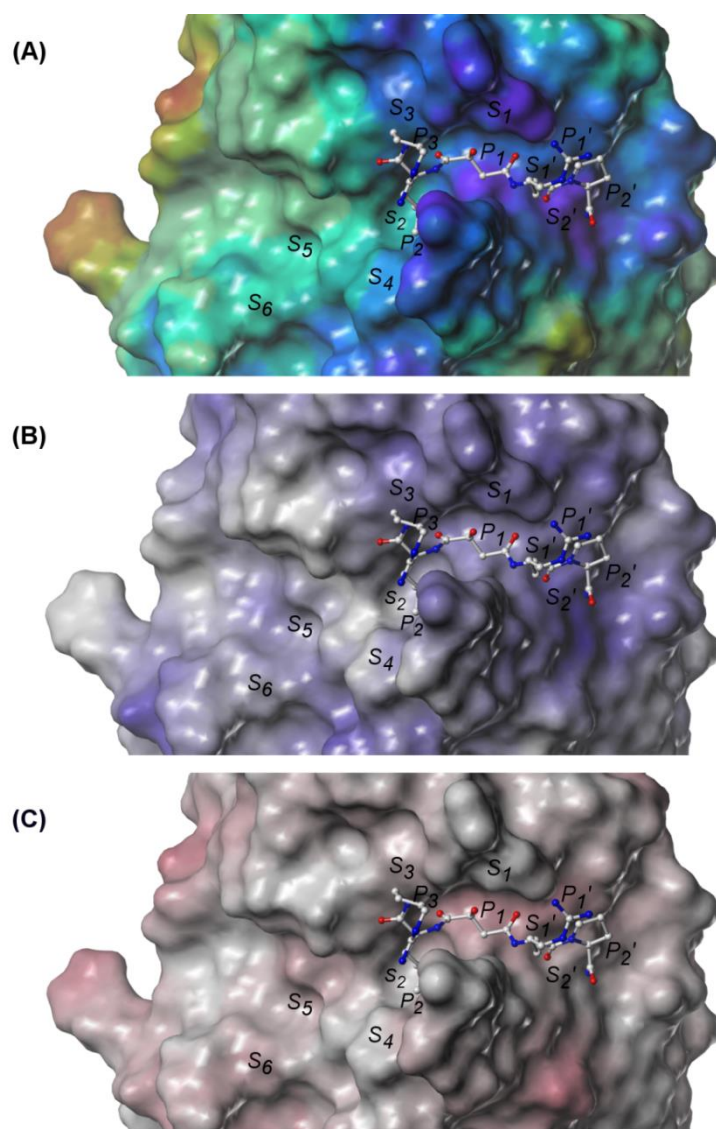


Figura 4.53. Representações da superfície^(a) da Lb^{pro} do vírus da FA^(b) colorida de acordo com algumas propriedades^(c): (A) potencial eletrostático variando de -57 até -626 kcal/mol elétron (vermelho e roxo, respectivamente). (B) densidade HBA variando de 0.000 até 0.087 Å⁻² (cinza e azul, respectivamente). (C) densidade de HBD variando de 0.000 até 0.128 Å⁻² (cinza e red, respectivamente). As cavidades S₂ a S₃ foram definidas com base nos resíduos do inibidor E64-R-P-NH₂ (P₂' a P₃), enquanto que as cavidades S₄ a S₆ foram definidas considerando a posição dos resíduos da Arg198(P₄), Gln197(P₅) e Val196(P₆) da região CTE como observado na estrutura 1QOL (GUARNE, 1998). A proteína e o inibidor estão mostrados como superfície e *ball and sticks*, respectivamente. Figura preparada com o programa SYBYL-X (SYBYL, 2013).

(a) Superfície acessível ao solvente gerada utilizando o algoritmo de Connolly (CONNOLLY, 1983) do programa SYBYL-X (SYBYL, 2013).

(b) A estrutura Lb^{pro} do vírus da FA corresponde ao PDB 4QBB (resolução 1,6 Å (STEINBERGER et al., 2014))

(c) Propriedades calculadas usando a ferramenta MOLCAD do SYBYL-X (SYBYL, 2013) após adicionar todos os átomos de hidrogênio e calcular as cargas parciais tipo Gasteiger-Hueckel (PURCELL; SINGER, 1967).

De modo análogo a protease do DENV, a profundidade das cavidades do sítio da Lb^{pro} do vírus da FA foi calculada. A coloração das cavidades do sítio ativo da Lb^{pro} variou de azul escuro a um amarelo claro, indicando um sítio raso com porções expostas ao solvente (Figura

4.54). Ainda, a região entre as cavidades S_1 e S_2 foi a que apresentou a coloração mais amarelada, dentre elas. Estas cavidades são descritas como menos expostas ao solvente (GUARNE, 1998). Como já apresentado, a profundidade do sítio ativo é uma das características utilizadas para medir sua *druggability*. De modo que sítios ativos rasos podem ser um indicativo de sítios não *druggables*.

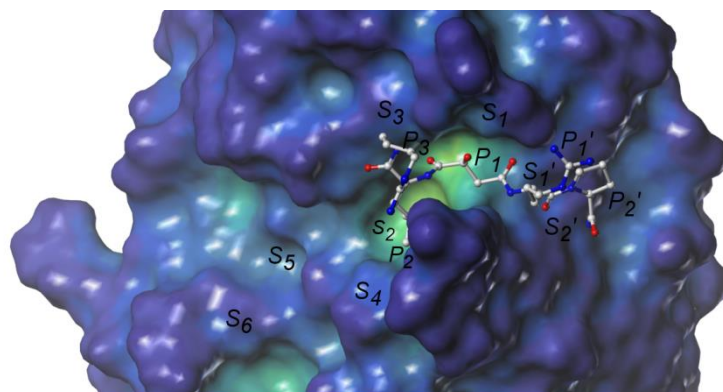


Figura 4.54. Representação da profundidade^(a) da superfície^(b) da Lb^{pro} do vírus da FA^(c). Superfície da Lb^{pro} colorida de acordo com a profundidade de cada região variando de -0,3 Å (azul) a 10 Å (vermelho). As cavidades S_2' a S_3 foram definidas com base nos resíduos do inibidor E64-R-P-NH₂ (P_2' a P_3) e as cavidades S_4 a S_6 foram definidas considerando a posição dos resíduos da Arg198(P_4), Gln197(P_5) e Val196(P_6) da região CTE como observado na estrutura 1QOL (GUARNE, 1998). A proteína e o inibidor estão mostrados como superfície e *ball and sticks*, respectivamente. Figura preparada com o programa SYBYL-X (SYBYL, 2013).

(a) A profundidade foi calculada usando a ferramenta MOLCAD do programa SYBYL-X (SYBYL, 2013).

(b) Superfície acessível ao solvente gerada como descrito na Figura 4.53.

(c) A estrutura Lb^{pro} do vírus da FA corresponde ao PDB 4QBB (resolução 1,6 Å (STEINBERGER et al., 2014))

A análise da *druggability* deste sítio ativo da Lb^{pro} foi feita de modo análogo ao descrito para a protease do DENV. Sete sítios foram identificados e classificados de acordo com seu valor de *drugScore* (denominados **1** a **7**). Estes sítios estão mostrados na Figura 4.55 e seus respectivos valores de *drugScore*, de profundidade e da razão superfície/volume estão apresentados na Tabela 4.44. Dos sete sítios de ligação identificados, um sítio (**5**) se encontra na região do sítio ativo da Lb^{pro} do vírus da FA. Este sítio de ligação está localizado próximo as cavidades S_1' , S_1 e S_2 (Figura 4.52 (B)). De modo análogo ao observado para a protease do DENV, este sítio teve um baixo valor de *drugScore* (0,33) que também é justificado pelo fato dele ser raso. Contudo, o sítio **5** cobre a área da cadeia principal dos resíduos P_1' , P_1 e P_2 e a cadeia lateral do resíduo P_2 , sendo, portanto, toda a área não exposta ao solvente deste

inibidor. Os sítios de ligação com maior valor de *drugScore* (**1** e **2**) se encontram na face oposta do sítio ativo da Lb^{pro} do vírus da FA (Figura 4.55).

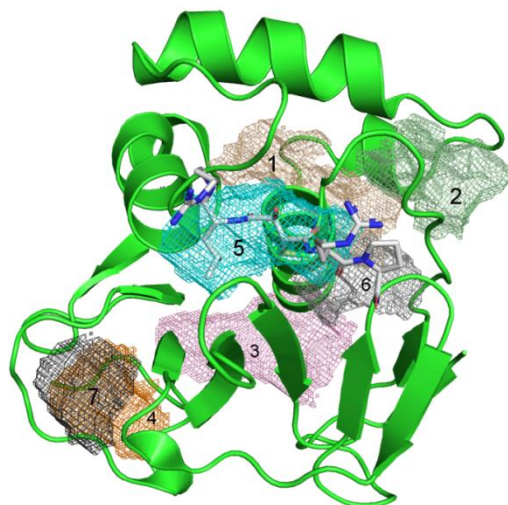


Figura 4.55. Sítios de ligação identificados^(a) para a Lb^{pro} do vírus da FA^(b). Os sítios de ligação estão mostrados como superfície de malha (bege, verde, rosa, laranja, ciano, cinza e preto, respectivamente, sítios de ligação **1** a **7**). A proteína (verde) e o inibidor (cinza) estão mostrados como *cartoon* e *sticks*, respectivamente. Figura preparada com *PyMOL* (SCHRÖDINGER, 2013).

^(a) Identificados utilizando a ferramenta *DoGSiteScorer* (VOLKAMER et al., 2012) e classificados de acordo com seus valores de *drugScore* (variando de 0,73 a 0,20).

^(b) A estrutura Lb^{pro} do vírus da FA corresponde ao PDB 4QBB (resolução 1,6 Å (STEINBERGER et al., 2014)).

Tabela 4.44. Valores de *DrugScore*, da profundidade e da razão superfície/volume de cada um dos sete sítios de ligação identificados para a Lb^{pro} do vírus da FA (estrutura 4QBB).

Sítio de ligação	<i>DrugScore</i>	Profundidade	Superfície/volume
1	0,73	18,65	1,23
2	0,57	14,87	1,71
3	0,56	13,45	1,60
4	0,38	9,62	1,38
5	0,33	7,95	1,54
6	0,29	10,22	0,78
7	0,20	7,12	2,28

Embora o sítio ativo da Lb^{pro} do vírus da FA também seja superficial, ainda foi possível identificar um sítio de ligação próximo às cavidades S_1' , S_1 e S_2 . Ainda, estas

cavidades apresentam um alto potencial eletrostático e uma grande densidade de doadores/aceptores de ligação de hidrogênio. Sendo assim, as cavidades S_I' , S_I e S_2 do sítio ativo da Lb^{pro} do vírus da FA foram escolhidas como cavidades alvo para os procedimentos de busca virtual.

4.4.1.3. Escolha da conformação do receptor e considerações sobre sua flexibilidade conformacional

Conforme apresentado na Introdução, além de se escolher uma conformação relevante do receptor, torna-se importante também considerar, dentro do possível, aspectos relacionados à sua flexibilidade conformacional. No entanto, diferentemente da protease do DENV, não há, dentro do nosso conhecimento, relatos sobre a flexibilidade conformacional envolvendo resíduos do sítio ativo da Lb^{pro} do vírus da FA. Sendo assim, não foram incluídos na busca virtual aspectos relacionados à sua flexibilidade. Ainda, conforme apresentado no item 4.4.1, as formas monoméricas obtidas usando a sLb^{pro} (2JQG, 1QMY e 4QBB) foram escolhidas para serem utilizadas na busca virtual de inibidores desta protease. No período referente à criação do primeiro modelo de busca virtual (Modelo I - FA), não havia sido descrita, qualquer estrutura da 3D da sLb^{pro} com ligantes. Assim, a 2JQG da sLb^{pro} sem ligantes foi escolhida como referência, pelos motivos já expostos no item 4.4.1. Entretanto, para o segundo modelo de busca virtual (Modelo II – FA), a estrutura 3D da sLb^{pro} complexada com o inibidor E64-R-P-NH₂ (4QBB, resolução 1,60 Å (STEINBERGER et al., 2014)) já estava disponível sendo escolhida como referência.

4.4.2. Modelos de busca virtual: Modelo I – FA

Conforme apresentado acima, durante a criação deste primeiro modelo de busca virtual (denominado Modelo I – FA), não estavam disponíveis informações sobre as interações entre inibidores e os resíduos do sítio desta protease. Assim, visando encontrar

inibidores bem como obter mais informações sobre as interações no sítio ativo da sLb^{pro}, uma série “*in house*” de 52 compostos foi utilizada para a busca virtual. Estes foram previamente selecionados em nosso grupo como potenciais inibidores covalentes da cruzaina (MALVEZZI; DE REZENDE; AMARAL, 2009), uma cisteína protease da família das papaínas que apresenta 53% de identidade em relação à Lb^{pro}. Embora as cavidades do sítio ativo da Lb^{pro} tenham características diferentes das demais proteases da família das papaínas (vide item 4.4.1), o fato desta série estar disponível em nosso laboratório e dos compostos terem grupos eletrofílicos reativos frente à cisteína proteases nos fez pensar que seria possível encontrar dentre estes algum inibidor da Lb^{pro}. Por estes motivos, todos os compostos desta série foram ancorados no sítio ativo da Lb^{pro} e subsequentemente analisados por inspeção visual como descrito item 3.5.5.5 (Materiais e Métodos). Os compostos selecionados por estes filtros foram submetidos a ensaios de inibição enzimática. A Figura 4.56 mostra uma representação esquemática dos filtros de seleção aplicados, acompanhado do valor da porcentagem de redução do número de compostos nesta etapa.

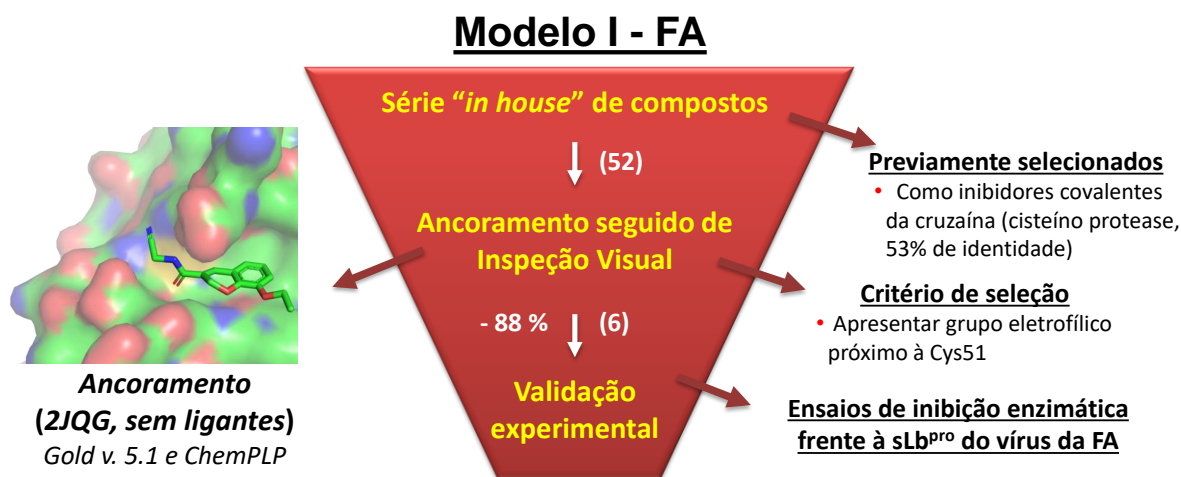


Figura 4.56. Representação esquemática dos filtros de seleção aplicados à série “*in house*” de compostos usada no primeiro modelo de busca virtual (**Modelo I – FA**). Entre parênteses são indicados os números de compostos selecionados nesta etapa e ao lado a correspondente porcentagem de redução.

4.4.2.1. Ancoramento seguido por Inspeção Visual

O único filtro usado nesta busca virtual foi o ancoramento acompanhado pela inspeção visual. Procedendo-se como descrito item 3.5.5.5 (Materiais e Métodos), os 52 compostos da série foram ancorados no sítio ativo da sLb^{pro} do vírus da FA usando a estrutura 2JQG. Todas as poses de todos os compostos foram, a seguir, analisadas por inspeção visual. Sabendo-se que estes compostos foram previamente selecionados como potenciais inibidores covalentes da cruzaina, apenas compostos que mostram suas poses de ancoramento com grupo eletrofílico próximos à Cys51 foram selecionados como potenciais inibidores covalentes da Lb^{pro} do vírus da FA. De acordo com este critério, seis compostos (~ 11% da série inicial) foram selecionados como potenciais inibidores covalentes da Lb^{pro} do vírus da FA. A Tabela 4.45 mostra a representação das estruturas dos seis compostos (denominados **FA-01** a **FA-06**) selecionados pelo Modelo I – FA que foram adquiridos e submetidos a testes de inibição enzimática frente à sLb^{pro} do vírus da FA.

Tabela 4.45. Representação da estrutura 2D e peso molecular dos 6 compostos (FA-01 a FA-06) selecionados pelo Modelo I – FA como potenciais de inibidores covalentes da Lb^{pro} do vírus da FA e submetidos ao teste de inibição enzimática.

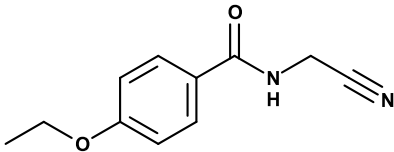
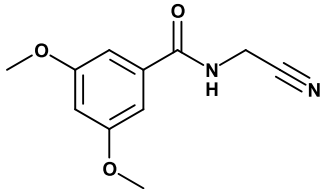
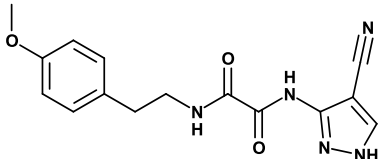
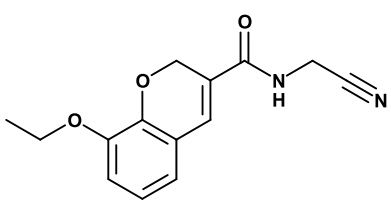
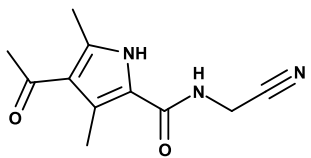
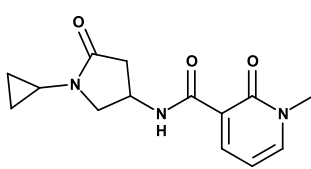
#Ref	Estrutura	Peso Molecular
FA-01		204,9
FA-02		220,23
FA-03		313,32

Tabela 4.45. (Continua).

#Ref	Estrutura	Peso Molecular
FA-04		258,28
FA-05		219,24
FA-06		275,31

4.4.2.2. Validação experimental

De modo análogo ao realizado para protease do DENV, a verdadeira utilidade dos modelos de busca virtual foi comprovada com sua validação experimental, através de ensaios enzimáticos. Assim, os 6 compostos selecionados pelo Modelo I - FA foram submetidos a ensaios de inibição enzimática frente à Lb^{pro} do vírus da FA. Todos os procedimentos experimentais usados nesta etapa foram realizados nos laboratórios da Prof^a. Antonia Tavares do Amaral (IQ-USP, São Paulo) e estão descritos no item 3.6.2.

Ensaio de inibição enzimática frente à Lb^{pro} do vírus da FA

Diferentemente do apresentado para protease do DENV, pretendeu-se, inicialmente, determinar os valores de IC₅₀ frente à sLb^{pro} do vírus da FA para todos os compostos selecionados pelo Modelo I – FA. Esta alteração se fez necessária, pois, além de encontrar inibidores, o Modelo I - FA foi construído visando obter mais informações sobre as interações no sítio ativo da sLb^{pro}.

