UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

ANACLETO SILVA DE SOUZA

Estudos em quimioinformática baseados na estrutura do ligante e do receptor de
candidatos a agentes antichagásicos

São Carlos
2019

ANACLETO SILVA DE SOUZA

Estudos em quimioinformática baseados na estrutura do ligante e do receptor de
candidatos a agentes antichagásicos

Tese apresentada ao programa de pós-
graduação em Física do Instituto de Física
de São Carlos da Universidade de São
Paulo para obtenção do título de Doutor
em Ciências

Área de concentração: Física Aplicada
Opção: Biomolecular
Orientador: Prof. Dr. Adriano D.
Andricopulo

Versão corrigida

(versão original disponível na Unidade que aloja o Programa)

São Carlos

2019

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Souza, Anacleto Silva de
Estudos em quimioinformática baseados na estrutura do
ligante e do receptor de candidatos a agentes
antichagásicos / Anacleto Silva de Souza; orientador
Adriano Defini Andricopulo - versão corrigida -- São
Carlos, 2019.
196 p.

Tese (Doutorado - Programa de Pós-Graduação em Física
Aplicada Biomolecular) -- Instituto de Física de São
Carlos, Universidade de São Paulo, 2019.

1. Inteligência artificial. 2. Trypanosoma cruzi. 3.
Doença de Chagas. 4. Cruzaína. 5. Planejamento de
fármacos. I. Andricopulo, Adriano Defini, orient. II.
Título.

À minha mãe, pelo sinônimo de esforço e dedicação. Amor eterno!

AGRADECIMENTOS

Primeiramente, agradeço a Deus por tudo e por ter me dado direção e sabedoria para enfrentar os grandes desafios e vencê-los. Agradeço aos meus familiares, Marilda de Fátima da Silva (mãe), Natan Silva de Souza (irmão), Fernando Luis da Silva (irmão), Renan Silva Bittencourt (irmão), pela força e por toda ajuda e apoio. Agradeço à Rafaella Radmylan Silva do Amaral (noiva), por estar comigo o tempo todo e ter me ajudado não somente no doutorado e também em outras áreas da minha vida, e sempre ter me motivado a continuar firme nos objetivos.

Agradeço a Valdenice, Bruna, Ewerton, Rita, Janaína por sempre apoiar nos estudos.

Agradeço ao prof. Dr. Adriano Defini Andricopulo, pela oportunidade de ter trabalhado no seu laboratório (Laboratório de Química Medicinal e Computacional), pela paciência, pela ajuda e por orientar-me nesse trabalho de doutorado.

Agradeço ao Dr. Leonardo Luiz Gomes Ferreira, por me ajudar com sugestões, pelo apoio no doutorado e pela ajuda na correção dessa tese.

Agradeço ao grupo LQMC por me receber, pelo apoio, amizade e carinho, prof. Adriano, Alex, Leonardo, Rafael, Renata, Simone, Luma, Maria, Mariana, Carol e Aldo.

Agradeço ao grupo NACA, em especial aos meus amigos Anderson, Everlin, Bruno, Naiara, Weverton, Alessandro, Leonardo, Hedieh, Richard, Thales, Higor, Washington, Juliana, Érika, Cássio, Luiz, Inês e prof. Valmor pela amizade, incentivo e motivação.

Agradeço à Universidade de São Paulo (USP), ao Instituto de Física de São Carlos (IFSC) e ao Laboratório de Química Medicinal e Computacional (LQMC).

Agradeço a CAPES pela bolsa de doutorado, e às agências de fomento CNPq e FAPESP (CEPID, CIBFar).

“O sucesso é a soma de pequenos esforços repetidos dia após dia.”

Robert Collier

RESUMO

SOUZA, A.S. **Estudos em quimiinformática baseados na estrutura do ligante e do receptor de candidatos a agentes antichagásicos.** 2019. 196 p. Tese (Doutorado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

A doença de Chagas é uma doença tropical negligenciada e um grande problema de saúde mundial, principalmente na América Latina. O agente etiológico da doença é o protozoário *Trypanosoma cruzi*. Segundo a Organização Mundial da Saúde (OMS), cerca de 8 milhões de pessoas estão infectadas com o parasita e 25 milhões vivem áreas de risco de contaminação. A farmacoterapia consiste nos fármacos nifurtimox e benzonidazol. Ambos apresentam sérios efeitos adversos como dermatite, neurotoxicidade e hepatite, e baixa eficácia na fase crônica da doença. Nesse contexto, há urgência no desenvolvimento de novos fármacos seguros e eficazes para o tratamento da doença. Dentre os alvos moleculares validados para o desenvolvimento de novos fármacos para a doença de Chagas, destaca-se a cruzaina, principal cisteíno-protease do parasita. A cruzaina é expressa em todo o ciclo de vida do parasita e está envolvida em diversos processos biológicos do protozoário, como nutrição, reprodução, invasão, reconhecimento molecular e evasão. A alta atividade que muitos inibidores apresentam contra a enzima, não se traduz, para grande parte destes inibidores, em alta atividade e seletividade em ensaios fenotípicos. Isso ocorre devido a fatores como propriedades farmacocinéticas inadequadas. Nesse contexto, essa tese de doutorado aborda dois temas: o primeiro consiste na construção de modelos de redes neurais artificiais ($r^2 = 0,81$ e $r^2_{\text{pred}} = 0,81$) e KPLS, na sigla em inglês para *Kernel-Based Partial Least Squares* ($N = 3$, $r^2 = 0,81$ e $r^2_{\text{pred}} = 0,85$). Esses modelos foram utilizados para investigar propriedades estruturais, fragmentos privilegiados e propriedades físico-químicas associadas à atividade em ensaios fenotípicos *in vitro*. Em seguida, foram desenvolvidos modelos de floresta randômica (número de árvores = 100, $r^2 = 0,97$ e $r^2_{\text{pred}} = 0,77$) e KPLS radial ($N = 4$, $r^2 = 0,88$ e $r^2_{\text{pred}} = 0,88$) para avaliar as propriedades moleculares e estruturais relacionados à seletividade. Posteriormente, essas informações foram combinadas para a realização de triagens virtuais em larga escala. Por fim, um conjunto de estruturas selecionadas foram docadas no sítio ativo cruzaina e submetidas à predição de permeabilidade através de membrana e absorção, distribuição, metabolismo, excreção e toxicidade (ADME-Tox). Esses estudos conduziram à descoberta de 260 novos compostos, os quais foram avaliados *in silico* quanto a atividade biológica, seletividade, e quanto aos pontos de clivagem pela enzima CYP3A4. O segundo tema consiste em simulações de dinâmica molecular do complexo K777-cruzaina em pH 5,5 e 8,0 combinadas a modelagem por redes neurais artificiais (RNAs). De acordo com as RNAs, a mudança nos estados de protonação de duas histidinas, causada por mudanças no pH pode deslocar alças, hélices- α e folhas- β . Estes deslocamentos causam uma alteração do ambiente químico no sítio de interação, alterando a afinidade do inibidor

pela enzima (energia livre de ligação). Em pH 5,5, a HIS162 interage principalmente com os resíduos ASP161, GLY163 e ASN182, e a CYS25 permanece disponível para realizar um ataque nucleofílico ao inibidor. Em pH 8,0, no entanto, a HIS162 perde as interações com o ASP161 e a GLY163, e interage com a cadeia lateral da CYS25 e com a cadeia principal da SER183. Estas alterações podem bloquear o ataque nucleofílico da CYS25 ao inibidor K777. Este estudo contribuiu para o avaliar a influência do pH na alteração dos elementos de estrutura secundária da enzima e sua relação com a afinidade dos inibidores pela cruzaina. Por fim, foram desenvolvidos modelos de QSAR 2D e 3D a partir de uma série de oxadiazóis ($r^2_{\text{pred KPLS}}$ e $r^2_{\text{pred CoMFA}}$ de 0,81 e 0,82, respectivamente). A integração de métodos de QSAR 2D e 3D e as estratégias de docagem molecular possibilitaram mapear os subsítios do sítio ativo da cruzaina quanto a ligações de hidrogênio, ligações de halogênio e interações do tipo π - π e CH- π . Estes estudos fornecem informações quanto ao espaço químico a ser explorado e alterações estruturais relevantes para o planejamento e avaliação de novos agentes antichagásicos e inibidores da cruzaina.

Palavras-chave: Inteligência artificial. *Trypanosoma cruzi*. Doença de Chagas. Cruzaina. Planejamento de fármacos.

ABSTRACT

SOUZA, A.S. **Ligand- and structure-based chemoinformatics studies of candidates to antichagasic agents.** 2019. 196 p. Tese (Doctor in Science) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2019.

Chagas disease is a neglected tropical disease and a major global health problem, especially in Latin America. The etiological agent of the disease is the protozoan *Trypanosoma cruzi*. According to the World Health Organization (WHO), approximately 7 million people are infected with the parasite and 25 million live at risk of infection. Pharmacotherapy consists of the drugs nifurtimox and benznidazole. Both have serious adverse effects, such as dermatitis, neurotoxicity and hepatites, and low efficacy in the chronic phase of the disease. In this context, there is an urgent need for the development of novel, safe and effective drugs for the treatment of the disease. Among the molecular targets validated for the development of new drugs for Chagas disease, cruzain, the major cysteine protease of the parasite can be highlighted. Cruzain is expressed in the whole lifecycle of the parasite and is involved in several protozoan biological processes such as nutrition, reproduction, invasion, molecular recognition and evasion. The high activity that many inhibitors present against the enzyme does not translate, for most of these inhibitors, into high activity and selectivity in phenotypic assays. This is due to factors such as unfavorable pharmacokinetic properties. In this context, this doctoral thesis deals with two themes: the first is the construction of artificial neural network models ($r^2 = 0.81$ and $r^2_{\text{pred}} = 0.81$) and Kernel-Based Partial Least Squares ($N = 3$, $r^2 = 0.81$ and $r^2_{\text{pred}} = 0.85$). These models were used to investigate structural properties, fragments and physicochemical properties associated the *in vitro* activity in phenotypic assays. Next, random forest models were developed (number of trees = 100, $r^2 = 0.97$ and $r^2_{\text{pred}} = 0.77$) and radial KPLS ($N = 4$, $r^2 = 0.88$ and $r^2_{\text{pred}} = 0.88$) to evaluate the molecular and structural properties related to selectivity. Subsequently, this information was combined for the development of a large-scale virtual screening. Finally, a set of selected structures were docked in the active site of cruzain and subjected to prediction of membrane permeability, and absorption, distribution, metabolism, excretion and toxicity (ADME-Tox). These studies led to the discovery of 260 novel compounds, which were evaluated *in silico* for biological activity, selectivity and cleavage by the CYP3A4 enzyme. The second theme consists of molecular dynamics simulations of the K777-cruzain complex at pH 5.5. and 8.0 combined with artificial neural network modeling (ANN). According to the ANNs, the change in the protonation states of histidine residues caused by changes in pH can displace loops, α -helices and β -sheets. These shifts cause changes in the chemical environment of the interaction site, altering the affinity of the inhibitor by the enzyme (free binding energy). At pH 5.5, HIS162 interacts primarily with residues ASP161, GLY163 and ASN182, and CYS25 remains available to make the nucleophilic attack to the inhibitor. At pH 8.0, however, HIS162 loses interactions with ASP161 and GLY163, and interacts with the CYS25 side chain and the SER183 main chain. These

changes can block the nucleophilic attack by CYS25 to the K777 inhibitor. This study contributed to evaluate the role played by the pH on the alteration of secondary structure elements of the enzyme and its relationship with the affinity of the inhibitors by cruzain. Finally, 2D and 3D QSAR models were developed for oxadiazole derivatives (r^2_{pred} KPLS and r^2_{pred} CoMFA of 0.81 and 0.82, respectively). The integration of 2D and 3D QSAR methods and molecular docking strategies enabled the mapping of the the subsites of the cruzain active site regarding hydrogen bonding, halogen bonding and π - π and CH- π interactions. These studies provide information on the chemical space to be explored and structural changes relevant for the design and evaluation of new antichagasic agents and cruzain inhibitors.

Keywords: Artificial intelligence. *Trypanosoma cruzi*. Chagas disease. Cruzain. Drug design.

LISTA DE ABREVIATURAS

WHO – *World Health Organization*

OMS – Organização Mundial da Saúde

ADME-Tox – *absorption, distribution, metabolism, excretion and toxicity*

PDB – *Protein Data Bank*

IC₅₀ – Concentração de composto para inibir o crescimento celular em 50%

SI – Índice de seletividade

CYP450 – Citocromo P450

TRP – Triptofano

PRO – Prolina

ASN - Asparagina

CYS – Cisteína

HIS – Histidina

GLY – Glicina

LYS – Lisina

GLU – Glutamato

MET – Metionina

RP – *Ensemble recursive partitioning*

KPLS – *Kernel-based partial least squares regression*

PLS – *Partial least squares regression*

LR – *linear regression*

PCA – *Principal component analysis*

pIC₅₀ – Logaritmo de 1/IC₅₀

pSI – Logaritmo de SI

IgG – Imunoglobulina

HIV – Human Immunodeficiency Virus

AIDS – Acquired Immunodeficiency Syndrome

TcNTR – *T. cruzi* nitrorredutase

FDA – *Food and Drug Administration*

DNDi – *Drugs for Neglected Disease initiative*

CoMFA – *Comparative Molecular Field Analysis*

CoMSIA – *Comparative Molecular Similarity Index Analysis*

CHARMM – *Chemistry at HARvard Macromolecular Mechanics*

AMBER – *Assisted Model Building with Energy Refinement*

OPLS – *Optimized Potential for Liquid Simulations*

GROMOS – *GRoningen Molecular Simulation*

LBVS – *ligand-based virtual screening*

NCEs – *new chemical entities*

LBDD – *ligand-based drug design*

SBDD – *structure-based drug design*

VS – *virtual screening*

HTS – *High-Throughput Screening*

QSAR – *quantitative structure-activity relationships*

SAR – *structure-activity relationships*

MW – peso molecular

$a\text{Log}P$ – lipofilicidade

HBD – número de doadores de ligação de hidrogênio

HBA – número de aceptores de ligação de hidrogênio

RB – número de ligações rotacionáveis

PSA – área de superfície polar

Estate – estados eletro-topológicos

MR – refratividade molar

Polar – polarização

HAC – número de átomos pesados

RC – número de anéis

RNAs – redes neurais artificiais

ANNs – *artificial neural networks*

GLIDE – *Grid-based Ligand Docking with Energetics*

HTVS – *high-throughput virtual screening*

SP – *standard precision*

XP – *extra precision*

vdW – termo de energia de Van der Waals

Coul – termo de energia de Coulomb

GOLD – *Genetic Optimization for Ligand Docking*

ΔG – energia livre de Gibbs

Caco-2 – permeabilidade em células de Caco-2

$\log P_{\text{bb}}$ – coeficiente de partição hematocefálica

AH – ligação em albumina humana

PAO – porcentagem de absorção oral

hERG – *Human ether-a-go-go-related gene*

pH – potencial hidrogeniônico

VMD – *visual molecular dynamics*

MM-PBSA – *molecular mechanics Poisson-Boltzmann surface area*

MR – taxa de momento

$q^2 - r^2$ para o conjunto teste r^2_{pred}

r^2 – coeficiente de correlação da série treinamento

RMSE – erro quadrático médio de predição

SD – desvio padrão

AME – erro médio absoluto

ERA – erro absoluto relativo

RQER – erro relativo quadrático médio de predição

NN – número de neurônios na camada escondida

RF – *random forest*

SUMÁRIO

1	INTRODUÇÃO	21
1.1	A DOENÇA DE CHAGAS	21
1.2	CRUZAÍNA	28
1.3	INIBIDORES DA CRUZAÍNA.....	30
1.4	RELAÇÃO DO PH NO DESENVOLVIMENTO DE NOVOS INIBIDORES DA CRUZAÍNA	32
1.5	PLANEJAMENTO DE NOVOS FÁRMACOS.....	35
1.6	PROPRIEDADES FARMACOCINÉTICAS.....	39
1.7	DINÂMICA MOLECULAR E A CRUZAÍNA	40
2	OBJETIVOS	43
3	MATERIAIS E MÉTODOS	45
3.1	TRIAGEM VIRTUAL EM LARGA ESCALA.....	45
3.1.1	<i>Busca no banco de dados .....</i>	<i>46</i>
3.1.2	<i>Análise por componentes principais.....</i>	<i>47</i>
3.1.3	<i>Descritores físico-químicos .....</i>	<i>47</i>
3.1.4	<i>Redes neurais artificiais de retropropagação.....</i>	<i>48</i>
3.1.5	<i>Descritores moleculares de impressão digital e KPLS .....</i>	<i>50</i>
3.1.6	<i>Floresta randômica.....</i>	<i>50</i>
3.1.7	<i>Regressão linear baseada em dados experimentais .....</i>	<i>51</i>
3.1.8	<i>Triagem virtual em larga escala baseada no ligante.....</i>	<i>52</i>
3.1.9	<i>Triagem virtual baseada na estrutura do alvo molecular .....</i>	<i>54</i>
3.1.9.1	<i>Preparação dos ligantes.....</i>	<i>54</i>
3.1.9.2	<i>Preparação da cruzaina .....</i>	<i>54</i>
3.1.9.3	<i>Docagem molecular.....</i>	<i>55</i>
3.1.10	<i>Filtragem baseada na predição da seletividade.....</i>	<i>58</i>
3.1.11	<i>Teste físico de permeabilidade de membrana .....</i>	<i>59</i>
3.1.12	<i>Filtragem baseada nas características ADME-Tox.....</i>	<i>60</i>
3.1.13	<i>Filtragem baseada na distância da reta ideal.....</i>	<i>61</i>
3.1.14	<i>Predição dos pontos de clivagem no enzima CYP3A4.....</i>	<i>61</i>
3.2	DINÂMICA MOLECULAR E REDES NEURAIS ARTIFICIAIS.....	62
3.2.1	<i>Predição dos estados de protonação do ligante .....</i>	<i>63</i>
3.2.2	<i>Predição dos estados de protonação dos aminoácidos.....</i>	<i>63</i>
3.2.3	<i>Construção dos sistemas apo e holo em diferentes pH.....</i>	<i>64</i>
3.2.4	<i>Cálculo da variação da energia livre de Gibbs.....</i>	<i>65</i>
3.2.5	<i>Modelos de redes neurais baseados em dinâmica molecular .....</i>	<i>65</i>
3.3	MAPEAMENTO DO SÍTIO ATIVO DA CRUZAÍNA.....	67
3.3.1	<i>Conjunto de dados .....</i>	<i>67</i>

3.3.2	<i>Construção dos modelos de QSAR 2D</i>	68
3.3.2.1	QSAR 2D	68
3.3.2.2	QSAR 3D	69
3.3.3	<i>Docagem molecular</i>	72
4	RESULTADOS E DISCUSSÃO	73
4.1	CONJUNTO 1 – ESTUDOS DE QSAR	73
4.1.1	<i>Análise por componentes principais – conjuntos treinamento e teste</i>	73
4.1.2	<i>Análise estatística</i>	74
4.1.2.1	Descritores físico-químicos e redes neurais artificiais.....	74
4.1.2.2	Descritores de impressão digital molecular e KPLS.....	79
4.1.3	<i>Relações das propriedades moleculares/estruturais e a atividade biológica</i>	80
4.1.3.1	QSAR baseado em redes neurais artificiais.....	80
4.1.3.2	QSAR baseado em KPLS molprint2D	82
4.1.3.3	Estudos de bioisosterismo baseado em fragmentos moleculares	83
4.2	CONJUNTO 2 – ESTUDOS DE QSAR	89
4.2.1	<i>Análise por componentes principais – conjuntos treinamento e teste</i>	89
4.2.2	<i>Análise estatística</i>	89
4.2.2.1	Descritores físico-químicos e floresta randômica.....	89
4.2.2.2	Regressão por mínimos quadrados de pSI versus pIC ₅₀	93
4.2.2.3	Descritores de impressão digital molecular e KPLS.....	94
4.2.3	<i>Relações das propriedades moleculares/estruturais e a atividade biológica</i>	96
4.2.3.1	QSAR baseado em floresta randômica	96
4.2.3.2	QSAR baseado em KPLS radial.....	97
4.3	TRIAGEM VIRTUAL EM LARGA ESCALA	98
4.3.1	<i>Espaço químico e biológico – Energia livre de Gibbs</i>	98
4.3.2	<i>Espaço químico e biológico – pIC₅₀ e pSI estimados</i>	100
4.3.3	<i>Predição das propriedades ADME-Tox</i>	103
4.4	DINÂMICA MOLECULAR E REDES NEURAS ARTIFICIAIS	108
4.4.1	<i>Equilíbrios de fase NVT, NPT, R_g e RMSD</i>	108
4.4.2	<i>Estudos estruturais e termodinâmicos</i>	110
4.4.3	<i>Estudos de redes neurais artificiais baseados em dinâmica molecular</i>	114
4.5	MAPEAMENTO DO SÍTIO ATIVO DA CRUZAÍNA	122
4.5.1	<i>Conjunto de dados</i>	122
4.5.2	QSAR 2D	123
4.5.2.1	Estudos de QSAR 2D	123
4.5.2.2	Ligação de halogênio e metila	125
4.5.3	<i>Hipótese farmacofórica baseada na estrutura do ligante</i>	126
4.5.4	QSAR 3D baseado no GALAHAD	127
4.5.5	<i>Capacidade preditiva dos modelos 2D e 3D</i>	129

5	CONCLUSÃO.....	131
	REFERÊNCIAS.....	132
	ANEXOS	149
	ANEXO A – Representação do conjunto 1 (pIC_{50})	149
	ANEXO B – Representação do conjunto 2 (pIC_{50} e pSI).....	157
	ANEXO C – Representação na forma SMILES dos 260 compostos obtidos através da triagem virtual em larga escala.....	163
	ANEXO D – Conjunto de dados de inibidores da cruzafina utilizado na construção dos modelos de QSAR 2D e 3D.....	195

1 INTRODUÇÃO

1.1 A doença de Chagas

A doença de Chagas é uma doença tropical negligenciada causada pelo protozoário *Trypanosoma cruzi*, descrita em 1909 pelo cientista brasileiro Carlos Justiniano Ribeiro Chagas.¹ Ele foi o único cientista na história a descrever completamente uma doença infecciosa, incluindo sua anatomia patológica, transmissão, agente etiológico, formas clínicas e estudos epidemiológicos.¹ Segundo a Organização Mundial da Saúde (OMS), cerca de 25 milhões pessoas vivem em áreas de risco de contaminação, e por volta de 8 milhões estão infectadas com o protozoário.² Além disso, é estimado que a doença cause por volta de 10 mil mortes por ano.² Na América Latina, essa doença é endêmica em 21 países, principalmente Brasil, Argentina e México.³ O fluxo migratório de latino-americanos infectados pelo parasita tem auxiliado na disseminação da doença nos Estados Unidos, Europa, Ásia e Oceania, fenômeno definido como globalização da doença de Chagas no mundo.⁴ Os Estados Unidos, diferentemente dos países da Europa e da Ásia, possui várias espécies de triatomíneos atuando como vetores de propagação do *T. cruzi*.⁵ No entanto, o número de infectados por transmissão vetorial é inferior ao número de imigrantes.⁵ Estima-se que o número de pessoas infectadas pelo *T. cruzi* nos Estados Unidos seja maior do que 300 mil;⁶ na Europa, o número de casos está entre 68 e 120 mil, dos quais 30% podem desenvolver os sintomas característicos da doença na sua forma determinada, e 90% possuam a forma indeterminada da doença de Chagas;⁷ No Japão, é estimado que o número de casos seja superior a 4 mil;⁸ Na Oceania, estima-se que mais de 1900 pessoas sejam portadoras do parasita.⁹ O mapa a seguir, descreve a epidemiologia da doença de Chagas no mundo (fig. 1).

Figura 1 - Mapa epidemiológico da doença de Chagas no mundo entre 2006-2009.

Fonte: Adaptada de WORLD ...²

A transmissão vetorial do parasita ocorre pelo contato com as fezes de insetos infectados com o parasita, sendo transmitida para humanos e mamíferos (ratos, cães e gatos) por mais de 130 espécies de insetos da subfamília Triatominae.¹⁰ Entretanto, as espécies *Triatoma infestans*, *Rhodnius prolixus* e *Triatoma dimidiata* são os principais vetores de transmissão do parasita.¹⁰ Esses três vetores são popularmente conhecidos como “barbeiro” e são amplamente distribuídos por florestas e áreas secas, as quais se estendem do sul dos Estados Unidos da América¹¹ até regiões da Argentina e Chile.³ A transmissão não vetorial ocorre via transfusão sanguínea, transplantes de órgãos ou consumo de sucos infectados pelo parasita, sendo esse tipo de transmissão responsável pela propagação do parasita em países não endêmicos e endêmicos.¹² Além disso, é estimado que a taxa de transmissão do protozoário por mulheres grávidas infectadas para a criança esteja entre 4 e 7%, e a taxa de transmissão por transplante de coração, fígado e rim por doadores infectados estão entre 75–100, 0–29, e 0–19%, respectivamente.³

O parasita *T. cruzi* possui um ciclo de vida complexo (fig. 2) que envolve o inseto triatomíneo, o parasita e um mamífero.³ O ciclo se inicia quando o triatomíneo

se alimenta do sangue de um mamífero e libera fezes contendo as formas tripomastigotas do *T. cruzi*.³ Posteriormente, na corrente sanguínea os tripomastigotas infectam as células do mamífero e se diferenciam em formas amastigotas.³ No interior da célula, os amastigotas se dividem por fissão binária e, após a replicação, se transformam em tripomastigotas metacíclicos.³ A partir daí, os parasitas rompem as células hospedeiras, entram na corrente sanguínea, e invadem diferentes células e tecidos do hospedeiro.³ A seguir, o inseto se alimenta do sangue de um mamífero contendo tripomastigotas metacíclicos.³ No intestino médio do inseto, o parasita se diferencia na forma epimastigota, e no intestino posterior, se converte em tripomastigota metacíclico.³ O ciclo de transmissão então é fechado quando o inseto se alimenta do sangue de um mamífero, defeca tripomastigotas metacíclicos e esses entram na circulação sanguínea do hospedeiro.³

Figura 2 - Ciclo de vida do *Trypanosoma cruzi*.

Fonte: Elaborada pelo autor.

A forma amastigota do parasita pode ser encontrada particularmente em tecidos, músculos cardíacos, e células do sistema digestivo e células fagocíticas.¹³ Durante o ciclo de vida do parasita, após a invasão da célula hospedeira, ocorre a conversão da forma tripomastigota para a forma amastigota.¹⁴ As alterações morfológicas mais importantes nessa etapa são o desaparecimento do flagelo livre,

a formação de uma membrana ondulada e a redução do tamanho do diâmetro da célula do parasita de 17 para 4 μm .¹⁵ Estudos mostraram que as amastigotas intracelulares ficam localizadas próximas à mitocôndria.¹⁶ Essa proximidade faz com que o parasita se aloje em pontos estratégicos para sua nutrição e replicação.¹³ Os mecanismos moleculares que envolvem a relação entre o parasita e as células a serem invadidas e/ou a residência parasitária no citosol ainda não foram completamente esclarecidos.¹³ No entanto, o metabolismo de ácidos graxos está entre as principais rotas celulares de mamíferos que pode estar relacionada à propagação da forma amastigota de *T. cruzi* no meio intracelular.¹⁷ Além disso, a transformação da forma tripomastigota para amastigota envolveria o aumento da ação enzimática de proteases, as quais estão associadas ao processo de remodelamento e biotransformação da membrana plasmática.^{16, 17} Isso seria explicado pelo aumento expressivo da síntese lipídica, a qual estaria envolvida na melhor adaptação do parasita ao interior da célula.^{16, 17} Assim, o metabolismo de ácidos graxos na célula hospedeira parece ser fundamental para o *T. cruzi*.¹⁶ Além disso, estudos mostraram que há distribuições lipídicas diferentes na membrana celular do hospedeiro e do parasita.^{16, 17} Por outro lado, ocorre um aumento expressivo de glicose nas células infectadas com o *T. cruzi*, e um aumento no catabolismo e anabolismo nas mitocôndrias.¹⁷ Estudos mostraram que a diminuição de concentração de glicose no meio extracelular estaria associada com a inibição do crescimento parasitário e sugerem que o parasita acesse as reservas de glicose do hospedeiro para alimentar seu próprio metabolismo e se reproduzir no citoplasma.¹⁵ Portanto, a relação de diferenciação da forma tripomastigota para amastigota, metabolismo e biogênese de lipídios e carboidratos podem estar associados à mitocôndria da célula hospedeira e são fatores que aparentemente são fundamentais para propagação do parasita.

Os métodos clínicos e laboratoriais permitem classificar a doença de Chagas em duas fases: a fase aguda e a fase crônica¹⁸ (fig. 3). A fase aguda é caracterizada pela alta taxa de carga parasitária na corrente sanguínea do hospedeiro.¹⁸ Essa fase é caracterizada por sintomas inespecíficos (febre, diarreia, náuseas) e pode durar de 2 a 4 meses em média, dependendo da virulência do parasita, de fatores genéticos e da idade do indivíduo.¹⁸ A evolução natural para a fase crônica pode não ocorrer, mesmo quando não tratada, sendo marcada pelo desaparecimento espontâneo dos sintomas.¹⁸ A fase crônica é caracterizada pela baixa taxa de parasitemia e altas

taxas de anticorpo IgG anti-*T. cruzi* e pode ser classificada em assintomática (indeterminada) ou sintomática (determinada).¹⁹ Na fase indeterminada, o indivíduo não manifesta mudanças radiológicas no sistema digestivo e no eletrocardiograma.²⁰ No entanto, o portador da doença de Chagas crônica determinada pode apresentar quadros clínicos severos após 10 a 30 anos depois da infecção, dentre eles, o aumento de tamanho do esôfago (megaesôfago, em 25% dos casos), cólon (megacólon, 30% dos pacientes) ou do coração (cardiomegalia, 80% dos infectados).^{18, 20} Os sintomas graves da cardiomegalia chagásica crônica são a cardiomiopatia e arritmias ventriculares que causam a morte do indivíduo.^{18, 21}

Figura 3 - Diagrama esquemático da evolução da doença de Chagas.
Fonte: Adaptada de CHATELAIN.¹⁸

Em relação à patogenicidade da doença de Chagas na sua forma determinada, ainda há questões em aberto.¹⁸ No entanto, a relação parasita e resposta imune do hospedeiro explicariam as diferentes manifestações clínicas para os portadores da doença no estágio crônico.²² A resposta imunológica eficaz levaria a redução significativa dos parasitas na corrente sanguínea e nos tecidos, e consequente ausência de sintomas em pacientes com a doença de Chagas crônica e indeterminada.²² Por sua vez, uma resposta imunológica deficiente pode levar à permanência do *T. cruzi* nos tecidos.²² O mecanismo de defesa seria o responsável pelas lesões graves e uma resposta inflamatória local, causando a longo prazo as sintomatologias específicas da doença na sua forma determinada, mesmo na ausência (ou quase ausência) do parasita nos tecidos.²² Mesmo que o sistema imune esteja relacionado com os sintomas graves da doença, estudos revelaram

que deficiências imunológicas poderiam agravar o quadro clínico, como nos casos de pacientes imunodeprimidos infectados pelo HIV, o vírus causador da AIDS.²³

O tratamento para a doença de Chagas se limita a dois nitrocompostos: o derivado de nitrofurano conhecido como nifurtimox, e o derivado nitroimidazólico chamado de benzonidazol (fig. 4), comercializados respectivamente, como Lampit® e Rochagan®.²⁴ Ambos apresentam melhores resultados em *T. cruzi* no meio extracelular do que no meio intracelular,²⁵ atuando pela formação de metabólitos nucleofílicos e/ou radicais livres gerados pela ação de nitroredutases do tipo I e/ou II.²⁶ No mecanismo I (descrição clássica) as nitroredutases do tipo II (TcNTR II) atuam sobre o grupo nitro produzindo um radical intermediário carregado negativamente.²⁶ Esse intermediário possui elevado potencial redox e sofre uma reação com a molécula de oxigênio, regenerando parcialmente esse composto. Assim, ocorre a transferência de um elétron para o oxigênio molecular formando um superóxido (O_2^-), e esse pode causar um estresse oxidativo no sistema celular do parasita.²⁶ No mecanismo II, as nitroredutases do tipo I reduzem o grupo nitro para o intermediário hidroxiamina, a qual liga-se a lipídios, proteínas e DNA, levando a um efeito tripanocida no meio intra e/ou extracelular.²⁶ A ação do nifurtimox pode ocorrer pelos dois mecanismos e ação do benzonidazol pelo mecanismo II (fig. 4). O mecanismo de ação do benzonidazol pode ocorrer pela modificação covalente das macromoléculas por intermediários nitroreduzidos ou por outras interações de nitroredução com os componentes do *T. cruzi* (fig. 4).²⁷ Inicialmente, a nitroredutase do tipo I reduz o benzonidazol a um intermediário nitroso seguido da redução para hidroxiamina.²⁷ Posteriormente, ocorrem consecutivas reações não enzimáticas na formação dos intermediários nitrênio, hidroxí, dihidro-diidroxi-imidazol.²⁷ Por fim, ocorre uma conversão química, formando o derivado guanidínico e glioxial dialdeído. Esse último é altamente reativo, e pode se ligar a ácidos nucleicos (DNA ou RNA), proteínas e pequenas moléculas do parasita, causando sua morte.²⁷

Figura 4 - Farmacoterapia atual para a doença de Chagas e ação dos fármacos A) Nifurtimox e B) Benznidazol.

Fonte: Adaptada de CERECETTO; GONZÁLEZ.²⁶

Em 2003, a Roche doou a patente do Rochagan® e sua tecnologia de produção para o Brasil, para o Laboratório Farmacêutico do Estado de Pernambuco (LaFEP).²⁸ Entretanto, devido ao aumento da demanda, ocorreu escassez desse medicamento, e o LaFEP passou a responsabilidade da produção do princípio ativo para a Nortec Química.²⁸ Recentemente, o benznidazol foi aprovado pela agência reguladora de alimentos e fármacos dos Estados Unidos (U.S. FDA, do inglês *United States Food & Drug Administration*), via aprovação acelerada, para o tratamento de crianças de 2 a 12 anos de idade portadores da doença de Chagas.²⁹ No entanto, ambos os fármacos não são indicados para mulheres em fase de gestação e pacientes com disfunção renal e/ou hepática. Além disso, devido ao seu mecanismo de ação, esses fármacos podem causar diversos efeitos adversos graves, tais como perda de peso, alterações neurológicas, anorexia, hipersensibilidade, dermatite, depressão da medula óssea, náuseas, vômitos, cólicas intestinais e diarreia.²⁴ A toxicidade do nifurtimox inclui dano testicular e neurotoxicidade.²⁴ Devido a este perfil de toxidez e por apresentar relatos de resistência a cepas de *T. cruzi*, o nifurtimox teve seu uso descontinuado no Brasil nos anos 80 e, posteriormente, em outros países latino-americanos, tais como Argentina, Chile e Uruguai.³⁰ Assim, o benznidazol é a primeira linha de tratamento para a doença de Chagas e única alternativa terapêutica em alguns países.^{24,27} O

nifurtimox não é um fármaco aprovado pelo FDA devido a sua alta toxicidade, e não está disponível no mercado farmacêutico, sendo obtido somente via OPAS (Organização Pan-americana da Saúde) e OMS.^{24,28} Portanto, segundo a iniciativa de desenvolvimento de fármacos para doenças negligenciadas (DNDi, do inglês *Drugs for Neglected Disease initiative*) é urgente o desenvolvimento de novos candidatos a fármacos que sejam eficazes e seguros nas duas fases clínicas da doença (aguda e crônica) e que possam ser administrados por via oral.³¹

1.2 Cruzaína

As proteases são classificadas de acordo com o seu mecanismo de catálise e especificidade pelo substrato.³² Os aminoácidos envolvidos no mecanismo de hidrólise de ligações peptídicas as classificam em cinco categorias principais: serina, cisteína, treonina e metalo-proteases, e proteases aspárticas.³² As proteases são abundantes no *T. cruzi* e são essenciais para vários processos biológicos do parasita.³³ Em particular, a cruzipaína (EC 3.4.22.51) é uma glicoproteína. Ela é a principal cisteíno protease e pertence à família da papaína.³⁴ No entanto, a proteína recombinante sem o domínio C-terminal é nomeada cruzaína.³⁵ No *T. cruzi*, a cruzipaína possui polimorfismo genético, possuindo apenas cinco isoformas expressas nos diferentes estágios morfológicos do *T. cruzi*.³⁶ Essas isoformas apresentam poucas variações nos seus resíduos de aminoácidos e cada forma morfológica do *T. cruzi* apresenta apenas uma ou duas isoformas.³⁶ Por exemplo, a cruzipaína A é expressa exclusivamente na forma epimastigota.³⁶ As cruzipaínas B e E são expressas na forma tripomastigota.³⁶ As cruzipaínas C e D são descritas somente na forma amastigota.³⁶ Estudos mostraram que essa baixa variedade genética de um determinado alvo biológico é um fator imprecindível para o planejamento de fármacos.³⁷ Dessa maneira, a baixa variedade genética da cruzaína a torna especialmente interessante para a descoberta de inibidores.

A cruzaína está envolvida nas reações de proteólise essenciais para o parasita. A principal função dessa glicoproteína no *T. cruzi* está relacionada a sua nutrição. Para diferentes pH (2,75 e 9,15), os valores de k_{cat}/k_m da proteína endógena comparados aos valores da proteína recombinante (ou seja, o domínio catalítico sem o C-terminal) não são afetados.³⁸ A protease pode digerir a

imunoglobulina G humana anti-*T. cruzi* (IgG anti-*T.cruzi*).³⁹ Esses resultados sugerem que a enzima pode estar envolvida na proteólise intracelular de IgG endocitada, protegendo o protozoário contra a resposta imune do hospedeiro.³⁹ A cruzaina é expressa nas três formas morfológicas do ciclo de vida do *T. cruzi*. No entanto, a catálise enzimática é 10 vezes maior na forma epimastigota.⁴⁰ Em epimastigotas, ela está localizada em um compartimento membranoso pré-lisossomal, o reservossomo.⁴¹⁻⁴³ Essa organela possui pH de 5,5 e contém enzimas digestivas que atuam em outras proteínas durante o processo de alteração morfológica da forma tripomastigota do *T. cruzi*.⁴⁴ Além disso, estudos de localização celular por microscopia eletrônica revelaram que outras isoformas da enzima estão localizadas na superfície celular e no bolsão flagelar nas formas amastigotas e tripomastigotas do *T. cruzi*, respectivamente.^{40,43} Nas formas amastigotas, a cruzaina estaria ligada à âncora de fosfatidil-inositol (GPI) na membrana celular do parasita.⁴⁵ Outros estudos mostraram que a enzima pode ser secretada para o ambiente extracelular por tripomastigotas durante a invasão de células hospedeiras.⁴⁶

O domínio catalítico das cisteína proteases contém a tríade de resíduos ASN182 (ou ASN175), CYS25 e HIS162 (ou HIS159), que são essenciais para a catálise enzimática (fig. 5). O modelo de mecanismo de catálise enzimática para a cruzaina foi proposto por Zhai⁴⁷ e complementado pelos estudos de Arafet e colaboradores.⁴⁸ Inicialmente, na enzima livre os resíduos CYS25 e HIS162 são neutros em relação à carga.⁴⁷ O processo de formação do par iônico tiolato-imidazólico pode ser ativado pela mudança conformacional induzida pelo substrato. Essa ativação seria provocada pela transferência de prótons da CYS25 para HIS162.⁴⁷ Depois da ativação do par iônico, o processo de catálise enzimática é dividido em duas etapas consecutivas: acilação e desacilação.⁴⁸ Na acilação, o resíduo HIS162 é responsável pela transferência de prótons para o átomo N1 da cadeia peptídica. Subsequentemente, o carbono carbonílico do peptídeo sofre um ataque nucleofílico pela CYS25, a qual é carregada negativamente.⁴⁸ No processo de desacilação, também conhecido como reação catalisada pela base, a água é ativada pelo resíduo HIS162.⁴⁸ A água ataca o peptídeo e o produto da reação é gerado.⁴⁸ Finalmente, os autores mostraram que os resíduos GLN19, ASN182 e TRP184 são relevantes para modular o valor de pK_a do resíduo titulável.⁴⁸

Figura 5 - Modelo de mecanismo de catálise proposto por Zhai e Arafet e colaboradores.

Fonte: Elaborada pelo autor.

1.3 Inibidores da cruzaina

Atualmente, os esforços para desenvolver novos inibidores da cruzaina são baseados em diversas abordagens. Podemos destacar a docagem molecular, as triagens virtuais baseadas estrutura do alvo molecular e do ligante, e os modelos de QSAR.⁴⁹ Dentre muitos compostos bioativos, o inibidor irreversível K11777 (fig. 6) da cruzaina (ou K777) pertence à classe das vinilsulfonas e apresenta boa atividade frente à cruzaina ($K_D = 1930$ nM).⁵⁰ A estrutura do complexo K777/cruzaina foi resolvida a 1,95 Å de resolução (PDB ID 2OZ2).⁵⁰ Nessa estrutura cristalográfica, é mostrada a ligação irreversível formada pelo ataque nucleofílico da CYS25 no C27 do grupo vinil.⁵⁰ O inibidor ocupa quatro subsítios da enzima: em S_1 , o grupo homofenilalanil fica exposto ao solvente, e o grupo sulfona recebe ligação de hidrogênio do resíduo GLN19; em S_2 , o grupo fenilalanil realiza interações hidrofóbicas com esse subsítio; em S_3 , o grupo N-metil-piperazina fica exposto ao solvente; e em S_1' , os resíduos HIS162 e TRP184 fazem interação π - π , e o aminoácido TRP184 doa ligação de hidrogênio para o grupo sulfona.⁵⁰

Figura 6 - Inibidor K777 ligado na cruzaina.
Fonte: Elaborada pelo autor.

Esse projeto foi interrompido devido a problemas de tolerabilidade observados em cães e primatas em baixas concentrações do inibidor.⁵¹ Esse análogo dipeptídico foi capaz de eliminar o parasita de células infectadas, interrompendo o ciclo de vida do parasita no nível do complexo de Golgi.⁵² Testes do K777 em microsossomos hepáticos humanos mostraram que o anel N-metil-piperazina é o principal local de metabolismo.⁵³ Testes em modelos *in vivo* demonstraram que o inibidor diminuía o dano do miocárdico causado pelo parasita.⁵⁴ Assim, esses resultados mostram claramente que a cruzaina é um alvo validado para doença de Chagas, com base em estudos de biologia estrutural e celular, bioquímicos e testes biológicos.⁵³⁻⁵⁴

Novos inibidores da cruzaina têm demonstrado serem promissores para futuros testes clínicos. Na literatura foram descritos diversos inibidores da cruzaina, sendo esses classificados em covalentes e não-covalentes. Os inibidores covalentes sofrem um ataque nucleofílico da CYS25, formando uma nova ligação química. Essa classe de inibidores é estruturalmente diversa, incluindo, entre outros arcabouços, grupos epóxidos, vinilsulfonas e as tiossemicarbazonas. Apresentam, frequentemente, valores de IC₅₀ na faixa nanomolar (fig. 7).^{49,56} Os inibidores não-covalentes, no entanto, não sofrem um ataque nucleofílico da CYS25 e se ligam de maneira reversível a enzima, por exemplo, os derivados de benzimidazóis e oxadiazóis. Os inibidores não-covalentes também possuem arcabouços diversificados e uma distribuição mais ampla de potência, com valores de IC₅₀ entre micro- e nanomolar (fig. 7).^{49,56}

Figura 7- Diferentes classes de inibidores biologicamente ativos na cruzaina.

Fonte: Adaptada de MARTINEZ-MAYORGA et al.⁴⁹

Embora haja grande diversidade estrutural dentre os inibidores da cruzaina, os quais apresentam potência na escala de nM, descobrir novos inibidores que sejam potentes em ensaios fenotípicos e apresentem propriedades farmacocinéticas adequadas, ainda é um grande desafio no planejamento de novos fármacos para a doença de Chagas.

1.4 Relação do pH no desenvolvimento de novos inibidores da cruzaina

Os estudos iniciais da cruzaina foram baseados na hidrólise da azocaseína em extratos de epimastigota de *T. cruzi*.⁵⁷ Essa enzima possui atividade 10 vezes maior em epimastigotas do que nas outras formas do *T. cruzi*.⁵⁸ Além disso, a enzima está localizada em um compartimento pré-lisossomal chamado reservosomo, o qual possui pH de 5,5 que está diretamente relacionado com a alta atividade catalítica dessa enzima.⁵⁹ Embora a cruzaina seja expressa nesse compartimento, outras isoformas da cruzaina podem ser encontradas ancoradas em glicosilfosfatidilinositol (GPI) na superfície extracelular das formas amastigotas⁶⁰ e no bolsão flagelar nas formas tripomastigotas.⁴⁰ Dessa maneira, o controle do pH se torna fundamental para o desenvolvimento de novos inibidores da cruzaina (fig. 8).

Figura 8 - Localização celular da cruzaina.

Fonte: Elaborada pelo autor.

Estudos cinéticos demonstraram que diferentes espécies de substratos não apresentaram a mesma velocidade de reação enzimática de acordo com o pH. Por exemplo, o pH ótimo de catálise enzimática para azocaseína, CVZ-Arg-OMe- β -Nam e Bz-Pro-Phe-Arg-pNa é de 5, 7 e 8,7, respectivamente.⁶¹ Os autores sugeriram que esse efeito estaria diretamente relacionado aos agentes caotrópicos, os quais afetariam a estrutura da cruzaina e não propriamente à natureza desses substratos.⁶¹

Baseado em estudos cinéticos, os inibidores K777 e WRR-483 apresentaram mudanças significativas nas constantes cinéticas de k_{cat}/K_i ($\text{s}^{-1} \cdot \text{M}^{-1}$) com o aumento do pH de 5,5 para 8.⁶²⁻⁶³ Isso mostrou uma dependência da eficiência catalítica em relação ao pH. Outros estudos demonstraram que derivados oxiguanidínicos também apresentaram constantes cinéticas de k_{cat}/K_i ($\text{s}^{-1} \cdot \text{M}^{-1}$) que variaram de acordo com o pH (fig. 9).

Figura 9 - Dependência da constante cinética de inibidores da cruzaina em meios ácidos e básicos.
 Fonte: Elaborada pelo autor.

Considerando que a cruzaina se encontra em diferentes compartimentos celulares e o funcionamento desses compartimentos depende diretamente do pH, o planejamento de novos inibidores baseado em ensaios bioquímicos em diferentes pH é fundamental para o desenvolvimento de novos fármacos para a doença de Chagas.

1.5 Planejamento de novos fármacos

As estratégias em química medicinal têm avançado de maneira significativa, contribuindo para o processo de P&D para a descoberta de candidatos a novos fármacos.⁶⁴ O objetivo é gerar novas entidades químicas (NCEs, *new chemical entities*) com mecanismos de ação determinados e propriedades farmacodinâmicas e farmacocinéticas adequadas.⁶²⁻⁶³ O sucesso de fármacos aprovados pelo FDA (*Food and Drug Administration*) ao longo dos anos está ligado a estratégias de planejamento de fármacos, incluindo os métodos de modelagem molecular.⁶⁴ Estes métodos são relevantes no estudo de propriedades farmacodinâmicas e farmacocinéticas, gerando resultados de maneira mais rápida, diminuindo a taxa de falsos positivos e demandando investimentos mais modestos em comparação a testes biológicos convencionais.⁶²⁻⁶⁴ Dentre essas estratégias, se destacam os seguintes métodos de planejamento: (i) baseados na estrutura do ligante (LBDD, *ligand-based drug design*) e (ii) baseados na estrutura do receptor (SBDD, *structure-based drug design*).

Os métodos em LBDD utilizam as informações contidas nas estruturas dos ligantes e as associa à propriedade biológica de interesse, gerando modelos de relações quantitativas entre estrutura e atividade 2D e 3D (QSAR, do inglês *Quantitative Structure-Activity Relationships*).^{33,65-66} Esses modelos são desenvolvidos a partir da aplicação de métodos estatísticos (fig. 10). Esses modelos são utilizados para direcionar a otimização de compostos líderes, gerar hipóteses farmacofóricas, prever a atividade biológica e atuar como filtros moleculares na triagem de bases de dados.

Figura 10 - Diagrama esquemático da geração de modelos de QSAR 2D e 3D.
Fonte: Elaborada pelo autor.

Os métodos em SBDD utilizam informações estruturais da macromolécula para prever o modo de ligação de ligantes no sítio de interação (fig. 11).^{64,69} A docagem molecular é utilizada para prever as conformações dos ligantes e suas afinidades com o receptor, a partir de bibliotecas virtuais de compostos.⁷⁰ Para tanto, a docagem molecular utiliza diversos algoritmos de busca conformacional. Dentre esses algoritmos, se encontram os randômicos (por exemplo, os algoritmos genéticos e de Monte Carlo), os sistemáticos (por exemplo, os incrementais) e os de simulação (dinâmica molecular e minimização de energia).⁷¹ Para avaliar as conformações preditas, são empregadas funções de pontuação. As funções de pontuação geram valores estimados de afinidade e são aplicadas para avaliar um determinado modo de ligação dos ligantes no sítio de interação.⁷¹ As funções de pontuação podem ser classificadas em empíricas, baseadas em campos de força de mecânica molecular e baseadas no conhecimento.⁷¹ Embora os métodos LBDD e SBDD apresentem diferentes abordagens, a integração dessas abordagens tem auxiliado a extração de informações fundamentais para a descoberta de novos fármacos.⁶⁸⁻⁷¹

Figura 11 - Diagrama esquemático da aplicação de métodos SBDD na obtenção de um modelo farmacofórico baseado na estrutura do receptor de estrógeno.

