



UFG

UNIVERSIDADE FEDERAL DE GOIÁS

FACULDADE DE FARMÁCIA

PAULO RICARDO PIMENTA DA SILVA RAMOS

**Integração de abordagens computacionais para identificar novos inibidores da
proteína NS5 do vírus Zika**

Goiânia - GO
Dezembro, 2020

TERMO DE CIÊNCIA E DE AUTORIZAÇÃO PARA DISPONIBILIZAR VERSÕES ELETRÔNICAS DE TRABALHO DE CONCLUSÃO DE CURSO DE GRADUAÇÃO NO REPOSITÓRIO INSTITUCIONAL DA UFG

Na qualidade de titular dos direitos de autor, autorizo a Universidade Federal de Goiás (UFG) a disponibilizar, gratuitamente, por meio do Repositório Institucional (RI/UFG), regulamentado pela Resolução CEPEC nº 1204/2014, sem ressarcimento dos direitos autorais, de acordo com a Lei nº 9.610/98, o documento conforme permissões assinaladas abaixo, para fins de leitura, impressão e/ou *download*, a título de divulgação da produção científica brasileira, a partir desta data.

O conteúdo dos Trabalhos de Conclusão dos Cursos de Graduação disponibilizado no RI/UFG é de responsabilidade exclusiva dos autores. Ao encaminhar(em) o produto final, o(s) autor(a)(es)(as) e o(a) orientador(a) firmam o compromisso de que o trabalho não contém nenhuma violação de quaisquer direitos autorais ou outro direito de terceiros.

1. Identificação do Trabalho de Conclusão de Curso de Graduação (TCCG):

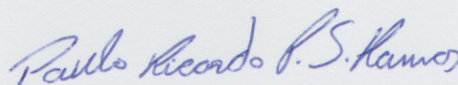
Nome(s) completo(s) do(a)(s) autor(a)(es)(as): Paulo Ricardo Pimenta Da Silva Ramos

Título do trabalho: Integração de abordagens computacionais para identificar novos inibidores da proteína NS5 do vírus Zika

2. Informações de acesso ao documento:

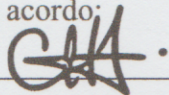
Concorda com a liberação total do documento SIM NÃO¹

Independente da concordância com a disponibilização eletrônica, é imprescindível o envio do(s) arquivo(s) em formato digital PDF do TCCG.



Assinatura(s) do(a)(s) autor(a)(es)(as)

Ciente e de acordo:



Assinatura do(a) orientador(a)²

Data: 25 / 01 / 2021

¹ Neste caso o documento será embargado por até um ano a partir da data de defesa. Após esse período, a possível disponibilização ocorrerá apenas mediante: a) consulta ao(à)(s) autor(a)(es)(as) e ao(à) orientador(a); b) novo Termo de Ciência e de Autorização (TECA) assinado e inserido no arquivo do TCCG. O documento não será disponibilizado durante o período de embargo.

Casos de embargo:

- Solicitação de registro de patente;
- Submissão de artigo em revista científica;
- Publicação como capítulo de livro.

PAULO RICARDO PIMENTA DA SILVA RAMOS

**Integração de abordagens computacionais para identificar novos inibidores da
proteína NS5 do vírus Zika**

Trabalho de conclusão de curso apresentado à
Faculdade de Farmácia da Universidade Federal de
Goiás, como parte dos requisitos para obtenção do
título de Farmacêutico.

Área de concentração: Química Medicinal

Orientadora: Profa. Dra. Carolina Horta Andrade

Coorientadora: Dra. Melina Mottin

Goiânia
Dezembro, 2020

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UFG.

Ramos, Paulo Ricardo Pimenta da Silva
Integração de abordagens computacionais para identificar novos inibidores da proteína NS5 do vírus Zika [manuscrito] / Paulo Ricardo Pimenta da Silva Ramos. - 2020.
LXXIII, 73 f.: il.

Orientador: Profa. Dra. Carolina Horta Andrade; co-orientadora Dra. Melina Mottin .

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Goiás, Faculdade Farmácia (FF), Farmácia, Goiânia, 2020.

Bibliografia.

Inclui siglas, mapas, abreviaturas, símbolos, gráfico, tabelas, lista de figuras, lista de tabelas.

1. Zika vírus. 2. Dengue vírus. 3. NS5. 4. docking molecular. 5. QSAR. I. Andrade, Carolina Horta, orient. II. Título.

CDU 615.1

PAULO RICARDO PIMENTA DA SILVA RAMOS

**Integração de abordagens computacionais para identificar novos inibidores da proteína
NS5 do vírus Zika**

Trabalho de Conclusão de Curso apresentado para a
obtenção de grau de Bacharel em Farmácia à
Faculdade de Farmácia da Universidade Federal de
Goiás

Data da aprovação: 14 / 12 /2020

Membros da Banca:



Prof. Dra. Carolina Horta Andrade
Orientadora
Farmacêutica, Dra
Faculdade de Farmácia – Universidade Federal de Goiás



Prof. Dr. Bruno Junior Neves
Farmacêutico, Dr
Faculdade de Farmácia – Universidade Federal de Goiás



Dra. Joyce Villa Verde Bastos Borba
Farmacêutica, Dra
Universidade Estadual de Campinas

EPÍGRAFE

“Soldados! Não vos entregueis a esses brutais... que vos desprezam... que vos escravizam... que arregimentam as vossas vidas... que ditam os vossos atos, as vossas ideias e os vossos sentimentos! Que vos fazem marchar no mesmo passo, que vos submetem a uma alimentação regrada, que vos tratam como gado humano e que vos utilizam como bucha de canhão! Não sois máquina! Homens é que sois! E com o amor da humanidade em vossas almas! Não odieis! Só odeiam os que não se fazem amar... os que não se fazem amar e os inumanos!

Soldados! Não batalheis pela escravidão! Lutai pela liberdade! No décimo sétimo capítulo de São Lucas está escrito que o Reino de Deus está dentro do homem – não de um só homem ou grupo de homens, mas dos homens todos! Está em vós! Vós, o povo, tendes o poder – o poder de criar máquinas. O poder de criar felicidade! Vós, o povo, tendes o poder de tornar esta vida livre e bela... de fazê-la uma aventura maravilhosa.”

Charlie Chaplin

DEDICATÓRIA

Dedico este trabalho aos meus queridos pais que sempre me apoiaram, amaram e fizeram o possível e o impossível para me proporcionar o privilégio de estudar.

AGRADECIMENTOS

À Deus em primeiro lugar, por sempre me guiar pelos melhores caminhos, muitas vezes espinhosos, mas necessários, pelo amor e compaixão hoje e sempre.

À minha família por todo apoio, suporte e amor.

À minha prezada e querida orientadora Prof. Dra. Carolina Horta Andrade, pela dedicação, compreensão, amizade e por ter me aberto as portas do LabMol, desde a minha iniciação científica no início de 2019, a realização deste TCC e atualmente ao mestrado, me guiado e orientado durante toda minha trajetória acadêmica.

À minha prezada e querida coorientadora Prof. Dra. Melina Mottin, pelo companheirismo, amizade e por ter me apresentado ao LabMol, me aceito como orientando, pela grande paciência em me ajudar, ensinar, corrigir, orientar e também por fazer parte de toda a minha trajetória acadêmica.

Aos colegas do LabMol, que sempre estiveram prontos para me auxiliar, pelo conhecimento compartilhado e pela amizade que me proporcionaram.

Aos meus amigos Tiago Abreu, Fábio Fernando e Victor Carrijo, que sempre me apoiaram.

A todos os meus amigos e familiares que fazem parte da minha vida, me apoiam e que mesmo indiretamente contribuíram para o meu êxito.

Às agências de fomento CNPq, CAPES pelo apoio financeiro.

A todos que contribuíram de alguma forma para a realização deste trabalho.

SUMÁRIO

1. INTRODUÇÃO.....	14
1.1 Epidemiologia.....	14
1.2 Transmissão.....	16
1.3 Sintomas	16
1.4 Genoma viral.....	18
1.5 Ciclo de replicação dos flavivírus	18
1.6 Proteína não-estrutural 5 (NS5)	20
1.6.1 Inibidores conhecidos da NS5	22
1.6.2 Análise sequencial dos sítios de ligação	24
1.6.3 Conservação dos sítios ativos	25
1.7 Estratégias de planejamento de fármacos auxiliado por computador (CADD) 26	
1.7.1 Relações quantitativas entre estrutura e atividade (QSAR).....	27
1.7.1.1 Princípios.....	27
1.7.1.2 Descritores moleculares	28
1.7.1.3 Métodos de aprendizado de máquina	29
1.7.1.4 Boas práticas de desenvolvimento e validação em QSAR.....	30
1.7.2 Docking molecular	31
1.7.2.1 Triagem Virtual	33
1.7.2.2 Análise de interações proteína-ligante	33
2 JUSTIFICATIVA.....	35
3 OBJETIVOS.....	36
3.1 Objetivos gerais	36
3.2 Objetivos específicos	36
4 METODOLOGIA.....	37
4.1 Desenvolvimento de modelos de QSAR para dados fenotípicos de ZIKV ...	37
4.1.1 Conjunto de dados.....	37
4.1.2 Preparo e padronização do conjunto de dados	37
4.1.3 Balanceamento do conjunto de dados.....	38
4.1.4 Geração e validação dos modelos de QSAR	38
4.1.5 Descritores moleculares	38
4.1.6 Método de aprendizado de máquina	39

4.1.7	Validação cruzada externa <i>5-fold</i>	39
4.1.8	Definição do domínio de aplicabilidade (DA)	39
4.1.9	Avaliação do desempenho dos modelos de QSAR	40
4.2	Triagem Virtual	40
4.2.1	Busca por compostos ativos nas bases de dados PubChem e ChEMBL	42
4.2.2	Modelo de QSAR fenotípico para ZIKV	42
4.2.3	Preparo da proteína	42
4.2.4	Docking Molecular	43
4.2.4.1	Preparo dos ligantes	43
4.2.5	Preparo dos <i>grids</i>	43
4.2.6	Validação do <i>docking</i> (<i>Redocking</i>)	44
4.2.7	Busca por similaridade em base de dados comerciais (eMolecules).....	44
4.2.8	Modelos Bayesianos.....	44
4.2.9	Análises automatizadas das interações ligante-proteína.....	45
5	RESULTADOS E DISCUSSÃO	46
5.1	Construção e validação dos modelos de QSAR para dados fenotípicos	46
5.2	Seleção de compostos inibidores de NS5 de DENV nas bases de dados PubChem e ChEMBL	47
5.3	Redocking para os ligantes co-cristalizados	47
5.4	Seleção de estruturas privilegiadas para guiar a triagem virtual	48
5.5	Triagem virtual	49
6	CONCLUSÕES.....	63
7	BIBLIOGRAFIA.....	64

LISTA DE FIGURAS

Figura 1- Arvore filogenética dos flavivírus	15
Figura 2- Mapa mostrando áreas com risco de infecção por ZIKV	17
Figura 3- Síndrome congênita do ZIKV	18
Figura 4- Partícula viral e genoma do ZIKV	19
Figura 5- Ciclo de replicação dos flavivírus	20
Figura 6- Proteína não estrutural 5 (NS5) completa com os domínios RdRP e MTase	21
Figura 7- Estrutura 3D do domínio NS5 polimerase (RdRP) de ZIKV.....	22
Figura 8- Estrutura 3D do domínio MTase da NS5 de ZIKV	23
Figura 9- Inibidores nucleosídeos e não nucleosídicos de RdRP	25
Figura 10- Representação da superfície da MTase (a) e RdRP (b) de NS5 de ZIKV	26
Figura 11- Representação geral do processo de construção de um modelo de QSAR	29
Figura 12- Fluxograma da busca por compostos privilegiados e triagem virtual.....	42
Figura 13- Validação da pose de docking molecular	49
Figura 14- Sequência de filtros aplicados para a seleção de estruturas privilegiadas para os domínios MTase e RdRP	50
Figura 15- Triagem virtual para a NS5 RdRP e MTase de ZIKV.....	52
Figura 16- Porcentagem de interações realizadas pelos resíduos do sítio GTP com os compostos.....	53
Figura 17- Porcentagem de interações realizadas pelos resíduos do sítio SAM com os compostos.....	54
Figura 18- Porcentagem de interações realizadas pelos resíduos do sítio RNA com os compostos	55
Figura 19- Estrutura dos ligantes cristalográficos presentes nos sítios SAM e GTP	61
Figura 20- Inibidor de RdRP de DENV NITD104	62
Figura 21- Interações intermoleculares em 3D e 2D para os compostos priorizados	67

LISTA DE TABELAS

Tabela 1 - Representação genérica de uma matriz de dados para um estudo de QSAR.....	32
Tabela 2 - Características estatísticas dos modelos de QSAR desenvolvidos.....	50
Tabela 3 - Resultados da inspeção visual baseada em Química Medicinal através do programa nAPOLI para todos os 12 compostos selecionados.....	60-64

LISTA DE ABREVIATURAS E SIGLAS

μM	Micro molar
ADMET	Absorção, Distribuição, Metabolismo, Excreção e Toxicidade
AUC	<i>Area Under The Curve</i> / Área sob a curva
CADD	<i>Computer-Assisted Drug Design</i> / Planejamento de fármacos auxiliado por computador
CCR	<i>Correct classification rate</i> / Taxa de classificação correta
DA	Domínio de aplicabilidade
DC-SIGN	<i>Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin</i> / Molécula de adesão intercelular específica para células dendríticas-3-agarradora não integrina
DENV	Vírus Dengue
DT	Limiar de distância
EC ₅₀	Concentração média que causa 50% da resposta máxima
FN	Falso negativo
FP	Falso positivo
GTP	Guanosina Trifosfato
HCV	Vírus da hepatite C
HSPG	<i>Heparan Sulfate Proteoglycan</i> / Proteoglicanos de heparam sulfato
HTS	<i>High throughput screening</i> / Triagem de alto rendimento
IC ₅₀	Concentração média que causa 50% da inibição máxima
JEV	Vírus da Encefalite Japonesa
k	<i>Cohen's kappa</i> / Kappa de Cohen
Kcal.mol ⁻¹	Quilocalorias por mol
kNN	<i>k-nearest neighbor</i> / Vizinho mais próximo
LBDD	<i>Ligand-Based Drug Design</i> / Planejamento de fármacos baseado nos ligantes
LogP	Coeficiente de partição óleo/água
MTase	Metiltransferase
NCATS	<i>National Center for Advancing Translational Sciences</i> / Centro Nacional para o Avanço das Ciências Translacionais
NPV	<i>Negative Predictive Value</i> / Valor preditivo negativo
NS5	Proteína Não-Estrutural 5

OECD	<i>Organization for Economy, Cooperation and Development /</i> Organização para Economia, Cooperação e Desenvolvimento
OMS	Organização Mundial da Saúde
P&D	Pesquisa e desenvolvimento
PDB	<i>Protein Data Bank /</i> Banco de dados de proteínas
pH	Potencial hidrogeniônico
PPV	<i>Positive Predictive Value /</i> Valor Preditivo Positivo
qHTS	<i>Quantitative High Throughput Screening /</i> Triagem quantitativa de alto rendimento
QSAR	<i>Quantitative Structure-Activity Relationship /</i> Relações quantitativas entre estrutura e atividade
RdRP	Polimerase
RE	Retículo endoplasmático
RF	<i>Random Forest /</i> Floresta randômica
RMSD	<i>Root Mean Square Deviation /</i> Desvio Médio Quadrático
RNA	Ácido ribonucleico
SAH	S-adenosil-homocisteína
SAM	S-adenosil-metionina
SBDD	<i>Structure Based Drug Design /</i> Planejamento de fármacos baseado na estrutura
SBG	Síndrome de Guillain-Barré
SE	Sensibilidade
SP	Especificidade
TAM	Tyro3, Axl, e Mer
TIM	<i>T cell/transmembrane, immunoglobulin, and mucin /</i> Mucina e imunoglobulina de células T
VN	Verdadeiro negativo
VP	Verdadeiro positivo
VS	Triagem virtual
WNV	Vírus do Nilo Ocidental
XP	Extra Precisão
YFV	Vírus Da Febre Amarela
ZIKV	Vírus Zika

RESUMO

O vírus Zika (ZIKV), causador de surtos e epidemias recentes em 2007, 2014 e 2015, é transmitido ao ser humano principalmente através da picada de fêmeas do mosquito *Aedes aegypti*. Além disso, a transmissão através de sangue, relação sexual e lactação foram descritas na literatura. A infecção por ZIKV apresenta diversas consequências principalmente neurológicas, até o momento, não há antivirais para combater a infecção nem vacinas para prevenir a infecção. A proteína não estrutural 5 (NS5), composta pelos domínios metiltransferase (MTase) e RNA polimerase (RdRP), possui um papel essencial na síntese e estabilidade do RNA viral, além da inibição do sistema imune do hospedeiro, sendo um alvo promissor para o desenvolvimento de novos fármacos antivirais. Neste trabalho, realizamos a integração de abordagens computacionais como modelos de relações quantitativas entre estrutura e atividade (QSAR) baseados em aprendizado de máquina, *docking molecular* e busca por similaridade para identificar novos inibidores da proteína NS5 de ZIKV. Devido à disponibilidade de dados na literatura e à elevada similaridade sequencial entre os sítios de NS5 de do vírus dengue (DENV) e de ZIKV, inicialmente, buscamos por inibidores da proteína NS5 de DENV nas bases de dados PubChem e ChEMBL, para guiar a triagem de novos inibidores de NS5 de ZIKV. Foram encontrados 145 inibidores de NS5 de DENV descritos na literatura. Realizamos a triagem virtual desses compostos através de um modelo de QSAR para dados fenotípicos de compostos testados contra ZIKV. Realizamos também estudos de *docking molecular* nos sítios ativos dos domínios MTase e RdRP de NS5 de ZIKV. Um total de 32 compostos foram priorizados nesta etapa. Posteriormente, foi realizada uma busca por similaridade estrutural aos *hits* virtuais encontrados no banco de dados comercial eMolecules e as etapas de QSAR e *docking* foram realizadas novamente com os compostos similares. Esta análise resultou em 4.953 compostos similares que também passaram pelo filtro de QSAR para ZIKV e *docking*. Após esta etapa, foram selecionados 176 compostos como inibidores promissores da NS5 de ZIKV. Esses compostos foram submetidos a modelos bayesianos de aprendizado de máquina para predição de atividade em ZIKV e citotoxicidade, resultando em 44 compostos preditos como ativos e não citotóxicos para células de mamíferos. Estes compostos foram então submetidos ao programa nAPOLI para análises de interações ligante-proteína. Essa etapa permitiu a seleção final de 14 compostos promissores que serão adquiridos e validados experimentalmente, através de ensaios enzimáticos na NS5 e celulares com o ZIKV que serão realizado em colaboração.

Palavras-chave: Zika vírus, Dengue vírus, NS5, *docking molecular*, QSAR, antivirais

ABSTRACT

The Zika virus (ZIKV), which caused recent outbreaks and epidemics in 2007, 2014 and 2015, is transmitted to humans mainly through the bite of female *Aedes aegypti* mosquitoes. In addition, transmission through blood, intercourse and lactation has been described in the literature. ZIKV infection has several consequences, mainly neurological serious, so far, there are no antivirals to fight infection or vaccines to prevent infection. Non-structural protein 5 (NS5), composed by the methyltransferase (MTase) and RNA polymerase (RdRP) domains, plays an essential role in the synthesis and stability of viral RNA, in addition to inhibiting the host's immune system, being a promising target for development of new antiviral. In this work, we perform the integration of computational approaches as quantitative structure activity relationships (QSAR) based on machine learning methods, molecular docking and search for similarity to identify new inhibitors of the NS5 protein from ZIKV. Due to the availability of data in the literature and the high sequential similarity between the NS5 sites of the dengue virus (DENV) and ZIKV, we initially searched for inhibitors of DENV NS5 protein in the PubChem and ChEMBL databases, to guide the screening of new NS5 inhibitors from ZIKV. 145 DENV NS5 inhibitors have been found in literature. We performed the virtual screening of these compounds using ZIKV phenotypic QSAR models. We also carry out molecular docking studies in the binding sites of the MTase and RdRP domains of ZIKV NS5. A total of 32 compounds were prioritized at this stage. Subsequently, a similarity search as performed in the commercial eMolecules database and the QSAR and docking steps were performed with the similar compounds. This analysis resulted in 4,953 similar compounds that also passed through the QSAR and docking filters. After this step, 176 compounds were selected as promising inhibitors of ZIKV NS5. We also submitted These compounds to Bayesian machine learning models to predict ZIKV activity and cytotoxicity, resulting in 44 compounds predicted to be active and non-cytotoxic to mammalian cells. These compounds were then submitted to the nAPOLI software for analysis of ligand-protein interactions. This step allowed the final selection of 14 promising compounds that will be acquired and experimentally validated, through ZIKV NS5 enzymatic assays and celular assays, that will be performed with collaborators.

Key-words: Zika virus, Dengue virus, NS5, molecular docking, QSAR, antivirals

1. INTRODUÇÃO

O vírus Zika (ZIKV) pertence à família *Flaviviridae* e ao gênero *Flavivirus*, que abrange também outros vírus patogênicos como o vírus da encefalite japonesa (JEV) e o vírus dengue (DENV) (MUSSO et al., 2015). É classificado como um arbovírus, ou seja, é um agente viral transmitido por vetores artrópodes, como mosquitos e carrapatos, que tem sua origem zoonótica. Evolutivamente, o ZIKV, assemelha-se a outros arbovírus como com o DENV, vírus da febre amarela (YFV) e vírus do Nilo Ocidental (WNV) (JUNIOR et al., 2015), como pode ser observado na Figura 1.