Procedendo-se como descrito 3.6.2.1 (Materiais e Métodos), determinaram-se, tentativamente, os valores de IC_{50} frente à sLb^{pro} do vírus da FA para os compostos **FA-01** a **FA-06** a partir de suas curvas dose-resposta. No entanto, **FA-01** e **FA-03** não se apresentaram solúveis nas condições do experimento. Adicionalmente, **FA-02** e **FA-06** apresentaram fluorescência intrínseca em $\lambda_{ex} = 320$ nm e $\lambda_{em} = 420$ nm, não sendo possível determinar seus valores de IC_{50} por este protocolo (item 3.6.2.1). Assim, as curvas dose-resposta só puderam ser obtidas para **FA-04** e **FA-05** e estão apresentadas na Figura 4.57.

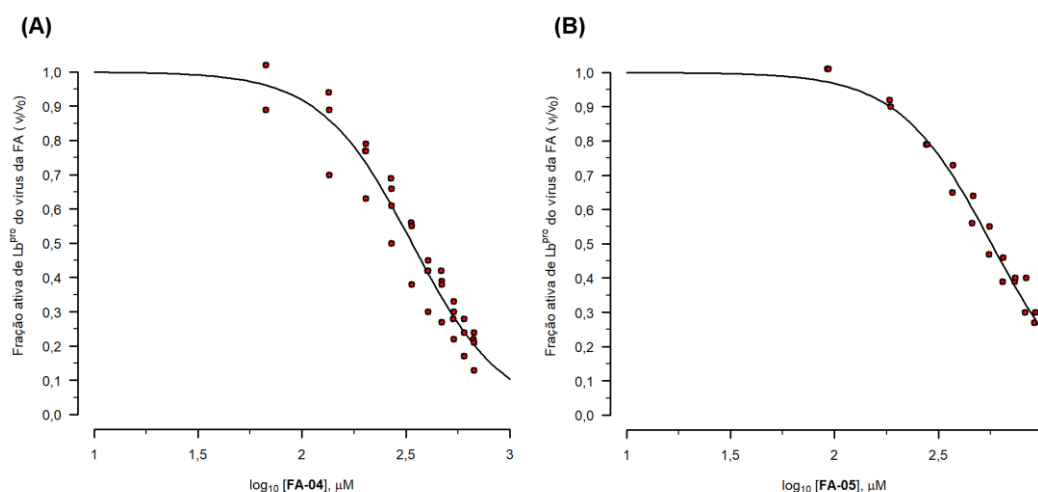


Figura 4.57. Curvas dose-resposta^(a) dos compostos **FA-04** e **FA-05**. Fração ativa^(b) de sLb^{pro} do vírus da FA em função dos valores do logaritmo da concentração^(c) (em μM) de cada composto: **FA-04** (A) e **FA-05** (B). Experimentos realizados, pelo menos, em duplicata.

- (a) Construída utilizando-se a opção LL.2 para curvas dose-resposta do pacote “drc” (RITZ; STREIBIG; RITZ, C. & STREIBIG, 2006) disponível no programa R (R CORE TEAM, 2013).
- (b) Valores de velocidade inicial de formação de produto com (v_i), e sem (v_0), o composto, determinados a partir das retas de melhor ajuste dos valores de fluorescência^(d) em função do tempo (s). Estes foram determinados usando um dos compostos: **FA-04** e **FA-05**.
- (c) Os valores da concentração de cada composto utilizados no teste variaram de 0,00 a 1000,00 μM .
- (d) Valores de fluorescência obtidos em solução tampão pH=7,8 (Borato 50 mM), contendo sLb^{pro} do vírus da FA sorotipo Ok1 (8,4 pM, disponibilizada pelo grupo do Prof Juliano), com 5 minutos de pré-incubação na presença de 2,5 mM de DTT, de KVQRKLKGAGQSSQ-Abz (3 μM , sintetizado pelo grupo do Prof Juliano) e de Triton-X100 (0,01 % v/v), à $37 \pm 0,1^\circ C$ e utilizando-se $\lambda_{em} = 320$ nm e $\lambda_{ex} = 420$ nm.

Os valores de IC_{50} dos compostos **FA-04** a **FA-05** foram iguais a $337,5 \pm 9,4 \mu M$ e a $568,4 \pm 11,5 \mu M$, respectivamente, ou seja, tem uma baixa atividade inibitória. Cabe ressaltar que estes valores foram obtidos na presença de Triton-X 100 visando-se evitar o mecanismo de inibição promíscuo (MALVEZZI et al., 2008). Adicionalmente, **FA-04** a **FA-05** tiveram a

reversibilidade da inibição avaliada objetivando verificar se estes atuam, ou não, por um mecanismo de inibição covalente (COPELAND, 2005).

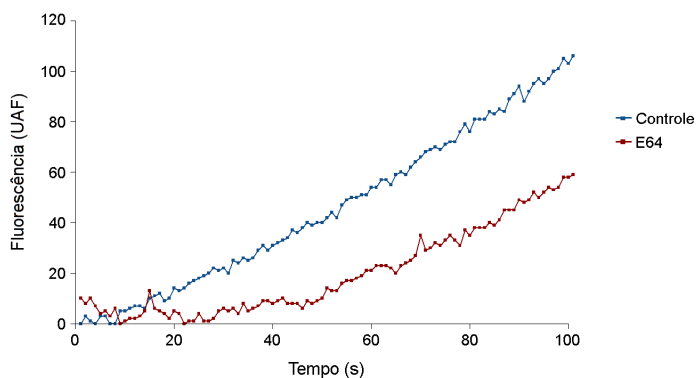


Figura 4.58. Curvas do progresso de reação^(a) obtidas imediatamente após diluição da solução^(b) de sLb^{pro} com E64. Valores de fluorescência^(c) em função do tempo (s). Mostrando as curvas dos experimentos controle (sem E64) e com E64 em azul e vermelho, respectivamente. Experimento realizado uma única vez.

- (a) Construída utilizando-se o programa *R* (R CORE TEAM, 2013).
- (b) Solução tampão pH=7,8 (Borato 50 mM) com sLb^{pro} (169 pM, disponibilizada pelo grupo do Prof Juliano) com 5 minutos de pré-incubação na presença de DTT (2,5 mM) e com E64 (497 µM) mantida em banho de gelo, sob agitação a cada 5 min, por 30 min. O experimento controle foi feito nas mesmas condições adicionando, no entanto, somente DMSO à solução tampão.
- (c) Valores de fluorescência obtidos em solução tampão pH=7,8 (Borato 50 mM), contendo sLb^{pro} do vírus da FA do sorotipo Ok1 (1,69 pM, disponibilizada pelo grupo do Prof Juliano), KVQRKLKGAGQSSQ-Abz (3 µM, sintetizado pelo grupo do Prof Juliano), Triton-X 100 (0,01 % v/v) com, ou sem, E64 (5,0 µM) à 37±0,1°C e utilizando-se $\lambda_{em} = 320\text{nm}$ e $\lambda_{ex} = 420\text{ nm}$.

Na Figura 4.58, observa-se que para o inibidor E64, usado para validar o protocolo experimental (vide item 3.6.2.2), a inclinação da curva do progresso da reação, obtida imediatamente após a diluição, aumenta lentamente com o tempo aproximando-se tardiamente da curva controle. Estes resultados indicam que E64 apresenta mecanismo de inibição lentamente reversível (COPELAND, 2005). No entanto, na literatura (ARAFET et al., 2014) sugere-se que E64 deveria apresentar mecanismo de inibição totalmente irreversível. Esta falta de concordância poderia ser explicada, considerando-se que o tempo de incubação (30 minutos) não foi suficiente para atingir 100 % de inibição enzimática pelo mecanismo covalente. Ainda assim, o protocolo experimental foi considerado validado e, assim, aplicado aos compostos **FA-04** a **FA-05**. As curvas de progresso de reação para **FA-04** a **FA-05** foram obtidas e estão apresentadas nas Figura 4.59 e Figura 4.60, respectivamente.

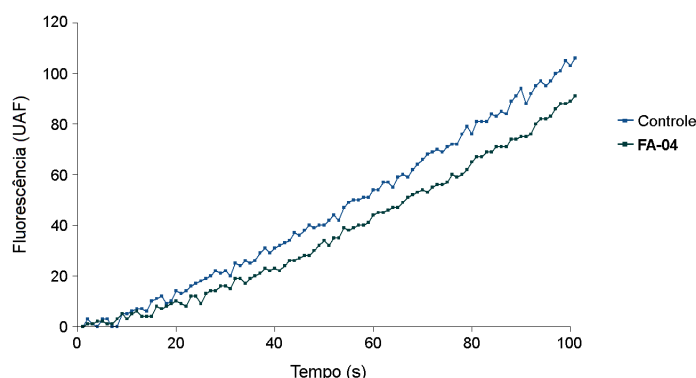


Figura 4.59. Curvas do progresso de reação^(a) obtidas imediatamente após diluição da solução^(b) de sLb^{pro} com FA-04. Valores de fluorescência^(c) em função do tempo (s). Mostrando as curvas dos experimentos controle (sem FA-04) e com FA-04 em azul e verde, respectivamente. Experimento realizado uma única vez.

- (a) Construída utilizando-se o programa *R* (R CORE TEAM, 2013).
- (b) Solução tampão pH=7,8 (Borato 50 mM) com sLb^{pro} (169 pM, disponibilizada pelo grupo do Prof Juliano) com 5 minutos de pré-incubação na presença de DTT (2,5 mM) e com FA-04 (5,3 μM) mantida em banho de gelo, sob agitação a cada 5 min, por 30 min. O experimento controle foi feito nas mesmas condições adicionando, no entanto, somente DMSO à solução tampão.
- (c) Valores de fluorescência obtidos em solução tampão pH=7,8 (Borato 50 mM), contendo sLb^{pro} do vírus da FA do sorotipo Ok1 (1,69 pM, disponibilizada pelo grupo do Prof Juliano), KVQRKLKGAGQSSQ-Abz (3 μM, sintetizado pelo grupo do Prof Juliano), Triton-X 100 (0,01 % v/v) com, ou sem, FA-04 (53 nM) à 37±0,1°C e utilizando-se $\lambda_{em} = 320\text{nm}$ e $\lambda_{ex} = 420\text{ nm}$.

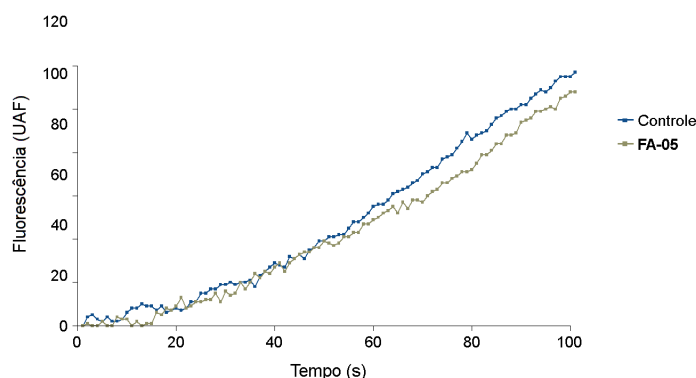


Figura 4.60. Curvas do progresso de reação^(a) obtidas imediatamente após diluição da solução^(b) de sLb^{pro} com FA-05. Valores de fluorescência^(c) em função do tempo (s). Mostrando as curvas dos experimentos controle (sem FA-05) e com FA-05 em azul e cinza, respectivamente. Experimento realizado uma única vez.

- (a) Construída utilizando-se o programa *R* (R CORE TEAM, 2013).
- (b) Solução tampão pH=7,8 (Borato 50 mM) com sLb^{pro} (169 pM, disponibilizada pelo grupo do Prof Juliano) com 5 minutos de pré-incubação na presença de DTT (2,5 mM) e com FA-05 (6,8 μM) mantida em banho de gelo, sob agitação a cada 5 min, por 30 min. O experimento controle foi feito nas mesmas condições adicionando, no entanto, somente DMSO à solução tampão.
- (c) Valores de fluorescência obtidos em solução tampão pH=7,8 (Borato 50 mM), contendo sLb^{pro} do vírus da FA do sorotipo Ok1 (1,69 pM, disponibilizada pelo grupo do Prof Juliano), KVQRKLKGAGQSSQ-Abz (3 μM, sintetizado pelo grupo do Prof Juliano), Triton-X 100 (0,01 % v/v) com, ou sem, FA-04 (68 nM) à 37±0,1°C e utilizando-se $\lambda_{em} = 320\text{nm}$ e $\lambda_{ex} = 420\text{ nm}$.

Para FA-04 e FA-05, a concentração de compostos na solução mais concentrada (solução mãe) do complexo enzima-inibidor foi menor do que a usada para E64 (5,3 nM de

FA-04 e 6,8 nM de **FA-05** contra 497 μ M de E64). Isto ocorreu pela impossibilidade de se aumentar a concentração destes compostos na solução concentrada do complexo enzima-inibidor (100 μ L) sem um aumento significativo do volume total de DMSO. Considerando tanto as baixas concentrações usadas bem como o curto tempo de incubação (30 minutos), esta avaliação foi considerada ainda preliminar. Nesta avaliação preliminar foi observado que as inclinações das curvas do progresso da reação para os compostos **FA-04** e **FA-05** (Figura 4.59 e Figura 4.60) aumentam rapidamente com o tempo aproximando-se da curva controle. Assim, estes resultados, embora preliminares, sugerem que **FA-04** e **FA-05** não se comportam como inibidores covalentes como proposto e planejado, pois não se observou inibição irreversível, ou lentamente, reversível.

4.4.2.3. Análise das interações dos inibidores compostos **FA-04** e **FA-05** com resíduos do sítio ativo da Lb^{pro} do vírus da FA

Embora os compostos **FA-04** e **FA-05** apresentem uma baixa atividade inibitória frente à Lb^{pro} do vírus da FA, como descrito acima, os estudos de ancoramento da série de 52 compostos potenciais inibidores também foram feitos para obter mais informações sobre as interações no sítio ativo da Lb^{pro}. Desta forma, a melhor pose de ancoramento, apresentando maior valor de escore, de cada composto foi usada para obter alguma informação sobre as interações no sítio ativo desta protease. As interações entre **FA-04** e **FA-05** e os resíduos do sítio ativo Lb^{pro}, preditas pelo ancoramento, estão apresentadas na Figura 4.61.

Estes dois compostos formam ligações de hidrogênio com a *His148* catalítica e com o *Glu147* do sítio *S*₁. Adicionalmente, os dois compostos formam interações hidrofóbicas com resíduos da cavidade *S*₂ (*Leu143* e *Ala149*), no entanto, **FA-04** forma uma interação hidrofóbica (*Leu178*) a mais do que **FA-05**. Ainda, para **FA-04** foi observada ligação de hidrogênio com a *Gly98* (cavidade *S*₂), enquanto que, para **FA-05** foi observada ligação de hidrogênio com *Asn46* (cavidade oxânion).

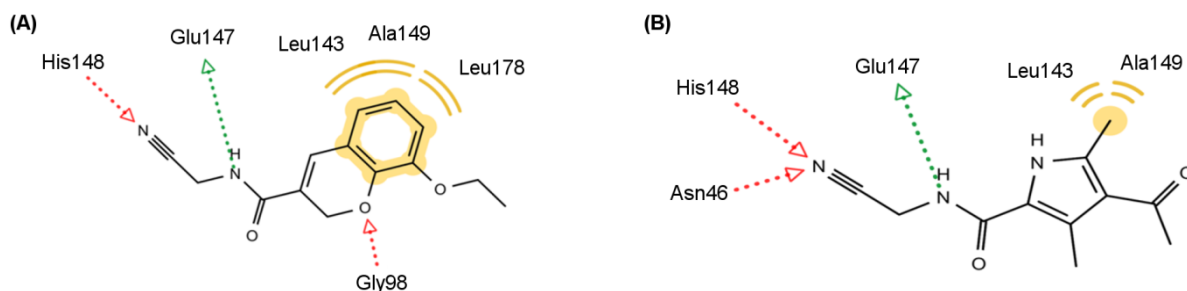


Figura 4.61. Representação esquemática, em 2D, das interações^(a) previstas para a melhor pose de ancoramento^(b) de cada um dos compostos com os resíduos do sítio ativo da sLb^{pro} do vírus da FA^(c). (A) e (B) mostram as interações previstas para melhor pose de ancoramento de **FA-04** ($IC_{50} = 337,5 \pm 9,4 \mu M$) e de **FA-05** ($IC_{50} = 568,4 \pm 11,5 \mu M$), respectivamente. As interações *HBD* e *HBA* estão representadas pelas setas verdes e vermelhas, respectivamente, e as interações hidrofóbicas pelas esferas amarelas.

- (a) Interações identificadas com o auxílio do programa *LigandScout* (WOLBER; LANGER, 2005), usando os valores de corte padrão do programa.
 (b) Pose com maior valor de escore obtida usando a função *ChemPLP* e o programa *GOLD* (JONES et al., 1997).
 (c) A estrutura sLb^{pro} do vírus da FA corresponde ao PDB 2JQG (obtida por RMN (CENCIC et al., 2007)).

O fato de **FA-04** fazer mais interações hidrofóbicas que **FA-05** poderia explicar por que o primeiro apresenta a maior atividade. Adicionalmente, para este composto, podemos observar que a cavidade S_2 é ocupada pelo anel aromático de **FA-04** (Figura 4.61(A)). Considerando que a cavidade S_2 da Lb^{pro} tem uma restrição a aminoácidos aromáticos (SANTOS et al., 2009), parece razoável propor que **FA-04** apresentaria uma maior atividade se seu anel aromático fosse substituído por um grupo hidrofóbico, mas, não aromático. Com base no observado, os estudos de ancoramento sugeriram como favoráveis interações com os seguintes resíduos: *Gly98*, *Leu143*, *Glu147*, *His148*, *Ala149* e *Leu178*. Estas interações serão, na medida do possível, consideradas na construção de um segundo modelo de busca virtual para selecionar inibidores da Lb^{pro} do vírus da FA. Por fim, cabe ressaltar que a importância dos resíduos *Asn46*, *Gly98*, *Leu143*, *Glu147*, *Ala149* e *Leu178* para as interações Lb^{pro}-ligante puderam ser confirmadas com a publicação da estrutura da sLb^{pro} complexada com E64-R-P-NH₂ (4QBB (STEINBERGER et al., 2014)). Assim, exceto pela interação com a *His148*, todas as interações previstas pelos estudos de ancoramento entre Lb^{pro} e **FA-04/FA-05** são encontradas na estrutura 4QBB.

4.4.3. Modelos de busca virtual: Modelo II - FA

Considerando a baixa atividade inibitória dos compostos selecionados pelo Modelo II – FA, foi proposto um segundo modelo de busca virtual de inibidores da Lb^{pro} do vírus da FA, denominado Modelo II - FA. Para este segundo modelo, tentou-se considerar, dentro do possível, as interações com os resíduos do sítio ativo sugeridas como favoráveis nos estudos de ancoragem realizados para Modelo I – FA (item 4.4.2.3). Além disso, no Modelo II – FA foram introduzidas algumas modificações em relação ao primeiro modelo, sendo estas: (i) no banco de dados de moléculas virtuais usado na busca virtual; (ii) no número e no tipo de filtros de seleção aplicados ao banco de dados e (iii) nos critérios de cada um destes filtros.

A primeira mudança em relação ao Modelo I – FA aplicada na construção deste Modelo II – FA foi usar o *ZINC-Curated* visando selecionar compostos sem grupos eletrofílicos. Esta mudança foi feita considerando-se que os compostos **FA-04** e **FA-05**, selecionados pelo primeiro modelo, não parecem ser inibidores covalentes da Lb^{pro} (vide pg. 341), não sendo mais necessário considerar apenas compostos com grupos eletrofílicos. A segunda mudança foi a inclusão de um número maior de filtros de seleção com naturezas diferentes. Assim, procedendo-se como descrito item 3.5.5.6 (Materiais e Métodos) e partindo-se do banco de dados *ZINC-Curated*, foram aplicados três filtros de seleção na seguinte sequência: farmacofórico, similaridade e inspeção visual. Ainda, a partir da publicação, em 11/2014, da estrutura da sLb^{pro} complexada com o inibidor E64-R-P-NH₂ (4QBB, resolução 1,6 Å (STEINBERGER et al., 2014)), foi possível incluir também as informações sobre as interações entre este inibidor e os resíduos do sítio ativo da Lb^{pro} na construção do modelo farmacofórico. Por fim, de modo análogo ao primeiro modelo, os compostos selecionados por este segundo modelo (Modelo II – FA) foram testados como inibidores da Lb^{pro} do vírus da FA. A Figura 4.62 mostra uma representação esquemática da

sequência de aplicação de cada filtro bem como os valores da porcentagem de redução do número de compostos em cada etapa.