Fonte: Adaptada de ANDRICOPULO; SALUM; ABRAHAM.⁶⁶

O desenvolvimento de um novo fármaco tem um custo na ordem de bilhões de dólares.⁷²⁻⁷³ Entretanto, uma boa porcentagem dos custos é atribuída a moléculas que falham ao longo do processo de P&D.⁷³⁻⁷⁴ Os métodos computacionais têm contribuído para mitigar esse problema, e tem demonstrado bons resultados na busca de novas estruturas químicas em bases de dados.⁷³⁻⁷⁴ A busca por novos compostos em bases de dados virtuais é conhecida por triagem ou ensaio virtual (VS, do inglês *virtual screening*).⁷⁷ Essa estratégia tem como objetivo a seleção de compostos promissores para a realização de ensaios *in vitro*.⁷⁶⁻⁷⁷ Assim, essa abordagem é capaz de identificar novos compostos e prever propriedades biológicas relevantes. Diferentemente da triagem experimental em larga escala (HTS, do inglês *High-Throughput Screening*), os compostos são codificados virtualmente, e conseqüentemente exploram o espaço químico de uma maneira mais ampla do que um ensaio em larga escala convencional. Esses compostos são obtidos a partir de grandes bibliotecas virtuais, tais como PubChem,⁸⁰ DrugBank⁸¹ e ZINC.⁸²

O ZINC, por exemplo, é um banco de dados livre para acesso público comumente usado para realização de triagem virtual na fase de descoberta de fármacos.⁸² Esse banco de dados é de fácil uso para especialistas em diversas áreas como química, física, biologia, computação e bioquímica.⁸⁰⁻⁸² Assim, essa base de dados integra dados de propriedades biológicas, propriedades químicas,

disponibilidade comercial de compostos, disponibilidade de consulta geral e uma interface gráfica de fácil uso.⁸⁰⁻⁸² Na versão 12 do banco de dados da ZINC (ZINC 12), as estruturas podem ser filtradas pelas propriedades físico-químicas (*i.e.*, log *P*, massa molecular, número de doadores de ligação de hidrogênio, número de aceptores de ligação de hidrogênio, etc). De acordo com esses critérios, o banco de dados pode ser separado em cinco categorias: “*lead-likeness*”, “*fragment-likeness*”, “*drug-likeness*”, “*shards*” e “*all purchasable*”.⁸⁵ O subgrupo “*lead-likeness*” possui moléculas que satisfazem a regra dos 3:⁸⁶ peso molecular entre 250 e 350 Da, cLog *P* entre 1 e 3, e número de ligações rotacionáveis menor ou igual a 7, totalizando 6.053.287 de moléculas.⁸⁷ O subconjunto *drug-likeness*, diferentemente do *lead-likeness*, possui características físico-químicas desejáveis para transpor membranas biológicas e satisfazem a “regra dos 5 de Lipinski”,⁸⁸ tais como peso molecular menor que 500 Da, número de doadores de ligação de hidrogênio menor ou igual a 5, número de aceptores de ligações de hidrogênio menor ou igual a 10 e cLog *P* menor que 5, contendo 17.900.742 de estruturas.^{85,87} O subgrupo “*fragment-likeness*”⁸⁹ contém moléculas menores quando comparado ao subgrupo “*lead-likeness*”, totalizando 847.909 compostos.^{83,85} Por fim, os subgrupos “*all purchasable*” e “*shards*” contêm 22.724.825 e 635.159 compostos, respectivamente.^{83,85} Nessa tese, foi escolhido o subgrupo “*drug-likeness*”, pois este está diretamente relacionado com a biodisponibilidade dos fármacos, propriedades de solubilidade desejáveis, distribuição na corrente sanguínea e penetração pelas membranas biológicas.⁸⁸⁻⁹⁰ Essas características farmacocinéticas são essenciais, pois a forma amastigota do *T. cruzi* se encontra no interior da célula do hospedeiro. Além disso, o objetivo deste trabalho foi identificar moléculas que estão diretamente associadas a estabilidade metabólica e permeabilidade por membranas.

A partir das bibliotecas virtuais, pode-se aplicar os métodos LBDD e SBDD. A triagem virtual em larga baseada na estrutura do ligante (LBVS, do inglês *ligand-based virtual screening*) utiliza o princípio do método LBDD ou “aprendizado de máquina”, a partir do qual é desenvolvida uma função matemática (ou modelo de QSAR).^{72, 77} Após a construção dos modelos, são calculados os descritores moleculares dos compostos da biblioteca virtual.^{78,93} Posteriormente, a partir desses descritores é realizada a predição da propriedade-alvo.^{75-76,92} Assim, as estruturas são classificadas e aquelas que apresentam os melhores resultados são mantidas.^{68,78,95} Os compostos que ocupam as primeiras posições do *ranking* são

consideradas terem maior probabilidade de apresentar atividade.^{68,78,95} Assim sendo, essa abordagem tem sido considerada valiosa no processo de planejamento de fármacos.^{77,79,95} A triagem virtual em larga escala baseada na estrutura do alvo molecular (SBVS, do inglês *structure-based virtual screening*), no entanto, se baseia nas estratégias SBDD para predizer as poses e as afinidades de compostos no sítio ativo da macromolécula de interesse.⁹⁶ Tradicionalmente, é utilizada a docagem molecular para classificar as estruturas químicas mais ativas através das predições de afinidade.^{64,69,96} Os compostos no topo do *ranking* são selecionados e testados frente a enzima de interesse para determinação de parâmetros experimentais de potencia e afinidade.^{64,69,96}

Os métodos de planejamento baseados na estrutura do ligante e do receptor são geralmente realizados de forma integrada na indústria farmacêutica e na academia.^{73,97} A utilização dessa abordagem integrada maximiza a probabilidade de sucesso nos testes *in vitro*, sendo possível explorar um vasto espaço químico na descoberta de novas NCEs.

1.6 Propriedades farmacocinéticas

A fase farmacodinâmica é uma característica fundamental associada à potência de um fármaco ao se ligar ao alvo molecular. Entretanto, esse parâmetro não nos informa sobre a biodisponibilidade de um fármaco.⁹⁸ Grande parte dos compostos em fase clínica de desenvolvimento é descontinuada devidos problemas farmacocinéticos, tais como baixa biodisponibilidade, alta toxicidade e interações medicamentosas indesejáveis. Assim, considerar aspectos farmacocinéticos associados a absorção, distribuição, metabolismo, excreção e toxicidade (ADME-Tox, do inglês *absorption, distribution, metabolism, excretion and toxicity*) tem se tornado imprescindível para a indústria farmacêutica.⁹⁸ Conciliar de forma integrada os perfis farmacodinâmicos e farmacocinéticos de um novo protótipo é uma estratégia para diminuir a taxa de insucesso de candidatos a fármacos na fase clínica de desenvolvimento.

As propriedades ADME-Tox são parâmetros fundamentais para tornar um determinado composto biodisponível para que ele exerça sua função terapêutica.⁹⁸ A absorção é regida por um balanço entre a permeabilidade, ou sua capacidade de

cruzar bicamada lipídica, e a solubilidade.⁹⁹ A distribuição do fármaco no organismo está diretamente associada à sua ligação às proteínas plasmáticas, sendo essas proteínas responsáveis pelo transporte da molécula até o alvo de interesse.¹⁰⁰ O metabolismo é uma característica determinante para a biodisponibilidade de um fármaco.¹⁰⁰ As estimativas *in vitro* para metabolismo são geralmente feitas em microsossomos de fígado (metabolismo de primeira passagem) e hepatócitos (metabolismo de segunda passagem).¹⁰⁰ Um fármaco com baixa estabilidade metabólica exigirá múltiplas doses para manter a janela terapêutica. Por outro lado, um fármaco muito estável possui maior tempo de meia-vida e pode se tornar inseguro. Outros fatores podem influenciar o metabolismo de um fármaco, tais como as interações medicamentosas, as quais podem diminuir sua eficácia.¹⁰¹ Por fim, a toxicidade também é um parâmetro primordial que avalia a segurança de um fármaco. Assim, a determinação de cada propriedade ADME-Tox *in silico* e posterior filtragem de bancos de dados de acordo com essas propriedades, é uma estratégia útil na identificação e seleção de compostos promissores para a realização de testes experimentais.

1.7 Dinâmica molecular e a cruzaina

A dinâmica molecular tornou-se um método importante na pesquisa e desenvolvimento de novos fármacos,^{67,100} por oferecer condições de observar interações entre o ligante e o alvo molecular, considerando a flexibilidade do receptor.¹⁰² Assim, as aplicações da dinâmica molecular tem ganhado destaque nos esforços de descoberta de novos inibidores enzimáticos.^{67,100-102}

A dinâmica molecular utiliza as soluções das equações diferenciais ordinárias de Newton (eq. 1) para encontrar as conformações de equilíbrio entre o ligante e a macromolécula.¹⁰² Essa abordagem computacional adota os átomos como esferas e ligam essas esferas como sistemas massa-mola (oscilador harmônico clássico).¹⁰⁰⁻¹⁰¹ Isso permite que os átomos no sistema apenas oscilem dentro de uma distância previamente especificada.¹⁰²

$$\vec{F}_i(r) = m_i \frac{d^2 \vec{r}}{dt^2} \quad (1)$$

Onde, $\vec{F}_i(r)$, t , m_i e $\vec{r}_i$ são o vetor da força resultante, tempo, massa e vetor posição, respectivamente.

O campo de força compreende os termos da energia potencial das ligações químicas – considerando suas hibridizações, ângulos de torsão e diédricos – e os termos das interações fracas – potenciais de van der Waals e de Coulomb.¹⁰² Os termos de ligação química, comumente chamados de parte ligada, consideram o tipo de ligação química e os ângulos de torsão como termos harmônicos, e os diédricos como uma função cosseno.¹⁰³ A parte que considera as interações fracas, chamadas de parte não ligada, considera os potenciais de Lennard-Jones e a lei de Coulomb.¹⁰³ Baseado nessas informações, diversos campos de força são aplicados em dinâmica molecular, como o CHARMM (Chemistry at HARvard Macromolecular Mechanics),¹⁰⁵ AMBER (Assisted Model Building with Energy Refinement),¹⁰⁶ OPLS (Optimized Potential for Liquid Simulations)¹⁰⁷ e GROMOS (GRONingen Molecular Simulation).¹⁰⁸ Diante dessa diversidade de campos de força, é importante que o sistema a ser estudado seja conhecido detalhadamente para que se obtenha resultados mais precisos.^{67,100,107-109}

Diversos estudos de dinâmica molecular foram realizados acerca da cruzaina, desde análises estruturais até estratégias combinadas com métodos LBDD e SBDD. Dentre esses trabalhos, se destacam os estudos: (1) para investigar as conformações da apoenzima e da holoenzima em pH 5,5;¹¹² (2) para aperfeiçoar a potência de derivados de quinolinas através da combinação de docagem molecular e dinâmica molecular;¹¹³ (3) e para determinar as interações intermoleculares de inibidores baseados em α -ceto e combiná-la com resultados obtidos em análises CoMFA (Comparative Molecular Field Analysis) e CoMSIA (Comparative Molecular Similarity Index Analysis).¹¹⁴ Todos os resultados demonstraram que as propriedades físico-químicas dos ligantes e estruturais da cruzaina estão diretamente relacionadas entre si. No entanto, ainda não há descrição que correlacionem mudanças estruturais da cruzaina dependentes de pH e a atividade de inibidores da enzima.

2 OBJETIVOS

O presente trabalho de doutorado tem como objetivo a construção de dois conjuntos de dados e o desenvolvimento de modelos de QSAR, e a realização de triagens virtuais baseadas na estrutura do ligante e do receptor, e em características físico-químicas e de ADME-Tox. Outro objetivo consiste na investigação da influência do pH no modo de ligação dos inibidores usando estratégias de dinâmica molecular acoplado a redes neurais artificiais. Esta tese também inclui a construção de modelos de QSAR para mapear o sítio ativo da cruzaina.

Dentre os objetivos específicos da triagem virtual destacam-se:

- ✓ Construção das bases de dados para elaboração dos modelos de QSAR utilizando as variáveis dependentes pIC_{50} (conjunto 1) e pSI (conjunto 2)
- ✓ Aplicação de métodos de redes neurais artificiais (RNAR), regressão por KPLS, e métodos de árvore de decisão aleatória;
- ✓ Caracterização estrutural e elaboração de bioisómeros baseados em KPLS 2D combinado com predição das redes neurais;
- ✓ Filtragem da base de dados da ZINC12 (categoria “*drug-likeness*”) baseado na floresta randômica;
- ✓ Aplicação dos métodos de busca e predição da energia de ligação por docking molecular;
- ✓ Filtragem baseada na predição da seletividade, permeabilidade, e características ADME-Tox;
- ✓ Predição dos pontos de clivagem pela isoforma CYP3A4.

Dentre os objetivos específicos da dinâmica molecular destacam-se:

- ✓ Estudos dos estados de protonação do ligante e do alvo molecular;
- ✓ Simulações de dinâmica molecular das configurações apo e holo da cruzaina em diferentes condições de pH;
- ✓ Definição das coordenadas de hélices- α , folhas- β e alças, e dos deslocamentos dessas estruturas secundárias;

- ✓ Estudos quantitativos das relações entre a estrutura e atividade do complexo enzima/inibidor através das redes neurais artificiais;

Dentre os objetivos específicos do mapeamento do sítio ativo por modelagem de QSAR 2D e 3D e docagem molecular destacam-se:

- ✓ Desenvolvimento de um conjunto de dados para a construção de modelos de QSAR;
- ✓ Aplicação de métodos de QSAR 2D e 3D;
- ✓ Mapeamento das interações intermoleculares no sítio ativo da cruzaina usando estratégias de docagem molecular;
- ✓ Avaliação da capacidade preditiva dos modelos gerados;

3 MATERIAIS E MÉTODOS

3.1 TRIAGEM VIRTUAL EM LARGA ESCALA

Na figura 12 é mostrado o diagrama resumido do trabalho. Inicialmente, foi feita uma ampla busca de periódicos no banco de dados *Web of Science*. Em seguida, foram selecionados os artigos que continham estruturas biologicamente ativas na forma amastigota intracelular de *T. cruzi*. A partir das informações de IC_{50} e SI (*selectivity index*), foram construídos dois conjuntos para construção de modelos de QSAR: o conjunto 1 (variável dependente IC_{50}) e o conjunto 2 (variável dependente SI). Ambos os conjuntos foram submetidos a uma análise de componentes principais (PCA, do inglês *principal component analysis*), para verificar a distribuição do espaço químico e biológico através do mapa de similaridade estrutural. Para cada conjunto, foram selecionados os conjuntos treinamento e teste para construção dos modelos de QSAR. O conjunto 1 foi usado para treinar as redes neurais de retropropagação e a construir o modelo KPLS *molprint* 2D. O conjunto 2 foi utilizado para construir o modelo de regressão linear considerando pIC_{50} e pSI como variáveis dependentes e independentes, respectivamente. Além disso, esse conjunto foi usado para treinar a árvore de decisão randômica e desenvolver o modelo KPLS radial. Todos esses modelos foram aplicados como filtros ao banco de dados ZINC 12. Posteriormente, foi realizado uma triagem virtual baseada no alvo molecular, utilizando-se os compostos filtrados como entrada para prever as energias de ligação no sítio ativo da cruzaina. Em seguida, foi realizada uma filtragem baseada no teste de permeabilidade física e nas propriedades ADME-Tox.

Figura 12 - Diagrama esquemático do processo de triagem virtual.

Fonte: Elaborada pelo autor.

3.1.1 Busca no banco de dados

Inicialmente, foi realizada uma busca de periódicos no banco de dados *Web of Science* a partir das palavras-chaves “*Chagas disease AND inhibitors*” and “*Trypanosoma cruzi AND inhibitors*”. Essa busca retornou 973 e 1893 artigos, respectivamente. A primeira filtragem foi baseada nos artigos que continham somente compostos testados em amastigotas intracelulares. Nessa seleção, não foi considerado o tempo de exposição dos compostos contra o parasita e nem a linhagem de células que foram infectadas. Assim, foram construídos dois conjuntos para as construções dos modelos de QSAR, os conjuntos 1 e 2, com variáveis dependentes IC₅₀ e SI.

No conjunto 1, os periódicos que não possuíam moléculas com valores de IC₅₀ determinados (concentração necessária para inibir 50% o crescimento do parasita), e/ou elementos da família B e/ou carboidratos na sua composição foram eliminados, restando 39 artigos.¹¹³⁻¹⁵¹ Todos os compostos que não possuíam propriedade biológica definida, *i.e.*, IC₅₀ > ou < ou ~ ou apresentavam IC₅₀ > 10⁻⁴ M, ou eram estruturas químicas duplicadas foram ignorados. Todas as estruturas foram construídas na plataforma Canvas 2.0.¹⁵⁴ O IC₅₀ foi convertido para pIC₅₀, *i.e.*, escala logarítmica na base 10 de 1/IC₅₀. Na etapa de construção das estruturas químicas, os estereoisômeros foram mantidos na sua forma racêmica para simplificar os modelos. Assim, o banco de dados possui 363 compostos bioativos em amastigota (disponível no material suplementar na forma de *SMILES*).

Para construção do conjunto 2, foram selecionados os periódicos que possuíam estruturas químicas com valores de índice de seletividade (SI) determinado experimentalmente, permanecendo 26 artigos.^{113,115,118,120-121,123,125-127,130-134,136-139,142-144,146-148,150,153} O índice de seletividade é a relação entre a concentração de composto necessária para inibir em 50% o crescimento de células saudáveis ($IC_{50}^{célula}$) e a concentração de composto necessária para inibir 50% do crescimento do parasita ($IC_{50}^{amastigota}$), ou seja, $SI = IC_{50}^{célula} / IC_{50}^{amastigota}$. O SI foi convertido em pSI, escala logarítmica na base 10 de SI. Nesse conjunto, foram considerados os compostos em mistura racêmica para simplificação dos modelos de QSAR.

3.1.2 Análise por componentes principais

Para verificar a similaridade dos compostos nos conjuntos 1 e 2, foram realizados estudos de similaridade estrutural a partir da análise de componentes principais (PCA), pelo método UNITY no programa SYBYL X 2.0.¹⁵⁶ Nesse método, os compostos do conjunto de dados são representados a partir da impressão digital UNITY. Essas representações são os descritores moleculares e são obtidos pelo comprimento e pelo número de ocorrência de um determinado fragmento na estrutura. A partir dos descritores, é construída uma matriz de correlação e são determinados os dois primeiros componentes para gerar as coordenadas iniciais do mapa. As distâncias de similaridade de Tanimoto entre as impressões digitais UNITY foram usados para determinar todos os pontos. Para ambos os conjuntos, foi estabelecido que compostos que apresentassem similaridade menor que 80% fossem adicionados à margem do mapa.

3.1.3 Descritores físico-químicos

Para os dois conjuntos, foram calculados os descritores físico-químicos: peso molecular (MW), lipofilicidade ($aLogP$), número de doadores de ligação de hidrogênio (HBD), número de aceptores de ligação de hidrogênio (HBA), número de ligações rotacionáveis (RB), área de superfície polar (PSA), estados eletrotológicos (Estate), refratividade molar (MR), polarização (polar), número de

átomos pesados (HAC) e de anéis (RC), utilizando-se o programa Canvas.¹⁵⁴ O efeito lipofílico é um balanço das propriedades hidrofóbicas e hidrofílicas.¹⁵⁷ Esta relação é importante para um fármaco cruzar membranas biológicas e ligar-se ao seu alvo molecular. Assim, encontrar padrões de propriedades físico-químicas, através da correlação entre pIC_{50} e pSI , é um aspecto importante para que um potencial candidato a fármaco chegue ao seu alvo molecular e interaja de modo seletivo, produzindo o efeito esperado.

3.1.4 Redes neurais artificiais de retropropagação

As redes neurais artificiais foram usadas para correlacionar os descritores físico-químicos com pIC_{50} . Para treinar as redes neurais artificiais, foi utilizado o programa WEKA 3.6 (Waikato Environment of Knowledge Analysis, desenvolvido pela Universidade de Waikato, versão 3.6).¹⁵⁸ Esse programa é de acesso gratuito e possui algoritmos de aprendizado de máquina para aplicação em mineração de dados.¹⁵⁸ Os descritores físico-químicos e propriedades biológicas são utilizados como atributos de entrada e saída, respectivamente. Todas as informações relacionadas a características físico-químicas previamente determinadas e as atividades biológicas foram adicionadas a um arquivo ARFF, e posteriormente foi utilizado o algoritmo de redes neurais artificiais (RNAs) de retropropagação em multicamadas. Este algoritmo é capaz de encontrar um padrão comum entre os descritores físico-químicos do banco de dados através de um conjunto treinamento. Posteriormente, os parâmetros são utilizados para predição dos compostos excluídos, para validação externa.

As RNAs são semelhantes às redes neurais biológicas sendo formadas pela unidade básica, o *perceptron* ou neurônios, possuindo os dendritos, o corpo celular e os axônios (fig. 13).¹⁵⁷⁻¹⁵⁸ Todos os dendritos são somados para gerar uma função de aprendizagem (corpo da célula).¹⁵⁷⁻¹⁵⁸ As entradas são as variáveis independentes, e os valores de saída são as variáveis dependentes. O sinal (axônio) pode ser propagado ou inibido se o valor retornado pela função de ativação for maior ou menor que um valor predeterminado.¹⁵⁷⁻¹⁵⁸ Um conjunto de *perceptrons* pode gerar camadas, sendo conhecidos como *perceptrons* multicamadas (MLP).¹⁶¹ O MLP possui vários neurônios interconectados entre si.¹⁶¹ Esta conexão é responsável

pela propagação do sinal nas camadas intermediárias, ou seja, a entrada no neurônio produz um sinal de saída e esse torna-se uma entrada para um neurônio posterior.¹⁶¹

Figura 13- Diagrama esquemático das redes neurais artificiais.

Fonte: Elaborada pelo autor.

Em particular, o algoritmo de retropropagação utiliza duas etapas: a parte de avanço e a parte de retrocesso.¹⁶² Na parte de avanço, os pesos são aleatoriamente determinados e o valor da atividade biológica é predito.¹⁶² Na parte de retrocesso, os erros associados entre os valores experimentais e preditos são fundamentais para corrigir os pesos de entrada até a primeira camada intermediária.¹⁶² Nessa correção, o ajuste é feito pela derivada parcial em relação ao peso da função de ativação.¹⁶² Para derivadas positivas, o peso de entrada aumenta o valor do erro.¹⁶² Caso contrário, essa diferença diminui a relação entre o valor predito e experimental.¹⁶²

A principal limitação do algoritmo de retropropagação é a convergência da rede devido a valores baixos e altos na taxa de aprendizagem.¹⁶² Para reduzir essa limitação, o termo momento garante estabilidade das RNAs.¹⁶² Nesse estudo, a função logística foi usada como função de ativação, cujo taxa de aprendizado variou de 0,1 a 0,4 e momento de 0,1 a 0,4. Por fim, o modelo foi melhorado variando o número de neurônios (1 a 10) na camada oculta.

3.1.5 Descritores moleculares de impressão digital e KPLS

Para os conjuntos 1 e 2 foram estudados os principais arcabouços estruturais que contribuem positivamente para a morte do parasita (pIC_{50}) e a seletividade (pSI), através do método KPLS baseado no descritor de impressão digital molecular.¹⁶³ Esse descritor é calculado por meio da representação SMILES das estruturas do conjunto de dados.¹⁶³ Estes descritores podem ser classificados, dentre outros, em dendritic, radial, molprint2D e linear.¹⁶⁴ Esses quatro descritores permitem a visualização dos mapas de contribuição atômica pelo método KPLS (*kernel-based partial least square*).¹⁶⁵ O descritor linear utiliza a informação dos fragmentos lineares e do fechamento do anel para converter a estrutura química na sequência numérica binária.¹⁶⁴ O *dendritic* usa a mesma informação do descritor linear, porém, adiciona as informações das partes ramificadas da molécula na etapa de construção da sequência binária.¹⁶⁴ O radial identifica todos os átomos pesados e codifica o composto pelo crescimento radial.¹⁶⁴ Por fim, o molprint2D usa o mesmo princípio do descritor linear, porém, o algoritmo codifica até duas ligações químicas.¹⁶⁴ Além disso, ele considera a frequência e a distância em relação ao tipo de átomo codificado.¹⁶⁴ Em seguida, o método KPLS é aplicado aos quatro descritores moleculares de impressão digital variando-se o parâmetro de número de fatores (N).¹⁶⁵ Por fim, o melhor modelo pode ser verificado através da melhor função de pontuação, a qual relaciona r^2 e q^2 , por $q^{2*}(1-|q^2 - r^2|)$.¹⁶⁶ No presente trabalho, foram variados os tipos de descritores de impressão digital (molprint2D, radial, linear e *dendritic*) e o N foi limitado a 4.

3.1.6 Floresta randômica

Para o conjunto 2, foi avaliado a combinação dos descritores físico-químicos para gerar um conjunto de árvores de decisão, a floresta randômica. A floresta randômica é criada pelo método de *bagging*, na qual são geradas diversas amostragens aleatórias (fig. 14).^{74,165} Assim, através dessas amostragens, são criadas diversas árvores de decisão, as quais são utilizadas em conjunto para prever a atividade biológica de interesse.^{74,165} O nó de cada árvore de decisão gerada representa um teste de um determinado atributo, o ramo representa a sua saída e a

folha representa o rótulo da atividade predita após a tomada de decisão da árvore.^{76,}
167

Figura 14 - Diagrama esquemático da floresta randômica.
Fonte: Elaborada pelo autor.

Para construir a floresta aleatória, foi utilizado o programa WEKA 3.6 (*Waikato Environment of Knowledge Analysis*, desenvolvido pela Universidade de Waikato, versão 3.6).¹⁵⁸ A floresta aleatória foi utilizada para detectar padrões que podem influenciar significativamente a seletividade (pSI). Além disso, devido à quantidade de árvores geradas, o desempenho de predição da propriedade biológica torna-se mais eficiente e estável que outros métodos de aprendizado de máquina na etapa de triagem virtual. Para o conjunto 2, foram consideradas a construção de 100 árvores aleatórias baseadas nas informações dos descritores físico-químicos, ensacamento de 100% e 100 estruturas para predição interna.

3.1.7 Regressão linear baseada em dados experimentais

Considerando a definição de SI, foi aplicado a função logarítmica em ambos os lados da equação 2, a partir da qual foi obtido uma equação da reta relacionando pSI e $pIC_{50}^{amastigota}$.

$$SI = \frac{IC_{50}^{célula}}{IC_{50}^{amastigota}} \quad (2)$$

$$\text{Log } SI = \text{Log } IC_{50}^{célula} - \text{Log } IC_{50}^{amastigota} \quad (3)$$

Usando a definição de pSI, pIC_{50} e adotando $\text{Log } IC_{50}^{célula} = \alpha$, é possível obter a equação da reta (eq. 4):

$$pSI = pIC_{50}^{amastigota} - \alpha \quad (4)$$

onde, o coeficiente angular é igual a 1 e o coeficiente linear é dado por α . Assim, será definida a constante citotóxica, α , a qual possui a dimensão da medida de toxicidade de um determinado composto em células saudáveis.

3.1.8 Triagem virtual em larga escala baseada no ligante

Na pesquisa de novos compostos com atividade anti-*T. cruzi* foram aplicados os modelos de QSAR no banco de dados ZINC 12,^{80-81,83,166} sub-grupo *drug-likeness*. Os mesmos descritores utilizados para construção dos modelos de QSAR, foram aplicados nessa base de dados. Na figura 15, são mostradas duas etapas de filtragem. A primeira filtragem foi realizada a partir das informações extraídas da floresta randômica, considerando pIC_{50} maior que 6 e pSI maior que 2 (descrito na sessão resultados e discussão). Assim, foram selecionados por volta de 2,8 milhões de estruturas químicas. Na segunda filtragem, as estruturas selecionadas foram submetidas à predição direta de pIC_{50} , pelas redes neurais artificiais e KPLS molprint2D, e predição indireta de pIC_{50} . Os valores de pSI foram preditos pela floresta randômica e de pIC_{50} através da equação da reta obtida experimentalmente (eq. 8, descrito na sessão resultados e discussão). Em seguida, foi analisado o consenso das predições dos modelos. Todas as estruturas que apresentavam desvio com relação a média maior que 0,4 foram eliminados. Dessa maneira, permaneceram cerca de 1,0 milhão de compostos, os quais foram preparados para a triagem virtual baseada na estrutura do alvo molecular.

Filtragem 1

Filtragem 2

Figura 15 - Etapas de aplicação dos métodos de filtragem baseados em QSAR.
Fonte: Elaborada pelo autor.

3.1.9 Triagem virtual baseada na estrutura do alvo molecular

3.1.9.1 Preparação dos ligantes

Para simular a docagem molecular, é importante preparar os ligantes em sua conformação 3D e determinar os estados de protonação adequados para o pH fisiológico. Assim, o LigPrep¹⁶⁹ foi usado para essa finalidade. Primeiro, os estados de protonação dos ligantes em pH 8,0 foram preditos pelo módulo Epik.¹⁷⁰ Em seguida, foi realizado o processo de minimização de energia aplicando água como solvente e o campo de força OPLS3.¹⁷¹ O campo de força OPLS3 foi escolhido porque possui um bom modelo de carga e boa precisão nos parâmetros de torção em comparação com os campos de forças convencionais.¹⁷¹

3.1.9.2 Preparação da cruzaina

Nos estudos de docagem molecular, a estrutura da cruzaina (PDB ID 3KKU, resolução 1,28 Å) foi selecionada.¹⁷² Os critérios de escolha foram baseados na ocupação do ligante co-cristalizado no sítio ativo da enzima, natureza não covalente do inibidor e boa resolução estrutural. A proteína foi preparada usando o módulo *Protein Preparation Wizard*.¹⁷³ As baixas ocupâncias foram selecionadas. A estrutura foi submetida à análise de perda das cadeias laterais e conformação das principais cadeias contendo regiões de *loops* através do protocolo Prime.¹⁷⁴ Todas as moléculas de águas que apresentavam distâncias maiores que 5 Å de qualquer heteroátomo foram eliminadas. Os estados de protonação das cadeias laterais dos aminoácidos foram preditos pelo método de Jensen¹⁷⁵ em pH 7,0. Esse pH foi escolhido, pois a cruzaina está localizada na superfície celular na forma amastigota de *T. cruzi*.⁴⁰ Por fim, a macromolécula foi minimizada através do campo de força OPLS3.

3.1.9.3 Docagem molecular

Após a preparação da cruzaina, foram melhorados os parâmetros para reprodução dos dados experimentais, comparando a conformação do ligante docado com a do ligante co-cristalizado. O programa GLIDE (do inglês *Grid-based Ligand Docking with Energetics*) possui diversas opções desempenho e precisão do ligante docado.¹⁷⁶⁻¹⁷⁷ O programa utiliza uma sequência de filtros para encontrar posições favoráveis para que um determinado ligante ocupe o sítio ativo do alvo molecular.¹⁷⁶⁻¹⁷⁷ Assim, é construída uma grade, na qual são geradas diversas conformações considerando a flexibilidade dos compostos.¹⁷⁴⁻¹⁷⁵ Todas as conformações geradas fora da grade são excluídas e as conformações no seu interior são avaliadas através da função de pontuação.¹⁷⁴⁻¹⁷⁵ A precisão da amostragem é dada pelos modos escolhidos, ou seja, cada modo determinará a flexibilidade para gerar o espaço amostral de conformações.¹⁷⁴⁻¹⁷⁵ Para estudos de triagem virtual em larga escala, o modo HTVS (do inglês *high-throughput virtual screening*) pode ser utilizado com pouca flexibilidade.¹⁷⁴⁻¹⁷⁵ O modo SP (do inglês *standard precision*) possui melhor flexibilidade conformacional e consegue prever a conformação de milhares de compostos.¹⁷⁴⁻¹⁷⁵ Por fim, o modo XP (*extra precision*) pode ser empregado para gerar uma ampla amostragem conformacional.¹⁷⁴⁻¹⁷⁵ Dessa maneira, o modo XP diminui a probabilidade de encontrar falsos positivos e aumenta a chance de se obter resultados mais confiáveis.¹⁷⁶⁻¹⁷⁷ Por fim, após gerar as poses, os ligantes são minimizados e as orientações são avaliadas pela função de regressão GlideScore ou GScore (eq. 5).¹⁷⁴⁻¹⁷⁵

$$GS = 0,05.vdW + 0,15.Coul + Lipo + Hbond + Metal + Retr + RotB + Sitio \quad (5)$$

Onde GS é o valor da energia livre estimada (GlideScore), e vdW e Coul são as energias de van der Waals e Coulomb calculados de grupos iônicos, tais como, metais, carboxilatos e guanidínicos. Lipo é o termo lipofílico calculado a partir de interações hidrofóbicas, Hbond é o termo devido a ligação de hidrogênio, e metal é o termo de energia para ligação metálica na qual são considerados os átomos aniônicos e os átomos considerados grandes aceptores polares. Retr é o termo energético para as penalidades geradas no sítio de ligação, RotB é a penalidade para

ligações rotacionáveis congeladas, e Sitio é o termo energético para as interações polares excluindo os átomos que fazem ligação de hidrogênio.¹⁷⁴⁻¹⁷⁵

O programa GOLD (*Genetic Optimization for Ligand Docking*) é um programa que usa um algoritmo genético para gerar as conformações dos ligantes no sítio de ligação.¹⁷⁶⁻¹⁷⁷ Inicialmente, é gerada uma ampla amostragem de populações. Cada indivíduo possui as informações dos ângulos e esses são codificados em genes. Posteriormente, esses indivíduos se cruzam, par a par, trocando as informações dos genes, podendo sofrer mutações. As informações trocadas geram novas poses, e essas são avaliadas por uma função de pontuação. Assim, é realizada uma seleção dos indivíduos aptos a dar continuidade nas informações geradas pelo cruzamento dos pais. Esse processo é repetido diversas vezes até alcançar o número máximo de iterações ou a diferença mínima do estado energético estipulado pelo operador.¹⁷⁶⁻¹⁷⁷

A partir dessas informações, a preparação dos parâmetros de docagem foi dividida em 4 etapas (fig. 16): os dois primeiros passos foram baseados no algoritmo do GLIDE, o terceiro passo pelo GOLD e o quarto novamente pelo GLIDE. Em todos os passos do GLIDE, foi definido o centróide do ligante (14, 12, -15) como o centro da grade (cúbica, aresta de 12 Å). Posteriormente, a simulação de docagem molecular foi realizada no modo HTVS (1ª etapa) e SP (2ª etapa), e as poses avaliadas pela função empírica GScore. Em seguida, as moléculas filtradas foram submetidas ao programa GOLD, no qual o sítio ativo foi definido por uma esfera de raio 12 Å e centrada no átomo de enxofre da CYS25 (-5, -33, 0,5) e utilizando o ligante co-cristalizado como referência para gerar as poses. A função de aptidão usada para avaliar as poses do algoritmo genético foi baseada em campo de força, GoldScore. Os operadores genéticos foram mantidos no modo automático, com eficiência de 30%. A energia livre de ligação foi calculada baseada em modelos de regressão linear obtido de Verdonk e colaboradores.¹⁷⁹ Por fim, a última etapa da docagem molecular foi realizada novamente no programa GLIDE no modo XP.

Figura 16 - Etapas de aplicação dos métodos de docagem molecular.

Fonte: Elaborada pelo autor.

Na sequência das operações dos métodos de docagem molecular, o número de estruturas foi reduzido de 1.000.000 para 945.632, 944.551, 425.341 e 412.953 compostos, respectivamente (fig. 17a). Após realizar os quatro processos de filtragem foram analisados o consenso das funções de pontuação, através da energia livre de Gibbs, para o modo XP (GLIDE, pelo Gscore) e para o modo HTVS (GOLD, pelo GoldScore convertido em energia livre pelo método de regressão linear).¹⁷⁹ Para isso foi realizado o cálculo da média das energias de ligação e feito o cálculo do desvio com relação à média. O primeiro critério foi excluir todos os compostos que apresentavam desvio médio absoluto maior que 1,2 kcal/mol e, posteriormente, eliminar pelo segundo critério todas as estruturas químicas que apresentavam valores médios maiores que -6,5 kcal/mol. Assim, as 412.953 estruturas químicas foram reduzidas para 108.273 e 51.369, respectivamente (fig. 17b).

Parte 1

Parte 2

Figura 17 - Diagrama esquemático dos processos de filtragem baseados em docagem molecular; (parte 1) filtro baseado na cavidade enzimática e (parte 2) filtro baseado no consenso das funções de pontuação.

Fonte: Elaborada pelo autor.

3.1.10 Filtragem baseada na predição da seletividade

Após o processo de filtragem baseado em diferentes métodos de docagem molecular, foram utilizados os modelos de floresta randômica e KPLS radial para prever o índice de seletividade dos 51.369 compostos filtrados. No processo de seleção dos compostos seletivos, foi feita uma análise com relação à média dos valores de pSI calculados. Durante a análise realizada dos valores preditos, foi

verificado que os compostos que possuíam pSI médio maior ou igual a 0.995 apresentavam, para a maioria, valores de pSI preditos positivos para ambos os modelos. Assim, o número de estruturas foi reduzido para 30.566.

3.1.11 Teste físico de permeabilidade de membrana

Considerando que a forma amastigota se encontra no interior da célula hospedeira, procuramos selecionar compostos que possam atravessar a bicamada lipídica e alcançar o alvo molecular. Os modelos de QSAR para predição de permeabilidade são rápidos e consideram somente a estrutura dos compostos.^{98,180} Entretanto, os modelos baseados em testes físicos oferecem melhor precisão para permeabilidade, pois abordam as ligações de hidrogênio intramoleculares. Essas ligações podem afetar significativamente a permeabilidade.¹⁸¹

A base biofísica da permeabilidade passiva em membrana abordada no presente trabalho é baseada em mecânica molecular, pois utiliza os parâmetros de entrada baseados em campo de força, considerando as cargas parciais e um modelo implícito de solvente (fig. 18).¹⁸¹ O primeiro passo consiste na predição do pKa e gerar todos os estados de protonação pelo módulo do Epik.¹⁷⁰ Posteriormente, são geradas diversas conformações e, para cada uma delas são determinadas as energias em água e clorofórmio.¹⁸¹ Após essa etapa, as conformações são submetidas a um processo de filtragem energética, no qual todas aquelas que apresentarem energia maior que 5 kcal/mol são eliminadas.¹⁸¹ Por fim, são determinados $\Delta G_{\text{transf.}}$, $\Delta G_{\text{C/W}}$, K_{barreira} , D_{barreira} , e P_m para cada conformação.¹⁸¹ O $\Delta G_{\text{transf.}}$ é calculado pela diferença de energia das conformações na água e no clorofórmio.¹⁸¹ O $\Delta G_{\text{C/W}}$ é determinado pela soma de $\Delta G_{\text{transf.}}$, ΔG_{penal} (penalidade da conformação) e $\Delta G_{\text{est.}}$ (penalidades dos estados de tautomerização/deionização).¹⁸¹

Figura 18 - Diagrama esquemático do teste físico de permeabilidade de membrana.

Fonte: Elaborada pelo autor.

As moléculas que foram filtradas pelo modelo de floresta randômica e KPLS radial foram submetidas ao teste físico de permeabilidade. Nessa etapa, o número de estruturas foi reduzido de 30.566 para 1.120 compostos.

3.1.12 Filtragem baseada nas características ADME-Tox

Para detectar previamente possíveis problemas associados a propriedades de ADME-Tox foram determinados através do módulo QikProp as propriedades de permeabilidade em células de Caco-2 (Caco-2), coeficiente de partição hematocefálica ($\log P_{bb}$), ligação em albumina humana (AH), porcentagem de absorção oral (PAO) e predição de pIC_{50} para bloqueio de canais hERG¹⁸². Após

calcular essas propriedades, todas as moléculas que apresentavam Caco-2 maior que 500 nm/s, $\text{Log}P_{\text{bb}}$ entre -3,0 e 1,2, AH entre -1,5 e 1,5, PAO maior que 80% e hERG maior que -5 foram consideradas aptas para o ensaio virtual.¹⁸² Assim, o número de estruturas foi reduzido para 260 compostos.

3.1.13 Filtragem baseada na distância da reta ideal

Após a análise da correlação dos compostos, foram selecionados os compostos pelo critério da distância da curva ideal (eq. 6), na qual os pontos correspondentes à energia livre e atividade biológica estão próximo dessa reta.

O cálculo da distância $d_{P,t}$ do ponto P (x,y) a reta t: $y - x = 0$ é definido como:

$$d_{P,t} = \frac{|y-x|}{\sqrt{2}} \quad (6)$$

Onde (1) x e y são $\Delta G^{\text{GoldScore}}$ e $\Delta G^{\text{XP GlideScore}}$, e (2) x e y são $\text{pIC}_{50}^{\text{RNA}}$ e $\text{pIC}_{50}^{\text{KPLS molprint2D}}$. Para (1), foi considerada uma distância menor que 0,35 e, para (2), uma distância menor que 0,2. Em seguida, as estruturas que não possuíam a mesma orientação no sítio ativo, foram excluídas.

3.1.14 Predição dos pontos de clivagem no enzima CYP3A4

As enzimas do citocromo P450 são fundamentais no metabolismo de fármacos.¹⁸¹⁻¹⁸³ Essas enzimas possui um grupo prostético porfirínico heme, e promovem o metabolismo de fármacos através de reações de desalquilação e hidroxilação, dentre outrass.¹⁸⁶ Portanto, para se ter melhor controle metabólico do fármaco, é importante predizer os principais pontos de clivagem dessas moléculas. Diversos estudos computacionais tem mostrado que melhores resultados de predição de reatividade estão associados à isoforma CYP3A4.¹⁸⁷ Isso ocorre porque não há uma preferência de orientação dos compostos no sítio de ligação dessa enzima,¹⁸⁸ sendo a reatividade dependente do acesso ao centro do Fe^{3+} do grupo prostético. Dessa maneira, o módulo *P450 Site of metabolism* combina o método de docagem molecular com mudanças estruturais da CYP3A4 por meio de uma série

de minimizações, utilizando o protocolo de ajuste induzido do Glide e o modo de refinamento do Prime.¹⁸⁹ Assim, o ligante pode deslizar através do sítio ativo em diversas conformações e alcançar o átomo de Fe^{3+} .¹⁸⁹ Posteriormente, são aplicadas as regras de reatividade intrínseca, baseadas nos modelos de Hammett e Taft,¹⁹⁰ na qual valores positivos e negativos são muito reativos e pouco reativos, respectivamente.¹⁸⁹ Por fim, as pontuações são calculadas, através da combinação linear da acessibilidade ao Fe^{3+} e da reatividade, e são representadas pelos círculos verdes.¹⁸⁹ A probabilidade de ocorrer clivagem no átomo aumenta proporcionalmente com o tamanho do círculo.¹⁸⁹

3.2 DINÂMICA MOLECULAR E REDES NEURAIS ARTIFICIAIS

Nessa etapa do trabalho, escolhemos a estrutura da cruzaina ligada ao inibidor K777. A escolha foi motivada, pelos avanços desse composto na fase pré-clínica,^{52,189} por apresentar parâmetros cinéticos definidos em pH 5,5 e 8,0,⁶³ e possuir estrutura cristalográfica com boa resolução (PDB ID 2OZ2, resolução de 1,95 Å).⁵⁰ A primeira etapa foi encontrar os valores do pK_a e determinar os estados de protonação do ligante co-cristalizado e das cadeias laterais dos aminoácidos protonáveis (ASP, ARG, LYS, GLU, GLN e CYS). Posteriormente, foram construídas as caixas de água para as configurações do sistema, simulando a cruzaina apo e holo nos meios ácidos e básicos, totalizando quatro sistemas diferentes. A seguir, os sistemas foram neutralizados através de adição de íons Na^+ para cada condição, e equilibrados nos sistemas canônicos (NVT, número de partículas, volume e temperatura constantes, respectivamente) e isotérmico-isobárico (NPT, número de partículas, pressão e temperatura constantes, respectivamente). Depois disso, foram calculadas as trajetórias das conformações através da dinâmica molecular e as energias de ligação dos complexos. Por fim, os deslocamentos das estruturas secundárias foram determinados e acoplados às redes neurais artificiais para os estudos quantitativos entre a estrutura e atividade do complexo holo nos diferentes pH.

3.2.1 Predição dos estados de protonação do ligante

Para simular as condições experimentais em diferentes pH, o inibidor K777 foi submetido ao protocolo de predição do pK_a no módulo Epik.¹⁷⁰ Após a predição, as estruturas foram minimizadas utilizando-se a água como solvente e o campo de força OPLS3.¹⁷¹ Foi observado que para os meios ácido e básico, o ligante co-cristalizado manteve o mesmo estado de protonação no grupo piperazina. Isso demonstra que o substrato e/ou inibidor não afeta a eficiência catalítica, a qual pode ser afetada por mudança dos estados de protonação dos aminoácidos em diferentes pH.

3.2.2 Predição dos estados de protonação dos aminoácidos

Devido ao tamanho dos átomos de hidrogênio, é difícil prever a protonação dos aminoácidos de estruturas que foram obtidas por cristalografia de raio-X. Portanto, foi necessário prever esses estados, utilizando-se o método de Jensen.¹⁷⁵ Para os meios ácidos e básicos, resíduos de GLN foram desprotonados (neutros), ARG e LYS foram protonadas (carga positiva), e ASP e GLU foram desprotonados (carga negativa), o nitrogênio da ALA1 no N-terminal foi protonado (carga positiva) e no C-terminal foi considerado o íon carboxilato ($-\text{COO}^-$) no resíduo GLY215. As histidinas, na sua maioria, foram os únicos aminoácidos a variarem os estados de protonação nos dois pH. As configurações das quatro histidinas foram: em pH 5,5, a HIS43 foi protonada em ambos os nitrogênios (HIP43, carga positiva) e, no pH 8,0, foi protonada somente no N δ 1 (HID43, carga neutra); nos dois pH, a HIS106 foi protonada em ambos os nitrogênios (HIP106 ou HIS106⁺, carga positiva); no pH 5,5, HIS115 foi protonada em ambos os nitrogênios (HIP115 ou HIS115⁺, carga positiva), e no pH 8,0, somente o N ϵ foi protonado (HIE115, neutra); por fim, o par catalítico CYS25 ($pK_a = 9,8$)⁴⁷ e HIS162 (N δ 1 protonado, com $pK_{a1} = 5,35$ e $pK_{a2} = 9,13$)^{47,192} foram mantidos neutros.

3.2.3 Construção dos sistemas apo e holo em diferentes pH

Toda a construção e cálculos de dinâmica do sistema com a apo cruzaina em pH 5,5 e 8,0 foram feitos no programa GROMACS 5.1.5.¹⁹³ Na configuração apo, o ligante K777 e as águas cristalográficas foram excluídos, e na configuração holo, o inibidor foi extraído para construção da topologia. O campo de força usado nos quatro sistemas foi o OPLS-AA^{107,194} e a topologia do ligante co-cristalizado (carga +1) foi feita no servidor web LigParGen,¹⁹⁵ utilizando modelo de carga 1.14*CM1A.¹⁹⁶ Todos os estados de protonação dos aminoácidos em pH 5,5 e pH 8,0 foram modelados nas configurações apo e holo pelo GROMACS.¹⁹³ Posteriormente, a proteína foi alocada no centro da caixa cúbica contendo modelos de água TIP3P.¹⁹⁷ No pH 5,5, as configurações apo e holo foram neutralizadas com adição de 11 e 10 átomos de Na⁺ respectivamente, e no pH 8,0, esses sistemas foram neutralizados com a adição de 13 e 12 átomos de Na⁺, respectivamente. Em seguida, os quatro sistemas foram minimizados na janela de 0,01 kcal.mol⁻¹, e submetidos ao equilíbrio de fase canônico (NVT constantes) e isotérmico-isobárico (NPT constantes), ambos em 2 ns e com passos de 2 fs, seguindo o tempo de equilíbrio elaborado por Hoelz e colaboradores.¹¹² No equilíbrio de fase NVT e NPT foram mantidas as condições periódicas de contorno e a temperatura de 300 K constante acoplados ao termostato de Nosé-Hoover¹⁹⁶⁻¹⁹⁷ e pressão isotrópica foram aplicados usando o barostato de Parrinello-Rahman.²⁰⁰ Para a configuração holo, a coordenada do ligante foi restringida nos diferentes meios. As interações eletrostáticas foram submetidas ao método de malhas de partículas de Ewald (PME, do inglês particle mesh of Ewald)¹⁹⁹⁻²⁰⁰. Nesse método, as distâncias de corte foram de 1 nm. Após os equilíbrios de fase, em todos os sistemas foram calculadas as trajetórias por 300 ns com passos de 2 fs, sem restrições de posição. Por fim, foram analisados o desvio quadrático médio (referência 0 ns), o grau de compactação da enzima, as ocupâncias de ligação de hidrogênio na enzima (considerando distância de 3 Å entre o heteroátomo e o hidrogênio) durante toda a trajetória no programa VMD.²⁰³

3.2.4 Cálculo da variação da energia livre de Gibbs

O cálculo da energia livre de ligação, ΔG , foi baseado no algoritmo de Kumari e colaboradores para MM-PBSA (do inglês molecular mechanics/Poisson-Boltzmann surface area).²⁰⁴ Esse método aproxima a contribuição de solvatação a um modelo de solvente contínuo, analisando separadamente os termos entálpicos e entrópicos.²⁰⁵ Assim, o método MM-PBSA aproxima ΔG a um ciclo termodinâmico modelando um solvente contínuo no modelo de Poisson-Boltzmann, e projetando em uma função linear contendo as informações da área de superfície de interação com o solvente.²⁰⁵ As cargas, raio e concentração de íons positivos foram de +1, 0,95 Å e a 0,15 M, respectivamente. As cargas, raio e concentração de íons negativos foram de -1, 1,81 Å e a 0,15 M, respectivamente. A constante dielétrica do soluto, do solvente e no vácuo foi de 32, 80 e 1, respectivamente. O raio da sonda do solvente foi de 1,4 Å. O método para mapear as cargas na grade foi o sp14 e o modelo usado para construir os limites dielétricos e iônicos foi o smol. No cálculo do solvente não polar, foi considerada a tensão superficial, raio da sonda e corte de energia de 0,023 kJ.mol⁻¹.Å⁻², 1,4 Å, 3,85 kJ.mol⁻¹, respectivamente. Por fim, foi considerado o termo de contribuição atrativa não-polar determinado pelo raio da sonda, pela densidade de solvente na grade e espaçamento da grade cúbica para o cálculo das integrais de volume de 1,20 Å, 0,03 Å³ e 0,7 Å, respectivamente. O intervalo de tempo analisado para os cálculos de ΔG foi de 1 ns.