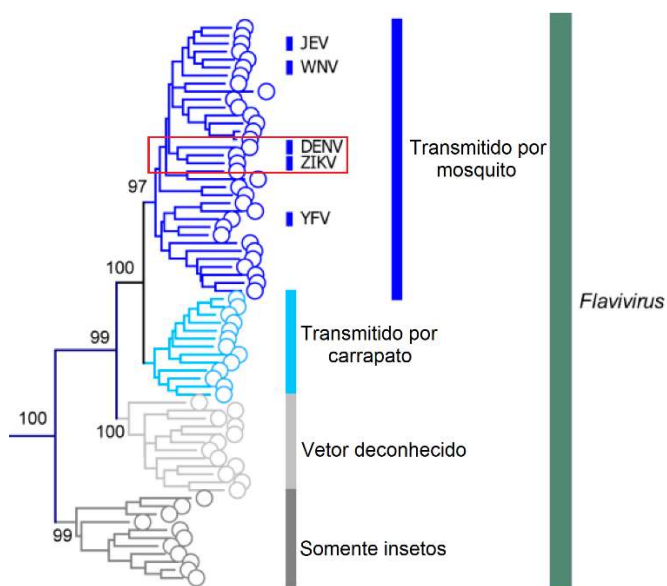


Figura 1: Arvore filogenética dos flavivírus. O grupo representado em azul escuro são transmitidos por mosquito e tem seus principais representantes, o vírus da encefalite japonesa (JEV), vírus do Nilo Ocidental (WNV), vírus Dengue (DENV), vírus Zika (ZIKV) e o vírus da febre amarela (YFV). O grupo representado em azul claro, são vírus transmitidos por carrapato. O grupo representado em cinza claro não tem seu vetor conhecido. E o grupo representado em cinza escuro tem sua transmissão por insetos específicos. Destacado em vermelho estão DENV e o ZIKV. Adaptado de (KLEINERT et al., 2019).

1.1 Epidemiologia

O ZIKV foi inicialmente identificado em 1947 na floresta de Zika na Uganda, África, em macacos *Rhesus*. A descoberta foi ao acaso, durante estudos sobre a febre amarela em mosquitos da espécie *Aedes africanus* coletados nesta floresta, porém sem indicativo de que o vírus fosse um patógeno humano (DICK, 1953). Posteriormente, os pesquisadores verificaram que este vírus passou a infectar humanos em Uganda (1952) (DICK, 1953) e na Nigéria (1953) (MACNAMARA, 1954). Nos anos de 1975 a 1977 foram observados indícios da presença do ZIKV em diversos

países africanos como Serra Leoa, Senegal, Tanzânia, Egito, Costa do Marfim e Gabão (JUNIOR et al., 2015). Nos anos de 1977 e 1978 foram encontrados os primeiros indicativos da circulação do ZIKV fora do continente africano, como na Indonésia (OLSON et al., 1981).

Em 2017, houve um surto de ZIKV nas ilhas do Yap, no Pacífico Ocidental, que representou a primeira transmissão do vírus fora da África e da Ásia (DUFFY et al., 2009). Posteriormente, o ZIKV se espalhou por diversas ilhas do pacífico (WORLD HEALTH ORGANIZATION, 2015).

Em 2014 o ZIKV chegou às Américas e em 2015 ao Brasil. A sua rápida disseminação nas Américas foi ocasionada pela presença do vetor *Aedes aegypti*. Os primeiros casos foram relatados na região Nordeste do Brasil e a sequência do vírus identificada pertencia a linhagem asiática (ECDC, 2016). O Brasil foi o país das Américas mais afetado, em que o vírus se mostrou bastante agressivo e se espalhou rapidamente, atingindo principalmente região Nordeste (ALIOTA et al., 2017).

Até junho de 2020, foram notificados 4.666 casos prováveis da doença em todo o Brasil (MS, 2020). O quadro alarmante associado à doença agravou-se devido à comprovação das correlações entre o ZIKV e complicações neurológicas, principalmente a microcefalia, observada em fetos e bebês de mães infectadas durante a gestação, e a síndrome de Guillain-Barré em adultos (HAZIN et al., 2016). Com isso, em fevereiro de 2016, a Organização Mundial da Saúde (OMS) declarou a epidemia de ZIKV como emergência global de saúde pública (COSTELLO et al., 2016). A Figura 2 apresenta o mapa com áreas de risco de transmissão de ZIKV em 2019.



Figura 2: Mapa mostrando áreas com risco de infecção por ZIKV. Em azul escuro, são destacados os países e territórios com transmissão atual ou anterior do ZIKV. Em azul claro, países e territórios com vetor, mas sem nenhum caso conhecido de infecção pelo ZIKV. Em cinza, países e territórios sem casos conhecidos de infecção pelo ZIKV e sem vetor. Adaptado de (WHO, 2019).

1.2 Transmissão

A transmissão do ZIKV ocorre predominantemente através da picada de seu principal vetor, o mosquito *Aedes aegypti*, que mantém hábitos diurnos e tornou-se extremamente adaptado aos ambientes urbanos (RODRIGUES-ALVES et al., 2020). O *A. aegypti* também é vetor de diversos outros vírus no território brasileiro, como o DENV, YFV e o vírus Chikungunya (CHIKV) (MARCHETTE; GARCIA; RUDNICK, 1969; LI et al., 2012). Outros tipos de transmissão como a sexual (ATKINSON et al., 2016; MUSSO et al., 2015a) e a transplacentária (BESNARD et al., 2014; MUSSO et al., 2015) também foram relatadas, proporcionando o aumento acelerado do número de infectados por ZIKV.

1.3 Sintomas

A infecção por ZIKV pode ser assintomática ou sintomática. No entanto, seus principais sintomas são leves e podem durar de dois a sete dias, incluindo febre, conjuntivite, artralgia, mialgia e erupção cutânea generalizada, que pode ser pruriginosa (DUFFY et al., 2009). Também foi observado dor de cabeça, dor retro-orbital, edema periférico e distúrbios gastrointestinais (SCHRAM, 2016). Atualmente, evidências científicas e epidemiológicas associam a infecção por ZIKV à microcefalia,

embasadas pela detecção do RNA viral em líquido amniótico ou no tecido cerebral de fetos e bebês afetados (TEIXEIRA et al., 2020).

A síndrome congênita do ZIKV (Figura 3) abrange as alterações resultantes da infecção congênita pelo ZIKV como complicações neuropsicomotoras, problemas oftalmológicos e auditivos, desproporção craniofacial, epilepsia e paralisia cerebral (GULLAND, 2016).

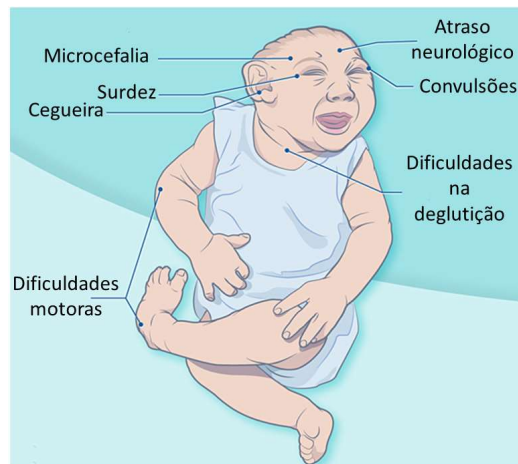


Figura 3: Síndrome congênita do ZIKV. Adaptado de (CDC, 2017).

Em Pernambuco (2016) foram relatados os primeiros casos de alterações oftálmicas em bebês, associadas ao ZIKV. Foram observados problemas como alterações maculares, alterações do nervo óptico, coloboma de íris, luxação de cristalino (VENTURA et al., 2016), microftalmia, anoftalmia, catarata congênita e calcificações intraoculares (HONEIN et al., 2017). Em adultos, o sintoma ocular mais frequente é a conjuntivite não purulenta. Entretanto problemas mais graves como a ruptura do epitélio pigmentar da retina macular externa (RPE) e a iridociclite, foram relatados em indivíduos imunocomprometidos e saudáveis (BRUNO M., 2016; HENRY et al., 2017; KODATI et al., 2017).

Outra complicação ocasionada pelo ZIKV é a síndrome de Guillain-Barré (SBG) é uma doença autoimune que ocorre através da destruição da bainha de mielina dos neurônios e interfere na condução do impulso nervoso até os músculos. Trabalhos recentes estabelecem uma correlação da SBG com a infecção por ZIKV (NÓBREGA et al., 2018). Os primeiros casos de SGB associado à infecção por ZIKV ocorreram na Polinésia Francesa em 2013 (CAO-LORMEAU et al., 2016), e em Porto Rico, em 2016 (DIRLIKOV et al., 2016). No Brasil, em 2015, as primeiras associações de ZIKV com

a SGB ocorreram após a confirmação da circulação vírus no país, principalmente na região Nordeste no qual os casos de SGB cresceram abruptamente em comparação ao mesmo período de 2014 (NÓBREGA et al., 2018).

1.4 Genoma viral

O genoma do ZIKV é constituído por uma fita simples de RNA com orientação negativa que codifica três proteínas estruturais: proteína de membrana (M), proteína de envelope (E) e proteína de capsídeo (C), dispostas sobre uma membrana lipídica, e sete proteínas não-estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5) (Figura 4) (MOTTIN et al., 2018).

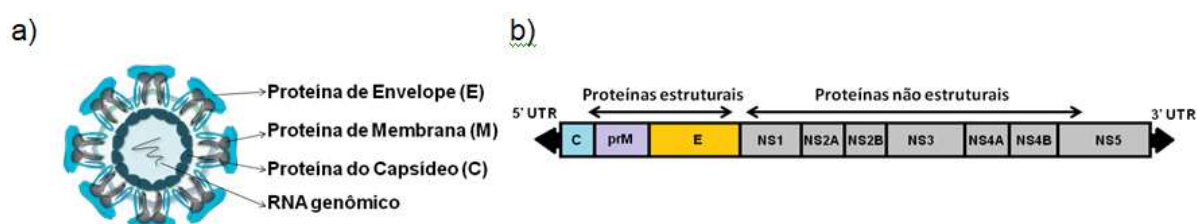


Figura 4: Partícula viral e genoma do ZIKV. (a) Representação da partícula viral de ZIKV formada pelas proteínas estruturais. (b) Organização do genoma viral com todas as proteínas transcritas, estruturais e não-estruturais. Adaptado de (MOTTIN et al., 2018).

1.5 Ciclo de replicação dos flavivírus

O ciclo de replicação viral (Figura 5) é iniciado através da ancoragem da proteína E com os receptores de TAM (do inglês, *Tyros3, Axl, and Mer*), TIM (do inglês, *T cell/transmembrane, immunoglobulin, and mucin*), receptor de manose, DC-SIGN (do inglês, *Dendritic Cell-Specific Intercellular adhesion molecule-3-Grabbing Non-integrin*) e HSPG (do inglês, *Heparan Sulfate Proteoglycan*) a membrana da célula hospedeira (NEUFELDT et al., 2018). Após a ancoragem, o vírus penetra na célula através de endocitose.

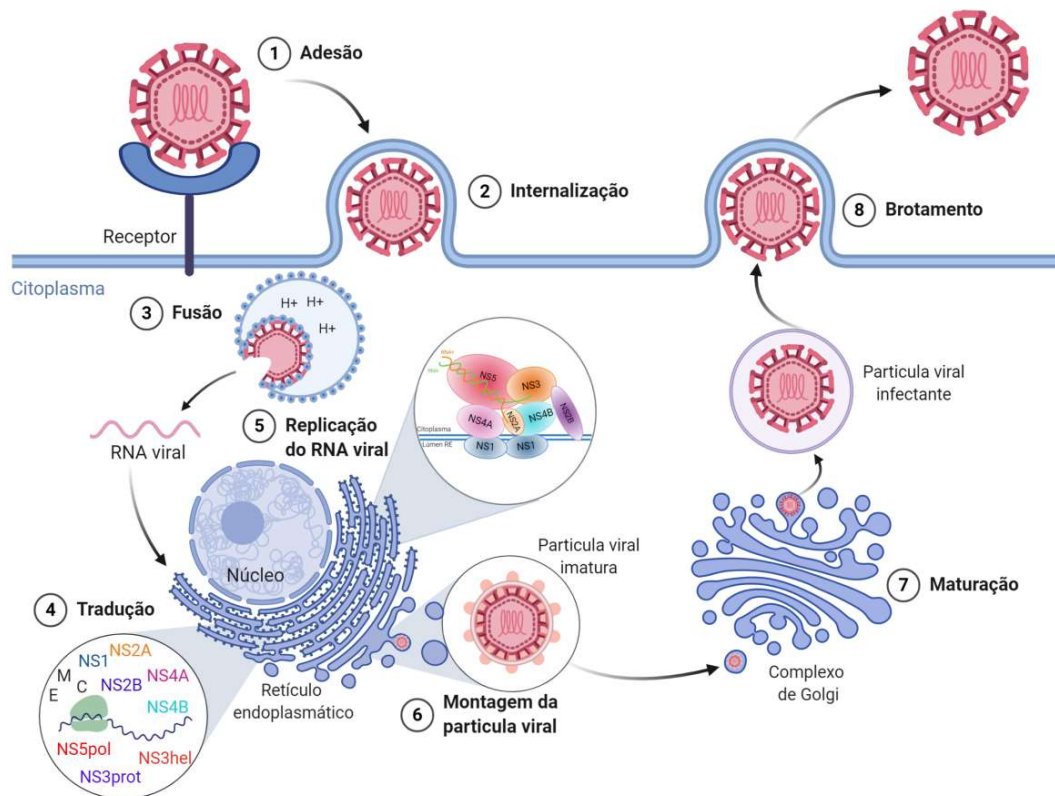


Figura 5: Ciclo de replicação dos flavivírus. **1** Após a picada do mosquito infectado, o vírus penetra na pele e se desenvolve nos linfonodos regionais, por reconhecimento da glicoproteína de envelope aos receptores de membrana TIM, receptor de manose, DC-SIGN e HSPG. **2** Uma vez iniciada a internalização viral, a via dependente de clatrina é ativada, expondo o agente viral a um endossomo de pH ácido, **3** que facilita o rearranjo estrutural da proteína E, assim como a fusão, liberando o material genético viral no citoplasma, **4** a tradução das proteínas virais pelos ribossomos do hospedeiro irá **5** realizar a replicação do genoma viral **6** e montagem de uma nova partícula viral **7** que passará por processos de maturação no complexo de Golgi e **8** seu respectivo brotamento, para infectar novas células e outro mosquito. Adaptado de (SOUSA, 2020).

No endossomo, a proteína E muda sua conformação por influência do pH e o envelope viral se funde com a membrana do endossomo. O capsídeo libera o RNA genômico no citoplasma. No retículo endoplasmático (RE), o RNA é traduzido em uma poliproteína viral. As proteases virais realizam a clivagem das proteínas e liberam os polipeptídeos funcionais que atuam nas primeiras etapas de tradução e auxiliam na síntese do RNA, que é utilizado como molde para a síntese de novas fitas de RNA genômico. Ao final desse processo, inicia-se a montagem do vírion na membrana do RE. As proteínas de capsídeo e o RNA genômico são então envelopados pelas glicoproteínas E e M, contidas na membrana para formar partículas imaturas, que são, então, transportadas pela via secretora. O amadurecimento acontece na rede Trans-

Golgi e por último os vírions maduros são liberados por brotamento (Figura 5) (KUIVANEN et al., 2017; RONG et al., 2017).

1.6 Proteína não-estrutural 5 (NS5)

A etapa de síntese do RNA genômico é realizada pela proteína não-estrutural 5 (NS5), constituída por dois domínios: polimerase (RdRP) e metiltransferase (MTase) (ZHAO et al., 2017). A Figura 6 apresenta a representação tridimensional em cartoon da NS5 com seus domínios MTase e RdRP.

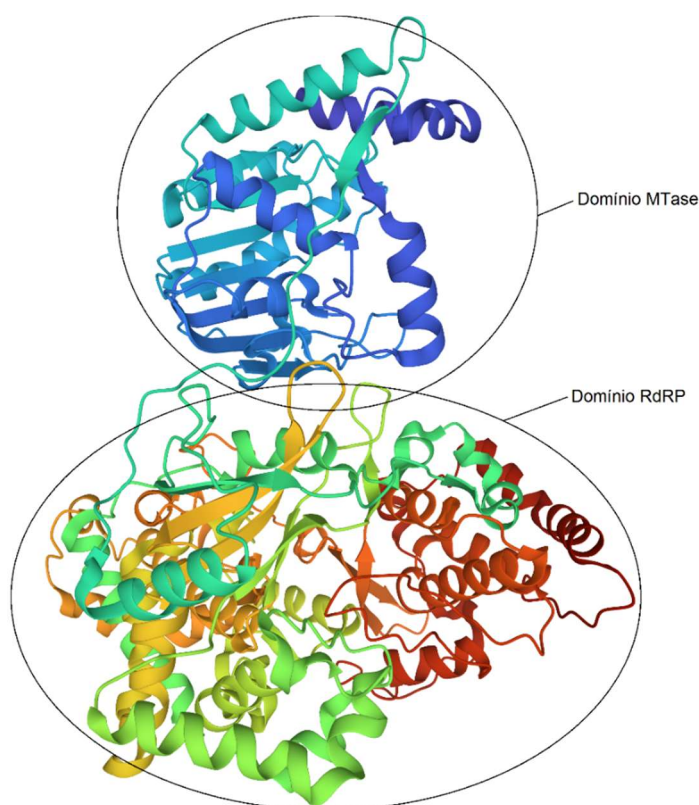


Figura 6: Proteína não estrutural 5 (NS5) completa com os domínios RdRP e MTase. A figura foi gerada no servidor online do PDB, utilizando o código PDB: 5TFR (UPADHYAY et al., 2017)

Como em outros flavivírus, o domínio RdRP de NS5 de ZIKV se assemelha a uma mão direita, com as três regiões principais (Figura 7): dedos (resíduos 321-488 e 542-608), palma (resíduos 489-541 e 609-714) e polegar (resíduos 715-903) (GODOY et al., 2017).

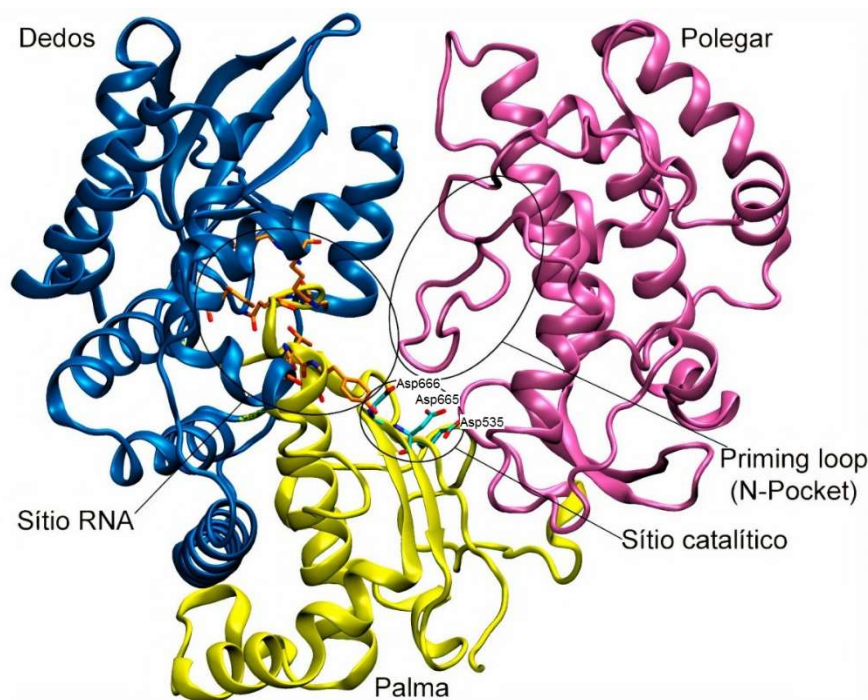


Figura 7: Estrutura 3D do domínio NS5 polimerase (RdRP) de ZIKV. As regiões foram coloridas da seguinte forma: palma, em amarelo; dedos, em azul e polegar, em rosa. Em destaque, os sítios de RNA, *N-pocket* (alostérico) e catalítico. O *priming loop* do sítio *N-pocket* também está destacado circulado em preto. Os resíduos dos sítios de RNA e catalítico estão representados em bastão e coloridos com os carbonos em laranja e azul ciano, respectivamente. Figura gerada com o software VMD, versão 1.9.3, com o código PDB ID: 5WZ3 (DUAN et al., 2017)

O RNA molde entra pelo sítio de RNA, que tem formato de túnel e é localizado entre as regiões dedos e polegar. O bolso alostérico ou *N-Pocket* é um túnel que faz o posicionamento dos nucleotídeos na extremidade 3' do RNA molde. Nesse bolso, o *priming loop* regula a ligação do RNA molde e a entrada de nucleotídeos para a polimerização (GODOY et al., 2017). No ZIKV, o sítio catalítico é formado principalmente por uma tríade de aspartatos (Asp535, Asp665 e Asp666) que catalisa a síntese do RNA de fita dupla (GODOY et al., 2017) (Figura 7).

O domínio MTase de ZIKV é altamente conservado quando comparado aos domínios MTases de outros flavivírus. Este domínio é responsável pela síntese e adição de um *cap* na extremidade 5' do RNA genômico, para proteção contra ribonucleases e evasão do sistema imune. Para construção do *cap*, são utilizados dois sítios de ligação: sítio de *S*-adenosil-metionina (SAM) e guanosina trifosfato (GTP) (Figura 8). A transferência do grupo metil do doador SAM ao RNA é realizada através da ligação dos resíduos da tétrade catalítica (Lys61 – Asp146 – Lys182 – Glu218) com

o GTP ancorado em seu sítio. O subproduto da doação é um S-adenosil-homocisteína (SAH) (ZHAO et al., 2017).

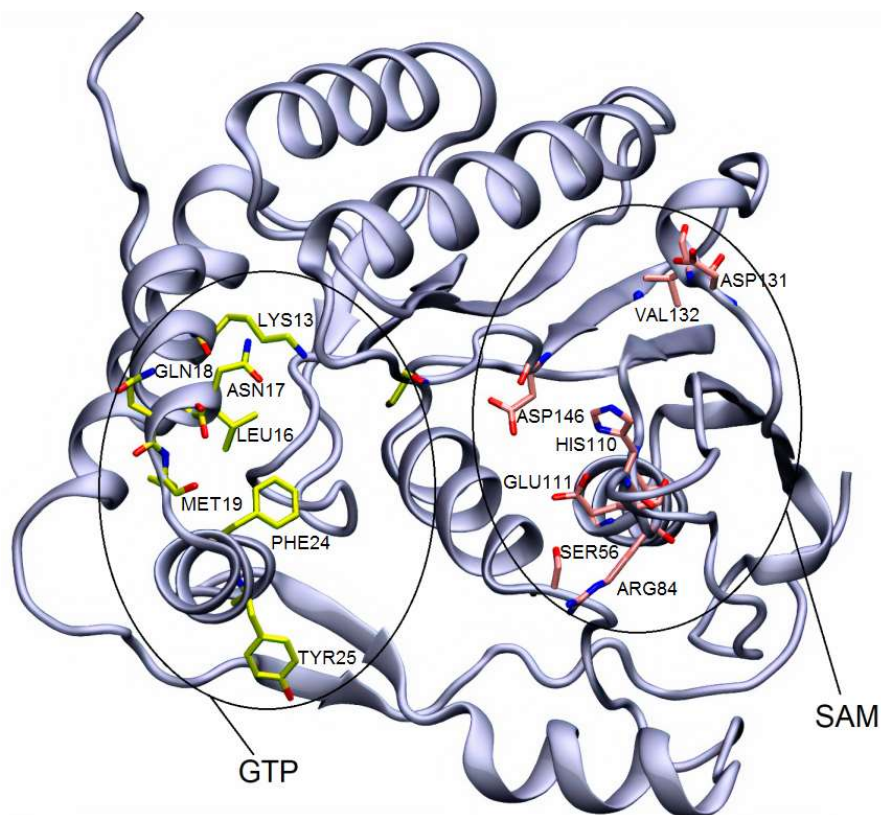


Figura 8: Estrutura 3D do domínio MTase da NS5 de ZIKV. Representada no formato de *cartoon* e colorido em azul. Os resíduos do sítio de SAM estão destacados em rosa e do sítio GTP, em amarelo. Figura gerada com o software VMD, versão 1.9.3, com o código PDB ID: 5KQS (COLOMA et al., 2016)

Enquanto que o domínio RdRP tem papel fundamental na replicação do genoma viral, o domínio MTase promove a proteção e estabilidade para o RNA genômico. Assim, ambos são alvos promissores para pesquisa e descoberta de novos fármacos (GODOY et al., 2017).