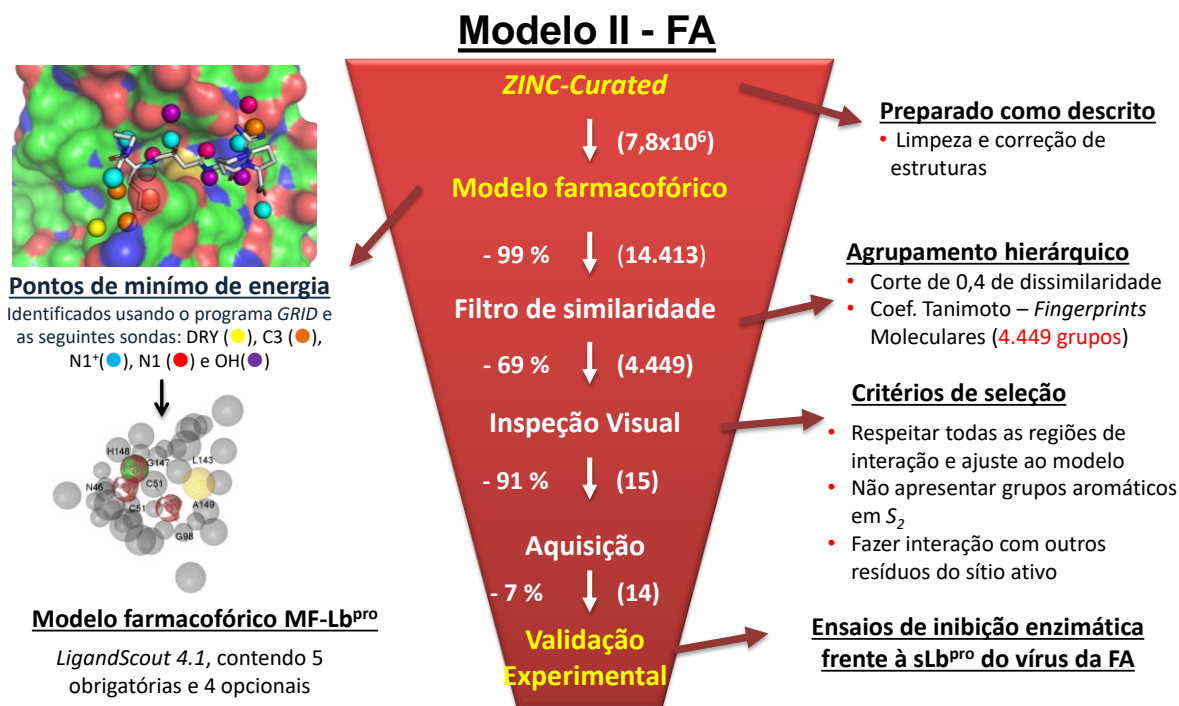


Figura 4.62. Representação esquemática da sequência de filtros de seleção aplicados ao banco *ZINC-Curated* usada no segundo modelo de busca virtual (**Modelo II – FA**). Entre parênteses são indicados os números de compostos selecionados em cada etapa e ao lado a porcentagem de redução por cada filtro.

4.4.3.1. Filtro farmacofórico

A primeira etapa do procedimento de busca virtual foi a aplicação de um filtro farmacofórico ao banco de dados *ZINC-Curated*. Como descrito na Introdução desta tese (item 1.7.1), modelos farmacofóricos podem também ser construídos usando coordenadas de pontos de mínimo de energia obtidos a partir da estrutura do receptor sem ligantes. Assim, procedendo-se como descrito no item 3.5.5.6 (Materiais e Métodos), o modelo farmacofórico foi gerado usando pontos de mínimo de energia obtidos pelo programa *GRID* (GOODFORD, 1985). Estes pontos de mínimo de energia foram obtidos usando a estrutura 4QBB e sondas moleculares de diferentes naturezas, a saber: hidrofóbica (“DRY” e “C3”), iônica (“N1⁺” e “O⁻”), aceptoras de ligação de hidrogênio e/ou doadoras de ligação de hidrogênio (“N1”, “O” e “OH”). A Figura 4.63 (A) mostra todos os pontos de mínimo de energia identificados. Em

seguida, aqueles próximos aos resíduos do sítio ativo da sLb^{pro} foram selecionados e sobrepostos ao inibidor E64-R-P-NH₂ (Figura 4.63 (B)).

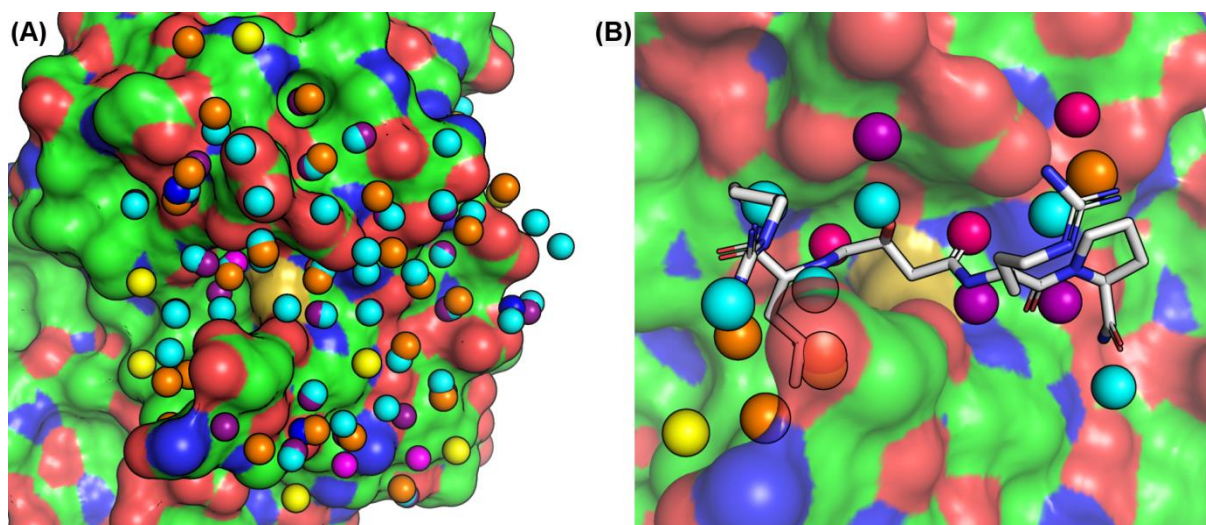


Figura 4.63. Representação dos pontos de mínimo de energia^(a) obtidos^(b). (A) Mostrando todos os pontos de mínimo de energia. (B) Mostrando aqueles próximos ao sítio ativo da sLb^{pro(c)} sobrepostos ao inibidor E64-R-P-NH₂. O sítio ativo da sLb^{pro} e o inibidor E64-R-P-NH₂ estão representados como superfície (colorida de acordo com o tipo de átomo) e como *sticks*, respectivamente. As esferas correspondem aos pontos de mínima energia identificados pelas sondas “DRY” (amarelo), “C3” (laranja), “N1+” (azul), “N1” (ciano), “O-” (vermelho), “O” (magenta) e “OH” (roxo).

^(a) Representando as regiões de interação(ões) favorável(is)

^(b) Obtidos a partir dos campos de interação moleculares calculados pelo programa GRID (GOODFORD, 1985) aplicando-se os valores máximos de energia de corte descritos no item 3.5.5.4.

^(c) Estrutura 4QBB preparada pelo programa SYBYL-X (SYBYL, 2013) como descrito no item 3.5.5.2.

Analisando a Figura 4.63 (B), podemos observar que não foram identificados pontos de mínimo de energia representando regiões de interação iônica positiva (sonda “N1+”) ou negativa (sonda “O-”), embora fosse esperado identificar regiões de interação iônica positiva próximos aos resíduos *Asp49* e *Glu147*. Nas cavidades *S*₁, *S*₁’ e *S*₂’ há um predomínio de pontos de mínimo de energia representando regiões de ligação de hidrogênio (esferas ciano, rosa e roxas). Ainda, há um predomínio de pontos de mínimo de energia representando interações hidrofóbicas (esferas amarelas e laranjas) próximos à cavidade *S*₂. Por fim, pode-se observar que há uma concordância entre os pontos de mínimo identificados e os átomos do inibidor E64-R-P-NH₂. As regiões associadas a cada ponto de mínimo de energia foram analisadas e selecionadas priorizando, dentro do possível, tanto aquelas capazes de estabelecer interações com mais de um resíduo bem como as observadas experimentalmente (estrutura

4QBB vide Figura 4.52) e/ou preditas pelo ancoramento (**FA-04** e **FA-05** vide Figura 4.61). Estas regiões foram usadas para construir um modelo farmacofórico, denominado MF-Lb^{pro}, que está apresentado na Figura 4.64. O modelo MF-Lb^{pro} contém nove regiões de interação, sendo elas: uma *HBD* (não-direcional obrigatória), sete *HBA* (três não-direcionais obrigatórias e quatro direcionais opcionais) e uma hidrofóbica (obrigatória).

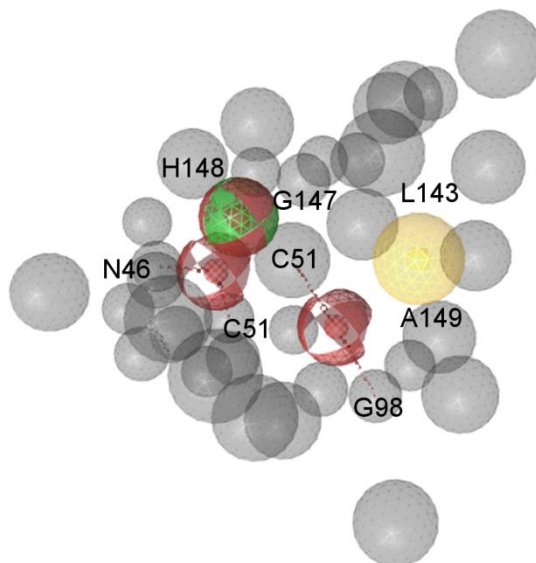


Figura 4.64. Representação esquemática do modelo farmacofórico^(a) MF-Lb^{pro}. Este modelo contendo 9 regiões de interação^(b), sendo 5 obrigatórias (1 *HBD* e 3 *HBA* e 1 hidrofóbica) e 4 opcionais (4 *HBA*) e várias regiões de exclusão de volume^(c). As regiões *HBD* e *HBA*, hidrofóbicas e de exclusão de volume estão apresentadas em verde, vermelho, amarelo e cinza, respectivamente. As setas indicam regiões de interação direcionais e as esferas indicam região de interação ou exclusão de volume não-direcionais. As regiões hachuradas correspondem a regiões de interação opcionais.

- (a) Gerado pelo programa *LigandScout* (WOLBER; LANGER, 2005) a partir dos pontos de mínimo de energia^(d) representando regiões de interação(ões) favorável(eis) no sítio ativo da Lb^{pro(e)}.
- (b) Todas as regiões de interação tem um raio de 1,5 Å.
- (c) Definidas a partir das coordenadas dos átomos dos resíduos que formam o sítio ativo de cada estrutura.
- (d) Obtidos como descrito na Figura 4.63.
- (e) Estrutura 4QBB prepara pelo programa *SYBYL-X* (SYBYL, 2013) como descrito no item 3.5.5.2.

O modelo farmacofórico MF-Lb^{pro} foi aplicado ao banco *ZINC-Curated* selecionando 14.413 (redução de 99,86 %) compostos capazes de satisfazer todas as características obrigatórias do modelo (três regiões doadoras de ligação de hidrogênio, uma região aceptora de ligação de hidrogênio e uma região hidrofóbica). Em seguida, os 14.413 compostos foram selecionados e submetidos a um filtro de similaridade.

4.4.3.2. Filtro de similaridade

De modo análogo ao descrito para o Modelo I – DENV (item 4.3.2.3), um filtro de similaridade foi aplicado visando excluir compostos similares mantendo, no entanto, a diversidade estrutural (Bologa and Oprea, 2012). Procedendo-se como descrito item 3.5.5.6 (Materiais e Métodos), os compostos foram agrupados com base nos valores do coeficiente de Tanimoto. O valor de corte de 60 % de similaridade usado resultou em 4.449 grupos. A seguir, um representante de cada grupo foi selecionado para a próxima etapa. Assim, 4.449 compostos (redução de 69,13 %) foram selecionados.

4.4.3.3. Filtro de inspeção visual

Como discutido na Introdução e de modo análogo ao realizado para a protease do DENV, os 4.449 compostos selecionados no filtro de similaridade foram analisados por inspeção visual para verificar seus ajustes no sítio ativo da Lb^{pro} do vírus da FA. Esta análise foi feita procedendo-se como descrito no item 3.5.5.8 (Materiais e Métodos). Numa primeira etapa, foram selecionados apenas os compostos capazes de satisfazer todas as características propostas pelo modelo MF-Lb^{pro}, inclusive regiões de interação definidas como opcionais. Assim, dos 4.449 compostos selecionados, apenas 177 (4 %) foram mantidos. Numa segunda etapa, cada um dos 177 compostos foi inspecionado dentro do sítio ativo da Lb^{pro} do vírus da FA. Além do ajuste no sítio ativo, nesta inspeção visual foram incluídos outros critérios. Dentre eles, verificar se o composto não apresentava grupamento aromático na região de interação hidrofóbica. Este critério foi incluído considerando que a cavidade S_2 da Lb^{pro} tem uma restrição a aminoácidos aromáticos (SANTOS et al., 2009). Uma descrição detalhada de todos os critérios usados está apresentada no item 3.5.5.8.

Nesta etapa, 15 compostos foram selecionados como potenciais inibidores da Lb^{pro} do vírus da FA, representando uma redução de 91 % em relação aos 177 compostos selecionados na etapa anterior. Destes 15 potenciais inibidores, 14 (93,33 %) estavam disponíveis e foram

adquiridos para validação experimental do Modelo II - FA. Por fim, a aplicação de todos os filtros de seleção representou uma redução total de mais de 99 % em relação ao banco *ZINC-Curated* ($\sim 7,8 \times 10^6$ compostos). A Tabela 4.46 mostra a representação das estruturas dos quatorze compostos (denominados **FA-07** a **FA-20**) selecionados pelo Modelo II – FA que foram adquiridos e submetidos a testes de inibição enzimática frente à sLb^{pro} do vírus da FA.

Tabela 4.46. Representação da estrutura 2D dos 14 compostos (FA-07 a FA-20) selecionados pelo Modelo II – FA como potenciais inibidores da Lb^{pro} do vírus da FA. Estes foram adquiridos e submetidos ao teste de inibição enzimática frente a esta protease. Estas estão acompanhadas dos seus correspondentes códigos ZINC, valores de peso molecular e empresa fornecedora.

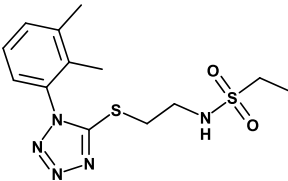
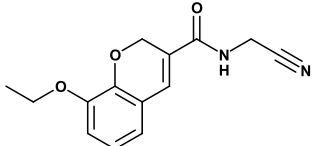
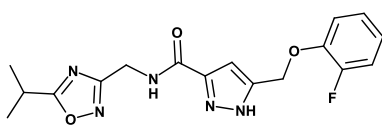
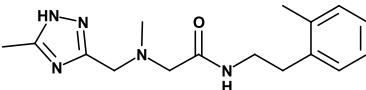
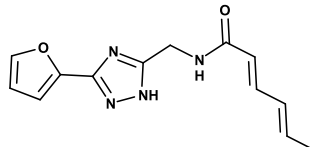
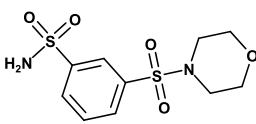
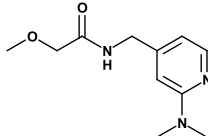
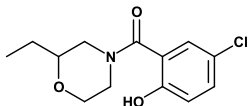
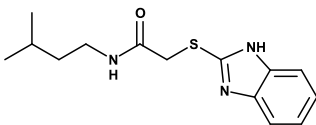
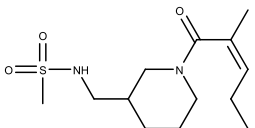
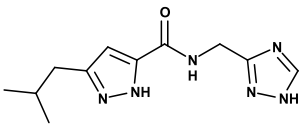
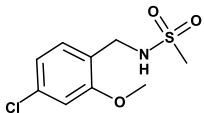
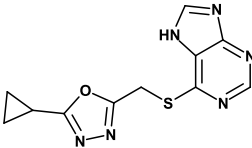
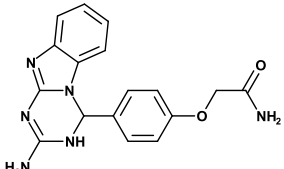
#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
FA-07		68044297	341,45	Enamine
FA-08		71905440	247,28	Enamine
FA-09		67846966	359,35	ChemBridge
FA-10		75107777	301,39	Enamine
FA-11		69842754	258,28	Enamine
FA-12		22409626	306,36	Enamine
FA-13		40512559	223,27	Enamine

Tabela 4.46. (Continua).

#Ref	Estrutura	Código ZINC	Peso Molecular	Vendedor
FA-14		30597143	269,72	Enamine
FA-15		03020332	277,39	UORSY
FA-16		69434864	288,41	Enamine
FA-17		75136897	248,28	Enamine
FA-18		69808851	249,71	Enamine
FA-19		71905439	274,30	Enamine
FA-20		09318924	336,35	UORSY

4.4.3.4. Validação experimental

De modo análogo realizado para o Modelo I - FA, a verdadeira utilidade do modelo Modelo II – FA foi comprovada com sua validação experimental, através de ensaios enzimáticos. Assim, os 14 compostos selecionados pelo Modelo II – FA foram submetidos a ensaios de inibição enzimática frente à sLb^{pro} do vírus da FA. Todos os procedimentos experimentais usados nesta etapa foram feitos nos laboratórios da Prof^a. Antonia Tavares do Amaral (IQ-USP, São Paulo) e estão descritos no item 3.6.2.

Ensaio de inibição enzimática frente à Lb^{pro} do vírus da FA

De modo análogo ao realizado para protease do DENV e como descrito no item 3.6.2.1 (Materiais e Métodos), nos ensaios de inibição enzimática foram determinados valores de porcentagem de inibição da sLb^{pro} do vírus da FA. Ainda, cabe ressaltar que nestes ensaios foi utilizado um substrato comercial contendo o AMC (7-Amido-4-metil-cumarina) como grupo fluorogênico, ao invés do grupo Abz (amino-benzoil) presente no substrato usado nos ensaios para o Modelo I – FA. Esta modificação foi feita visando evitar que os compostos adquiridos não pudessem ser analisados, por apresentarem fluorescência intrínseca em $\lambda_{ex} = 320$ nm e $\lambda_{em} = 420$ nm, como observado para os compostos **FA-02** e **FA-06**. O AMC foi escolhido como grupo fluorogênico pelo fato deste ser mais amplamente usado na literatura.

Considerando a modificação feita no protocolo experimental e visando comparar os compostos **FA-07** a **FA-20** com os selecionados pelo Modelo I – FA, o valor da porcentagem de inibição da sLb^{pro} do vírus da FA do composto **FA-04** (em 200 μ M) foi determinado, sendo igual a 42 ± 5 %. Em seguida, o valor de porcentagem de inibição da sLb^{pro} do vírus da FA de cada composto (em 200 μ M) foi determinado, sendo apresentados na Tabela 4.47. Diferentemente dos compostos selecionados pelo Modelo I – FA, todos os compostos selecionados pelo Modelo II – FA foram solúveis e não apresentam fluorescência intrínseca, nas condições do experimento. O composto mais ativo desta série foi o **FA-17** (inibição da sLb^{pro} igual a 40 ± 3 %). Este composto apresentou, dentro do erro experimental, um valor de porcentagem de inibição comparável ao do **FA-04** (42 ± 5 %). Sendo assim, nenhum dos 14 compostos selecionados pelo Modelo II – FA apresentou alta atividade inibitória frente à Lb^{pro} do vírus da FA. Ainda, três compostos parecem ter ativado a enzima, ao invés de inibi-la, ou seja, a atividade da enzima na presença do composto foi maior do que na ausência, sendo eles: compostos **FA-07** (ZINC68044297), **FA-08** (ZINC71905440) e **FA-12** (ZINC22409626).