3.2.5 Modelos de redes neurais baseados em dinâmica molecular

Para verificar as contribuições positivas e negativas dos deslocamentos das estruturas secundárias nos pH 5,5 e 8,0 na holo cruzaína, foram aplicadas as redes neurais artificiais. Para isso, foram definidos os aminoácidos que compõem cada estrutura secundária – sendo referência, coordenadas de hélices- α , folhas- β e alças no tempo zero (fig. 19). Assim, as hélices 1, 2, 3, 4, 5 e 6 são rotuladas de 25 a 40, 50 a 54, 62 a 64, 69 a 78, 121 a 132 e 202 a 204, respectivamente. As folhas- β 1, 2, 3, 4, 5, 6, 7, 8, 9 e 10 foram definidas de 4 a 6, 82 a 84, 108 a 111, 114 a 117, 137 a 139, 162 a 166, 168 a 170, 177 a 180, 194 a 197 e 210 a 213. Por fim, as alças foram rotuladas de 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16 e 17 nos

intervalos de 1 a 3, 7 a 24, 40 a 50, 57 a 61, 65 a 67, 78 a 81, 85 a 107, 112 a 113, 118 a 123, 133 a 134, 140 a 161, somente o resíduo 167, 171 a 176, 181 a 193, 198 a 201, 205 a 209 e 214 a 215. Os resíduos CYS25, HIS162 e ASN182 (triáde catalítica) estão localizadas na hélice 1, folha 6 e alça 14, respectivamente (fig. 19).

Figura 19 - Definição das estruturas secundárias.

Fonte: Elaborada pelo autor.

Após definir as coordenadas de referência das estruturas secundárias, foram calculados os deslocamentos estruturais definidos pela equação 7:

$$d^{es}(t_j) = \frac{1}{N} \sum_{\text{átomo } i} \sqrt{[x_i(t_j) - x(t_0)]^2 + [y_i(t_j) - y(t_0)]^2 + [z_i(t_j) - z(t_0)]^2} \quad (7)$$

Onde $x_i(t_j)$, $y_i(t_j)$ e $z_i(t_j)$ são as coordenadas de uma estrutura secundária decorrido no tempo t_j (j varia de 1 a 300, com passo de 1), e $x(t_0)$, $y(t_0)$ e $z(t_0)$ são as coordenadas de referência do átomo i no tempo t_0 , d^{es} é o deslocamento médio da estrutura secundária rotulada, e N é o número de átomos totais calculados para cada estrutura secundária rotulada. Devido à quantidade de dados a serem calculados, foi aplicada programação python.

Depois de calculados todos os deslocamentos de estruturas secundárias e as energias de ligação pelo método MM-PBSA, as redes neurais artificiais de retropropagação de multicamadas foram utilizadas para fazer os estudos quantitativos entre estrutura do complexo cruzaina-K777 e atividade biológica. Cada deslocamento associado a uma determinada estrutura secundária foi considerado como variável independente e a energia livre de Gibbs como variável dependente. Nos meios ácidos e básicos, a taxa de aprendizado, momento e número de neurônios foram de 0,3, 0,2 e 13, respectivamente (fig. 20).

Figura 20- Estudos estruturais baseados em redes neurais artificiais.
 Fonte: Elaborada pelo autor.

3.3 MAPEAMENTO DO SÍTIO ATIVO DA CRUZAÍNA

3.3.1 Conjunto de dados

O conjunto de dados original contém 50 compostos, contendo predominantemente derivados oxadiazóis.²⁰⁶ A seleção dos 40 inibidores não covalentes da cruzaína foi baseada no método *leave-one-out*, utilizando o modelo CoMFA no conjunto original para verificar e retirar os *outliers* presentes. A coleção de compostos possui boa diversidade estrutural e distribuição dos dados em 3 ordens de magnitude. Estas características são desejáveis para a produção de modelos de QSAR estatisticamente significativos. Os valores de IC_{50} (concentração de composto necessária para inibir em 50% a atividade catalítica da enzima cruzaína) foram convertidos em pIC_{50} (escala logarítmica na base 10 de $1/IC_{50}$, $\log_{10}1/IC_{50}$). A representação da diversidade estrutural e a distribuição dos dados experimentais são mostradas na figura 21.

Figura 21 - A) Representação da diversidade estrutural. B) Distribuição da propriedade alvo.
Fonte: Elaborada pelo autor.

3.3.2 Construção dos modelos de QSAR 2D

3.3.2.1 QSAR 2D

A seleção adequada dos conjuntos treinamento e teste pode afetar significativamente a qualidade dos modelos de QSAR. Para identificar a melhor combinação de conjuntos, foi aplicado o método de QSAR automatizado (AutoQSAR, *automated QSAR*), disponível no módulo *Small-Molecule Drug Discovery*. Na figura 22 é mostrado o processo esquemático do método. O AutoQSAR gera automaticamente 497 descritores físico-químicos e topológicos por meio do Canvas *fingerprint*. Posteriormente, são geradas seleções randômicas para seleção das séries treinamento e teste e aplicados os métodos de aprendizado de máquina. Esses métodos incluem: RP (*ensemble recursive partitioning*), kernel-PLS (*kernel-based partial least square*), PLS (*partial least square*), MLR (*multiple linear regression*), PCR (*principal component regression*) e Bayes (naïve Bayes). Gerados os modelos de QSAR, eles são ranqueados por uma função de pontuação. A inspeção e análises dos valores preditos e estimados *versus* valores experimentais das séries treinamento e teste são realizadas pelo operador. Assim, foram inseridos no algoritmo a seleção randômica de 80% do conjunto de dados para a construção dos modelos e o restante foi utilizado como série teste.

Figura 22 - Diagrama esquemático do algoritmo do AutoQSAR.

Fonte: Elaborada pelo autor.

O máximo par de correlações foi mantido em 0,80 e o número de modelos mantidos para cada método de aprendizado de máquina foi de 99. O critério de seleção do melhor modelo foi baseado na melhor pontuação (valor que contém informações de $q^2 = r^2_{\text{pred}}$ e r^2), seguido das análises de q^2 e r^2 . Para visualizar os mapas de contribuições atômicas do melhor modelo 2D, foram reproduzidas as informações no Canvas KPLS *visualization*. Foram calculados os descritores *fingerprints* e mantidos o método de aprendizado e o número de componentes do melhor modelo gerado pelo AutoQSAR.

3.3.2.2 QSAR 3D

Preparação dos ligantes 3D

Para gerar os modelos de QSAR 3D é necessário preparar os ligantes na sua conformação tridimensional. Para isso, foi utilizado o LigPrep para converter as representações 2D em 3D. Em seguida, foi realizada uma busca conformacional em larga escala usando o MacroModel, aplicando água como solvente, janela de energia de 5 kcal/mol e campo de força OPLS3 (fig. 23). O campo de força OPLS3 foi escolhido, pois sua precisão nos parâmetros de torção é de 2 ordens de magnitude em relação aos campos de força convencionais. O objetivo da busca conformacional é encontrar nos ligantes a conformação de mínima energia global (fig. 23). Posteriormente, foram calculadas as cargas de Gasteiger-Hückel no programa SYBYL 2.0.

Figura 23 - Diagrama esquemático das estratégias de minimização dos ligantes.
 Fonte: Elaborada pelo autor.

Alinhamento molecular e método CoMFA

Para gerar as características 3D e associá-las a propriedade alvo foi aplicado o método CoMFA. Este método requer que as moléculas do conjunto de dados sejam alinhadas para a comparação dos descritores estereoquímicos (energia de Lennard-Jones) e eletrostáticos (energia de Coulomb) entre os diferentes compostos. Para isso, o método de alinhamento molecular baseado em algoritmo genético com atribuição linear (GALAHAD, *genetic algorithm with linear assignment for the hypermolecular alignment of datasets*) foi aplicado. Este método assume que a estrutura de mínima energia se liga ao sítio ativo. Basicamente, o algoritmo genético (*genetic algorithm*) gera diversas conformações variando o ângulo diedro das estruturas químicas. É um método não supervisionado, isto é, não depende da propriedade biológica. Uma vez obtidos, os confôrmeros de mínima energia são alinhados por suas características farmacofóricas. Este processo segue duas etapas: (i) o algoritmo genético opera sobre as coordenadas internas gerando uma diversidade de conformações; (ii) ocorre o alinhamento 3D seguindo as atribuições lineares para avaliar o nível de sobreposição molecular (STERIC), características farmacofóricas (HBOND) e a soma das energias das conformações (ENERGY). Na

sequência, uma hipermolécula é criada a partir das características farmacofóricas comuns. Depois de gerada a hipótese farmacofórica, esta é usada como molde para o alinhamento tridimensional das moléculas do conjunto de dados. Para gerar o modelo de QSAR 3D baseado no GALAHAD, as três estruturas mais potentes da série foram selecionadas, e para o algoritmo genético, foi estabelecido o número 100 para população e 240 para geração (fig. 24). Os demais parâmetros foram mantidos no modo padrão. Depois de realizado o alinhamento molecular do conjunto de dados a partir do GALAHAD, foi aplicado o método CoMFA. O conjunto alinhado foi submetido a uma grade virtual reticulada, de espaçamento de 2 Å, e 4 Å de extensão além das superfícies das estruturas. No cálculo dos descritores estereoquímicos (energia de Lennard-Jones) e eletrostáticos (energia de Coulomb), foi utilizado o carbono sp^3 com carga líquida +1.

Figura 24 - Método GALAHAD aplicado nas três estruturas mais potentes da série para obtenção do farmacóforo.

Fonte: Elaborada pelo autor.

3.3.3 Docagem molecular

Para os estudos de docagem molecular, foi selecionada a estrutura da cruzaina com o código PDB ID 4XUI, resolução de 2,51 Å. A escolha teve como critério a ocupação do ligante co-cristalizado nos quatro subsítios da enzima. A docagem molecular foi utilizada para gerar as poses dos ligantes que possuem melhor pontuação, com o objetivo de compreender e extrair informações relevantes dos modelos de QSAR 2D e 3D. As proteínas foram preparadas usando o módulo *Protein Preparation Wizard*. As altas ocupâncias foram selecionadas e o campo de força OPLS3 foi aplicado para minimizar a estrutura 3D da macromolécula. Esse campo de força possui melhor precisão torsional (2 ordens de magnitude) e apresenta melhor modelo de cargas em relação aos campos de força convencionais. A estrutura da enzima foi submetida às predições das perdas das cadeias laterais e das conformações das cadeias principais contendo as regiões de *loops* pelo protocolo do Prime. O pH foi mantido em 5,5 para simular a interação no reservossomo. Por fim, foram mantidos os demais parâmetros no modo padrão. A partir das estruturas preparadas, o centróide do ligante (14, 12, -15) foi definido como o sítio de interação para construção do GRID (cúbico, aresta de 12 Å). Posteriormente, as poses da simulação de docagem molecular foram avaliadas pela função empírica GlideScore XP (*extra precision*).

4 RESULTADOS E DISCUSSÃO

4.1 CONJUNTO 1 – ESTUDOS DE QSAR

4.1.1 Análise por componentes principais – conjuntos treinamento e teste

O mapa de similaridade estrutural (fig. 25) representa a distribuição do espaço químico e biológico. Esse mapa mostra a grande diversidade química e o histograma apresenta 5 ordens de magnitude para a propriedade biológica (pIC_{50}), variando de 4 a 9. A diversidade química e biológica observada no mapa e no histograma satisfazem as condições adequadas para o desenvolvimento dos modelos de QSAR. Baseado no mapa e na distribuição da atividade biológica no histograma, os conjuntos treinamento (280 estruturas químicas, 77%) e teste (83 compostos, 23%) foram selecionados. A partir desse conjunto, as séries treinamento e teste foram mantidas na construção dos modelos de QSAR.

Figura 25 - Similaridade estrutural dos compostos ativos na forma amastigota de *T. cruzi* e a distribuição da atividade biológica do conjunto de dados e da série treinamento e teste.

Fonte: Elaborada pelo autor.

4.1.2 Análise estatística

4.1.2.1 Descritores físico-químicos e redes neurais artificiais

Na Figura 26 é mostrada a distribuição dos descritores físico-químicos MW, *aLogP*, HBA, HBD, RB e HAC para o conjunto de dados, com suas respectivas análises estatísticas. O MW apresentou média de 402,97 e desvio 119,25, apresentando distribuição majoritária entre 283,72 e 522,22. O descritor *aLogP* mostrou média de 4,07 e desvio de 2,10; a maioria dos compostos estão concentrados entre 1,97 e 6,17. Os descritores HBA e HBD mostraram médias de aproximadamente 3 e 1 com desvio de 2 e 1, respectivamente. Isso mostra que o número de grupos doadores e aceptores estão predominantemente entre 0 e 5 para HBA e 0 e 2 para HBD, respectivamente. O descritor RB, apresentou média de 6 ligações rotacionáveis e a maioria dos compostos se concentram na faixa entre 2 e 10. HAC mostrou média de 28 átomos com desvio de aproximadamente 9, mostrando que a maioria dos compostos possuem número de átomos entre 20 e 37.

Figura 26 - Distribuição dos descritores físico-químicos do conjunto de dados.
Fonte: Elaborada pelo autor.

Na Figura 27 é apresentada a distribuição dos descritores físico-químicos RC, PSA, Estate, MR e Polar. O RC apresentou média por volta de 4 anéis e um desvio de aproximadamente 1,5, com a maioria dos compostos apresentando de 2 a 5 anéis. O PSA tem média de 73,11 e um desvio de 38,63, podendo a maioria estar com área de superfície polar entre 34,48 e 111,74. A maioria dos compostos anti-mastigota também apresentou Estate entre 44,66 e 86,34, com média de 65,5. Por

fim, os descritores MR e Polar apresentaram média de 109,64 e 49,56, estando a maioria dos compostos entre 75,97 e 143,31, e 33,56 e 65,56, respectivamente.

Figura 27 - Distribuição dos descritores físico-químicos do conjunto de dados.

Fonte: Elaborada pelo autor.

Todos os descritores físico-químicos foram usados para treinar as redes neurais artificiais. O critério de melhor modelo é dado pelo valor da função de pontuação, de acordo com a referência.¹⁶⁶ A Tabela 1 mostra o desempenho do

modelo em função da taxa de aprendizado (0,1 a 0,4). O melhor resultado apresentado pelas redes neurais apresentou pontuação de 0,80, seguido dos outros parâmetros estatísticos, como q^2 e r^2 com valores de 0,85 e 0,79, respectivamente.

Tabela 1 - Desempenho dos modelos de redes neurais artificiais em função da taxa de aprendizado.

Conjunto treinamento						Conjunto teste					
LR	r^2	MEA	RMSE	EAR	RQER	q^2	MEA	RMSE	EAR	RQER	Pontuação
0,1	0,79	0,65	0,82	65	68	0,85	0,60	0,75	59	61	0,80
0,2	0,80	0,58	0,76	59	64	0,78	0,66	0,84	65	68	0,76
0,3	0,79	0,60	0,77	60	65	0,78	0,67	0,89	66	72	0,77
0,4	0,80	0,58	0,77	59	64	0,77	0,69	0,94	68	76	0,75

LR = taxa de aprendizado; $q^2 = r^2$ para o conjunto teste (r^2_{pred}); r^2 = coeficiente de correlação da série treinamento; RMSE = erro quadrático médio de predição; SD = desvio padrão; AME= erro médio absoluto; RAE= erro absoluto relativo; RQER= erro relativo quadrático médio de predição; Pontuação = $(1-|(r^2-q^2)|)*q^2$.

Fonte: Elaborada pelo autor.

Em seguida, foi avaliado o parâmetro momento para verificar as melhores condições de convergência para as redes neurais artificiais, com variação de 0,1 a 0,4 e com passo de 0,1. A tabela 2 mostra que não há diferenças nos valores de pontuação para os modelos gerados. O critério de seleção do melhor modelo foi o parâmetro RMSE. Assim, a rede mostrou melhor convergência e desempenho com valor de momento de 0,2, e q^2 e r^2 de 0,85 e 0,79, respectivamente.

Tabela 2 - Desempenho dos modelos de redes neurais artificiais em função da taxa de convergência.

Conjunto treinamento						Conjunto teste					
MR	r^2	MEA	RMSE	EAR	RQER	q^2	MEA	RMSE	EAR	RQER	Pontuação
0,1	0,78	0,64	0,81	64	68	0,84	0,58	0,73	57	59	0,79
0,2	0,79	0,65	0,82	65	68	0,85	0,60	0,75	59	61	0,80
0,3	0,79	0,66	0,83	66	70	0,85	0,62	0,77	61	63	0,80
0,4	0,79	0,67	0,84	67	71	0,85	0,62	0,78	62	63	0,80

MR = taxa de momento; $q^2 = r^2$ para o conjunto teste (r^2_{pred}); r^2 = coeficiente de correlação da série treinamento; RMSE = erro quadrático médio de predição; SD = desvio padrão; AME= erro médio absoluto; RAE= erro absoluto relativo; RQER= erro relativo quadrático médio de predição.

Fonte: Elaborada pelo autor.

Finalmente, foi realizada a variação do número de neurônios na camada escondida de 1 a 10, adicionando 1 neurônio para gerar os modelos. Os parâmetros da taxa de aprendizado e momento, obtidos no melhoramento do modelo, foram mantidos constantes e iguais a 0,1 e 0,2, respectivamente. A Tabela 3 mostra que o modelo contendo 7 neurônios na camada escondida apresentaram os melhores resultados de pontuação, q^2 e r^2 de 0,81, 0,81 e 0,81, respectivamente.

Tabela 3 - Desempenho dos modelos de redes neurais artificiais em função do número de neurônios na camada escondida.

Conjunto treinamento							Conjunto teste				
MR	r^2	MEA	RMSE	EAR	RQER	q^2	MEA	RMSE	EAR	RQER	Pontuação
1	0,51	0,86	1,03	87	68	x	x	x	x	x	x
2	0,72	0,69	0,85	69	72	x	x	x	x	x	x
3	0,75	0,69	0,86	69	72	x	x	x	x	x	x
4	0,77	0,68	0,84	68	70	x	x	x	x	x	x
5	0,78	0,64	0,81	65	68	0,80	0,64	0,82	63	66	0,78
6 [#]	0,79	0,65	0,82	65	68	0,85	0,60	0,75	59	61	0,80
7	0,81	0,57	0,73	56	62	0,81	0,59	0,76	58	62	0,81
8	0,79	0,68	0,85	68	71	0,77	0,69	0,90	68	73	0,76
9	0,80	0,65	0,82	65	68	0,82	0,64	0,81	63	66	0,80
10	0,82	0,58	0,75	58	63	0,79	0,62	0,83	61	67	0,77

[#]modo padrão; NN = número de neurônios na camada escondida; $q^2 = r^2$ para o conjunto teste (r^2_{pred}); r^2 = coeficiente de correlação da série treinamento; RMSE = erro quadrático médio de predição; SD = desvio padrão; AME= erro médio absoluto; RAE= erro absoluto relativo; RQER= erro relativo quadrático médio de predição.

Fonte: Elaborada pelo autor

A partir do modelo 7, da tabela 3, verificou-se a relação entre os valores preditos e experimentais (fig. 28). Este gráfico mostrou uma boa relação dos dados experimentais e previstos para as séries treinamento e teste. Portanto, as RNAs de retro-propagação podem ser usadas para prever a atividade biológica de novos compostos anti-*T. cruzi*.

Figura 28 - Gráfico dos valores preditos e estimados pelas redes neurais e experimentais dos valores de pIC_{50} .

Fonte: Elaborada pelo autor.

4.1.2.2 Descritores de impressão digital molecular e KPLS

Os estudos estatísticos foram realizados a partir do método KPLS, avaliando os melhores descritores radial, linear, *molprint2D* e *dendritic*. Os resultados dos melhores modelos são apresentados na tabela 4. O critério de melhor modelo foi baseado no melhor valor da função de pontuação. As pontuações do KPLS aplicado aos descritores *dendritic*, linear, radial e *molprint2D* foram de 0,76, 0,78, 0,80 e 0,82, respectivamente. O melhor modelo foi o KPLS *molprint2D*, o qual apresentou pontuação, q^2 , r^2 e N de 0,82, 0,84, 0,81 e 3, respectivamente.

Tabela 4 - Desempenho dos modelos KPLS em função dos descritores estruturais *dendritic*, linear, radial e *molprint2D*.

FD	Pontuação	SD	q^2	RMSE	r^2	N
Dendritic	0,76	0,53	0,82	0,40	0,89	3
Linear	0,78	0,51	0,83	0,41	0,89	3
Radial	0,80	0,54	0,81	0,54	0,80	2
Molprint2D	0,82	0,50	0,84	0,52	0,81	3

FD = descriptor de impressão digital molecular; $q^2 = r^2$ para o conjunto teste (r^2_{pred}); r^2 = coeficiente de correlação do conjunto treinamento; RMSE = erro quadratic médio de predição do conjunto teste; SD = desvio padrão; pontuação = $(1 - |(r^2 - q^2)|) * q^2$.

Fonte: Elaborada pelo autor.

Posteriormente, o melhor modelo (KPLS *molprint2D*) foi utilizado para verificar a correlação do conjunto treinamento e teste, através do gráfico dos valores preditos e estimados em função dos valores experimentais (fig. 29). Assim, o gráfico mostrou boa correlação dos dados experimentais com os valores preditos, conforme representados pelos parâmetros estatísticos q^2 e r^2 , 0,84 e 0,81, respectivamente. Portanto, o modelo de redes neurais artificiais pode ser utilizado para prever a atividade biológica de novos compostos na triagem virtual, diante da grande diversidade estrutural e variedade na propriedade biológica.

Figura 29 - Gráfico dos valores preditos e estimados, pelo modelo KPLS molprint2D, contra os dados experimentais de pIC_{50} .

Fonte: Elaborada pelo autor.

4.1.3 Relações das propriedades moleculares/estruturais e a atividade biológica

4.1.3.1 QSAR baseado em redes neurais artificiais

As entradas aplicadas nas redes neurais artificiais são responsáveis pelo ajuste dos pesos em cada neurônio. Cada peso pode ter valor positivo ou negativo. Um valor positivo associado a um descritor físico-químico pode representar um aumento proporcional na atividade biológica, *i.e.*, diminuir o valor numérico desse descritor implica diminuir a atividade biológica, e aumentar o valor numérico implica aumentar o valor da atividade biológica julgado por cada neurônio. Diferentemente do peso positivo, o valor negativo guia um neurônio para contribuir de forma inversamente proporcional, ou seja, um eventual aumento da atividade biológica é provocado pela diminuição de uma propriedade físico-química, e uma diminuição da atividade biológica é causada pelo aumento do valor da entrada julgado pelo neurônio. A Tabela 5 mostra os pesos representativos dos descritores moleculares para os sete neurônios contidos na camada oculta.

Tabela 5 - Pesos dos descritores moleculares associados aos 7 neurônios na camada escondida.

Neurônio	MW	aLogP	HBA	HBD	RB	HAC	RC	PSA	Estate	MR	Polar
1	0,5	1,9	0,9	3,3	0,7	0,5	-0,3	-0,3	-0,3	1,0	0,6
2	2,8	-4,0	-2,6	-0,1	-0,2	3,0	-2,2	-2,8	1,0	2,4	3,7
3	-1,4	1,8	0,5	2,1	-0,3	-0,2	2,4	1,6	-0,3	-1,1	-0,6
4	0,6	0,1	0,1	1,1	0,1	0,3	0,4	0,3	0,3	0,1	0,8
5	1,4	-4,9	-6,7	-0,3	-0,8	3,3	4,6	1,1	-0,2	4,9	5,5
6	0,6	2,8	0,9	3,2	-2,0	-0,6	0,7	-5,2	-3,3	1,5	-0,8
7	1,6	3,8	5,4	2,4	-2,7	-2,2	-2,5	-0,1	0,9	-1,1	-2,8

Descritores: peso molecular (MW), efeito lipofílico (aLogP), número de doadores (HBD) e aceptores (HBA) de ligação de hidrogênio, número de ligações rotacionáveis (RB), área de superfície polar (PSA), estado eletro-topológico (Estate), refratibilidade molar (MR), polarização (polar), número de átomos pesados (HAC) e número de anéis (RC).

Fonte: Elaborada pelo autor.

Os seguintes resultados foram obtidos para as redes neurais: para o peso molecular (MW), 6 neurônios mostraram valores positivos e apenas 1 neurônio teve valor negativo; para o efeito lipofílico (aLogP), 5 neurônios apresentaram pesos positivos e 2 neurônios pesos negativos; para o número de doadores de ligação de hidrogênio (HBD) 5 neurônios apresentaram pesos positivos e 2 apresentaram pesos negativos. O número de aceptores de ligação de hidrogênio (HBA) foi o mesmo do HBD, e o número de ligações rotacionáveis (RB) apresentou 2 neurônios positivos e 5 negativos. O descritor HAC teve 4 neurônios positivos e 3 neurônios negativos, e o RC mostrou 4 neurônios com valores positivos e 3 neurônios negativos. Para o PSA e o Estate, foram obtidos 3 neurônios positivos e 4 negativos; a refratibilidade molar (MR) teve 5 neurônios positivos e 2 negativos; e finalmente, no descritor Polar, 4 neurônios tiveram pesos positivos e 3 neurônios mostraram pesos negativos.

Baseados na análise dos pesos das redes neurais foram verificados os votos majoritários associados aos pesos dos neurônios. A maioria dos pesos positivos estão relacionados aos descritores MW, aLogP, HBD, HBA, HAC, RC, MR e Polar. Os descritores com pesos negativos foram os descritores RB, PSA e Estate. Todos os descritores que possuem valores de pesos majoritariamente positivos estão relacionados às propriedades farmacocinéticas adequadas para um fármaco atravessar a membrana biológica e se ligar ao seu alvo molecular. Podemos destacar os descritores MW (85,7% dos neurônios), aLogP (71,4% dos neurônios), HBA (71,4% dos neurônios) e HBD (71,4% dos neurônios). Apenas o descritor RB apresentou uma grande maioria dos votos com pesos negativos, sendo 5 neurônios

com pesos negativos (71,4% dos neurônios). Os descritores HAC, RC, PSA, Estate e Polar mostraram que o voto da maioria diferiu de apenas 1 neurônio. No entanto, pode-se associar os descritores HAC (4 pesos positivos, 3 pesos negativos) e RC (4 pesos positivos, 3 pesos negativos) ao MW e $a\text{Log}P$. O descritor PSA (4 pesos negativos e 3 pesos positivos) e o Estate (4 pesos negativos e 3 pesos positivos), no entanto, apresentaram uma relação antagônica aos descritores MW e $a\text{Log}P$ e estão associados com os descritores HBA, HBD e Polar. Por fim, o descritor Polar (4 pesos positivos e 3 negativos) está associado aos efeitos de polarização causados pelos átomos eletronegativos na molécula e, portanto, pode favorecer a atividade contra a forma amastigota de *T. cruzi*.

4.1.3.2 QSAR baseado em KPLS molprint2D

Os estudos de QSAR foram feitos, através dos mapas de contribuição atômica do modelo KPLS molprint2D. O modelo mostra a contribuição favorável ou desfavorável de um átomo e/ou fragmento em relação à atividade biológica (pIC_{50}). O mapa é representado por um código de cores, sendo as contribuições positivas dadas pela cor vermelha, neutras pela cor branca e negativas pela cor azul. Este mapa é gerado através do modelo, combinando as informações da função matemática com a estrutura química. Todas as análises foram baseadas nas projeções do modelo para o conjunto de treinamento.

De acordo com o modelo, grupos de anéis aromáticos heterocíclicos apresentam contribuições positivas. Os hidrocarbonetos alifáticos podem ser favoráveis ou desfavoráveis para a atividade. Os compostos contendo átomos de halogênio apresentam condições favoráveis, neutras e desfavoráveis para o flúor, cloro e iodo, e o bromo, respectivamente. Os grupos ciclohexano e ciclopropano podem apresentar contribuições positivas ou negativas para a atividade biológica. Os grupos hidroxila e metoxi apresentam contribuição negativa nos ensaios fenotípicos. O anel piperazina pode apresentar contribuição favorável ou desfavorável, sendo dependente da estrutura de cada molécula testada nos ensaios. Em geral, os anéis fundidos podem ser desfavoráveis para a atividade biológica. Na figura 30, são apresentados os mapas de contribuição dos compostos mais ativos e menos ativos contra a forma amastigota de *T. cruzi*.

Figura 30 - Mapas de contribuição dos compostos mais ativos e menos ativos na forma amastigota de *T. cruzi*.

Fonte: Elaborada pelo autor.

4.1.3.3 Estudos de bioisosterismo baseado em fragmentos moleculares

As análises do mapa de contribuição atômica foram realizadas através de 29 compostos selecionados na série treinamento (244, 302, 303, 304, 308, 309, 311, 313, 319, 320, 329, 332, 336, 337, 338, 339, 342, 344, 348, 350, 351, 352, 353, 355, 356, 357, 359, 361, 362), os quais apresentaram valores de pIC_{50} maiores que 6. Assim, foram construídos os fragmentos moleculares que apresentaram contribuição positiva para a atividade biológica. Em seguida, os descritores físico-químicos foram calculados e aplicados às redes neurais artificiais para prever a atividade biológica. Com as informações das atividades biológicas experimentais e previstas pelas redes neurais, um mapa de calor foi construído para analisar o perfil físico-químico dos fragmentos moleculares, através das curvas de níveis. Dessa maneira, foram explorados os intervalos físico-químicos importantes desses fragmentos através de métodos de aprendizado de máquina.

Na figura 31 são mostrados os mapas de calor dos descritores físico-químicos, MW, $aLogP$, HBD e HBA. A partir do gradiente apresentado nesse mapa, é possível estimar as características relevantes dos fragmentos moleculares e sua

relação com a propriedade biológica. No mapa é mostrado que os fragmentos próximos a 360 Da são os mais ativos. Para o $a\text{Log}P$, os fragmentos mais ativos se concentram majoritariamente entre 1 e 5. Para os descritores HBA e HBD, os fragmentos que apresentam um grupo doador de ligação de hidrogênio e o número de grupos aceptores de ligação de hidrogênio entre 3 e 6 apresentaram os melhores valores de pIC_{50} .

Figura 31 - Mapas de calor dos aspectos físico-químicos MW, $a\text{Log}P$, HBD e HBA associados aos fragmentos moleculares.

Fonte: Elaborada pelo autor.

Na figura 32 são apresentados quatro mapas de calor representando os descritores HAC, RB, RC e PSA dos fragmentos moleculares preditos. Os fragmentos com número de átomos pesados (HAC) maior do que 15 representam os melhores resultados de atividade preditos. Os fragmentos com número de ligações rotacionáveis (RB) maior do que 4 foram mais ativos. De acordo com as redes neurais, o aumento do número de anéis (RC) também melhora o perfil dos

compostos anti-*T. cruzi*, onde fragmentos com número de anéis maior do que 3 foram, em geral, os mais ativos. Por fim, a maioria dos fragmentos com valores de área de superfície polar (PSA) entre 50 e 100 mostrou-se mais ativa.

Figura 32 - Mapas de calor dos aspectos físico-químicos HAC, RB, RC e PSA associados aos fragmentos moleculares.

Fonte: Elaborada pelo autor.

A figura 33 mostra três mapas de calor dos descritores Estate, MR e Polar. Os estados eletro-topológicos (Estate), refratibilidade molar (MR) e polarização (Polar) entre 35 e 63, 85 e 115, e 36 e 53, respectivamente, foram os melhores perfis identificados para esses fragmentos.

Figura 33 - Mapas de calor dos aspectos físico-químicos Estate, MR e Polar associados aos fragmentos moleculares.

Fonte: Elaborada pelo autor.

Os fragmentos que apresentaram os critérios descritos no mapa de calor são apresentados na figura 34, com os respectivos valores de atividade biológica predita. Os arcabouços são diversos, contendo dentre outros, derivados de piridina, pirimidina, piperazina e triazóis.

Figura 34 - Estrutura dos principais arcabouços identificados no modelo KPLS, molprint 2D e predição de pIC_{50} pelas redes neurais artificiais.

Fonte: Elaborada pelo autor.

Considerando as análises do modelo KPLS molprint2D, foram propostas algumas substituições bioisostéricas para alguns grupos funcionais (fig. 35). O grupo sulfona é bioisóstero do grupo amídico e apresenta boa atividade predita contra o parasita em comparação ao grupo amida (pIC_{50} de 5,97 e 5,47, respectivamente). Porém, essa substituição aumenta em aproximadamente quatro vezes o efeito lipofílico ($aLogP$ de 0,38 e 1,57, respectivamente). O mesmo fenômeno ocorre com a substituição do grupo triazol (pIC_{50} = 5,77) pela piperazina (pIC_{50} = 5,47). Essa substituição bioisostérica melhora o perfil lipofílico em aproximadamente duas ordens de magnitude ($aLogP$ de -0,25 e 1,57, respectivamente). Além disso, a partir do modelo 2D, foram obtidas possíveis substituições de anéis, as quais mostram aspectos físico-químicos fundamentais para a atividade biológica. Dentre esses

anéis, podemos destacar a piperazina, cicloexano, triazol, pirimidina, piridina e oxadiazol.

Bioisosterismo

Bioisosterismo entre anéis

Figura 35 - Substituições bioisostéricas de grupos químicos e sistemas de anéis.

Fonte: Elaborada pelo autor.

4.2 CONJUNTO 2 – ESTUDOS DE QSAR

4.2.1 Análise por componentes principais – conjuntos treinamento e teste

Na figura 36, é mostrada a distribuição do espaço químico e biológico pelo mapa de similaridade estrutural. O mapa apresenta uma grande diversidade de compostos biologicamente ativos e apresenta boa distribuição dos dados, com o valor de pSI variando de -4 a 5. O conjunto apresentado mostra-se adequado para o desenvolvimento de modelos de QSAR. Assim, considerando esse mapa, foram selecionados os conjuntos treinamento (183 compostos, 79%) e teste (48 compostos, 21%) para construção dos modelos e validação externa, respectivamente.

Figura 36 - Similaridade estrutural dos compostos ativos na forma amastigota de *T. cruzi* e a distribuição da atividade biológica do conjunto de dados e da série treinamento e teste.
Fonte: Elaborada pelo autor.

4.2.2 Análise estatística

4.2.2.1 Descritores físico-químicos e floresta randômica

Na figura 37 é mostrada a distribuição dos descritores físico-químicos MW, *aLogP*, HBA, HBD, RB e HAC para o conjunto de dados, com suas respectivas análises estatísticas. O MW tem média de 416,12 Da e desvio 119,49, apresentando distribuição majoritária entre 296,63 e 535,61. O descritor *aLogP* apresentou média

de 4,18 e desvio de 2,18, e a maioria dos compostos estão concentrados entre 2 e 6,36. Os descritores HBA e HBD mostraram médias de aproximadamente 3 e 1 com desvio de 2 e 1, respectivamente. Isso mostra que o perfil dos grupos doadores e aceptores de ligação de hidrogênio são predominantes entre 0 e 2 para HBA e 0 e 2 para HBD, respectivamente. O descritor RB, apresentou média de 5 ligações rotacionáveis e a maioria dos compostos se concentram na faixa entre 2 e 8. O HAC mostrou média de 29 átomos com desvio de aproximadamente 8, mostrando que a maioria dos compostos possuem número de átomos entre aproximadamente 21 e 38.

Figura 37 - Distribuição dos descritores físico-químicos (MW, aLogP, HBA, HBD, RB e HAC) do conjunto de dados.

Fonte: Elaborada pelo autor.

A figura 38 mostra a distribuição dos descritores físico-químicos RC, PSA, Estate, MR e Polar. O RC apresentou média por volta de 4 anéis e um desvio de aproximadamente 1, na qual a maioria dos compostos possui entre 3 e 5 anéis. O PSA médio possui valor de 69,74 e um desvio de 33,47, variando entre 36,27 e 103,21. O descritor Estate possui média de 65,51 e possui extremos com relação à média de 46,78 e 84,24. Por fim, os descritores MR e Polar apresentaram média de

115,08 e 52,71, e os compostos encontram-se, em sua maioria, entre 80,35 e 149,81, e 36,16 e 69,26, respectivamente.

Figura 38 - Distribuição dos descritores físico-químicos (RC, PSA, Estate, MR e Polar) do conjunto de dados.

Fonte: Elaborada pelo autor.

Os onze descritores foram usados como entrada para treinar a floresta randômica e posteriormente, foi realizada a validação externa pelo cálculo do

coeficiente da correlação do conjunto teste (q^2). Assim, a correlação entre os valores preditos e experimentais para a série treinamento (r^2) e teste (q^2) de pSI e seus respectivos erros quadráticos médios foram de $0,97 \pm 0,28$ e $0,77 \pm 0,85$, respectivamente (fig. 39).

Figura 39 - Gráfico dos valores preditos e estimados pela floresta randômica e experimentais dos valores de pIC_{50} .

Fonte: Elaborada pelo autor.

4.2.2.2 Regressão por mínimos quadrados de pSI versus pIC_{50}

Com base nos dados experimentais, foi determinada a correlação entre a seletividade (pSI) e a atividade biológica (pIC_{50}). O gráfico apresentado na figura 40 mostra a boa correlação dos dados, com r^2 e r de Person igual a 0,79 e 0,88, respectivamente. Através da equação da reta, obtida pelo método dos mínimos quadrados, foi observada a relação constante entre os dois parâmetros biológicos (eq. 8).

$$pSI = (0,92 \pm 0,03).pIC_{50} + (-3,99 \pm 0,18) \quad (8)$$

Essa correlação mostra que a seletividade não depende do tipo de célula e nem do tempo decorrido de exposição, e sim das características físico-químicas dos compostos contra a forma amastigota de *T. cruzi*. Isso é demonstrado, quando é

analisada a curva de ajuste experimental com a equação 3. Além disso, é possível estimar o valor de pSI, a partir dos valores de pIC_{50} obtidos de dados experimentais e/ou preditos. A partir do coeficiente linear, é possível identificar o parâmetro envolvido com os efeitos citotóxicos do fármaco na célula, a constante citotóxica α . O valor de α está diretamente relacionado com a concentração necessária para inibir em 50% o crescimento das células viáveis. Assim, a partir da regressão linear obtida do gráfico (fig. 40), foi determinado $pIC_{50}^{crítico} = -3,99$ ou $IC_{50}^{crítico} = 10^4$ M. Para valores de $pIC_{50}^{amastigota}$ maiores que 3,99, o fármaco apresenta seletividade positiva, ou seja, é mais seletivo para o parasita do que para a célula infectada. No entanto, valores de $pIC_{50}^{amastigota}$ menores que 3,99, o composto apresenta seletividade negativa, ou seja, o fármaco mais seletivo para a célula infectada do que para o parasita. Por fim, $pIC_{50}^{amastigota}$ tendendo a 3,99, o limite na qual ocorre a eliminação de parasitas amastigotas e também da célula, numa proporção de 50% cada.

Figura 40 - Ajuste experimental da seletividade e atividade biológica.
Fonte: Elaborada pelo autor.

4.2.2.3 Descritores de impressão digital molecular e KPLS

Nessa sessão, são apresentados os resultados dos melhores modelos na tabela 6. O critério de melhor modelo foi baseado no melhor valor da função de pontuação. As pontuações do KPLS aplicadas aos descritores *dendritic*, linear, radial

e molprint2D foram de 0,78, 0,79, 0,84 e 0,82, respectivamente. O melhor modelo foi o KPLS radial, o qual apresentou pontuação, q^2 , r^2 e N de 0,84, 0,86, 0,84 e 3, respectivamente.

Tabela 6 - Desempenho dos modelos KPLS em função dos descritores estruturais *dendritic*, linear, radial e molprint2D.

FD	Pontuação	RMSE	q^2	SD	r^2	N
Dendritic	0,78	0,57	0,82	0,51	0,87	3
Linear	0,79	0,61	0,79	0,64	0,79	2
Radial	0,84	0,50	0,86	0,55	0,84	3
Molprint2D	0,82	0,55	0,83	0,60	0,83	3

FD = descriptor de impressão digital molecular; $q^2 = r^2$ para o conjunto teste (r^2_{pred}); r^2 = coeficiente de correlação do conjunto treinamento; RMSE = erro quadratic médio de predição do conjunto teste; SD = desvio padrão; Pontuação = $(1 - |(r^2 - q^2)|) * q^2$.

Fonte: Elaborada pelo autor.

Posteriormente, o melhor modelo (KPLS radial) foi utilizado para verificar a correlação do conjunto treinamento e teste, através do gráfico dos valores preditos e estimados em função dos valores experimentais (fig. 41). O gráfico apresentou boa correlação dos dados experimentais com os valores preditos. Portanto, o modelo KPLS radial pode ser utilizado para prever a atividade biológica de novas estruturas, dentro da diversidade estrutural e faixa de valor da propriedade biológica estudadas.

Figura 41 - Gráfico dos valores preditos e estimados, pelo modelo KPLS radial, contra os dados experimentais de pSI.

Fonte: Elaborada pelo autor.

4.2.3 Relações das propriedades moleculares/estruturais e a atividade biológica

4.2.3.1 QSAR baseado em floresta randômica

As entradas aplicadas na floresta randômica foram utilizadas na geração de critérios para guiar o desenvolvimento de um fármaco de alta ou baixa seletividade. A floresta é composta por 100 árvores aleatórias e de diferentes tamanhos. O peso dos descritores físico-químicos baseado na média da diminuição da impureza mostrou que MW, aLogP, HBA, HBD, RB, HAC, RC, PSA, Estate, MR e Polar, foram de 2,29, 2,72, 2,64, 1,79, 4,73, 4,37, 7,68, 3,20, 4,52, 6,35 e 4,58 respectivamente. O número de anéis, refratibilidade molar, o estado eletro-topológico, o número de átomos pesados, polarização e área de superfície polar possuem pesos médios significativos em relação aos demais descritores, mostrando que eles são fundamentais para definir a seletividade. A massa, o número de doadores de ligação de hidrogênio e o efeito lipofílico, no entanto, demonstraram possuir pouca contribuição para a seletividade em relação aos demais descritores físico-químicos.

Os nós das árvores são representados por MW, aLogP, HBA, HAC, PSA, Estate e MR com valores de 1749, 1634, 657, 710, 943, 860, 851, respectivamente. Isso mostra que para definir um critério de classificação, a floresta randômica precisa utilizar esses descritores diversas vezes em relação os outros aspectos físico-químicos. No entanto, os descritores HBD, Polar, RB e RC apresentaram média nos nós de 474, 645, 769 e 273, mostrando que para classificar a seletividade de um composto, será necessário utilizar poucas informações sobre esses descritores, pois apresentam melhor descrição da propriedade molecular.

Foi realizada a análise da árvore, e definido os critérios para pSI maior que 2,2 e pIC₅₀ maior 6. Por exemplo, para obter um composto com pSI = 2,21, foi analisado o critério de classificação de uma das árvores de decisão aleatória e obtido: aLogP entre 1,90 e 2,11; RC menor 3,5; e MW menor que 305,35. Assim, essa análise foi realizada para todas as 100 árvores de decisão, a qual manteve 11 critérios dos descritores físico-químicos para que os compostos fossem classificados com relação a seletividade pela série treinamento: (1) $279 < MW < 485$; (2) $0,9 < alogP < 5,6$; (3) $1 \leq HBA \leq 6$; (4) HBD 0 ou 1; (5) $4 \leq RB \leq 8$; (6) $21 \leq HAC \leq 36$; (7) 3

$\leq RC \leq 5$; (8) $19 < PSA < 145$; (9) $45 < Estate < 91$; (10) $85 < MR < 132$; e (11) $37 < Polar < 64$.

4.2.3.2 QSAR baseado em KPLS radial

Os mapas de contribuição atômica foram gerados a partir do modelo KPLS radial. O modelo 2D mostra diversidade de contribuições positivas e negativas de grupos químicos para a atividade biológica. Na maioria dos compostos, os grupos nitrilas, nitrotriazóis, triazóis, piridinas, pirimidinas, quinazolininas e aminas primárias são grupos que apresentam, na maioria dos compostos, contribuição positiva para a seletividade. Os átomos de enxofre foram majoritariamente positivos ou neutros, e os halogênios cloro e flúor foram, em grande parte, positivos para a seletividade. No entanto, os átomos de bromo e iodo podem ser neutros ou negativos para a atividade biológica. Os grupos amina e sulfona apresentaram contribuição positiva ou neutra para a seletividade. Na figura 42 é mostrado o mapa de contribuição atômica dos compostos menos seletivos até os compostos mais seletivos testados em amastigotas.

Figura 42 - Mapas de contribuição dos compostos seletivos entre célula saudável e a forma amastigota de *Trypanosoma cruzi*.

Fonte: Elaborada pelo autor.

4.3 TRIAGEM VIRTUAL EM LARGA ESCALA

Os resultados apresentados nessa seção são baseados nos dados finais da triagem virtual, mostrando as distribuições do espaço químico e biológico, das propriedades biológicas preditas, tais como energia livre de Gibbs, pIC_{50} e pSI preditos na forma amastigota. Além disso, será abordada a diversidade estrutural dos novos arcabouços e a distribuição dos parâmetros preditos para absorção, distribuição, metabolismo, excreção e toxicidade. Por fim, serão mostrados os 20 melhores compostos, com a predição dos pontos de clivagem pela enzima CYP3A4 e as conformações preditas no sítio ativo da cruzaina.

4.3.1 Espaço químico e biológico – Energia livre de Gibbs

A figura 43 mostra a distribuição do espaço químico e biológico dos 260 compostos obtidos nas estratégias de triagem virtual. Os dados de $\Delta G^{GoldScore}$ foram obtidos através da função de pontuação GoldScore aplicada na equação de Verdonk e colaboradores.¹⁷⁹ O $\Delta G^{XP\ GlideScore}$ foi obtido diretamente pela função de pontuação empírica GlideScore Extra Precision (XP GlideScore). Para verificar a relação do espaço químico e a energia livre obtido pelos diferentes métodos de docagem molecular, foi construído o mapa de similaridade estrutural por análise de componentes principais (PCA) usando método UNITY. Para o $\Delta G^{GoldScore}$, a energia dos compostos possui intervalos entre -9 e -6 kcal.mol⁻¹, e a maioria dos inibidores possuem energia entre -8 e -7 kcal.mol⁻¹. O $\Delta G^{XP\ GlideScore}$, no entanto, possui intervalo restringido entre -8 e -6, e a maioria dos compostos possui energia entre -7 e -6 kcal.mol⁻¹.

Figura 43 - Diversidade química dos compostos e sua relação com a energia livre de Gibbs obtidos pelos diferentes métodos de docagem molecular.

Fonte: Elaborada pelo autor.

O intervalo de energia comum aos dois métodos permite encontrar compostos com grande probabilidade de interagir com o sítio ativo da cruzaina. Partindo dessas informações, foram obtidos cinco compostos que estavam nos intervalos entre -8 e -7 kcal.mol⁻¹, das energias livre $\Delta G^{\text{GoldScore}}$ e $\Delta G^{\text{XP GlideScore}}$ (ou $\Delta G^{\text{XP Gscore}}$). Para encontrar os inibidores com maior probabilidade de serem ativos contra a cruzaina, foi determinada uma reta ideal (coeficientes angular e linear igual a 1 e 0, respectivamente), através da qual foram selecionadas quatro estruturas com distância menor que 0,35 e que possuíam mesma orientação no sítio ativo, pelos

dois métodos. Na figura 44, são mostradas essas estruturas e as orientações preditas, com os respectivos valores de $\Delta G^{\text{GoldScore}}$ e $\Delta G^{\text{XP Gscore}}$ estimados.

Figura 44 - Predição das orientações dos potenciais inibidores da cruzaina.

Fonte: Elaborada pelo autor.

4.3.2 Espaço químico e biológico – pIC_{50} e pSI estimados

A distribuição do espaço químico e a relação dos valores de pIC_{50} predito pelas redes neurais artificiais e o modelo KPLS molprint2D mostraram que estão localizados nos intervalos de 6 a 7 e 5 a 6, respectivamente. Na predição pelas redes neurais, o mapa mostrou que ambas as faixas de 5 a 6 e 6 a 7 estão quase homogeneamente distribuídas. Na predição de pIC_{50} pelo modelo KPLS molprint2D, a faixa predominante de predição, de 5 a 6, foi distribuída em quase todo o mapa. Na figura 45, é mostrada a PCA e a distribuição da atividade contra o parasita.