1.6.1 Inibidores conhecidos da NS5

Os inibidores nucleosídicos e nucleotídicos do domínio RdRP ligam-se aos sítios catalítico ou de RNA. No vírus da hepatite C (HCV), na região do sítio catalítico, um ribonucleosídeo 5'-trifosfato é adicionado à extremidade 3' da cadeia crescente de RNA. Portanto, uma abordagem racional para o desenvolvimento de inibidores da RdRP é a utilização de nucleosídicos que atuem como terminadores da cadeia e bloqueiem os passos da replicação. Um estudo anterior identificou diversos compostos nucleosídicos com atividade contra a RdRP de HCV (DING et al., 2005;

VARAPRASAD et al., 2007) (Figura 9). Em particular, o composto **(3)** que apresentou potência em faixa submicromolar ($EC_{50} = 0,6 \mu\text{M}$). O inibidor nucleosídico sofosbuvir **(4)** (Figura 9) obteve sucesso clínico para HCV, e depende da ativação por quinases do hospedeiro (SOFIA et al., 2012). O sofosbuvir teve sua atividade comprovada contra outros flavivírus como o ZIKV, em que, foi verificada a inibição do domínio RdRP de forma dose-dependente com um valor de IC_{50} de $0,38 \pm 0,03 \mu\text{M}$ (SACRAMENTO et al., 2017). A atividade do sofosbuvir também foi descrita contra DENV (XU et al., 2017) com um valor de EC_{50} de $1,4 \pm 0,2 \mu\text{M}$ e contra YFV, em que, foi verificado sua atividade com EC_{50} de $5 \mu\text{M}$ (DE FREITAS et al., 2019). Os inibidores não-nucleosídeos são amplamente desejados com o intuito de inibir o domínio RdRP dos flavivírus, por meio da ligação ao bolso alostérico (*N-pocket*) (POWDRILL; BERNATCHEZ; GÖTTE, 2010). Um estudo de relação estrutura-atividade (SAR, do inglês *structure-activity relationships*) inicial na estrutura do composto inibidor de HCV não-nucleosídeo **(5)** (Figura 9) levou a uma série de compostos ativos como o composto SB-711845 **(6)** com $IC_{50} = 0,010 \mu\text{M}$ (LE POGAM et al., 2006). O composto **(6)** mostrou boa biodisponibilidade em modelos animais e uma alta relação de distribuição fígado/plasma. Porém, a adição de albumina sérica humana no ensaio gerou uma diminuição substancial da potência, sugerindo uma possível ligação a proteínas plasmáticas. A ligação às proteínas plasmáticas foi reduzida pela adição de uma cadeia lateral de oxiacetamido no C-7 do anel benzotiadiazina, o que resultou no composto SB-750330 **(7)**, que obteve uma maior potência e foi selecionado como candidato clínico (SHAW et al., 2009). Outros inibidores não-nucleosídicos de RdRP foram avaliados em ensaios celulares contra DENV e obtiveram valores de EC_{50} de $11,09 \pm 0,21 \mu\text{M}$ **(8)** e $7,61 \pm 0,36 \mu\text{M}$ **(9)** (Figura 9) (PELLICCIA et al., 2017).

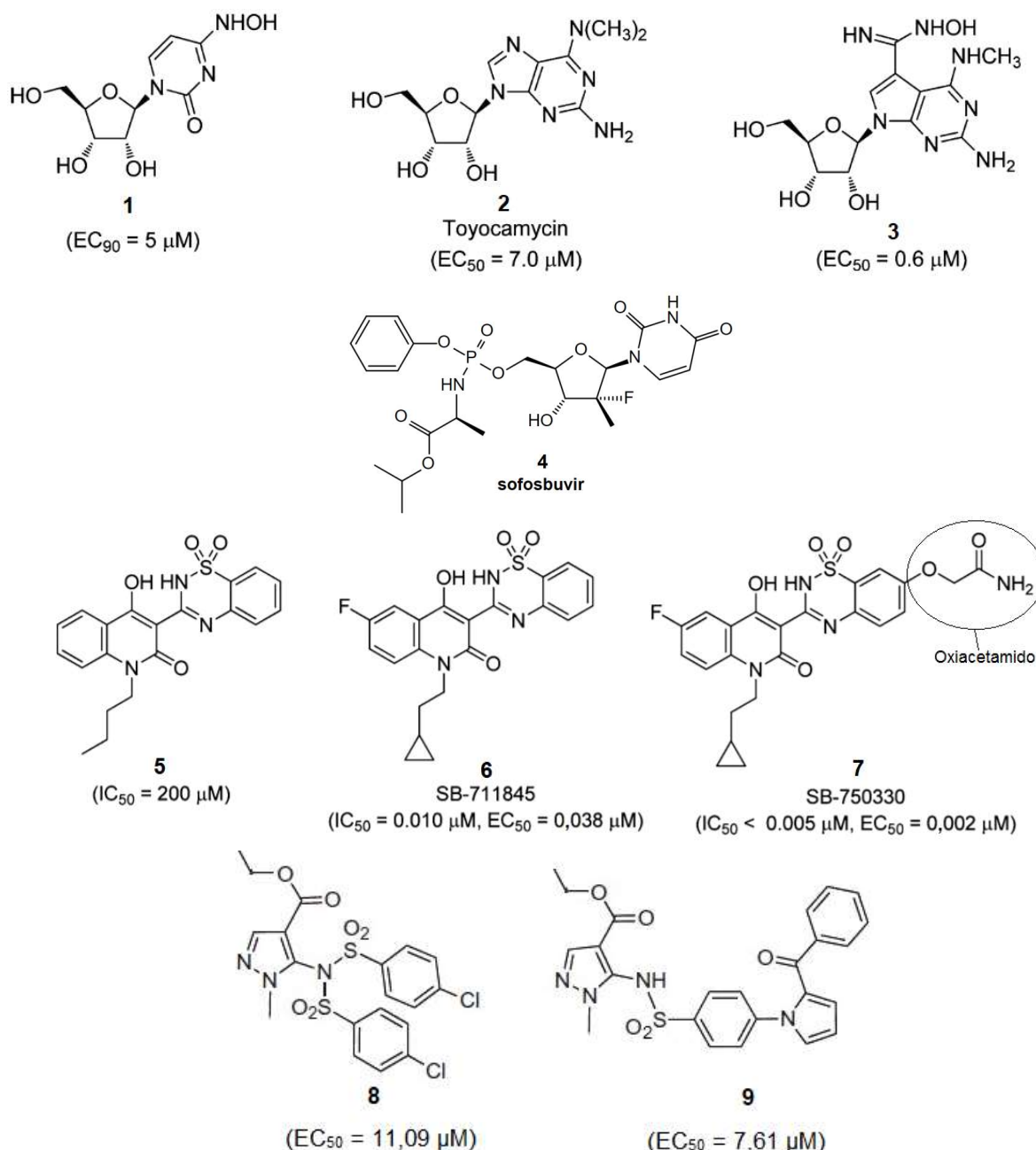


Figura 9: Inibidores nucleosídeos e não nucleosídicos de RdRP. Estruturas 1, 2 e 3 inibidores nucleosídeos identificados com modificação da base nitrogenada. Estrutura 4 sofosbuvir. Estruturas 5, 6 e 7 inibidores de RdRP de HCV não-nucleosídicos derivados da benzotiadiazina. Estruturas 8 e 9 inibidores não-nucleosídicos de RdRP de DENV. Adaptado de (SOFIA et al., 2012; PELLICCIA et al., 2017).

1.6.2 Análise sequencial dos sítios de ligação

A análise da estrutura primária e terciária de proteínas é uma abordagem em que se alinham as estruturas primárias e identifica a similaridade, comparando também a posição espacial e o tipo de resíduo. As proteínas NS5 dos flavivírus são altamente conservadas, principalmente na região dos sítios de ligação, conforme pode ser observado na Figura 10 (DUAN et al., 2017).

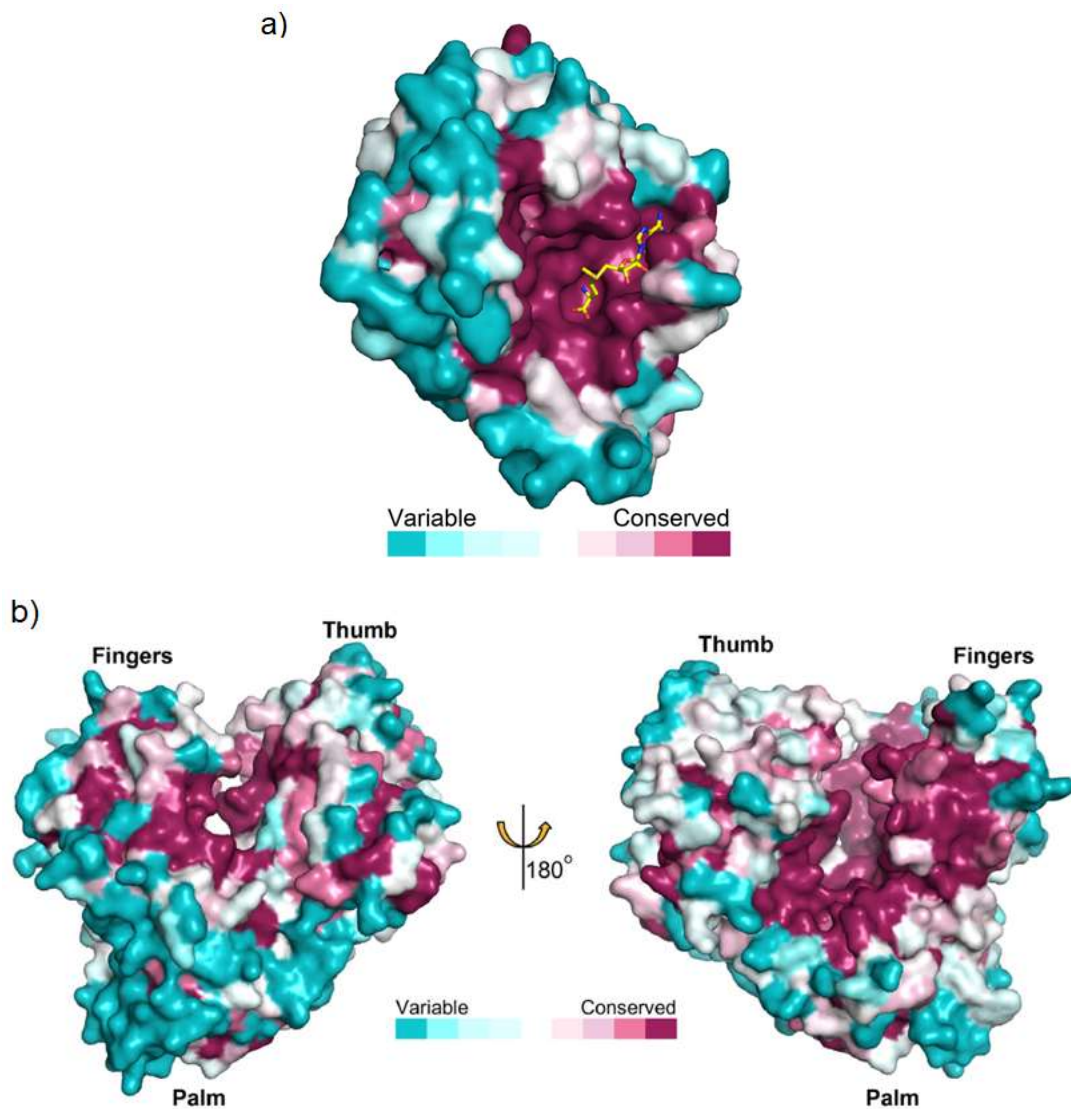


Figura 10: Representação da superfície da MTase (a) e RdRP (b) de NS5 de ZIKV. Os domínios foram coloridos de acordo com a conservação, com base no alinhamento das sequências de 73 MTases de flavivírus. (DUAN et al., 2017).

1.6.3 Conservação dos sítios ativos

Os três sítios ativos do domínio RdRP de NS5 de ZIKV apresentam alta conservação se comparados a DENV. Para o sítio *N-pocket*, dos 14 resíduos constituintes, apenas três sofreram variações. São eles, Met765, His800, Gln802 para DENV, que correspondem respectivamente aos resíduos, Leu707, Lys802, Glu804 em ZIKV. Para o sítio de RNA, dos dez resíduos constituintes, apenas um sofreu variação, sendo uma Gly604 em DENV que foi substituída por Val607 em ZIKV. Para o sítio catalítico, nenhum dos seis resíduos sofreu variação (LIM et al., 2016; DUAN et al., 2017).

No domínio MTase existe uma grande conservação entre os sítios de ligação de NS5 de DENV e ZIKV. Para o sítio SAM, nenhum dos nove resíduos constituintes sofreu variação. Para o sítio GTP, dos oito resíduos constituintes, apenas um sofreu variação, o resíduo Asp026 da MTase de DENV foi substituído por Tyr025 na MTase de ZIKV (COUTARD et al., 2014; COLOMA et al., 2016; ZHOU et al., 2017).

A região dos sítios de ligação da proteína NS5 é bastante conservada. A maior parte dos sítios não variou e, dos que variaram, poucos resíduos foram alterados. As características dos resíduos divergentes foram mantidas em todos os sítios da proteína, exceto pela substituição de Gln802 um resíduo polar, por Glu804, um resíduo com carga, no sítio *N-Pocket* (DUAN et al., 2017).

1.7 Estratégias de planejamento de fármacos auxiliado por computador (CADD)

O planejamento de fármacos auxiliado por computador (CADD, do inglês *Computer-Assisted Drug Design*) (MARSHALL, 1987) é uma abordagem que utiliza recursos computacionais (como bancos de dados, algoritmos e programas, serviços Web ou softwares de visualização 3D) para promover racionalmente o processo de descoberta e otimização de protótipos de fármacos (MACALINO et al., 2015). As estratégias computacionais tem se difundido como métodos de menor custo, mais rápidos e com maiores taxas de recuperação de compostos ativos, quando comparado a experimentos como triagens biológicas em alta vazão (HTS, do inglês *High throughput screening*) (REYNOLDS, 2014).

As estratégias de CADD podem ser classificadas em planejamento de fármacos baseado na estrutura (SBDD, do inglês, *Structure Based Drug Design*) e planejamento de fármacos baseado nos ligantes (LBDD, do inglês *Ligand-Based Drug Design*, LBDD) (OOMS, 2000). Quando as informações sobre o alvo biológico estão disponíveis, utiliza-se as estratégias de SBDD, como o *de novo design*, empregado quando se tem apenas a estrutura tridimensional do alvo. Já a estratégia LBDD é utilizada quando temos apenas as informações de ligantes bioativos, como modelos farmacofóricos (JORGENSEN, 2004, 2009; SILVA, 2015).

1.7.1 Relações quantitativas entre estrutura e atividade (QSAR)

O QSAR (do inglês, *Quantitative Structure-Activity Relationship*) é um método que relaciona a estrutura de um composto com sua atividade biológica (HANN; GREEN, 1999). Por meio de métodos estatísticos, é possível desenvolver modelos matemáticos capazes de prever a atividade biológica de compostos a partir da relação estabelecida entre estrutura e atividade biológica. (CRAMER, 2012; DUFFY et al., 2012; ALVES et al., 2017).

Os dados gerados por modelos de QSAR podem ser utilizados de várias formas, como: predição da atividade de novos compostos ainda não testados (e.g., potência, seletividade), separação de compostos com propriedades indesejáveis (e.g., modelos para predição de propriedades farmacocinéticas de absorção, distribuição, metabolismo, excreção e de toxicidade (ADMET), na otimização de compostos e obtenção de análogos mais potentes, através de modificações moleculares racionais baseadas nos modelos de QSAR (CHERKASOV et al., 2014).

A construção de modelos de QSAR teve uma grande evolução nos últimos anos. As análises que antes eram restritas a um pequeno grupo de compostos com estrutura semelhante (séries congêneres), passaram a ser aplicadas a grandes conjuntos de dados com alta diversidade estrutural, através da adição de vários métodos matemáticos, estatísticos e de aprendizado de máquina (CHERKASOV et al., 2014).

1.7.1.1 Princípios

Para a construção do modelo QSAR, duas informações são necessárias: (i) um conjunto de compostos com atividade biológica definida, e (ii) o cálculo dos descritores moleculares desses compostos. Diante disso, a relação pode ser estabelecida por meio da aplicação de métodos estatísticos, que são definidos genericamente pela equação 1:

$$P_i = k'(D_1, D_2, \dots, D_n) \quad (1)$$

Em que P_i é a atividade biológica ou propriedade da molécula, D são as propriedades estruturais ou descritores calculados e k' é o peso atribuído aos descritores moleculares por algoritmos computacionais através do cálculo de propriedades moleculares (TROPSHA, 2010).

O esquema geral de construção de um modelo de QSAR está representado na Figura 11.

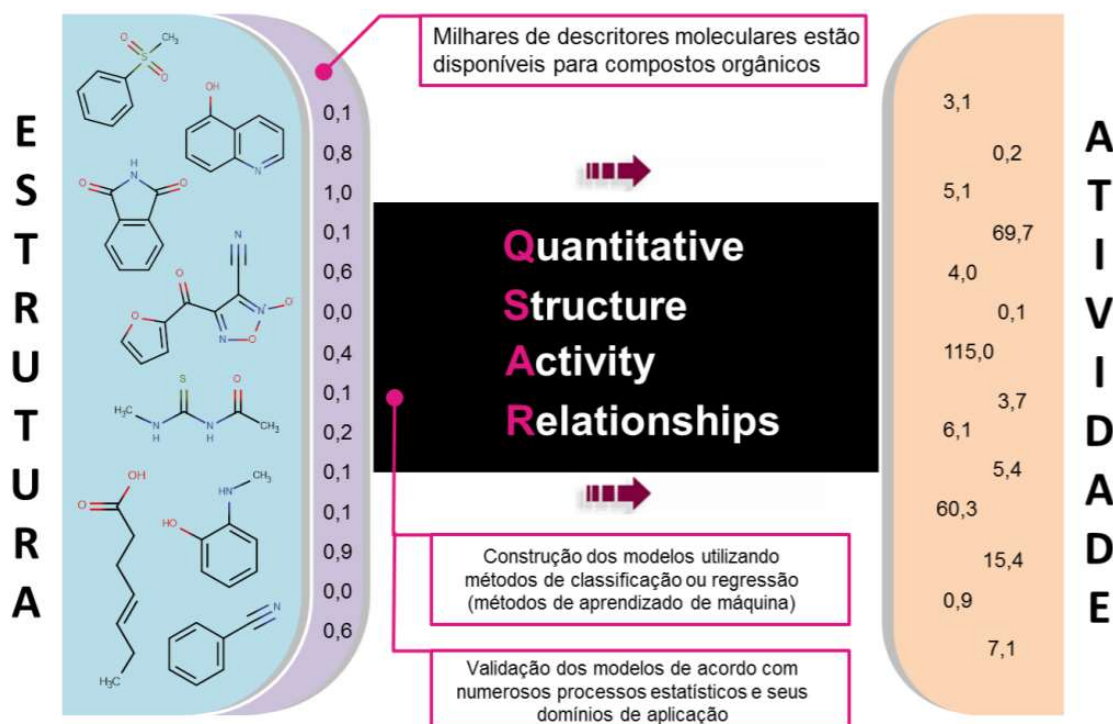


Figura 11: Representação geral do processo de construção de um modelo de QSAR (Adaptado de (TROPSHA, 2010)).

1.7.1.2 Descritores moleculares

Um descritor molecular é o resultado final do processo matemático e lógico de conversão de informações químicas codificadas pela representação simbólica de um composto em um número útil. O descritor pode ser ainda uma representação de algum experimento padronizado como o coeficiente de partição octanol/água ($\log P$), que é uma medida quantitativa para a lipofilicidade de compostos (TODESCHINI; CONSONNI, 2000).

Existe uma diversa variedade de descritores moleculares conhecidos, que podem ser classificados quanto a sua dimensionalidade ou natureza. Os descritores 1D (unidimensionais), são baseados em propriedades físico-químicas e fórmula molecular (ex. $\log P$, pK_a , refratividade molar, e outros); os 2D (bidimensionais) estão relacionados a representação 2D da molécula e considera as propriedades calculadas a partir dela (por exemplo, número de átomos de nitrogênio e oxigênio, número de

ligações, índices de conectividade); os 3D (tridimensionais) são dependentes da conformação da molécula, ou seja, da disposição de átomos no espaço tridimensional (por exemplo, área de superfície acessível ao solvente, volume molecular); já os descritores 4D incluem flexibilidade conformacional e liberdade de alinhamento por meio de conjunto em descritores tridimensionais convencionais encontrados em métodos tradicionais 3D-QSAR (XUE; BAJORATH, 2000).

No que se refere aos descritores classificados conforme a natureza, temos os descritores topológicos (ex. índice de conectividade molecular de Randic); constitucionais, que são derivados da composição química dos compostos (ex. peso molecular, número de grupos -OH); geométricos, que estão relacionados as coordenadas 3D da molécula (ex., área de superfície polar, volume molecular); eletrostáticos, baseados nas movimentações de cargas parciais (ex. índice de polaridade); quanto-mecânicos, que são baseados nas funções de onda que descrevem o movimento ondulatório dos elétrons (ex. orbitais moleculares); e por fim, impressões digitais moleculares (*fingerprints*) que avaliam a frequência de fragmentos na molécula (ex., Avalon, Morgan, etc) (TODESCHINI, R. CONSONNI, 2009; ROGERS; HAHN, 2010).