Tabela 4.47. Valores de porcentagem de inibição (em 200 μM)^(a) e dos correspondentes desvios padrão^(b) de cada um dos compostos (FA-07 a FA-21) selecionados pelo Modelo II – FA.

#Ref	Estrutura	Inibição da sLb ^{pro} do vírus da FA (% , em 200 μM)
FA-07		A.E. ^(c)
FA-08		A.E. ^(c)
FA-09		25 \pm 9
FA-10		20 \pm 3
FA-11		7 \pm 1
FA-12		A.E. ^(c)
FA-13		16 \pm 4
FA-14		15 \pm 1
FA-15		20 \pm 10

^(a) Determinados em solução tampão pH=7,8 (Borato 50 mM) contendo sLb^{pro} do vírus da FA sorotipo Ok1 (8,4 pM, disponibilizada pelo grupo do Prof Juliano), com 5 minutos de pré-incubação na presença de 2,5 mM de DTT, Triton-X 100 (0,001 % v/v), Z-LR-MCA (100 μM) e cada um dos compostos (200 μM).

^(b) Considerando pelo menos três experimentos independentes.

^(c) A.E. significa ativação da enzima, ou seja, a atividade enzimática na presença do composto foi maior do que a observada na ausência do composto.

Tabela 4.47. (Continua).

#Ref	Estrutura	Inibição da sLb ^{pro} do vírus da FA (% , em 200 μ M)
FA-16		4 \pm 1
FA-17		40 \pm 3
FA-18		2 \pm 6
FA-19		27 \pm 8
FA-20		5 \pm 13

^(a) Determinados em solução tampão pH=7,8 (Borato 50 mM) contendo sLb^{pro} do vírus da FA sorotipo Ok1 (8,4 pM, disponibilizada pelo grupo do Prof Juliano), com 5 minutos de pré-incubação na presença de 2,5 mM de DTT, Triton-X 100 (0,001 % v/v), Z-LR-MCA (100 μ M) e cada um dos compostos (200 μ M).

^(b) Considerando pelo menos três experimentos independentes.

Dentre as possíveis justificativas para a falta de sucesso desse modelo, quanto à atividade inibitória, podemos considerar a não inclusão de uma etapa de avaliação de desempenho do modelo farmacofórico devido a pouca informação disponível sobre compostos ativos e inativos frente à Lb^{pro} do vírus da FA. Independentemente da razão do insucesso do Modelo II – FA, os resultados obtidos até este momento não foram plenamente satisfatórios, desta forma, outro modelo de busca virtual será, futuramente, proposto considerando as informações sobre os compostos inativos obtidos até o presente momento.

5. CONCLUSÕES

Criação do banco *ZINC-Curated*

As ferramentas estatísticas e quimiométricas usadas permitiram analisar e avaliar extensivamente a diversidade estrutural e das propriedades físico-químicas do banco de dados virtuais de moléculas construído nesta tese, denominado *ZINC-Curated*. Nestas análises, observou-se uma certa diminuição da diversidade estrutural e das propriedades físico-químicas deste banco quando comparado ao original (*ZINC-2014*). No entanto, as características dos compostos do banco *ZINC-Curated* são mais próximas às observadas para os compostos do banco DB (representando o espaço químico dos fármacos). Dentre estas, a solubilidade em água e a estabilidade metabólica. Portanto, concluímos que o uso dos filtros de propriedade *drug-like* e estruturais para construir o banco *ZINC-Curated* enriqueceu este banco com compostos com características de fármacos. Adicionalmente, todos os compostos selecionados a partir do banco *ZINC-Curated* e, por nós testados, foram solúveis e estáveis nas condições dos experimentos de inibição (pH 7,8 ou pH 9,0).

Busca virtual de inibidores da protease do vírus da Dengue (NS2B/NS3pro DENV)

O uso de estratégias *LBDD* e *SBDD* permitiu construir quatro modelos de busca virtual de inibidores da protease do vírus da Dengue (Modelo I – DENV a Modelo IV – DENV). A aplicação destes quatro modelos a banco de dados comerciais (*ZINC-Original* e *NCI*) e ao *ZINC-Curated* resultou na identificação de oito inibidores desta protease (**DV-27**, **DV-29**, **DV-30**, **DV-33**, **DV-53**, **DV-54**, **DV-55** e **DV-57** com $IC_{50} < 100 \mu M$). O inibidor mais ativo identificado (**DV-57**) apresenta atividade frente à protease do DENV2 e do DENV3 ($IC_{50} = 4,7 \pm 0,3 \mu M$ para DENV2 e $IC_{50} = 3,5 \pm 0,2 \mu M$ para DENV3). Ainda, este inibidor apresenta um mecanismo de inibição misto, ou seja, não competitivo (K_i e αK_i iguais a 10,4 e 6,1 μM) com maior afinidade pela enzima na presença de substrato. A comparação entre o **DV-57** com um inibidor contendo a mesma subestrutura sugere que **DV-57** poderá ser

melhor estrutura de partida. Considerando apenas o modelo que selecionou o **DV-57**, a taxa de acerto foi de ~ 22 %. Este resultado é bastante importante e significativo, principalmente se considerarmos que o alvo analisado é uma protease.

Adicionalmente, as mudanças realizadas nos modelos de busca virtual (descrita em detalhes nesta tese) permitiram melhorar, significativamente, os resultados. A comparação dos resultados obtidos pelos modelos I - DENV, III - DENV e IV - DENV permite concluir que a inclusão da flexibilidade conformacional na criação de modelos farmacofóricos, em conjunto com a avaliação do desempenho destes modelos, permitiu obter resultados significativamente melhores.

Ainda, o uso das simulações de dinâmica molecular permitiu estudar alguns aspectos relacionados à flexibilidade conformacional da NS2B/NS3pro do DENV, resultando em uma publicação (PICCIRILLO, et al. 2016). De acordo com o sugerido, nosso estudo mostra que a conformação fechada é predominante, mesmo na ausência de ligante. Sendo assim, esta é a mais relevante para ser usada nas estratégias de *SBDD*. Ainda, conseguimos identificar outras conformações da NS2B/NS3pro do DENV (Conf-1, Conf-2 e Conf-5) que podem ser relevantes para busca de inibidores usando *SBDD*, sendo algumas incluídas em nossos modelos de busca virtual. Além disso, os estudos permitiram identificar quais resíduos do sítio ativo desta protease formam interações estáveis com o inibidor Bz-nKRR-H, sendo eles: *Asp81* (NS2B), *His51*, *Asp129* e *Gly151*. O conhecimento destas interações forneceu subsídios para as etapas de seleção dos compostos usando inspeção visual (Modelos II - DENV a IV - DENV). Por fim, concluímos que estes estudos contribuíram para o conhecimento da interação Bz-nKRR-H – NS2B/NS3pro do DENV, que poderá ser usado em outras buscas virtuais.

Busca virtual de inibidores da protease do vírus da Febre Aftosa (Lb^{pro} do FA)

O uso de estratégias *SBDD* permitiu construir dois modelos de busca virtual de inibidores da protease do vírus da Febre Aftosa (Modelo I – FA e Modelo II - FA). A aplicação destes dois modelos a uma série *in house* de compostos e ao banco *ZINC-Curated* resultou na identificação de dois compostos com alguma atividade inibitória frente a esta protease (**FA-04** e **FA-05** com IC_{50} igual a $337,5 \pm 9,4 \mu M$ e a $568,4 \pm 11,5 \mu M$, respectivamente). Os resultados preliminares da reversibilidade de inibição sugerem que **FA-04** e **FA-05** não atuam por um mecanismo covalente. Ainda assim, a atividade destes compostos frente à Lb^{pro} foi bastante baixa. Embora não desejados, estes resultados podem ser atribuídos ao fato de que Lb^{pro} é uma protease e, principalmente, que poucos inibidores desta protease foram descritos na literatura.

Ainda assim, os estudos de ancoramento realizados e apresentados nesta tese permitiram identificar resíduos do sítio ativo favoráveis a interação, sendo eles: *Asn46*, *Gly98*, *His148*, *Glu147*, *Leu143*, *Ala149* e *Leu178*. A importância dos resíduos *Asn46*, *Gly98*, *Leu143*, *Glu147*, *Ala149* e *Leu178* para as interações Lb^{pro} -ligante pode ser confirmada com a publicação da estrutura da sLb^{pro} complexada com E64-R-P-NH₂ (4QBB (STEINBERGER et al., 2014)). Assim, os estudos realizados nesta tese puderam contribuir para o entendimento das interações Lb^{pro} -ligante.

6. REFERÊNCIAS

- ABEL, R. et al. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. **Journal of the American Chemical Society**, v. 130, n. 9, p. 2817–2831, mar. 2008.
- ABEL, R. et al. Advancing Drug Discovery through Enhanced Free Energy Calculations. **Accounts of Chemical Research**, v. 50, n. 7, p. 1625–1632, 18 jul. 2017.
- ABIEC. **Exportações Brasileiras de Carne Bovina**. Disponível em: <<http://www.abiec.com.br/download/anual-2016.pdf>>. Acesso em: 4 ago. 2017.
- AMARAL, A. T. -D. et al. A evolução da Química Medicinal no Brasil: avanços nos 40 anos da Sociedade Brasileira de Química. **Química Nova**, v. 40, n. 6, p. 694–700, 3 jul. 2017.
- ANDREWS, P. R.; CRAIK, D. J.; MARTIN, J. L. Functional group contributions to drug-receptor interactions. **Journal of medicinal chemistry**, v. 27, n. 12, p. 1648–1657, 1984.
- ARAFET, K. et al. Quantum mechanics/molecular mechanics studies of the mechanism of falcipain-2 inhibition by the epoxysuccinate E64. **Biochemistry**, v. 53, n. 20, p. 3336–3346, 2014.
- ARAVAPALLI, S. et al. Inhibitors of Dengue virus and West Nile virus proteases based on the aminobenzamide scaffold. **Bioorganic and Medicinal Chemistry**, v. 20, n. 13, p. 4140–4148, jul. 2012.
- BABINE, R. E.; BENDER, S. L. Molecular Recognition of Protein - Ligand Complexes : Applications to Drug Design. **Chemical Reviews**, v. 97, n. 5, p. 1359–1472, 1997.
- BAELL, J. B.; HOLLOWAY, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. **Journal of Medicinal Chemistry**, v. 53, n. 7, p. 2719–2740, 2010.
- BAELL, J.; WALTERS, M. A. Chemistry: Chemical con artists foil drug discovery. **Nature**, v. 513, n. 7519, p. 481–483, 2014.
- BARLOW, D. J.; THORNTON, J. M. Ion-pairs in proteins. **Journal of Molecular Biology**, v. 168, n. 4, p. 867–885, ago. 1983.
- BARONI, M. et al. A common reference framework for analyzing/comparing proteins and ligands. Fingerprints for Ligands and Proteins (FLAP): Theory and application. **Journal of Chemical Information and Modeling**, v. 47, n. 2, p. 279–294, mar. 2007.
- BASTOS LIMA, A. et al. Dual inhibitors of the dengue and West Nile virus NS2B-NS3 proteases: Synthesis, biological evaluation and docking studies of novel peptide-hybrids. **Bioorganic and Medicinal Chemistry**, v. 23, n. 17, p. 5748–5755, 2015.
- BEHNAM, M. A. M. et al. C-terminal residue optimization and fragment merging: Discovery of a potent peptide-hybrid inhibitor of dengue protease. **ACS Medicinal Chemistry Letters**, v. 5, n. 9, p. 1037–1042, 2014.
- BEHNAM, M. A. M. et al. Discovery of Nanomolar Dengue and West Nile Virus Protease Inhibitors Containing a 4-Benzyloxyphenylglycine Residue. **Journal of Medicinal Chemistry**, v. 58, n. 23, p. 9354–9370, 2015.
- BEHNAM, M. A. M. et al. The Medicinal Chemistry of Dengue Virus. **Journal of Medicinal Chemistry**, v. 59, n. 12, p. 5622–5649, 23 jun. 2016.
- BEHNEN, J. et al. Experimental and Computational Active Site Mapping as a Starting Point to Fragment-Based Lead Discovery. **ChemMedChem**, v. 7, n. 2, p. 248–261, 2012.
- BENET, L. Z. et al. BDDCS, the Rule of 5 and drugability. **Advanced Drug Delivery Reviews**, v. 101, p. 89–98, 2016.
- BENTO, A. P. et al. The ChEMBL bioactivity database: An update. **Nucleic Acids Research**, v. 42, n. D1, p. D1083–D1090, jan. 2014.
- BERNSTEIN, F.; KOETZLE, T.; WILLIAMS, J. ET AL. The Protein Data Bank. **Eur. J. Biochem.**, v. 80, n. 1, p. 319–324, jan. 1977.
- BERTHOLD, M. R. et al. **KNIME: The Konstanz Information Miner**. Studies in Classification, Data Analysis, and Knowledge Organization (GfKL 2007). **Anais...** Springer, 2007
- BETZ, M. et al. Impact of Surface Water Layers on Protein-Ligand Binding: How Well Are Experimental Data Reproduced by Molecular Dynamics Simulations in a Thermolysin Test Case? **Journal of Chemical Information and Modeling**, v. 56, n. 1, p. 223–233, 2016.
- BHATT, S. et al. The global distribution and burden of dengue. **Nature**, v. 496, n. 7446, p. 504–7, 2013.
- BISSANTZ, C.; KUHN, B.; STAHL, M. A medicinal chemist's guide to molecular interactions. **Journal of Medicinal Chemistry**, v. 53, n. 14, p. 5061–5084, jul. 2010.
- BODENREIDER, C. et al. A fluorescence quenching assay to discriminate between specific and nonspecific inhibitors of dengue virus protease. **Analytical Biochemistry**, v. 395, n. 2, p. 195–204, 2009.
- BOLDESCU, V. et al. Broad-spectrum agents for flaviviral infections: dengue, Zika and beyond. **Nature Publishing Group**, v. 16, 2017.
- BOLOGA, C. G.; OLAH, M. M.; OPREA, T. I. Chemical Database Preparation for Compound Acquisition

- or Virtual Screening. In: **Bioinformatics and Drug Discovery**. Totowa, NJ: Humana Press, 2006. v. 316p. 375–388.
- BOLOGA, C. G.; OPREA, T. I. Compound Collection Preparation for Virtual Screening. In: LARSON, R. S. (Ed.). **Bioinformatics and Drug Discovery**. Totowa, NJ: Humana Press, 2012. p. 125–143.
- BOSTRÖM, J.; NORRBY, P. O.; LILJEFORS, T. Conformational energy penalties of protein-bound ligands. **Journal of computer-aided molecular design**, v. 12, n. 4, p. 383–396, 1998.
- BRIAN Y FENG et al. High-throughput assays for promiscuous inhibitors. **Nature Chemical Biology**, v. 1, n. 3, p. 146–148, ago. 2005.
- BRITO, B. P. et al. Review of the Global Distribution of Foot-and-Mouth Disease Virus from 2007 to 2014. **Transboundary and Emerging Diseases**, v. 64, n. 2, p. 316–332, abr. 2017.
- BROWN, F. K. et al. The evolution of drug design at Merck Research Laboratories. **Journal of Computer-Aided Molecular Design**, v. 31, n. 3, p. 255–266, 23 mar. 2017.
- BRUS, B. et al. Discovery, Biological Evaluation, and Crystal Structure of a Novel Nanomolar Selective Butyrylcholinesterase Inhibitor. **Journal of Medicinal Chemistry**, v. 57, n. 19, p. 8167–8179, 9 out. 2014.
- BUCKLE, D. R. et al. Glossary of Terms used in Medicinal Chemistry Part II. **Iupac**, n. 919, p. 1–42, 2014.
- CAO, Y. et al. ChemmineR: A compound mining framework for R. **Bioinformatics**, v. 24, n. 15, p. 1733–1734, 2008.
- CAPDEVILLE, R. et al. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. **Nature Reviews Drug Discovery**, v. 1, n. 7, p. 493–502, jul. 2002.
- CARLSON, H. A. et al. CSAR 2014: A Benchmark Exercise Using Unpublished Data from Pharma. **Journal of Chemical Information and Modeling**, v. 56, n. 6, p. 1063–1077, 27 jun. 2016.
- Catalyst. San Diego, CA, USA© Accelrys Software Inc., , 2005.
- CENCIC, R. et al. Investigating the Substrate Specificity and Oligomerisation of the Leader Protease of Foot and Mouth Disease Virus using NMR. **Journal of Molecular Biology**, v. 373, n. 4, p. 1071–1087, nov. 2007.
- CERETO-MASSAGUÉ, A. et al. Molecular fingerprint similarity search in virtual screening. **Methods**, v. 71, n. C, p. 58–63, 2015.
- CERUTTI, D. S. et al. Simulations of a protein crystal with a high resolution X-ray structure: Evaluation of force fields and water models. **Journal of Physical Chemistry B**, v. 114, n. 40, p. 12811–12824, 2010.
- CHANDRAMOULI, S. et al. Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. **J Virol**, v. 84, n. 6, p. 3059–3067, mar. 2010.
- CHEFFER, A. et al. Lipophilicity as a determinant of binding of procaine analogs to rat $\alpha 3\beta 4$ nicotinic acetylcholine receptor. **Journal of Neuroscience Research**, v. 90, n. 8, p. 1607–1614, ago. 2012.
- CHEMICALCOMPUTINGGROUPINC. **Molecular Operating Environment (MOE)Scientific Computing & Instrumentation**1010 Sherbooke St. West, Suite {#}910, Montreal, QC, Canada, H3A 2R7Chemical Computing Group Inc., , 2004.
- CHEN, W. N. et al. The dengue virus NS2B-NS3 protease retains the closed conformation in the complex with BPTI. **FEBS Letters**, v. 588, n. 14, p. 2206–2211, jun. 2014.
- CHEN, W. N. et al. Sensitive NMR Approach for Determining the Binding Mode of Tightly Binding Ligand Molecules to Protein Targets. **Journal of the American Chemical Society**, v. 138, n. 13, p. 4539–4546, 2016.
- CHEVILLARD, F. et al. In silico prediction of aqueous solubility: A multimodel protocol based on chemical similarity. **Molecular Pharmaceutics**, v. 9, n. 11, p. 3127–3135, 2012.
- CHOUDHURY, C.; PRIYAKUMAR, U. D.; SASTRY, G. N. Dynamics based pharmacophore models for screening potential inhibitors of mycobacterial cyclopropane synthase. **Journal of Chemical Information and Modeling**, v. 55, n. 4, p. 848–860, 2015.
- CIEPLAK, P. et al. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. **Journal of Computational Chemistry**, v. 16, n. 11, p. 1357–1377, nov. 1995.
- CIORDIA, M. et al. Application of Free Energy Perturbation for the Design of BACE1 Inhibitors. **Journal of Chemical Information and Modeling**, v. 56, n. 9, p. 1856–1871, 26 set. 2016.
- CLARK, D. E.; PICKETT, S. D. Computational methods for the prediction of “drug-likeness”. **Drug Discovery Today**, v. 5, n. 2, p. 49–58, fev. 2000.
- CLERCQ, E. DE. Antivirals and antiviral strategies. **Nature Reviews Microbiology**, v. 2, n. 9, p. 704–721, 2004.
- CONNOLLY, M. L. Solvent-accessible surfaces of proteins and nucleic acids. **Science (New York, N.Y.)**, v. 221, n. 4612, p. 709–713, 1983.
- COPELAND, R. A. Lead Optimization and Structure-Activity Relationships for Resersible Inhibitors. In: **Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists**. 1st. ed. [s.l.] John Wiley & Sons, 2005, 2005. p. 384.
- COPELAND, R. A. The drug–target residence time model: a 10-year retrospective. **Nature Reviews Drug Discovery**, v. 15, n. 2, p. 87–95, 18 dez. 2015.