Figura 45 - Diversidade química dos compostos e sua relação com a atividade biológica predita.
Fonte: Elaborada pelo autor.

A distribuição da seletividade também foi associada ao mapa de similaridade estrutural (fig. 46). Segundo a predição pela floresta randômica, os valores estão localizados entre 0,5 a 3,0 e, a predição por KPLS radial, entre -0,5 a 2,0. No primeiro, os compostos se concentram entre 1,0 e 2,0, e no segundo, entre 0,5 e 1,0.

Figura 46 - Diversidade química dos compostos e sua relação com a seletividade predita.

Fonte: Elaborada pelo autor.

Após associar a distribuição dos 260 compostos no mapa de similaridade com a atividade biológica e a seletividade pelos métodos de aprendizado de máquina, foi realizada a análise da distância dos parâmetros biológicos na reta ideal (eq. 6). Foram selecionadas 9 moléculas, cujas distâncias foram menores que 0,2, e avaliado posteriormente as orientações conformacionais dos ligantes pelos diferentes métodos. Na figura 47 são mostrados nove compostos promissores inibidores da cruzaina com os valores de energia livre determinados pelos métodos de docagem molecular e a predição da atividade biológica contra o *T.cruzi*.

Figura 47 - Predição das poses e da energia livre dos potenciais inibidores da cruzaina pelos diferentes métodos de docagem molecular.

Fonte: Elaborada pelo autor.

4.3.3 Predição das propriedades ADME-Tox

As distribuições e os dados estatísticos das propriedades ADME-Tox para o conjunto de 260 compostos são mostrados através dos histogramas (fig. 48).

Figura 48 - Distribuição das propriedades farmacocinéticas da biblioteca virtual.

Fonte: Elaborada pelo autor.

A distribuição das propriedades ADME-Tox da biblioteca virtual tem o seguinte perfil: $cLogP$ possui valores mínimo e máximo de 1 e 5,31, respectivamente, estando majoritariamente distribuído entre 2 e 5; mais de 150 compostos podem ser 100% absorvido oralmente, sendo que o valor mínimo é de 83,44% e média de 97,73%; a taxa de permeabilidade em células Caco-2 estão majoritariamente nas faixas entre 500 e 4000 nm.s⁻¹, possuindo média de 1630 nm.s⁻¹. O coeficiente de partição hematocefálica possui valores mínimo e máximo de -1,06 e 0,77, respectivamente, na qual a média é -0,35; O pIC₅₀ para bloqueio de canal K⁺

hERG estão majoritariamente entre -5 e -3, possuindo média de -4,25; e a média da predição de ligação em albumina humana está localizada próximo da origem, com valor de 0,02.

Considerando as moléculas obtidas através da distância da reta ideal para a energia livre de Gibbs e a atividade biológica, foi elaborada a tabela 7 contendo as informações das propriedades farmacocinéticas para as 13 estruturas. Todos os parâmetros preditos estão dentro das condições aceitáveis para ADME-Tox. Porém, o melhor composto é aquele que possui baixa toxicidade e apresenta boa distribuição através da albumina humana. Com base nesse critério, os compostos ZINC 91701517, ZINC 69762537, ZINC 74542092 e ZINC 91701517 apresentam valores de pIC_{50}^{hERG} por volta de -3 e valores próximo de 0 para pIC_{50}^{AH} .

Tabela 7 - Propriedades farmacocinéticas preditas para os 13 compostos mais promissores com atividade anti-*T. cruzi*.

ZINC	cLogP	AO	Caco-2	Log P_{bb}	hERG	AH
 12904725	4,82	100	1303	0,01	-4,71	0,18
 26413486	2,40	96	989	-0,56	-4,81	0,00
 60857051	2,47	100	4402	-0,24	-4,72	0,01
 64344145	3,57	100	950	0,30	-4,74	0,31
 68650118	3,08	92	782	-0,70	-3,72	-0,04
 69762537	2,20	94	664	-0,70	-3,24	-0,02
 70874454	3,76	100	1795	-0,61	-4,93	0,42
 74542092	2,74	100	4012	-0,01	-3,72	-0,06
 75435037	2,98	95	601	-0,59	-4,71	0,14

continua

continuação						
ZINC	cLogP	AO	Caco-2	Log P_{bb}	hERG	AH
 79586889	3,03	100	1097	-0,77	-4,67	-0,04
 87166849	2,39	100	1470	-0,55	-4,97	0,15
 91701517	2,99	100	1205	-0,60	-3,63	0,33
 92964267	3,64	100	1326	-0,20	-4,66	0,38

AO = absorção oral (%), Caco-2 = permeabilidade em células Caco-2 (nm/s), Log P_{bb} = coeficiente de partição hematocefálica, hERG = pIC₅₀ para bloqueio de canais de K⁺ de hERG, e AH = pIC₅₀ de ligação em albumina humana.

Fonte: Elaborada pelo autor.

Identificar os pontos de clivagem é fundamental para um planejamento mais qualificado quanto ao processo de depuração metabólica. Nesse contexto, as dez melhores estruturas, cujo valor de pIC₅₀^{RNA} e pSI^{RF} foram maiores que 6 e 1,8 respectivamente, foram avaliadas *in silico* quanto à metabolização pela enzima CYP3A4 (fig. 49). Quanto maior o círculo verde, maior a probabilidade de o composto interagir nesse ponto com o Fe³⁺ da enzima. O derivado ciano piridínico (ZINC 92964227) e o derivado benzimidazólico (ZINC 44049212) apresentam poucos pontos de clivagem, em relação aos demais (fig. 49). O modelo de clivagem para o derivado piridínico (ZINC 31601820) mostrou que o nitrogênio da piridina pode interagir com o Fe²⁺, justificando o tamanho da esfera nessa região.

Figura 49 - Predição dos pontos de clivagem pelo receptor CYP3A4.

Fonte: Elaborada pelo autor.

4.4 DINÂMICA MOLECULAR E REDES NEURAIS ARTIFICIAIS

4.4.1 Equilíbrios de fase NVT, NPT, R_g e RMSD

Durante a preparação da macromolécula na configuração apo e holo, a proteína precisa estar em um ambiente equiparável ao sistema biológico em que se encontra. Porém, encontrar o estado relaxado no processo de mudança dos estados de protonação dos aminoácidos, devido a mudança do pH e alteração da carga líquida no sistema aquoso torna fundamental a manutenção de NVT e NPT constantes. Nesse contexto, foram equilibrados os sistemas apo e holo em diferentes sistemas. Em todos os processos de equilíbrio NVT e NPT em meios diferentes, a temperatura foi mantida constante (fig. 50).

Meio ácido (pH 5,5)

Meio básico (pH 8,0)

Figura 50 - Equilíbrio de fase das configurações apo e complexo holo em meios ácidos e básicos.
Fonte: Elaborada pelo autor.

Após aplicar as fases de equilíbrio, a dinâmica molecular foi realizada para quatro sistemas: (1) apo e (2) holo em pH 5,5, e (3) apo e (4) holo em pH 8,0. Os quatro sistemas foram submetidos a dinâmica de 300 ns. Para verificar o grau de compactação da proteína em função do tempo e o desvio quadrático médio, foram calculados o raio de giro (R_g) e o RMSD, tomando a conformação inicial em $t=0$ ns como referência. O R_g e o RMSD para os quatro sistemas são mostrados na figura 51. Pode-se observar que as grandes mudanças conformacionais em função do tempo da conformação da apoenzima está relacionada com o estado aberto, tanto em meio ácido, quanto em meio alcalino. No entanto, as baixas mudanças estruturais da holoenzima, nos dois meios, está relacionada com o estado fechado da cruzaina. Nesse caso, o inibidor seria responsável por esse efeito na cavidade enzimática.

Meio ácido (pH 5,5)

Meio básico (pH 8,0)

Figura 51 - RMSD e R_g das configurações apo e complexo holo em meios ácidos e básicos.

Fonte: Elaborada pelo autor.

4.4.2 Estudos estruturais e termodinâmicos

As análises realizadas nessa etapa do trabalho foram baseadas nos estudos das ocupâncias de ligação de hidrogênio, desvio quadrático médio e a energia livre de Gibbs baseado no método MM-PBSA. Conforme mencionado na seção materiais e métodos, as baixas ocupâncias de ligação de hidrogênio foram definidas como sendo menor ou igual a 0,1, e as altas ocupâncias maior que 0,1. De acordo com a tabela 8, as baixas ocupâncias de ligação de hidrogênio diminuem e as altas ocupâncias aumentam, mudando do pH 5,5 para 8,0. Na configuração holo, no

entanto, não houve mudanças significativas no processo de mudança do meio ácido para o meio básico.

Tabela 8 - Número de ocupâncias de ligação de hidrogênio nos diferentes resíduos da cruzafina nas configurações apo e holo em diferentes pH.

Ocupância – HB	Apo (pH 5,5)	Holo (pH 5,5)	Apo (pH 8,0)	Holo (pH 8,0)
Alta	108	126	121	126
Baixa	668	524	569	558

As baixas ocupâncias de ligação de hidrogênio foram consideradas menores ou iguais a 0,1.

Fonte: Elaborada pelo autor.

Essas informações não foram suficientes para afirmar algo em relação a mudança da eficiência catalítica, para os diferentes inibidores, conforme descrito no trabalho de Jones e colaboradores⁶³. Dessa maneira, foram isolados o número de ocupância de ligação de hidrogênio dos únicos aminoácidos que são alterados no processo de mudança do pH considerando somente as histidinas que sofreram alteração no padrão de interação. Na tabela 9, é mostrada a mudança do número de ocupâncias de ligação de hidrogênio das quatro histidinas (HIS43, HIS106, HIS115 e HIS162). Na configuração apo, a HIS43 manteve o número de ocupâncias baixas e altas de 0 e 2, respectivamente, para os meios ácidos e básicos. As outras histidinas, no entanto, sofreram alteração nas altas e baixas ocupâncias: a HIS106 mudou de 12 para 5 (baixa) e 1 para 0 (alta); a HIS115 de 2 para 6 (baixa) e 1 para 0 (alta); a HIS162 manteve 3 (baixa) e mudou de 1 para 0 (alta). Na configuração holo, a mudança do meio ácido para o básico causa: na HIS43 mudança de 1 para 0 (alta) e mantém 3 (baixa); a HIS106 muda de 0 para 1 (alta) e 8 para 4 (baixa); a HIS115 passa de 2 para 1 (alta) e aumenta de 2 para 4 (baixa); por fim, a HIS162 manteve 0 (alta) e diminuiu de 4 para 3 (baixa).

Tabela 9 - Número de ocupâncias de ligação de hidrogênio nas quatro histidinas.

pH	Ocupância	Apo				holo			
		HIS43	HIS106	HIS115	HIS162	HIS43	HIS106	HIS115	HIS162
5,5	alta	0	1	2	1	1	0	2	0
	baixa	2	12	2	3	3	8	2	4
8,0	alta	0	0	1	0	0	1	1	0
	baixa	2	5	6	3	3	4	4	3

As baixas ocupâncias de ligação de hidrogênio foram consideradas menores ou iguais a 0,1.

Fonte: Elaborada pelo autor.

Na figura 52, é mostrado um esquema dos aminoácidos que doam ou recebem ligação de hidrogênio das cadeias laterais das quatro histidinas na apoenzima. Os aminoácidos nas cores preto e vermelho representam a manutenção e a alteração do modo de interação com um determinado aminoácido, respectivamente. Mudando o pH de 5,5 para 8,0, a HIS106 muda substancialmente o número de interações, perdendo 8 com os resíduos TYR84, THR85, ASP87, SER88, THR102 e SER104 e ganhando 1 interação com o TYR89. A HIS115, por outro lado, mantém os pontos de interação, porém recebe 4 novas interações com os resíduos ASN69, GLU73 e SER210. Por fim, a HIS162 mantém a interação com GLN19 e CYS25, porém muda a interação com o TRP184 para o ASP161.

Figura 52 - Ocupância de ligação de hidrogênio dos aminoácidos nas quatro histidinas na cruzaína (apoenzima) nos meios ácidos e básicos.

Fonte: Elaborada pelo autor.

A holo cruzaína, quando muda do meio ácido para o básico, a HIS43 muda somente o ponto de interação da GLY215 para a cadeia principal da ALA41. A HIS106 perde as interações com TYR84, TYR89 e TY103, porém mantém as interações com SER88, THR102 e THR103. A HIS115 mantém as ligações de hidrogênio com os resíduos ASN69, GLU73, GLU117 e SER210, porém a cadeia lateral com o ASN69 fica mais favorável para receber ligação de hidrogênio. Essas

alterações mudam os pontos de interação da cadeia lateral da HIS162, dos resíduos ASP161 e GLY163 para CYS25 e SER183 (fig. 53).

Figura 53 - Ocupância de ligação de hidrogênio dos aminoácidos nas quatro histidinas na cruzaína (holoenzima) nos meios ácidos e básicos.

Fonte: Elaborada pelo autor.

Na figura 54, são mostrados os gráficos RMSD da holo enzima e a energia livre baseada em MM-PBSA. A relação entre a as eficiências catalíticas nos meios ácidos e básicos e a média temporal da energia livre de Gibbs em pH 5,5 e 8,0 é uma constante (aproximadamente 6) (eq. 9):

$$\frac{k_{\text{inat}}/K_i^{\text{pH } 5,5}}{k_{\text{inat}}/K_i^{\text{pH } 8,0}} \sim \frac{\Delta G^{\text{pH } 5,5}}{\Delta G^{\text{pH } 8,0}} \sim 6 \quad (9)$$

Figura 54 - Desvio quadrático médio e a relação da energia livre de Gibbs da holo enzima nos meios ácidos e básicos.

Fonte: Elaborada pelo autor.

Além disso, a energia livre experimental em pH 5,5 é $-7,85 \text{ kcal.mol}^{-1}$. O resultado da dinâmica molecular mostra, em pH 5,5, que a energia de ligação do inibidor K777 é de $-7,20 \text{ kcal.mol}^{-1}$. Os valores de $\Delta G^{\text{XP GlideScore pH 5,5}}$ e $\Delta G^{\text{XP GlideScore pH 8,0}}$ foram substancialmente diferentes e iguais a $-10,28$ e $-9,46 \text{ kcal.mol}^{-1}$ (relação de 1:1).

De fato, as mudanças estruturais estão diretamente ligadas à energia livre de Gibbs, conforme mostrado na figura 54. Um aumento no deslocamento das estruturas secundárias pode diminuir a atividade do inibidor. No entanto, esses gráficos não mostram uma correlação entre os mecanismos moleculares de interação e a afinidade entre a enzima e o inibidor.

4.4.3 Estudos de redes neurais artificiais baseados em dinâmica molecular

Para verificar a relação dos mecanismos moleculares de interação com a afinidade calculada pela dinâmica molecular, foram construídos modelos de redes neurais artificiais de retropropagação. Na figura 55, é mostrado o gráfico de correlação da energia livre predita pelas redes neurais e a calculada teoricamente pela dinâmica molecular, ambos em função do tempo, para os meios ácido e básico. Nos pH 5,5 e 8,0, os coeficientes de correlação obtidos foram de 0,93 e 0,94, respectivamente.

Figura 55 - Correlação dos dados de energia livre preditos e experimentais.
Fonte: Elaborada pelo autor.

A tabela 10 mostra a relação dos pesos dos deslocamentos de alças e sua contribuição para a afinidade do inibidor no pH 5,5. Os pesos dos deslocamentos de estruturas de hélices positivos para as alças 1, 5, 11, 13, 14,15, 16 e 17 e valores negativos para as alças 2, 3, 4, 6, 7, 8, 9, 10 e 12. Para os pesos positivos, o aumento do deslocamento favorece a atividade, e a diminuição do deslocamento diminui a afinidade do inibidor pela enzima. Assim, a alta flexibilidade das alças favorece energeticamente a interação do ligante no sítio ativo da cruzaína.

Tabela 10 - Pesos dos deslocamentos de alças em pH 5,5.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
a-1	0,6	-1,8	1,2	1,1	-1,4	1,9	-1,5	-1,7	0,4	1,0	2,0	0,4	1,3
a-2	-0,2	-1,3	-0,3	0,6	-1,4	0,2	0,0	0,4	1,1	-0,5	-2,8	-0,5	-1,4
a-3	-0,3	0,5	1,7	3,0	-1,1	-0,3	-0,6	-0,4	1,6	-0,9	-1,3	0,5	-0,9
a-4	0,4	1,7	-1,5	-2,6	-0,5	-1,0	-1,8	0,4	-1,0	-2,0	-0,1	0,3	0,9
a-5	0,3	-1,0	3,9	0,2	0,6	-3,8	0,6	3,1	-0,8	0,6	0,9	0,0	-0,1
a-6	-1,2	0,4	-0,8	-1,9	0,1	-2,2	-0,2	-1,3	-1,8	0,1	-0,8	0,1	-0,5
a-7	-0,3	0,5	-0,6	0,8	-0,3	1,8	-0,3	-0,7	-1,8	-1,5	-0,5	0,3	2,5
a-8	0,2	-4,5	0,7	0,1	-2,7	-1,7	-1,0	-0,9	1,6	1,0	-1,0	-0,6	0,1
a-9	0,1	-0,6	-1,0	1,9	1,4	0,2	-1,3	-2,1	1,2	-0,7	1,0	-0,3	-0,3
a-10	0,2	-0,6	-0,7	0,4	-0,2	2,5	0,1	-2,7	-0,2	-0,2	-0,1	1,0	-0,7
a-11	0,3	1,0	-0,7	0,1	-1,1	0,9	-0,7	2,7	1,1	-1,0	-1,3	0,0	0,9
a-12	-0,4	0,5	-0,7	-0,2	-0,1	0,7	0,1	-1,0	-0,1	-2,1	-1,4	-0,2	-2,0
a-13	0,1	0,4	0,2	0,6	0,2	-0,3	-0,2	-2,3	1,1	-0,6	-0,6	0,0	-1,6
a-14	0,4	-1,1	1,3	1,2	0,7	0,2	-0,1	-0,6	0,1	1,0	1,5	0,9	1,1
a-15	-0,2	-0,1	-0,2	0,3	-2,3	-1,3	0,1	1,8	-0,4	0,8	1,1	1,0	0,5
a-16	-0,4	1,4	0,0	-3,4	1,3	2,2	-1,0	1,4	-1,5	-2,0	-1,4	0,6	1,3
a-17	0,6	0,2	1,5	2,1	0,3	2,3	0,9	2,0	-0,3	2,4	0,5	-0,3	2,1

a = alça.

Fonte: Elaborada pelo autor.

Na tabela 11, as estruturas de folhas- β em pH 5,5 apresentam os pesos positivos em 2, 4, 6, 7 e 8 e pesos negativos nas estruturas 1, 3, 5, 9 e 10. Os pesos positivos mostram que eventuais aumentos do RMSD dessas estruturas geram estabilidade do inibidor no sítio ativo, decorrente da mudança estrutural das folhas- β . No entanto, em pH 8,0, o deslocamento dessas estruturas diminui a atividade biológica do inibidor no sítio de interação.

Tabela 11. Pesos dos deslocamentos de folhas- β em pH 5,5.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
f-1	-0,9	-1,8	1,3	-0,4	-0,3	2,5	0,0	-0,6	-0,9	3,2	-0,7	0,4	0,8
f-2	0,4	-0,1	-0,6	-1,0	-0,1	0,9	0,7	-0,6	0,7	-0,4	1,6	0,9	0,2
f-3	-0,4	1,0	-0,5	0,0	1,7	-0,6	-1,0	-3,0	-2,8	-1,0	0,6	-0,7	0,9
f-4	-0,3	1,6	0,9	1,1	0,7	-1,6	-1,4	1,4	0,7	-0,4	1,0	-0,5	0,1
f-5	-0,1	1,3	-0,7	2,2	-1,1	-2,2	-0,2	-2,7	2,3	1,5	-0,7	-0,3	-0,9
f-6	0,2	1,0	-1,1	-0,9	2,4	1,0	-0,2	2,6	1,3	-1,8	1,8	0,8	2,9
f-7	0,5	1,2	0,5	1,0	0,1	0,9	2,5	2,1	0,0	2,8	0,4	1,7	2,2
f-8	0,6	-1,6	0,9	1,0	0,2	-0,5	0,5	-1,1	1,3	-0,6	1,3	0,5	1,4
f-9	-0,3	2,4	-0,4	0,5	0,2	-0,4	-0,3	-0,6	-2,1	0,9	-2,2	-1,4	-0,3
f-10	-0,1	-3,4	-0,8	-2,2	1,4	3,0	-2,0	0,7	1,3	-0,2	0,4	-1,1	0,2

f = folha- β

Fonte: Elaborada pelo autor.

Na tabela 12, os pesos positivos associados a estruturas de hélices em pH 5,5, estão nas hélices 1, 2 e 5. Os pesos negativos estão somente nas hélices 3, 4 e 6. Assim, a rigidez das hélices 3, 4 e 6 e a flexibilidade das hélices 1, 2 e 5 mantém

o ligante no sítio ativo da enzima. Por outro lado, a flexibilidade das hélices 3 e 4, e a rigidez das hélices 1, 2, 5 e 6 tornam o inibidor instável no sítio de ligação.

Tabela 12 - Pesos dos deslocamentos de hélices em pH 5,5.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
h-1	0,2	0,2	-0,2	1,0	1,1	0,1	-0,8	3,4	0,9	-2,2	0,3	0,2	1,5
h-2	0,2	1,6	0,0	0,1	-1,3	-2,4	1,5	1,9	0,1	-0,3	0,3	1,9	1,0
h-3	-0,2	-1,3	-0,7	-0,7	0,0	0,6	-0,1	-1,1	0,4	-1,8	-4,0	0,2	1,3
h-4	-0,4	-3,3	-0,7	-0,6	0,2	0,6	1,0	-0,8	-1,0	-1,1	-1,1	-0,2	-1,2
h-5	0,4	2,3	1,3	0,8	0,7	0,0	0,2	1,7	2,3	0,5	1,4	-0,1	-0,2
h-6	0,1	-2,5	0,4	-0,2	-0,3	-1,8	1,5	-1,2	0,1	0,7	0,0	-0,1	0,1

h = hélice

Fonte: Elaborada pelo autor.

A relação dos pesos dos deslocamentos de alças e sua contribuição para a afinidade do inibidor no pH 8,0 são mostrados nas tabelas 13, 14 e 15, respectivamente. Na tabela 13, as alças 1, 3, 4, 7, 8, 9, 10, 11, 15, 16 e 17 apresentam pesos positivos e as alças 2, 5, 6, 12, 13 e 14 apresentam pesos negativos. A rigidez das alças 1, 3, 4, 7, 8, 9, 10, 11, 15, 16 e 17 desfavorece o modo de ligação do inibidor e deslocamentos das alças 2, 5, 6, 12, 13 e 14 faz o inibidor ter menos afinidade pela cavidade enzimática.

Tabela 13 - Pesos dos deslocamentos de alças em pH 8,0.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
a-1	0,7	1,6	1,2	1,4	0,8	-0,7	0,8	1,1	1,6	0,6	0,8	1,7	-0,3
a-2	1,0	0,0	-0,1	0,4	-0,6	-0,3	-0,3	-0,7	-0,4	-0,9	-0,2	-1,5	0,6
a-3	2,1	-0,8	0,6	0,3	0,1	-0,8	0,3	0,9	-0,1	0,5	0,5	0,0	0,7
a-4	-0,6	-0,2	-0,1	-1,1	0,7	-3,5	0,0	0,7	1,5	0,4	0,5	0,5	1,2
a-5	-0,1	-1,2	0,1	-0,4	-0,4	0,9	-0,2	0,7	-1,2	-0,1	0,2	-0,6	-0,8
a-6	-0,9	-1,6	-0,4	-1,9	0,0	0,7	-0,2	-1,1	-1,7	0,6	0,7	-1,2	-2,0
a-7	-0,3	0,5	0,1	0,4	-0,9	-1,5	0,2	1,0	-0,6	0,9	1,1	0,0	0,7
a-8	0,6	-0,2	0,5	0,6	-1,1	2,9	0,2	-0,3	-2,3	0,3	-0,6	2,5	1,8
a-9	0,2	-1,2	0,1	1,2	1,6	-1,7	-0,2	-0,6	1,9	0,1	-0,5	-0,2	1,3
a-10	0,7	0,2	0,1	0,6	1,8	0,1	0,1	0,5	0,8	-0,1	1,1	-1,0	-0,1
a-11	-0,1	0,2	0,1	-0,8	1,6	1,9	0,2	-0,4	0,6	0,2	0,7	1,9	0,4
a-12	0,3	-2,6	-0,1	-0,5	-0,7	-0,1	-0,3	0,1	0,6	-0,6	1,4	-0,8	0,2
a-13	-1,0	-0,5	-0,2	-0,5	1,6	-1,3	-0,2	-0,4	0,7	0,1	-0,3	0,4	0,3
a-14	-0,7	-0,3	-0,5	0,3	-1,0	-1,9	-0,6	-0,2	-0,3	0,2	-1,4	2,8	0,2
a-15	-0,4	0,6	0,2	0,2	0,9	0,3	0,1	1,2	0,2	0,1	-0,4	-0,7	1,0
a-16	0,3	1,7	0,1	1,5	1,9	3,1	-0,2	-1,7	1,6	-0,7	0,4	1,8	-0,1
a-17	0,7	1,6	0,5	0,9	1,3	-1,7	0,3	0,8	2,0	0,2	0,3	-0,5	-1,2

a = alça.

Fonte: Elaborada pelo autor.

Na tabela 14, as estruturas de folhas- β foram simuladas em pH 8,0, e segundo as redes neurais, os pesos das folhas 1, 2, 3, 4, 6, 9 e 10 são majoritariamente negativos e mostram que os deslocamentos desses elementos de estruturas secundárias são desfavoráveis para a afinidade do inibidor pela enzima. A rigidez, no entanto, torna o ligante mais estável no sítio ativo. As folhas 5, 7 e 8, no entanto, apresentam o efeito inverso, ou seja, a rigidez torna o inibidor menos potente e o deslocamento o torna mais potente.

Tabela 14 - Pesos dos deslocamentos de folhas- β em pH 8,0.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
f-1	1.0	0.4	0.1	0.8	-1.3	-0.8	-0.2	-1.9	-1.0	-0.1	0.3	-0.6	-0.1
f-2	1.4	0.3	0.3	-0.9	-2.0	0.8	0.1	-0.4	-1.3	-0.4	-0.5	-0.7	0.2
f-3	-1.2	-1.2	-0.3	-1.4	0.6	0.7	-0.5	1.1	1.6	-0.6	-0.7	1.7	1.2
f-4	-1.2	-1.0	-0.2	-0.1	1.6	-0.2	-0.3	1.9	-0.3	0.5	0.8	0.7	-0.8
f-5	0.3	-0.1	0.3	-0.4	-0.6	1.6	0.1	-0.6	1.2	-0.5	-0.3	0.7	0.3
f-6	-1.5	-1.6	0.0	0.5	2.3	0.3	-0.2	-0.9	0.9	0.6	-0.2	-0.8	-2.6
f-7	0.8	0.5	0.4	1.1	1.6	0.0	0.1	0.0	0.8	0.2	0.5	-0.2	-0.2
f-8	-1.2	0.0	0.0	0.5	-0.5	0.4	0.1	-0.5	0.5	-0.5	-0.3	0.0	-0.5
f-9	-1.3	-0.5	-0.2	-0.7	-1.2	1.0	-0.3	-0.7	-1.6	-0.2	-1.1	-1.0	1.3
f-10	-0.4	1.8	0.0	0.9	-2.5	-2.0	-0.2	0.1	2.1	0.0	-2.0	-1.0	-1.5

f = folha- β

Fonte: Elaborada pelo autor.

As estruturas de hélices apresentaram pesos negativos nas estruturas 1, 3 e 4, e valores positivos em 2, 5 e 6, no pH 8,0 (tab. 15). Assim, para os valores negativos, os deslocamentos desfavorecem a atividade biológica e a rigidez favorece a ligação do inibidor no sítio ativo da enzima. Para os valores positivos, as estruturas flexíveis são fundamentais para aumentar a atividade biológica e a rigidez desfavorece a atividade biológica.

Tabela 15 - Pesos dos deslocamentos de hélices- α em pH 8,0.

Neurônio	1	2	3	4	5	6	7	8	9	10	11	12	13
h-1	-1.8	2.7	-0.1	0.0	-0.2	0.5	-0.1	-0.6	-1.3	0.2	0.6	2.8	-1.9
h-2	-0.3	0.5	0.3	0.7	0.5	1.6	0.3	-0.1	0.1	-0.9	1.3	-0.1	1.3
h-3	0.6	-0.2	-0.5	-1.2	-0.3	1.0	-0.1	-1.3	-1.5	0.1	0.8	0.2	-1.2
h-4	-0.3	-1.0	-0.2	-0.7	2.1	-0.2	-0.3	-1.2	1.4	-0.2	0.4	-1.8	0.4
h-5	1.6	-2.1	0.4	0.9	0.2	-0.7	0.2	0.5	1.1	0.0	1.1	-1.4	0.2
h-6	0.1	-0.4	0.8	1.6	-0.3	1.2	0.4	1.1	0.4	0.4	-0.5	-0.3	-1.4

h = hélice

Fonte: Elaborada pelo autor.

A partir dessas informações, foi construído um mapa compondo o voto da maioria dos neurônios no processo de classificação de cada estrutura secundária, variando o pH. A cor verde representa o voto majoritário dos neurônios com peso positivo, e a cor azul representa o voto majoritário dos neurônios que classificam os deslocamentos das estruturas secundárias com peso negativo (fig. 56).

Figura 56 - Mapa de redes neurais artificiais aplicado aos deslocamentos das estruturas secundárias. Fonte: Elaborada pelo autor.

A representação do mapa das redes neurais mostrou que o padrão é diversificado para cada estrutura secundária. Essa alteração de padrão é causada pela mudança do estado de protonação de apenas dois resíduos de histidinas (HIS43 e HIS115), provocado pelo aumento de pH e apresentam os seguintes resultados (fig. 56): (1) em pH 5,5, 8 estruturas de alças são verdes e 9 são azuis, e no pH 8,0, 11 neurônios são verdes e apenas 6 são azuis, *i.e.*, o aumento do pH faz com que o padrão de mobilidade da maioria das alças aumente a afinidade de interação com a enzima; (2) em pH 5,5, 5 estruturas de folhas-β são classificadas em verde e 5 estruturas em azul, e no pH 8,0, 3 estruturas ficam verdes e 7 são azuis, ou seja, o aumento do pH faz com que a rigidez desses elementos de estruturas secundárias aumente a afinidade do inibidor pela enzima; (3) em pH 5,5 e

8,0 as hélices possuem a mesma quantidade de cores, sendo 3 verdes e 3 azuis. Porém a hélice que contém a cisteína catalítica muda o padrão de verde para azul. Em geral, o aumento do RMSD das hélices pode aumentar a afinidade dos inibidores e a diminuição do RMSD diminui a afinidade do inibidor pela enzima.

Na figura 57, são mostradas as distribuições do RMSD dos elementos de estrutura secundária que compõem a tríade catalítica no sítio ativo. Segundo os histogramas, mesmo no sítio ativo, a alça 14 apresenta amplitude de deslocamento maior do que hélices- α e folhas- β nos dois pH. Esse movimento é mais intenso no meio ácido do que no meio básico. De acordo com a referência, isso é esperado para esse elemento de estrutura secundária, por apresentar maior fator de temperatura.²⁰⁷ De acordo com a distribuição do RMSD ao longo da dinâmica molecular, a alça 14, folha 6 e a hélice 1, em pH 5,5, possuem RMSD médio de $(2,49 \pm 0,50)$, $(1,23 \pm 0,20)$ e $(1,29 \pm 0,12)$ Å, e em pH 8,0, $(2,15 \pm 0,56)$, $(1,37 \pm 0,22)$ e $(1,21 \pm 0,12)$ Å, respectivamente nos meios correspondentes. Assim, a alça 14 em meio ácido tende a ser mais flexível do que no meio básico. A folha- β 6, tende a ser menos flexível em meio ácido do que em básico. A hélice- α 1 tende a ser mais flexível em meio ácido do que em meio básico. De acordo com o histograma (fig. 57) e o mapa de redes neurais artificiais (fig. 56), a mobilidade da alça 14 é fundamental para manutenção do ajuste induzido favorável da folha- β 6 e da hélice- α 1 em meio ácido. No entanto, quando a enzima se encontra em meio básico, a flexibilidade dos três elementos de estrutura secundária torna o ajuste do inibidor no sítio catalítico desfavorável para a atividade biológica. Consequentemente, o ambiente químico que comporta a tríade catalítica CYS25, HIS162 e ANS182 é afetado com essa mudança estrutural. Os resultados das ocupâncias de ligação de hidrogênio da histidina catalítica (HIS162) com os resíduos ao redor, nos meios ácido e básico (fig. 53) revelam que em pH 5,5, a flexibilidade da hélice- α 1 faz com que a CYS25 realize um ataque nucleofílico ao inibidor K777. No entanto, de acordo com as redes neurais artificiais, em pH 8,0, a diminuição da flexibilidade da alça 14 combinado com a flexibilidade da folha- β 6 e da hélice- α 1 apontadas pelos histogramas são cruciais para diminuir a afinidade do inibidor K777 pela enzima. Essa relação do meio básico justifica a ocupância de ligação de hidrogênio da HIS162 ser predominante com a CYS25, ASN182 e SER183, causando um bloqueio de um possível ataque nucleofílico e, consequentemente, a diminuição da eficiência catalítica do inibidor.

Figura 57 - Distribuição do RMSD das estruturas secundárias que compõem a tríade CYS25 (hélice- α 1), HIS162 (folha- β 6) e ASN182 (alça 14).

Fonte: Elaborada pelo autor.

Para complementar com as informações obtidas pelos métodos computacionais, a figura 58 ilustra o comportamento dos resíduos catalíticos CYS25 e HIS162 em meio ácido e básico. Em meio ácido ($t = 146$ ns), os dois resíduos ficam próximos para formar o par iônico a uma distância de 2,60 Å ($\Delta G = -11,25$ kJ.mol⁻¹). No entanto, o ambiente químico favorece a CYS25 flexível a se aproximar do ligante a 3,32 Å ($t = 147$ ns) e, conseqüentemente, se afastar da HIS 162, tornando a interação exergônica ($\Delta G = -36,84$ kJ.mol⁻¹). No meio básico, o

deslocamento dos elementos de estrutura secundária muda o ambiente químico do sítio ativo, fazendo com que a formação do par iônico seja mais favorável ($\Delta G = -18,69 \text{ kJ.mol}^{-1}$) do que o ataque nucleofílico da cisteína catalítica ao inibidor ($\Delta G = -8,39 \text{ kJ.mol}^{-1}$). A formação do par iônico sendo favorável, torna a CYS25 menos acessível ao inibidor, fazendo com que a eficiência catalítica seja reduzida em 6 vezes, corroborando com os dados experimentais ⁶³.

Meio ácido

Meio básico

Figura 58 - Comportamento da CYS25 e HIS162 nos meios ácido e básico.
Fonte: Elaborada pelo autor.

4.5 MAPEAMENTO DO SÍTIO ATIVO DA CRUZAÍNA

4.5.1 Conjunto de dados

O conjunto de dados possui boa diversidade estrutural, tanto no arcabouço quanto nos substituintes (conjunto de dados presente no anexo). Além de possuir boa diversidade química, os dados apresentam boa amplitude de atividade biológica,

com valores de pIC_{50} variando entre 4,1 e 6,7. A série treinamento selecionada pelo AutoQSAR é composta por 8 inibidores (20% do conjunto de dados): **3, 6, 15, 21, 24, 29, 32 e 38**.

4.5.2 QSAR 2D

O AutoQSAR produziu uma grande diversidade de modelos, como mostrado na tabela 16. Todos os modelos de QSAR variaram somente no método KPLS e mantiveram constante o número ótimo de componentes ($N = 1$). O melhor modelo foi escolhido baseado nos respectivos valores de *score*, seguido dos maiores valores de q^2 (r^2_{pred}). Portanto, a análise foi fundamentada no modelo KPLS molprint2D, com valores de 0,82, 0,83 e 0,82 para *score*, q^2 e r^2 , respectivamente. Os erros RMSE e SD do modelo foram de 0,30 e 0,26, respectivamente.

Tabela 16 - Diversidade de modelos construídos pelo AutoQSAR.

Modelo	Score	q^2	r^2	SD	RMSE	N
KPLS molprint2D	0,82	0,83	0,82	0,26	0,30	1
KPLS_linear	0,81	0,81	0,80	0,27	0,30	1
KPLS dendritic	0,80	0,80	0,91	0,28	0,28	2

$q^2 = r^2$ para a série teste (r^2_{pred}); r^2 = coeficiente de determinação da série treinamento; RMSE = erro quadrático médio da série teste; SD = desvio padrão (do modelo).

Fonte: Elaborada pelo autor.

4.5.2.1 Estudos de QSAR 2D

A discussão será fundamentada nos estudos de QSAR e nas estratégias de docagem molecular para os inibidores da cruzaina. Interações importantes dos inibidores da cruzaina foram identificadas. O grupo amino possui, na maioria dos compostos da série, contribuição positiva para a atividade, podendo ser um doador ou aceptor de ligação de hidrogênio. Os três compostos mais potentes apresentaram padrões comuns de contribuição positiva no grupo pirazol, mostrando que o grupo pode atuar como doador ou aceptor de ligação de hidrogênio, e/ou fazer interações π - π com aminoácidos no subsítio da enzima. Em contraste, contribuições negativas nas estruturas com baixa atividade são representadas por diidroindenos, piridinas ligadas a morfolinás, carbonos carbonílicos de amidas e cicloheptanos (fig. 59).

Figura 59 - Visualização das contribuições atômicas do Canvas.
 Fonte: Elaborada pelo autor.

As glicolamidas **27** e **37** (fig. 60A) apresentaram contribuições negativas e neutras no NH_2 dos grupos aminobenzeno e aminopiridina, respectivamente. Foram investigadas como as propriedades moleculares destas estruturas estão associadas ao modo de interação no sítio ativo da cruzaina. As análises das poses dos ligantes (fig. 60B) revelaram que o grupo amina pode fazer ligação de hidrogênio com a cadeia principal do resíduo ASP161. Para a glicolamida **27**, o nitrogênio pertencente à piridina pode diminuir a ligação de hidrogênio com este resíduo de aminoácido. Entretanto, para o inibidor **37**, o carbono 3 do anel benzênico teve contribuição neutra, fazendo com que a amina possua contribuição positiva para a ligação de hidrogênio.

Figura 60 - A) Glicolamidas **27** e **37** e suas contribuições para a atividade biológica. B) Docagem molecular das estruturas no sítio ativo da cruzaina. C) Energia de ligação de hidrogênio determinada por DFT.

Fonte: Elaborada pelo autor.

4.5.2.2 Ligação de halogênio e metila

Após investigar as poses dos ligantes resultantes da docagem molecular, identificaram-se compostos que podem fazer ligação de halogênio no sítio ativo da cruzaina. O campo de força OPLS3 identificou a região eletrofílica associada ao halogênio do ligante e a região nucleofílica presente no correspondente resíduo da macromolécula. A região eletrofílica identificada na simulação de docagem molecular é conhecida como σ -hole, que se encontra como uma densidade de carga positiva na porção terminal da extensão da ligação covalente. Neste estudo, exemplifica-se o caso do oxadiazol **18** (fig. 61), que apresentou ligação de halogênio de um dos átomos Cl ao resíduo HIS162 (distância de 3,49 Å, ângulo do σ -hole de 164,9°).

Para investigar a contribuição positiva do oxigênio no mapa de contribuição do composto líder, estudos de docagem molecular foram realizados. Uma interação do tipo CH- π do oxadiazol **1** com o resíduo TRP184 foi identificada (fig. 61). O oxigênio do grupo metoxi permitiu angulação (144,9°) e distância (3,07 Å) adequadas para que essa interação ocorra.

Figura 61 - Ligação de halogênio do inibidor 18 à HIS162, com ângulo e distância.

Fonte: Elaborada pelo autor.

4.5.3 Hipótese farmacofórica baseada na estrutura do ligante

Para gerar a hipótese farmacofórica, os três compostos mais ativos da série foram utilizados (fig. 62). Todos os modelos farmacofóricos construídos pelo algoritmo genético apresentaram significância estatística semelhante, isto é, todos apresentaram o mesmo peso farmacofórico e diferenças conformacionais finais desprezíveis. Assim, foi escolhido o modelo farmacofórico que apresentou menor energia ($-0,16 \text{ kcal.mol}^{-1}$).

Figura 62 - A) Modelo farmacofórico gerado pelo GALAHAD. B) Docagem molecular dos 3 ligantes mais potentes da série que foram usados na geração do farmacóforo. Os resíduos de aminoácidos realizando interações são mostrados.

Fonte: Elaborada pelo autor.

O farmacóforo (fig. 62A) apresentou três grupos doadores de ligação de hidrogênio, um grupo aceptor de ligação de hidrogênio, dois grupos que podem ser doadores ou aceptores de ligação de hidrogênio e três grupos que podem fazer interação hidrofóbica ou aromática. Para mapear o sítio da cruzaina, estas informações foram utilizadas com os resultados fornecidos pelas simulações de docagem molecular (fig. 62B).

Para investigar a relevância das características farmacofóricas, foram avaliados os pesos associados às análises conformacionais geradas pelo GALAHAD, a partir do raio de tolerância (r). O raio de tolerância descreve o volume disponível para que um grupo que contenha aquelas características ocupe aquele ponto farmacofórico; esses são os pesos associados a cada característica advinda das estruturas químicas. Estes pesos são importantes na eventual realização de um vHTS, quando pesos são atribuídos aos resíduos de aminoácidos mais importantes com o objetivo de avaliar novos arcabouços. Na figura 62, observou-se que o grupo amino é o mais importante deste modelo (esfera 1, $r = 0,93 \text{ \AA}$) e pode se comportar como grupo doador ou aceptor de ligação de hidrogênio. Esta característica também está presente no pirazol (esfera 9, $r = 0,76 \text{ \AA}$). O anel benzênico está diretamente ligado as características aromáticas ou hidrofóbicas (esfera 2, $r = 0,54 \text{ \AA}$). No anel oxadiazol, observou-se duas esferas que são aceptoras de ligação de hidrogênio (esferas 3 e 5, $r = 0,37$ e $0,76 \text{ \AA}$, respectivamente). Características aromáticas também estão presentes nos anéis oxadiazol e pirazol (esferas 4 e 8, $r = 0,24$ e $0,32 \text{ \AA}$, respectivamente). Esses grupos podem participar de interações π - π com os resíduos HIS162 e TRP184 que estão presentes no subsítio S1'. Por fim, o oxigênio da amida (esfera 6, $r = 0,44 \text{ \AA}$) e o nitrogênio (esfera 7, $r = 0,30 \text{ \AA}$) são aceptores e aceptores/doadores de ligação de hidrogênio, respectivamente.

4.5.4 QSAR 3D baseado no GALAHAD

Para construir o modelo CoMFA, o farmacóforo foi empregado como molde para fazer o alinhamento individual dos inibidores a partir do GALAHAD (fig. 63A). O alinhamento molecular obtido nesse experimento apresentou melhor sobreposição na região contendo o núcleo benzoaminas-oxadiazol (lado direito, fig. 63A). O

melhor perfil de sobreposição nessa região ocorre devido a baixos graus de liberdade, desfavorecendo a rotação desse grupo. No lado esquerdo na figura 63A, entretanto, os substituintes apresentam um maior número de ligações rotacionáveis em relação à sobreposição do grupo benzoamina-oxadiazol, e ocupam um volume maior do espaço 3D.

Figura 63 - A) Alinhamento molecular baseado nas características farmacofóricas dos ligantes. B) Mapas de contorno para os inibidores **1** e **38** dos campos estereoquímico e eletrostático obtidos a partir do método CoMFA StDev*Coeff.

Fonte: Elaborada pelo autor.

A partir do alinhamento molecular resultante do farmacóforo, o modelo CoMFA foi gerado utilizando a série treinamento e teste selecionados a partir do algoritmo do AutoQSAR. Para refinar o modelo, foi aplicado o método de focalização de região para ajustar os pesos dos descritores, aumentando os mais relevantes e atenuando os menos importantes (tab. 17).

Tabela 17 - Diversidade dos modelos 3D acompanhado pela capacidade de predição.

Treinamento (%)	SDC	q^2	SEP	r^2	SEE	N	r^2_{pred}
80	0,6	0,76	0,34	0,91	0,21	2	0,57
	0,9	0,77	0,34	0,93	0,19	4	0,82

SDC = coeficiente de desvio padrão; q^2 = coeficiente de correlação com validação cruzada *leave-one-out*; r^2 = coeficiente de correlação sem validação cruzada; N = número ótimo de componentes; SEP = erro padrão de predição; SEE = erro padrão da estimativa; r^2_{pred} = coeficiente de correlação de predição.

Fonte: Elaborada pelo autor.

A partir do melhor modelo ($r^2_{pred} = 0,82$; $q^2 = 0,77$; e $r^2 = 0,93$), foram efetuadas as projeções tridimensionais estereoquímicas e eletrostáticas (fig. 63B)

dos inibidores **1** ($pIC_{50} = 6,7$) e **38** ($pIC_{50} = 4,1$). Estes inibidores foram selecionados por apresentarem diferença nos extremos de atividade. Segundo o mapa de contorno, o grupo metoxifenila fica próximo aos poliedros 3 (grupos eletronegativos favoráveis) e 4 (grupos volumosos favoráveis). Em contraste, a amina do inibidor **38** apresenta carga parcial positiva nos hidrogênios, o que faz que este grupo seja menos favorável para a atividade.

A simulação de docagem molecular (fig. 64) do composto líder mostrou que as regiões 1 e 2 estão relacionadas aos resíduos CYS25, TRP26, MET68. Todavia, os resíduos LEU160 e ASP161 pertencem a região 2. As cargas parciais positivas das cadeias principais dos resíduos de aminoácidos contidos nessas regiões são complementares às informações extraídas no mapa de contorno. Na região 1, o resíduo TRP184 pode fazer ligação π - π com o benzeno do inibidor **1**. Observa-se ainda que a região 4, que corresponde ao subsítio S_1 , não foi explorada nesta série. O planejamento de novos inibidores com grupos volumosos nesta região poderá ser favorável para a atividade biológica.

Figura 64 - Docagem molecular mostrando as regiões do mapa de contorno CoMFA (estereoquímico e eletrostático). Regiões em azul (região 1-3) mostram onde os grupos eletronegativos são favoráveis para a atividade biológica e a região em verde (região 4), favorável para grupos volumosos.

Fonte: Elaborada pelo autor.

4.5.5 Capacidade preditiva dos modelos 2D e 3D

Após a construção dos modelos de QSAR 2D e 3D, foram gerados os gráficos dos valores de pIC_{50} preditos e estimados *versus* experimentais nas séries

treinamento e teste (fig. 65). Ambos os modelos apresentaram boa correlação entre as séries treinamento e teste e foram validados internamente (r^2_{KPLS} e r^2_{CoMFA} , 0,82 e 0,77, respectivamente) e externamente (q^2_{KPLS} e $r^2_{\text{pred CoMFA}}$, 0,83 e 0,82, respectivamente). Nesse trabalho foram construídos modelos robustos que podem ser utilizados nas predições de novos inibidores da cruzaina, dentro da diversidade estrutural estudada.

Figura 65 - Gráfico dos valores preditos e estimados versus experimentais dos valores de pIC₅₀ dos modelos KPLS molprint2D e CoMFA.

Fonte: Elaborada pelo autor.

5 CONCLUSÃO

A pesquisa de doutorado possibilitou investigar os aspectos moleculares fundamentais para o desenvolvimento de inibidores da cruzaina, um alvo validado para a doença de Chagas. Uma das grandes estratégias utilizadas nesse trabalho foi o uso da inteligência artificial. Assim, na primeira parte, os modelos construídos possibilitaram encontrar fragmentos moleculares e as características desejáveis para a melhoria da potência, da seletividade e das propriedades farmacocinéticas. Como consequência, foram encontrados sub-estruturas e grupos químicos, caracterizados como bioisóteros. Esses componentes estruturais têm impacto no desenvolvimento de novos compostos com atividade anti-*T. cruzi* e com boas propriedades farmacocinéticas.

Os métodos de inteligência artificial revelaram aspectos moleculares e estruturais associados à cruzaina e sua relação com o pH. Isso possibilitou descobrir que o deslocamento dos elementos estruturais pode afetar significativamente a potência do composto no sítio ativo. O planejamento de novos fármacos, levando em consideração esses aspectos bioquímicos possibilitará maior sucesso em ensaios pré-clínicos e clínicos.

A terceira parte da pesquisa revelou os resíduos fundamentais para o reconhecimento molecular entre um composto e a cruzaina. Além disso, foi descoberta a interação CH- π , a qual não é descrita na literatura. Outras interações, tais como as ligações de halogênio e hidrogênio também foram relevantes e descritas nesse trabalho. Essas informações são fundamentais, pois o planejamento de compostos baseados nos modelos de QSAR e docagem molecular podem guiar o planejamento de inibidores mais potentes através de triagens virtuais e predição de novos compostos com atividade anti-*T. cruzi*.

A combinação dos aspectos farmacocinéticos e farmacodinâmicos (primeira parte) e dos estados do inibidor e do ambiente químico da cavidade enzimática (segunda parte), bem como os resíduos que fazem parte no reconhecimento molecular e as interações intermoleculares no complexo enzima-inibidor (terceira parte) são guias úteis em química medicinal para o planejamento de novos inibidores para o tratamento da doença de Chagas.