1.7.1.3 Métodos de aprendizado de máquina

O aprendizado de máquina consiste em métodos de inteligência artificial que dizem respeito ao estudo e construção de sistemas que podem aprender com os dados (ALVES et al., 2017). É desenvolvido, geralmente em três etapas: (i) representação dos dados, (ii) otimização da hipótese, e (iii) generalização (WELLING, 2011). O aprendizado de máquina pode ser usado na construção de modelos de QSAR, em que os dados são representados em uma matriz (Tabela 1), na qual o eixo X representa os descritores e o Y à atividade/propriedade.

Tabela 1: Representação genérica de uma matriz de dados para um estudo de QSAR.

Compostos	Atividade	Descritor 1	Descritor 2	Descritor 3 ...	Descritor n
Composto 1	y ₁	X ₁₁	X ₁₂	X ₁₃ ...	X _{1n}
Composto 2	y ₂	X ₂₁	X ₂₂	X ₂₃ ...	X _{2n}
Composto 3	y ₃	X ₃₁	X ₃₂	X ₃₃ ...	X _{3n}
...
Composto n	y _n	X _{n1}	X _{n2}	X _{n3} ...	X _{nn}

Em seguida é construída uma hipótese (equação) que estabelece uma relação entre os descritores e a atividade biológica (TROPSHA, 2010). O conjunto de dados é dividido em conjunto treinamento e conjunto teste, que são utilizados para a construção e validação dos modelos, respectivamente. Por fim, os modelos gerados caso sejam estatisticamente robustos e preditivos, tem a habilidade de prever corretamente a atividade/propriedade de compostos não utilizados em sua construção (WELLING, 2011).

A aplicação de modelos de aprendizado de máquina em QSAR visa classificar corretamente os dados dispostos para sua construção, treinamento e validação, e por meio de parâmetros matemáticos, descrever a atividade de novos compostos inseridos no modelo (PLEWCZYNSKI, 2011).

1.7.1.4 Boas práticas de desenvolvimento e validação em QSAR

Para garantir que os modelos desenvolvidos sejam confiáveis, a OECD (do inglês, *Organization for Economy, Cooperation and Development*) publicou em 2007 um protocolo de princípios para a validação de modelos de QSAR para fins regulatórios. São eles: (i) atividade biológica definida; (ii) utilização de algoritmos claros; (iii) ter domínio de aplicabilidade definido (DA); (iv) conter robustez e preditividade adequadas; e (v) realizar a interpretação mecanística do modelo, ou seja, para encontrar a relação entre o descritor e a atividade biológica, o objetivo é compreender melhor o mecanismo de ação de uma estrutura química ou aprofundar o conhecimento biológico sobre uma propriedade estudada.(OECD, 2007).

A validação de modelos de QSAR é realizada através de duas abordagens: a validação interna e a validação externa. Para a validação interna, parte dos compostos são utilizados na construção do modelo e parte para avaliação. Para a validação externa os compostos que serão confrontados com o modelo não são utilizados em sua construção (GOLBRAIKH et al., 2014).

O domínio de aplicabilidade (DA) está diretamente relacionado aos compostos do conjunto treinamento, de forma que, se o modelo for utilizado para predição de estruturas muito distantes quimicamente do conjunto treinamento, os resultados não serão confiáveis. Portanto, determinar qual o espaço químico coberto pelo modelo e qual a aplicabilidade do mesmo além do espaço químico, é de extrema importância

para a confiabilidade dos resultados (TROPSHA; GRAMATICA; GOMBAR, 2003; OECD, 2007).

1.7.2 Docking molecular

O *docking molecular*, acoplamento ou ancoragem molecular é uma abordagem que possibilita o acoplamento de um ligante em um sítio de ligação de uma macromolécula, com o objetivo de prever o complexo tridimensional proteína-ligante, e também as principais interações, o modo de ligação e a afinidade de ligação entre eles (MORRIS; LIM-WILBY, 2008). O *docking* é uma abordagem bastante utilizada no SBDD para triagem virtual de conjuntos de dados. Desde sua primeira utilização na década de 1980 (KUNTZ et al., 1982), o *docking* mostrou-se bastante importante para a pesquisa e desenvolvimento (P&D) de fármacos, seja na indústria ou academia. Sua utilização aumentou concomitantemente com a evolução do poder de processamento dos computadores e da disponibilidade de estruturas 3D de proteínas em bases de dados (MORRIS; LIM-WILBY, 2008). As aplicações do *docking* no processo de P&D de fármacos são várias como: busca por compostos que se liguem a um determinado sítio de ligação; determinação de qual a conformação mais estável do ligante; eleição de ligantes ou inibidores mais promissores para atuar como fármaco; e determinação da estrutura 3D proteína-ligante de um composto que tenha apresentado atividade *in vitro*. É importante ressaltar que o *docking*, como qualquer método, tem suas limitações, são elas: scores similares podem ser obtidos de poses diferentes, logo poses incorretas podem resultar em previsões de baixa precisão reduzindo a taxa de acerto da triagem, baixa precisão das funções de score reduzida a amostragem de ambos, ligante e conformações da macromolécula, inaptidão para modelar o solvente, entropia e flexibilidade do alvo com precisão. (GUIDO; OLIVA; ANDRICOPULO, 2008; MENG et al., 2014).

No processo de *docking* existem dois componentes básicos, os algoritmos de busca e os algoritmos de pontuação (YOUNG, 2009). Os algoritmos de busca exploram as diferentes orientações e conformações de um ligante dentro do sítio de uma macromolécula e podem ser divididos em três categorias básicas: métodos sistemáticos, métodos estocásticos e métodos determinísticos (KITCHEN et al., 2004; VERLI, 2014).

Os algoritmos sistemáticos tem como objetivo explorar todos os graus de liberdade do ligante por meio de construção incremental, busca conformacional, ou por bancos de dados. A construção incremental pode ser feita de várias maneiras, acoplando fragmentos da região do sítio e crescendo-os através de ligações covalentes ou dividindo os fragmentos em uma parte rígida (fragmento de núcleo) e uma parte flexível (cadeia lateral) e inserindo um núcleo rígido no sítio de ligação e gradualmente adicionando cadeias laterais (DESJARLAIS et al., 1986; YOUNG, 2009). Outro método de pesquisa sistemática é o uso de uma biblioteca de conformações pré-geradas (LEACH; KUNTZ, 1992; YOUNG, 2009).

Os algoritmos estocásticos fazem alterações aleatórias em um único ligante ou em uma população de ligantes. Cada pose do ligante recém-obtida é avaliada com base em uma função de probabilidade predefinida que verifica se a nova pose é melhor que a anterior e se irá avalia mantê-la ou descartá-la (HART; READ, 1992; GOODSELL et al., 1993). Dois métodos randômicos populares são a simulação de Monte Carlo e algoritmo genético (KITCHEN et al., 2004).

Dentre os métodos determinísticos, dinâmica molecular é a mais utilizada. No entanto, essa abordagem por vezes pode não transpor barreiras de alta energia inviabilizando que o ligante adquira uma conformação que tenha um mínimo de energia (NAKAJIMA et al., 1997). Diferentes métodos foram desenvolvidos para explorar melhor a situação da energia livre de ligação. Alguns métodos aumentam a temperatura da simulação (DI NOLA; ROCCATANO; BERENDSEN, 1994). Outros métodos podem manipular a energia potencial da superfície e torná-la mais suave, permitindo uma exploração completa da superfície (BROOIJMANS; KUNTZ, 2003).

A função de pontuação (score) são um conjunto de algoritmos que estimam a energia de afinidade entre ligante e macromolécula, e estes podem ser classificados em três categorias: funções de pontuação baseadas no campo de força, empíricas e baseadas no conhecimento. (RASHID et al., 2013).

As funções baseadas no campo de força (da mecânica molecular) geralmente quantificam um somatório de duas energias, a energia de interação receptor-ligante e a energia interna do ligante. A maioria das funções de pontuação do campo de força

considera apenas uma única conformação de proteína, o que torna possível omitir cálculos de energia da proteína e simplificar a pontuação (KITCHEN et al., 2004).

As funções de pontuação empíricas são baseadas na energia de interação e nas informações estruturais, os coeficientes destes vários termos podem ser obtidos por meio de análise de regressão. Por conseguinte, através da soma dos termos individuais é possível realizar a aproximação da energia de ligação (ELDRIDGE et al., 1997).

As funções de pontuação baseadas em conhecimento visam reproduzir a estrutura experimental em vez de energias de ligação. Nessas funções os complexos proteína-ligantes são modelados através da energia potencial de pares de ligações atômicos. Essas interações são definidas de acordo com um ambiente molecular em que os átomos se encontram dependendo do seu tipo atômico (MUEGGE, 2000). Portanto, como os métodos empíricos, as funções de pontuação baseadas em conhecimento tentam capturar implicitamente os efeitos de ligação que são difíceis de modelar explicitamente (YOUNG, 2009).

1.7.2.1 Triagem Virtual

A triagem virtual (VS, do inglês *Virtual Screening*) abrange o processo de triagem de bibliotecas de compostos usando modelos computacionais para avaliar e/ou selecionar compostos com as características desejadas (SCHNEIDER, 2010).. É uma abordagem bem estabelecida no moderno processo de planejamento e desenvolvimento de fármacos, bem como na avaliação de compostos potencialmente perigosos para o ambiente (ZHU et al., 2014). A triagem virtual é uma alternativa rápida e barata para a seleção de *hits*, o que reduz o número de compostos selecionados para avaliação experimental (KESSEL, 2011; POLGAR; M. KESERU, 2011).

1.7.2.2 Análise de interações proteína-ligante

A análise dos resultados de *docking* é baseada em conceitos da Química Medicinal e pode ser realizada através de inspeção visual (manual) ou de forma automatizada. Com o auxílio de *softwares* de visualização de complexos proteína-ligante, como o *Visual Molecular Dynamics* (VMD) (HUMPHREY; DALKE; SCHULTEN, 1996) ou PyMOL (DELANO, 2002), é possível observar as interações de

hidrogênio (acceptor ou doador de ligação de H), interações do tipo empilhamento de anéis, cátion- π , pontes salinas e interações metálicas, além de calcular ângulos e distâncias de interação. Na inspeção visual das poses de *docking*, deve-se observar as poses que apresentem: (a) a maioria dos heteroátomos do ligante fazendo ligação de hidrogênio, interações eletrostáticas favoráveis, ou ligações de halogênio com a proteína; (b) grupamentos hidrofóbicos do ligante empacotados por resíduos hidrofóbicos da proteína; (c) anéis aromáticos do ligante fazendo interações de T-stacking, base-stacking ou cátion- π com a proteína e (d) ligante faz interações fortes com regiões conservadas da proteína. Deve-se rejeitar poses que apresentem: (a) um ou mais grupos hidrofóbicos grandes (i-propila, t-butila ou benzila) do ligante expostos ao solvente; (b) mais de um grupamento polar “inserido” em bolsas hidrofóbicas da proteína; e (c) mais de uma interação eletrostática repulsiva com grupos carregados/polares da proteína (PERRYMAN et al., 2014).

Além disso, pode-se utilizar servidores automatizados para analisar as interações e posteriormente fazer uma análise visual final das poses mais promissoras. O programa nAPOLI é um servidor automatizado que faz análises de interação proteína-ligante em larga escala, através de um conjunto de algoritmos (MARTÍNEZ; ANDREANI; MARTÍNEZ, 2007; DURRANT; MCCAMMON, 2011; FABRI et al., 2011), que realizam a filtragem e validação de dados, construção de modelos baseados em gráficos bipartidos rotulados por propriedades físico-químicas atômicas e interações, cálculo das interações proteína-ligantes e clusterização de ligantes semelhantes e, por fim, superposição para busca por equivalências de resíduos e conservação de interações entre complexos. Diferentemente da maioria dos servidores, o programa nAPOLI permite a inspeção de várias poses de *docking* ao mesmo tempo; faz comparações com ligantes co-cristalizados em proteínas; agrupa os ligantes cristalográficos para comparação de suas interações e busca por interações não covalentes conservadas. O servidor apresenta os resultados na forma de representações gráficas abrangentes e relatórios dos resíduos de interação (FASSIO et al., 2019).

2 JUSTIFICATIVA

A infecção por ZIKV é uma doença negligenciada e está enquadrada pela OMS nas viroses emergentes com potencial de causar surtos e epidemias (WORLD HEALTH ORGANIZATION, 2020; XIA et al., 2020), mesmo após vários surtos com graves consequências neurológicas para a população infectada. Após sua chegada ao Brasil em 2015, o ZIKV ainda mantém uma alta taxa de prevalência e incidência, sendo que em 2019 foram notificados ao Ministério da Saúde 10.715 casos prováveis (taxa de incidência 5,1 casos por 100 mil habitantes). Até o momento, não há vacinas para a prevenção ou fármacos para o tratamento da infecção viral. Conseqüentemente, é urgente a descoberta de fármacos com ação antiviral para tratamento de infecções causadas por ZIKV. Neste aspecto, a proteína viral NS5 tem um papel extremamente importante na replicação do material genético viral e sobrevivência do vírus no hospedeiro e, portanto, é alvo potencial na pesquisa e desenvolvimento de novos fármacos. Atualmente, existem diversas estruturas 3D das proteínas de ZIKV obtidas experimentalmente disponíveis em bancos de dados. Da mesma forma existem diversos dados de ensaios biológicos de compostos testados para o ZIKV disponíveis em quimiotecas de acesso livre para o uso em estudos computacionais LBDD e SBDD.

Métodos computacionais são importantes ferramentas para auxiliar o planejamento racional de fármacos, através do emprego de metodologias como a triagem virtual, e são essenciais para o planejamento de fármacos contra ZIKV, de forma a reduzir a quantidade de compostos a serem sintetizados e testados e diminuindo custos. Da mesma forma, a integração de abordagens computacionais com avaliações experimentais eleva as chances de sucesso de candidatos a fármacos antivirais.

3 OBJETIVOS

3.1 Objetivos gerais

O presente trabalho teve como objetivo geral identificar potenciais inibidores da proteína NS5 de ZIKV, através da integração de metodologias computacionais de *docking molecular*, análise de similaridade em modelos de QSAR baseados em métodos bayesianos e de aprendizado de máquina.

3.2 Objetivos específicos

- Construir e validar modelos de QSAR para dados fenotípicos de ZIKV para a triagem virtual;
- Verificar a conservação de sítios de ligação a ligantes entre as proteínas NS5 de DENV e ZIKV;
- Buscar nas bases de dados ChEMBL e PubChem por compostos com ação inibitória em NS5 de DENV;
- Realizar a seleção de estruturas priorizadas utilizando os filtro de QSAR fenotípico e *docking molecular* na proteína NS5 de ZIKV;
- Buscar em bibliotecas comerciais por compostos que apresentem similaridade estrutural aos compostos priorizados na etapa anterior;
- Realizar a triagem virtual dos compostos similares, utilizando os filtros de QSAR fenotípico, *docking molecular* na proteína NS5 de ZIKV e de modelos bayesianos, a fim de, priorizar compostos com potencial inibitório contra a proteína NS5 de ZIKV, para posterior validação experimental (ensaios celulares e enzimáticos).

4 METODOLOGIA

4.1 Desenvolvimento de modelos de QSAR para dados fenotípicos de ZIKV

4.1.1 Conjunto de dados

Os dados utilizados para a construção dos modelos de QSAR foram obtidos na base de dados *PubChem*, bioensaio AID 1224857 (<https://pubchem.ncbi.nlm.nih.gov/bioassay/1224857>) realizados no *National Center for Advancing Translational Sciences* (NCATS).

Os compostos foram testados por meio de um ensaio de triagem quantitativa de alto desempenho (qHTS, do inglês *quantitative high throughput screening*) para avaliação da atividade inibitória de pequenas moléculas na caspase-3 induzida pelo ZIKV em células progenitoras neurais derivadas de iPSC humano (hNPCs). O aumento na atividade da caspase-3 indica uma ativação da via de apoptose causada pela infecção pelo vírus Zika em certos tipos de células neuronais. Ao utilizar um limiar de atividade de IC₅₀ de 10µM, esse conjunto de dados apresentou 45 compostos ativos (IC₅₀ ≤ 10µM) e 5.304 inativos (>10µM);

4.1.2 Preparo e padronização do conjunto de dados

Todas as estruturas químicas e respectivos dados biológicos foram preparadas de acordo com os protocolos desenvolvidos por Fourches e colaboradores (FOURCHES; MURATOV; TROPSHA, 2010, 2015; FOURCHES; MURATOV; ALEXANDER, 2016). Todas as etapas de preparação e padronização das estruturas foram realizadas através do workflow desenvolvido em nosso laboratório KSAR 1.6.2, implementado na plataforma KNIME 3.2.2 (BERTHOLD et al., 2006). Em resumo, átomos de hidrogênio explícitos foram adicionados, os sais foram removidos e quimiotipos específicos foram normalizados, como grupos nitro e anéis aromáticos. Polímeros, compostos organometálicos e misturas foram removidos. Ademais, duplicatas foram revisadas e excluídas quando necessário. Foram utilizados os seguintes critérios nas análises de duplicatas: (i) se os dados relatados para as duplicadas fossem iguais, um composto era mantido no conjunto de dados e o outro excluído; (ii) se os dados fossem diferentes entres si, ambos eram excluídos.

4.1.3 Balanceamento do conjunto de dados

O balanceamento do conjunto de dados foi realizado através do método *undersampling linear*, que consiste em reduzir os compostos da classe majoritária (não inibidores) a classe minoritária (inibidores). Ao contrário de outros métodos *undersampling* que realizam amostragem randômica para o balanceamento, o *undersampling linear* retém a maioria das estruturas representativas da classe majoritária, cobrindo assim o espaço químico original. Para isso, foi utilizado o método kNN (do inglês, *k-nearest neighbor*) (ALTMAN, 1992). As amostras da classe majoritária foram selecionadas linearmente através da faixa de *k*-distâncias, resultando em um conjunto balanceado de ativos e inativos obedecendo à proporção de 1:1, ou seja, 45 compostos ativos e 45 inativos.

4.1.4 Geração e validação dos modelos de QSAR

Os modelos de QSAR foram construídos utilizando o conjunto de dados obtido após as etapas de preparo, padronização e balanceamento. Os modelos foram desenvolvidos e validados de acordo com as boas práticas de modelagem por QSAR descritas na literatura (TROP SHA, 2010; CHERKASOV et al., 2014).

4.1.5 Descritores moleculares

Os descritores moleculares 2D utilizados foram do tipo impressão digital (*fingerprints*) e foram são calculadas usando o programa de código aberto RDKit (<http://www.rdkit.org>),

Neste trabalho foram calculados quatro descritores moleculares do tipo fingerprint: Avalon, MACCS, Morgan e FeatMorgan. Os descritores MACCS representam 166 fragmentos predefinidos associadas a um padrão denominado SMARTS (DURANT et al., 2002). Os descritores Avalon utilizam um gerador de impressão digital que enumera certos caminhos e classes de recursos do gráfico molecular (GEDECK; ROHDE; BARTELS, 2006). Os descritores Morgan são um tipo de impressão digital circular que codifica ambientes de átomos circulares até determinado raio de um átomo central usando o algoritmo de Morgan e invariantes de recurso.(MORGAN, 1965). Os descritores FeatMorgan constituem fragmentos circulares baseados no algoritmo Morgan (FCFP-like) e os combinam com características farmacofóricas. O farmacóforo é um conjunto de propriedades

estéricas e eletrônicas que são fundamentais para a interação com alvos biológicos (MORGAN, 1965).

4.1.6 Método de aprendizado de máquina

O algoritmo *Random Forest* (RF) foi utilizado para construção dos modelos de QSAR. O RF é um método de empacotamento de árvore que pode criar um grande número de árvores de decisão correlacionadas, e a previsão final é definida pela maioria dos votos de um conjunto de árvores de decisão (BREIMAN, 2001; SVETNIK et al., 2003). O algoritmo de RF foi executado utilizando o recurso Weka 3.7 do [KNIME](#)

4.1.7 Validação cruzada externa 5-fold

O modelo de QSAR foi submetido à validação cruzada externa utilizando o método *5-folds* (CHERKASOV et al., 2014), que consiste na separação aleatória do conjunto de dados em cinco subconjuntos de tamanhos iguais, sendo que quatro desses foram utilizados para a geração do modelo (80% dos compostos) e a quinta parte (20% dos compostos) para a validação externa. Esse processo foi repetido com todos os cinco subconjuntos individualmente permitindo que cada uma fosse utilizado como conjunto de validação externa e todos os modelos foram gerados a partir de seus respectivos conjuntos de modelagem. Os conjuntos de validação não foram utilizados na construção dos modelos.

4.1.8 Definição do domínio de aplicabilidade (DA)

O cálculo do DA foi realizado a partir da distância euclidiana dos compostos a serem preditos com os vizinhos mais próximos do conjunto de treinamento de cada modelo, para o cálculo da distância foi utilizado o algoritmo k-NN. A distância foi comparada a um limiar predefinido, e a predição seria considerada menos confiável, se a mesma fosse maior que o limiar de distância. Para calcular o limiar de distância (D_T) foi utilizada a equação 2:

$$D_T = \bar{y} + Z\sigma \quad (2)$$

Na equação, o \bar{y} corresponde a distância Euclidiana média em relação aos cinco vizinhos mais próximos ao conjunto treinamento, σ é o desvio padrão das

distancias e Z é um padrão aleatório que controla o nível de significância. O valor de Z foi definido em 0,5 (GOLBRAIKH et al., 2003).

4.1.9 Avaliação do desempenho dos modelos de QSAR

Para avaliar a preditividade dos modelos, foram calculadas as seguintes métricas: sensibilidade (SE), especificidade (SP), taxa de classificação correta (do inglês *correct classification rate*, CCR), valor preditivo positivo (do inglês *positive predictive value*, PPV) e o valor preditivo negativo (do inglês *negative predictive value*, NPV), conforme as equações 3, 4, 5, 6 e 7:

$$SE = \frac{VP}{VP + FN} \quad (3)$$

$$SP = \frac{VN}{VN + FP} \quad (4)$$

$$CCR = \frac{SE + SP}{2} \quad (5)$$

$$PPV = \frac{VP}{VP + FP} \quad (6)$$

$$NPV = \frac{VN}{VN + FN} \quad (7)$$

Onde VP e VN, correspondem respectivamente aos verdadeiros positivos e verdadeiros negativos, FP e FN correspondem respectivamente aos falsos positivos e falsos negativos.