- COZZINI, P. et al. Target flexibility: An emerging consideration in drug discovery and design. **Journal of Medicinal Chemistry**, v. 51, n. 20, p. 6237–6255, 2008.
- CREGAR-HERNANDEZ, L. et al. Small Molecule Pan-dengue and West Nile Virus NS3 Protease Inhibitors. **Antivir Chem Chemother.**, v. 21, n. 5, p. 209–218, 2012.
- CRUCIANI, G. et al. Molecular fields in quantitative structure-permeation relationships: The VolSurf approach. **Journal of Molecular Structure: THEOCHEM**, v. 503, n. 1–2, p. 17–30, 2000.
- CUMMING, J. G. et al. Chemical predictive modelling to improve compound quality. **Nature Reviews Drug Discovery**, v. 12, n. 12, p. 948–962, 29 nov. 2013.
- DABBOUSEH, N. M.; JENSEN, D. M. Future therapies for chronic hepatitis C. **Nature Reviews Gastroenterology & Hepatology**, v. 10, n. 5, p. 268–276, 19 fev. 2013.
- DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **J. Chem. Phys.**, v. 98, n. 12, p. 10089, 1993.
- DAYLIGHT INC. **SMARTS: A Language for Describing Molecular Patterns**. Disponível em: <<http://www.daylight.com/dayhtml/doc/theory/theory.smarts.html>>.
- DE ALMEIDA, H. et al. New Binding Site Conformations of the Dengue Virus NS3 Protease Accessed by Molecular Dynamics Simulation. **PLoS ONE**, v. 8, n. 8, p. e72402, jan. 2013.
- DE LA CRUZ, L. et al. Binding of low molecular weight inhibitors promotes large conformational changes in the dengue virus ns2b-ns3 protease: Fold analysis by pseudocontact shifts. **Journal of the American Chemical Society**, v. 133, n. 47, p. 19205–19215, nov. 2011.
- DE LA CRUZ, L. et al. Binding mode of the activity-modulating C-terminal segment of NS2B to NS3 in the dengue virus NS2B-NS3 protease. **FEBS Journal**, v. 281, n. 6, p. 1517–1533, mar. 2014.
- DE SOUSA, L. R. F. et al. Flavonoids as noncompetitive inhibitors of Dengue virus NS2B-NS3 protease: Inhibition kinetics and docking studies. **Bioorganic and Medicinal Chemistry**, v. 23, n. 3, p. 466–470, 2015.
- DE VIVO, M. et al. Role of Molecular Dynamics and Related Methods in Drug Discovery. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4035–4061, 2016.
- DENG, J. 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. 6278–6293, jul. 2012.
- DI MARTINO, G. P. et al. An Automated Docking Protocol for hERG Channel Blockers. **Journal of Chemical Information and Modeling**, v. 53, n. 1, p. 159–175, 28 jan. 2013.
- DIAZ-SAN SEGUNDO, F. et al. Foot-and-mouth disease vaccines. **Veterinary Microbiology**, v. 206, p. 102–112, jul. 2017.
- DOAK, B. C. et al. How Beyond Rule of 5 Drugs and Clinical Candidates Bind to Their Targets. **Journal of Medicinal Chemistry**, v. 59, n. 6, p. 2312–2327, 2016.
- DUNBAR, J. B. et al. CSAR data set release 2012: Ligands, affinities, complexes, and docking decoys. **Journal of Chemical Information and Modeling**, v. 53, n. 8, p. 1842–1852, 2013.
- ERBEL, P. 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. 372–373, 12 abr. 2006.
- ESWAR, N. et al. Comparative Protein Structure Modeling Using Modeller. **Curr protocols in bioinformatics**, v. 5, p. 5.6.1–5.6.30, nov. 2006.
- FALLER, B. et al. Evolution of the physicochemical properties of marketed drugs: Can history foretell the future? **Drug Discovery Today**, v. 16, n. 21–22, p. 976–984, 2011.
- FALSINI, M. et al. The 1,2,4-Triazolo[4,3-*b*]pyrazin-3-one as a Versatile Scaffold for the Design of Potent Adenosine Human Receptor Antagonists. Structural Investigations to Target the A_{2A} Receptor Subtype. **Journal of Medicinal Chemistry**, v. 60, n. 13, p. 5772–5790, 13 jul. 2017.
- FEHER, M.; SCHMIDT, J. M. Property distributions: Differences between drugs, natural products, and molecules from combinatorial chemistry. **Journal of Chemical Information and Computer Sciences**, v. 43, n. 1, p. 218–227, 2003.
- FERREIRA, R. S. et al. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. **Journal of Medicinal Chemistry**, v. 53, n. 13, p. 4891–4905, 2010.
- FILTER 2.1.1: OpenEye Scientific Software Inc., Santa Fe, NM.**, [s.d.]. Disponível em: <<http://www.eyesopen.com>>
- FLIPSE, J.; SMIT, J. M. The Complexity of a dengue vaccine: A review of the human antibody response. **PLoS Neglected Tropical Diseases**, v. 9, n. 6, p. 1–18, jun. 2015.
- 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**, v. 24, n. 3, p. 195–212, mar. 2010.
- FRISCH, M. J. et al. **Gaussian 09, Revision**, 2009.
- FUJITA, T. In memoriam professor Corwin Hansch: Birth pangs of QSAR before 1961. **Journal of Computer-Aided Molecular Design**, v. 25, n. 6, p. 509–517, 2011.
- FUJITA, T.; WINKLER, D. A. Understanding the Roles of the “two QSARs”. **Journal of Chemical**

Information and Modeling, v. 56, n. 2, p. 269–274, 2016.

GANESH, V. K. et al. Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. **Bioorganic and Medicinal Chemistry**, v. 13, n. 1, p. 257–264, jan. 2005.

GIBBS, A.; STEELE, R.; TOUNGE, B. NMR structure of an inhibitor bound dengue NS3 protease. 2014.

GLEESON, M. P. et al. Probing the links between in vitro potency, ADMET and physicochemical parameters. **Nature Reviews Drug Discovery**, v. 10, n. 3, p. 197–208, 1 mar. 2011.

GLEN, R. C. Connecting the virtual world of computers to the real world of medicinal chemistry. **Future medicinal chemistry**, v. 3, n. 4, p. 399–403, 2011.

GOHLKE, H.; HENDLICH, M.; KLEBE, G. Predicting binding modes, binding affinities and “hot spots” for protein-ligand complexes using a knowledge-based scoring function. **Perspectives in Drug Discovery and Design**, v. 20, n. 1, p. 115–144, 2000.

GOHLKE, H.; KLEBE, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. **Angewandte Chemie - International Edition**, v. 41, n. 15, p. 2644–2676, 2002.

GOLDBERG, F. W. et al. Designing novel building blocks is an overlooked strategy to improve compound quality. **Drug Discovery Today**, v. 20, n. 1, p. 11–17, jan. 2015.

GOODFORD, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. **Journal of Medicinal Chemistry**, v. 28, n. 7, p. 849–857, jul. 1985.

GOUVEA, I. E. 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. 187–196, jan. 2007.

GRUBMAN, M.; BAXT, B. Foot-and-mouth disease. **Clinical microbiology reviews**, v. 17, n. 2, p. 465–493, 2004.

GRUBMAN, M. J. et al. Evading the host immune response: how foot-and-mouth disease virus has become an effective pathogen. **FEMS immunology and medical microbiology**, v. 53, n. 1, p. 8–17, jun. 2008.

GUAN, S. H.; BELSHAM, G. J. Separation of foot-and-mouth disease virus leader protein activities; identification of mutants that retain efficient self-processing activity but poorly induce eIF4G cleavage. **Journal of General Virology**, v. 98, n. 4, p. 671–680, 2017.

GUARNE, A. Structure of the foot-and-mouth disease virus leader protease: a papain-like fold adapted for self-processing and eIF4G recognition. **The EMBO Journal**, v. 17, n. 24, p. 7469–7479, 15 dez. 1998.

GUARNE, A. et al. Structure of the foot-and-mouth disease virus leader protease: A papain-like fold adapted for self-processing and eIF4G recognition. **EMBO Journal**, v. 17, n. 24, p. 7469–7479, 15 dez. 1998.

GUARNÉ, A. et al. Structural and biochemical features distinguish the foot-and-mouth disease virus leader proteinase from other papain-like enzymes. **Journal of Molecular Biology**, v. 302, n. 5, p. 1227–40, 2000.

GUARNÉ, A. et al. Crystallization and preliminary X-ray diffraction studies of the Lb proteinase from foot-and-mouth disease virus. **Protein Science**, v. 5, n. 9, p. 1931–1933, set. 1996.

GUZMAN, M. G.; HARRIS, E. Dengue. **The Lancet**, v. 385, n. 9966, p. 453–465, 2015.

HAIDER, N. Functionality pattern matching as an efficient complementary structure/reaction search tool: An open-source approach. **Molecules**, v. 15, n. 8, p. 5079–5092, 2010.

HALGREN, T. A. et al. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. **Journal of Medicinal Chemistry**, v. 47, n. 7, p. 1750–1759, mar. 2004.

HANSCH, C.; STEWARD, A. R. The Use of Substituent Constants in the Analysis of the Structure-Activity Relationship in Penicillin Derivatives. **Journal of Medicinal Chemistry**, v. 7, n. 6, p. 691–694, nov. 1964.

HARVEY, A. L.; EDRADA-EBEL, R.; QUINN, R. J. The re-emergence of natural products for drug discovery in the genomics era. **Nature Reviews Drug Discovery**, v. 14, n. 2, p. 111–129, 2015.

HAWKINS, P. C. D. et al. How to do an evaluation: Pitfalls and traps. **Journal of Computer-Aided Molecular Design**, v. 22, n. 3–4, p. 179–190, 2008.

HAWKINS, P. C. D. et al. Conformer generation with OMEGA: Algorithm and validation using high quality structures from the protein databank and cambridge structural database. **Journal of Chemical Information and Modeling**, v. 50, n. 4, p. 572–584, 2010.

HAWKINS, P. C. D.; NICHOLLS, A. Conformer generation with OMEGA: Learning from the data set and the analysis of failures. **Journal of Chemical Information and Modeling**, v. 52, n. 11, p. 2919–2936, nov. 2012.

HAWKINS, P. C. D.; SKILLMAN, A. G.; NICHOLLS, A. Comparison of shape-matching and docking as virtual screening tools. **Journal of Medicinal Chemistry**, v. 50, n. 1, p. 74–82, 2007.

HERSEY, A. **ChEMBL20**. [s.l.: s.n.]. Disponível em: <ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_20>.

HILL, A. P.; YOUNG, R. J. Getting physical in drug discovery: A contemporary perspective on solubility and hydrophobicity. **Drug Discovery Today**, v. 15, n. 15–16, p. 648–655, ago. 2010.

- HODGSON, J. ADMET - turning chemicals into drugs. **Nature Biotechnology**, v. 19, n. 8, p. 722–726, ago. 2001.
- HOPKINS, A. L. et al. The role of ligand efficiency metrics in drug discovery. **Nature Reviews Drug Discovery**, v. 13, n. 2, p. 105–121, 31 jan. 2014.
- HOPKINS, A. L.; GROOM, C. R. The druggable genome. **Nature Reviews Drug Discovery**, v. 1, n. 9, p. 727–730, 2002.
- HU, B.; LILL, M. A. Exploring the potential of protein-based pharmacophore models in ligand pose prediction and ranking. **Journal of Chemical Information and Modeling**, v. 53, n. 5, p. 1179–1190, 2013.
- HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, fev. 1996.
- IMMING, P. Medicinal Chemistry: Definitions and Objectives, Drug Activity Phases, Drug Classification Systems. **The Practice of Medicinal Chemistry**, p. 63–72, 2008.
- INDEX, R.; DEVELOPMENT, T. R.; TEAM, C. R. **A Language and Environment for Statistical Computing** Foundation for Statistical Computing Vienna, Austria, 2009. Disponível em: <<http://www.r-project.org>>
- IRWIN, J. J. et al. ZINC: A free tool to discover chemistry for biology. **Journal of Chemical Information and Modeling**, v. 52, n. 7, p. 1757–1768, 2012.
- IRWIN, J. J. et al. An Aggregation Advisor for Ligand Discovery. **Journal of Medicinal Chemistry**, v. 58, n. 17, p. 7076–7087, 10 set. 2015.
- IRWIN, J. J.; SHOICHET, B. K. Docking Screens for Novel Ligands Conferring New Biology. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4103–4120, 2016.
- ISHIKI, H. € M.; AMARAL, A. € T. -D. Three-Dimensional Quantitative Structure-Activity Relationship Study of Antitumor 2-Formylpyridine Thiosemicarbazones Derivatives as Inhibitors of Ribonucleotide Reductase. **QSAR & Combinatorial Science**, v. 28, n. 11, p. 1334–1345, 2009.
- ISHIKI, H. M.; GALEMBECK, S. E. AMARAL, A. T. -D. **Application of two alignments procedures in CoMFA analysis of nevirapine derivative**. Rational Approaches To Drug Design, Barcelona,. **Anais...**2001
- JADHAV, A. et al. Quantitative analyses of aggregation, autofluorescence, and reactivity artifacts in a screen for inhibitors of a thiol protease. **Journal of Medicinal Chemistry**, v. 53, n. 1, p. 37–51, 2010.
- JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. **Journal of molecular biology**, v. 267, n. 3, p. 727–48, 1997.
- JONES, G.; WILLETT, P.; GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. **Journal of Molecular Biology**, v. 245, n. 1, p. 43–53, jan. 1995.
- JORGENSEN, W. L. et al. Comparison of simple potential functions for simulating liquid water. **The Journal of Chemical Physics**, v. 79, n. 2, p. 926, 1983.
- JÖST, C. et al. Promiscuity and selectivity in covalent enzyme inhibition: A systematic study of electrophilic fragments. **Journal of Medicinal Chemistry**, v. 57, n. 18, p. 7590–7599, 2014.
- KAWAKAMI, Y. et al. The rationale for E2020 as a potent acetylcholinesterase inhibitor. **Bioorganic & Medicinal Chemistry**, v. 4, n. 9, p. 1429–1446, set. 1996.
- KENNY, P. W.; MONTANARI, C. A. Inflation of correlation in the pursuit of drug-likeness. **Journal of Computer-Aided Molecular Design**, v. 27, n. 1, p. 1–13, 2013.
- KIAT, T. S. et al. Inhibitory activity of cyclohexenyl chalcone derivatives and flavonoids of fingerroot, *Boesenbergia rotunda* (L.), towards dengue-2 virus NS3 protease. **Bioorganic and Medicinal Chemistry Letters**, v. 16, n. 12, p. 3337–3340, jun. 2006.
- KIM, S. et al. PubChem substance and compound databases. **Nucleic Acids Research**, v. 44, n. D1, p. D1202–D1213, 4 jan. 2016.
- KIRCHMAIR, J. et al. The Protein Data Bank (PDB), Its Related Services and Software Tools as Key Components for In Silico Guided Drug Discovery. **Journal of Medicinal Chemistry**, v. 51, n. 22, p. 7021–7040, 27 nov. 2008.
- KIRCHMAIR, J. et al. How to optimize shape-based virtual screening: Choosing the right query and including chemical information. **Journal of Chemical Information and Modeling**, v. 49, n. 3, p. 678–692, 2009.
- KIRCHMAIR, J. et al. Predicting drug metabolism: experiment and/or computation? **Nature Reviews Drug Discovery**, v. 14, n. 6, p. 387–404, 2015.
- KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature Reviews Drug Discovery**, v. 3, n. 11, p. 935–949, nov. 2004.
- KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. **Drug Discovery Today**, v. 11, n. 13–14, p. 580–594, jul. 2006.
- KLEBE, G. Applying thermodynamic profiling in lead finding and optimization. **Nature Reviews Drug Discovery**, v. 14, n. 2, p. 95–110, 2015.
- KNEHANS, T. 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. 263–274, 23 mar. 2011.

KOCH, O. et al. Molecular Dynamics Reveal Binding Mode of Glutathionylspermidine by Trypanothione Synthetase. **PLoS ONE**, v. 8, n. 2, p. e56788, jan. 2013.

KOH-STENTA, X. et al. Identification of covalent active site inhibitors of dengue virus protease. **Drug Design, Development and Therapy**, v. 9, p. 6389–6399, 2015.

KORB, O. et al. Are predefined decoy sets of ligand poses able to quantify scoring function accuracy? **Journal of Computer-Aided Molecular Design**, v. 26, n. 2, p. 185–197, 2012.

KRZYWINSKI, M.; ALTMAN, N. Points of significance: Nonparametric tests. **Nature methods**, v. 11, n. 5, p. 467–469, 2014.

KUBINYI, H. Similarity and Dissimilarity: A Medicinal Chemist's View. **Perspectives in Drug Discovery and Design**, v. 9, n. 0, p. 225–252, 1998.

KUBINYI, H. Chance favors the prepared mind - from serendipity to rational drug design. **Journal of receptor and signal transduction research**, v. 19, n. 1–4, p. 15–39, 1999.

KUBINYI, H. Opinion: Drug research: myths, hype and reality. **Nature Reviews Drug Discovery**, v. 2, n. 8, p. 665–668, ago. 2003.

KUBINYI, H. In search of new leads. **Industrial Pharmacy**, n. 1, p. 7–10, 2004a.

KUBINYI, H. Changing paradigms in drug discovery. **The chemical theatre of biological systems: Proceedings of the international Beilstein workshop**, p. 51–72, 2004b.

KUBINYI, H. Success Stories of Computer-Aided Design. **Computer Applications in Pharmaceutical Research and Development**, v. 1, p. 377–424, 2006.

KUBINYI, H. The Changing Landscape in Drug Discovery. **Discovery**, p. 24–46, 2007.

KUHN, B. et al. A Real-World Perspective on Molecular Design. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4087–4102, 2016.

KUHNERT, M.; DIEDERICH, W. Structure-Based Drug Design in Medicinal Chemistry: The Devil is in the Detail. **Synlett**, v. 27, n. 5, p. 641–649, 4 jan. 2016.

LAI, H. et al. Design, synthesis and characterization of novel 1,2-benzisothiazol-3(2H)-one and 1,3,4-oxadiazole hybrid derivatives: Potent inhibitors of Dengue and West Nile virus NS2B/NS3 proteases. **Bioorganic and Medicinal Chemistry**, v. 21, n. 1, p. 102–113, jan. 2013.

LAI, H.; SRIDHAR PRASAD, G.; PADMANABHAN, R. Characterization of 8-hydroxyquinoline derivatives containing aminobenzothiazole as inhibitors of dengue virus type 2 protease in vitro. **Antiviral Research**, v. 97, n. 1, p. 74–80, jan. 2013.

LARSON, R. S. Bioinformatics and Drug Discovery Second. In: LARSON, R. S. (Ed.). **Methods in Molecular Biology**. Methods in Molecular Biology. Totowa, NJ: Humana Press, 2012. v. 531p. 588.

LEACH, A. R.; HANN, M. M. Molecular complexity and fragment-based drug discovery: Ten years on. **Current Opinion in Chemical Biology**, v. 15, n. 4, p. 489–496, 2011.

LESCAR, J. 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, 2008.

LIEBESCHUETZ, J. W.; COLE, J. C.; KORB, O. Pose prediction and virtual screening performance of GOLD scoring functions in a standardized test. **Journal of Computer-Aided Molecular Design**, v. 26, n. 6, p. 737–748, jun. 2012.

LIM, S. P. et al. Ten years of dengue drug discovery: Progress and prospects. **Antiviral Research**, v. 100, n. 2, p. 500–519, nov. 2013.

LIPINSKI, C. A. et al. 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, jan. 1997.

LIPINSKI, C. A. Drug-like properties and the causes of poor solubility and poor permeability. **Journal of Pharmacological and Toxicological Methods**, v. 44, n. 1, p. 235–249, 2000.

LIPINSKI, C. A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 46, n. 1–3, p. 3–26, mar. 2001.

LIPINSKI, C. A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 64, n. SUPPL., p. 4–17, dez. 2012.

LIPINSKI, C. A. Rule of five in 2015 and beyond: Target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. **Advanced Drug Delivery Reviews**, v. 101, p. 34–41, jun. 2016.