REFERÊNCIAS

- 1 CHAGAS, C. Nova tripanozomíase humana: estudos sobre a morfologia e o ciclo evolutivo do *Schizotrypanum cruzi* n. gen., n. sp., agente etiológico de nova entidade morbida do homem. **Memórias do Instituto Oswaldo Cruz**, v. 1, n. 2, p. 159–218, 1909.
- 2 WORLD HEALTH ORGANIZATION. **American Trypanosomiasis (Chagas disease)**. Disponível em: <<http://www.who.int/chagas/en/>>. Acesso em: 9 nov. 2018.
- 3 PÉREZ-MOLINA, J. A.; MOLINA, I. Chagas disease. **Lancet**, v. 6736, n. 17, p. 82–94, 2017.
- 4 SCHMUNIS, G. A.; YADON, Z. E. Chagas disease: a Latin American health problem becoming a world health problem. **Nature**, v. 465, p. S6–S7, 2010.
- 5 MONTGOMERY, S. P. et al. Perspective piece what do we know about chagas disease in the United States? **American Journal of Tropical Medicine and Hygiene**, v. 95, n. 6, p. 1225–1227, 2016.
- 6 MANNE-GOEHLER, J. Chagas disease in the United States: out of the shadows. **Clinical Infectious Diseases**, v. 64, n. 9, p. 1189–1190, 2017.
- 7 ANTINORI, S. et al. Chagas disease in Europe : a review for the internist in the globalized world. **European Journal of Internal Medicine**, v. 43, p. 6–15, 2017. doi: 10.1016/j.ejim.2017.05.001.
- 8 IMAI, K. et al. Chronic Chagas disease with advanced cardiac complications in Japan: case report and literature review. **Parasitology International**, v. 64, n. 5, p. 240–242, 2015.
- 9 PINTO, A.; PETT, S.; JACKSON, Y. Identifying Chagas disease in Australia: an emerging challenge for general practitioners. **Tropical Medicine and International Health**, v. 19, n. 2, p. 212–218, 2014.
- 10 MONTEIRO, F. A. et al. Evolution, systematics, and biogeography of the triatominae, vectors of chagas disease. **Advances in Parasitology**, v. 99, p. 265–344, 2018. doi: 10.1016/bs.apar.2017.12.002.
- 11 HARRIS, N. et al. Autochthonous Chagas disease in the southern United States: a case report of suspected residential and military exposures. **Zoonoses and Public Health**, v. 64, n. 6, p. 491–493, 2017.
- 12 COURA, J. R. The main sceneries of chagas disease transmission. The vectors, blood and oral transmissions - a comprehensive review. **Memórias do Instituto Oswaldo Cruz**, v. 110, n. 3, p. 277–282, 2015.
- 13 MANDAL, S. Epidemiological aspects of Chagas disease - a review. **Journal of Ancient Diseases and Preventive Remedies**, v. 02, n. 02, p. 1–7, 2014.

- 14 VIANNA MARTINS, A. et al. Biology of *Trypanosoma cruzi*: an update. **Infectio**, v. 16, n. 1, p. 45–58, 2012.
- 15 TYLER, K. M.; ENGMAN, D. M. The life cycle of *Trypanosoma cruzi*. **International Journal for Parasitology**, v. 31, n. 5–6, p. 472–481, 2001.
- 16 LENTINI, G.; PACHECO, N. S.; BURLEIGH, B. A. Targeting host mitochondria : a role for the *Trypanosoma cruzi* amastigote flagellum. **Cellular Microbiology**, v. 20, n. 2, p. 1–8, 2018.
- 17 SALTO, M. L. et al. Formation and remodeling of inositolphosphoceramides during differentiation of *Trypanosoma cruzi* from Trypomastigote to Amastigote. **Eucaryotic Cell**, v. 2, n. 4, p. 756–768, 2003.
- 18 CHATELAIN, E. Chagas disease research and development: is there light at the end of the tunnel? **Computational and Structural Biotechnology Journal**, v. 15, p. 98–103, 2017. doi: 10.1016/j.csbj.2016.12.002.
- 19 GEORG, I. et al. Evolution of anti-*Trypanosoma cruzi* antibody production in patients with chronic Chagas disease: correlation between antibody titers and development of cardiac disease severity. **PLoS Neglected Tropical Diseases**, v. 11, n. 7, p. 1–22, 2017.
- 20 BERN, C. Chagas' disease. **New England Journal of Medicine**, v. 373, n. 5, p. 456–466, 2015.
- 21 PÉREZ-MOLINA, J. A.; MOLINA, I. Chagas disease cardiomyopathy treatment remains a challenge – authors' reply. **Lancet**, v. 391, n. 10136, p. 2209–2210, 2018.
- 22 DE BONA, E. et al. Autoimmunity in chronic Chagas disease: a road of multiple pathways to cardiomyopathy? **Frontiers in Immunology**, v. 9, n. AUG, p. 1–8, 2018. doi: 10.3389/fimmu.2018.01842.
- 23 HERNÁNDEZ, C. et al. Chagas disease (*Trypanosoma cruzi*) and HIV co-infection in Colombia. **International Journal of Infectious Diseases**, v. 26, p. 146–148, 2014. doi: 10.1016/j.ijid.2014.04.002.
- 24 DIAS, J. C. P. et al. II Consenso brasileiro em doença de Chagas, 2015. **Epidemiologia e Serviços de Saúde**, v. 25, n. 21, p. 1–10, 2016. doi: 10.5123/S1679-49742016000500002.
- 25 CASTRO, J. A.; DE MECCA, M. M.; BARTEL, L. C. Toxic side effects of drugs used to treat Chagas' disease (American trypanosomiasis). **Human & Experimental Toxicology**, v. 25, n. 8, p. 471–479, 2006.
- 26 CERECETTO, H.; GONZÁLEZ, M. Antiparasitic prodrug nifurtimox: revisiting its activation mechanism. **Future Microbiology**, v. 6, n. 8, p. 513–514, 2011.

- 27 KRATZ, J. M. et al. Clinical and pharmacological profile of benznidazole for treatment of Chagas disease. **Expert Review of Clinical Pharmacology**, v. 11, n. 10, p. 943–957, 2018.
- 28 ALPERN, J. D.; LOPEZ-VELEZ, R.; STAUFFER, W. M. Access to benznidazole for Chagas disease in the United States - cautious optimism? **PLoS Neglected Tropical Diseases**, v. 11, n. 9, p. 10–14, 2017.
- 29 MULLARD, A. 2017 FDA drug approvals. **Nature Reviews Drug Discovery**, v. 17, n. 2, p. 81–85, 2018.
- 30 COURA, J. R.; DE CASTRO, S. L. A critical review on Chagas disease chemotherapy. **Memorias do Instituto Oswaldo Cruz**, v. 97, n. 1, p. 3–24, 2002.
- 31 DNDI. **Antiproteoal activity profiling of approved drugs**: a starting point toward drug repositioning. Disponível em: <<https://www.dndi.org/diseases-projects/open-innovation/antiproteoal-activity-profiling/>>. Acesso em: 24 ago. 2018.
- 32 TAVANO, O. L. et al. Biotechnological applications of proteases in food technology. **Comprehensive Reviews in Food Science and Food Safety**, v. 17, n. 2, p. 412–436, 2018.
- 33 FERREIRA, L. G.; OLIVA, G.; ANDRICOPULO, A. D. From medicinal chemistry to human health: current approaches to drug discovery for cancer and neglected tropical diseases. **Anais da Academia Brasileira de Ciências**, v. 90, p. 645–661, 2018. doi: 10.1590/0001-3765201820170505.
- 34 DUSCHAK, V. G.; COUTO, A. S. Cruzipain, the major cysteine protease of *Trypanosoma cruzi*: a sulfated glycoprotein antigen as relevant candidate for vaccine development and drug target . a review. **Current Medicinal Chemistry**, v. 16, n. 24, p. 3174–3202, 2009.
- 35 EAKIN, A. E. et al. The sequence, organization, and expression of the major cysteine protease (cruzain) from *Trypanosoma cruzi*. **Journal of Biological Chemistry**, v. 267, n. 11, p. 7411–7420, 1992.
- 36 LIMA, A. P. C. et al. Identification of new cysteine protease gene isoforms in *Trypanosoma cruzi*. **Molecular and Biochemical Parasitology**, v. 67, n. 2, p. 333–338, 1994.
- 37 BALANT, L. P. et al. Relevance of genetic polymorphism in drug metabolism in the development of new drugs. **European Journal of Clinical Pharmacology**, v. 36, n. 6, p. 551–554, 1989
- 38 STOKA, V. et al. The high stability of cruzipain against pH-induced inactivation is not dependent on its C-terminal domain. **FEBS Letters**, v. 469, n. 1, p. 29–32, 2000.
- 39 BONTEMPI, E.; CAZZULO, J. J. Digestion of human immunoglobulin G by the major cysteine proteinase (cruzipain) from *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 70, n. 3, p. 337–342, 1990.

- 40 SOUTO-PADRON, T. et al. Cysteine proteinase in *Trypanosoma cruzi*: immunocytochemical localization and involvement in parasite-host cell interaction. **Journal of Cell Science**, v. 96, n. 2, p. 485–490, 1990.
- 41 CAMPETELLA, O.; MARTINEZ, J. A major cysteine proteinase is developmentally regulated in *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 67, n. 1-2, p. 145–149, 1990.
- 42 TOMAS, A. M.; MILES, M. A.; KELLY, J. M. Overexpression of cruzipain, the major cysteine proteinase of *Trypanosoma cruzi*, is associated with enhanced metacyclogenesis. **European Journal of Biochemistry**, v. 244, n. 2, p. 596–603, 1997.
- 43 SANT'ANNA, C. et al. Subcellular proteomics of *Trypanosoma cruzi* reservosomes. **Proteomics**, v. 9, n. 7, p. 1782–1794, 2009.
- 44 SOARES, M. J.; SOUTO-PADRIN, T.; SOUZA, W. D. E. Identification of a large pre-lysosomal compartment in the pathogenic protozoon *Trypanosoma cruzi*. **Journal of Cell Science**, v. 102, Pt. 1, p. 157–167, 1992.
- 45 GARG, N. et al. Glycosylphosphatidylinositols are required for the development of *Trypanosoma cruzi* amastigotes. **Infection and immunity**, v. 65, n. 10, p. 4055–4060, 1997.
- 46 APARICIO, I. M.; SCHARFSTEIN, J.; LIMA, A. P. C. A. A new cruzipain-mediated pathway of human cell invasion by *Trypanosoma cruzi* requires trypomastigote membranes. **Infection and immunity**, v. 72, n. 10, p. 5892–5902, 2004.
- 47 ZHAI, X.; MEEK, T. D. Catalytic mechanism of cruzain from *Trypanosoma cruzi* as determined from solvent kinetic isotope effects of steady-state and pre-steady-state kinetics. **Biochemistry**, v. 57, n. 22, p. 3176–3190, 2018.
- 48 ARAFET, K.; FERRER, S.; MOLINER, V. Computational study of the catalytic mechanism of the cruzain cysteine protease. **ACS Catalysis**, v. 7, n. 2, p. 1207–1215, 2017.
- 49 MARTINEZ-MAYORGA, K. et al. Cruzain inhibitors : efforts made, current leads and a structural outlook of new hits. **Drug Discovery Today**, v. 20, n. 7, p. 890–898, 2015.
- 50 KERR, I. D. et al. Vinyl sulfones as antiparasitic agents and a structural basis for drug design. **Journal of Biological Chemistry**, v. 284, n. 38, p. 25697–25703, 2009.
- 51 DNDI. **K777 (Chagas)**. Disponível em: <<https://www.dndi.org/diseases-projects/portfolio/completed-projects/k777/>>. Acesso em: 27 nov. 2018.
- 52 ENGEL, J. C. et al. Cysteine protease inhibitors cure an experimental *Trypanosoma cruzi* infection. **Journal of Experimental Medicine**, v. 188, n. 4, p. 725–734, 1998.

- 53 JACOBSEN, W.; CHRISTIANS, U.; BENET, L. Z. In vitro evaluation of the disposition of a novel cysteine protease inhibitor. **Drug Metabolism and Disposition**, v. 28, n. 11, p. 1343–1351, 2000.
- 54 BARR, S. C. et al. A cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi* a cysteine protease inhibitor protects dogs from cardiac damage during infection by *Trypanosoma cruzi*. **Antimicrobial Agents and Chemotherapy**, v. 49, n. 12, p. 5160–5162, 2005.
- 55 CAZZULO, J. J.; STOKA, V.; TURK, V. The major cysteine proteinase of *Trypanosoma cruzi*: a valid target for chemotherapy of Chagas disease. **Current Pharmaceutical Design**, v. 7, n. 12, p. 1143–1156, 2001.
- 56 DUSCHAK, V. G. Recent patents on anti-infective drug discovery. **Recent Patents on Anti-Infective Drug Discovery**, v. 11, p. 74–173, 2016. doi: 10.2174/1574891X11666161024165.304.
- 57 ITOW, S.; CAMARGO, E. P. Proteolytic activities in cell extracts of *Trypanosoma cruzi*. **Journal of Protozoology**, v. 24, n. 4, p. 591–595, 1977.
- 58 CAMPETELLA, O. et al. A major cysteine proteinase is developmentally regulated in *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 67, n. 1–2, p. 145–149, 1990.
- 59 MURTA, A. C. M. et al. Structural and functional identification of GP57/51 antigen of *Trypanosoma cruzi* as a cysteine proteinase. **Molecular and Biochemical Parasitology**, v. 43, n. 1, p. 27–38, 1990.
- 60 FRESNO, M. et al. *Trypanosoma-cruzi* - identification of a membrane cysteine proteinase linked through a GPI anchor. **Brazilian Journal of Medical and Biological Research**, v. 27, n. 2, p. 431–437, 1994.
- 61 CAZZULO, J. J. et al. Some kinetic properties of a cysteine proteinase (cruzipain) from *Trypanosoma cruzi*. **Biochimica et Biophysica Acta (BBA)** - protein structure and molecular enzymology, v. 1037, n. 2, p. 186–191, 1990.
- 62 CHEN, Y. T. et al. In vitro and in vivo studies of the trypanocidal properties of WRR-483 against *Trypanosoma cruzi*. **PLoS Neglected Tropical Diseases**, v. 4, n. 9, p. 1–11, 2010.
- 63 JONES, B. D. et al. Synthesis and evaluation of oxyguanidine analogues of the cysteine protease inhibitor wr-483 against cruzain. **ACS Medicinal Chemistry Letters**, v. 7, n. 1, p. 77–82, 2016.
- 64 LIAO, C. et al. Software and resources for computational medicinal chemistry. **Future Medicinal Chemistry**, v. 3, n. 8, p. 1057–1085, 2011.
- 65 SONG, C. M.; LIM, S. J.; TONG, J. C. Recent advances in computer-aided drug design. **Briefings in Bioinformatics**, v. 10, n. 5, p. 579–591, 2009.

- 66 ANDRICOPULO, A.; SALUM, L.; ABRAHAM, D. Structure-based drug design strategies in medicinal chemistry. **Current Topics in Medicinal Chemistry**, v. 9, n. 9, p. 771–790, 2009.
- 67 PLEWCZYNSKI, D.; SPIESER, S. A. H.; KOCH, U. Performance of machine learning methods for ligand-based virtual screening. **Combinatorial Chemistry and High Throughput Screening**, v. 12, n. 4, p. 358–368, 2009.
- 68 ZHANG, W. et al. When drug discovery meets web search: learning to rank for ligand-based virtual screening. **Journal of Cheminformatics**, v. 7, n. 1, p. 1–13, 2015.
- 69 ŚLEDŹ, P.; CAFLISCH, A. Protein structure-based drug design: from docking to molecular dynamics. **Current Opinion in Structural Biology**, v. 48, p. 93–102, 2018. doi.org/10.1016/j.sbi.2017.10.010.
- 70 TERFLOTH, L.; SPYCHER, S.; GASTEIGER, J. **Drug discovery**: an overview. Deutschland: Wiley-VCH Verlag GmbH and Co. KGaA, 2018.
- 71 LIONTA, E. et al. Structure-based virtual screening for drug discovery: principles, applications and recent advances. **Current Topics in Medicinal Chemistry**, v. 14, n. 16, p. 1923–1938, 2014.
- 72 CHAYAN ACHARYA et al. Recent advances in ligand-based drug design: relevance and utility of the conformationally sampled pharmacophore approach. **Current Computational-Aided Drug Design**, v. 7, n. 1, p. 10–22, 2011.
- 73 ANDRICOPULO, A. D. Structure- and ligand-based drug design: advances and perspectives. **Current Topics in Medicinal Chemistry**, v. 9, n. 9, p. 754, 2009.
- 74 HUGHES, J. P. et al. Principles of early drug discovery. **British Journal of Pharmacology**, v. 162, n. 6, p. 1239–1249, 2011.
- 75 LINDNER, J. R.; LINK, J. Molecular imaging in drug discovery and development. Circulation. **Cardiovascular Imaging**, v. 11, n. 2, p. 1–11, 2018.
- 76 REKER, D.; SCHNEIDER, G. Active-learning strategies in computer-assisted drug discovery. **Drug Discovery Today**, v. 20, n. 4, p. 458–465, 2015.
- 77 KOEPPEN, H. et al. Ligand-based virtual screening. In: SOTRIFTER, C. (Ed.) **Virtual screening**: principles, challenges, and practical guidelines. New York: Wiley, 2011. p. 61–85.
- 78 RIPPHAUSEN, P.; NISIUS, B.; BAJORATH, J. State-of-the-art in ligand-based virtual screening. **Drug Discovery Today**, v. 16, n. 9–10, p. 372–376, 2011.
- 79 BAIG, M. H. et al. Computer aided drug design : success and limitations structure of target protein propose new lead or optimize existing lead. **Current Pharmaceutical Design**, v. 22, n. 5, p. 572–581, 2016.

- 80 KIM, S. et al. PubChem substance and compound databases. **Nucleic Acids Research**, v. 44, p. D1202–D1213, 2016. doi: 10.1093/nar/gkv951.
- 81 WISHART, D. S. et al. DrugBank 5.0: a major update to the DrugBank database for 2018. **Nucleic Acids Research**, v. 46, n. D1, p. D1074–D1082, 2018.
- 82 STERLING, T.; IRWIN, J. J. ZINC 15 - Ligand discovery for everyone. **Journal of Chemical Information and Modeling**, v. 55, n. 11, p. 2324–2337, 2015.
- 83 IRWIN, J. J. Using ZINC to acquire a virtual screening library. **Current Protocols in Bioinformatics**, v. 22, p. 1–23, 2008. doi: 10.1002/0471250953.bi1406s22.
- 84 IRWIN, J. J.; SHOICHET, B. K. ZINC – a free database of commercially available compounds for virtual screening. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 177–182, 2005.
- 85 IRWIN, J. J. et al. ZINC: a free tool to discover chemistry for biology. **Journal of Chemical Information and Modeling**, v. 52, n. 7, p. 1757–1768, 2012.
- 86 TEAGUE, S. J. et al. The design of leadlike combinatorial libraries. **Angewandte Chemie International Edition**, v. 38, n. 24, p. 3743–3748, 1999.
- 87 ZINC12. **Physical property subsets**. Disponível em: <<http://zinc.docking.org/browse/subsets/>>. Acesso em: 5 nov. 2018.
- 88 LIPINSKI, C. A. Lead- and drug-like compounds: the rule-of-five revolution. **Drug Discovery Today: technologies**, v. 1, n. 4, p. 337–341, 2004.
- 89 CARR, R. A. E. et al. Fragment-based lead discovery: leads by design. **Drug Discovery Today**, v. 10, n. 14, p. 987–992, 2005.
- 90 LIPINSKI, C. A. Rule of five in 2015 and beyond: target and ligand structural limitations, ligand chemistry structure and drug discovery project decisions. **Advanced Drug Delivery Reviews**, v. 101, p. 34–41, 2016. doi: 10.1016/j.addr.2016.04.029.
- 91 _____. Drug-like properties and the causes of poor solubility and poor permeability. **Journal of Pharmacological and Toxicological Methods**, v. 44, n. 1, p. 235–249, 2000.
- 92 POLINSKY, A. Lead-likeness and drug-likeness. **Practice of Medicinal Chemistry**, p. 244–254, 2008. doi: 10.1016/B978-0-12-374194-3.00012-3.
- 93 KUMAR, V.; KRISHNA, S.; SIDDIQI, M. I. Virtual screening strategies: recent advances in the identification and design of anti-cancer agents. **Methods**, v. 71, p. 64–70, 2015. doi: 10.1016/j.ymeth.2014.08.010.
- 94 HOLLIDAY, J. D.; SANI, N.; WILLETT, P. Ligand – based virtual screening using a genetic algorithm with data fusion. **Match Communications in Mathematical and in Computer Chemistry**, v. 80, p. 623–638, 2018.

- 95 BERENGER, F.; VU, O.; MEILER, J. Consensus queries in ligand-based virtual screening experiments. **Journal of Cheminformatics**, v. 9, n. 1, p. 1–13, 2017.
- 96 FERREIRA, L. G. et al. Molecular docking and structure-based drug design strategies. **Molecules**, v. 20, n. 7, p. 13384–13421, 2015.
- 97 PRATHIPATI, P.; DIXIT, A.; SAXENA, A. Computer-aided drug design: integration of structure-based and ligand-based approaches in drug design. **Current Computer Aided-Drug Design**, v. 3, n. 2, p. 133–148, 2007.
- 98 MALTAROLLO, V. G. et al. Applying machine learning techniques for ADME-Tox prediction: a review. **Expert Opinion on Drug Metabolism and Toxicology**, v. 11, n. 2, p. 259–271, 2015.
- 99 WANG, J.; HOU, T. Advances in computationally modeling human oral bioavailability. **Advanced Drug Delivery Reviews**, v. 86, n. 1, p. 11–16, 2015.
- 100 MOROY, G. et al. Toward in silico structure-based ADMET prediction in drug discovery. **Drug Discovery Today**, v. 17, n. 1–2, p. 44–55, 2012.
- 101 VOLPE, D. A.; BALIMANE, P. V. Application of in vitro CYP and transporter assays to predict clinical drug – drug interactions. **Bioanalysis**, v. 10, n. 9, p. 619–623, 2018.
- 102 GANESAN, A.; COOTE, M. L.; BARAKAT, K. Molecular dynamics-driven drug discovery: leaping forward with confidence. **Drug Discovery Today**, v. 22, n. 2, p. 249–269, 2017.
- 103 MORTIER, J. et al. The impact of molecular dynamics on drug design: Applications for the characterization of ligand-macromolecule complexes. **Drug Discovery Today**, v. 20, n. 6, p. 686–702, 2015.
- 104 ZHAO, H.; CAFLISCH, A. Molecular dynamics in drug design. **European Journal of Medicinal Chemistry**, v. 91, p. 4–14, 2015. doi.org/10.1016/j.ejmech.2014.08.004.
- 105 BROOKS, B. R. et al. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. **Journal of Computational Chemistry**, v. 4, n. 2, p. 187–217, 1983.
- 106 BAYLY, C. I. et al. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. **Journal of the American Chemical Society**, v. 117, n. 19, p. 5179–5197, 1995.
- 107 JORGENSEN, W. L.; TIRADO-RIVES, J. The OPLS potential functions for proteins. energy minimizations for crystals of cyclic peptides and crambin. **Journal of the American Chemical Society**, v. 110, n. 6, p. 1657–1666, 1988.

- 108 SCOTT, W. R. P. et al. The GROMOS biomolecular simulation program package. **Journal of Physical Chemistry A**, v. 103, n. 19, p. 3596–3607, 1999.
- 109 NERENBERG, P. S.; HEAD-GORDON, T. New developments in force fields for biomolecular simulations. **Current Opinion in Structural Biology**, v. 49, p. 129–138, 2018. doi: 10.1016/j.sbi.2018.02.002.
- 110 LIU, X. et al. Molecular dynamics simulations and novel drug discovery. **Expert Opinion on Drug Discovery**, v. 13, n. 1, p. 23–37, 2018.
- 111 DURRANT, J. D.; MCCAMMON, J. A. Molecular dynamics simulations and drug discovery. **BMC Biology**, v. 9, n. 1, p. 23–37, 2011.
- 112 HOELZ, L. V. B. et al. Molecular dynamics simulations of the free and inhibitor-bound cruzain systems in aqueous solvent: insights on the inhibition mechanism in acidic pH. **Journal of Biomolecular Structure and Dynamics**, v. 34, n. 9, p. 1969–1978, 2016.
- 113 MARTINS, L. C. et al. Investigation of the binding mode of a novel cruzain inhibitor by docking, molecular dynamics, ab initio and MM/PBSA calculations. **Journal of Computer-Aided Molecular Design**, v. 32, n. 5, p. 591–605, 2018.
- 114 SARAIVA, Á. P. B. et al. Molecular description of α -keto-based inhibitors of cruzain with activity against Chagas disease combining 3D-QSAR studies and molecular dynamics. **Chemical Biology and Drug Design**, v. 92, n. 2, p. 1475–1487, 2018.
- 115 GÓMEZ-AYALA, S. et al. Synthesis, structural elucidation and in vitro antiparasitic activity against *Trypanosoma cruzi* and *Leishmania chagasi* parasites of novel tetrahydro-1-benzazepine derivatives. **Bioorganic and Medicinal Chemistry**, v. 18, n. 13, p. 4721–4739, 2010.
- 116 KING-KELLER, S. et al. Chemical validation of phosphodiesterase C as a chemotherapeutic target in *Trypanosoma cruzi*, the etiological agent of Chagas' disease. **Antimicrobial Agents and Chemotherapy**, v. 54, n. 9, p. 3738–3745, 2010.
- 117 MAGARACI, F. et al. Azasterols as Inhibitors of Sterol 24-Methyltransferase in *Leishmania* Species and *Trypanosoma cruzi*. **Journal of Medicinal Chemistry**, v. 46, n. 22, p. 4714–4727, 2003.
- 118 JONES, S. M. et al. Analogues of thiolactomycin as potential antimalarial agents. **Journal of Medicinal Chemistry**, v. 48, n. 19, p. 5932–5941, 2005.
- 119 ZUCCOTTO, F. et al. The structure-based design and synthesis of selective inhibitors of *Trypanosoma cruzi* dihydrofolate reductase. **Bioorganic and Medicinal Chemistry Letters**, v. 9, n. 10, p. 1463–1468, 1999.

120 JONCKERS, T. H. M. et al. Synthesis, cytotoxicity, and antiplasmodial and antitrypanosomal activity of new neocryptolepine derivatives. **Journal of Medicinal Chemistry**, v. 45, n. 16, p. 3497–3508, 2002.

121 SZAJNMAN, S. H. et al. Bisphosphonates derived from fatty acids are potent inhibitors of *Trypanosoma cruzi* farnesyl pyrophosphate synthase. **Bioorganic and Medicinal Chemistry Letters**, v. 13, n. 19, p. 3231–3235, 2003.

122 DA ROSA, R. et al. Design and synthesis of a new series of 3,5-disubstituted isoxazoles active against *Trypanosoma cruzi* and *Leishmania amazonensis*. **European Journal of Medicinal Chemistry**, v. 128, p. 25–35, 2017. doi: 10.1016/j.ejmech.2017.01.029.

123 DE AZEREDO, C. M. O. et al. Biological activity of the azlactone derivative EPA-35 against *Trypanosoma cruzi*. **FEMS Microbiology Letters**, v. 364, n. 4, p. 1–7, 2017.

124 GUERRA, A. et al. Novel Imidazo[4,5-c][1,2,6]thiadiazine 2,2-dioxides as antiproliferative *trypanosoma cruzi* drugs: computational screening from neural network, synthesis and in vivo biological properties. **European Journal of Medicinal Chemistry**, v. 136, p. 223–234, 2017. doi: 10.1016/j.ejmech.2017.04.075.

125 SYKES, M. L.; AVERY, V. M. Development and application of a sensitive, phenotypic, high-throughput image-based assay to identify compound activity against *Trypanosoma cruzi* amastigotes. **International Journal for Parasitology: drugs and drug resistance**, v. 5, n. 3, p. 215–228, 2015.

126 DE MENEZES, D. R. et al. Hydroxamic acid derivatives: a promising scaffold for rational compound optimization in Chagas disease. **Journal of Enzyme Inhibition and Medicinal Chemistry**, v. 31, n. 6, p. 964–973, 2016.

127 OLMO, F. et al. An in vitro iron superoxide dismutase inhibitor decreases the parasitemia levels of *Trypanosoma cruzi* in BALB/c mouse model during acute phase. **International Journal for Parasitology: drugs and drug resistance**, v. 5, n. 3, p. 110–116, 2015.

128 PAPADOPOULOU, M. V. et al. 3-Nitrotriazole-based piperazides as potent antitrypanosomal agents. **European Journal of Medicinal Chemistry**, v. 103, p. 325–334, 2015. doi: 10.1016/j.ejmech.2015.08.042.

129 _____. Discovery of potent nitrotriazole-based antitrypanosomal agents: in vitro and in vivo evaluation. **Bioorganic and Medicinal Chemistry**, v. 23, n. 19, p. 6467–6476, 2015.

130 NEITZ, R. J. et al. Tetrafluorophenoxymethyl ketone cruzain inhibitors with improved pharmacokinetic properties as therapeutic leads for Chagas' disease. **Bioorganic and Medicinal Chemistry Letters**, v. 25, n. 21, p. 4834–4837, 2015.

- 131 SANGENITO, L. S. et al. Nelfinavir and lopinavir impair *Trypanosoma cruzi* trypomastigote infection in mammalian host cells and show anti-amastigote activity. **International Journal of Antimicrobial Agents**, v. 48, n. 6, p. 703–711, 2016.
- 132 SANTOS, G. B. et al. Semisynthesis of new aphidicolin derivatives with high activity against *Trypanosoma cruzi*. **Bioorganic and Medicinal Chemistry Letters**, v. 26, n. 4, p. 1205–1208, 2016.
- 133 DE VITA, D. et al. In vitro screening of 2-(1H-imidazol-1-yl)-1-phenylethanol derivatives as antiprotozoal agents and docking studies on *Trypanosoma cruzi* CYP51. **European Journal of Medicinal Chemistry**, v. 113, p. 28–33, 2016. doi: 10.1016/j.ejmech.2016.02.028.
- 134 EBERLE, C. et al. Synthesis, inhibition potency, binding mode, and antiprotozoal activities of fluorescent inhibitors of trypanothione reductase based on mepacrine-conjugated diaryl sulfide scaffolds. **ChemMedChem**, v. 4, n. 12, p. 2034–2044, 2009.
- 135 PALMA, A. et al. Synthesis and in vitro activity of new tetrahydronaphtho[1,2-b]azepine derivatives against *Trypanosoma cruzi* and *Leishmania chagasi* parasites. **Bioorganic and Medicinal Chemistry Letters**, v. 19, n. 8, p. 2360–2363, 2009.
- 136 HERRERA, C. et al. In vitro activity of thienyl-2-nitropropene compounds against *Trypanosoma cruzi*. **Memorias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 980–985, 2009.
- 137 ROSSO, V. S. et al. Synthesis and biological evaluation of new 2-alkylaminoethyl-1,1-bisphosphonic acids against *Trypanosoma cruzi* and *Toxoplasma gondii* targeting farnesyl diphosphate synthase. **Bioorganic and Medicinal Chemistry**, v. 19, n. 7, p. 2211–2217, 2011.
- 138 EBERLE, C. et al. Improved inhibitors of trypanothione reductase by combination of motifs: synthesis, inhibitory potency, binding mode, and antiprotozoal activities. **ChemMedChem**, v. 6, n. 2, p. 292–301, 2011.
- 139 SEALEY-CARDONA, M. et al. Kinetic characterization of squalene synthase from *Trypanosoma cruzi*: selective inhibition by quinuclidine derivatives. **Antimicrobial Agents and Chemotherapy**, v. 51, n. 6, p. 2123–2129, 2007.
- 140 FRANCK, X. et al. Biological evaluation of substituted quinolines. **Bioorganic and Medicinal Chemistry Letters**, v. 14, n. 14, p. 3635–3638, 2004.
- 141 KHABNADIDEH, S. et al. Design, synthesis and evaluation of 2,4-diaminoquinazolines as inhibitors of trypanosomal and leishmanial dihydrofolate reductase. **Bioorganic and Medicinal Chemistry**, v. 13, n. 7, p. 2637–2649, 2005.
- 142 SZAJNMAN, S. H. et al. Synthesis and biological evaluation of 1-amino-1,1-bisphosphonates derived from fatty acids against *Trypanosoma cruzi* targeting farnesyl pyrophosphate synthase. **Bioorganic and Medicinal Chemistry Letters**, v. 15, n. 21, p. 4685–4690, 2005.

143 _____. Synthesis and biological evaluation of 2-alkylaminoethyl-1,1-bisphosphonic acids against *Trypanosoma cruzi* and *Toxoplasma gondii* targeting farnesyl diphosphate synthase. **Bioorganic and Medicinal Chemistry**, v. 16, n. 6, p. 3283–3290, 2008.

144 BRINGMANN, G. et al. Synthesis and antiprotozoal activities of simplified analogs of naphthylisoquinoline alkaloids. **European Journal of Medicinal Chemistry**, v. 43, n. 1, p. 32–42, 2008.

145 BLANCO, M. C. et al. Synthesis of novel polysubstituted (2SR,4RS)-2-heteroaryltetrahydro-1,4-epoxy-1-benzazepines and cis-2-heteroaryl-4-hydroxytetrahydro-1H-1-benzazepines as antiparasitic agents. **European Journal of Medicinal Chemistry**, v. 86, p. 291–309, 2014. doi: 10.1016/j.ejmech.2014.08.055.

146 OLMO, F. et al. Synthetic single and double aza-scorpian macrocycles act as inhibitors of the antioxidant enzymes iron superoxide dismutase and trypanothione reductase in *Trypanosoma cruzi* with promising results in a murine model. **RSC Advances**, v. 4, n. 110, p. 65108–65120, 2014.

147 BRAGA, S. F. P. et al. Synthesis and evaluation of the antiparasitic activity of bis-(arylmethylidene) cycloalkanones. **European Journal of Medicinal Chemistry**, v. 71, p. 282–289, 2014. doi: 10.1016/j.ejmech.2013.11.011.

148 PAPADOPOULOU, M. V. et al. Novel 3-nitrotriazole-based amides and carbinols as bifunctional antichagasic agents. **Journal of Medicinal Chemistry**, v. 58, n. 3, p. 1307–1319, 2015.

149 KEENAN, M. et al. Design, structure-activity relationship and in vivo efficacy of piperazine analogues of fenarimol as inhibitors of *Trypanosoma cruzi*. **Bioorganic and Medicinal Chemistry**, v. 21, n. 7, p. 1756–1763, 2013.

150 CARVALHO, S. A. et al. Design and synthesis of new (E)-cinnamic N-acylhydrazones as potent antitrypanosomal agents. **European Journal of Medicinal Chemistry**, v. 54, p. 512–521, 2012. doi.org/10.1016/j.ejmech.2012.05.041.

151 GALIANA-ROSELLÓ, C. et al. In vitro and in vivo antileishmanial and trypanocidal studies of new N-benzene- and N-naphthalenesulfonamide derivatives. **Journal of Medicinal Chemistry**, v. 56, n. 22, p. 8984–98, 2013.

152 UPADHAYAYA, R. S. et al. New antiprotozoal agents: Their synthesis and biological evaluations. **Bioorganic and Medicinal Chemistry Letters**, v. 23, n. 9, p. 2750–2758, 2013.

153 SILVA JÚNIOR, E. F. et al. Design, synthesis, molecular docking and biological evaluation of thiophen-2-iminothiazolidine derivatives for use against *Trypanosoma cruzi*. **Bioorganic and Medicinal Chemistry**, v. 24, n. 18, p. 4228–4240, 2016.

154 CANVAS. **Schrödinger release 2016-3**. 2016. Disponível em: <<https://www.schrodinger.com/canvas>>. Acesso em: 20 nov. 2018

155 RODRÍGUEZ, J. et al. New potent 5-nitroindazole derivatives as inhibitors of *Trypanosoma cruzi* growth: synthesis, biological evaluation, and mechanism of action studies. **Bioorganic and Medicinal Chemistry**, v. 17, n. 24, p. 8186–8196, 2009.

156 SYBYL-X. 2013. Disponível em: <<https://www.certara.com/pressreleases/certara-enhances-sybyl-x-drug-design-and-discovery-software-suite/>>. Acesso em: 20 nov. 2018

157 LIEN, E. J.; WANG, P. H. Lipophilicity, molecular weight, and drug action: reexamination of parabolic and bilinear models. **Journal of Pharmaceutical Sciences**, v. 69, n. 6, p. 1979–1981, 1980.

158 HALL, M. et al. The WEKA data mining software: an update. **SIGKDD Explorations**, v. 11, n. 1, p. 10–18, 2009.

159 SCHMIDHUBER, J. Deep learning in neural networks: an overview. **Neural Networks**, v. 61, p. 85–117, 2015. doi.org/10.1016/j.neunet.2014.09.003.

160 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *Nature*, v. 521, p. 436–444, 2015. doi.org/10.1038/nature14539.

161 CYBENKO, G. Approximation by superpositions of a sigmoidal function. **Mathematics of Control, Signals, and Systems**, v. 2, n. 4, p. 303–314, 1989.

162 RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. **Nature**, v. 323, p. 533–536, 1986.

163 CERETO-MASSAGUÉ, A. et al. Molecular fingerprint similarity search in virtual screening. **Methods**, v. 71, n. C, p. 58–63, 2015.

164 DUAN, J. et al. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. **Journal of Molecular Graphics and Modelling**, v. 29, n. 2, p. 157–170, 2010.

165 AN, Y.; SHERMAN, W.; DIXON, S. L. Kernel-based partial least squares: Application to fingerprint-based QSAR with model visualization. **Journal of Chemical Information and Modeling**, v. 53, n. 9, p. 2312–2321, 2013.

166 DIXON, S. L. et al. AutoQSAR: an automated machine learning tool for best-practice quantitative structure–activity relationship modeling. **Future Medicinal Chemistry**, v. 8, n. 15, p. 1825–1839, 2016.

167 CHEN, C.; LIAW, A.; BREIMAN, L. **Using Random Forest to learn imbalanced data**. 2004. Disponível em: <<https://statistics.berkeley.edu/sites/default/files/tech-reports/666.pdf>>. Acesso em: 20 nov. 2018

168 AWALE, M.; JIN, X.; REYMOND, J. L. Stereoselective virtual screening of the ZINC database using atom pair 3D-fingerprints. **Journal of Cheminformatics**, v. 7, n. 1, p. 1–15, 2015.

169 LIGPREP. **Schrödinger release 2016-3.** 2016. Disponível em: <<https://www.schrodinger.com/ligprep>>. Acesso em: 23 nov. 2018

170 SHELLEY, J. C. et al. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. **Journal of Computer-Aided Molecular Design**, v. 21, n. 12, p. 681–691, 2007.

171 HARDER, E. et al. OPLS3: a force field providing broad coverage of drug-like small molecules and proteins. **Journal of Chemical Theory and Computation**, v. 12, n. 1, p. 281–296, 2016.

172 FERREIRA, R. S. et al. Complementarity between a docking and a high-throughput screen in discovering new cruzain inhibitors. **Journal of Medicinal Chemistry**, v. 53, n. 13, p. 4891–4905, 2010.

173 PROTEIN PREPARATION WIZARD. **Schrödinger release 2016-3.** 2016. Disponível em: <<https://www.schrodinger.com/protein-preparation-wizard>>. Acesso em: 23 nov. 2018

174 PRIME. **Schrödinger release 2016-3.** 2016. Disponível em: <<https://www.schrodinger.com/prime>>. Acesso em: 23 nov. 2018

175 ROSTKOWSKI, M. et al. Graphical analysis of pH-dependent properties of proteins predicted using PROPKA. **BMC Structural Biology**, v. 11, 2011. doi: 10.1186/1472-6807-11-6.

176 FRIESNER, R. A. et al. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. **Journal of Medicinal Chemistry**, v. 49, n. 21, p. 6177–6196, 2006.

177 GLIDE. **Glide manual.** 2016. Disponível em: <http://gohom.win/ManualHom/Schrodinger/Schrodinger_2015-2_docs/glide/glide_user_manual.pdf>. Acesso em: 20 nov. 2018.

178 GOLD. **Gold manual.** 2017. Disponível em: <<https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/gold.pdf>>. Acesso em: 20 nov. 2018.

179 VERDONK, M. L. et al. Improved protein-ligand docking using GOLD. **Proteins: structure, function and genetics**, v. 52, n. 4, p. 609–623, 2003.

180 WARING, M. J. Lipophilicity in drug discovery. **Expert Opinion on Drug Discovery**, v. 5, n. 3, p. 235–248, 2010.

181 LEUNG, S. S. F. et al. Testing physical models of passive membrane permeation. **Journal of Chemical Information and Modeling**, v. 52, n. 6, p. 1621–1636, 2012.

182 QIKPROP. **QikProp manual.** 2016. Disponível em: <<https://www.schrodinger.com/qikprop>>. Acesso em: 20 nov. 2018

183 ZANGER, U. M.; SCHWAB, M. Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation. **Pharmacology and Therapeutics**, v. 138, n. 1, p. 103–141, 2013.

184 LYNCH, T.; PRICE, A. The effect of cytochrome p450 metabolism on drug response, interactions, and adverse effects. **American Family Physicians**, v. 76, n. 3, p. 1–6, 2007.

185 JIN, S. E.; HA, H.; SHIN, H. K. Effects of traditional herbal formulae on human CYP450 isozymes. **Chinese Journal of Integrative Medicine**, v. 23, n. 1, p. 62–69, 2017.

186 HASHIZUME, T. et al. N -dealkylation and hydroxylation of Ebastine by human liver cytochrome P450 abstract. **Drug Metabolism & Disposition**, v. 26, n. 6, p. 566–571, 1998.

187 SEVRIUKOVA, I. F.; POULOS, T. L. Understanding the mechanism of cytochrome P450 3A4: recent advances and remaining problems. **Dalton Transactions**, v. 42, n. 9, p. 3116–3126, 2014.

188 _____. Current approaches for investigating and predicting cytochrome P450 3A4-ligand interactions. **Advances in Experimental Medicine and Biology**, v. 851, p. 83–105, 2015. doi: 10.1007/978-3-319-16009-2_3.

189 P450 SITE OF METABOLISM. **P450 site of metabolism manual**. 2016. Disponível em: <http://gohom.win/ManualHom/Schrodinger/Schrodinger_2012_docs/cyp450som/cyp450som_user_manual.pdf>. Acesso em: 24 nov. 2018.

190 HANSCH, C.; LEO, A.; TAFT, R. W. A survey of hammett substituent constants and resonance and field parameters. **Chemical Reviews**, v. 91, n. 2, p. 165–195, 1991.

191 MCKERROW, J. H. et al. Two approaches to discovering and developing new drugs for Chagas disease. **Memorias do Instituto Oswaldo Cruz**, v. 104, p. 263–269, 2009. doi: 10.1590/S0074-02762009000900034.

192 LEE, G. M. et al. Mapping inhibitor binding modes on an active cysteine protease via nuclear magnetic resonance spectroscopy. **Biochemistry**, v. 51, n. 50, p. 10087–10098, 2012.

193 ABRAHAM, M. J. et al. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. **SoftwareX**, v. 1–2, n. 1, p. 19–25, 2015.

194 KAMINSKI, G. A. et al. Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides. **Journal of Physical Chemistry B**, v. 105, n. 28, p. 6474–6487, 2001.

195 DODDA, L. S. et al. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. **Nucleic Acids Research**, v. 45, n. W1, p. W331–W336, 2017.

196 _____. 1.14*CM1A-LBCC: localized bond-charge corrected CM1A charges for condensed-phase simulations. **Journal of Physical Chemistry B**, v. 121, n. 15, p. 3864–3870, 2017.

197 JORGENSEN, W. L. et al. Strengthening of traditional paddy seed selection practices of tribal farm families with improved knowledge and skills in Koraput district, Odisha. **Journal of Chemical Physics**, v. 79, n. 2, p. 926–935, 1983.

198 NOSÉ, S. A unified formulation of the constant temperature molecular dynamics methods. **Journal of Chemical Physics**, v. 81, n. 1, p. 511–519, 1984.

199 HOOVER, W. G.; HOLIAN, B. L. Kinetic moments method for the canonical ensemble distribution. **Physics Letters A**, v. 211, n. 5, p. 253–257, 1996.

200 PARRINELLO, M.; RAHMAN, A. Crystal structure and pair potentials: A molecular-dynamics study. **Physical Review Letters**, v. 45, n. 14, p. 1196–1199, 1980.

201 TOUKMAJI, A. et al. Efficient particle-mesh Ewald based approach to fixed and induced dipolar interactions. **Journal of Chemical Physics**, v. 113, n. 24, p. 10913–10927, 2000.

202 ESSMANN, U. et al. A smooth particle mesh Ewald method. **Journal of Chemical Physics**, v. 103, n. 19, p. 8577–8593, 1995.

203 HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, 1996.

204 RASHMI KUMARI et al. g-mmpbsa -a GROMACS tool for high-throughput MM-PBSA calculations. **Journal of Chemical Information and Modeling**, v. 54, n. 7, p. 1951–1962, 2014.

205 HOMEYER, N.; GOHLKE, H. Free energy calculations by the molecular mechanics Poisson-Boltzmann surface area method. **Molecular Informatics**, v. 31, n. 2, p. 114–122, 2012.

206 FERREIRA, R. S. et al. Divergent modes of enzyme inhibition in a homologous structure - activity series. **Journal of Medicinal Chemistry**, v. 52, n. 16, p. 5005–5008, 2009.

207 YUAN, Z.; ZHAO, J.; WANG, Z.-X. Flexibility analysis of enzyme active sites by crystallographic temperature factors. **Protein Engineering, Design and Selection**, v. 16, n. 2, p. 109–114, 2003.