4.2 Triagem Virtual

Em uma primeira busca por estrutura privilegiadas, realizou-se a integração de diversas abordagens computacionais para a busca por possíveis inibidores da proteína NS5 de ZIKV, com base nos inibidores conhecidos da NS5 de DENV. O fluxograma geral desta etapa é mostrado na Figura 12.

Primeiramente, foi realizada uma análise sequencial e comparação dos sítios de ligação das proteínas NS5 de ZIKV e DENV. Posteriormente foi realizada a busca por compostos inibidores de NS5 de DENV nas bases de dados PubChem (KIM et al.,

2016) e ChEMBL (GAULTON et al., 2012). Os compostos foram agrupados em inibidores de MTase e inibidores de RdRP. Em seguida, esses inibidores foram submetidos ao modelo de QSAR fenotípico de ZIKV, previamente desenvolvido e validado. Os compostos preditos como ativos foram submetidos aos cálculos de *docking* nos sítios de ligação da proteína NS5 de ZIKV, utilizando o programa GLIDE v 11.2 (SCHRÖDINGER, 2019). Finalmente, os compostos que apresentaram bons scores de *docking* (scores entre -6,1 e -9,6 kcal·mol⁻¹) foram submetidos a uma busca por similaridade estrutural (mínimo de 80% de similaridade) na base de dados de compostos comerciais *eMolecules* (“*eMolecules*”, 2006).

Na etapa de triagem virtual VS (Figura 12), os compostos similares aos compostos que deram origem ao estudo, foram submetidos ao modelo de QSAR fenotípico de ZIKV.

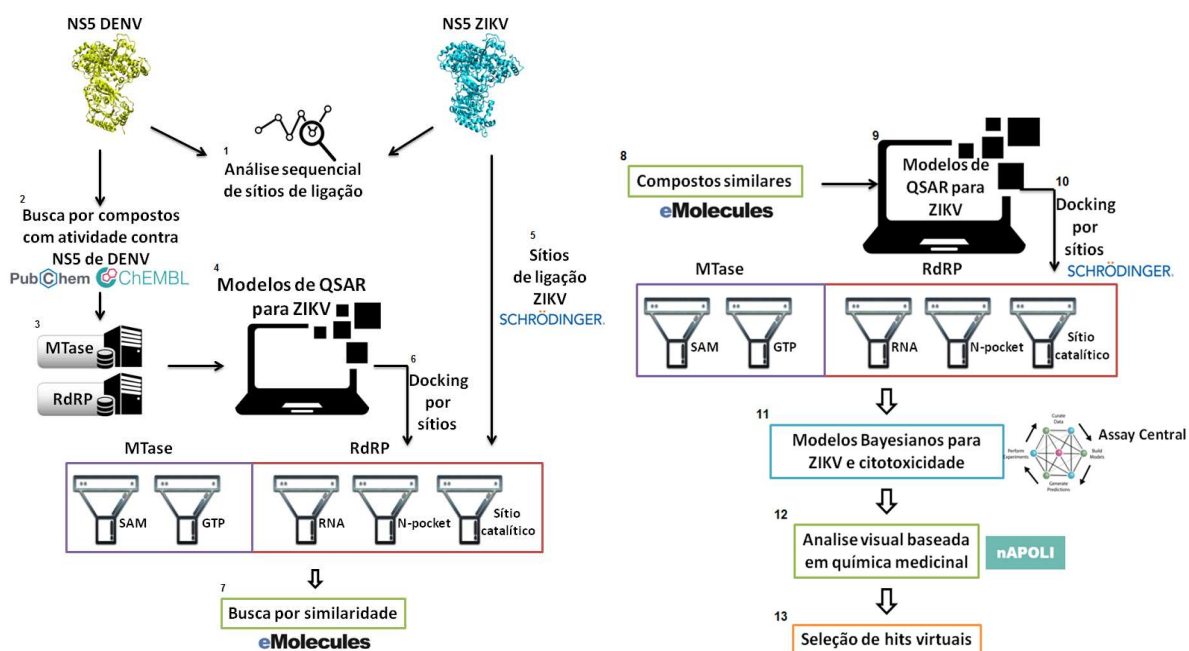


Figura 12: Fluxograma da busca por compostos privilegiados. (1) Análise sequencial dos sítios de ligação das proteínas NS5 de DENV e ZIKV. (2) Busca por compostos inibidores de NS5 de DENV. (3) Agrupamento dos compostos encontrados por domínio da NS5. (4) Submissão dos inibidores ao modelo de QSAR fenotípico de ZIKV. (5) Preparação dos *grids* com a proteína NS5 de ZIKV. (6) Realização dos *dockings* nos sítios da proteína NS5 de ZIKV. (7) Busca por similaridade na base de dados *eMolecules*. Fluxograma da etapa de triagem virtual. (8) Busca por similaridade. (9) Submissão dos compostos similares aos modelos de QSAR para ZIKV. (10) Realização dos *dockings* para os sítios da proteína NS5 de ZIKV. (11) Submissão dos melhores resultados do *docking* aos modelos bayesianos de aprendizado de máquina para ZIKV e citotoxicidade. (12) Análise de interações ligante-proteína através no programa nAPOLI. (13) Seleção de hits virtuais.

Após a predição pelos modelos de QSAR para ZIKV, os compostos preditos como ativos foram submetidos aos cálculos de *docking* molecular nos respectivos sítios e domínios da NS5 de ZIKV. Os melhores compostos selecionados no *docking* foram então submetidos aos modelos bayesianos de ZIKV e citotoxicidade, disponíveis no servidor AssayCentral, da empresa *Collaborations Pharmaceuticals, Inc.* Finalmente, os compostos preditos como ativos e não-citotóxicos, foram então submetidos ao programa nAPOLI para análise de interações ligante-proteína e priorização dos hits virtuais. Todas as etapas envolvidas nas duas triagens virtuais seguem descritas nos subitens abaixo.

4.2.1 Busca por compostos ativos nas bases de dados PubChem e ChEMBL

Inicialmente, foi realizada uma busca nas bases de dados PubChem (KIM et al., 2016) e ChEMBL (GAULTON et al., 2012) por ensaios que testaram inibidores para a proteína NS5 MTase e RdRP de DENV. O limiar de atividade utilizado para a seleção de compostos foi de 50 uM estabelecido pelo PubChem.

Os bioensaios utilizados para a coleta de dados do domínio RdRP foram: PubChem AID: 528347, 528350, 642356, 663478, 1168646, 1168659, 1229713, 1229715, 1277364, 1277368, 1277373, 1293709, 1301571 e 1301576.

Os bioensaios utilizados para coleta de dados do domínio MTase foram: PubChem AID: 458597, 458600, 458602, 458604, 1226489 e 1338232.

4.2.2 Modelo de QSAR fenotípico para ZIKV

Após a seleção dos compostos nas bases de dados PubChem e ChEMBL, as moléculas foram submetidas ao modelo de QSAR fenotípico de ZIKV, conforme detalhado no item 2.1.

4.2.3 Preparo da proteína

A primeira etapa para realização do *docking* foi a escolha da estrutura 3D da proteína alvo no banco de dados de proteínas *Protein Data Bank* (PDB). As estruturas selecionadas foram PDB ID (5WXB e 5KQS), ambas em forma holo, que possuem boa resolução (1,7 e 1,5 Å, respectivamente), para os sítios SAM e GTP do domínio MTase e a estrutura PDB ID (5WZ3), em sua forma apo, foi escolhida para os três sítios do domínio RdRP, com uma boa resolução (1,8 Å). Para preparar a proteína foi

utilizada a ferramenta *Protein Preparation Wizard* (SCHRÖDINGER, 2018a), através da qual hidrogênios foram adicionados, estados de protonação dos resíduos da proteína em $\text{pH } 7,4 \pm 0,5$ foram calculados, e a energia foi minimizada usando o campo de força OPLS-2005. Os artefatos de cristalografia foram removidos e águas estruturais com distâncias maiores que 5 Å das cadeias laterais foram removidas.

4.2.4 Docking Molecular

Nesta etapa, foi utilizado o programa Glide versão 11.2 (SCHRÖDINGER, 2019) pertencente ao pacote de programas do Maestro. Foi utilizado o modo de pontuação extra precisão (XP), que consiste na representação da forma e propriedades do receptor em um grid, por diferentes conjuntos de campos que fornecem um score progressivamente mais preciso da pose do ligante. Uma pontuação exaustiva de torções de ligantes foi calculada e gera uma coleção de conformações que foram examinadas durante o processo de *docking*. Dadas essas conformações de ligantes, as poses iniciais são realizadas de maneira determinística em todo o espaço, para localizar as mais promissoras. A partir de poses selecionadas pela triagem inicial, o ligante é refinado no espaço de torção no campo do receptor usando OPLS3 e OPLS-2005. Finalmente, um pequeno número de poses foi minimizado dentro do campo do receptor com total flexibilidade do ligante (minimização *pós-docking* ou PDM) (SCHRÖDINGER, 2019). O *docking* foi realizado considerando os ligantes flexíveis e a proteína rígida.

4.2.4.1 Preparo dos ligantes

O preparo dos ligantes foi realizado utilizando a ferramenta *LigPrep* (SCHRÖDINGER, 2018b) do Maestro, na qual calculamos os estados de protonação dos ligantes para o $\text{pH } 7,4 \pm 0,5$. Foram gerados confôrmeros e tautômeros para os compostos. Além disso, a quiralidade dos estereoisômeros foi mantida.

4.2.5 Preparo dos grids

Os grids da proteína foram preparados utilizando a ferramenta *Receptor Grid Generation*. Para o domínio RdRP, os *grids* foram centralizados nos resíduos: Asp535, Asp665 e Asp666 para o sítio catalítico (DUAN et al., 2017); Lys403, Asn494, Phe487, Gln605, Ser603, Arg483, Tyr609, Thr608, Val607, Trp797, para o sítio RNA (DUAN et

al., 2017); e Met763, Leu767, Leu513, Tyr768, Cys711, His713, His800, Trp805, Lys802, Thr796, Arg731, Ser798, Glu804 e Arg739 para o sítio *N-pocket* (DUAN et al., 2017). Para o domínio MTase os *grids* foram centralizados nos resíduos: Ser56, Lys61, Cys82, Arg84, Gly86, Trp87, Thr104, Lys105, His110, Glu111, Lys105, Asp131, Val132, Phe133, Cys145, Asp146, Ile147, Gly148, Glu149, Arg160, Arg163, Val164, Lys181, Leu182 e Glu217, para o sítio SAM (HERCIK et al., 2017); e Trp12, Lys13, Ala14, Arg15, Leu16, Asn17, Gln18, Glu23, Phe24, Tyr25, Tyr27, Lys28, Ser150, Ser151, Ser152, para o sítio GTP (ZHANG et al., 2017). O tamanho do grid foi estabelecido em 20 Å e as demais opções foram mantidas padrão.

4.2.6 Validação do *docking* (Redocking)

Para validar a acurácia do método de *docking* selecionado, pode-se realizar o *redocking* de um ligante co-cristalizado, na sua proteína de origem, para avaliar se o programa é capaz de reproduzir a pose experimental. Por intermédio dos ligantes cristalográficos S-Adenosil-L-homocisteína (SAH) e M7G das respectivas proteínas, 5WXB e 5KQS, do domínio MTase, foi feito o processo de *redocking*. A pose obtida pelo *redocking* foi analisada por inspeção visual e comparada com a da proteína original. O desvio médio quadrático (*root mean square deviation*, RMSD) foi calculado, entre as distancias dos átomos do ligante na pose de *docking* e no cristal. O valor de RMSD precisa ser inferior a 2 Å (DETERING; VARANI, 2004), dado que, quanto mais se aproximar de zero, mais próximo da pose original.

4.2.7 Busca por similaridade em base de dados comerciais (eMolecules)

Os compostos priorizados na primeira etapa da triagem virtual foram submetidos a uma busca por similaridade estrutural na base de dados de compostos comerciais *eMolecules* (“eMolecules”, 2006). Compostos que apresentavam 80% de similaridade estrutural com os inibidores experimentais foram selecionados.

4.2.8 Modelos Bayesianos

Os compostos priorizados na etapa de *docking* foram submetidos aos modelos para predição de atividade anti-ZIKV e citotoxicidade, através do servidor Assay Central™ (EKINS; CLARK,). Um valor de predição > 0.5, é considerado ativo.

O repositório de modelos bayesianos Assay Central (EKINS; CLARK,) para curagem de dados, construções de modelos e realização de testes, foi descrita

anteriormente em outros trabalhos (EKINS et al., 2008, 2010; PETRIE et al., 2013; YASUDA et al., 2013; RUSSO et al., 2018; DALECKI et al., 2019; WANG et al., 2019; ZORN et al., 2019).

4.2.9 Análises automatizadas das interações ligante-proteína

O programa nAPOLI (FASSIO et al., 2019) foi utilizado para realizar as análises automatizadas de interação ligante-proteína das poses de *docking* obtidas. Os resultados do *docking* foram analisados em função das interações dos ligantes com resíduos dos sítios de ligação e comparados com as interações dos ligantes co-cristalizados em proteínas cristalográficas. As interações utilizadas para realizar a priorização de ligantes foram: ligações de hidrogênios realizadas com os heteroátomos dos ligantes; grupamentos hidrofóbicos do ligante empacotados por resíduos hidrofóbicos da proteína; anéis aromáticos do ligante fazendo interações de *T-stacking*, *base-stacking* ou *cátion- π* com a proteína. Para o domínio MTase foram utilizadas as interações descritas para os ligantes cristalográficos SAH e M7G para os sítios SAM e GTP, respectivamente, descritos no item 2.5. E para o domínio RdRP as interações foram analisadas em função dos resíduos dos sítios de ligação, descritos no item 2.5.1.

5 RESULTADOS E DISCUSSÃO

5.1 Construção e validação dos modelos de QSAR para dados fenotípicos

Os modelos de QSAR para dados fenotípicos de ZIKV foram construídos através da combinação do método de aprendizado de máquina RF e os *fingerprints* moleculares de MACCS, Avalon, Morgan, FeatMorgan e descritores 2D do RDKit com a finalidade de distinguir entre compostos ativos e inativos contra o ZIKV. De acordo com os resultados estatísticos da validação cruzada externa de *5-folds* (Tabela 2), todos os modelos individuais gerados apresentaram performance estatística satisfatória, com valores de CCR variando entre 0,82-0,87; SE de 0,84-0,91; SP de 0,78-0,87; VPP de 0,81-0,87; VPN 0,84-0,90; κ de 0,64-0,73; e AUC de 0,88-0,93. O modelo de Descritores-RF, composto pelos descritores 2D do RDKit, foi o melhor individual para predição de atividade contra ZIKV e obteve CCR = 0,87, SE = 0,87, SP = 0,87 e AUC = 0,89. Os modelos de consenso foram construídos através dos valores médios previstos obtidos após a combinação dos modelos individuais. E o modelo de consenso rigoroso foi construído através da combinação dos modelos individuais com a condição restritiva do DA. Se o composto estava fora do DA para qualquer um dos modelos, o resultado foi especificado como não confiável.

Tabela 2: Características estatísticas dos modelos de QSAR desenvolvidos.

Modelo	CCR	Cohen's kappa (κ)	Se	Sp	VPP	VPN	AUC	Cobertura
MACCS-RF	0,82	0,64	0,84	0,80	0,81	0,84	0,88	0,62
FeatMorgan-RF	0,84	0,69	0,91	0,78	0,80	0,90	0,91	0,70
Avalon-GBM	0,84	0,69	0,87	0,82	0,83	0,86	0,93	0,66
Descritores 2D-RF	0,87	0,73	0,87	0,87	0,87	0,87	0,89	0,78
Consenso	0,91	0,81	0,95	0,86	0,87	0,95	0,91	1,00
Consenso rigoroso	0,87	0,75	0,91	0,84	0,88	0,88	0,87	0,63

CCR: taxa de classificação correta; SE: sensibilidade; SP: especificidade; VPP: valor preditivo positivo; e VPN: valor preditivo negativo; κ : Cohen's kappa; Cobertura: porcentagem de compostos do conjunto de teste dentro do domínio de aplicabilidade; Em negrito, os modelos de consenso construídos através dos valores médios previstos obtidos após a combinação dos modelos individuais e o modelo de consenso rigoroso construído através da combinação dos modelos individuais com a condição restritiva do DA.

O modelo consenso foi o que apresentou melhores resultados dentre todos, CCR = 0,91, SE = 0,95, SP = 0,86 e AUC = 0,91, e, portanto, foi escolhido para a VS.

5.2 Seleção de compostos inibidores de NS5 de DENV nas bases de dados PubChem e ChEMBL

Uma busca por bioensaios que testaram compostos com atividade para a proteína NS5 de DENV foi realizada nas plataformas PubChem e ChEMBL. Os dados foram preparados e separados em bases de dados por domínio. Os bioensaios reportados no PubChem também foram reportados na ChEMBL, dessa forma trabalhamos apenas com os dados do PubChem.

Com os compostos ativos selecionados, foram excluídas as duplicatas e restaram 92 compostos ativos para o domínio RdRP. Após a seleção de ativos e exclusão de duplicatas restaram 52 compostos ativos para o domínio MTase.

5.3 Redocking para os ligantes co-cristalizados

As proteínas selecionadas para elaboração dos grids dos sítios do domínio MTase continham substratos cristalográficos em seus sítios de ligação, SAM e GTP. Estes substratos foram utilizados em um processo de *redocking* para prever a capacidade de acerto em posicionar a molécula docada. Para o *redocking* no sítio SAM foi utilizado o substrato cristalográfico SAH (Figura 13), e obteve-se um valor de RMSD de 0,96 Å. Já para o sítio GTP, foi utilizado o ligante cristalográfico M7G (Figura 13), obtendo-se um valor de RMSD obtido de 0,94 Å. Ambos os resultados são menores que 2 Å e desta forma asseguram o poder preditivo do protocolo de *docking* utilizado neste trabalho. O valor de *docking score* variou entre -15,0 a -7,6 Kcal/mol⁻¹ para os ligantes de ambos os sítios.

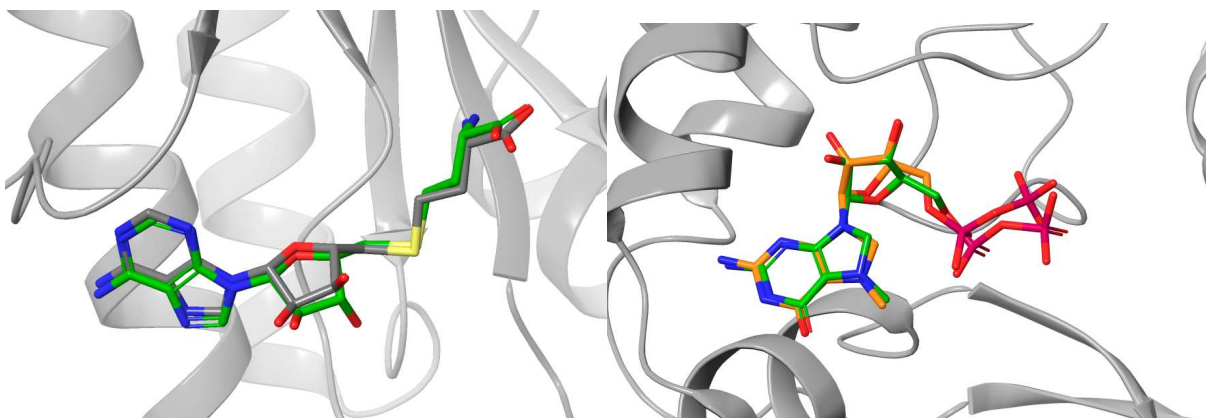


Figura 13: Validação da pose de *docking molecular* para o sítio SAM a esquerda. Em verde o ligante SAH em comparação com sua conformação original cristalográfica (cinza). RMSD = 0,96 Å. PDB ID: 5KQS. Validação da pose de *docking molecular* para o sítio GTP a direita. Em verde o ligante M7G em comparação com sua conformação original (laranja). RMSD = 0,94 Å. PDB ID: 5KQS.

5.4 Seleção de estruturas privilegiadas para guiar a triagem virtual

Na seleção de estruturas privilegiadas para MTase de ZIKV foram utilizados dois filtros, o QSAR para ZIKV, previamente construído e validado e o *docking molecular* para os sítios SAM e GTP.

Primeiramente foi utilizado o filtro de QSAR, com o auxílio do programa KMINE, que permitiu a seleção de 18 compostos inibidores de MTase de DENV. Posteriormente foi realizado o *docking* com o programa Glide no modo extra precisão (XP) para o sítio SAM, em que foram selecionados 05 compostos com *docking score* menor que -7 Kcal/mol^{-1} . Os mesmos parâmetros da seleção de estruturas para o sítio SAM (*docking score*, posição e interações), foram utilizados para a realização do *docking* para o sítio GTP que permitiu a seleção de 07 compostos. Na Figura 14 está representada a sequência de filtros aplicados nesta triagem.

Na seleção de estruturas privilegiadas para RdRP de ZIKV foram utilizados os inibidores de RdRP de DENV e os filtros de QSAR para ZIKV e *docking* para todos os sítios de ligação do domínio RdRP. Os critérios de seleção foram os mesmos para todos os sítios, O valor de *docking score* escolhido para o corte foi de -7 Kcal/mol^{-1} e a avaliação de interações e posição foi realizada. Na Figura 14 está representada a sequência de filtros aplicados nesta triagem.