LIU, H. H. et al. Identification of novel thiadiazoloacrylamide analogues as inhibitors of dengue-2 virus NS2B/NS3 protease. **Bioorganic and Medicinal Chemistry**, v. 22, n. 22, p. 6344–6352, nov. 2014.

LUO, D. et al. Crystal structure of the NS3 protease-helicase from dengue virus. **Journal of virology**, v. 82, n. 1, p. 173–83, jan. 2008.

MAGALHÃES, M. N.; DE LIMA, A. C. P. Variáveis Aleatórias Contínuas. In: **Noções de Probabilidade e Estatística**. 7, 3a reim. ed. São Paulo: Editora da Universidade de São Paulo, 2015. p. 177–221.

MAIBAUM, J.; FELDMAN, D. L. Case History on Tekturna®/Rasilez® (Aliskiren), a Highly Efficacious

Direct Oral Renin Inhibitor as a New Therapy for Hypertension. In: **Annual Reports in Medicinal Chemistry**. [s.l.: s.n.]. v. 44p. 105–127.

MAKARA, G. M. On Sampling of Fragment Space. **Journal of Medicinal Chemistry**, v. 50, n. 14, p. 3214–3221, 2007.

MALVEZZI, A. **Modelos de virtual Screening de inibidores da cruzaina: desenvolvimento e validação experimental**. [s.l.] Universidade de São Paulo, São Paulo, 2008.

MALVEZZI, A. et al. Uncovering false positives on a virtual screening search for cruzain inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 18, n. 1, p. 350–354, jan. 2008.

MALVEZZI, A. et al. MPO Inhibitors Selected by Virtual Screening. **Molecular Informatics**, v. 30, n. 6–7, p. 605–613, jun. 2011.

MALVEZZI, A. et al. The Cysteine-Rich Protein Thimet Oligopeptidase as a Model of the Structural Requirements for S-glutathiolation and Oxidative Oligomerization. **PLoS ONE**, v. 7, n. 6, p. e39408, 25 jun. 2012.

MALVEZZI, A.; DE REZENDE, L.; AMARAL, A. T. -D. Pharmacophore Model of Cruzain Inhibitors. **QSAR & Combinatorial Science**, v. 28, n. 8, p. 781–784, ago. 2009.

MALVEZZI, A.; DO AMARAL, A. T. Ion pair stabilization effects on a series of procaine structural analogs. **European Journal of Pharmaceutical Sciences**, v. 41, n. 5, p. 631–635, 2010.

MANGLIK, A. et al. Structure-based discovery of opioid analgesics with reduced side effects. **Nature**, v. 537, n. 7619, p. 185–190, 17 ago. 2016.

MANNHOLD, R. et al. Use of MIF-based VolSurf Descriptors in Physicochemical and Pharmacokinetic Studies. In: CRUCIANI, G. (Ed.). **Molecular Interaction Fields: Applications in Drug Discovery and ADME Prediction**. Weinheim, FRG: Wiley-VCH Verlag GmbH & Co. KGaA, 2006. v. 27p. 171–196.

MARTIN, Y. C. Let's not forget tautomers. **Journal of Computer-Aided Molecular Design**, v. 23, n. 10, p. 693–704, 2009.

MarvinSketch. ChemAxon, , 2017.

MATTOS, C.; RINGE, D. Multiple binding modes. In: KUBINYI, H. (Ed.). **3D QSAR in drug design: Theory, methods and applications**. [s.l.] Springer Netherlands, 1993. p. 226–254.

MAURI, A. et al. Mixtures, metabolites, ionic liquids: a new measure to evaluate similarity between complex chemical systems. **Journal of Cheminformatics**, v. 8, n. 1, p. 1–3, 2016.

MAYEN, F. L. Foot and mouth disease in Brazil and its control - An overview of its history, present situation and perspectives for eradication. **Veterinary Research Communications**, v. 27, n. 2, p. 137–148, 2003.

MAYR, L. M.; BOJANIC, D. Novel trends in high-throughput screening. **Current Opinion in Pharmacology**, v. 9, n. 5, p. 580–588, 2009.

MCGANN, M. FRED pose prediction and virtual screening accuracy. **Journal of chemical information and modeling**, v. 51, n. 3, p. 578–596, mar. 2011.

MCGANN, M. FRED and HYBRID docking performance on standardized datasets. **Journal of Computer-Aided Molecular Design**, v. 26, n. 8, p. 897–906, 2012.

MCGANN, M.; NICHOLLS, A.; ENYEDY, I. The statistics of virtual screening and lead optimization. **Journal of Computer-Aided Molecular Design**, v. 29, n. 10, p. 923–936, 2015.

MCGAUGHEY, G.; PATRICK WALTERS, W. Modeling & Informatics at Vertex Pharmaceuticals Incorporated: our philosophy for sustained impact. **Journal of Computer-Aided Molecular Design**, v. 31, n. 3, p. 293–300, 29 mar. 2017.

MCGOVERN, S. L. et al. A Common Mechanism Underlying Promiscuous Inhibitors from Virtual and High-Throughput Screening. **Journal of Medicinal Chemistry**, v. 45, n. 8, p. 1712–1722, abr. 2002.

MCGOVERN, S. L. et al. A specific mechanism of nonspecific inhibition. **Journal of Medicinal Chemistry**, v. 46, n. 20, p. 4265–4272, 2003.

MEANWELL, N. A. Improving drug candidates by design: A focus on physicochemical properties as a means of improving compound disposition and safety. **Chemical Research in Toxicology**, v. 24, n. 9, p. 1420–1456, 2011.

MERGET, B.; SOTRIFFER, C. A. Slow-onset inhibition of Mycobacterium tuberculosis InhA: Revealing molecular determinants of residence time by MD simulations. **PLoS ONE**, v. 10, n. 5, p. 1–24, 2015.

MIGNANI, S. et al. Compound high-quality criteria: A new vision to guide the development of drugs, current situation. **Drug Discovery Today**, v. 21, n. 4, p. 573–584, 2016.

MILANI, P. et al. Synthesis of L-cysteine-based boron compounds and their evaluation as proteasome inhibitors. **New J. Chem. New J. Chem**, v. 38, n. 38, p. 4859–4871, 2014.

MILLETTI, F. et al. New and original pKa prediction method using grid molecular interaction fields. **Journal of Chemical Information and Modeling**, v. 47, n. 6, p. 2172–2181, 2007.

MINISTÉRIO DA AGRICULTURA. **Classificação de risco para febre aftosa e zona livre da doença - 2014 Programa Nacional de Erradicação e Prevenção da Febre Aftosa-PNEFA**, 2014. Disponível em: <[http://www.agricultura.gov.br/arq_editor/file/Classificação de risco_agosto_2014.pdf](http://www.agricultura.gov.br/arq_editor/file/Classificação%20de%20risco_agosto_2014.pdf)>. Acesso em: 5 abr. 2016

- MISRA, M. et al. Dissecting the Specificity of Adenosyl Sulfamate Inhibitors Targeting the Ubiquitin-Activating Enzyme. **Structure**, v. 25, n. 7, p. 1120–1129.e3, jul. 2017.
- MOUSTAKAS, D. T. et al. Development and validation of a modular, extensible docking program: DOCK 5. **Journal of Computer-Aided Molecular Design**, v. 20, n. 10–11, p. 601–619, 2006.
- MUEGGE, I.; BERGNER, A.; KRIEGL, J. M. Computer-aided drug design at Boehringer Ingelheim. **Journal of Computer-Aided Molecular Design**, v. 31, n. 3, p. 275–285, 20 mar. 2017.
- MUELLER, N. H. et al. Identification and biochemical characterization of small-molecule inhibitors of west nile virus serine protease by a high-throughput screen. **Antimicrobial Agents and Chemotherapy**, v. 52, n. 9, p. 3385–3393, 2008.
- MULLARD, A. New drugs cost US\$2.6 billion to develop. **Nature Reviews Drug Discovery**, v. 13, n. 12, p. 877–877, 1 dez. 2014.
- MURPHY, R. B. et al. WScore: A Flexible and Accurate Treatment of Explicit Water Molecules in Ligand-Receptor Docking. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4364–4384, 2016.
- NEWMAN, D. J.; CRAGG, G. M. Natural Products As Sources of New Drugs over the 30 Years from 1981 to 2010. **Journal of Natural Products**, v. 75, n. 3, p. 311–335, 23 mar. 2012.
- NGUYEN, C. N. et al. Thermodynamics of water in an enzyme active site: Grid-based hydration analysis of coagulation factor xa. **Journal of Chemical Theory and Computation**, v. 10, n. 7, p. 2769–2780, jul. 2014.
- NGUYEN, T. et al. In Vitro Evaluation of Novel Inhibitors against the NS2B-NS3 Protease of Dengue Fever Virus Type 4. **Molecules**, v. 18, n. 12, p. 15600–15612, 13 dez. 2013.
- NICHOLLS, A. et al. Molecular Shape and Medicinal Chemistry: A Perspective. **Journal of Medicinal Chemistry**, v. 53, n. 10, p. 3862–3886, 2010.
- NITSCHKE, C. et al. Retro peptide-hybrids as selective inhibitors of the Dengue virus NS2B-NS3 protease. **Antiviral Research**, v. 94, n. 1, p. 72–79, 2012.
- NITSCHKE, C. et al. Thiazolidinone-peptide hybrids as dengue virus protease inhibitors with antiviral activity in cell culture. **Journal of Medicinal Chemistry**, v. 56, n. 21, p. 8389–8403, 2013.
- NITSCHKE, C. et al. Biochemistry and medicinal chemistry of the dengue virus protease. **Chemical Reviews**, v. 114, n. 22, p. 11348–11381, 26 nov. 2014.
- NITSCHKE, C.; STEUER, C.; KLEIN, C. D. Arylcianoacrylamides as inhibitors of the Dengue and West Nile virus proteases. **Bioorganic and Medicinal Chemistry**, v. 19, n. 24, p. 7318–7337, 2011.
- NITTINGER, E. et al. Evidence of water molecules - A statistical evaluation of water molecules based on electron density. **Journal of Chemical Information and Modeling**, v. 55, n. 4, p. 771–783, 2015.
- NOBLE, C. G. et al. Strategies for development of dengue virus inhibitors. **Antiviral Research**, v. 85, n. 3, p. 450–462, 2010.
- NOBLE, C. G. et al. Ligand-Bound Structures of the Dengue Virus Protease Reveal the Active Conformation. **Journal of Virology**, v. 86, n. 1, p. 438–446, jan. 2012.
- NOBLE, C. G.; SHI, P. Y. Structural biology of dengue virus enzymes: Towards rational design of therapeutics. **Antiviral Research**, v. 96, n. 2, p. 115–126, nov. 2012.
- NOGUEIRA SANTOS, J. A. et al. Foot and mouth disease leader protease (Lbpro): Investigation of prime side specificity allows the synthesis of a potent inhibitor. **Biochimie**, v. 94, n. 3, p. 711–718, mar. 2012.
- NORDER, H. et al. Picornavirus non-structural proteins as targets for new anti-virals with broad activity. **Antiviral Research**, v. 89, n. 3, p. 204–218, 2011.
- NORMILE, D. Surprising New Dengue Virus Throws A Spanner in Disease Control Efforts. **Science**, v. 342, n. October, p. 2013, 2013.
- NORMILE, D. Dengue vaccine trial poses public health quandary. **Science**, v. 345, n. 6195, p. 367–368, 2014.
- O'BOYLE, N. M. et al. Open Babel: An Open chemical toolbox. **Journal of Cheminformatics**, v. 3, n. 10, p. 33, jan. 2011.
- O'NEIL, M. **The Merck Index, An Encyclopedia Of Chemicals, Drugs and Biologicals**. 12th. ed. Whitehouse Station, NJ: Merck Research Laboratories of Merck & CO., INC., 2006.
- OIE. **Map of FMD official status**. Disponível em: <<http://www.oie.int/en/animal-health-in-the-world/official-disease-status/fmd/en-fmd-carte/>>. Acesso em: 4 ago. 2017.
- OLAH, M. et al. WOMBAT: World of Molecular Bioactivity. In: OPREA, T. I. (Ed.). **Chemoinformatics in Drug Discovery**. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2005. p. 221–239.
- OPREA, T. I. et al. Is There a Difference between Leads and Drugs? A Historical Perspective. **Journal of Chemical Information and Modeling**, v. 41, n. 5, p. 1308–1315, 2001.
- OPREA, T. I. Virtual screening in lead discovery: A viewpoint. **Molecules**, v. 7, n. 1, p. 51–62, jan. 2002.
- OTTO, H.; SCHIRMEISTER, T. Cysteine Proteases and Their Inhibitors. **Chemical reviews**, v. 97, n. 1, p. 133–172, 1997.
- OWENS, P. K. et al. A decade of innovation in pharmaceutical R&D: the Chorus model. **Nature Reviews Drug Discovery**, v. 14, n. 1, p. 17–28, 2014.

- PARKINSON, T.; PRYDE, D. C. Small molecule drug discovery for Dengue and West Nile viruses: applying experience from hepatitis C virus. **Future medicinal chemistry**, v. 2, n. 7, p. 1181–203, jul. 2010.
- PASTOR, M.; CRUCIANI, G. A novel strategy for improving ligand selectivity in receptor-based drug design. **Journal of medicinal chemistry**, v. 38, n. 23, p. 4637–4647, nov. 1995.
- PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. **Journal of Computational Chemistry**, v. 26, n. 16, p. 1781–1802, 2005.
- PICCIRILLO, E. et al. Conformational flexibility of DENV NS2B/NS3pro: From the inhibitor effect to the serotype influence. **Journal of Computer-Aided Molecular Design**, v. 30, n. 3, p. 251–270, 29 mar. 2016.
- PIRES, J. R. et al. Investigation of 5-Nitrofurán Derivatives: Synthesis, Antibacterial Activity, and Quantitative Structure–Activity Relationships. **Journal of Medicinal Chemistry**, v. 44, n. 22, p. 3673–3681, out. 2001.
- PURCELL, W. P.; SINGER, J. A. A brief review and table of semiempirical parameters used in the Hueckel molecular orbital method. **Journal of Chemical & Engineering Data**, v. 12, n. 2, p. 235–246, 1967.
- R CORE TEAM. **R: A Language and Environment for Statistical Computing** Vienna, Austria, 2013.
- RAMINELLI, C. **Estudo da relação quantitativa entre a estrutura química e atividade citotóxica de séries de derivados de bases de Mannich**. [s.l.] Universidade de São Paulo, São Paulo, 2001.
- RAREY, M. et al. A fast flexible docking method using an incremental construction algorithm. **Journal of molecular biology**, v. 261, n. 3, p. 470–89, 1996.
- RAUT, R. et al. A small molecule inhibitor of dengue virus type 2 protease inhibits the replication of all four dengue virus serotypes in cell culture. **Virology Journal**, v. 12, n. 1, p. 16, 2015.
- RAVINA, E. **The Evolution of Drug Discovery: From Traditional Medicines to Modern Drugs**. [s.l.] John Wiley & Sons, 2011.
- RECHFELD, F. et al. Thienquinolines as Novel Disruptors of the PKC ϵ /RACK2 Protein–Protein Interaction. **Journal of Medicinal Chemistry**, v. 57, n. 8, p. 3235–3246, 24 abr. 2014.
- RISHTON, G. M. Reactive compounds and in vitro false positives in HTS. **Drug Discovery Today**, v. 2, n. 9, p. 382–384, 1997.
- RISHTON, G. M. Nonleadlikeness and leadlikeness in biochemical screening. **Drug Discovery Today**, v. 8, n. 2, p. 86–96, jan. 2003.
- RITZ, C.; STREIBIG, J. C. Bioassay Analysis using R. **Journal of Statistical Software**, v. 12, n. 5, p. 1–17, 2005.
- RITZ, C.; STREIBIG, J. C.; RITZ, C. & STREIBIG, J. C. Bioassay Analysis using R. **Journal of statistical software**, v. 12, n. 5, p. 1–17, 2006.
- ROCHE, O. et al. Development of a Virtual Screening Method for Identification of “Frequent Hitters” in Compound Libraries. **Journal of Medicinal Chemistry**, v. 45, n. 1, p. 137–142, jan. 2002.
- ROLLINGER, J. M.; STUPPNER, H.; LANGER, T. Virtual screening for the discovery of bioactive natural products. **Progress in drug research. Fortschritte der Arzneimittelforschung. Progres des recherches pharmaceutiques**, v. 65, p. 211, 213–49, 2008.
- RONDEAU, J.-M.; SCHREUDER, H. Protein Crystallography and Drug Discovery. In: WERMUTH, C. G. (Ed.). **The Practice of Medicinal Chemistry**. 3. ed. [s.l.] Elsevier, 2008. p. 605–634.
- SADOWSKI, J.; GASTEIGER, J. From atoms and bonds to 3-dimensional atomic coordinates - Automatic model builders. **Chemical Reviews**, v. 93, n. 7, p. 2567–2581, nov. 1993.
- SALOMON-FERRER, R.; CASE, D. A.; WALKER, R. C. An overview of the Amber biomolecular simulation package. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 3, n. 2, p. 198–210, mar. 2013.
- SANT’ANNA, C. M. R. Glossário de termos usados no planejamento de fármacos (recomendações da IUPAC para 1997). **Química Nova**, v. 25, n. 3, p. 505–512, maio 2002.
- SANTOS, J. A. N. et al. Hydrolytic properties and substrate specificity of the foot-and-mouth disease leader protease. **Biochemistry**, v. 48, n. 33, p. 7948–7958, 25 ago. 2009.
- SCHAEFER, B.; KISKER, C.; SOTRIFFER, C. A. Molecular dynamics of Mycobacterium tuberculosis KasA: Implications for inhibitor and substrate binding and consequences for drug design. **Journal of Computer-Aided Molecular Design**, v. 25, n. 11, p. 1053–1069, nov. 2011.
- SCHAEFFER, L. The Role of Functional Groups in Drug – Receptor Interactions. In: WERMUTH, C. G. (Ed.). **The Practice of Medicinal Chemistry**. 3. ed. [s.l.] Elsevier Ltd, 2008. p. 464–480.
- SCHOMBURG, K. T.; RAREY, M. Benchmark data sets for structure-based computational target prediction. **Journal of Chemical Information and Modeling**, v. 54, n. 8, p. 2261–2274, 2014.
- SCHRÖDINGER, L. **The PyMOL Molecular Graphics System**, 2013.
- SCHÜLLER, A. et al. Tripeptide inhibitors of dengue and West Nile virus NS2B-NS3 protease. **Antiviral Research**, v. 92, n. 1, p. 96–101, 2011.
- SCHULZ-GASCH, T.; STAHL, M. Binding site characteristics in structure-based virtual screening: Evaluation of current docking tools. **Journal of Molecular Modeling**, v. 9, n. 1, p. 47–57, 2003.