ANEXOS

ANEXO A – Representação do conjunto 1 (pIC₅₀)

Representação na forma SMILES	Estrutura	pIC ₅₀
<chem>c1c(Br)ccc(c12)NC(CC(C2)O)c3ccccc3</chem>	1	4,01
<chem>Cc1cc(ccc1)C(CC(C2)O)Nc(c23)cccc3</chem>	2	4,01
<chem>c1ccccc1-c2cc(on2)-c3ccc(N)cc3</chem>	3	4,04
<chem>c1ccc(OC)cc1C2CC(O3)Cc4ccc(c5c4C23)cccc5</chem>	4	4,05
<chem>c1ccccc1\C=C\C(=O)N\N=C\c2ccc([N+][O-])=O)cc2</chem>	5*	4,09
<chem>O=P(O)(O)C(N)(P(=O)(O)O)CCCCC</chem>	6	4,11
<chem>CCCCCCCCCCC1(C)C(O)=C(C(=O)S1)CCCC</chem>	7	4,11
<chem>[O-][N+](=O)c1ccc(cc1)S(=O)(=O)Nc2cncn2</chem>	8	4,12
<chem>C1CN(C)CCN1Cc2cnc([nH]2)-c3c(ccc(Br)c3)Sc4ccccc4</chem>	9	4,12
<chem>c1ccccc1CN(C2)C(C)(C)Cc(c23)cc(O)cc3OC</chem>	10	4,12
<chem>c1ccccc1/C=C2/C(=O)\C(CC2)=C\c3ccccc3</chem>	11*	4,14
<chem>c1ccccc1\C=C\C(=O)N\N=C\c2cc(OC)c(O)cc2</chem>	12	4,14
<chem>O=P(O)(O)C(N)(P(=O)(O)O)CCCCC</chem>	13	4,14
<chem>CCCCCCCCCCC1(CC)C(O)=C(C)C(=O)S1</chem>	14*	4,14
<chem>c1ccccc1\C=C\C(=O)N\N=C\c2cc(OC)c(OC)c(c2)OC</chem>	15	4,15
<chem>c1cncc(c12)cccc2/C=C3/C(=O)\C(CCC3)=C\c4cccc(c45)cncc5</chem>	16	4,16
<chem>Clc1cccc(Cl)c1C(CC(C2)O)Nc(c23)cccc3</chem>	17	4,16
<chem>CC(C)=CCC/C(C)=C/CC1(C)C(O)=C(C)C(=O)S1</chem>	18	4,17
<chem>CCCCCCCCCCC1(C)C(O)=C(C)C(=O)S1</chem>	19	4,17
<chem>CCCC(=O)c1cc(c(cc1)OC)OC2CCOC2</chem>	20*	4,17
<chem>ONC(=O)C(C1)ON=C1c2ccc(cc2)OCC</chem>	21	4,18
<chem>O=P(O)(O)C(O)(P(=O)(O)O)CCCCCCC</chem>	22*	4,18
<chem>c1ccc(OC)cc1N(CC2)CCN2CC(O)COC(C)(c3c4cccc3)c(c45)c(nc6c5c c(Br)cc6)N(CC7)CCN7C(c8ccccc8)c9ccccc9</chem>	23	4,19
<chem>COc(cc1)ccc1-c2ccn(n2)C(CO)c3c(ccc(F)c3)Oc(nc(c45)ccc(c5)[N+][O-])=O)c(c4)Cc6ccccc6</chem>	24	4,19
<chem>CCCCCCCCCCC1(C)C(O)=C(C(=O)S1)Cc2ccccc2</chem>	25	4,19
<chem>CCC(=O)N(C1)CCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5</chem>	26	4,19
<chem>CC1=NC(C)(C)Cc(c12)c(Br)c(OC)cc2OC</chem>	27	4,2
<chem>c1ccccc1\C=C\C(=O)N\N=C\c(c2)ccc(c23)OCO3</chem>	28	4,21
<chem>CCOC(=O)CC(C(=O)OCC)NC(=O)c1ccc(cc1)N(CCCC(=O)OCC)Cc(n2)cnc(c23)nc(N)nc3N</chem>	29*	4,21
<chem>c1cc(Br)ccc1C(CC23)N(O3)c4c(C2)cc(Cl)cc4</chem>	30	4,22
<chem>n1ccccc1Sc2cccc(c2CNC(=O)CN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC</chem>	31	4,24
<chem>CCCCCCCCCCC1(C)C(O)=C(C(=O)S1)CCC</chem>	32	4,25
<chem>c1ccccc1C(CC23)N(O3)c4c(C2)cc(Br)cc4</chem>	33	4,25
<chem>C1CN(C)CCN1Cc2cnc([nH]2)-c3c(Nc4ccccc4)cccc3Sc5ccccc5</chem>	34*	4,27
<chem>CCCCCCCCCCC1(C)C(O)=C(C(=O)S1)CC</chem>	35	4,27
<chem>O=P(O)(O)C(O)(P(=O)(O)O)CCCCCCCCCCCCC</chem>	36*	4,27
<chem>CCOC(=O)CC(C(=O)OCC)NC(=O)c1ccc(cc1)N(CCCCCCCC)Cc(n2)cnc(c23)nc(N)nc3N</chem>	37	4,27
<chem>c1cccc(c12)sc(c2)C3(CCCCC3)N(CC4)CCN4Cc5cn(nn5)-c6c(Sc7ccccc7)cccc6Sc8ccccc8</chem>	38	4,28
<chem>c1cc(C)cc(C2)c1N(O3)C(CC23)c4c(Cl)cccc4</chem>	39	4,28
<chem>CC(C)(C)[Si](C)(C)OC(CC1)CC(CC2)C1(C)C(CC3)C2C(C34C)CCC4C(C)(O)c5cccn5</chem>	40	4,29
<chem>c1cncc(Cl)c1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC</chem>	41	4,3

5		
N1CCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(C)CC5	42	4,3
Cc(c1)ccc(c12)NC(CC(C2)O)c(cc3)ccc3C	43*	4,3
CCCCCCCCCCCCCCCCC1(CC)C(O)=C(C)C(=O)S1	44	4,31
C1CN(CC2)CC(C12)Cc3ccc(cc3)-c4cccc4	45	4,33
c1cc(Cl)ccc1C(CC23)N(O3)c4c(C2)cccc4	46*	4,34
c1cc(C)cc(C2)c1N(O3)C(CC23)c(cc4)ccc4C	47	4,34
Cc(c1)ccc(c12)NC(CC(C2)O)c3c(Cl)cccc3	48	4,34
CCCCCCCCCCCCCCCCC1(C)C(O)=C(C(=O)S1)CC	49*	4,37
c1csc(c1C)C(CC(C2)O)Nc(c23)ccc(Cl)c3	50	4,38
C1NCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(C)CC5	51	4,38
c1cccc(c12)NC(CC(C2)O)c3ccc(Cl)cc3	52	4,38
c1cccc1\C=C\C(=O)N\N=C\c2cc([N+](=[O-])=O)c(O)c(c2)OC	53	4,4
c1cc(Cl)ccc1C(CC23)N(O3)c4c(C2)cc(F)cc4	54	4,4
CCOC(=O)CC(C(=O)OCC)NC(=O)c1ccc(cc1)N(CCCCCC)Cc(n2)cnc(c23)nc(N)nc3N	55*	4,41
c1cnccc1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	56	4,46
n1cccc1Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC	57*	4,47
Cc1cc(ccc1)C(N2)CC(O)Cc(c2c34)ccc3cccc4	58	4,48
n1cccc1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	59	4,48
c1cncc(c12)cccc2/C=C3/C(=O)\C(CC3)=C\c4cccc(c45)cncc5	60	4,5
ONC(=O)C(C1)ON=C1c2c(OCC)cccc2	61	4,52
COc1cc(OS(=O)(=O)C(F)(F)F)cc(c12)CC(C)(C)N(C2)Cc3cccc3	62	4,55
c1cc([N+](=[O-])=O)ccc1/C=C/c(cc2)nc(c23)cccc3	63*	4,55
CCCCCCCCCCCCCCCCC1(C)C(O)=C(C)C(=O)S1	64	4,55
c1cccc1Sc2cccc(c2CNCCN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC	65	4,56
c1cccc(c1C#N)N(CC2)CCN2C(=O)CN3CCN(CC3)c(c4C#N)cccc4	66	4,59
c1n[nH]c(c12)cc(cc2)NS(=O)(=O)c3ccc(Cl)cc3	67	4,59
c1ccc2ccc3ccc(c4ccc1c2c34)CNCCNCCN(CCNC5)CCNCc(n6)cccc56	68*	4,6
FC(F)(F)c1ccc(cc1)Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)c(c(cc5)OC	69	4,61
c1ccc(OC)cc1C(N2)CC(O)Cc(c2c34)ccc3cccc4	70	4,61
s1c(Br)ccc1C(CC23)N(O3)c4c(C2)cc(Cl)cc4	71*	4,62
c1cccc2ccc(c3c12)CC(O)CC(N3)c4cc(Cl)ccc4	72	4,63
CCC(=O)N1CCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)C(C(O)CC5	73	4,64
c1cc(F)ccc1/C=C/c(cc2)nc(c23)cccc3	74	4,66
O=P(O)(O)C(O)(P(=O)(O)O)CCC	75*	4,67
c1nccn1CCCNc(nnc2Cl)c(c23)cccc3O	76	4,68
c1cccc2ccc(c3c12)CC(O)CC(N3)c4cccc4	77	4,71
C1CN(C)CCN1CCNCCc2c(cccc2Sc(cc3)ccc3C(F)(F)F)Nc(c4)ccc(c45)n c6c(c5)cc(cc6)OC	78	4,72
n1cncn1CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3	79	4,74
c1ccc(Br)cc1C2CC(O3)Cc4ccc(c5c4C23)cccc5	80	4,74
C1NCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	81*	4,74
c1cc(I)cc(c2)c1n(C)c(c2c34)nc3cccc4	82	4,74
O=P(O)(O)C(O)(P(=O)(O)O)CCCCC	83	4,74
c1cccc1Cn2c(-c3cccc3)nc(c24)N(S(=O)(=O)N=C4N)Cc5cccc5	84	4,77
CN(C)c(cc1)ccc1/C=C/c(cc2)nc(c23)cccc3	85	4,77
C1CCCCN1C(=O)/C=C/c(cc2)nc(c23)cccc3	86*	4,77
c1c(Cl)ccc(c2)c1n(C)c(c2c34)nc3cccc4	87	4,77
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(s2)nnc2-c3ccc(Cl)cc3	88	4,78
c1cc(Cl)ccc1C2CC(O3)Cc4ccc(c5c4C23)cccc5	89	4,79
c1cc(C#N)cc(c2)c1n(C)c(c2c34)nc3cccc4	90*	4,79
c1ccc(C(F)(F)F)c(c2)c1n(C)c(c2c34)nc3cccc4	91	4,79
C1CN(C)CCN1CC(=O)NCc2c(cccc2Sc3ccccn3)Nc(c4)ccc(c45)nc6c(c	92	4,8

5)cc(cc6)OC		
CCOC(=O)c1ccc(cc1)C(C23)=CN(CC2)CC3	93	4,8
c1ccc(Br)c(c2)c1n(C)c(c2c34)nc3cccc4	94*	4,8
N#CSCCOc1ccc(cc1)Oc2ccccc2	95	4,8
CC(C)(C1)NCC(c12)c(OC)cc(OC)c2-c3ccc(O)c(c34)cccc4	96	4,82
c1ccc(OC)cc1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	97*	4,83
c1cccc2ccc(c3c12)CC(O)CC(N3)c4ccc(F)cc4	98	4,83
c1cccc1CN(C(C)(C)C2)Cc(c23)c(OC)cc(c3Br)OC	99	4,83
c1cccc(c12)cc3c(cccc3)c2CNCCCNCN(CCN4)CCNCC(n5)cccc45	100	4,85
O=P(O)(O)C(P(=O)(O)O)CNN1CCCC1	101	4,85
c1ccc(Cl)cc1C2CC(O3)Cc4ccc(c5c4C23)cccc5	102	4,88
c1ncccc1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	103*	4,88
c1cccc(c12)cc3c(cccc3)c2CNCCN(CCN4)CCNCC(n5)cccc45	104	4,89
c1cccc1CN(C2)C(C)(C)Cc(c23)c(I)ccc3OC	105	4,89
C1COCCN1C(=O)/C=C/c(cc2)nc(c23)cccc3	106	4,89
C1NCCCC1C(C(=O)OC(C)(C)C)O\N=C(c2c3cccc2)\c(c34)c(C(F)(F)F)nc5c4cc(Br)cc5	107	4,91
CN(C)Cc1cnc([nH]1)-c2c(Sc3cccc3)cccc2Sc4cccc4	108	4,91
c1cc(Cl)cc(c2)c1n(C)c(c2c34)nc3cccc4	109*	4,92
CCN(CC)Cc1cnc([nH]1)-c2c(Sc3cccc3)cccc2Sc4cccc4	110*	4,94
c1cccc1Sc2cccc(c2CNC(=O)CN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC	111	4,95
c1cccc1Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC	112	4,95
O=Cc1cc(O)c(cc1)OC(F)F	113	4,96
c1cccc1Sc(ccc(Br)c2)c2-n(nn3)cc3CN4CCN(CC4)C5(CCCCC5)c(c6)sc(c67)cccc7	114*	4,97
c1cccc2ccc(c3c12)CC(O)CC(N3)c4c(Cl)cccc4	115	4,97
CC1=NC(C)(C)Cc(c12)c(c(OC)cc2OC)-c3ccc(cc3)OC	116	4,97
c1cccc1Sc2cccc(c2CNC(=N)CN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC	117	4,99
CCCC(=O)c1cc(c(CC(F)F)cc1)Oc2ccccc2	118	4,99
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c3ncccn3	119	5
c1cccc2ccc(c3c12)CC(O)CC(N3)c4ccc(Br)cc4	120	5
O=P(O)(O)C(P(=O)(O)O)CNCCCCCCC	121	5
CC(C)CCNCC(P(=O)(O)O)P(=O)(O)O	122*	5
CC(C)(C)NCC(P(=O)(O)O)P(=O)(O)O	123	5
CCCC(O)c1cc(c(cc1)OC)OC2CCOCC2	124	5
C1CCCc(c12)sc(c2C(=O)OCC)\N=c(\n3CC=C)sc3-c4cccc4	125*	5,01
C1CN(CC2)CC(C12)(O)C#Cc3cc(ccc3)OCc4cccc4	126	5,02
COc(cc1)cc(c2)c1nc(c23)ccc(c3)Nc(c4CNC(=O)CN5CCN(C)CC5)cccc4Sc(c6)ccc(c67)cccc7	127	5,03
C1CN(C)CCN1CC(=O)NCC2c(cccc2Sc(cc3)ccc3C(F)(F)F)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC	128	5,03
C1CCCc(c12)sc(c2C(=O)OC(C)C)NC(=S)Nc3cccc3	129*	5,06
c1cccc1CN(C(C)(C)C2)Cc(c23)c(OC)cc(c3I)OC	130	5,07
c1cccc(c12)nc(cc2)CNCCCNCCN(CCN3)CCNCC(n4)cccc34	131	5,08
c1cc(Br)cc2c1nc(C(F)(F)F)c(c23)C(C)(OCC4CO4)c5c3cccc5	132	5,08
c1c(Br)ccc(c12)nc(C(F)(F)F)c(c2Br)C(O)c3ccccc3	133*	5,09
[O-][N+](=O)c1ncn(n1)CC(=O)NC2CCN(CC2)Cc3ccccc3	134	5,1
CC(C)(C)OC(=O)NC(C(C)(C)C)C(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6	135	5,1
FC(F)Oc(cc1)c(OC(F)F)cc1C(=O)CCCc2ccccc2	136	5,11
c1cccc2C(c(c3c12)cccc3)COC(=O)NC(CC(C)C)C(=O)O\N=C(c4c5ccc4)\c(c56)c(nc7c6cc(Br)cc7)N8CCN(CC8)c9ccccc9	137*	5,12
n1c[nH]cc1CCNc(nnc2Cl)c(c23)cccc3O	138	5,13
c1cccc2c1C(C)(OC(=O)N(C)C)c(c23)c(OC)nc4c3cc(Br)cc4	139	5,13
CC(C)(C)OC(=O)NC(CCSC)C(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3c	140	5,13

c(Br)cc4)N5CCN(CC5)c6ccccc6		
CC(C)(C)OC(=O)N1CCCC1C(=O)O\N=C(c2c3cccc2)\c(c34)c(-n5ccnc5)nc6c4cc(Br)cc6	141	5,13
N1CCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	142	5,13
c1cccc1CN(CC2)CCN2CC(O)COC(C)(c3c4cccc3)c(c45)c(OC)nc6c5cc(Br)cc6	143*	5,14
c1cccc2c1C(=N/O)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6	144	5,15
C1CCCCc(c12)sc(c2C(=O)OC(C)C)\N=C(SCC3=O)\N3c4cccc4	145	5,16
c1cc(Br)cc2c1nc(OCC(F)(F)F)c(c23)C(C)(O)c4c3cccc4	146*	5,16
c1cccc1C(n(cn2)cc2[N+][O-])=O)c(c3)c(OC)nc(c34)ccc(Br)c4	147	5,16
c1cccc2c1C(=O)c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6	148	5,16
CC(C)(C)OC(=O)NCC(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6	149*	5,17
c1cccc1\C=C\C(=O)N\N=C\c2c(O)cccc2	150	5,18
c1cccc1Sc2cccc(Sc3cccc3)c2-c([nH]4)nc4C[N+](C)(C)Cc5cnc([nH]5)-c6c(Sc7cccc7)cccc6Sc8cccc8	151	5,2
c1ccc(OC)cc1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	152	5,21
CC(C)(C)OC(=O)NC(C)C(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6	153	5,21
O=P(O)(O)C(P(=O)(O)O)CNN1CCCCC1	154*	5,21
CC(C)(C)[Si](C)(C)OC(CC1)CC(CC2)C1(C)C(CC3)C2C(C34C)CCC4C(C)(O)c5ccncc5	155	5,21
C1CCCCC1CCC(=O)c2cc(c(cc2)OC)OC3CCOC3	156	5,21
C1CCCCc(c12)sc(c2C(=O)OC(C)C)\N=c(\n3-c4cccc4)sc3-c5cccc5	157	5,22
c1cccc(c12)nc(cc2)CNCCN(CCNC3)CCNCC(c4)cccc34	158	5,22
c1ccc(OC)cc1-c2ccn(n2)C(CO)c3c(ccc(F)c3)Oc(nc(c45)ccc(c5)[N+][O-])=O)c(c4)Cc6cccc6	159	5,22
c1cc(F)cc(c2)c1n(C)c(c2c34)nc3cccc4	160*	5,22
o1cccc1C(=O)N2CCN(CC2)C(=O)OC(Cn3ccnc3)c4ccc(Cl)cc4	161	5,23
c1cccc1C(c2cccc2)N3CCN(CC3)CC(O)COC(C)(c4c5cccc4)c(c56)c(OC)nc7c6cc(Br)cc7	162	5,23
COc(cc1)ccc1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	163	5,24
FC(F)Oc(cc1)c(OC(F)F)cc1C(=O)CC2CCCCC2	164	5,24
C1CCCCC1CCC(=O)c2cc(c(cc2)OC(F)F)OC3CCOC3	165	5,24
c1cccc1CCC(=O)c2cc(c(cc2)OC)OC3CCOC3	166*	5,25
OCC1(C)CCCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4	167	5,26
Cc1cc(n(O)c(=O)c1)C2CCCCC2	168	5,27
FC(F)(F)c1cc(ccc1)-c2ccn(n2)CC(O)COC(C)(c3c4cccc3)c(c45)c(OC)nc6c5cc(Br)cc6	169*	5,27
OCc1ccc(cc1)-c(on2)cc2-c(c3)ccc(c34)OCO4	170	5,28
c12cccc(n2)CNCCN(CCNC1)CCNCC3ccnc(c34)c5c(cc4)cccn5	171	5,28
CCCCC(=O)c1cc(OC)c(cc1)OC2CCOCC2	172*	5,28
c12cccc(n2)CNCCN(CCNC1)CCNCC3cccc(n3)CNCCN(CCNC4)CCNCC(n5)cccc45	173	5,29
O=C/C=C/C(cc1)nc(c12)cccc2	174	5,3
O=P(O)(O)C(P(=O)(O)O)CNCCCC	175	5,32
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c3ccccc3	176	5,33
c1cccc(OC)c1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	177*	5,36
C1CCCCC1CC(=O)c2cc(c(cc2)OC(F)F)OC3CCOC3	178	5,36
c1cccc(c12)sc(c2)C3(CCCCC3)N(CC4)CCN4Cc5c[nH]c(n5)-c6c(ccc(Br)c6)Sc7cccc7	179	5,37
[O-][N+](=O)c1ncn1CC(=O)Nc2ccc(cc2)Oc3ccccc3	180	5,38
c12cccc(n2)CNCCN(CCNC1)CCNCC(c3ccc4c5c36)ccc6ccc5ccc4	181	5,38

<chem>O=P(O)(O)C(P(=O)(O)O)CNCCCCCCCC</chem>	182	5,39
<chem>c1cccc(c2)c1n(C)c(c2c34)nc3cccc4</chem>	183*	5,4
<chem>CSc(cc1)cc(c2)c1n(C)c(c2c34)nc3cccc4</chem>	184	5,4
<chem>c1ccc(OC)c(c2)c1n(C)c(c2c34)nc3cccc4</chem>	185*	5,4
<chem>CCCC(=O)c1cc(OC)c(cc1)OC2CCOCC2</chem>	186	5,4
<chem>[O-][N+](=O)C(\C)=C\c1ccsc1</chem>	187	5,41
<chem>c1cccc(c12)ncc(c2)C(O)(CCC)c3cc(c(cc3)OC(F)F)Oc4ccccn4</chem>	188	5,42
<chem>CCCCC(=O)c1cc(OC)c(cc1)OC2CCN(CC2)C(=O)OC(C)(C)C</chem>	189	5,42
<chem>FC(F)(F)c1cc(ccc1)S(=O)(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5</chem>	190	5,43
<chem>CCCC(=O)c1cc(OC)c(cc1)OC2CCOC2</chem>	191*	5,43
<chem>c12cccc(n2)CNCCN(CCNC1)CCNCc(n3)ccc4ccc(c5c34)ccc(n5)CNCCN(CCNC6)CCNCc(n7)cccc67</chem>	192	5,44
<chem>CC(C)CC(=O)c1cc(c(cc1)OC(F)F)OC2CCCC2</chem>	193	5,44
<chem>OCC1(C)C(O)CCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4</chem>	194	5,46
<chem>[O-][N+](=O)c1nccn1CC(=O)Nc2ccc(cc2)Oc3ccc(F)cc3</chem>	195*	5,48
<chem>[O-][N+](=O)c1nccn1CC(C)(O)c2ccc(cc2)-n3ccnc3</chem>	196	5,48
<chem>CCCC(=O)c1cc(OC)c(cc1)OC2CCN(CC2)C(=O)OC(C)(C)C</chem>	197	5,48
<chem>N1CCCN(C1=O)C(C(C)C)C(=O)NC(Cc2cccc2)CC(O)C(Cc3cccc3)N</chem>	198	5,49
<chem>C(=O)COc(c4C)c(C)ccc4</chem>	199	5,49
<chem>c1cccc(c12)nccc2CNCCN(CCNC3)CCNCc(n4)cccc34</chem>	200*	5,5
<chem>COc(cc1)ccc1CC(N)C(=O)NC(C2CO)C(O)C(O2)n(cn3)c(c34)ncnc4N(C)C</chem>	201	5,51
<chem>C1CN(CC2)C=C(C12)C#Cc3ccc(cc3)OCc4cccc4</chem>	202	5,51
<chem>c1cccc1CCC(=O)c2cc(c(cc2)OC(F)F)OC3CCOC3</chem>	203	5,51
<chem>c1cccc1CCCC(=O)c2cc(c(cc2)OC(F)F)OC3CCOC3</chem>	204*	5,54
<chem>c1cccc(OC)c1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5</chem>	205	5,55
<chem>c1cccc(c1C#N)N(CC2)CCN2C(=O)Cn(n3)cnc3[N+](=[O-])=O</chem>	206*	5,55
<chem>c1cnccc1CNCCCNCCN(CCNC2)CCNCc(n3)cccc23</chem>	207	5,57
<chem>c1cccc1-c(cc2)ccc2C(=O)OC(Cn3ccnc3)c4ccc(cc4)-c5cccc5</chem>	208	5,57
<chem>FC(F)(F)c1ccc(cc1)N(CC2)CCN2C(=O)CN3CCN(CC3)c(cc4)ccc4C(F)(F)F</chem>	209	5,59
<chem>[O-][N+](=O)c1nccn1CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3</chem>	210*	5,6
<chem>[O-][N+](=O)c1nccn1CC(=O)Nc(n2)nccc2-c3cccc3</chem>	211	5,6
<chem>c1cccc(OC)c1NC(=O)NCC(O)COC(C)(c2c3cccc2)c(c34)c(OC)nc5c4c</chem>	212	5,62
<chem>c(Br)cc5</chem>	213	5,62
<chem>[O-][N+](=O)c1nccn1CC(=O)N2CCN(CC2)c3ccc(cc3)OC</chem>	214	5,62
<chem>CCCC(=O)ON=C(c1c2cccc1)\c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5</chem>	215	5,63
<chem>c1cc(Cl)c(Cl)cc1C[N+](C)(C)Cc2cnc([nH]2)-c3c(Sc4cccc4)cccc3Sc5cccc5</chem>	216	5,63
<chem>c1cccc1\C=C\C(=O)CC(c2cccc2)C3(C)C(=O)OC(=N3)c4cccc4</chem>	217*	5,63
<chem>Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C(C)C)NCc(n3)ccn(c34)nc4</chem>	218	5,64
<chem>c1cccc(c12)sc(c2)C3(CCCC3)N(CC4)CCN4Cc5c[nH]c(n5)-c6c(Sc7cccc7)cccc6Sc8cccc8</chem>	219	5,67
<chem>FC(F)Oc(cc1)c(OC(F)F)cc1C(=O)CCc2cccc2</chem>	220	5,68
<chem>c1nccn1CC(O)COC(C)(c2c3cccc2)c(c34)c(OC)nc5c4cc(Br)cc5</chem>	221	5,7
<chem>CCCCC(=O)c1cc(c(cc1)OC)OC2CCOCC2</chem>	222*	5,71
<chem>C1CCCCN1C(CO)c2c(ccc(F)c2)Oc(nc(c34)ccc(c4)[N+](=[O-])=O)c(c3)Cc5cccc5</chem>	223	5,72
<chem>c1nccn1CCc2c(ccc(F)c2)Oc(c(c3)Cc4cccc4)nc(c35)ccc(c5)-n(nn6)cc6-c7cccc7</chem>	224	5,74
<chem>c1cccc(c12)ncc(c2)C(CCC)c3cc(OC)c(cc3)Oc4ccccn4</chem>	225*	5,77
<chem>CCCCC(=O)c1cc(OC)c(cc1)OC2CCOC2</chem>	226	5,77
<chem>c1nccn1CC(c2ccc(F)cc2)OC(=O)N3CCN(CC3)c(cc4)ccc4[N+](=[O-])=O</chem>	227	5,77
<chem>OCC1(C)/C(=N/O)CCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4</chem>		
<chem>COc(cc1)cc(c2)c1n(C)c(c2c34)nc3cccc4</chem>		

C1CCCCC1CC(=O)c2cc(c(cc2)OC)OC3CCOC3	228*	5,77
c1ccc(O)c(C)c1C(=O)NC(CSc2cccc2)[C@H](O)CN(C(C3)C(=O)NC(C)(C)C)CC(C34)CCCC4	229	5,8
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NC(=O)c3ccncc3	230	5,8
CCCC(=O)c1cc(c(CC(F)F)cc1)OC2CCOCC2	231	5,8
CCCCC(=O)OC(C)(c1c2cccc1)c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	232*	5,84
c1cccc2c1C(C)(O)c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	233	5,86
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(cc3)ccc3C	234	5,88
c1n[nH]c(c12)cc(cc2)NS(=O)(=O)c(cc3)ccc3F	235	5,89
[O-][N+](=O)C(\C)=C\c1csc(c12)cccc2	236	5,89
c1ccc([N+])([O-])=O)cc1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5ccccc5	237	6
[O-][N+](=O)C(\C)=C\c1cc(Br)cs1	238	6
[O-][N+](=O)\C=C\c(cc1)nc(c12)cccc2	239	6
N#Cc(c1)ccc2nc(c3c12)n(C)c4c(c3)cccc4	240*	6
c1cccc2nc(c3c12)n(C)c4c(c3)ccc(c4)OC	241	6
c1cccc2c1C(C)(OC(=O)C)c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	242*	6,03
c1ncn1CC(c2ccc(Cl)cc2)OC(=O)N3CCN(CC3)c(cc4)ccc4[N+](O-)=O	243	6,06
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)Cc(cc3)ccc3C(F)(F)F	244	6,06
c1cc(Cl)ccc1C(c2cncnc2)N3CCN(CC3)S(=O)(=O)C	245	6,06
O=P(O)(O)C(P(=O)(O)O)CNCCCCC	246	6,08
o1cccc1C(=O)N2CCN(CC2)c(cc3)ccc3NC(=O)OC(Cn4ccnc4)c5ccc(F)cc5	247*	6,1
C1CC(O)(CO)C(C2)CC(C123)CCC4C3(C)CCC5C4(C)COS(=O)O5	248	6,11
c1ncncc1C(c(cc2)ccc2C(F)(F)F)N3CCN(CC3)S(=O)(=O)C	249	6,11
c1cc(Cl)ccc1C(C)(c2cccc2)OCCC3CCCN3C	250	6,12
c1ncn1CC(c2ccc(F)cc2)OC(=O)c3ccc(cc3)-c4ccccc4	251*	6,15
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NC(=O)c(c3)ccc(c34)nccc4	252	6,15
n1cncn1CC(C)(O)c2ccc(cc2)-c3ccccc3	253	6,15
[O-][N+](=O)C(\C)=C\c1ccc(s1)Br	254	6,15
O=P(O)(O)C(P(=O)(O)O)CNCCCCCCCCCCCC	255*	6,17
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NC(=O)c(cc3)cc(c34)scn4	256	6,19
c1cccc1C(=O)OCC2(C)C(O)CCC3(C)C2CCC(C345)CC(C4)C(O)(CC5)COC(=O)c6ccccc6	257	6,22
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(cc3)ccc3C(F)(F)F	258	6,22
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(ccc2)cc2-c(cs3)nc3C	259	6,26
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)N3CCCCC3	260*	6,26
O=P(O)(O)C(P(=O)(O)O)CNCCCCC	261	6,27
c1ncccc1C(c(cc2)ccc2C#N)N(CC3)CCN3C(=O)c4c(C)ncs4	262	6,29
Nc1nc(N)nc(c12)ccc(c2)C#Cc3ccc(cc3)OCc4ccccc4	263	6,3
c1ncn1CC(c2ccc(Cl)cc2)OC(=O)c3ccc(cc3)-c4ccccc4	264	6,32
c1cccc2c1C(C)(OCC(O)CN)c(c23)c(OC)nc4c3cc(Br)cc4	265	6,33
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(n2)sc2-c3cc(F)c(F)cc3	266*	6,34
c1cccc1C(n2ccnc2)c(c3)c(OC)nc(c34)ccc(Br)c4	267	6,34
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C(C)C)NCc(c3)ccc(c34)nccc4	268*	6,42
c1ncccc1C(c(cc2)ccc2C#N)N3CCN(CC3)S(=O)(=O)C	269	6,44
CCC1(O)C(=O)OCc(c2=O)c1cc3n2Cc(c34)cc5c(n4)cccc5	270	6,47
c1cccc1CC(Cn(n2)cnc2[N+](O-)=O)NS(=O)(=O)c(cc3)ccc3C	271	6,47
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)Cc3cc(Cl)c(Cl)cc3	272*	6,53
c1ncccc1C(c(cc2)ccc2C#N)N(CC3)CCN3C(=O)C4CC4	273	6,55
c1cccc2c1C(=O)c(c23)c(Nc4ccncc4)nc5c3cc(Br)cc5	274	6,6
c1ncn1C(CO)c2c(ccc(F)c2)Oc(nc(c34)ccc(c4)[N+](O-	275	6,6

])=O)c(c3)Cc5ccccc5		
c1ncccc1C(c(cc2)ccc2C#N)N(CC3)CCN3C(=O)C(C)(C)C	276*	6,6
CC(=O)N1CCN(CC1)C(c2ccc(Cl)cc2)c3ccnc3	277	6,62
n1ccncc1C(c(cc2)ccc2C(F)(F)F)N3CCN(CC3)c4c(OC)cccc4	278	6,62
CCCCCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	279	6,66
[O-][N+](=O)c1ncn(n1)CC(=O)c2ccc(cc2)-c3ccccc3	280*	6,73
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(ccc2)cc2-c3ccccc3	281	6,74
Nc1nc(N)nc(c12)ccc(c2)/C=C/c3ccc(cc3)OCc4cccc4	282	6,74
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(c3)ccc(Cl)c3Cl	283	6,77
C1CC1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cncnc4	284	6,77
CCCCCCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	285*	6,82
Br\C=C\c(cc1)nc(c12)cccc2	286	6,82
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NCc3ccncc3	287	6,84
N#Cc1ccc(cc1)-c2ccc(cc2)NC(=O)Cn(n3)cnc3[N+](=[O-])=O	288	6,86
CN(C)C(=O)N1CCN(CC1)C(c2ccc(Cl)cc2)c3ccnc3	289	6,92
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)C	290*	6,96
CCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	291	6,96
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(n2)sc2-c3ccc(Cl)cc3	292	6,99
[O-][N+](=O)c1ncn(n1)CC(C)(O)c2ccc(cc2)-c3ccccc3	293	6,99
CC(C)(C)C(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cncnc3	294	7,09
c1ncccc1C(c(cc2)ccc2C(F)(F)F)N3CCN(CC3)S(=O)(=O)C	295	7,11
c1cc(Cl)cc(F)c1C(c2cncnc2)N3CCN(CC3)c4cccc4	296*	7,11
c1cc(O)ccc1CCc(c2)ccc(c23)nc(N)nc3N	297	7,11
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccc4	298	7,14
c1cccc1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	299	7,15
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(cc2)ccc2Cc3ccccc3	300*	7,17
c1cccc1C(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4ccnc4	301	7,17
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccccc3	302	7,19
c1cc(Cl)cc(F)c1C(c2cncnc2)N3CCN(CC3)c4c(OC)cccc4	303	7,19
Nc1nc(N)nc(c12)ccc(c2)CCc3ccc(cc3)OCc4cccc4	304	7,23
CC(C)(C)NC(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cncnc3	305*	7,27
c1ncccc1C(c(cc2)ccc2C#N)N3CCN(CC3)c4nccs4	306	7,3
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(cc2)ccc2-c3ccc(Cl)cc3	307*	7,35
Fc1c(F)cc(F)c(F)c1OCC(=O)C(COC)n(nn2)cc2C(C)(C(C)C)NCc(cc3)cc(c34)scn4	308	7,36
CCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	309	7,38
CCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	310	7,39
c1nccn1CC(c2ccc(F)cc2)OC(=O)Nc3ccc(cc3)N4CCN(CC4)S(=O)(=O)CC	311	7,4
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccc(F)cc3	312*	7,44
s1cnc(C)c1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	313	7,44
CC(C)C(=O)N1CCN(CC1)C(c2ccnc2)c3ccc(cc3)OC(C)C	314	7,46
c1cc(Cl)cc(F)c1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)C	315	7,46
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)c4cccc4	316	7,47
[O-][N+](=O)c1ncn(n1)CC(O)(Cn2cncn2)c3ccc(cc3)-c4cccc4	317*	7,48
CC(C)(C)C(=O)N1CCN(CC1)C(c(cc2)ccc2C(F)(F)F)c3ccnc3	318	7,52
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2cc(ccc2)OCc3ccccc3	319	7,55
C1CCCCC1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	320	7,55
c1ncccc1C(c(cc2)c(F)cc2C(F)(F)F)N3CCN(CC3)S(=O)(=O)C	321*	7,55
CC(C)S(=O)(=O)N(CC1)CCN1C(c2ccc(Cl)cc2)c3ccnc3	322	7,55
CC(C)Oc(cc1)ccc1C(c2ccnc2)N3CCN(CC3)c4c(OC)cccc4	323	7,55
C1CC1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	324	7,59
c1cc(Cl)cc(F)c1C(c2ccnc2)N3CCN(CC3)c4nccn4	325*	7,59
c1ncncc1C(c(cc2)c(F)cc2C(F)(F)F)N3CCN(CC3)S(=O)(=O)C	326	7,6
CC(C)(C)C(=O)N1CCN(CC1)C(c(cc2)c(F)cc2C(F)(F)F)c3ccnc3	327	7,62
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)c4ccc(Cl)cc4	328	7,62
CC(C)(C)OC(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cncnc3	329	7,62

Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NCc(c3)c cc(c34)nccc4	330*	7,66
c1cc(Cl)ccc1NC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cncnc4	331	7,66
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCOCC)n(nn2)cc2C(C)(C3CCCC3)NCc (n4)ccn(c45)nccc5	332	7,7
CC(C)(C)OC(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cccnc3	333	7,7
c1cc(Cl)ccc1C(c2cccnc2)N3CCN(CC3)c4c(OC)cccc4	334	7,7
CC(C)(C)C(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cccnc3	335*	7,72
CC(C)c1ccc(cc1)S(=O)(=O)N(CC2)CCN2C(c3c(F)cc(Cl)cc3)c4cncnc4	336	7,74
c1cc(Cl)cc(F)c1C(c2cccnc2)N3CCN(CC3)c4cccc4	337	7,74
CC(C)(C)OC(=O)N1CCN(CC1)C(c(cc2)ccc2C(F)(F)F)c3cccnc3	338	7,77
c1ncccc1C(c(cc2)c(F)cc2C(F)(F)F)N3CCN(CC3)c(on4)nc4C	339	7,77
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C3CCCC3)NCc(c4)ccc(c45)nccc5	340*	7,82
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C(C)C) NCc(cc3)cc(c34)scn4	341	7,92
CC(C)c1ccc(cc1)S(=O)(=O)N(CC2)CCN2C(c3ccc(Cl)cc3)c4cccnc4	342	7,92
C1CN(CC2)CC(C12)(O)C#Cc3c(Cc4cccc4)nc(cc3)CCCCOC	343	7,96
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C3CC CC3)NCc(cc4)cc(c45)scn5	344	8
c1cc(Cl)ccc1C(c2cccnc2)N3CCN(CC3)S(=O)(=O)C4CCCCC4	345*	8
c1cc(Cl)ccc1NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cncnc4	346	8
c1cc(Cl)ccc1C(c2cccnc2)N3CCN(CC3)S(=O)(=O)c4cccc4	347	8,05
c1cccc1NC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccnc4	348	8,05
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NCc(n3) ccn(c34)nccc4	349*	8,1
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3	350	8,1
c1cc(Cl)cc(F)c1C(c2cccnc2)N3CCN(CC3)S(=O)(=O)c4cccc4	351	8,1
c1cccc1OC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccnc4	352	8,1
C1CN(CC2)CC(C12)(O)C#Cc3c(Cc4cccc4)nc(cc3)N(C5)CC(O)C5O C	353	8,1
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCOCC)n(nn2)cc2C(C)(C3CCCC3)NCc (cc4)cc(c45)scn5	354*	8,15
CC(C)c1ccc(cc1)NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cncnc4	355	8,15
c1cccc1NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cccnc4	356	8,22
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C3CC CC3)NCc(cc4)cc(c45)ccc5	357	8,3
c1cccc1OC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cccnc4	358*	8,3
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C3CCCC3)NCc(cc4)cc(c45)scn5	359	8,52
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCS(=O)(=O)C)n(nn2)cc2C(C)(C(C)C) NCc(cc3)cc(c34)ccc4	360*	8,52
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C3CCCC3)NCc(n4)ccn(c45)nccc5	361	8,52
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCOCC)n(nn2)cc2C(C)(C(C)C)NCc(c3) ccc(c34)nccc4	362	8,7
Fc1c(F)cc(F)c(F)c1OCC(=O)C(CCCC)n(nn2)cc2C(C)(C(C)C)NCc(cc3) cc(c34)scn4	363*	8,7

* Conjunto teste

ANEXO B – Representação do conjunto 2 (pSI e pIC₅₀).

Representação na forma SMILES	Estrutura	pIC ₅₀	pSI
<chem>[O-][N+](=O)c1nccn1CC(=O)c2ccc(cc2)-c3ccccc3</chem>	1	1,85	-3,17
<chem>n1cncn1CC(=O)c2ccc(cc2)-c3ccccc3</chem>	2	1,64	-2,31
<chem>Cc1ccc(cc1)N(CC2)CCN2C(=O)CN3CCN(CC3)c(cc4)ccc4C</chem>	3	1,53	-2,15
<chem>[O-][N+](=O)c1nccn1CC(=O)N2CCN(CC2)c3ncccn3</chem>	4	3,85	-1,83
<chem>[O-][N+](=O)c1nccn1CC(=O)Nc(n2)sc2-c3cc(F)c(F)cc3</chem>	5	2,4	-1,81
<chem>c1cc(Cl)ccc1N(CC2)CCN2C(=O)CN3CCN(CC3)c4ccc(Cl)cc4</chem>	6*	2,22	-1,72
<chem>[O-][N+](=O)c1nccn1CC(=O)N2CCN(CC2)C(=O)C3CC3</chem>	7	3,83	-1,46
<chem>c1c(Br)ccc(c12)nc(C(F)(F)F)c(c2Br)C(O)c3ccccc3</chem>	8	5,09	-0,73
<chem>c1cc(F)cc(F)c1C(=O)Cn(n2)cnc2[N+](=[O-])=O</chem>	9	2,92	-0,69
<chem>c1cc(F)cc(F)c1C(O)(Cn2cncn2)Cn(n3)cnc3[N+](=[O-])=O</chem>	10	2,93	-0,66
<chem>c1ccccc1Sc2cccc(c2CNC(=O)CN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC</chem>	11*	4,95	-0,57
<chem>C1CN(CC2)CC(C12)Cc3ccc(cc3)-c4ccccc4</chem>	12	4,33	-0,36
<chem>c1cc(Br)cc2c1nc(C(F)(F)F)c(c23)C(C)(OCC4CO4)c5c3ccccc5</chem>	13*	5,08	-0,32
<chem>CCN(CC)Cc1cnc([nH]1)-c2c(Sc3ccccc3)cccc2Sc4ccccc4</chem>	14	4,94	-0,32
<chem>CCC(=O)N1CCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5</chem>	15	4,64	-0,31
<chem>C1CN(C)CCN1Cc2cnc([nH]2)-c3c(Nc4ccccc4)cccc3Sc5ccccc5</chem>	16	4,27	-0,28
<chem>FC(F)(F)c1ccc(cc1)Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC</chem>	17	4,61	-0,25
<chem>c1cccc2c1C(=O)c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccc6</chem>	18	5,16	-0,2
<chem>CC(C)(C1)NCc(c12)c(OC)cc(OC)c2-c3ccc(O)cc3</chem>	19*	3,84	-0,2
<chem>C1CN(C)CCN1CCNCc2c(cccc2Sc(cc3)ccc3C(F)(F)F)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC</chem>	20	4,72	-0,12
<chem>c1ccccc1CN(CC2)CCN2CC(O)COC(C)(c3c4ccccc3)c(c45)c(OC)nc6c5cc(Br)cc6</chem>	21	5,14	-0,11
<chem>FC(F)(F)c1ccc(cc1)Sc2cccc(c2CNC(=O)CN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC</chem>	22	3,98	-0,1
<chem>c1ccccc1Sc2cccc(c2CNC(=N)CN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC</chem>	23	4,99	-0,1
<chem>c1cc(Cl)c(Cl)cc1C[N+](C)(C)Cc2cnc([nH]2)-c3c(Sc4ccccc4)cccc3Sc5ccccc5</chem>	24	5,62	-0,09
<chem>c1cncccc1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5</chem>	25*	4,46	-0,06
<chem>c1cc(Cl)cc(c2)c1n(C)c(c2c34)nc3ccccc4</chem>	26	4,74	-0,06
<chem>C1CN(C)CCN1Cc2cnc([nH]2)-c3c(Sc4ccccc4)cccc3Sc5ccccc5</chem>	27	3,94	-0,05
<chem>c1c(Cl)ccc(c2)c1n(C)c(c2c34)nc3ccccc4</chem>	28	4,77	-0,04
<chem>c1ccc(OC)cc1N(CC2)CCN2CC(O)COC(C)(c3c4ccccc3)c(c45)c(n c6c5cc(Br)cc6)N(CC7)CCN7C(c8ccccc8)c9ccccc9</chem>	29	4,19	-0,02
<chem>COc(cc1)ccc1CC(N)C(=O)NC(C2CO)C(O)C(O2)n(cn3)c(c34)ncnc4N(C)C</chem>	30*	5,5	-0,01
<chem>c1cc(C#N)cc(c2)c1n(C)c(c2c34)nc3ccccc4</chem>	31	4,79	-0,01
<chem>c1ccc(C(F)(F)F)c(c2)c1n(C)c(c2c34)nc3ccccc4</chem>	32	4,79	-0,01
<chem>c1cc(C)cc(C2)c1N(O3)C(CC23)c4c(Cl)cccc4</chem>	33	4,28	-0,01
<chem>CN(C)Cc1cnc([nH]1)-c2c(Nc3ccccc3)cccc2Sc4ccccc4</chem>	34	3,79	0,01
<chem>CC(C)(C)OC(=O)N1CCCC1C(=O)O\N=C(c2c3ccccc2)\c(c34)c(-n5ccnc5)nc6c4cc(Br)cc6</chem>	35*	5,13	0,04
<chem>C1CN(C)CCN1Cc2cnc([nH]2)-c3c(ccc(Br)c3)Sc4ccccc4</chem>	36	4,12	0,04
<chem>C1CN(C)CCN1CC(=O)NCc2c(cccc2Sc(cc3)ccc3C(F)(F)F)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC</chem>	37	5,03	0,06
<chem>C1CN(C)CCN1CC(=O)NCc2c(cccc2Sc3ccccc3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC</chem>	38	4,8	0,08
<chem>c1cccc(c12)sc(c2)C3(CCCCC3)N(CC4)CCN4Cc5cn(nn5)-</chem>	39*	4,28	0,09

c6c(Sc7cccc7)cccc6Sc8cccc8			
CSc(c1)cc(c2)c1n(C)c(c2c34)nc3cccc4	40	5,4	0,1
c1ccc(OC)c(c2)c1n(C)c(c2c34)nc3cccc4	41	5,4	0,1
c1cccc2ccc(c3c12)CC(O)CC(N3)c4cc(Cl)ccc4	42*	4,63	0,12
n1cccc1Sc2cccc(c2CNC(=O)CN(C)C)Nc(c3)ccc(c34)nc5c(c4)c c(cc5)OC	43	4,24	0,16
O1CC1(C)c2ccc(cc2)-n3ccnc3	44	3,73	0,17
CC(C)(C1)NCc(c12)c(OC)cc(OC)c2-c3ccc(O)c(c34)cccc4	45	4,82	0,18
n1cccc1Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc 5)OC	46*	4,47	0,21
N1CCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	47	5,13	0,21
C1NCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	48	4,74	0,21
c1cccc2C(c(c3c12)cccc3)COC(=O)NC(CC(C)C)C(=O)O\N=C(c 4c5cccc4)\c(c56)c(nc7c6cc(Br)cc7)N8CCN(CC8)c9ccccn9	49	5,12	0,26
FC(F)(F)c1cc(ccc1)S(=O)(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc (F)c4)c(c3)Cc5ccccc5	50*	5,43	0,33
c1cccc(OC)c1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5ccccc5	51	5,54	0,33
c1ccc([N+])([O-])=O)cc1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc 5ccccc5	52	6	0,34
c1cc(F)cc(c2)c1n(C)c(c2c34)nc3cccc4	53*	5,22	0,4
Cc1cc(ccc1)C(N2)CC(O)Cc(c2c34)ccc3cccc4	54	4,48	0,42
c1cccc2c1C(C)(OCC(O)CN)c(c23)c(OC)nc4c3cc(Br)cc4	55*	6,33	0,46
c1cc(Br)cc2c1nc(OCC(F)(F)F)c(c23)C(C)(O)c4c3cccc4	56	5,16	0,46
CC1=NC(C)(C)Cc(c12)c(Br)c(OC)cc2OC	57	4,2	0,46
c1cccc1C(c2cccc2)N3CCN(CC3)CC(O)COC(C)(c4c5cccc4)c(c56)c(OC)nc7c6cc(Br)cc7	58	5,23	0,48
c1cccc2c1C(=O)c(c23)c(Nc4ccncc4)nc5c3cc(Br)cc5	59	6,6	0,48
c1cccc1-c(cc2)ccc2C(=O)OC(Cn3ccnc3)c4ccc(cc4)-c5ccccc5	60	5,57	0,51
c1ccc2ccc3ccc(c4ccc1c2c34)CNCCNCCN(CCNC5)CCNCc(n 6)cccc56	61*	4,6	0,53
c1cccc(c12)sc(c2)C3(CCCCC3)N(CC4)CCN4Cc5c[nH]c(n5)- c6c(ccc(Br)c6)Sc7cccc7	62	5,37	0,53
c1cccc1CN(C2)C(C)(C)Cc(c23)cc(O)cc3OC	63	4,12	0,54
c1cccc2nc(c3c12)n(C)c4c(c3)ccc(c4)OC	64	6	0,54
OCC1(C)CCCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4	65	5,26	0,55
c1nccn1CC(O)COC(C)(c2c3cccc2)c(c34)c(OC)nc5c4cc(Br)cc5	66	5,67	0,57
c1ccc(OC)cc1C(N2)CC(O)Cc(c2c34)ccc3cccc4	67*	4,61	0,6
C1CCCCN1C(CO)c2c(ccc(F)c2)Oc(nc(c34)ccc(c4)[N+])([O-])=O)c(c3)Cc5ccccc5	68	5,7	0,61
c1ccc(OC)cc1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c 3)Cc5ccccc5	69	4,83	0,64
CC(C)(C)OC(=O)NC(C)C(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3 cc(Br)cc4)N5CCN(CC5)c6ccccn6	70	5,21	0,64
FC(F)(F)c1cc(ccc1)- c2ccn(n2)CC(O)COC(C)(c3c4cccc3)c(c45)c(OC)nc6c5cc(Br)cc 6	71	5,27	0,65
CC(C)(C)OC(=O)NC(C)(C)C(=O)O\N=C(c1c2cccc1)\c(c23) c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccn6	72*	5,1	0,66
c1cc(C)cc(C2)c1N(O3)C(CC23)c(cc4)ccc4C	73	4,34	0,66
c1cccc(c1C#N)N(CC2)CCN2C(=O)CN3CCN(CC3)c(c4C#N)ccc c4	74	4,59	0,69
c1cccc1Sc(ccc(Br)c2)c2- n(nn3)cc3CN4CCN(CC4)C5(CCCCC5)c(c6)sc(c67)cccc7	75	4,97	0,69
CCC(=O)N(C1)CCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)	76	4,19	0,72