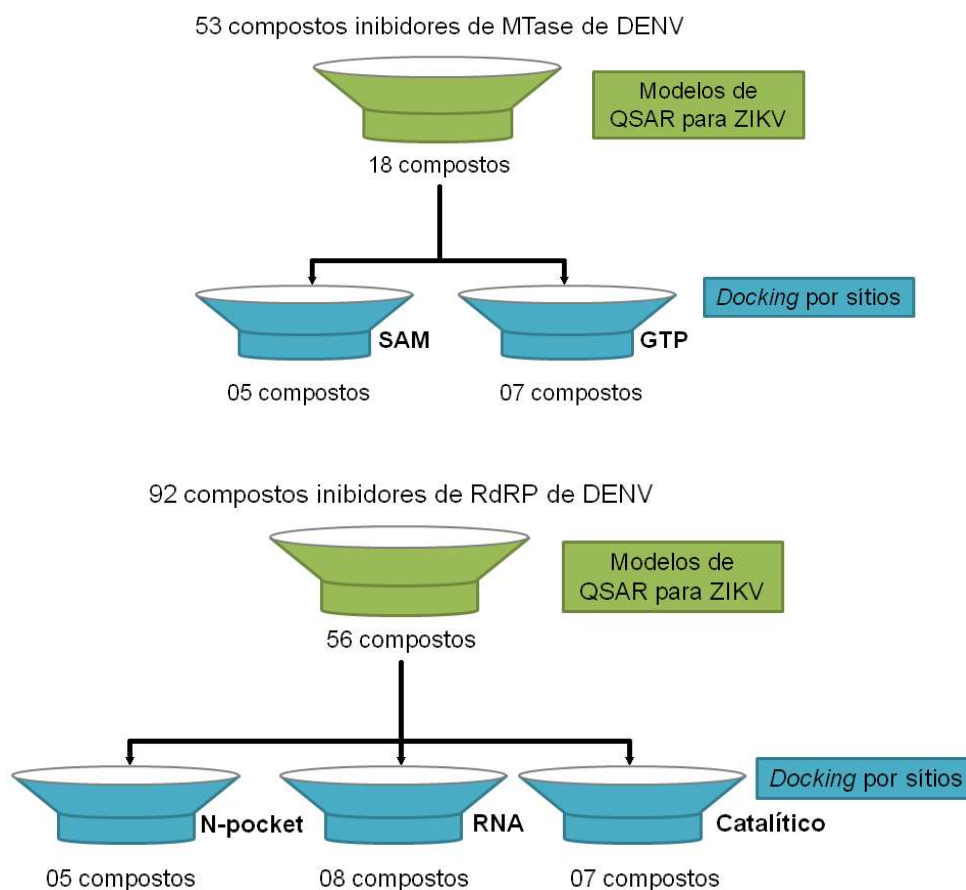


Figura 14: Sequência de filtros aplicados para a seleção de estruturas privilegiadas para os domínios MTase e RdRP.

O *docking* no sítio de RNA resultou em 08 compostos selecionados, para o sítio N-pocket o resultado foi de 05 compostos e para o sítio catalítico foram selecionados 07 compostos.

5.5 Triagem virtual

Os compostos resultantes de todas as etapas anteriores não são comerciais e por esta razão não são de fácil aquisição para validação experimental. Uma estratégia para solucionar esse problema é a busca por compostos similares em bases de dados comerciais como a eMolecules (“eMolecules”, 2006). Através dessa busca podemos encontrar compostos idênticos ou com alta similaridade e de fácil obtenção.

Os resultados de *docking* obtidos para cada sítio de ligação foram separados em conjuntos de dados. Para todos os compostos selecionados na etapa anterior, realizamos uma busca por compostos similares na base de dados comerciais eMolecules. Os dados obtidos foram separados por sítios e os resultados desta etapa

constituíram cinco bancos de dados com compostos similares aos compostos resultantes do *docking* para cada sítio de ligação.

Na busca por similaridade entre os resultados dos *dockings* para os sítios do domínio MTase, os 05 compostos obtidos para o sítio SAM geraram 1844 similares. E para o sítio GTP, os 07 compostos resultaram em 1234 compostos similares.

Para o domínio RdRP os resultados de *docking* para os sítios foram RNA 08 compostos, N-Pocket 05 compostos e catalítico 07 compostos, estes, geraram respectivamente 786, 1256 e 933 compostos após a busca por similaridade.

Na triagem virtual para os compostos similares aos inibidores de MTase de DENV, os mesmos foram submetidos aos mesmos filtros de QSAR fenotípico para ZIKV e *docking*. Uma vez que os mesmos não foram submetidos a nenhum tipo de triagem virtual e precisam ser avaliados nos mesmos filtros anteriores para a predição de sua atividade.

Os dois bancos de dados obtidos para os similares foram submetidos ao filtro fenotípico de QSAR fenotípico para ZIKV com o auxílio do software KNIME. Nessa etapa, foram selecionados para os sítios de ligação, 211 compostos para SAM e 323 compostos para GTP. Os *dockings* dos similares foram realizados com os mesmos arquivos de GRID desenvolvidos anteriormente com a programa Glide (SCHRÖDINGER, 2019) da interface Maestro . Os conjuntos de ligantes, agora separados por sítios, foram preparados com a ferramenta LigPrep (SCHRÖDINGER, 2018b) para obtenção de, confôrmeros, estados de protonação em pH fisiológico e retirada de sais. Os parâmetros de seleção e corte foram os mesmos utilizados anteriormente no item 5.3. Na Figura 15 está representada a sequência de filtros aplicados nesta triagem.

Através dos *dockings* dos similares puderam-se selecionar para os sítios de ligação de MTase, 73 compostos para SAM e 44 compostos para GTP.

Na triagem virtual para o domínio RdRP os compostos similares aos inibidores de RdRP de DENV foram submetidos aos filtros de QSAR para ZIKV e *docking molecular*. Uma vez que esses compostos não foram submetidos a nenhuma triagem virtual, precisam ser avaliados nos mesmos filtros anteriores para predição de sua atividade.

A submissão no QSAR fenotípico para ZIKV foi realizado com o auxílio do programa KNIME. Nessa etapa, foram selecionados para os sítios de ligação, 786 compostos para N-pocket, 477 compostos para RNA e 342 compostos para o sítio catalítico. Os *dockings* dos similares foram realizados com os mesmos arquivos de GRID desenvolvidos anteriormente com a ferramenta Glide Dock (SCHRÖDINGER, 2019) do software Maestro. Os conjuntos de ligantes, agora separados por sítios, foram preparados com a ferramenta LigPrep (SCHRÖDINGER, 2018b) para obtenção de, confôrmeros, estados de protonação em pH fisiológico e retirada de sais. Os parâmetros de seleção e corte foram os mesmos utilizados anteriormente no item 5.3. Na figura 15 está representada a sequência dos filtros utilizados nesta triagem.

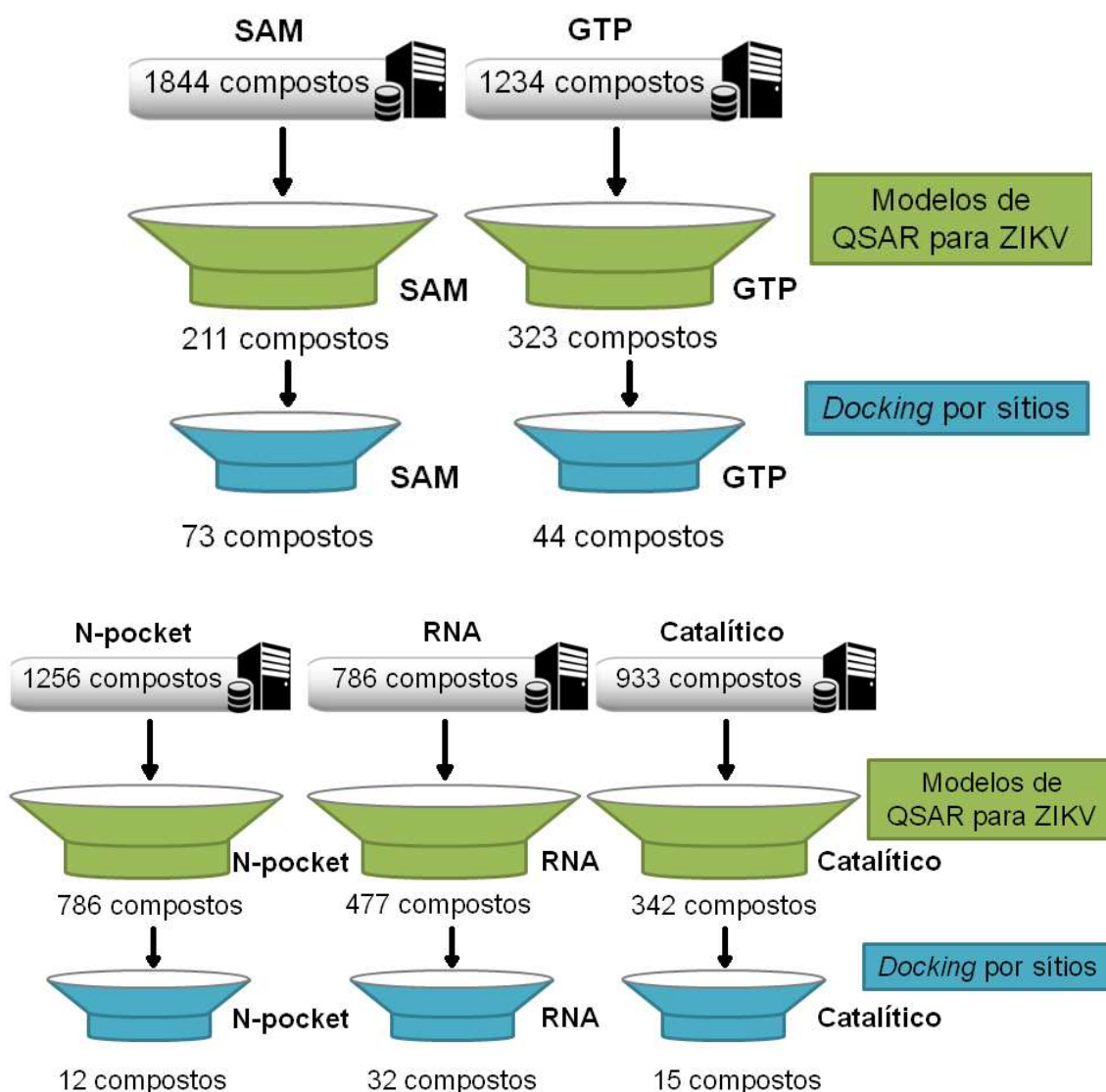


Figura 15: Triagem virtual para a NS5 RdRP de ZIKV nos sítios de RNA, catalítico e *N-pocket*. E para a NS5 MTase nos sítios SAM e GTP.

Através dos *dockings* dos similares puderam-se selecionar para os sítios de ligação da RdRP, 477 compostos para RNA, 786 compostos para N-Pocket e 342 compostos para o sítio Catalítico

Os melhores hits selecionados foram priorizados através do software Assay Central (Collaborations Pharmaceuticals, Inc.). Esse software utiliza modelos bayesianos de aprendizado de máquina para realizar as predições de atividade em ZIKV e de citotoxicidade.

Todos os conjuntos dados separados por sítios de ligação foram submetidos aos modelos bayesianos de aprendizado de máquina e foram priorizados 20 compostos para SAM, 13 compostos para GTP e 08 compostos para RNA.

Todos os *hits* selecionados pelos modelos bayesianos de aprendizado de máquina foram submetidos ao programa nAPOLI, suas interações foram determinadas e comparadas com as interações conservadas de ligantes cristalográficos. Para as proteínas que não continham ligantes cristalográficos, as interações foram priorizadas utilizando as informações a respeito dos resíduos de sítios de ligação disponíveis em literatura.

O programa nAPOLI permite a inspeção automatizada das poses obtidas no *docking*. É possível verificar as interações por resíduos de todo o banco de dados inserido. Desta forma, foi possível visualizar as interações mais presentes de todos os conjuntos de dados por sítio, SAM, GTP e RNA. Como mostrado nas figuras 22, 23 e 24. Para o sítio catalítico, o único composto selecionado, também foi selecionado para o sítio de RNA e com *docking score* melhor, desta forma, sua análise no programa nAPOLI foi realizada apenas no sítio RNA.

Verificamos que os 13 compostos selecionados para o sítio GTP realizaram diversas interações importantes para o complexo proteína-ligante (Figura 16). As interações essenciais (Trp12, Lys13, Ala14, Arg15, Leu16, Asn17, Gln18, Glu23, Met19, Phe24, Tyr25, Tyr27, Lys28, Ser150, Ser151, Ser152) são interações que observamos no substrato GTP (Figura 19 estrutura 13) co-cristalizado com a proteína MTase (PDB ID 5GOZ) e foram utilizadas para priorizar os compostos. Todos os compostos realizaram interações com o resíduo Phe24, um resíduo importante que realiza interações com o substrato cristalográfico GTP. Ademais, os resíduos

importantes que mais interagiram foram Met19, Leu16, Lys28, Lys13 e Ser150 que interagiram respectivamente com 12 (92,31%), 9 (69,23%), 8 (61,54%), 4 (30,77%) e 4 (30,77%) ligantes.

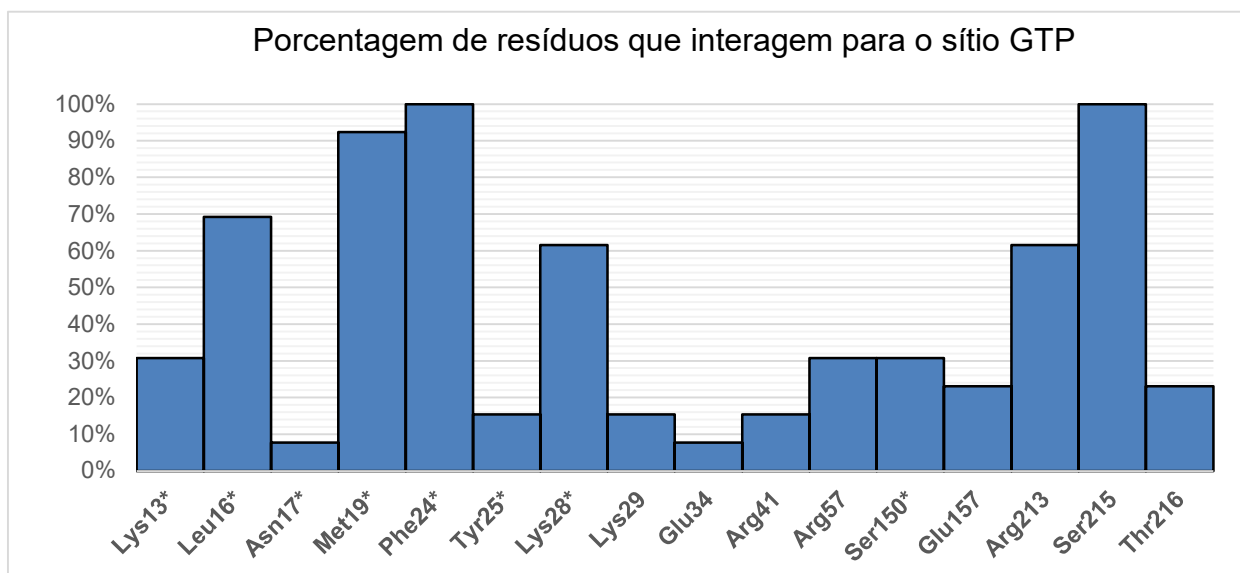


Figura 16: Porcentagem de interações realizadas pelos resíduos do sítio GTP com os compostos. Os resíduos marcados com * fazem interações com os ligantes cristalográficos analisados.

Podemos observar que o grupo de 21 compostos do sítio SAM as interações importantes (Figura 17) (Ser56, Lys61, Cys82, Arg84, Gly86, Trp87, Thr104, Lys105, His110, Glu111, Lys105, Asp131, Val132, Phe133, Cys145, Asp146, Ile147, Gly148, Glu149, Arg160, Arg163, Val164, Lys181, Leu182 e Glu217) para o complexo proteína/ligante foram obtidas através do substrato SAM (Figura 19 estrutura 10), subproduto SAH (Figura 19 estrutura 12), e do inibidor sinefugina (Figura 19 estrutura 11), ambos co-cristalizados e disponíveis sob os códigos PDB ID 5M5B (COUTARD et al., 2017), PDB ID 5WXB (HERCIK et al., 2017) e PDB ID 5MRK (ZHOU et al., 2017). Os resíduos importantes que realizaram a maioria das interações foram Arg84, Glu111, Asp131, Val132 e Asp146 que interagiram com 7 (35%), 19 (95%), 12 (60%), 14 (90%) ligantes respectivamente e foram utilizadas para priorizar os compostos.

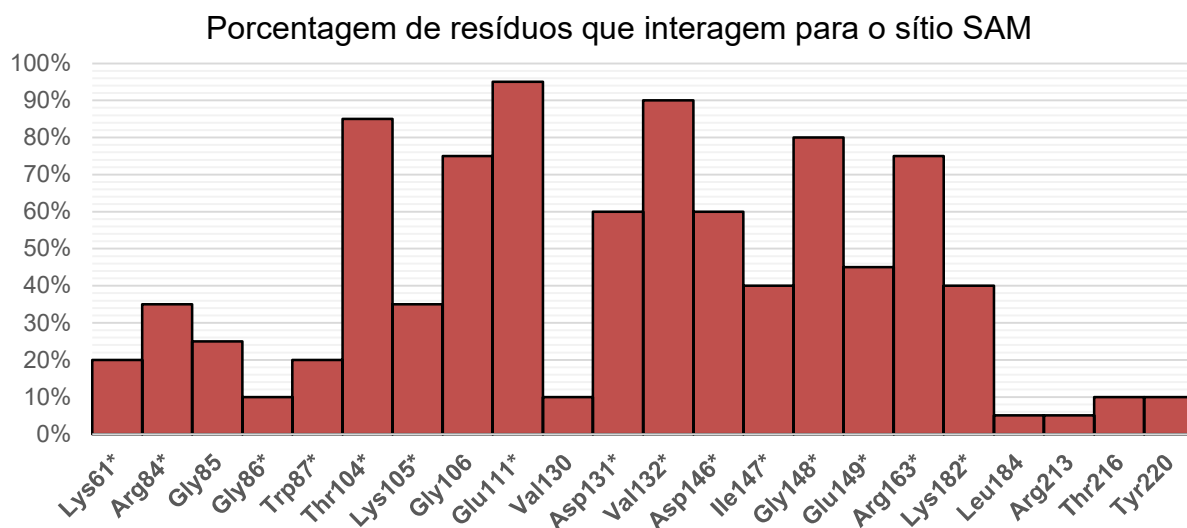


Figura 17: Porcentagem de interações realizadas pelos resíduos do sítio SAM com os compostos. Os resíduos marcados com * fazem interações com os ligantes cristalográficos analisados.

Até o momento não existem ligantes cristalográficos disponíveis na literatura para avaliarmos as interações importantes para o sítio RNA da RdRP de ZIKV. As interações com resíduos que foram considerados importantes (Lys403, Asn494, Phe487, Gln605, Ser603, Arg483, Tyr609, Thr608, Val607, Trp797) para o complexo proteína/ligante foram os utilizados da formação do GRID (DUAN et al., 2017). Os resíduos importantes que mais realizaram interações foram (Figura 18), Tyr609, Thr608, Trp797 e Asn494 que interagiram com 7 (87,5%), 4 (50%), 5 (62,5%) e 2 (25%) ligantes respectivamente e foram utilizadas para priorizar os compostos

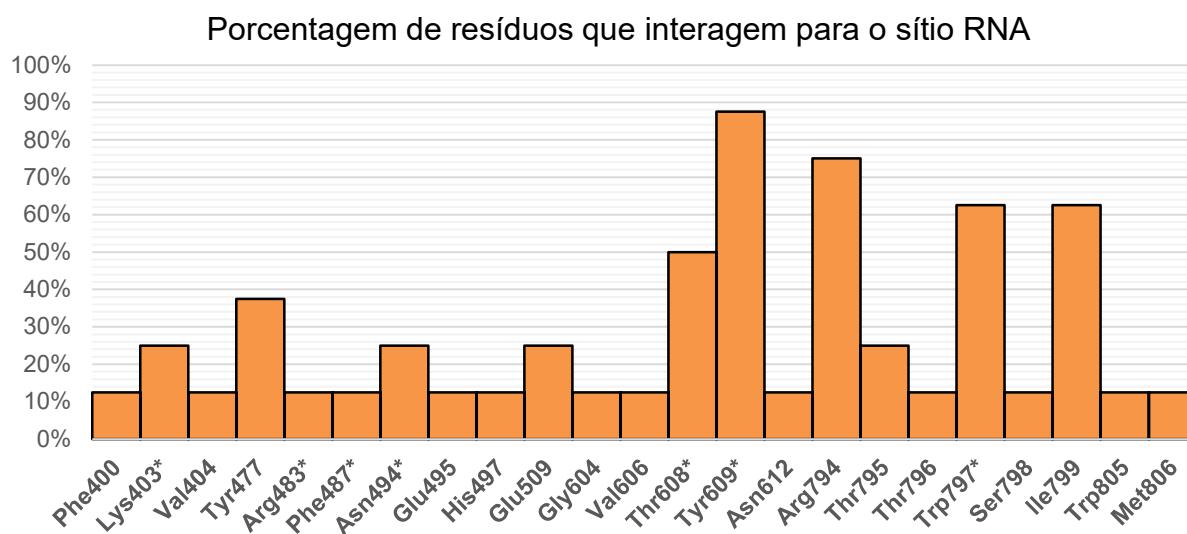


Figura 18: Porcentagem de interações realizadas pelos resíduos do sítio RNA com os compostos. Os resíduos marcados com * fazem interações com os ligantes cristalográficos analisados.

Através das análises no programa nAPOLI foi possível priorizar 05 compostos para o sítio SAM, 03 compostos para GTP e 04 compostos para RNA. Todos os compostos priorizados estão dispostos na tabela 6, juntamente dos valores de *docking score* e o *QSAR consensus score*, que é o valor resultante da votação realizada pelos modelos de QSAR para ZIKV desenvolvidos, sendo 1 o valor máximo resultado da aprovação em todos os modelos. Foi verificado que o único composto selecionado para o sítio catalítico se repetia no sítio de RNA, em função disso, foram selecionados ao todo 14 compostos promissores contra os sítios de ligação dos domínios MTase e RdRP da proteína NS5 de ZIKV. Dispostos na Tabela 3:

Tabela 3: Lista de hits virtuais obtidos a partir da VS utilizando QSAR, *docking* e análises automatizadas .

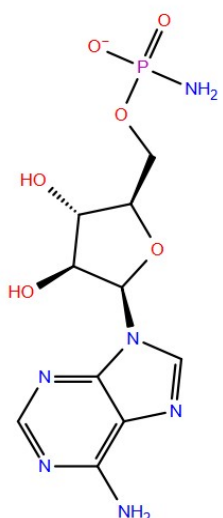
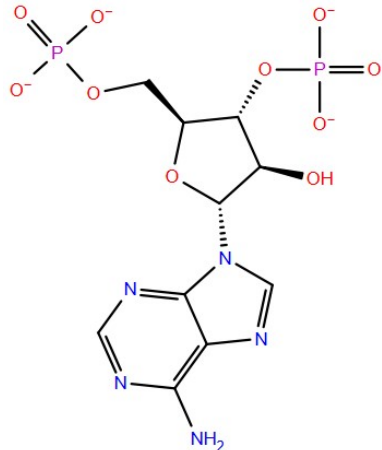
Diagrama 2D de interação	QSAR Consensus Score	Docking Score	eMolecules ID	Sítio
	1	-10,9	46010483	GTP
	1	-10,6	50418361	GTP

Tabela 3: Resultados da inspeção visual baseada em Química Medicinal através do programa nAPOLI para todos os 12 compostos selecionados.