- SCHULZE, U. et al. R&D productivity: on the comeback trail. **Nature Reviews Drug Discovery**, v. 13, n. 5, p. 331–332, 22 abr. 2014.
- Scientific American Blog**. Disponível em: <<https://blogs.scientificamerican.com/%0A>>. Acesso em: 4 ago. 2016.
- SCIOR, T. et al. Recognizing pitfalls in virtual screening: A critical review. **Journal of Chemical Information and Modeling**, v. 52, n. 4, p. 867–881, 23 abr. 2012.
- SCOTTI, M. T. et al. Variable-selection approaches to generate QSAR models for a set of antichagasic semicarbazones and analogues. **Chemometrics and Intelligent Laboratory Systems**, v. 154, p. 137–149, 2016.
- SHIRYAEV, S. A et al. 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, 2011.
- SIPPL, W. Pharmacophore Identification and Pseudo-Receptor Modeling. In: **The Practice of Medicinal Chemistry**. [s.l: s.n.]. p. 572–586.
- SMITH, R. D. et al. CSAR Benchmark Exercise 2013: Evaluation of Results from a Combined Computational Protein Design, Docking, and Scoring/Ranking Challenge. **Journal of Chemical Information and Modeling**, v. 56, n. 6, p. 1022–1031, 2016.
- SOTRIFER, C. A. Accounting for induced-fit effects in docking: what is possible and what is not? **Current topics in medicinal chemistry**, v. 11, n. 2, p. 179–191, 2011.
- SPASIC, A.; SERAFINI, J.; MATHEWS, D. H. The Amber ff99 force field predict relative free energy changes for RNA helix formation. **Journal of Chemical Theory and Computation**, v. 8, n. 7, p. 2497–2505, 2013.
- SPRENT, P. **Applied Nonparametric Statistical Methods**. 1. ed. Dordrecht: Springer Netherlands, 1988.
- SRINIVASAN, J. et al. Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. **Theoretical Chemistry Accounts**, v. 101, n. 6, p. 426–434, 1999.
- STANK, A. et al. Protein Binding Pocket Dynamics. **Accounts of Chemical Research**, v. 49, n. 5, p. 809–815, 2016.
- STEINBERGER, J. et al. Foot-and-mouth disease virus leader proteinase: Structural insights into the mechanism of intermolecular cleavage. **Virology**, v. 468, p. 397–408, nov. 2014.
- STEINBRECHER, T. et al. Predicting the Effect of Amino Acid Single-Point Mutations on Protein Stability—Large-Scale Validation of MD-Based Relative Free Energy Calculations. **Journal of Molecular Biology**, v. 429, n. 7, p. 948–963, abr. 2017.
- STEINBRECHER, T. B. et al. Accurate Binding Free Energy Predictions in Fragment Optimization. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2411–2420, 23 nov. 2015.
- STEUER, C. et al. Synthesis and biological evaluation of α -ketoamides as inhibitors of the Dengue virus protease with antiviral activity in cell-culture. **Bioorganic and Medicinal Chemistry**, v. 19, n. 13, p. 4067–4074, jul. 2011.
- STILL, W. C. et al. Semianalytical treatment of solvation for molecular mechanics and dynamics. **Journal of the American Chemical Society**, v. 112, n. 16, p. 6127–6129, 1990.
- STUMPFE, D. et al. Recent progress in understanding activity cliffs and their utility in medicinal chemistry. **Journal of Medicinal Chemistry**, v. 57, n. 1, p. 18–28, 2014.
- STUMPFE, D.; BAJORATH, J. Similarity searching. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 1, n. 2, p. 260–282, mar. 2011.
- SWINNEY, D. C.; ANTHONY, J. How were new medicines discovered? **Nature Reviews Drug Discovery**, v. 10, n. 7, p. 507–519, 2011.
- SYBYL. SYBYL1699 South Hanley Rd., St. Louis, Missouri, 63144, USATripes International, , 2009.
- SYBYL. SYBYL-X1699 South Hanley Rd., St. Louis, Missouri, 63144, USATripes, a Certara Companie, , 2013.
- TAYLOR, W. R. Residual colours: a proposal for aminochromography. **Protein engineering**, v. 10, n. 7, p. 743–746, 1997.
- TEAGUE et al. The Design of Leadlike Combinatorial Libraries. **Angewandte Chemie (International ed. in English)**, v. 38, n. 24, p. 3743–3748, 16 dez. 1999.
- TETKO, I. V. et al. Virtual computational chemistry laboratory - Design and description. **Journal of Computer-Aided Molecular Design**, v. 19, n. 6, p. 453–463, jun. 2005.
- TIEW, K. C. et al. Inhibition of dengue virus and west Nile virus proteases by click chemistry-derived benz[d]isothiazol-3(2H)-one derivatives. **Bioorganic and Medicinal Chemistry**, v. 20, n. 3, p. 1213–1221, fev. 2012.
- TIMIRI, A. K.; SINHA, B. N.; JAYAPRAKASH, V. Progress and prospects on DENV protease inhibitors. **European Journal of Medicinal Chemistry**, v. 117, p. 125–143, 2016.
- TOMLINSON, S. M. et al. Structure-based discovery of dengue virus protease inhibitors. **Antiviral Research**, v. 82, n. 3, p. 110–114, jun. 2009.

TOMLINSON, S. M.; WATOWICH, S. J. Anthracene-based inhibitors of dengue virus NS2B-NS3 protease. **Antiviral Research**, v. 89, n. 2, p. 127–135, 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. 245–252, 2012.

TSE, M. T. All-oral HCV therapies near approval. **Nature reviews Drug discovery**, v. 12, n. 6, p. 409–11, 2013.

TURK, B. Targeting proteases: successes, failures and future prospects. **Nature Reviews Drug Discovery**, v. 5, n. 9, p. 785–799, set. 2006.

U.S. Food and Drug Administration. Disponível em: <<https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm>>. Acesso em: 1 maio. 2017.

USP - United States Pharmacopeial Convention. Disponível em: <<http://www.usp.org/>>. Acesso em: 1 maio. 2017.

VARMUZA, K.; PETER, F. Multivariate Data 2.1. In: GROUP, T. & F. (Ed.). **Introduction to Multivariate Statistical Analysis in Chemometrics**. 1. ed. [s.l.] CRC Press, 2008.

VEBER, D. F. et al. Molecular properties that influence the oral bioavailability of drug candidates. **Journal of Medicinal Chemistry**, v. 45, n. 12, p. 2615–2623, jun. 2002.

VERDONK, M. L. et al. Improved protein-ligand docking using GOLD. **Proteins: Structure, Function, and Bioinformatics**, v. 52, n. 4, p. 609–623, 1 ago. 2003.

VIEGAS, C.; DA SILVA BOLZANI, V.; BARREIRO, E. J. OS produtos naturais e a química medicinal moderna. **Química Nova**, v. 29, n. 2, p. 326–337, 2006.

VISWANATHAN, U. et al. Identification of a Novel Inhibitor of Dengue Virus Protease through Use of a Virtual Screening Drug Discovery Web Portal. **Journal of Chemical Information and Modeling**, v. 54, n. 10, p. 2816–2825, 2014.

VOLKMER, A. et al. Combining global and local measures for structure-based druggability predictions. **Journal of Chemical Information and Modeling**, v. 52, n. 2, p. 360–372, 2012.

VON ITZSTEIN, M. et al. Rational design of potent sialidase-based inhibitors of influenza virus replication. **Nature**, v. 363, n. 6428, p. 418–423, 3 jun. 1993.

WADE, R. C. “Flu” and structure-based drug design. **Structure**, v. 5, p. 1139–1145, 1997.

WANG, R. et al. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. **Journal of Medicinal Chemistry**, v. 47, n. 12, p. 2977–2980, jun. 2004.

WARING, M. J. et al. An analysis of the attrition of drug candidates from four major pharmaceutical companies. **Nature Reviews Drug Discovery**, v. 14, n. 7, p. 475–486, 19 jun. 2015.

WARREN, G. L. et al. Essential considerations for using protein-ligand structures in drug discovery. **Drug Discovery Today**, v. 17, n. 23–24, p. 1270–1281, 2012.

WASSERMANN, A. M. et al. Dark chemical matter as a promising starting point for drug lead discovery. **Nature chemical biology**, v. 11, n. 12, p. 958–966, 2015.

WEBER, A. E.; THORNBERRY, N. Case History: JANUVIATM (Sitagliptin), a Selective Dipeptidyl Peptidase IV Inhibitor for the Treatment of Type 2 Diabetes. In: **Annual Reports in Medicinal Chemistry**. [s.l.: s.n.]. v. 42p. 95–109.

WENLOCK, M. C. et al. A comparison of physicochemical property profiles of development and marketed oral drugs. **Journal of Medicinal Chemistry**, v. 46, n. 7, p. 1250–1256, 2003.

WENLOCK, M. C.; BARTON, P. In silico physicochemical parameter predictions. **Molecular Pharmaceutics**, v. 10, n. 4, p. 1224–1235, abr. 2013.

WHITEHORN, J.; SIMMONS, C. P. The pathogenesis of dengue. **Vaccine**, v. 29, n. 42, p. 7221–7228, set. 2011.

WICHAPONG, K. et al. 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.

WIEDER, M. et al. Common Hits Approach: Combining Pharmacophore Modeling and Molecular Dynamics Simulations. **Journal of Chemical Information and Modeling**, v. 57, n. 2, p. 365–385, 2017.

WILLET, P. Similarity-based data mining in files of two-dimensional chemical structures using fingerprint measures of molecular resemblance. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 1, n. 3, p. 241–251, 2011.

WISHART, D. S. DrugBank: a comprehensive resource for in silico drug discovery and exploration. **Nucleic Acids Research**, v. 34, n. 90001, p. D668–D672, 1 jan. 2006.

WOLBER, G.; LANGER, T. LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 160–169, 2005.

WOLD, S. **Data analysis for chemists, applications to QSAR and chemical product design**. [s.l.] Oxford science publications, 1997. v. 38

WU, H. et al. Novel dengue virus NS2B/NS3 protease inhibitors. **Antimicrobial Agents and Chemotherapy**, v. 59, n. 2, p. 1100–1109, 2015.

WYLES, D. L. Antiviral resistance and the future landscape of hepatitis C virus infection therapy. **Journal of Infectious Diseases**, v. 207, n. SUPPL.1, p. 33–39, 2013.

YANG, C. C. et al. Novel dengue virus-specific NS2B/NS3 protease inhibitor, BP2109, discovered by a high-throughput screening assay. **Antimicrobial Agents and Chemotherapy**, v. 55, n. 1, p. 229–238, 2011.

YILDIZ, M. et al. Allosteric inhibition of the NS2B-NS3 protease from dengue virus. **ACS Chemical Biology**, v. 8, n. 12, p. 2744–2752, 2013.

YIN, Z. et al. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 16, n. 1, p. 40–43, jan. 2006.

ZHOU, G.-C. et al. Discovery and SAR studies of methionine–proline anilides as dengue virus NS2B-NS3 protease inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 23, n. 24, p. 6549–6554, dez. 2013.

ANEXO I

Cr terios “Drug” do programa *Filter*, usado para construir o banco *ZINC-Curated* mostrando em negrito os valores *default* que foram, por n s, alterados.

MIN_MOLWT	200	"Minimum molecular weight"
MAX_MOLWT	600	"Maximum molecular weight"
MIN_NUM_HVY	15	"Minimum number of heavy atoms"
MAX_NUM_HVY	35	"Maximum number of heavy atoms"
MIN_RING_SYS	0	"Minimum number of ring systems"
MAX_RING_SYS	5	"Maximum number of ring systems"
MIN_RING_SIZE	0	"Minimum atoms in any ring system"
MAX_RING_SIZE	20	"Maximum atoms in any ring system"
MIN_CON_NON_RING	0	"Minimum number of connected non-ring atoms"
MAX_CON_NON_RING	15	"Maximum number of connected non-ring atoms"
MIN_FCNGRP	0	"Minimum number of functional groups"
MAX_FCNGRP	18	"Maximum number of functional groups"
MIN_UNBRANCHED	0	"Minimum number of connected unbranched non-ring atoms"
MAX_UNBRANCHED	6	"Maximum number of connected unbranched non-ring atoms"
MIN_CARBONS	7	"Minimum number of carbons"
MAX_CARBONS	35	"Maximum number of carbons"
MIN_HETEROATOMS	2	"Minimum number of heteroatoms"
MAX_HETEROATOMS	20	"Maximum number of heteroatoms"
MIN_Het_C_Ratio	0.10	"Minimum heteroatom to carbon ratio"
MAX_Het_C_Ratio	1.0	"Maximum heteroatom to carbon ratio"
MIN_HALIDE_FRACTION	0.0	"Minimum Halide Fraction"
MAX_HALIDE_FRACTION	0.5	"Maximum Halide Fraction"
ADJUST_ROT_FOR_RING	true	"BOOLEAN for whether to estimate degrees of freedom in rings"
MIN_ROT_BONDS	0	"Minimum number of rotatable bonds"
MAX_ROT_BONDS	20	"Maximum number of rotatable bonds"
MIN_RIGID_BONDS	0	"Minimum number of rigid bonds"
MAX_RIGID_BONDS	35	"Maximum number of rigid bonds"
MIN_HBOND_DONORS	0	"Minimum number of hydrogen-bond donors"
MAX_HBOND_DONORS	6	"Maximum number of hydrogen-bond donors"
MIN_HBOND_ACCEPTORS	0	"Minimum number of hydrogen-bond acceptors"
MAX_HBOND_ACCEPTORS	8	"Maximum number of hydrogen-bond acceptors"
MIN_LIPINSKI_DONORS	0	"Minimum number of hydrogens on O & N atoms"
MAX_LIPINSKI_DONORS	5	"Maximum number of hydrogens on O & N atoms"
MIN_LIPINSKI_ACCEPTORS	0	"Minimum number of oxygen & nitrogen atoms"
MAX_LIPINSKI_ACCEPTORS	10	"Maximum number of oxygen & nitrogen atoms"
MIN_COUNT_FORMAL_CRG	0	"Minimum number formal charges"
MAX_COUNT_FORMAL_CRG	3	"Maximum number of formal charges"
MIN_SUM_FORMAL_CRG	-2	"Minimum sum of formal charges"
MAX_SUM_FORMAL_CRG	2	"Maximum sum of formal charges"
MIN_CHIRAL_CENTERS	0	"Minimum chiral centers"
MAX_CHIRAL_CENTERS	4	"Maximum chiral centers"
MIN_XLOGP	-5.0	"Minimum XLogP"
MAX_XLOGP	6.0	"Maximum XLogP"
MIN_SOLUBILITY	insoluble	"Minimum solubility"
PSA_USE_SandP	false	"Count S and P as polar atoms"
MIN_2D_PSA	0.0	"Minimum 2-Dimensional (SMILES) Polar Surface Area"
MAX_2D_PSA	160.0	"Maximum 2-Dimensional (SMILES) Polar Surface Area"
AGGREGATORS	false	"Eliminate known aggregators"
PRED_AGG	false	"Eliminate predicted aggregators"
GSK_VEBER	false	"PSA>140 or >10 rot bonds"
MAX_LIPINSKI	2	"Maximum number of Lipinski violations"
MIN_ABS	0.0	"Minimum probability F>10% in rats"
PHARMACOPIA	false	"LogP > 5.88 or PSA > 131.6"
ALLOWED_ELEMENTS	H,C,N,O,F,S,Cl,Br	
ELIMINATE_METALS	Sc,Ti,V,Cr,Mn,Fe,Co,Ni,Cu,Zn,Y,Zr,Nb,Mo,Tc,Ru,Rh,Pd,Ag,Cd	

#functional groups removed

RULE 0 quinone
 RULE 0 pentafluorophenyl_esters
 RULE 0 paranitrophenyl_esters
 RULE 0 HOBT_esters
 RULE 0 triflates
 RULE 0 lawesson_s_reagent
 RULE 0 phosphoramides
 RULE 0 beta_carbonyl_quat_nitrogen
 RULE 0 acylhydrazide
 RULE 0 cation_C_Cl_I_P_or_S
 RULE 0 phosphoryl
 RULE 0 alkyl_phosphate
 RULE 0 phosphinic_acid
 RULE 0 phosphanes
 RULE 0 phosphoranes
 RULE 0 imidoyl_chlorides
 RULE 0 nitroso
 RULE 0 N_P_S_Halides
 RULE 0 carbodiimide
 RULE 0 isonitrile
 RULE 0 triacyloxime
 RULE 0 cyanohydrins
 RULE 0 acyl_cyanides
 RULE 0 sulfonylnitrile
 RULE 0 phosphonylnitrile
 RULE 0 azocyanamides
 RULE 0 beta_azo_carbonyl
 RULE 0 polyenes
 RULE 0 saponin_derivatives
 RULE 0 acid_halide
 RULE 0 aldehyde
 RULE 0 alkyl_halide
 RULE 0 anhydride
 RULE 0 azide
 RULE 0 azo
 RULE 0 di_peptide
 RULE 0 michael_acceptor
 RULE 0 beta_halo_carbonyl
 RULE 0 nitro
 RULE 0 oxygen_cation
 RULE 0 peroxide
 RULE 0 phosphonic_acid
 RULE 0 phosphonic_ester
 RULE 0 phosphoric_acid
 RULE 0 phosphoric_ester
 RULE 0 sulfonic_acid
 RULE 0 sulfonic_ester
 RULE 0 tricarbo_phosphene
 RULE 0 epoxide
 RULE 0 sulfonyl_halide
 RULE 0 halopyrimidine
 RULE 0 perhalo_ketone
 RULE 0 aziridine
 RULE 0 alphahalo_amine
 RULE 0 halo_amine
 RULE 0 halo_alkene
 RULE 0 acyclic_NCN
 RULE 0 acyclic_NS
 RULE 0 SCN2

RULE 0 terminal_vinyl
 RULE 0 hetero_hetero
 RULE 0 hydrazine
 RULE 0 N_methoyl
 RULE 0 NS_beta_halothyl
 RULE 0 propiolactones
 RULE 0 iodoso
 RULE 0 iodoxy
 RULE 0 noxide
 RULE 0 phosphonamide
 RULE 0 alphahalo_ketone
 RULE 0 oxaziridine
 RULE 0 sulfonimine
 RULE 0 sulfinimine
 RULE 0 sulfinylthio
 RULE 0 disulfide
 RULE 0 enol_ether
 RULE 0 enamine
 RULE 0 organometallic
 RULE 0 dithioacetal
 RULE 0 isothiocyanate
 RULE 0 isocyanate
 RULE 0 carbamic_acid
 RULE 0 triazine
 RULE 0 nonacylhydrazone
 RULE 0 thiourea
 RULE 0 hemiketal
 RULE 0 hemiacetal
 RULE 0 ketal
 RULE 0 amination
 RULE 0 hemiaminal
 RULE 0 benzyloxycarbonyl_CBZ
 RULE 0 t_butoxycarbonyl_tBOC
 RULE 0 fluorenylmethoxycarbonyl_Fmoc
 RULE 0 trimethylsilyl_TMS
 RULE 0 t_butyldimethylsilyl_TBDMS
 RULE 0 triisopropylsilyl_TIPS
 RULE 0 t_butyldiphenylsilyl_TBDPS
 # functional groups reduced
 RULE 1 cytochalasin_derivatives
 RULE 4 cycloheximide_derivatives
 RULE 1 monensin_derivatives
 RULE 1 squalastatin_derivatives
 RULE 1 oxalyl
 RULE 6 alcohol
 RULE 4 alkene
 RULE 4 amide
 RULE 4 amino_acid
 RULE 2 amine
 RULE 4 primary_amine
 RULE 4 secondary_amine
 RULE 4 tertiary_amine
 RULE 2 carboxylic_acid
 RULE 6 halide
 RULE 0 iodine
 RULE 2 ketone
 RULE 4 phenol
 RULE 1 imine
 RULE 1 methyl_ketone

RULE 1 alkylaniline
RULE 4 sulfonamide
RULE 1 sulfonylurea
RULE 1 cyclopropyl
RULE 2 guanidine
RULE 1 hydroxamic_acid
RULE 1 oxime
RULE 3 lactone
RULE 3 lactam
RULE 1 thioester
RULE 1 carbonate
RULE 1 thiocarbamate
RULE 1 malonic
RULE 2 alkyne
RULE 4 aniline
RULE 4 aryl_halide
RULE 2 carbamate
RULE 3 ester
RULE 2 arenesulfonyl

RULE 5 ether
RULE 1 hydrazone
RULE 1 hydroxylamine
RULE 2 nitrile
RULE 2 sulfide
RULE 2 sulfone
RULE 2 sulfoxide
RULE 1 thioamide
RULE 1 thiol
RULE 2 urea
RULE 1 acetal
RULE 1 dioxolane_5MR
RULE 1 dioxane_6MR
RULE 1 tetrahydropyran_THP
RULE 1 methoxyethoxymethyl_MEM
RULE 2 benzyl_ether
RULE 2 t_butyl_ether
RULE 1 phthalimides_PHT

ANEXO II

Lista dos 58 descritores do *Volsurf+* usados na ACP

Sigla	Descritor
V	Molecular Volume
S	Molecular surface
R	Rugosity
G	Molecular Globularity
Flex, Flex_RB	Flexibility parameters
W1	Hydrophilic region volume at -0.2 kcal/mol
W8	Hydrophilic region volume at -6.0 kcal/mol
CW1	Capacity factor at W1 volume
CW8	Capacity factor at W8 volume
D1	Hydrophobic region volume at -0.2 kcal/mol
D8	Hydrophobic region volume at -6.0 kcal/mol
CD1	Capacity factor at D1 volume
CD8	Capacity factor at D1 volume
DD1	Difference of the hydrophobic at D1
DD8	Difference of the hydrophobic at D8
IW1	Interaction energy moment at W1
IW4	Interaction energy moment at W4
ID1	Interaction energy moment at D1
ID4	Interaction energy moment at D1
WO1	H-bond donor volume at -1 kcal/mol
WO6	H-bond donor volume at -6 kcal/mol
WN1	H-bond acceptor volume at -1 kcal/mol
WN6	H-bond acceptor volume at -6 kcal/mol
HL1	Hydrophilic/lipophilic balance (low energy)
HL2	Hydrophilic/lipophilic balance (high energy)
A	Amphiphilic moment
CP	Critical packing
POL	Polarizability
DIFF	Diffusivity
MW	Molecular Weight
LOGP n-Oct	Logarithm of the partition coefficient between 1-octanol and water
LOGP c-Hex	Logarithm of the partition coefficient between cyclohexane and water

Sigla	Descriptor
LgD7.5	Logarithm of the partition coefficient between 1-octanol and water at pH 7.5
PSA	Polar Surface Area
HSA	Hydrophobic Surface Area
PSAR	Polar and Surface Area Ratio
PHSAR	Polar and Hydrophobic Surface Area Ratio
NCC	Number of Charged Centers
AUS7.4	Available Uncharged Species at pH 7.4
%FU7	Percentage of unionised species at pH 7.
DRDRDR	Dry-Dry-Dry triplet
DRDRAC	Dry-Dry-Acceptor triplet
DRDRDO	Dry-Dry-Donor triplet
DRACAC	Dry-Acceptor-Acceptor triplet
DRACDO	Dry-Acceptor-Donor triplet
DRDODO	Dry-Donor-Donor triplet
ACACAC	Acceptor-Acceptor-Acceptor triplet
ACACDO	Acceptor-Acceptor-Donor triplet
ACDODO	Acceptor-Donor-Donor triplet
DODODO	Donor-Donor-Donor triplet
SOLY	Intrinsic solubility in mol/L at 25 °C
LgS7.5	Solubility at pH 7.5
CACO2	CACO2 cells permeability
SKIN	Skin permeability
PB	Percentage of protein binding
LgBB	Log blood-brain barrier distribution
MetStab	Metabolic Stability
VD	Volume of distribution
HTSflag	High throughput screening flag

ANEXO III

Definições estatísticas e equações matemáticas

Agrupamento hierárquico e dendograma

Em cada passo dos agrupamentos, os compostos com menor valor de dissimilaridade, ou seja, mais semelhantes, foram combinados resultando em dendrogramas horizontais (estruturas em árvore). Cada dendrograma obtido apresenta uma variedade de soluções de agrupamento. Assim, diferentes números de grupos podem ser gerados, dependendo do valor de corte aplicado para considerá-los dissimilar/similar.