C(CC4)CC(O)CC5			
c1cccc2ccc(c3c12)CC(O)CC(N3)c4cccc4	77*	4,71	0,75
c1cccc1CN(C2)C(C)(C)Cc(c23)c(l)ccc3OC	78	4,89	0,76
[O-][N+](=O)C(\C)=C\c1ccsc1	79	5,41	0,77
c1cccc2ccc(c3c12)CC(O)CC(N3)c4ccc(Br)cc4	80	5	0,82
c1nccn1CCCNc(nnc2Cl)c(c23)cccc3O	81*	4,68	0,85
COc(cc1)ccc1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	82	5,24	0,86
n1cncn1CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3	83	4,74	0,88
C1NC(C)(C)Cc(c12)c(l)ccc2OC	84	3,5	0,92
c1cccc2c1C(C)(OC(=O)N(C)C)c(c23)c(OC)nc4c3cc(Br)cc4	85*	5,13	0,94
CC(C)(C)OC(=O)NC(CCSC)C(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccn6	86	5,13	0,94
c1cccc1C(n(cn2)cc2[N+](O-)=O)c(c3)c(OC)nc(c34)ccc(Br)c4	87	5,16	0,96
c1cccc2c1C(=N/O)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccc cn6	88	5,15	0,96
c1cccc(c12)sc(c2)C3(CCCCC3)N(CC4)CCN4Cc5c[nH]c(n5)- c6c(Sc7cccc7)cccc6Sc8cccc8	89	5,63	0,96
c1cccc2ccc(c3c12)CC(O)CC(N3)c4c(Cl)cccc4	90*	4,97	0,96
c1nccn1CC(c2ccc(F)cc2)OC(=O)N3CCN(CC3)c(cc4)ccc4[N+](O-)=O	91	5,77	0,97
c1cccc(c12)cc3c(cccc3)c2CNCCCNCCN(CCNC4)CCNCc(n5)cc cc45	92	4,85	0,97
Nc1nc(N)nc(c12)ccc(c2)C#Cc3ccc(cc3)OCc4cccc4	93	6,3	0,99
c1ccc(OC)cc1NC(=O)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	94	5,21	1,01
c1cc(O)ccc1CCc(c2)ccc(c23)nc(N)nc3N	95*	7,11	1,02
c1ccc(OC)cc1- c2ccn(n2)C(CO)c3c(ccc(F)c3)Oc(nc(c45)ccc(c5)[N+](O-)=O)c(c4)Cc6cccc6	96	5,22	1,03
c1nccn1CC(c2ccc(F)cc2)OC(=O)c3ccc(cc3)-c4cccc4	97	6,15	1,04
FC(F)(F)c1ccc(cc1)N(CC2)CCN2C(=O)CN3CCN(CC3)c(cc4)cc c4C(F)(F)F	98	5,57	1,07
CC1=NC(C)(C)Cc(c12)c(c(OC)cc2OC)-c3ccc(cc3)OC	99*	4,97	1,08
[O-][N+](=O)C(\C)=C\c1cc(Br)cs1	100	6	1,1
OCc1ccc(cc1)-c(on2)cc2-c(c3)ccc(c34)OCO4	101	5,28	1,13
CCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	102*	6,96	1,13
[O-][N+](=O)c1nccn(n1)CC(=O)NC2CCN(CC2)Cc3cccc3	103	5,1	1,15
c1cccc1CN(C(C)(C)C2)Cc(c23)c(OC)cc(c3Br)OC	104	4,83	1,16
c12cccc(n2)CNCCN(CCNC1)CCNCc3cccc(n3)CNCCN(CCNC4) CCNCc(n5)cccc45	105	5,29	1,17
CCCCC(=O)OC(C)(c1c2cccc1)c(c23)c(- n4ccnc4)nc5c3cc(Br)cc5	106	5,84	1,17
[O-][N+](=O)c1nccn(n1)CC(=O)N2CCN(CC2)c3ncccn3	107	5	1,18
C1CN(CC2)C=C(C12)C#Cc3ccc(cc3)OCc4cccc4	108*	5,51	1,18
OCC1(C)C(O)CCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4	109	5,46	1,2
c1cccc(c12)cc3c(cccc3)c2CNCCN(CCNC4)CCNCc(n5)cccc45	110	4,89	1,22
[O-][N+](=O)C(\C)=C\c1ccc(s1)Br	111*	6,15	1,25
c1cc(Cl)ccc1C2CC(O3)Cc4ccc(c5c4C23)cccc5	112	4,79	1,25
c1cccc1CN(C(C)(C)C2)Cc(c23)c(OC)cc(c3Br)OC	113*	5,07	1,25
CCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	114	7,39	1,31
Nc1nc(N)nc(c12)ccc(c2)CCc3ccc(cc3)OCc4cccc4	115	7,23	1,31
[O-][N+](=O)\C=C\c(cc1)nc(c12)cccc2	116	6	1,34
[O-][N+](=O)c1nccn(n1)CC(=O)N2CCN(CC2)Cc3cc(Cl)c(Cl)cc3	117	6,53	1,39
CCCCCCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	118	6,82	1,4
CCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	119*	7,38	1,42
c1cccc(c12)nc(cc2)CNCCN(CCNC3)CCNCc(n4)cccc34	120	5,22	1,49
o1cccc1C(=O)N2CCN(CC2)c(cc3)ccc3NC(=O)OC(Cn4ccnc4)c5	121*	6,1	1,52

ccc(F)cc5			
c1ncn1CC2c(ccc(F)c2)Oc(c3)Cc4ccccc4)nc(c35)ccc(c5)- n(nn6)cc6-c7ccccc7	122	5,71	1,52
c1cccc2c1C(C)(OC(=O)C)c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	123	6,03	1,54
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c3ccccc3	124	5,33	1,57
[O-][N+](=O)c1ncn1CC(=O)Nc2ccc(cc2)Oc3ccc(F)cc3	125	5,48	1,57
c12cccc(n2)CNCCN(CCNC1)CCNCc3ccnc(c34)c5c(cc4)cccn5	126	5,28	1,59
c1cccc(c12)ncnc2CNCCN(CCNC3)CCNCc(c4)cccc34	127	5,49	1,66
[O-][N+](=O)c1ncn1CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3	128	5,59	1,69
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c3ccc(cc3)OC	129	5,62	1,74
c1cccc1\C=C\C(=O)N\N=C\c2c(O)cccc2	130*	5,18	1,8
c1cccc1C(n2ccnc2)c(c3)c(OC)nc(c34)ccc(Br)c4	131	6,34	1,83
s1c(Br)ccc1C(CC23)N(O3)c4c(C2)cc(Cl)cc4	132	4,62	1,84
CC(C)(C)[Si](C)(C)OC(CC1)CC(CC2)C1(C)C(CC3)C2C(C34C) CCC4C(C)(O)c5ccncc5	133	5,21	1,97
c1cccc1\C=C\C(=O)CC(c2ccccc2)C3(C)C(=O)OC(=N3)c4cccc c4	134	5,63	1,98
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(cc3)ccc3C	135	5,88	2,05
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(ccc2)cc2-c(cs3)nc3C	136*	6,26	2,09
c1ncn1C(CO)c2c(ccc(F)c2)Oc(nc(c34)ccc(c4)[N+](O-))=O)c(c3)Cc5ccccc5	137	6,6	2,1
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)N3CCCCC3	138	6,26	2,27
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(n2)scc2-c3cc(F)c(F)cc3	139	6,34	2,28
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(ccc2)cc2-c3ccccc3	140	6,74	2,4
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(cc3)ccc3C(F)(F)F	141	6,22	2,53
c1cccc1CC(Cn(n2)cnc2[N+](O-))=O)NS(=O)(=O)c(cc3)ccc3C	142	6,47	2,63
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(cc2)ccc2Cc3ccccc3	143*	7,17	2,72
[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)c(c3)ccc(Cl)c3Cl	144	6,77	2,81
c1cc(Cl)cc(F)c1C(c2cncnc2)N3CCN(CC3)c4ccccc4	145	7,11	2,91
[O-][N+](=O)c1ncn(n1)CC(C)(O)c2ccc(cc2)-c3ccccc3	146	6,99	2,93
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)c4ccc(Cl)cc4	147	7,62	2,98
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(n2)scc2-c3ccc(Cl)cc3	148	6,99	2,99
c1cccc1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	149	7,15	2,99
[O-][N+](=O)c1ncn(n1)CC(=O)c2ccc(cc2)-c3ccccc3	150	6,73	3,07
N#Cc1ccc(cc1)-c2ccc(cc2)NC(=O)Cn(n3)cnc3[N+](O-)=O	151	6,86	3,08
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccccc3	152*	7,19	3,15
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)c4ccccc4	153	7,47	3,29
CC(C)(C)OC(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3ccnc3	154	7,7	3,33
C1CCCCC1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	155*	7,55	3,35
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)OCc3ccccc3	156*	7,55	3,39
s1cnc(C)c1C(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	157	7,44	3,41
c1cc(Cl)ccc1NC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cncnc4	158	7,66	3,43
c1ncccc1C(c(cc2)c(F)cc2C(F)(F)F)N3CCN(CC3)S(=O)(=O)C	159	7,55	3,44
CC(C)Oc(cc1)ccc1C(c2ccnc2)N3CCN(CC3)c4c(OC)cccc4	160	7,55	3,44
[O-][N+](=O)c1ncn(n1)CC(=O)Nc(cc2)ccc2-c3ccc(Cl)cc3	161	7,35	3,45
c1ncccc1C(c(cc2)c(F)cc2C(F)(F)F)N3CCN(CC3)c(on4)nc4C	162*	7,77	3,48
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)c4c(OC)cccc4	163	7,7	3,48
CC(C)(C)C(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3ccnc3	164	7,72	3,5
c1cc(Cl)cc(F)c1C(c2ccnc2)N3CCN(CC3)c4ccccc4	165	7,74	3,51
c1cc(Cl)ccc1NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cncnc4	166	8	3,53
[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccc(Cl)cc3	167	8,1	3,56
[O-][N+](=O)c1ncn(n1)CC(O)(Cn2cncn2)c3ccc(cc3)-c4ccccc4	168	7,48	3,58
c1cc(Cl)cc(F)c1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)c4ccccc4	169	8,1	3,72
c1cc(Cl)ccc1C(c2ccnc2)N3CCN(CC3)S(=O)(=O)C4CCCC4	170	8	3,75
CC(C)c1ccc(cc1)S(=O)(=O)N(CC2)CCN2C(c3ccc(Cl)cc3)c4ccc nc4	171	7,92	3,82
c1cccc1OC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4ccnc4	172	8,1	3,85
c1cccc1NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4ccnc4	173	8,22	3,88

c1ccccc1OC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cccnc4	174*	8,3	4,07
Clc1c(Cl)ccc(c1)N(CC2)CCN2C(=O)CN3CCN(CC3)c(c4)ccc(Cl)c4Cl	175	2,74	-1,95
C1NCCCC1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(C)CC5	176	4,38	-1,23
n1ccccc1C(C)(O)C2CCC(C23C)C4C(CC3)C5(C)C(CC4)CC(O)CC5	177	4,48	-0,46
C1NCCCC1C(C(=O)OC(C)(C)C)O\N=C(c2c3cccc2)\c(c34)c(C(F)(F)F)nc5c4cc(Br)cc5	178	4,91	-0,32
c1ccccc1Sc2cccc(Sc3cccc3)c2-c([nH]4)nc4C[N+](C)(C)Cc5cnc([nH]5)-c6c(Sc7cccc7)cccc6Sc8cccc8	179*	5,2	-0,29
c1ccccc1Sc2cccc(c2CNC(=O)CN(C)C)Nc(c3)ccc(c34)nc5c(c4)c(c5)OC	180	3,76	-0,16
CC(C)(C1)NCc(c12)c(OC)cc(c2I)OC	181	2,96	-0,06
CN(C)Cc1cnc([nH]1)-c2c(Sc3cccc3)cccc2Sc4cccc4	182*	4,91	-0,04
COc(cc1)cc(c2)c1nc(c23)ccc(c3)Nc(c4CNC(=O)CN5CCN(C)CC5)cccc4Sc(c6)ccc(c67)cccc7	183	5,03	-0,04
c1ccccc1Sc2cccc(c2CNCCN(C)C)Nc(c3)ccc(c34)nc5c(c4)cc(cc5)OC	184	4,95	-0,01
c1ccc(Br)c(c2)c1n(C)c(c2c34)nc3cccc4	185	4,8	0,01
c1ccccc1Sc2cccc(c2CNCCN3CCN(C)CC3)Nc(c4)ccc(c45)nc6c(c5)cc(cc6)OC	186	4,56	0,05
COc(cc1)cc(c2)c1nc(c23)ccc(c3)Nc(c4CNC(=O)CN(C)C)cccc4Sc(c5)ccc(c56)cccc6	187	3,92	0,09
c1cc(Cl)cc(c2)c1n(C)c(c2c34)nc3cccc4	188	4,92	0,14
Cc(c1)ccc(c12)NC(CC(C2)O)c3c(Cl)cccc3	189	4,34	0,25
c1cccc(OC)c1NC(=O)NCC(O)COC(C)(c2c3cccc2)c(c34)c(OC)n c5c4cc(Br)cc5	190	5,6	0,33
COc(cc1)cc(c2)c1n(C)c(c2c34)nc3cccc4	191	5,77	0,37
c1cccc(c2)c1n(C)c(c2c34)nc3cccc4	192	5,4	0,44
O=C/C=C/c(cc1)nc(c12)cccc2	193	5,3	0,51
CCCC(=O)O\N=C(c1c2cccc1)\c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	194*	5,62	0,54
c1cccc(OC)c1NC(=S)Nc(c2)ccc(c23)nc(Oc(c4CCO)ccc(F)c4)c(c3)Cc5cccc5	195	5,36	0,57
c1ccccc1-c2cc(on2)-c3ccc(N)cc3	196	4,04	0,65
C1CN(CC2)CC(C12)(O)C#Cc3cc(ccc3)OCc4cccc4	197	5,02	0,67
c1csc(c1C)C(CC(C2)O)Nc(c23)ccc(Cl)c3	198*	4,38	0,68
CCCCCCCCOc(cc1)ccc1CCc(c2)ccc(c23)nc(N)nc3N	199	6,66	0,74
c1cccc2ccc(c3c12)CC(O)CC(N3)c4ccc(F)cc4	200	4,83	0,83
c1cccc2c1C(C)(O)c(c23)c(-n4ccnc4)nc5c3cc(Br)cc5	201	5,86	0,92
Nc1nc(N)nc(c12)ccc(c2)/C=C/c3ccc(cc3)OCc4cccc4	202	6,74	0,95
c1nccn1CC(c2ccc(Cl)cc2)OC(=O)N3CCN(CC3)c(cc4)ccc4[N+](O-)=O	203	6,06	0,97
CC(C)(C)OC(=O)NCC(=O)O\N=C(c1c2cccc1)\c(c23)c(nc4c3cc(Br)cc4)N5CCN(CC5)c6ccccn6	204*	5,17	0,97
c1cccc(c12)nc(cc2)CNCCCNCCN(CCNC3)CCNCc(n4)cccc34	205	5,08	1,02
[O-][N+](=O)C(\C)=C\c1csc(c12)cccc2	206	5,89	1,08
[O-][N+](=O)c1nccn1CC(=O)N2CCN(CC2)Cc(cc3)ccc3C(F)(F)F	207	6,06	1,12
[O-][N+](=O)c1nccn1CC(=O)Nc2ccc(cc2)Oc3cccc3	208	5,38	1,17
c1nccn1CC(c2ccc(Cl)cc2)OC(=O)c3ccc(cc3)-c4cccc4	209	6,32	1,18
CCOC(=O)c1ccc(cc1)C(C23)=CN(CC2)CC3	210	4,8	1,22
c12cccc(n2)CNCCN(CCNC1)CCNCc(c3ccc4c5c36)ccc6ccc5cc c4	211	5,38	1,3
c12cccc(n2)CNCCN(CCNC1)CCNCc(n3)ccc4ccc(c5c34)ccc(n5) CNCCN(CCNC6)CCNCc(n7)cccc67	212	5,44	1,42
n1c[nH]cc1CCNc(nnc2Cl)c(c23)cccc3O	213	5,13	1,46
c1cncnc1CNCCCNCCN(CCNC2)CCNCc(n3)cccc23	214*	5,55	1,54

<chem>OCC1(C)/C(=N\O)CCC2(C)C1CCC(C234)CC(C3)C(O)(CO)CC4</chem>	215	5,77	1,58
<chem>[O-][N+](=O)c1ncn(n1)CC(=O)Nc(n2)nc2-c3ccccc3</chem>	216*	5,6	1,83
<chem>c1cccc(c1C#N)N(CC2)CCN2C(=O)Cn(n3)cnc3[N+](O-)=O</chem>	217	5,55	1,87
<chem>n1cncn1CC(C)(O)c2ccc(cc2)-c3ccccc3</chem>	218	6,15	2,21
<chem>[O-][N+](=O)c1ncn(n1)CC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccc</chem>	219	7,14	2,3
<chem>c4n1ccncc1C(c(cc2)ccc2C(F)(F)F)N3CCN(CC3)c4c(OC)cccc4</chem>	220*	6,62	2,41
<chem>c1cccc1C(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccnc4</chem>	221*	7,17	2,95
<chem>c1cc(Cl)cc(F)c1C(c2cncnc2)N3CCN(CC3)c4c(OC)cccc4</chem>	222	7,19	2,97
<chem>[O-][N+](=O)c1ncn(n1)CC(=O)Nc2ccc(cc2)Oc3ccc(F)cc3</chem>	223	7,44	3,29
<chem>CC(C)(C)C(=O)N1CCN(CC1)C(c(cc2)ccc2C(F)(F)F)c3cccnc3</chem>	224	7,52	3,29
<chem>c1cc(Cl)cc(F)c1C(c2cccnc2)N3CCN(CC3)c4ncccn4</chem>	225	7,59	3,36
<chem>CC(C)(C)OC(=O)N1CCN(CC1)C(c2c(F)cc(Cl)cc2)c3cncnc3</chem>	226	7,62	3,43
<chem>CC(C)(C)C(=O)N1CCN(CC1)C(c(cc2)c(F)cc2C(F)(F)F)c3cccnc</chem>	227	7,62	3,48
<chem>3CC(C)(C)OC(=O)N1CCN(CC1)C(c(cc2)ccc2C(F)(F)F)c3cccnc3</chem>	228*	7,77	3,52
<chem>c1cc(Cl)ccc1C(c2cccnc2)N3CCN(CC3)S(=O)(=O)c4ccccc4</chem>	229*	8,05	3,68
<chem>c1cccc1NC(=O)N2CCN(CC2)C(c3ccc(Cl)cc3)c4cccnc4</chem>	230*	8,05	3,87
<chem>CC(C)c1ccc(cc1)NC(=O)N2CCN(CC2)C(c3c(F)cc(Cl)cc3)c4cnc</chem>	231	8,15	3,95
<chem>nc4</chem>			

* Conjunto teste

ANEXO C – Representação na forma SMILES dos 260 compostos obtidos através da triagem virtual em larga escala.

Tabela C1 – Energia de ligação na cruzaina obtido por docagem molecular (em kcal.mol⁻¹).

SMILES	ZINC	Estrutura	goldscore	Gscore XP
<chem>N#Cc1c(C)nc(C)cc1N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F</chem>	92964267	1	-7,66	-7,38
<chem>c1cc(Cl)cc(Cl)c1NC(=O)N(Cc2cccnc2)CCN3CCOCC3</chem>	20423620	2	-7,78	-5,52
<chem>c1cc(F)ccc1C[C@H](n2nnnc2C)C(=O)NCC3(CCC3)c4ccc(Cl)cc4</chem>	72087376	3	-7,87	-5,53
<chem>c1ncccc1CC(=O)NCc2c(nccc2)Oc(cc3C(F)(F)F)ccc3</chem>	12449018	4	-8,22	-6,80
<chem>o1cccc1CN(C[C@H]2CCCO2)C(=O)Nc3cn(C)c(=O)c(c34)cccc4</chem>	6932434	5	-7,74	-5,59
<chem>c1cccc(F)c1[C@H]2CN(CCO2)C(=O)Nc(s3)nnnc3-c4ccsc4</chem>	67280610	6	-7,76	-5,55
<chem>c1cc(Cl)ccc1CCNC(=O)[C@H]2CCCN(C2)C(=NC=N3)[C@H](C3=4)N=CN4</chem>	63492379	7	-8,08	-6,07
<chem>c1ncccc1CNC(=O)CSc(n2)nc(C(F)(F)F)c(c2c34)CCc3cccc4</chem>	646627	8	-8,44	-6,11
<chem>N#CC[C@H](C)Sc(cccc1)c1NC(=O)N(CCO2)C[C@H]2c3c(F)cccc3</chem>	66956124	9	-8,06	-5,68
<chem>o1cccc1CCC(=O)Nc(cc2)cc(c2F)S(=O)(=O)N3CCOCC3</chem>	44567226	10	-8,07	-5,70
<chem>C1COCCN1c(ccc(Cl)c2)c2NC(=O)N(C)CCOc3cccc3</chem>	71901309	11	-8,18	-6,04
<chem>n1cn(C)c(Cl)c1S(=O)(=O)Nc2c(cccc2)N3COCC3</chem>	75581011	12	-8,27	-6,04
<chem>FC(F)(F)c1cc(ccc1)Cn(c(n2)CCCC)c(c23)nc(C)c(c3)N</chem>	4906567	13	-8,15	-6,56
<chem>CC(=O)c1cc(ccc1)Nc2c(C(=O)OCC)cnc(c23)ccc(F)c3</chem>	956542	14	-8,49	-6,43
<chem>n1ncn(c12)nc(s2)NC(=O)CCCOc(cc3)ccc3C(C)(C)C</chem>	3459546	15	-7,63	-6,42
<chem>COc(cc1)ccc1CNC(=O)C2(CC2)c3cc(on3)-c4c(F)cccc4</chem>	56919886	16	-8,06	-5,99
<chem>FC(F)n1ccnc1CN(C)C(=O)NCc2ncc(o2)-c3cccc3</chem>	79517712	17	-7,75	-6,11
<chem>CCc1nnc(s1)NC(=O)N(CC)C[C@H](O2)COc(c23)cccc3</chem>	43344165	18	-7,80	-5,89
<chem>CC(C)(C)Oc(ncc1)cc1CNC(=O)[C@H](C2=O)[C@H](N2C)c3cccnc3</chem>	92119304	19	-7,77	-5,62
<chem>CCc1ccc(cc1)CNC(=O)CCc2c(noc2)-c3c(F)cccc3</chem>	75510592	20	-7,96	-5,61
<chem>c1cc(F)cc(F)c1CC(=O)N[C@H](C)c2nc(c s2)-c3ccncc3</chem>	87335960	21	-8,48	-6,28
<chem>CCc(o1)ccc1C(=O)Nc2c(cccc2)N(C[C@H]3C)C[C@H](O3)C</chem>	55685381	22	-7,89	-5,83
<chem>c1ncccc1CN(CCC#N)C(=O)Nc(ccc2)c(c23)ncccc3</chem>	80779799	23	-8,40	-6,42
<chem>c1cc(F)ccc1CNC(=O)[C@H](C(C)C)Sc(n2)nn(c23)cccn3</chem>	30718569	24	-8,26	-6,35
<chem>c1cccc(c12)CN(CC2)C(=O)[C@H](C(C)C)Sc(o3)nnnc3C(C)(C)C</chem>	77247970	25	-7,77	-5,72
<chem>CCc(cc1)ccc1[C@H](C)NC(=O)[C@H]</chem>	40130295	26	-7,88	-5,48

2CCCN(C2)c3ccc(Cl)nn3				
C1CCCCC1OCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F	93052953	27	-8,00	-6,28
c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCCOC[C@@H]3CCCO3	67386210	28	-7,49	-8,22
c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCCOC[C@H]3CCCO3	67386217	29	-7,45	-7,81
C[C@@H]1CN(C[C@@H](O1)C)c(cccc2)c2NC(=O)N(C)Cc3ncnn3C	72509269	30	-7,26	-6,17
c1cc(Cl)ccc1NC(=O)Cn(n2)c(C)cc2C(=O)N3CCOCC3	11927649	31	-8,12	-5,89
COC[C@@H]1CN(CCO1)C(=O)NCC2(CC2)c3cc(OC)ccc3	74142415	32	-8,29	-6,62
C1OCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F	93053722	33	-7,35	-6,39
C1OCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F	93053726	34	-7,63	-5,83
C1CC1c(n2C)nnn2CN(C)C(=O)Nc3c(F)ccc(c3)-n4cccc4	81067125	35	-7,91	-5,57
CCc1nc(on1)CN(C)C(=O)NCCc2cnn(c2)-c3ccccc3	69762537	36	-8,10	-6,28
FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)Cn3nnnc3C(C)C	91788667	37	-8,02	-5,69
CCCc1nnc(s1)NC(=O)[C@@H](C)c(c2)ccc(c23)OCCO3	68650118	38	-8,07	-5,72
CC(=O)c1ccc(nc1)N(C2)CCC[C@H]2Cn(n3)cc3CCO	23478649	39	-7,85	-5,84
C1CCn(c12)c(nn2)CNC(=O)CCc3ccc(o3)-c4cccs4	20869699	40	-7,94	-5,70
Cc(n1)sc(C)c1[C@H](C)N(C)C(=O)CCc2nnc(o2)-c3ccccc3	12302511	41	-7,19	-6,26
C1COCCN1C(=O)Nc2cc(ccc2)OCc3ccc(F)cc3	27343935	42	-7,59	-5,97
n1ccccc1CNC(=O)[C@@H]2CCCN(C2)c3ncnc(c34)sc(c4)CC	33078299	43	-7,92	-6,68
c1cc(F)c(F)cc1[C@H](C)N(C)C(=O)NCc(n2)n(c23)ccc3	79265262	44	-7,47	-5,91
C1CCCN1[C@H]2CCN(C2)C(=O)NCc3c(nccc3)OC4CCCC4	66178874	45	-8,13	-6,03
n1nc(C)n(c12)CCC[C@H]2NC(=O)CCc3nc(c(o3)-c4ccccc4	79594184	46	-7,55	-5,84
Cc(c1)cc(C)c(c12)ncc(C#N)c2N(CC3)CCC3OC[C@@H]4CCCCO4	81031040	47	-7,99	-6,62
Cc(c1)cc(C)c(c12)ncc(C#N)c2N(CC3)CCC3OC[C@H]4CCCCO4	81031042	48	-7,85	-6,35
C1CC[C@H]([C@H]1C)NC(=O)C[C@@H](C)NC2=NS(=O)(=O)c(c23)cccc3	80069368	49	-8,10	-5,70
C1OCC[C@]1(C(=O)OC)NC(=O)c2cc(ccc2)-c(cs3)nc3C	88220932	50	-7,99	-6,01
CC(C)(C)c1nc(on1)CCC(=O)N[C@H]2CCOc(c23)cccc3	33159238	51	-8,19	-5,98
CC(C)(C)c1nc(on1)CCC(=O)N[C@@H]2CCOC(c23)cccc3	33159237	52	-7,97	-5,98
O1COc(c12)ccc(c2)NC(=O)[C@H](C)N(C)Cc3c(F)cccc3Cl	3457326	53	-8,88	-6,69
c1cc(C)ccc1CNC(=O)CCc2nc(no2)-c3cc(F)ccc3	72332141	54	-7,98	-5,75
CC(C)(O)CN(C)C(=O)Cc1csc(n1)-c(c2)ccc(c23)OCO3	80822767	55	-7,93	-6,01

c1cccc(OC)c1C[C@H](C)N(C)C(=O)Nc2cncc(c23)cccc3	89357203	56	-7,48	-6,30
n1cccc(OC(C)C)c1C(=O)N2C[C@H](OC[C@H]2C)c(c3C)cc(C)cc3	79531837	57	-8,24	-5,89
c1cccc(c1CC)NC(=O)CCC(=O)N2CCC[C@H]2c3nccs3	19146445	58	-7,63	-6,20
O1CCN(C1=O)CCC(=O)N(CC2)CCC2OCc3ccc(F)cc3	92076090	59	-7,33	-6,05
CC(C)c1nc(on1)CCC(=O)N2CCN(CC2)c3cnccc3	48105992	60	-7,43	-6,95
c1ccc(F)c(F)c1[C@H](C2)[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956312	61	-7,90	-6,58
n1oc(C)nc1CNC(=O)N(CCc(c23)cccc2)[C@H]3Cc4cccc4	90593151	62	-7,59	-5,95
c1cc(Cl)ccc1CCNC(=O)Cc2csc(n2)N3CCC3=O	8133311	63	-7,95	-5,84
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(cc(F)c3)-c4cncnc4	11974390	64	-7,79	-5,80
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(cc(F)c3)-c4cncnc4	11974396	65	-8,01	-6,06
c1cccc(Cl)c1CNC(=O)CCn(c2=O)cnc(c23)sc(C)c3C	40024984	66	-7,76	-6,02
c1cc(Cl)cc(Cl)c1[C@H](C)NC(=O)CCn(c2=O)cnc(c23)scc3	12749371	67	-7,50	-6,26
COc1nccc(c1C#N)N[C@H](CC2=O)CN2CCc3cccc3	72153248	68	-7,13	-6,19
CC(C)n(nc1)c(C)c1C(=O)OCC(=O)N[C@H]2CCCc(c23)cccc3	92631569	69	-7,53	-6,65
n1[nH]c(C)nc1CCC(=O)N(C)Cc2csc(n2)-c3cccc3	72454569	70	-8,51	-6,48
c1c(F)ccc(F)c1[C@H](C2CC2)NC(=O)N(C)Cc3cnc(s3)CC	74893708	71	-7,84	-6,15
FC(F)(F)c1c(Cl)ccc(c1)NC(=O)[C@H](C2=O)CN2Cc3ccco3	7306435	72	-7,32	-6,08
Cc(n1)scc1CC(=O)N(C2)CCC[C@H]2c(c3)nc(C)cc3Cc4c(F)cccc4	72354119	73	-7,94	-6,46
CCS(=O)(=O)N(CC1)CCC1C(=O)NCCc(c2)ccc(c23)OCC3	74616895	74	-7,44	-5,92
CC(C)(C)NC(=O)CCSc(o1)nnc1-c2cccc(c23)cccc3	59188606	75	-8,34	-6,16
C1OCC[C@H]1CN(CC)C(=O)CCc2nc(no2)-c3cccc3	48263350	76	-7,75	-6,21
CC(C)(C)c1ccc(cc1)OC[C@H](O)Cn2c(C)cc(=O)n(c23)ncn3	45097288	77	-7,76	-5,60
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c23)cc(Cl)cc3	3448391	78	-7,89	-5,98
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c23)cc(Cl)cc3	3448394	79	-8,03	-5,80
CCOC(=O)[C@H]1CCCN(C1)C(=O)Nc2c(c3)ccc2OC3CCCC3	79682507	80	-7,93	-5,54
c1ccnc(OC)c1CNS(=O)(=O)c2cccc(c23)cncc3	77615954	81	-7,99	-5,63
CC(=O)c1cnc(s1)NC(=O)C2(CCCC2)c3c(F)cc(F)cc3	89811381	82	-7,38	-6,30
Cc(n1)scc1CCNC(=O)N(C)Cc(c2)ccc(c23)OCO3	44967753	83	-8,01	-5,67
c1cccc(OC)c1COC[C@H](O)CN2CCOc(c23)cccc3	60857051	84	-7,31	-7,00
C1OCC[C@H]1CN(C)C(=O)NC[C@H](C)c	89717833	85	-8,18	-6,36

2nc(no2)-c3ccccc3				
C1OCC[C@ @H]1CN(C)C(=O)NC[C@H](C)c2nc(no2)-c3ccccc3	89717835	86	-8,19	-6,17
C1CCCCC1(C#N)NC(=O)[C@ @H](C)Sc(c2)nc(c23)cccc3	5627281	87	-7,44	-5,88
c1cccc1OCC(=O)N(C2)CCC[C@ @H]2C(=O)Nc(cc3C)cnc3	79577108	88	-7,39	-6,23
c1cc(F)ccc1[C@ @H]2CN(CCO2)C(=O)NCCn3ccnc3	42896953	89	-7,98	-6,89
Cc1cn(nn1)CCNC(=O)CCSc(c2)ccc(c23)OCCO3	65528008	90	-7,87	-5,74
s1ccn(c12)cc(n2)CC(=O)NCc3cc(ccc3)OCc4ccccc4	59411307	91	-8,29	-6,24
C1CCCCC1NC(=O)COC(=O)c2cc(ccc2)Oc3ccccc3	3393368	92	-7,77	-6,26
c1cnc(OC)cc1CNC(=O)N(C2)CCC[C@ @H]2c(n3)oc(c34)cccc4	45494692	93	-8,06	-5,95
s1ccc(C)c1C(=O)[C@ @H]2CCCN(C2)C(=O)Nc(c(cc3)OC)cc3C	39393772	94	-8,05	-5,81
s1ccc(C)c1C(=O)[C@H]2CCCN(C2)C(=O)Nc(c(cc3)OC)cc3C	39393770	95	-7,71	-5,60
FC(F)COc(nc1)cc1CNC(=O)Cc2cccc(c23)cccn3	91046021	96	-7,56	-6,18
c1ccc(Cl)cc1[C@ @H](C)NC(=O)[C@H](C)C2(=O)CN2Cc3ccccc3	31601820	97	-8,11	-5,88
c1cccc(c1C(F)(F)F)CN(C2CC2)C(=O)NCCc3ncn(n3)C	92377367	98	-7,78	-5,52
CO[C@H]1CCCN(C1)C(=O)c2cc(ccc2)S(=O)(=O)NCC3CC3	76245437	99	-8,51	-6,39
CCCOc(nc1)cc1CNC(=O)N2CCC[C@ @H]2CN3CCOCC3	73875207	100	-7,70	-5,62
CCCCNC(=O)[C@H](C#N)c(n1)c(nc(c12)c(ccc2)N(C3)CCC[C@ @H]3C	4837255	101	-8,11	-5,94
c1cc(Cl)cc(Br)c1OCC(=O)NCc2ccc(o2)-c3csc(n3)C	12904725	102	-8,07	-5,76
C1CCCC1NC(=O)Cc2nc(on2)Cc3ccc(cc3)OCC	72552497	103	-7,81	-5,61
c1cccc(c12)N(C(=O)CC2)CC(=O)NCCc(n3)sc(c34)cccc4	74718345	104	-7,64	-6,68
COC[C@ @H](C)NC(=O)CCn(c(c12)nc1)ccc2)c(c1)-c3ccncc3	19456616	105	-7,58	-5,90
c1cccc(c12)oc(c2)[C@ @H](C)C(=O)NCc3cc(ncc3)OCC(F)F	91045925	106	-7,86	-5,51
c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@ @H](C)Oc3ccc(Cl)cc3	81348380	107	-7,69	-5,94
Cc1onc(CC)c1C(=O)NCCCO[C@ @H]2CCCc(c23)cccc3	70874454	108	-7,80	-6,29
c1cncc(c12)cccc2NC(=O)CCC(=O)N(C3)CCc(c34)sc4	54849997	109	-7,71	-6,30
c1cccc(c12)N(CC2)C(=O)NCCOCc3ccc(cc3)OC	93201872	110	-7,87	-6,17
c1ccnc(OC)c1C(=O)N(C2)CCC[C@ @H]2CCCc3ccc(F)cc3	71749519	111	-8,63	-6,69
O1CCOc(c12)ccc(c2)CN(C)C(=O)NCCCOc(c3C)cccc3	78518847	112	-7,71	-5,75
CCOC(=O)c1cnnc(c1)[C@ @H]2CCCN(C2)C(=O)NCC3(CC)CCC3	91701517	113	-7,86	-5,95
Cc1c(Cl)c(C)n(n1)CC(=O)NCc2ncc(o2)-c3ccccc3	80338882	114	-8,08	-5,91

s1cccc1[C@H](CC)NC(=O)N(CCO2)C[C@H]2c3cc(F)c(F)cc3	67406599	115	-7,43	-6,91
s1cccc1[C@@H](CC)NC(=O)N(CCO2)C[C@H]2c3cc(F)c(F)cc3	67406602	116	-7,41	-6,75
n1cccc1CN(CC)C(=O)NCc2cn(nn2)-c3cccc3	87166849	117	-7,75	-5,81
N#CC[C@H](C)N(C)C(=O)c1csc(n1)-c(c2)ccc(c23)OCO3	78420281	118	-7,82	-5,48
n1cccn1CC2(CC2)C(=O)N(C3)CCC[C@@H]3c(n4)[nH]c(c45)cccc5C	91321262	119	-7,84	-5,67
c1sccc1[C@H](C)C(=O)N[C@@H](C)c2cc(ccc2)N(C3=O)CCO3	78090111	120	-7,84	-5,60
COCCOc(cccc1)c1NC(=O)N(C)[C@H](C)c(c2)oc(c23)cccc3	66403521	121	-7,86	-5,74
COCCOc(cccc1)c1NC(=O)N(C)[C@H](C)c(c2)oc(c23)cccc3	66403524	122	-7,89	-6,84
c1cc(C)ccc1CNC(=O)CCc2ncc(o2)-c3ccc(F)cc3	12929356	123	-7,96	-6,04
FC(F)(F)[C@@]1(O)CCN(C1)C(=O)CCCc2ncc(o2)-c3cccc3	77989783	124	-7,96	-6,65
FC(F)(F)[C@]1(O)CCN(C1)C(=O)CCCc2ncc(o2)-c3cccc3	77989787	125	-7,73	-6,58
CCOC(=O)CSc(c1C#N)nc(c(c12)CCCC2)N3CCOCC3	903744	126	-7,57	-7,47
CC(C)NC(=O)COC(=O)c1nn(c(c12)CCC2)-c3c(F)cccc3	46509084	127	-8,20	-6,86
c1cnccc1[C@H](C)N(C)C(=O)NCCCO[C@@H]2CCCc(c23)cccc3	73221057	128	-8,30	-6,30
c1cccc1CCC(=O)N(C2)CC2c3nc(no3)-c4ccsc4	75156512	129	-8,52	-6,30
Cn(c1)nc(CC)c1CNC(=O)N2C[C@@H](OC[C@H]2C)c(c3C)cccc3	90824539	130	-7,95	-5,62
Cc1onc(C)c1CN(C)C(=O)N[C@@H](C)c(c2)oc(c23)c(OC)ccc3	79157214	131	-8,05	-5,67
c1ccc(C)c(c1C)NC(=O)CNC(C)(C)c(c2)ccc(c23)OCO3	64344145	132	-7,11	-6,70
CCc1nc(on1)CN(C)C(=O)NCc2csc(c23)ccc3	69772036	133	-8,27	-6,22
c1cccn(c12)c(nn2)[C@H](C)NC(=O)N(C)Cc(s3)ncc3CC	83175027	134	-7,71	-5,62
CC(C)c1nc([nH]n1)SCC(=O)N(C)C[C@@H](O2)COc(c23)cccc3	14110875	135	-7,90	-5,98
CC(C)CCO[C@H]1CCN(C1)C(=O)NCc2cn(nn2)-c3cccc3	87166632	136	-7,94	-5,75
CCc(n1)scc1CC(=O)NCCc2ccc(cc2)-n3cccn3	65520577	137	-7,80	-6,04
c1nc(C)nc(c12)CCC[C@@H]2NC(=O)COc(c3)ccc(c34)cccc4	72242521	138	-7,54	-6,10
o1cccc1C(=O)C[C@H]2COCCN2C(=O)c3n[nH]c(c34)cccc4	95439902	139	-7,37	-5,92
c1cccc(c12)OCC[C@@H]2C(=O)N[C@H](C)COc3ccc(F)cc3	76798231	140	-7,72	-6,08
c1cc(F)c(F)cc1[C@@H](C(=O)OC)NC(=O)[C@@H](C2)Sc(c23)cccc3	89364439	141	-7,83	-5,50
c1ncccc1C(=O)NCc2c(cc(C)cc2)OC[C@@H]3CCOC3	87455925	142	-7,81	-5,55
C1CCC[C@H](C)[C@H]12N([C@@H](C2)C(=O)OC)C(=O)CCNC(=O)c3cccc3	55448395	143	-7,78	-5,58
Cc1onc(C)c1CN(C)C(=O)NCCOc2cccc(c2	56960730	144	-8,01	-5,94

3)cccc3				
COc1cc(ccc1)NC(=O)N2CCC[C@H]2c (n3)sc(c34)cccc4	76128493	145	-7,81	-6,46
c1nccn1CCCNC(=O)N2CCC[C@H]2c(n 3)sc(c34)cccc4	42896682	146	-7,79	-5,99
c1cccc(c12)[C@H](CCC2(C)C)CNS(=O)(=O)CC3CCOCC3	93939051	147	-7,33	-6,19
s1ccn(c12)cc(n2)CNC(=O)N(C)CCOc(c3C) cccc3	78984628	148	-7,86	-5,83
c1ccc(F)c(F)c1CCC(=O)N[C@H]2CCCN(c2 3)nc(n3)CC	88197817	149	-8,00	-6,24
COc(cc1)cc(Cl)c1NC(=O)N(C)Cc(n2)sc(c2 3)cccc3	68561409	150	-7,80	-5,62
O1COc(c12)ccc(c2)CN(C)C(=O)NCCCc(n3)sc(c34)CCCC4	66204557	151	-8,10	-5,71
c1cc(C)ccc1CNC(=O)CCc2nnc(o2)- c3c(C)occ3	40756727	152	-7,78	-5,52
O1CCC[C@H]1Cn2c(SCC(=O)NC(C)C)nc c2-c3cccc3	8709934	153	-8,19	-5,86
O1CCC[C@H]1CNC(=O)Cc2csc(n2)SCc 3c(F)cccc3	8939348	154	-7,83	-5,73
c1cccc1C(=O)C[C@H]2COCCN2C(=O)C Cc3cscn3	95412325	155	-8,12	-6,17
c1cccc1C(=O)C[C@H]2COCCN2C(=O) CCc3cscn3	95412324	156	-7,83	-5,79
O1COc(c12)ccc(c2)CN(CC)C(=O)Nc3cccc(c34)cncc4	43162190	157	-8,25	-6,01
c1cccc(c12)OC[C@H](C)N2C(=O)c3ccc(=O)n(n3)CCCC	72532639	158	-7,97	-5,86
OCCOC(=O)c1cc(- c2cccc2)c(s1)N3CCOCC3	26413486	159	-7,61	-6,93
c1cccc(F)c1C(=O)C[C@H](C)C(=O)NCCC(n2)cn(c23)cccc3	79400264	160	-7,82	-5,97
c1cccc(c12)onc2CC(=O)NCc3cc(ncc3)OC 4CCC4	90013735	161	-8,71	-6,87
c1ccnc(c12)c(ccc2)CC(=O)NCCCO[C@@ H]3CCCC(c34)cccc4	72821379	162	-7,77	-6,07
c1cccc1[C@H](O2)CN(C[C@H]2C)C(=O)NCCOC3CCOCC3	74542092	163	-7,40	-7,06
n1cccc1CCNC(=O)N(CCO2)C[C@H]2c3c cc(Br)cc3	67589176	164	-7,82	-6,29
s1cccc1CN(CC=C)C(=O)N[C@H](C)c2nc(no2)C3CCCC3	80567565	165	-8,00	-5,72
c1ccn(C)c(=O)c1NC(=O)CCC(=O)N(C2)C Cc(c23)scc3	74073007	166	-7,62	-5,88
Cc1ccc(cc1)- c(nc(s2)C)c2CC(=O)Nc3ccnn3CCC	75598283	167	-8,42	-6,25
Cc1csc(=O)n1CC(=O)N(C2)CC2OCc3cccc (c34)cccc4	80941753	168	-7,48	-6,59
C1CC1c2ncc(cn2)NC(=O)N(C3)CC[C@H] 3Oc4cc(Cl)ccc4	90479947	169	-8,08	-6,01
C1CCCCC1OCC(=O)NC[C@H](C)c2nc(no2)-c3cccc3	89955773	170	-8,23	-5,94
c1ncccc1OCCCC(=O)N[C@H]2CCCOc(c23)cccc3	79880561	171	-8,73	-6,76
c1cccc(c12)OC[C@H](C2)C(=O)N[C@H](C)c(c3)oc(c34)c(OC)ccc4	74509023	172	-7,74	-6,72
c1ccc(F)cc1CN(CC)C(=O)c(c2)cnc(c23)[n H]nc3C	47382594	173	-7,49	-6,04

C1CCCCN1C(=O)c2cc(ccc2)NC(=O)N(C)C c3ncnn3C	72509551	174	-7,77	-5,58
n1ccncc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3))CCOC	92841199	175	-8,28	-6,18
o1nc(C)nc1COCC(=O)N[C@H](C2CCC2)c 3ccc(Cl)cc3	74566760	176	-8,24	-5,92
c1cccc(c12)OCC[C@H]2CC(=O)NCc(o3)c cc3C(=O)OCC	79586889	177	-7,56	-6,47
c1cc(C)c(C)cc1CNC(=O)Cn(cn2)cc2- c(n3)sc3CCC	65250562	178	-8,27	-6,10
c1cccc(COC)c1CNC(=O)N(C)Cc(n2)sc(c23))CCCC3	83177089	179	-7,58	-5,88
c1cccc(c12)OCC[C@H]2NC(=O)Cc3coc(n3)-c4cccc4	32207533	180	-7,69	-6,13
C1COCCN1C(=O)CSc(n2C[C@H]3CCC O3)ncc2-c4cccc4	11641712	181	-7,71	-6,70
C1COCCN1C(=O)CSc(n2C[C@H]3CCCO 3)ncc2-c4cccc4	11641718	182	-8,03	-6,78
c1cc(Cl)ccc1C[C@H](C)CC(=O)N[C@@ H]2CCCN(c23)nc(n3)C(C)C	92958887	183	-7,96	-5,84
N#Cc1cc(c(F)cc1)CNC(=O)N(CCO2)C[C@ H]2c3cccc3	75435037	184	-7,67	-5,66
c1cccc(c12)oc(=O)n2CCC(=O)N(C)CC(=O))Nc3cccc(C)c3C	7312811	185	-8,27	-5,88
C1OCC[C@H]1CN(C)C(=O)Nc(ccc2)cc2C OCc3ccco3	78493097	186	-7,72	-5,84
c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@H] (C)Oc3ccc(F)cc3	80760513	187	-7,72	-5,94
c1c(F)ccc(c12)CCC[C@H]2NC(=O)CCc 3nnc(o3)C4CCCC4	93061622	188	-7,84	-5,61
C1CCCC=C1CCNC(=O)Cn(c2=O)c(C)nc(c 23)onc3C	35507406	189	-7,44	-5,90
n1nc(Cl)ccc1SCC(=O)N[C@H](C)c(c2)ccc(c23)cccc3	10653606	190	-8,24	-6,17
c1nccn1C[C@H](C)NC(=O)N(C[C@H]2C) C[C@H](O2)c3ccc(F)cc3	80570599	191	-7,68	-5,93
c1ccnc(c12)c(ccc2)CC(=O)NCc3cc(ncc3)O C(C)(C)C	92531571	192	-7,96	-6,34
O1CCC[C@H]1CNC(=O)CCCc2nc(no2)- c(cc3)ccc3C	4433832	193	-7,84	-5,92
n1oc(C)cc1CN(C)C(=O)CCc2ccc(o2)- c3c(F)cccc3	25137544	194	-8,07	-5,75
c1cccc(OC)c1CNC(=O)CN2C(=O)CSc(c23))nccc3	6802489	195	-8,08	-6,06
O1CCC[C@H]1C(=O)NCc(n2)n(c(c23)cccc 3)Cc4c(C)ccc(C)c4	40197222	196	-7,67	-5,62
C1COCCC1OCCC(=O)Nc(n2)sc(c23)cccc 3OC	84520887	197	-7,67	-5,73
CN1C[C@H](CC1=O)c(n2)n(c(c23)cccc3))CCOc4cccc(c45)cccc5	40250743	198	-7,82	-6,60
CCc(n1)sc1CC(=O)N(CCO2)C[C@H]2c3c cc(F)cc3	55196763	199	-7,47	-6,52
c1cccc1C[C@H](C(=O)C)NC(=O)CSc(n 2)n(C)c(c23)ccc(Cl)c3	44049212	200	-7,84	-6,28
c1cc(Cl)ccc1CCNC(=O)COc2ncnc(c2c34)s c3CCC4	3599992	201	-8,00	-6,14
OCC[C@H]1COCCN1C(=O)c2c(OCC)cn (n2)-c3cccc3	92452672	202	-7,67	-6,02
O1CCC[C@H]1CNC(=O)CCc2ncc(o2)-	24924387	203	-8,37	-6,76

c(c3)ccc(C)c3C				
O1CCC[C@H]1CNC(=O)CCc2ncc(o2)-c(c3)ccc(C)c3C	24924391	204	-7,97	-6,01
Cc(o1)nc(C)c1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3	73959717	205	-7,68	-6,06
C1COCCC1OCCC(=O)Nc(ccc2)cc2-c(n3)oc(C)c3C	84530501	206	-8,25	-6,11
c1cccc(c12)CCC[C@H]2NC(=O)COC(=O)c3cccc(n3)OC	83939506	207	-8,50	-6,63
O1CCOc(c12)ccc(c2)CN(C)C(=O)N[C@H]3CCCN(c34)nc(n4)CC	87249506	208	-7,32	-6,31
COC[C@H]1CN(CCO1)c2c(C(=O)OCC)nc(c23)ccc(C)c3	78342925	209	-6,53	-7,55
O1CCC[C@H]1C(=O)NCCc(n2)n(c(c23)cccc3)Cc4c(C)cccc4	40198730	210	-8,06	-5,74
C1COCCN1C(=O)[C@H](CC)NC(=O)Cc2c oc(c23)cc4c(c3)CCC4	84223098	211	-8,04	-6,75
O1COc(c12)ccc(c2)CN[C@H](C)C(=O)N[C@H]3CCCc(c34)cccc4	59818114	212	-8,78	-6,44
s1ccnc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3)CCOC	92841424	213	-7,99	-6,11
C1CC1[C@H](O)CN(C)C(=O)c2nc(C(C)C)n(n2)-c3c(F)cccc3	87561073	214	-7,58	-6,26
CCOc(cc1)ccc1CC(=O)N[C@H](C)c(c2)oc(c23)c(OC)ccc3	25137832	215	-8,32	-6,07
COc(cc1)ccc1CC(=O)N[C@H](C)c(c2)oc(c23)c(ccc3)OCC	57908442	216	-8,35	-6,59
Cc1nc([nH]n1)SCC(=O)N2CCOc(c23)cc(cc3)N(C)C	74191425	217	-8,00	-6,10
n1ccncc1CN(C)C(=O)N[C@H](C)c2cc(cc2)-n3ccnc3	80551568	218	-7,83	-5,52
c1cc(F)ccc1[C@H](C[C@H]2C)CN2C(=O)Nc3ccn(n3)CCOC	77440429	219	-8,14	-5,97
n1occc1CC(=O)N[C@H](c2ccccc2)Cc3cc(Cl)cc3	81282781	220	-7,94	-5,54
Clc1cccc(F)c1CC(=O)NCCn(n2)c(=O)n(c23)cccc3	58244146	221	-7,28	-6,14
C1CO[C@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3	92414547	222	-7,60	-6,16
C1CO[C@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3	92414544	223	-7,50	-5,83
c1cccc(c12)CCC[C@H]2C(=O)NCc3ccc(nc3)N(CC)CC	40121323	224	-8,74	-6,52
N#Cc1cc(ccc1)-c2ncn2CCc3ccc(Cl)cc3	88521299	225	-7,46	-5,84
o1cccc1-c(no2)nc2CCC(=O)N[C@H](CCC3)[C@H]3Cc4ccccc4	78150926	226	-7,83	-6,02
n1ccccc1S[C@H](C)C(=O)NCc2cc(ccc2)-n3ccnc3	84977059	227	-8,23	-5,93
CC1(C)CN(CCO1)C(=O)Nc2cc(ccc2)OC[C@H]3CCCO3	78270310	228	-7,76	-5,62
n1cccc(O)c1C(=O)N(C2)CCC[C@H]2CCc3ccc(F)cc3	71735509	229	-7,17	-6,18
c1cc(Br)ccc1O[C@H](C)C(=O)NCc2c(nccc2)N3CCOCC3	31326556	230	-7,80	-5,54
c1cccc(c12)oc(c2)[C@H](C)NC(=O)COc3c(OC)cccc3OC	22314745	231	-8,25	-6,35
c1cc(F)ccc1OCC(=O)NCCc(n2)sc(c23)ccc c3	40323351	232	-7,91	-5,76

c1cccn(c12)cc(n2)CNC(=O)CCCN3c(=O)oc(c34)cccc4	7944046	233	-7,65	-5,75
CC(C)(C)OC(=O)N(C1)C[C@H](CO)c(c12)nn(c2)CC3CCCC3	91302842	234	-7,77	-6,17
n1ccccc1CNC(=O)[C@@H](CC2=O)CN2c(c3)ccc(C)c3C	9668971	235	-8,05	-6,02
Cc(o1)ccc1[C@@H](C)NC(=O)CN2C(=O)[C@H](C)Oc(c23)cccc3	23306977	236	-8,01	-5,74
c1ccccc1[C@@H]2CN(CCO2)C(=O)NCc3c(CC)nn(C)c3OC	93149538	237	-7,86	-5,53
COCc1cc(ncn1)NC(=O)N2C[C@@H](C[C@H]2C)c3ccc(F)cc3	93316218	238	-7,82	-5,88
Cn1nnnc1S[C@H](C)C(=O)N[C@@H](C)c(c2)ccc(c23)CCCC3	22784584	239	-8,25	-6,27
COCc1cc(ncn1)NC(=O)N2C[C@H](C[C@H]2C)c3ccc(F)cc3	93316214	240	-8,00	-5,80
n1oc(C)cc1CN(C)C(=O)CCc2cnn(c2)-c3cccc3	31657942	241	-7,83	-5,56
CCCN(n1)cc(Cl)c1C(=O)N(C2)CCC[C@@H]2Cc3ccccn3	72384932	242	-7,82	-5,88
s1ccnc1CN(C)C(=O)NCc2cc(ccc2)-n3cccn3	80696288	243	-7,88	-5,56
FC(F)(F)c1cc(ccc1)[C@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3	92036660	244	-7,57	-6,50
FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3	92036658	245	-7,58	-6,03
c1cccc(c12)sc(n2)CN(C)C(=O)N[C@@H]3CCN(c34)nc(n4)CC	87251107	246	-7,68	-5,77
CC(C)Cn(c1=O)c(=O)n(C)c(c12)nc3n2c(C)c(C)n3CCOC	4314324	247	-7,96	-6,77
c1cccc1C(=O)N(C2)CCC[C@H]2C(=O)OCCc3cccn3	73299335	248	-7,54	-5,89
COc1n(C)nc(CC)c1CN[C@H]2CCOc(c23)c(Cl)ccc3Cl	89211609	249	-7,36	-6,37
Clc1cnn(c1)CCNC(=O)CCN(nc2)c(c23)ccc(Cl)c3	40760451	250	-7,48	-6,01
CCN(CC)C(=O)CSc(o1)nnc1-c(c2)ccc(c23)cccc3	6416097	251	-7,93	-5,67
Cn(n1)cnc1CNC(=O)N(C2)CC[C@@H]2Oc3c(Cl)cccc3	91775455	252	-8,12	-5,80
C1COCCC1OCCC(=O)Nc2ccc(OC)c(c23)nccc3	84532176	253	-7,25	-6,53
C1COCCN1CCNC(=O)Cc2coc(c23)ccc4c3cccc4	21874011	254	-7,85	-6,25
c1c(Cl)ccc(Cl)c1NC(=O)[C@H](C)Sc(n2CC)nnc2C3CC3	7618357	255	-8,06	-5,72
C[C@@H](C1)[C@H]1C(=O)N2CCN(CC2)C(=O)NCCC(c3)oc(c34)cccc4	74380190	256	-8,04	-5,81
Cc(n1)sc1CCCC(=O)N(C2)CC[C@@H]2Oc3cc(Cl)ccc3	89624334	257	-7,91	-6,16
Cc(n1)sc1CCCC(=O)N(C2)CC[C@H]2Oc3cc(Cl)ccc3	89624337	258	-7,87	-5,89
c1cccc(c12)OCC[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956457	259	-7,94	-6,62
c1cccc(c12)OCC[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956458	260	-7,75	-6,51

Tabela C2 – Predição de pIC₅₀ na forma amastigota de *Trypanosoma cruzi*.