(Continuação)

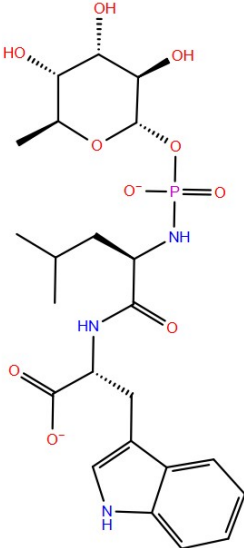
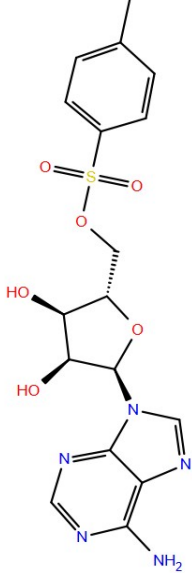
Diagrama 2D de interação	QSAR Consensus Score	Docking Score	eMolecules ID	Sítio
	1	-11,8	313080127	GTP
	0,75	-10,5	94219954	SAM

Tabela 3: Resultados da inspeção visual baseada em Química Medicinal através do programa nAPOLI para todos os 12 compostos selecionados.

(Continuação)

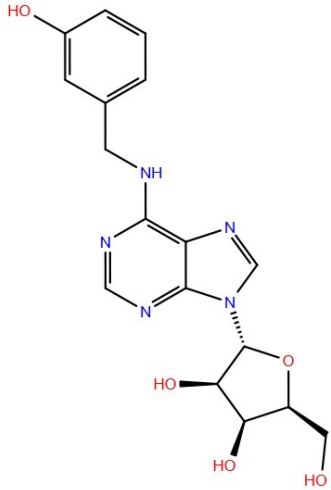
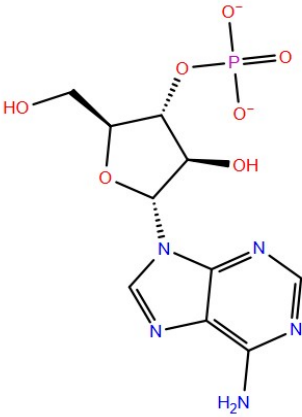
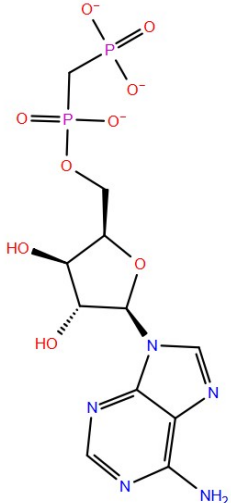
Diagrama 2D de interação	QSAR Consensus Score	Docking Score	eMolecules ID	Sítio
	1	-11,1	106928215	SAM
	0,75	-12,7	43339222	SAM
	0,75	-14,6	313512277	SAM

Tabela 3: Resultados da inspeção visual baseada em Química Medicinal através do programa nAPOLI para todos os 12 compostos selecionados.

(Continuação)

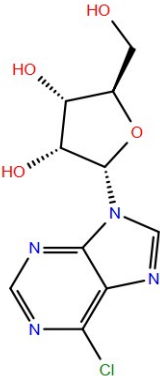
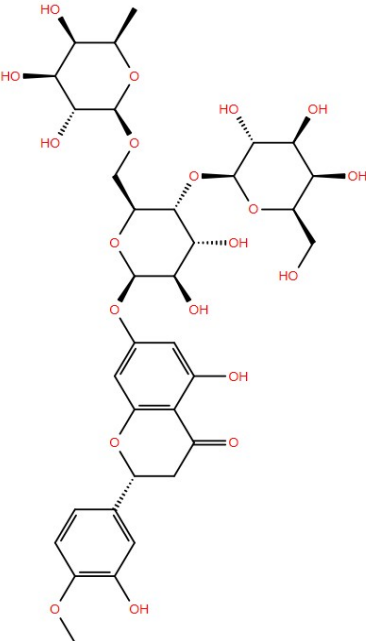
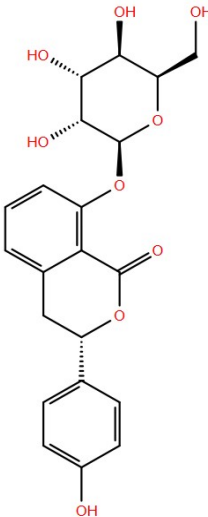
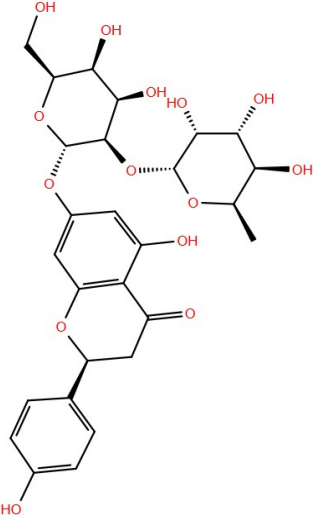
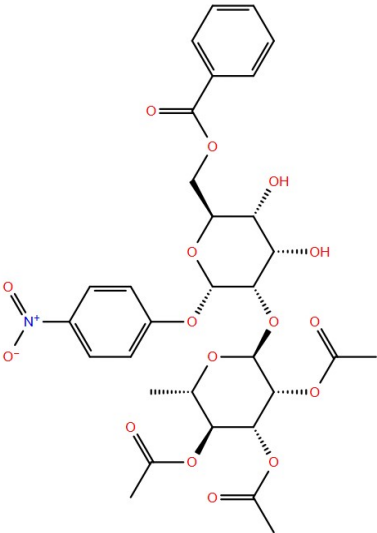
Diagrama 2D de interação	QSAR Consensus Score	Docking Score	eMolecules ID	Sítio
	1	-11,3	33509848	SAM
	1	-14,1	50578712	RNA
	1	-9,4	50153806	RNA

Tabela 3: Resultados da inspeção visual baseada em Química Medicinal através do programa nAPOLI para todos os 12 compostos selecionados.

Diagrama 2D de interação	QSAR Consensus Score	Docking Score	eMolecules ID	Sítio
	1	-11,0	23752469	RNA
	0,75	-10,1	17406067	RNA

A análise baseada em química medicinal revelou que o melhor composto priorizado para o sítio GTP continha as principais interações realizadas entre o substrato e a proteína (Figura 21A-B) visto que, a estrutura selecionada mimetiza a molécula de GTP (Figura 19 estrutura 13) através de semelhanças estruturais como a base nitrogenada presente nas duas moléculas, a presença do grupamento fosfato e o açúcar na extremidade que pode mimetizar a ribose do GTP (Figura 19 estrutura 13) (GEISS et al., 2009). As interações observadas foram, ligações de hidrogênio com os resíduos Asn17, Lys13, Ser215, Lys28, Arg213 e uma interação hidrofóbica com o resíduo Phe24.

Para o sítio SAM o composto (Figura 21C-D) também mimetiza o substrato SAM (Figura 19 estrutura 10) com uma base nitrogenada adenosil e uma ribose em sua estrutura e apresentou três interações importantes descritas no substrato SAM (PDB ID ID 5M5B) (COUTARD et al., 2017) e nos inibidores SAH (Figura 19 estrutura 12) (PDB ID 5WXB) (HERCIK et al., 2017) e sinefugina (Figura 19C) (PDB ID 5MRK) (ZHOU et al., 2017). São elas, ligações de hidrogênio com os resíduos Asp131, Val132 e Glu111. Além disso, uma quarta interação importante foi verificada, com o resíduo Asp146, porem intermediada por uma água cristalográfica.

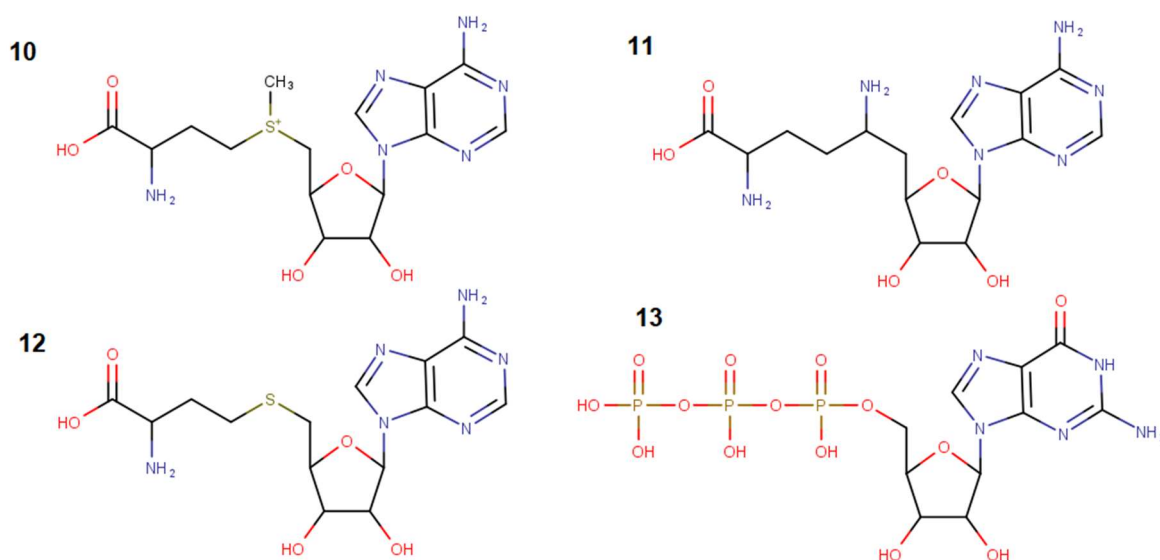


Figura 19: Estrutura dos ligantes cristalográficos presentes nos sítios SAM e GTP. Em 10) o substrato SAM, em 11) o inibidor sinefugina, em 12) o produto SAH, ambos para o sítio SAM. Em 13) o substrato GTP

O composto selecionado para o sítio de RNA (Figura 21E-F) apresenta em sua estrutura um açúcar de seis membros que pode mimetizar uma ribose presente na molécula de RNA e realiza três interações importantes já descritas em literatura com o inibidor NITD107 (Figura 20) de RdRP de DENV (NOBLE et al., 2013). São elas, ligações de hidrogênio com os resíduos Tyr609, Thr608, Trp797, e os mesmos são conservados entre os vírus DENV e ZIKV. É importante ressaltar que a interação com o resíduo Trp797, além de uma ligação de hidrogênio, também é constituída por uma interação do tipo π - π *stacking*.

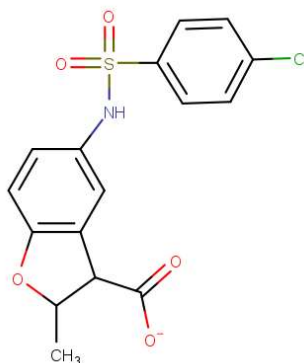


Figura 20: Inibidor de RdRP de DENV NITD104 (NOBLE et al., 2013)

Ao total, foram selecionados 12 compostos, que serão adquiridos na base de dados da eMolecules para posterior validação experimental através de ensaios enzimáticos com a proteína NS5 e ensaios celulares. Esses ensaios serão realizados no Laboratório de Imunofarmacologia IOC/FIOCRUZ, pelo grupo de pesquisa do Dr. Thiago Moreno Souza (FIOCRUZ, Brasil).

6 CONCLUSÕES

Neste trabalho, diferentes métodos computacionais foram integrados para a VS de potenciais inibidores da proteína NS5 de ZIKV. Foi observada uma alta conservação entre os resíduos dos sítios de ligação da NS5 de DENV e ZIKV, possibilitando a triagem de compostos inibidores de NS5 de DENV em modelos computacionais para ZIKV. Modelos de QSAR fenotípicos para ZIKV robustos e preditivos foram construídos e validados, com as seguintes métricas de validação: CCR = 0,91, SE = 0,95, SP = 0,86 e AUC = 0,91. Estes modelos foram utilizados em uma priorização de 145 compostos obtidos de bases de dados de ensaios biológicos, com atividade descrita em literatura contra a proteína NS5 de DENV. Posteriormente os compostos selecionados foram submetidos ao *docking* molecular para os sítios de ligação dos dois domínios da proteína NS5. Na etapa da triagem virtual, os hits selecionados foram utilizados em uma busca por similaridade na base de dados comerciais eMolecules e os compostos similares foram agrupados em bancos de dados separados por sítios. O processo de QSAR e *docking* foi refeito para os compostos similares. Os compostos selecionados na última etapa de *docking* foram submetidos aos modelos bayesianos de aprendizado de máquina para predição da atividade anti-ZIKV e citotoxicidade. E finalmente, os compostos resultantes foram submetidos ao programa nAPOLI para o estudo das interações moleculares proteína-ligantes resultando em 12 compostos promissores similares aos inibidores de NS5 de DENV e com estruturas análogas aos substratos cristalográficos e outros inibidores nucleotídeos e nucleosídeos para os sítios SAM, GTP e RNA.

Através do programa nAPOLI pode-se realizar a comparação dos compostos com outros ligantes cristalográficos disponíveis nas proteínas do PDB utilizadas e verificamos a presença de diversas interações importantes. As interações dos compostos também foram comparadas as de inibidores descritos em literatura e verificamos a presença das interações realizadas com os inibidores. É importante ressaltar que os compostos selecionados serão adquiridos e terão suas atividades avaliadas *in vitro*, em ensaios celulares para ZIKV e enzimáticos contra a NS5, realizados em colaboração.

7 BIBLIOGRAFIA

ALIOTA, M. T. et al. **Zika in the Americas, year 2: What have we learned? What gaps remain? A report from the Global Virus Network Antiviral Research**, 2017. .

ALTMAN, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. **The American Statistician**, v. 46, n. 3, p. 175–185, ago. 1992. Disponível em: <<http://www.tandfonline.com/doi/abs/10.1080/00031305.1992.10475879>>.

ALVES, V. et al. Quimioinformática: Uma Introdução. **Química Nova**, v. 41, n. 2, p. 202–212, 2017.

ATKINSON, B. et al. Detection of Zika Virus. v. 22, n. 5, p. 160107, 2016.

BERTHOLD, M. R. et al. KNIME: The konstanz information miner. **4th International Industrial Simulation Conference 2006, ISC 2006**, v. 11, n. 1, p. 58–61, 2006.

BESNARD, M. et al. Evidence of perinatal transmission of zika virus, French Polynesia, December 2013 and February 2014. **Eurosurveillance**, v. 19, n. 13, p. 8–11, 2014.

BREIMAN, L. RANDOM FORESTS. **Machine Learning**, v. 45, n. 1, p. 5–32, 2001.

BROOIJMANS, N.; KUNTZ, I. D. Molecular Recognition and Docking Algorithms. **Annual Review of Biophysics and Biomolecular Structure**, v. 32, n. 1, p. 335–373, jun. 2003. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.biophys.32.110601.142532>>.

BRUNO M., F. Zika virus-related hypertensive iridocyclitis. **Arquivos Brasileiros de Oftalmologia**, v. 387, n. 10015, p. 63, 2016. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0004-27492016000100020&lang=es>.

CAO-LORMEAU, V. M. et al. Guillain-Barré Syndrome outbreak associated with Zika virus infection in French Polynesia: A case-control study. **The Lancet**, 2016.

CDC. **About 1 in 10 U.S. pregnant women with confirmed Zika infection had a fetus or baby with birth defects in 2016**. Disponível em: <<https://www.cdc.gov/media/releases/2017/p0404-zika-pregnancy.html>>.

CHERKASOV, A. et al. QSAR Modeling: Where have you been? Where are you going to? **J Med Chem**, v. 57, p. 4977–5010, 2014.

COLOMA, J. et al. Structures of NS5 Methyltransferase from Zika Virus. **Cell Reports**, v. 16, n. 12, p. 3097–3102, 2016.

COSTELLO, A. et al. Defining the syndrome associated with congenital Zika virus infection. **Bulletin of the World Health Organization**, 2016.

COUTARD, B. et al. Assessment of Dengue virus helicase and methyltransferase as targets for fragment-based drug discovery. **Antiviral Research**, v. 106, n. 1, p. 61–70, 2014. Disponível em: <<http://dx.doi.org/10.1016/j.antiviral.2014.03.013>>.

COUTARD, B. et al. Zika Virus Methyltransferase: Structure and Functions for Drug Design Perspective. **Journal of Virology**, v. 91, n. 5, p. e02202-16, 2017.

CRAMER, R. D. The inevitable QSAR renaissance. **Journal of Computer-Aided Molecular Design**, v. 26, n. 1, p. 35–38, 30 jan. 2012. Disponível em: <<http://link.springer.com/10.1007/s10822-011-9495-0>>.

DALECKI, A. G. et al. High-throughput screening and Bayesian machine learning for copper-dependent inhibitors of *Staphylococcus aureus*. **Metalomics : integrated biometal science**, v. 11, n. 3, p. 696–706, mar. 2019.

DE FREITAS, C. S. et al. Yellow fever virus is susceptible to sofosbuvir both in vitro and in vivo. **PLOS Neglected Tropical Diseases**, v. 13, n. 1, p. e0007072, 30 jan. 2019. Disponível em: <<https://dx.plos.org/10.1371/journal.pntd.0007072>>.

DELANO, W. L. Pymol: An open-source molecular graphics tool. **CCP4 Newsletter On Protein Crystallography**, v. 40, p. 82–92, 2002.

DESJARLAIS, R. L. et al. Docking flexible ligands to macromolecular receptors by molecular shape. **Journal of Medicinal Chemistry**, v. 29, n. 11, p. 2149–2153, nov. 1986. Disponível em: <<https://pubs.acs.org/doi/abs/10.1021/jm00161a004>>.

DETERING, C.; VARANI, G. Validation of Automated Docking Programs for Docking and Database Screening against RNA Drug Targets. **Journal of Medicinal Chemistry**, v. 47, n. 17, p. 4188–4201, ago. 2004. Disponível em: <<https://pubs.acs.org/doi/10.1021/jm030650o>>.

DI NOLA, A.; ROCCATANO, D.; BERENDSEN, H. J. C. Molecular dynamics simulation of the docking of substrates to proteins. **Proteins: Structure, Function, and Genetics**, v. 19, n. 3, p. 174–182, jul. 1994. Disponível em: <<http://doi.wiley.com/10.1002/prot.340190303>>.

DICK, G. W. A. EPIDEMIOLOGICAL NOTES ON SOME VIRUSES ISOLATED IN UGANDA. **National Institute for Medical Research**, v. 67, n. 1944, p. 115, 1953.

DING, Y. et al. Synthesis of 2'- β -C-methyl toyocamycin and sangivamycin analogues as potential HCV inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 15, n. 3, p. 725–727, fev. 2005. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0960894X0401368X>>.

DIRLIKOV, E. et al. Guillain-Barré syndrome during ongoing Zika virus transmission — Puerto Rico, January 1–July 31, 2016. **Morbidity and Mortality Weekly Report**, v. 65, n. 34, p. 910–914, 2016.

DUAN, W. et al. The crystal structure of Zika virus NS 5 reveals conserved drug targets. **The EMBO Journal**, v. 36, n. 7, p. 919–933, 2017.

DUFFY, B. C. et al. Early phase drug discovery: Cheminformatics and computational techniques in identifying lead series. **Bioorganic & Medicinal Chemistry**, v. 20, n. 18, p. 5324–5342, set. 2012. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0968089612003598>>.

DUFFY, M. R. et al. Zika virus outbreak on Yap Island, Federated States of Micronesia. **New England Journal of Medicine**, v. 360, n. 24, p. 2536–2543, 2009.

DURANT, J. L. et al. Reoptimization of MDL Keys for Use in Drug Discovery. **Journal of Chemical Information and Computer Sciences**, v. 42, n. 6, p. 1273–1280, nov. 2002. Disponível em: <<https://pubs.acs.org/doi/10.1021/ci010132r>>.

DURRANT, J. D.; MCCAMMON, J. A. BINANA: A novel algorithm for ligand-binding characterization. **Journal of Molecular Graphics and Modelling**, v. 29, n. 6, p. 888–893, abr.

2011. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S1093326311000076>>.

EKINS, S. et al. Molecular characterization of CYP2B6 substrates. **Current drug metabolism**, v. 9, n. 5, p. 363–73, jun. 2008.

EKINS, S. et al. Analysis and hit filtering of a very large library of compounds screened against *Mycobacterium tuberculosis*. **Molecular bioSystems**, v. 6, n. 11, p. 2316–2324, nov. 2010.

EKINS, S.; CLARK, A. M. **ASSAY CENTRAL**. Disponível em: <<https://www.collaborationspharma.com/assay-central>>. Acesso em: 20 jul. 2020a.

EKINS, S.; CLARK, A. M. **Assay Central**.

ELDRIDGE, M. D. et al. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. **Journal of Computer-Aided Molecular Design**, v. 11, n. 5, p. 425–445, 1997.

eMolecules. **Choice Reviews Online**, 2006.

EUROPEAN CENTRE FOR DISEASE PREVENTION AND CONTROL. **Epidemiological update: Outbreaks of Zika virus and complications potentially linked to the Zika virus infection**. Disponível em: <<https://www.ecdc.europa.eu/en/news-events/epidemiological-update-outbreaks-zika-virus-and-complications-potentially-linked-zika-1>>.

FABRI, A. et al. CGAL - The Computational Geometry Algorithms Library. p. 6, 2011.

FASSIO, A. V. et al. nAPOLI: a graph-based strategy to detect and visualize conserved protein-ligand interactions in large-scale. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, p. 1–1, 2019.

FOURCHES, D.; MURATOV, E.; ALEXANDER, T. Trust, But Verify II: A Practical Guide to Chemogenomics Data Curation. **J Chem Inf Model**, v. 176, n. 1, p. 1243–1252, 2016.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Trust, but verify: On the importance of chemical structure curation in cheminformatics and QSAR modeling research. **Journal of Chemical Information and Modeling**, v. 50, n. 7, p. 1189–1204, 2010.

FOURCHES, D.; MURATOV, E.; TROPSHA, A. Curation of chemogenomics data. **Nature Chemical Biology**, v. 11, n. 8, p. 535, jul. 2015.

GAULTON, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. **Nucleic acids research**, v. 40, n. Database issue, p. D1100-7, jan. 2012.