Auto escalonamento

No auto escalonamento, o valor da média é subtraído dos valores das variáveis e os valores resultantes são divididos pelo desvio padrão. Este é expresso pela equação:

$$X'_{ij} = \frac{X_{ij} - \bar{X}_j}{S_j}$$

onde, X'_{ij} é o novo valor escalonado da observação i da variável j ; \bar{X}_j é o valor da média da variável j ; S_j é o valor do desvio padrão da variável j . Sendo que as observações i e as variáveis j representam os compostos e os descritores, respectivamente.

Curva dose-resposta usada para determinar os valores de IC₅₀

A função LL.2 do pacote “drc” (RITZ; STREIBIG, 2005) disponível no programa R (R CORE TEAM, 2013) foi usada para construir a curva dose-resposta, ou seja, gráfico dos valores da fração de enzima ativa (v_i/v_0) em função do logaritmo da concentração do composto, sendo v_i a velocidade inicial na presença do composto e v_0 a velocidade inicial na ausência do composto. Nestas curvas, a equação de reta de melhor ajuste está expressa pela equação:

$$f(x) = \frac{1}{1 + \exp(b(\log(x) - \log(e)))}$$

onde, $f(x)$ é a resposta/fração de enzima ativa (v_i/v_0), x é a dose (concentração do composto), e é o valor do IC₅₀ (dose responsável por produzir metade da resposta) e b é a inclinação relativa ao redor de e .

Distância Euclidiana

$$D_{eucl} = \sqrt{\sum_i^n (x_i - y_i)^2}$$

onde, x_i e y_i valor da variável i para os compostos x e y e n é o número total de variáveis.

Gráficos de caixa

Definição: Gráficos de caixa tipo *Tukey* são gráficos no qual as caixas indicam a distância interquartil ($Q_1 - Q_3$), a linha preta mostra o valor da mediana (Q_2), os *whiskers* (fio de bigode) são segmentos que estendem 1,5 vezes a distancia interquartil e os círculos representam os valores atípicos (*outliers*).

Gráficos de probabilidade Normal

Definição: Gráfico dos quantis teóricos obtidos da distribuição Normal padrão em função dos quantis amostrais, no qual se traça uma reta bissetriz que representa os quantis teóricos da amostra, calculados para uma distribuição Normal. Se os descritores/propriedades seguirem uma distribuição próxima da distribuição Normal seus correspondentes valores dos quantis amostrais estarão distribuídos próximos à reta bissetriz (MAGALHÃES; DE LIMA, 2015).

Medida de similaridade (MS2 e MS3)

.Os valores de distância Euclidiana foram transformados em uma medida de similaridade, usando a Equação 3 (VARMUZA; PETER, 2008).

$$s_{AB} = 1 - \left(\frac{d_{AB}}{d_{max}} \right) \quad \text{Equação 1}$$

onde, d_{AB} é valor da distância Euclidiana entre dois compostos A e B e d_{max} é o valor máximo da distância Euclidiana do conjunto de dados.

Root Mean Square Distance (RMSD)

Os valores de *RMSD*, ou seja, os valores da raiz quadrada da média dos quadrados das distâncias, são calculados pela equação:

$$RMSD = \sqrt{\frac{\sum \Delta^2}{\eta}}$$

onde Δ é o valor da distância entre um par de átomos correspondentes alinhados e η é o número de átomos na molécula.

Similaridade Hausdorff-like - HSL

O cálculo do *HSL* foi feito usando uma matriz com todos os valores de coeficiente de Tanimoto calculados entre dois bancos, na qual as linhas e colunas representam os compostos de cada um dos bancos. A seguir, a medida de *HLS* (MAURI et al., 2016), definida pela **Equação 2**, foi calculada somando-se o valor máximo do coeficiente de Tanimoto de cada linha com o valor máximo de cada coluna e dividindo-se o resultado pela soma do número de compostos dos dois bancos.

$$H_S = \frac{\sum_{a \in A} \max_{b \in B} [s_{ab}] + \sum_{b \in B} \max_{a \in A} [s_{ba}]}{n_A + n_B} \quad \text{Equação 2}$$

onde, s_{ab} e s_{ba} são qualquer medida de similaridade, par-a-par, entre o elemento a do conjunto A e o elemento b do conjunto B ; $\max_{b \in B} [s_{ab}]$ é a máxima similaridade entre o elemento a (do conjunto A) com o conjunto B ; $\max_{a \in A} [s_{ba}]$ é a máxima similaridade entre o elemento b (do conjunto B) com o conjunto A e, n_A e n_B é o número total de elementos no conjunto A e B , respectivamente. Sendo que a medida de similaridade, par-a-par, corresponde ao coeficiente de Tanimoto.

ANEXO IV – PUBLICAÇÃO

Conformational flexibility of DENV NS2B/NS3pro: from the inhibitor effect to the serotype influence

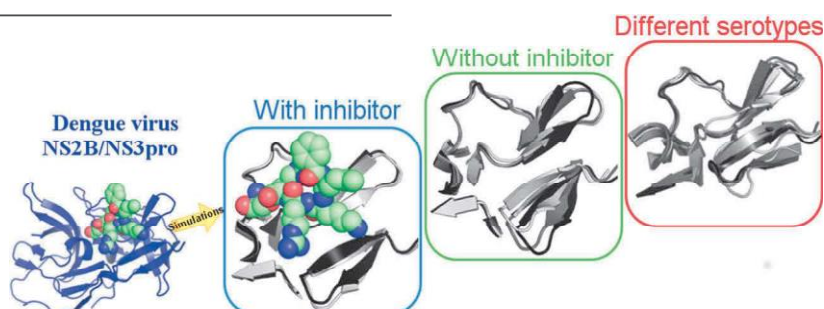
Erika Piccirillo¹ · Benjamin Merget² · Christoph A. Sotriffer² · Antonia T. do Amaral¹ 

Received: 10 September 2015 / Accepted: 11 February 2016 / Published online: 29 February 2016
© Springer International Publishing Switzerland 2016

Abstract The dengue virus (DENV) has four well-known serotypes, namely DENV1 to DENV4, which together cause 50–100 million infections worldwide each year. DENV NS2B/NS3pro is a protease recognized as a valid target for DENV antiviral drug discovery. However, NS2B/NS3pro conformational flexibility, involving in particular the NS2B region, is not yet completely understood and, hence, a big challenge for any virtual screening (VS) campaign. Molecular dynamics (MD) simulations were performed in this study to explore the DENV3 NS2B/NS3pro binding-site flexibility and obtain guidelines for further VS studies. MD simulations were done with and without the Bz-nKRR-H inhibitor, showing that the NS2B region stays close to the NS3pro core even in the ligand-free structure. Binding-site conformational

states obtained from the simulations were clustered and further analysed using GRID/PCA, identifying four conformations of potential importance for VS studies. A virtual screening applied to a set of 31 peptide-based DENV NS2B/NS3pro inhibitors, taken from literature, illustrated that selective alternative pharmacophore models can be constructed based on conformations derived from MD simulations. For the first time, the NS2B/NS3pro binding-site flexibility was evaluated for all DENV serotypes using homology models followed by MD simulations. Interestingly, the number of NS2B/NS3pro conformational states differed depending on the serotype. Binding-site differences could be identified that may be crucial to subsequent VS studies.

Graphical Abstract



Electronic supplementary material The online version of this article (doi:10.1007/s10822-016-9901-8) contains supplementary material, which is available to authorized users.

✉ Antonia T. do Amaral
atdamara@iq.usp.br

¹ Departamento de Química Fundamental, Instituto de Química, Universidade de São Paulo, São Paulo, Brazil

² Institut für Pharmazie und Lebensmittelchemie, Julius-Maximilians-Universität Würzburg, Würzburg, Germany

Keywords Dengue virus · NS2B/NS3 protease · Molecular dynamics simulations · Conformational flexibility

Abbreviations

2D-RMSD Two-dimensional root-mean-square deviations
Bz-nKRR-H Benzoyl-norleucine-lysine-arginine-arginine-aldehyde
DENV Dengue virus

SÚMULA CURRICULAR

1. DADOS PESSOAIS

- Nome: Erika Piccirillo
- Local e data de nascimento: São Paulo, 8 de março de 1988.

2. EDUCAÇÃO

- Colégio Objetivo, São Paulo, 2005.
- Universidade de São Paulo, São Paulo, 2011. Graduação em Farmácia e Bioquímica.

3. FORMAÇÃO COMPLEMENTAR

- Molecular modelling and computer simulations, Universidade Federal de Minas Gerais, UFMG, Brasil, 2014, Carga horária: 30 horas.
- Extensão universitária em Introdução a Programação, Instituto de Matemática e Estatística da Universidade de São Paulo, IME-USP, Brasil, 2013. Carga horária: 60 horas.
- Modelagem Molecular aplicada ao planejamento, 35ª Reunião Anual da Sociedade Brasileira de Química, 35ª RASBQ, Brasil, 2012, Carga horária: 6 horas.
- *Concepts and Techniques in Molecular Modeling*, Universidade Federal do ABC, UFABC, Brasil, 2011, Carga horária: 10 horas.
- *Continuum Solvation Models*, Universidade Federal do ABC, UFABC, Brasil, 2011. Carga horária: 10 horas.
- Planejamento de Candidatos à novos fármacos, 33ª Reunião Anual da Sociedade Brasileira de Química, 33ª RASBQ, Brasil, 2010, Carga horária: 6 horas.
- *New Approaches in Drug Discovery*, Instituto de Química da Universidade de São Paulo, IQ-USP, Brasil, Carga horária: 16 horas.

4. OCUPAÇÃO

- Bolsista de Iniciação Científica, FAPESP (N 2008/11138-5), de 01/02/2009 a 30/11/2011.
- Bolsista de Doutorado Direto, CNPq (N 140997/2012-1), de 01/03/2012 a 30/05/2012.
- Bolsista de Doutorado Direto, FAPESP (N 2012/06633-2), de 01/06/2012 a 30/06/2017.
- Bolsista PAE (Programa de Aperfeiçoamento de Ensino) da Pró-Reitoria de Pós-Graduação da Universidade de São Paulo, 1º Semestre de 2013.
- Bolsista BEPE (Bolsa de Estágio e Pesquisa no Exterior), FAPESP (N 2014/01614-5), de 30/03/2014 a 29/06/2014. Estágio de pesquisa realizado na Julius-Maximilians-Universität Würzburg, Würzburg, Alemanha, sob a co-orientação do Prof. Dr. Christoph A. Sotriffer.

5. PUBLICAÇÕES

Artigos Completos

- **PICCIRILLO, E.;** MERGET, B.; SOTRIFER, C. A.; AMARAL, A. T.-DO. Conformational flexibility of DENV NS2B/NS3pro: from the inhibitor effect to the serotype influence. *Journal of Computer-Aided Molecular Design*, v. 30, p. 251-270, 2016.

Resumos em Congressos

- **PICCIRILLO, E.;** AMARAL, A. T.-DO, *Using MD simulations binding modes and curated database to improve a VS model to search for Dengue virus NS2B/NS3PRO inhibitors*, In: 21st European Symposium on Quantitative Structure-Activity Relationship, Verona, Italy, 2016. (Apresentação de pôster)
- **PICCIRILLO, E.;** MERGET, B.; SOTRIFER, C. A.; AMARAL, A. T.-DO, *How to assess Dengue virus NS2B/NS3 pro binding- site flexibility in the search of for novel inhibitors?*, In: Gordon Research Seminar on Computer-Aided Drug Design, Mount Snow Resort, VT, United States, 2015. (Apresentação de pôster)
- **PICCIRILLO, E.;** MERGET, B.; SOTRIFER, C. A.; AMARAL, A. T.-DO, *How to assess Dengue virus NS2B/NS3 pro binding- site flexibility in the search of for novel inhibitors?*, In: Gordon Research Conference on Computer-Aided Drug Design, Mount Snow Resort, VT, United States, 2015 (Apresentação de pôster).
- **PICCIRILLO, E.;** AMARAL, A. T.-DO, *Improved virtual screening model to search of FMDV Lb^{pro} inhibitors using crystal active-site interactions and curated database*, In: 23rd International Congress of the IUBMB and 44th Annual Meeting of the Brazilian Society for Biochemistry and Molecular Biology (SBBq), Foz do Iguaçu, Paraná, Brazil, 2015. (Apresentação de pôster)
- **PICCIRILLO, E.;** PRADIE, N. A.; AMARAL, A. T.-DO, *Search of novel FMDV Lb^{pro} inhibitors: characterization of the inhibition mechanism of compounds previously selected by docking*, In: 7th Brazilian Symposium on Medicinal Chemistry - BrazMedChem2014, 2014. (Apresentação de pôster)
- **PICCIRILLO, E.;** PERON, L. M.; REZENDE DE, L.; AMARAL, A. T.-DO, *Search for novel inhibitors of dengue virus NS2B/NS3 protease by virtual screening.*, In: Gordon Research Conference Computer Aided Drug Design; Mount Snow Resort; VT; United States, 2013. (Apresentação de poster e 1 minuto de apresentação oral.
- **PICCIRILLO, E.;** REZENDE DE, L.; PERON, L. M.; AMARAL, A. T.-DO, *Virtual screening of potential inhibitors of dengue virus NS2B/NS3 protease: applying pharmacophore, similarity and docking filters*, In: 6th Brazilian Symposium on Medicinal Chemistry - BrazMedChem2012, Hotel Continental Canela; RS, Brazil, 2012. (Apresentação de pôster)
- **PICCIRILLO, E.;** MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Estudos da seletividade frente à cruzaína de um composto selecionado por Virtual Screening*, In: 35^a Reunião Anual da Sociedade Brasileira de Química, Hotel Monte Real Resort; Cidade: Águas de Lindoia , 2012. (Apresentação de pôster)
- **PICCIRILLO, E.;** MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Estudo do mecanismo de inativação frente à cruzaína de um inibidor selecionado por Virtual*

Screening, In: 34ª Reunião Anual da Sociedade Brasileira de Química, Centro de Convenções do Costão do Santinho Resort; Florianópolis, Santa Catarina; 2011. (Apresentação de pôster)

- **PICCIRILLO, E.**; MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Mechanism of Action of a cruzain inhibitor selected by virtual screening*, In: II Simpósio de Planejamento e Desenvolvimento de Novos Fármacos para Doenças Negligenciadas; University of São Paulo, São Paulo, SP, Brazil, 2011. (Apresentação de pôster)
- **PICCIRILLO, E.**; MALVEZZI, A.; REZENDE DE, L.; KATZIN, A. M.; TONHOSOLO, R.; AMARAL, A. T.-DO, *Homology model and docking studies on Plasmodium falciparum OPPs*, In: 5th Brazilian Symposium on Medicinal - BrazMedChem2010, Convention Center of Federal University of Ouro Preto; Ouro Preto, MG; Brazil, Chemistry, 2010. (Apresentação de pôster)
- **PICCIRILLO, E.**; MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Inibição time-dependent frente à Cruzaína de compostos selecionados por Virtual Screening*, In: 33ª Reunião Anual da Sociedade Brasileira de Química, Centro de Convenções do Hotel Monte Real Resort; Águas de Lindóia, 2010. (Apresentação de pôster)
- **PICCIRILLO, E.**; MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Study of the Mechanism of Action for a Cruzain Inhibitor Selected by Virtual Screening*, In: 5th Brazilian Symposium on Medicinal - BrazMedChem2010, Convention Center of Federal University of Ouro Preto; Ouro Preto, MG; Brazil, Chemistry, 2010. (Apresentação de pôster)
- **PICCIRILLO, E.**; MALVEZZI, A.; REZENDE DE, L.; AMARAL, A. T.-DO, *Caracterização do Mecanismo de ação de Inibidores da Cruzaína Selecionados por Virtual Screening*, In: 32ª Reunião Anual da Sociedade Brasileira de Química, Centro de Conveções; Fortaleza, Ceará; 2009. (Apresentação de pôster)

6. DEPOSITOS DE PEDIDOS DE PATENTE

- Depósito do pedido de patente junto ao Instituto Nacional da Propriedade Industrial (INPI), título: “Método de identificação de compostos antivirais, compostos antivirais, método de avaliação da funcionalidade dos compostos antivirais, uso dos compostos antivirais e composição farmacêutica antiviral”, inventores: Amaral, A. T.-Do; **Piccirillo, E.**; Juliano, M. A.; Kondo, M. Y. Número do registro: BR1020150006562 e data de depósito: 24/03/2015. Disponível on-line no site”(http://usp.technologypublisher.com/tech/Antiviral_Contra_A_Febre_Aftosa).

7. PRÊMIOS

- Melhor pôster na Sessão 2 - *Strategies in Drug Design*, no 6th *Brazilian Symposium on Medicinal Chemistry*, Divisão de Química Medicinal da Sociedade Brasileira de Química, 2012.
- Melhor pôster no II Simpósio de Planejamento e Desenvolvimento de Novos Fármacos para Doenças Negligenciadas, São Paulo, SP, Brasil, 2011.

8. OUTRAS ATIVIDADES

- Organização do Workshop da *OpenEye - Virtual Screening and Lead Discovery*. Workshop ministrado pelo Dr. Paul C. Hawkins, OpenEye, United States, no IQ-USP, São Paulo, Brasil, 2014.
- Co-orientação de alunos de iniciação científica com bolsas do CNPq: Rafael G. Frare (2014); Raul S. Rosa (2013) e Bernardo S. Pires (2012).