GEDECK, P.; ROHDE, B.; BARTELS, C. QSAR – How Good Is It in Practice? Comparison of Descriptor Sets on an Unbiased Cross Section of Corporate Data Sets. **Journal of Chemical Information and Modeling**, v. 46, n. 5, p. 1924–1936, 3 set. 2006. Disponível em: <<https://pubs.acs.org/doi/10.1021/ci050413p>>.

GEISS, B. J. et al. Analysis of Flavivirus NS5 Methyltransferase Cap Binding. **Journal of Molecular Biology**, v. 385, n. 5, p. 1643–1654, fev. 2009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0022283608014939>>.

GODOY, A. S. et al. Crystal structure of Zika virus NS5 RNA-dependent RNA polymerase. **Nature Communications**, v. 8, p. 1–6, 2017.

GOLBRAIKH, A. et al. Rational selection of training and test sets for the development of validated QSAR models. **Journal of Computer-Aided Molecular Design**, v. 17, n. 2–4, p. 241–253, 2003.

GOLBRAIKH, A. et al. Data Set Modelability by QSAR. **Journal of Chemical Information and Modeling**, v. 54, n. 1, p. 1–4, 27 jan. 2014. Disponível em: <<https://pubs.acs.org/doi/10.1021/ci400572x>>.

GOODSELL, D. S. et al. Automated docking in crystallography: Analysis of the substrates of aconitase. **Proteins: Structure, Function, and Genetics**, v. 17, n. 1, p. 1–10, set. 1993. Disponível em: <<http://doi.wiley.com/10.1002/prot.340170104>>.

GUIDO, R. V. C.; OLIVA, G.; ANDRICOPULO, A. D. Virtual screening and its integration with modern drug design technologies. **Current Medicinal Chemistry**, v. 15, n. 1, p. 37–46, 2008.

GULLAND, A. Zika virus may be linked to several birth defects, expert warns. **BMJ (Clinical research ed.)**, v. 352, n. March, p. i1322, 2016. Disponível em: <<http://dx.doi.org/doi:10.1136/bmj.i1322>>.

HANN, M.; GREEN, R. Chemoinformatics — a new name for an old problem? **Current Opinion in Chemical Biology**, v. 3, n. 4, p. 379–383, ago. 1999. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S136759319980057X>>.

HART, T. N.; READ, R. J. A multiple-start Monte Carlo docking method. **Proteins: Structure, Function, and Genetics**, v. 13, n. 3, p. 206–222, jul. 1992. Disponível em: <<http://doi.wiley.com/10.1002/prot.340130304>>.

HAZIN, A. N. et al. **Computed tomographic findings in microcephaly associated with Zika Virus** *New England Journal of Medicine* Massachusetts Medical Society, , jun. 2016. .

HENRY, C. R. et al. Chorioretinal Lesions Presumed Secondary to Zika Virus Infection in an Immunocompromised Adult. **JAMA Ophthalmology**, v. 135, n. 4, p. 386–389, 1 abr. 2017. Disponível em: <<https://doi.org/10.1001/jamaophthalmol.2017.0098>>.

HERCIK, K. et al. Structural basis of Zika virus methyltransferase inhibition by sinefungin. **Archives of Virology**, v. 162, n. 7, p. 2091–2096, 29 jul. 2017. Disponível em: <<http://link.springer.com/10.1007/s00705-017-3345-x>>.

HONEIN, M. A. et al. Birth defects among fetuses and infants of US women with evidence of possible zika virus infection during pregnancy. **JAMA - Journal of the American Medical Association**, v. 317, n. 1, p. 59–68, 2017.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. Sartorius products. **Journal of molecular graphics**, v. 14, n. October 1995, p. 33–38, 1996. Disponível em: <<https://www.tapbiosystems.com/tap/products/index.htm>>.

JORGENSEN, W. L. The Many Roles of Computation in Drug Discovery. **Science**, v. 303, n. 5665, p. 1813–1818, 2004.

JORGENSEN, W. L. Efficient Drug Lead Discovery and Optimization 2008.pdf. **Accounts of Chemical Research**, v. 42, n. 6, p. 724–733, 2009.

JUNIOR, P. et al. Vírus Zika: Revisão para Clínicos Zika Virus: A Review to Clinicians. **Ordem dos médicos**, v. 1, n. 6, p. 7, 2015.

KESSEL, M. The problems with today's pharmaceutical business—an outsider's view. **Nature Biotechnology**, v. 29, n. 1, p. 27–33, 1 jan. 2011. Disponível em: <<http://www.nature.com/articles/nbt.1748>>.

KIM, S. et al. PubChem substance and compound databases. **Nucleic Acids Research**, 2016.

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. Disponível em: <<http://www.nature.com/articles/nrd1549>>.

KLEINERT, R. D. V. et al. Yellow Fever: Integrating Current Knowledge with Technological Innovations to Identify Strategies for Controlling a Re-Emerging Virus. **Viruses**, v. 11, n. 10, p. 960, 17 out. 2019. Disponível em: <<https://www.mdpi.com/1999-4915/11/10/960>>.

KODATI, S. et al. Bilateral posterior uveitis associated with Zika virus infection. **The Lancet**, v. 389, n. 10064, p. 125–126, 2017. Disponível em: <[http://dx.doi.org/10.1016/S0140-6736\(16\)32518-1](http://dx.doi.org/10.1016/S0140-6736(16)32518-1)>.

KUIVANEN, S. et al. Obatoclox, saliphenylhalamide and gemcitabine inhibit Zika virus infection in vitro and differentially affect cellular signaling, transcription and metabolism. **Antiviral Research**, v. 139, p. 117–128, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.antiviral.2016.12.022>>.

KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions. **Journal of Molecular Biology**, v. 161, n. 2, p. 269–288, out. 1982. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/002228368290153X>>.

LE POGAM, S. et al. Selection and Characterization of Replicon Variants Dually Resistant to Thumb- and Palm-Binding Nonnucleoside Polymerase Inhibitors of the Hepatitis C Virus. **Journal of Virology**, v. 80, n. 12, p. 6146–6154, 15 jun. 2006. Disponível em: <<https://jvi.asm.org/content/80/12/6146>>.

LEACH, A. R.; KUNTZ, I. D. Conformational analysis of flexible ligands in macromolecular receptor sites. **Journal of Computational Chemistry**, v. 13, n. 6, p. 730–748, jul. 1992. Disponível em: <<http://doi.wiley.com/10.1002/jcc.540130608>>.

LI, M. Z. I. et al. Oral Susceptibility of Singapore Aedes (Stegomyia) aegypti (Linnaeus) to Zika Virus. **PLoS Neglected Tropical Diseases**, v. 6, n. 8, 2012.

LIM, S. P. et al. Potent Allosteric Dengue Virus NS5 Polymerase Inhibitors: Mechanism of Action and Resistance Profiling. **PLoS Pathogens**, v. 12, n. 8, p. 1–25, 2016.

MACALINO, S. J. Y. et al. Role of computer-aided drug design in modern drug discovery. **Archives of Pharmacal Research**, v. 38, n. 9, p. 1686–1701, 25 set. 2015. Disponível em: <<http://link.springer.com/10.1007/s12272-015-0640-5>>.

MACNAMARA, F. N. Zika virus: A report on three cases of human infection during an epidemic of jaundice in Nigeria. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 48, n. 2, p. 139–145, 1954.

MARCHETTE, N. J.; GARCIA, R.; RUDNICK, A. Isolation of Zika Virus from Aedes Aegypti Mosquitoes in Malaysia. **The American Society of Tropical Medicine and Hygiene**, v. 8, p. 411–415, 1969. Disponível em: <<http://www.ajtmh.org/content/journals/10.4269/ajtmh.1969.18.411>>.

MARSHALL, G. R. Computer-Aided Drug Design. **Annual Review of Pharmacology and Toxicology**, v. 27, n. 1, p. 193–213, abr. 1987. Disponível em: <<http://www.annualreviews.org/doi/10.1146/annurev.pa.27.040187.001205>>.

MARTÍNEZ, L.; ANDREANI, R.; MARTÍNEZ, J. M. Convergent algorithms for protein structural alignment. **BMC Bioinformatics**, v. 8, n. 1, p. 306, 22 dez. 2007. Disponível em: <<https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-8-306>>.

MENG, X.-Y. et al. Curr Comput Aided Drug Des. **Bone**, v. 23, n. 1, p. 1–7, 2014.

MINISTÉRIO DA SAÚDE. Boletim Epidemiológico. v. 51, p. 45, 2020. Disponível em: <<https://www.saude.gov.br/images/pdf/2020/July/14/Boletim-epidemiologico-SVS-28-v2.pdf>>.

MORGAN, H. L. The Generation of a Unique Machine Description for Chemical Structures—A Technique Developed at Chemical Abstracts Service. **Journal of Chemical Documentation**, v. 5, n. 2, p. 107–113, 1965.

MORRIS, G. M.; LIM-WILBY, M. Molecular Docking. In: [s.l: s.n.]p. 365–382.

MOTTIN, M. et al. The A – Z of Zika drug discovery. **Drug Discovery Today**, v. 23, n. 11, p. 1833–1847, 2018. Disponível em: <<https://doi.org/10.1016/j.drudis.2018.06.014>>.

MUEGGE, I. A knowledge-based scoring function for protein-ligand interactions: Probing the reference state. **Perspectives in Drug Discovery and Design**, v. 20, p. 99–114, 2000.

MUSSO, D. et al. Potential sexual transmission of Zika virus. **Emerging infectious diseases**, v. 21, 2015.

NAKAJIMA, N. et al. Flexible docking of a ligand peptide to a receptor protein by multicanonical molecular dynamics simulation. **Chemical Physics Letters**, v. 278, n. 4–6, p. 297–301, out. 1997. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0009261497010749>>.

NEUFELDT, C. J. et al. Rewiring cellular networks by members of the Flaviviridae family. **Nature Reviews Microbiology**, v. 16, n. 3, p. 125–142, 12 mar. 2018. Disponível em: <<http://www.nature.com/articles/nrmicro.2017.170>>.

NOBLE, C. G. et al. Conformational Flexibility of the Dengue Virus RNA-Dependent RNA Polymerase Revealed by a Complex with an Inhibitor. **Journal of Virology**, v. 87, n. 9, p. 5291–5295, 1 maio 2013. Disponível em: <<https://jvi.asm.org/content/87/9/5291>>.

NÓBREGA, M. E. B. da et al. Surto de síndrome de Guillain-Barré possivelmente relacionado à infecção prévia pelo vírus Zika, Região Metropolitana do Recife, Pernambuco, Brasil, 2015. **Epidemiologia e serviços de saúde : revista do Sistema Unico de Saude do Brasil**, v. 27, n. 2, p. e2017039, 2018.

OECD. Guidance Document on the Validation of (Quantitative) Structure-Activity Relationship [(Q)Sar] Models. **Transport**, v. 2, n. February, p. 1–154, 2007.

OLSON, J. G. et al. Zika virus, a cause of fever in central java, indonesia. **Transactions of the Royal Society of Tropical Medicine and Hygiene**, v. 75, n. 3, p. 389–393, 1981.

OOMS, F. Molecular Modeling and Computer Aided Drug Design. Examples of their Applications in Medicinal Chemistry. **Current Medicinal Chemistry**, v. 7, n. 2, p. 141–158, 1 fev. 2000. Disponível em:

<<http://www.eurekaselect.com/openurl/content.php?genre=article&issn=0929-8673&volume=7&issue=2&spage=141>>.

PELLICCIA, S. et al. Inhibition of dengue virus replication by novel inhibitors of RNA-dependent RNA polymerase and protease activities. **Journal of Enzyme Inhibition and Medicinal Chemistry**, v. 32, n. 1, p. 1091–1101, 1 jan. 2017. Disponível em: <<https://www.tandfonline.com/doi/full/10.1080/14756366.2017.1355791>>.

PERRYMAN, A. L. et al. Virtual screening with AutoDock Vina and the common pharmacophore engine of a low diversity library of fragments and hits against the three allosteric sites of HIV integrase: participation in the SAMPL4 protein–ligand binding challenge. **Journal of Computer-Aided Molecular Design**, v. 28, n. 4, p. 429–441, 4 abr. 2014. Disponível em: <<http://link.springer.com/10.1007/s10822-014-9709-3>>.

PETRIE, M. et al. Cross-reactivity studies and predictive modeling of “Bath Salts” and other amphetamine-type stimulants with amphetamine screening immunoassays. **Clinical toxicology (Philadelphia, Pa.)**, v. 51, n. 2, p. 83–91, fev. 2013.

PLEWCZYNSKI, D. Brainstorming: weighted voting prediction of inhibitors for protein targets. **Journal of Molecular Modeling**, v. 17, n. 9, p. 2133–2141, 21 set. 2011. Disponível em: <<http://link.springer.com/10.1007/s00894-010-0854-x>>.

POLGAR, T.; KESERU, G. Integration of Virtual and High Throughput Screening in Lead Discovery Settings. **Combinatorial Chemistry & High Throughput Screening**, v. 14, n. 10, p. 889–897, 1 dez. 2011. Disponível em: <<http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1386-2073&volume=14&issue=10&spage=889>>.

POWDRILL, M. H.; BERNATCHEZ, J. A.; GÖTTE, M. Inhibitors of the Hepatitis C Virus RNA-Dependent RNA Polymerase NS5B. **Viruses**, v. 2, n. 10, p. 2169–2195, 28 set. 2010. Disponível em: <<http://www.mdpi.com/1999-4915/2/10/2169>>.

RASHID, U. et al. Structure based virtual screening-driven identification of monastrol as a potent urease inhibitor. **Journal of Molecular Graphics and Modelling**, v. 43, p. 47–57, 2013. Disponível em: <<http://dx.doi.org/10.1016/j.jmgm.2013.04.006>>.

REYNOLDS, C. Impact of Computational Structure-based Methods on Drug Discovery. **Current Pharmaceutical Design**, v. 20, n. 20, p. 3380–3386, 31 maio 2014. Disponível em: <<http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1381-6128&volume=20&issue=20&spage=3380>>.

RODRIGUES-ALVES, M. L. et al. Historical Perspective and Biotechnological Trends to Block Arboviruses Transmission by Controlling Aedes aegypti Mosquitos Using Different Approaches. **Frontiers in Medicine**, v. 7, 23 jun. 2020. Disponível em: <<https://www.frontiersin.org/article/10.3389/fmed.2020.00275/full>>.

ROGERS, D.; HAHN, M. Extended-Connectivity Fingerprints. **J. Chem. Inf. Model**, p. 742–754, 2010.

RONG, Z. et al. A CRISPR screen defines a signal peptide processing pathway required by flaviviruses. **J Autism Dev Disord**, v. 47, n. 3, p. 549–562, 2017.

RUSSO, D. P. et al. Comparing Multiple Machine Learning Algorithms and Metrics for Estrogen Receptor Binding Prediction. **Molecular pharmaceuticals**, v. 15, n. 10, p. 4361–4370, 2018.

SACRAMENTO, C. Q. et al. The clinically approved antiviral drug sofosbuvir inhibits Zika virus replication. **Scientific Reports**, v. 7, n. 1, p. 40920, 18 mar. 2017. Disponível em: <<http://www.nature.com/articles/srep40920>>.

SCHNEIDER, G. Virtual screening: an endless staircase? **Nature Reviews Drug Discovery**, v. 9, n. 4, p. 273–276, abr. 2010. Disponível em: <<http://www.nature.com/articles/nrd3139>>.

SCHRAM, P. C. F. Zika virus and public health. **Journal of Human Growth and Development**, v. 26, n. 1, p. 7–8, 2016.

SCHRÖDINGER. **ProteinPrep | SchrödingerSchrödinger Release 2018-2**, 2018a. .

SCHRÖDINGER. **LigPrep | SchrödingerSchrödinger Release 2018-2**, 2018b. .

SCHRÖDINGER, L. **Schrödinger Release 2019-4: Glide**New York, NY, 2019. .

SHAW, A. N. et al. Substituted benzothiadiazine inhibitors of Hepatitis C virus polymerase. **Bioorganic & Medicinal Chemistry Letters**, v. 19, n. 15, p. 4350–4353, ago. 2009. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0960894X09007847>>.

SILVA, A. D. C. E. **Planejamento e identificação in silico de novos candidatos a protótipos de fármacos antitumorais**. 2015. Universidade Federal de Goiás, 2015. Disponível em: <<https://repositorio.bc.ufg.br/tede/handle/tede/5989>>.

SOFIA, M. J. et al. Nucleoside, nucleotide, and non-nucleoside inhibitors of hepatitis C virus NS5B RNA-dependent RNA-polymerase. **Journal of Medicinal Chemistry**, v. 55, n. 6, p. 2481–2531, 2012.

SOUSA, B. K. P. **Estratégias computacionais e experimentais para identificação de inibidores das proteínas helicase e protease dos vírus Dengue e Zika**. 2020. Dissertação (Mestrado em Ciências Farmacêuticas) – Programa De Pós-Graduação Em Ciências Farmacêuticas, Universidade Federal de Goiás, Goiânia, 2020.

SVETNIK, V. et al. Random forest: a classification and regression tool for compound classification and QSAR modeling. **Journal of chemical information and computer sciences**, v. 43, n. 6, p. 1947–58, 2003.

TEIXEIRA, G. A. et al. Analysis of the concept of the zika virus congenital syndrome. **Ciencia e Saude Coletiva**, v. 25, n. 2, p. 567–574, 2020.

TODESCHINI, R. CONSONNI, V. **Molecular descriptors for chemoinformatics**. 2. ed. [s.l.] Wiley-VCH Verlag GmbH & Co. KGaA, 2009.

TODESCHINI, R.; CONSONNI, V. Handbook of molecular descriptors. **Methods and Principles in Medicinal Chemistry Series**, v. 11, p. 11134–11136, 2000.

TROPSHA, A. Best Practices for QSAR Model Development, Validation, and Exploitation. **Molecular Informatics**, v. 29, n. 6–7, p. 476–488, jul. 2010.

TROPSHA, A.; GRAMATICA, P.; GOMBAR, V. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. **QSAR & Combinatorial Science**, v. 22, n. 1, p. 69–77, abr. 2003. Disponível em: <<http://doi.wiley.com/10.1002/qsar.200390007>>.

UPADHYAY, A. K. et al. Crystal structure of full-length Zika virus NS5 protein reveals a

conformation similar to Japanese encephalitis virus NS5. **Acta Crystallographica Section F Structural Biology Communications**, v. 73, n. 3, p. 116–122, 1 mar. 2017. Disponível em: <<http://scripts.iucr.org/cgi-bin/paper?S2053230X17001601>>.

VARAPRASAD, C. V. N. S. et al. Synthesis of pyrrolo[2,3-d]pyrimidine nucleoside derivatives as potential anti-HCV agents. **Bioorganic Chemistry**, v. 35, n. 1, p. 25–34, fev. 2007. Disponível em: <<https://linkinghub.elsevier.com/retrieve/pii/S0045206806000563>>.

VENTURA, C. V. et al. Ophthalmological findings in infants with microcephaly and presumable intra-uterus Zika virus infection. **Arquivos Brasileiros de Oftalmologia**, v. 79, n. 1, p. 1–3, 2016.

VERLI, H. **Bioinformática da Biologia à Flexibilidade Molecular**. [s.l.: s.n.]v. 53

WANG, P. F. et al. Halogen Substitution Influences Ketamine Metabolism by Cytochrome P450 2B6: In Vitro and Computational Approaches. **Molecular Pharmaceutics**, 2019.

WELLING, M. **A First Encounter with Machine Learning Max**. 1. ed. [s.l.] CA.: University of California, 2011.

WORLD HEALTH ORGANIZATION. **Zika virus outbreaks in the Americas**, 2015. . (Nota técnica).

WORLD HEALTH ORGANIZATION. Countries and territories with current or previous Zika virus transmission. 2019.

WORLD HEALTH ORGANIZATION. **Prioritizing diseases for research and development in emergency contexts**. Disponível em: <<https://www.who.int/activities/prioritizing-diseases-for-research-and-development-in-emergency-contexts>>.

XIA, S. et al. Inhibition of SARS-CoV-2 (previously 2019-nCoV) infection by a highly potent pan-coronavirus fusion inhibitor targeting its spike protein that harbors a high capacity to mediate membrane fusion. **Cell Research**, v. 30, n. 4, p. 343–355, 30 abr. 2020. Disponível em: <<http://www.nature.com/articles/s41422-020-0305-x>>.

XU, H.-T. et al. Evaluation of Sofosbuvir (β -D-2'-deoxy-2'- α -fluoro-2'- β -C-methyluridine) as an inhibitor of Dengue virus replication #. **Scientific Reports**, v. 7, n. 1, p. 6345, 24 dez. 2017. Disponível em: <<http://www.nature.com/articles/s41598-017-06612-2>>.

XUE, L.; BAJORATH, J. Molecular Descriptors in Chemoinformatics, Computational Combinatorial Chemistry, and Virtual Screening. **Combinatorial Chemistry & High Throughput Screening**, v. 3, n. 5, p. 363–372, 1 out. 2000. Disponível em: <<http://www.eurekaselect.com/openurl/content.php?genre=article&issn=1386-2073&volume=3&issue=5&page=363>>.

YASUDA, K. et al. Drug transporters on arachnoid barrier cells contribute to the blood-cerebrospinal fluid barrier. **Drug metabolism and disposition: the biological fate of chemicals**, v. 41, n. 4, p. 923–31, abr. 2013.

YOUNG, D. C. **Computational Drug Design**. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2009.

ZHANG, C. et al. Structure of the NS5 methyltransferase from Zika virus and implications in inhibitor design. **Biochemical and Biophysical Research Communications**, v. 492, n. 4, p. 624–630, 2017. Disponível em: <<https://doi.org/10.1016/j.bbrc.2016.11.098>>.

ZHAO, B. et al. Structure and function of the Zika virus full-length NS5 protein. **Nature Communications**, v. 8, p. 14762, 2017. Disponível em: <<http://dx.doi.org/10.1038/ncomms14762>>.

ZHOU, H. et al. The conformational changes of Zika virus methyltransferase upon converting SAM to SAH. **Oncotarget**, v. 8, n. 9, p. 14830–14834, 2017.

ZHU, H. et al. Big Data in Chemical Toxicity Research: The Use of High-Throughput Screening Assays To Identify Potential Toxicants. **Chemical Research in Toxicology**, v. 27, n. 10, p. 1643–1651, 20 out. 2014. Disponível em: <<https://pubs.acs.org/doi/10.1021/tx500145h>>.

ZORN, K. M. et al. Multiple Machine Learning Comparisons of HIV Cell-based and Reverse Transcriptase Data Sets. **Molecular pharmaceuticals**, v. 16, n. 4, p. 1620–1632, abr. 2019.