SMILES	ZINC	Estrutura	RNA	KPLS molprint2D	RL
<chem>N#Cc1c(C)nc(C)cc1N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F</chem>	92964267	1	6,20	5,79	6,77
<chem>c1cc(Cl)cc(Cl)c1NC(=O)N(Cc2cccn2)CCN3CCOCC3</chem>	20423620	2	6,80	6,79	7,29
<chem>c1cc(F)ccc1C[C@H](n2nnnc2C)C(=O)NCC3(CCC3)c4ccc(Cl)cc4</chem>	72087376	3	6,32	5,86	6,67
<chem>c1ncccc1CC(=O)NCc2c(nccc2)Oc(cc3C(F)(F)F)ccc3</chem>	12449018	4	6,15	5,61	6,70
<chem>o1cccc1CN(C[C@H]2CCCO2)C(=O)Nc3cn(C)c(=O)c(c34)cccc4</chem>	6932434	5	6,34	5,83	6,97
<chem>c1cccc(F)c1[C@H]2CN(CCO2)C(=O)Nc(s3)nnc3-c4ccsc4</chem>	67280610	6	6,28	5,97	6,28
<chem>c1cc(Cl)ccc1CCNC(=O)[C@H]2CCCN(C2)C(=NC=N3)[C@H](C3=4)N=CN4</chem>	63492379	7	5,66	5,44	6,35
<chem>c1ncccc1CNC(=O)CSc(n2)nc(C(F)(F)F)c(c2c34)CCc3cccc4</chem>	646627	8	6,15	6,31	6,68
<chem>N#CC[C@H](C)Sc(cccc1)c1NC(=O)N(CCO2)C[C@H]2c3c(F)ccc3</chem>	66956124	9	6,34	6,04	6,46
<chem>o1cccc1CCC(=O)Nc(cc2)cc(c2F)S(=O)(=O)N3CCOCC3</chem>	44567226	10	5,78	6,12	6,62
<chem>C1COCCN1c(ccc(Cl)c2)c2NC(=O)N(C)CCOc3cccc3</chem>	71901309	11	6,33	5,86	6,47
<chem>n1cn(C)c(Cl)c1S(=O)(=O)Nc2c(ccc2)N3CCOCC3</chem>	75581011	12	6,21	5,58	5,77
<chem>FC(F)(F)c1cc(ccc1)Cn(c(n2)CCCC)c(c23)nc(C)c(c3)N</chem>	4906567	13	6,56	5,67	6,53
<chem>CC(=O)c1cc(ccc1)Nc2c(C(=O)OC)Cnc(c23)ccc(F)c3</chem>	956542	14	5,60	5,58	6,42
<chem>n1ncn(c12)nc(s2)NC(=O)CCCOc(c3)ccc3C(C)(C)C</chem>	3459546	15	5,55	5,29	6,17
<chem>COc(cc1)ccc1CNC(=O)C2(CC2)c3cc(on3)-c4c(F)cccc4</chem>	56919886	16	5,95	5,36	6,54
<chem>FC(F)n1ccnc1CN(C)C(=O)NCc2nc(c(o2)-c3cccc3</chem>	79517712	17	6,00	5,36	6,50
<chem>CCc1nnc(s1)NC(=O)N(CC)C[C@H](O2)COc(c23)cccc3</chem>	43344165	18	5,01	5,57	6,04
<chem>CC(C)(C)Oc(ncc1)cc1CNC(=O)[C@H](CC2=O)[C@H](N2C)c3ccncc3</chem>	92119304	19	6,46	5,89	6,63
<chem>CCc1ccc(cc1)CNC(=O)CCc2c(noc2)-c3c(F)cccc3</chem>	75510592	20	6,08	5,67	6,25
<chem>c1cc(F)cc(F)c1CC(=O)N[C@H](C)c2nc(cs2)-c3ccncc3</chem>	87335960	21	6,35	5,85	6,46
<chem>CCc(o1)ccc1C(=O)Nc2c(cccc2)N(C[C@H]3C)C[C@H](O3)C</chem>	55685381	22	5,74	5,16	5,73
<chem>c1ncccc1CN(CCC#N)C(=O)Nc(ccc2)c(c23)nccc3</chem>	80779799	23	5,88	6,23	5,95
<chem>c1cc(F)ccc1CNC(=O)[C@H](C(C)C)Sc(n2)nn(c23)cccn3</chem>	30718569	24	5,97	5,61	6,38
<chem>c1cccc(c12)CN(CC2)C(=O)[C@H](C(C)C)Sc(o3)nnc3C(C)(C)C</chem>	77247970	25	5,73	5,30	6,18
<chem>CCc(cc1)ccc1[C@H](C)NC(=O)[C@H]2CCCN(C2)c3ccc(Cl)nn3</chem>	40130295	26	5,80	5,65	6,37
<chem>C1CCCCC1OCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F</chem>	93052953	27	5,90	5,42	6,34

<chem>c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCCCOC[C@@H]3CCCCO3</chem>	67386210	28	6,32	6,16	6,88
<chem>c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCCCOC[C@@H]3CCCCO3</chem>	67386217	29	6,32	6,16	6,88
<chem>C[C@@H]1CN(C[C@@H](O1)C)c(cccc2)c2NC(=O)N(C)Cc3ncnn3C</chem>	72509269	30	6,23	5,91	6,11
<chem>c1cc(Cl)ccc1NC(=O)Cn(n2)c(C)cc2C(=O)N3CCOCC3</chem>	11927649	31	6,29	6,17	6,39
<chem>COC[C@@H]1CN(CCO1)C(=O)NCC2(CCCC2)c3cc(OC)ccc3</chem>	74142415	32	5,81	5,42	5,97
<chem>C1OCCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F</chem>	93053722	33	6,11	5,85	6,39
<chem>C1OCCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc(nc3)ccc3C(F)(F)F</chem>	93053726	34	6,11	5,85	6,39
<chem>C1CC1c(n2C)nnc2CN(C)C(=O)Nc3c(F)ccc(c3)-n4cccc4</chem>	81067125	35	6,57	6,25	6,81
<chem>CCc1nc(on1)CN(C)C(=O)NCCc2cnn(c2)-c3cccc3</chem>	69762537	36	5,71	5,31	5,99
<chem>FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)Cn3nnnc3C(C)C</chem>	91788667	37	5,46	5,68	6,19
<chem>CCCc1nnc(s1)NC(=O)[C@@H](C)c(c2)ccc(c23)OCCO3</chem>	68650118	38	5,10	5,06	5,88
<chem>CC(=O)c1ccc(nc1)N(C2)CCC[C@H]2Cn(nn3)cc3CCO</chem>	23478649	39	5,62	5,42	6,02
<chem>C1CCn(c12)c(nn2)CNC(=O)CCc3cc(o3)-c4cccs4</chem>	20869699	40	6,50	5,45	6,04
<chem>Cc(n1)sc(C)c1[C@H](C)N(C)C(=O)CCc2nnc(o2)-c3cccc3</chem>	12302511	41	5,90	5,14	6,07
<chem>C1COCCN1C(=O)Nc2cc(ccc2)OCc3ccc(F)cc3</chem>	27343935	42	5,63	6,60	5,88
<chem>n1cccc1CNC(=O)[C@@H]2CCCCN(C2)c3ncnc(c34)sc(c4)CC</chem>	33078299	43	6,36	5,61	6,23
<chem>c1cc(F)c(F)cc1[C@H](C)N(C)C(=O)NCc(nn2)n(c23)cccc3</chem>	79265262	44	6,11	6,06	6,42
<chem>C1CCCN1[C@H]2CCN(C2)C(=O)NCc3c(nc3)OC4CCCC4</chem>	66178874	45	6,74	5,75	6,18
<chem>n1nc(C)n(c12)CCC[C@H]2NC(=O)CCc3ncc(o3)-c4cccc4</chem>	79594184	46	6,05	5,46	5,91
<chem>Cc(c1)cc(C)c(c12)ncc(C#N)c2N(C3)CCC3OC[C@H]4CCCCO4</chem>	81031040	47	6,43	5,74	6,42
<chem>Cc(c1)cc(C)c(c12)ncc(C#N)c2N(C3)CCC3OC[C@H]4CCCCO4</chem>	81031042	48	6,43	5,74	6,42
<chem>C1CC[C@H]([C@H]1C)NC(=O)C[C@@H](C)NC2=NS(=O)(=O)c(c23)cccc3</chem>	80069368	49	5,84	5,37	6,24
<chem>C1OCC[C@]1(C(=O)OC)NC(=O)c2cc(ccc2)-c(cs3)nc3C</chem>	88220932	50	5,56	5,26	6,29
<chem>CC(C)(C)c1nc(on1)CCC(=O)N[C@H]2CCOC(c23)cccc3</chem>	33159238	51	5,06	5,20	6,00
<chem>CC(C)(C)c1nc(on1)CCC(=O)N[C@H]2CCOC(c23)cccc3</chem>	33159237	52	5,06	5,20	6,00
<chem>O1COc(c12)ccc(c2)NC(=O)[C@H](C)N(C)Cc3c(F)ccc3Cl</chem>	3457326	53	6,87	6,00	6,75
<chem>c1cc(C)ccc1CNC(=O)CCc2nc(no2)-c3cc(F)ccc3</chem>	72332141	54	5,74	5,29	6,18
<chem>CC(C)(O)CN(C)C(=O)Cc1csc(n1)-c(c2)ccc(c23)OCO3</chem>	80822767	55	5,35	5,14	6,14
<chem>c1cccc(OC)c1C[C@H](C)N(C)C(=</chem>	89357203	56	6,03	5,71	5,97

O)Nc2cncc(c23)cccc3					
n1cccc(OC(C)C)c1C(=O)N2C[C@H](OC[C@H]2C)c(c3C)cc(C)cc3	79531837	57	5,98	5,54	6,64
c1cccc(c1CC)NC(=O)CCC(=O)N2CCC[C@H]2c3nccs3	19146445	58	6,51	5,48	6,00
O1CCN(C1=O)CCC(=O)N(CC2)C CC2OCc3ccc(F)cc3	92076090	59	6,27	5,61	6,28
CC(C)c1nc(on1)CCC(=O)N2CCN(CC2)c3ccncc3	48105992	60	6,20	5,49	5,94
c1ccc(F)c(F)c1[C@H](C2)[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956312	61	6,27	5,49	6,35
n1oc(C)nc1CNC(=O)N(CCc(c23)cc cc2)[C@H]3Cc4cccc4	90593151	62	6,06	5,64	6,05
c1cc(Cl)ccc1CCNC(=O)Cc2csc(n2) N3CCCC3=O	8133311	63	6,36	5,70	6,05
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(cc(F)c3)-c4cnnc4	11974390	64	5,62	5,39	6,32
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(cc(F)c3)-c4cnnc4	11974396	65	5,62	5,39	6,32
c1cccc(Cl)c1CNC(=O)CCn(c2=O)c nc(c23)sc(C)c3C	40024984	66	6,21	5,47	6,24
c1cc(Cl)cc(Cl)c1[C@@H](C)NC(=O)CCn(c2=O)cnc(c23)scc3	12749371	67	6,12	5,66	6,52
COc1nccc(c1C#N)N[C@H](CC2=O)CN2CCCc3cccc3	72153248	68	6,16	5,41	5,99
CC(C)n(nc1)c(C)c1C(=O)OCC(=O)N[C@H]2CCCc(c23)cccc3	92631569	69	6,08	5,52	6,00
n1[nH]c(C)nc1CCC(=O)N(C)Cc2cs c(n2)-c3cccc3	72454569	70	6,31	5,41	5,98
c1c(F)ccc(F)c1[C@@H](C2CC2)N C(=O)N(C)Cc3cnc(s3)CC	74893708	71	6,42	5,57	6,17
FC(F)(F)c1c(Cl)ccc(c1)NC(=O)[C@H](CC2=O)CN2Cc3ccco3	7306435	72	6,58	5,93	6,47
Cc(n1)scc1CC(=O)N(C2)CCC[C@H]2c(c3)nc(C)cc3Cc4c(F)cccc4	72354119	73	6,55	5,71	6,50
CCS(=O)(=O)N(CC1)CCC1C(=O)N CCc(c2)ccc(c23)OCC3	74616895	74	6,30	5,46	6,13
CC(C)(C)NC(=O)CCSc(o1)nnc1-c2cccc(c23)cccc3	59188606	75	5,89	5,22	5,79
C1OCC[C@H]1CN(CC)C(=O)C Cc2nc(no2)-c3cccc3	48263350	76	5,92	5,53	6,05
CC(C)(C)c1ccc(cc1)OC[C@H](O)Cn2c(C)cc(=O)n(c23)ncn3	45097288	77	5,39	5,27	5,96
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c23)cc(Cl)cc3	3448391	78	5,47	5,60	6,08
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c23)cc(Cl)cc3	3448394	79	5,47	5,60	6,08
CCOC(=O)[C@H]1CCCN(C1)C(=O)Nc2c(cccc2)OC3CCCC3	79682507	80	5,20	5,52	5,81
c1ccnc(OC)c1CNS(=O)(=O)c2cccc (c23)cncc3	77615954	81	5,39	5,90	5,72
CC(=O)c1cnc(s1)NC(=O)C2(CCC C2)c3c(F)cc(F)cc3	89811381	82	5,89	5,48	5,92
Cc(n1)scc1CCNC(=O)N(C)Cc(c2)c cc(c23)OCO3	44967753	83	5,85	5,23	5,98
c1cccc(OC)c1COC[C@H](O)CN 2CCOc(c23)cccc3	60857051	84	5,78	5,16	5,60
C1OCC[C@H]1CN(C)C(=O)NC[C@H](C)c2nc(no2)-c3cccc3	89717833	85	5,36	5,68	5,86

C1OCC[C@ @H]1CN(C)C(=O)NC[C@H](C)c2nc(no2)-c3ccccc3	89717835	86	5,36	5,68	5,86
C1CCCCC1(C#N)NC(=O)[C@ @H](C)Sc(cc2)nc(c23)cccc3	5627281	87	5,95	5,23	5,78
c1cccc1OCC(=O)N(C2)CCC[C@ @H]2C(=O)Nc(cc3C)cnc3	79577108	88	6,34	5,53	6,20
c1cc(F)ccc1[C@ @H]2CN(CCO2)C(=O)NCCCN3ccnc3	42896953	89	6,59	6,07	6,01
Cc1cn(nn1)CCNC(=O)CCSc(c2)cc(c23)OCCO3	65528008	90	5,77	5,25	6,14
s1ccn(c12)cc(n2)CC(=O)NCc3cc(c3)OCc4ccccc4	59411307	91	6,65	5,50	6,01
C1CCCCC1NC(=O)COC(=O)c2cc(ccc2)Oc3ccccc3	3393368	92	5,55	5,24	5,92
c1cnc(OC)cc1CNC(=O)N(C2)CCC[C@ @H]2c(n3)oc(c34)cccc4	45494692	93	5,34	5,62	6,10
s1ccc(C)c1C(=O)[C@ @H]2CCCN(C2)C(=O)Nc(c(cc3)OC)cc3C	39393772	94	5,83	5,45	6,17
s1ccc(C)c1C(=O)[C@H]2CCCN(C2)C(=O)Nc(c(cc3)OC)cc3C	39393770	95	5,83	5,45	6,17
FC(F)COc(ncc1)cc1CNC(=O)Cc2c(ccc(c23)cccn3	91046021	96	6,06	5,59	6,46
c1ccc(Cl)cc1[C@ @H](C)NC(=O)[C@H](CC2=O)CN2Cc3ccccc3	31601820	97	6,83	5,73	6,42
c1cccc(c1C(F)(F)F)CN(C2CC2)C(=O)NCc3ncn(n3)C	92377367	98	6,40	5,54	6,06
CO[C@H]1CCCN(C1)C(=O)c2cc(c2)S(=O)(=O)NCC3CC3	76245437	99	6,13	5,65	6,10
CCCOc(nccc1)c1CNC(=O)N2CCC[C@ @H]2CN3CCOCC3	73875207	100	6,12	5,72	5,94
CCCCNC(=O)[C@H](C#N)c(n1)c(n1)c(c12)cccc2)N(C3)CCC[C@ @H]3C	4837255	101	5,51	5,15	6,06
c1cc(Cl)cc(Br)c1OCC(=O)NCc2ccc(o2)-c3csc(n3)C	12904725	102	5,51	5,32	5,80
C1CCCC1NC(=O)Cc2nc(on2)Cc3c(cc3)OCC	72552497	103	5,26	5,29	5,72
c1cccc(c12)N(C(=O)CC2)CC(=O)NCCc(n3)sc(c34)cccc4	74718345	104	6,58	5,52	5,97
COC[C@ @H](C)NC(=O)CCn(c(c12)nccc2)c(c1)-c3ccncc3	19456616	105	6,44	5,37	5,85
c1cccc(c12)oc(c2)[C@ @H](C)C(=O)NCc3cc(ncc3)OCC(F)F	91045925	106	5,64	5,52	6,25
c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@ @H](C)Oc3ccc(Cl)cc3	81348380	107	5,80	5,59	6,36
Cc1onc(CC)c1C(=O)NCCCO[C@ @H]2CCCC(c23)cccc3	70874454	108	5,82	5,47	5,67
c1cncc(c12)cccc2NC(=O)CCC(=O)N(C3)CCc(c34)scc4	54849997	109	6,61	5,75	6,05
c1cccc(c12)N(CC2)C(=O)NCCOCC3ccc(cc3)OC	93201872	110	5,88	5,37	5,54
c1ccnc(OC)c1C(=O)N(C2)CCC[C@ @H]2CCc3ccc(F)cc3	71749519	111	6,14	5,52	5,93
O1CCOc(c12)ccc(c2)CN(C)C(=O)NCCCOc(c3C)cccc3	78518847	112	5,82	5,21	6,22
CCOC(=O)c1cnn(c1)[C@ @H]2CCCN(C2)C(=O)NCC3(CC)CCC3	91701517	113	5,95	5,73	5,69
Cc1c(Cl)c(C)n(n1)CC(=O)NCc2ncc(o2)-c3ccccc3	80338882	114	6,02	5,19	5,86

s1cccc1[C@H](CC)NC(=O)N(CCO2)C[C@H]2c3cc(F)c(F)cc3	67406599	115	6,33	5,80	6,34
s1cccc1[C@@H](CC)NC(=O)N(CCO2)C[C@H]2c3cc(F)c(F)cc3	67406602	116	6,33	5,80	6,34
n1cccc1CN(CC)C(=O)NCc2cn(nn2)-c3cccc3	87166849	117	5,99	5,82	5,64
N#CC[C@@H](C)N(C)C(=O)c1csc(n1)-c(c2)ccc(c23)OCO3	78420281	118	5,51	4,89	5,63
n1cccn1CC2(CC2)C(=O)N(C3)CC[C@H]3c(n4)[nH]c(c45)cccc5C	91321262	119	6,39	5,39	5,74
c1sccc1[C@H](C)C(=O)N[C@@H](C)c2cc(ccc2)N(C3=O)CCO3	78090111	120	6,19	5,34	5,93
COCCOc(cccc1)c1NC(=O)N(C)[C@@H](C)c(c2)oc(c23)cccc3	66403521	121	5,61	5,71	5,97
COCCOc(cccc1)c1NC(=O)N(C)[C@@H](C)c(c2)oc(c23)cccc3	66403524	122	5,61	5,71	5,97
c1cc(C)ccc1CNC(=O)CCc2ncc(o2)-c3ccc(F)cc3	12929356	123	6,22	5,35	5,68
FC(F)(F)[C@@]1(O)CCN(C1)C(=O)CCc2ncc(o2)-c3cccc3	77989783	124	5,96	5,12	6,08
FC(F)(F)[C@]1(O)CCN(C1)C(=O)CCc2ncc(o2)-c3cccc3	77989787	125	5,96	5,12	6,08
CCOC(=O)CSc(c1C#N)nc(c(c12)C CCC2)N3CCOCC3	903744	126	5,76	5,22	5,92
CC(C)NC(=O)COC(=O)c1nn(c(c12)CCC2)-c3c(F)cccc3	46509084	127	5,67	5,45	5,99
c1cnccc1[C@H](C)N(C)C(=O)NCCO[C@@H]2CCc(c23)cccc3	73221057	128	6,65	5,94	6,07
c1cccc1CCC(=O)N(C2)CC2c3nc(no3)-c4ccsc4	75156512	129	6,15	5,49	6,03
Cn(c1)nc(CC)c1CNC(=O)N2C[C@@H](OC[C@H]2C)c(c3C)cccc3	90824539	130	6,32	5,56	6,15
Cc1onc(C)c1CN(C)C(=O)N[C@@H](C)c(c2)oc(c23)c(OC)ccc3	79157214	131	4,92	5,44	6,05
c1ccc(C)c(c1C)NC(=O)CNC(C)(C)c(c2)ccc(c23)OCO3	64344145	132	6,15	5,93	5,77
CCc1nc(on1)CN(C)C(=O)NCc2csc(c23)cccc3	69772036	133	5,86	5,27	5,65
c1cccn(c12)c(nn2)[C@H](C)NC(=O)N(C)Cc(s3)ncc3CC	83175027	134	6,02	5,72	6,22
CC(C)c1nc([nH]n1)SCC(=O)N(C)C[C@@H](O2)COc(c23)cccc3	14110875	135	5,50	5,42	6,09
CC(C)CCO[C@H]1CCN(C1)C(=O)NCc2cn(nn2)-c3cccc3	87166632	136	5,99	5,92	5,76
CCc(n1)sc1CC(=O)NCCc2ccc(cc2)-n3cccn3	65520577	137	6,42	5,79	5,89
c1nc(C)nc(c12)CCC[C@@H]2NC(=O)COc(c3)ccc(c34)cccc4	72242521	138	6,14	5,38	5,98
o1cccc1C(=O)C[C@H]2COCCN2C(=O)c3n[nH]c(c34)cccc4	95439902	139	5,79	5,18	6,04
c1cccc(c12)OCC[C@@H]2C(=O)N[C@H](C)COc3ccc(F)cc3	76798231	140	5,88	5,32	5,70
c1cc(F)c(F)cc1[C@@H](C(=O)OC)NC(=O)[C@@H](C2)Sc(c23)cccc3	89364439	141	6,08	5,44	6,38
c1ncccc1C(=O)NCc2c(cc(C)cc2)OC[C@H]3CCOC3	87455925	142	6,02	5,39	5,83
C1CCC[C@@H]([C@H]12)N([C@@H](C2)C(=O)OC)C(=O)CCNC(=O)c3cccc3	55448395	143	5,43	5,16	5,94

<chem>Cc1onc(C)c1CN(C)C(=O)NCCOc2cccc(c23)cccc3</chem>	56960730	144	5,84	5,20	5,68
<chem>COCc1cc(ccc1)NC(=O)N2CCC[C@H]2c(n3)sc(c34)cccc4</chem>	76128493	145	6,36	6,04	5,90
<chem>c1nccn1CCCN(C(=O)N2CCC[C@H]2c(n3)sc(c34)cccc4</chem>	42896682	146	6,53	5,73	6,03
<chem>c1cccc(c12)[C@H](CCC2(C)C)CN S(=O)(=O)CC3CCOCC3</chem>	93939051	147	6,16	5,73	5,78
<chem>s1ccn(c12)cc(n2)CNC(=O)N(C)CC Oc(c3C)cccc3</chem>	78984628	148	6,30	5,42	5,70
<chem>c1ccc(F)c(F)c1CCC(=O)N[C@H]2CCN(c23)nc(n3)CC</chem>	88197817	149	6,17	5,57	6,38
<chem>COc(cc1)cc(Cl)c1NC(=O)N(C)Cc(n2)sc(c23)cccc3</chem>	68561409	150	5,73	5,51	5,96
<chem>O1COc(c12)ccc(c2)CN(C)C(=O)N CCCc(n3)sc(c34)CCCC4</chem>	66204557	151	6,25	5,23	5,93
<chem>c1cc(C)ccc1CNC(=O)CCc2nnc(o2)-c3c(C)occ3</chem>	40756727	152	5,40	5,20	5,93
<chem>O1CCC[C@H]1Cn2c(SCC(=O)NC(C)C)ncc2-c3cccc3</chem>	8709934	153	6,41	5,53	5,73
<chem>O1CCC[C@H]1CNC(=O)Cc2csc(n2)SCc3c(F)cccc3</chem>	8939348	154	6,36	5,83	5,96
<chem>c1cccc1C(=O)C[C@H]2COCCN2C(=O)CCc3cscn3</chem>	95412325	155	6,21	5,42	5,99
<chem>c1cccc1C(=O)C[C@H]2COCCN2C(=O)CCc3cscn3</chem>	95412324	156	6,21	5,42	5,99
<chem>O1COc(c12)ccc(c2)CN(CC)C(=O)Nc3cccc(c34)cncc4</chem>	43162190	157	5,75	5,63	5,73
<chem>c1cccc(c12)OC[C@H](C)N2C(=O)c3ccc(=O)n(n3)CCCC</chem>	72532639	158	5,73	5,00	5,98
<chem>OCCOC(=O)c1cc(-c2cccc2)c(s1)N3CCOCC3</chem>	26413486	159	5,81	5,42	5,49
<chem>c1cccc(F)c1C(=O)C[C@H](C)C(=O)NCCc(n2)cn(c23)cccc3</chem>	79400264	160	6,73	5,65	6,36
<chem>c1cccc(c12)onc2CC(=O)NCc3cc(ncc3)OC4CCC4</chem>	90013735	161	5,63	5,17	5,62
<chem>c1ccnc(c12)c(ccc2)CC(=O)NCCCO[C@H]3CCc(c34)cccc4</chem>	72821379	162	6,74	5,91	6,11
<chem>c1cccc1[C@H](O2)CN(C[C@H]2C)C(=O)NCCCOC3CCOCC3</chem>	74542092	163	6,43	6,00	6,16
<chem>n1cccc1CCNC(=O)N(CCO2)C[C@H]2c3ccc(Br)cc3</chem>	67589176	164	5,90	5,66	5,90
<chem>s1cccc1CN(CC=C)C(=O)N[C@H](C)c2nc(no2)C3CCCC3</chem>	80567565	165	5,90	5,32	5,80
<chem>c1ccn(C)c(=O)c1NC(=O)CCC(=O)N(C2)CCc(c23)scc3</chem>	74073007	166	6,56	5,87	6,32
<chem>Cc1ccc(cc1)-c(nc(s2)C)c2CC(=O)Nc3ccnn3CC C</chem>	75598283	167	6,30	5,24	5,71
<chem>Cc1csc(=O)n1CC(=O)N(C2)CC2OCc3cccc(c34)cccc4</chem>	80941753	168	6,36	5,29	6,02
<chem>C1CC1c2ncc(cn2)NC(=O)N(C3)CC[C@H]3OC4cc(Cl)ccc4</chem>	90479947	169	5,70	5,90	5,85
<chem>C1CCCCC1OCC(=O)NC[C@H](C)c2nc(no2)-c3cccc3</chem>	89955773	170	5,26	5,12	5,66
<chem>c1ncccc1OCCCC(=O)N[C@H]2CCCOc(c23)cccc3</chem>	79880561	171	5,86	5,54	5,81
<chem>c1cccc(c12)OC[C@H](C2)C(=O)N[C@H](C)c(c3)oc(c34)c(OC)ccc4</chem>	74509023	172	5,67	5,24	5,99

<chem>c1ccc(F)cc1CN(CC)C(=O)c(c2)cnc(c23)[nH]nc3C</chem>	47382594	173	6,24	5,15	5,86
<chem>C1CCCCN1C(=O)c2cc(ccc2)NC(=O)N(C)Cc3ncnn3C</chem>	72509551	174	6,28	5,43	6,18
<chem>n1ccncc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3)CCOC</chem>	92841199	175	6,23	6,13	5,84
<chem>o1nc(C)nc1COCC(=O)N[C@H](C2CCC2)c3ccc(Cl)cc3</chem>	74566760	176	5,34	5,67	5,84
<chem>c1cccc(c12)OCC[C@H]2CC(=O)Nc(c3)ccc3C(=O)OCC</chem>	79586889	177	5,27	5,20	5,56
<chem>c1cc(C)c(C)cc1CNC(=O)Cn(cn2)cc2-c(n3)sc3CC</chem>	65250562	178	6,42	5,45	5,67
<chem>c1cccc(COC)c1CNC(=O)N(C)Cc(n2)sc(c23)CCCC3</chem>	83177089	179	6,30	5,34	5,77
<chem>c1cccc(c12)OCC[C@H]2NC(=O)Cc3coc(n3)-c4cccc4</chem>	32207533	180	6,01	5,18	5,71
<chem>C1COCCN1C(=O)CSc(n2C[C@@H]3CCCO3)ncc2-c4cccc4</chem>	11641712	181	6,73	5,62	6,11
<chem>C1COCCN1C(=O)CSc(n2C[C@H]3CCCO3)ncc2-c4cccc4</chem>	11641718	182	6,73	5,62	6,11
<chem>c1cc(Cl)ccc1C[C@@H](C)CC(=O)N[C@@H]2CCCN(c23)nc(n3)C(C)C</chem>	92958887	183	6,18	5,66	6,09
<chem>N#Cc1cc(c(F)cc1)CNC(=O)N(CCO2)C[C@H]2c3cccc3</chem>	75435037	184	5,95	5,77	5,70
<chem>c1cccc(c12)oc(=O)n2CCC(=O)N(C)CC(=O)Nc3cccc(C)c3C</chem>	7312811	185	6,00	5,81	6,65
<chem>C1OCC[C@H]1CN(C)C(=O)Nc(ccc2)cc2COCC3CCCO3</chem>	78493097	186	6,02	6,35	5,75
<chem>c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@@H](C)Oc3ccc(F)cc3</chem>	80760513	187	5,87	5,54	6,15
<chem>c1c(F)ccc(c12)CCC[C@H]2NC(=O)CCc3nnc(o3)C4CCCC4</chem>	93061622	188	5,80	5,45	6,13
<chem>C1CCCC=C1CCNC(=O)Cn(c2=O)c(C)nc(c23)onc3C</chem>	35507406	189	5,14	5,18	5,94
<chem>n1nc(Cl)ccc1SCC(=O)N[C@H](C)c(c2)ccc(c23)cccc3</chem>	10653606	190	6,02	5,60	5,86
<chem>c1nccn1C[C@H](C)NC(=O)N(C[C@H]2C)C[C@H](O2)c3ccc(F)cc3</chem>	80570599	191	6,51	6,07	6,27
<chem>c1ccnc(c12)c(ccc2)CC(=O)NCc3cc(ncc3)OC(C)(C)C</chem>	92531571	192	5,96	5,59	5,74
<chem>O1CCC[C@H]1CN(C(=O)CCCc2nc(no2)-c(cc3)ccc3C</chem>	4433832	193	5,41	5,45	5,86
<chem>n1oc(C)cc1CN(C)C(=O)CCc2ccc(o2)-c3c(F)cccc3</chem>	25137544	194	5,97	5,37	6,19
<chem>c1cccc(OC)c1CNC(=O)CN2C(=O)CSc(c23)nccc3</chem>	6802489	195	6,13	5,22	6,06
<chem>O1CCC[C@H]1C(=O)NCc(n2)n(c(c23)cccc3)Cc4c(C)ccc(C)c4</chem>	40197222	196	6,37	5,86	6,30
<chem>C1COCCC1OCCC(=O)Nc(n2)sc(c23)cccc3OC</chem>	84520887	197	5,53	5,49	5,76
<chem>CN1C[C@H](CC1=O)c(n2)n(c(c23)cccc3)CCOc4cccc(c45)cccc5</chem>	40250743	198	5,89	5,30	6,02
<chem>CCc(n1)sc1CC(=O)N(CCO2)C[C@H]2c3ccc(F)cc3</chem>	55196763	199	6,15	5,34	5,80
<chem>c1cccc1C[C@H](C(=O)C)NC(=O)CSc(n2)n(C)c(c23)ccc(Cl)c3</chem>	44049212	200	6,58	5,74	6,34
<chem>c1cc(Cl)ccc1CCNC(=O)COc2ncnc(c2c34)sc3CCC4</chem>	3599992	201	6,13	5,58	5,74

<chem>OCC[C@@H]1COCCN1C(=O)c2c(OCC)cn(n2)-c3ccccc3</chem>	92452672	202	5,78	5,03	5,81
<chem>O1CCC[C@H]1CNC(=O)CCc2ncc(o2)-c(c3)ccc(C)c3C</chem>	24924387	203	5,78	5,92	5,71
<chem>O1CCC[C@H]1CNC(=O)CCc2ncc(o2)-c(c3)ccc(C)c3C</chem>	24924391	204	5,78	5,92	5,71
<chem>Cc(o1)nc(C)c1C(=O)N[C@@H](C2=O)CN2CCOc3ccccc3</chem>	73959717	205	5,81	4,96	5,77
<chem>C1COCCC1OCCC(=O)Nc(ccc2)cc2-c(n3)oc(C)c3C</chem>	84530501	206	5,79	6,10	5,77
<chem>c1cccc(c12)CCC[C@H]2NC(=O)COC(=O)c3ccccc(n3)OC</chem>	83939506	207	4,96	5,36	5,71
<chem>O1CCOc(c12)ccc(c2)CN(C)C(=O)N[C@@H]3CCCN(c34)nc(n4)CC</chem>	87249506	208	5,63	5,20	6,19
<chem>COC[C@H]1CN(CCO1)c2c(C(=O)OCC)cnc(c23)ccc(C)c3</chem>	78342925	209	5,96	6,00	5,81
<chem>O1CCC[C@H]1C(=O)NCCc(n2)n(c(c23)cccc3)Cc4c(C)cccc4</chem>	40198730	210	6,51	5,96	6,05
<chem>C1COCCN1C(=O)[C@H](CC)NC(=O)Cc2coc(c23)cc4c(c3)CCC4</chem>	84223098	211	6,36	5,81	6,36
<chem>O1COc(c12)ccc(c2)CN[C@H](C)C(=O)N[C@H]3CCCN(c34)cccc4</chem>	59818114	212	6,36	5,44	6,02
<chem>s1ccnc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3)CCOC</chem>	92841424	213	6,38	6,01	5,96
<chem>C1CC1[C@H](O)CN(C)C(=O)c2nc(C(C)C)n(n2)-c3c(F)cccc3</chem>	87561073	214	5,90	5,41	6,32
<chem>CCOc(cc1)ccc1CC(=O)N[C@@H](C)c(c2)oc(c23)c(OC)ccc3</chem>	25137832	215	5,73	5,20	5,85
<chem>COc(cc1)ccc1CC(=O)N[C@@H](C)c(c2)oc(c23)c(ccc3)OCC</chem>	57908442	216	5,73	5,20	5,85
<chem>Cc1nc([nH]n1)SCC(=O)N2CCOc(c23)cc(cc3)N(C)C</chem>	74191425	217	6,05	5,26	5,93
<chem>n1ccncc1CN(C)C(=O)N[C@@H](C)c2cc(ccc2)-n3ccnc3</chem>	80551568	218	6,38	5,98	5,89
<chem>c1cc(F)ccc1[C@@H](C)[C@H]2CN2C(=O)Nc3ccn(n3)CCOC</chem>	77440429	219	6,16	5,56	6,13
<chem>n1occc1CC(=O)N[C@@H](c2cccc2)Cc3ccc(Cl)cc3</chem>	81282781	220	5,90	5,68	5,58
<chem>Clc1cccc(F)c1CC(=O)NCCn(n2)c(=O)n(c23)cccc3</chem>	58244146	221	6,12	5,63	6,28
<chem>C1CO[C@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3</chem>	92414547	222	6,27	5,38	5,87
<chem>C1CO[C@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3</chem>	92414544	223	6,27	5,38	5,87
<chem>c1cccc(c12)CCC[C@H]2C(=O)NCc3ccc(nc3)N(CC)CC</chem>	40121323	224	6,34	5,57	5,46
<chem>N#Cc1cc(ccc1)-c2nccn2CCCc3ccc(Cl)cc3</chem>	88521299	225	6,03	5,88	5,58
<chem>o1cccc1-c(no2)nc2CCC(=O)N[C@H](CCC3)[C@H]3Cc4ccccc4</chem>	78150926	226	5,84	5,65	5,88
<chem>n1ccccc1S[C@H](C)C(=O)NCc2cc(ccc2)-n3ccnc3</chem>	84977059	227	6,43	5,94	5,88
<chem>CC1(C)CN(CCO1)C(=O)Nc2cc(ccc2)OC[C@H]3CCCO3</chem>	78270310	228	5,22	6,17	5,67
<chem>n1cccc(O)c1C(=O)N(C2)CCC[C@H]2CCc3ccc(F)cc3</chem>	71735509	229	5,68	5,65	5,92
<chem>c1cc(Br)ccc1O[C@H](C)C(=O)NCCc2c(nccc2)N3CCOCC3</chem>	31326556	230	6,08	5,25	5,92

<chem>c1cccc(c12)oc(c2)[C@H](C)NC(=O)COc3c(OC)cccc3OC</chem>	22314745	231	5,16	5,25	5,85
<chem>c1cc(F)ccc1OCC(=O)NCCc(n2)sc(c23)cccc3</chem>	40323351	232	6,18	5,53	5,35
<chem>c1cccn(c12)cc(n2)CNC(=O)CCc3c(=O)oc(c34)cccc4</chem>	7944046	233	6,34	5,59	6,36
<chem>CC(C)(C)OC(=O)N(C1)C[C@H](CO)c(c12)nn(c2)CC3CCCC3</chem>	91302842	234	5,47	5,58	5,74
<chem>n1cccc1CNC(=O)[C@@H](CC2=O)CN2c(c3)ccc(C)c3C</chem>	9668971	235	6,43	5,61	5,96
<chem>Cc(o1)ccc1[C@@H](C)NC(=O)CN2C(=O)[C@H](C)Oc(c23)cccc3</chem>	23306977	236	5,82	5,10	5,79
<chem>c1cccc1[C@@H]2CN(CCO2)C(=O)NCC3c(CC)nn(C)c3OC</chem>	93149538	237	5,83	5,53	5,89
<chem>COCC1cc(ncn1)NC(=O)N2C[C@@H](C[C@H]2C)c3ccc(F)cc3</chem>	93316218	238	5,24	5,61	6,00
<chem>Cn1nnnc1S[C@H](C)C(=O)N[C@@H](C)c(c2)ccc(c23)CCCC3</chem>	22784584	239	5,56	5,35	5,80
<chem>COCC1cc(ncn1)NC(=O)N2C[C@H](C[C@H]2C)c3ccc(F)cc3</chem>	93316214	240	5,24	5,61	6,00
<chem>n1oc(C)cc1CN(C)C(=O)CCc2cnn(c2)-c3cccc3</chem>	31657942	241	6,17	5,17	5,91
<chem>CCCN(n1)cc(Cl)c1C(=O)N(C2)CCC[C@@H]2Cc3ccccn3</chem>	72384932	242	6,31	5,22	5,80
<chem>s1ccnc1CN(C)C(=O)NCc2cc(c2)-n3cccc3</chem>	80696288	243	6,38	5,70	5,91
<chem>FC(F)(F)c1cc(ccc1)[C@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3</chem>	92036660	244	6,50	6,13	5,57
<chem>FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3</chem>	92036658	245	6,50	6,13	5,57
<chem>c1cccc(c12)sc(n2)CN(C)C(=O)N[C@@H]3CCc4nc(n4)CC</chem>	87251107	246	6,35	5,55	6,22
<chem>CC(C)Cn(c1=O)c(=O)n(C)c(c12)nc3n2c(C)c(C)n3CCOC</chem>	4314324	247	6,47	5,47	6,02
<chem>c1cccc1C(=O)N(C2)CCC[C@@H]2C(=O)OCCc3ccnc3</chem>	73299335	248	6,19	5,24	5,64
<chem>COc1n(C)nc(CC)c1CN[C@H]2CCOc(c23)c(Cl)ccc3Cl</chem>	89211609	249	6,10	5,32	5,95
<chem>Clc1cnn(c1)CCNC(=O)CCn(nc2)c(c23)ccc(Cl)c3</chem>	40760451	250	6,46	5,54	5,68
<chem>CCN(CC)C(=O)CSc(o1)nn1-c(c2)ccc(c23)cccc3</chem>	6416097	251	5,97	5,12	5,81
<chem>Cn(n1)cnc1CNC(=O)N(C2)CC[C@@H]2Oc3c(Cl)cccc3</chem>	91775455	252	5,76	5,61	5,69
<chem>C1COCCC1OCCC(=O)Nc2ccc(OC)c(c23)cccc3</chem>	84532176	253	5,66	6,01	5,86
<chem>C1COCCN1CCNC(=O)Cc2coc(c23)ccc4c3cccc4</chem>	21874011	254	6,56	5,52	5,84
<chem>c1c(Cl)ccc(Cl)c1NC(=O)[C@H](C)Sc(n2CC)nn2C3CC3</chem>	7618357	255	5,97	5,95	5,79
<chem>C[C@@H](C1)[C@H]1C(=O)N2CCN(CC2)C(=O)NCCc(c3)oc(c34)ccc4</chem>	74380190	256	6,44	5,70	5,86
<chem>Cc(n1)sc1CCCC(=O)N(C2)CC[C@@H]2Oc3cc(Cl)ccc3</chem>	89624334	257	6,30	5,27	5,81
<chem>Cc(n1)sc1CCCC(=O)N(C2)CC[C@@H]2Oc3cc(Cl)ccc3</chem>	89624337	258	6,30	5,27	5,81
<chem>c1cccc(c12)OCC[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4</chem>	89956457	259	5,89	5,19	5,63

<chem>c1cccc(c12)OCC[C@H]2C(=O)N Cc3cc(ncc3)OC4CCC4</chem>	89956458	260	5,89	5,19	5,63

RNA=redes neurais artificiais, RL=regressão linear

Tabela C3 – Predição da seletividade entre células sadias e a forma amastigota de *Trypanosoma cruzi*.

SMILES	ZINC	Estrutura	KPLS	RF
<chem>N#Cc1c(C)nc(C)cc1N(C2)CC[C@H]2Oc(nc3)c cc3C(F)(F)F</chem>	92964267	1	1,48	2,23
<chem>c1cc(Cl)cc(Cl)c1NC(=O)N(Cc2ccnc2)CCN3CCO CC3</chem>	20423620	2	0,90	2,72
<chem>c1cc(F)ccc1C[C@H](n2nnnc2C)C(=O)NCC3(CC C3)c4ccc(Cl)cc4</chem>	72087376	3	1,46	2,15
<chem>c1ncccc1CC(=O)NCc2c(nccc2)Oc(cc3C(F)(F)F)c cc3</chem>	12449018	4	1,36	2,18
<chem>o1cccc1CN(C[C@H]2CCCO2)C(=O)Nc3cn(C)c(=O)c(c34)cccc4</chem>	6932434	5	0,61	2,42
<chem>c1cccc(F)c1[C@H]2CN(CCO2)C(=O)Nc(s3)nn c3-c4ccsc4</chem>	67280610	6	1,24	1,79
<chem>c1cc(Cl)ccc1CCNC(=O)[C@H]2CCCN(C2)C(=NC=N3)[C@H](C3=4)N=CN4</chem>	63492379	7	1,15	1,85
<chem>c1ncccc1CNC(=O)CSc(n2)nc(C(F)(F)F)c(c2c34) CCc3cccc4</chem>	646627	8	0,84	2,16
<chem>N#CC[C@H](C)Sc(cccc1)c1NC(=O)N(CCO2)C [C@H]2c3c(F)cccc3</chem>	66956124	9	0,98	1,96
<chem>o1cccc1CCC(=O)Nc(cc2)cc(c2F)S(=O)(=O)N3CC OCC3</chem>	44567226	10	0,83	2,10
<chem>C1COCCN1c(ccc(Cl)c2)c2NC(=O)N(C)CCOc3cc ccc3</chem>	71901309	11	0,96	1,97
<chem>n1cn(C)c(Cl)c1S(=O)(=O)Nc2c(cccc2)N3CCOCC 3</chem>	75581011	12	1,58	1,32
<chem>FC(F)(F)c1cc(ccc1)Cn(c(n2)CCCC)c(c23)nc(C)c(c3)N</chem>	4906567	13	0,87	2,01
<chem>CC(=O)c1cc(ccc1)Nc2c(C(=O)OCC)cnc(c23)ccc(F)c3</chem>	956542	14	0,96	1,92
<chem>n1ncn(c12)nc(s2)NC(=O)CCCOc(cc3)ccc3C(C)(C)C</chem>	3459546	15	1,19	1,68
<chem>COc(cc1)ccc1CNC(=O)C2(CC2)c3cc(on3)- c4c(F)cccc4</chem>	56919886	16	0,81	2,02
<chem>FC(F)n1ccnc1CN(C)C(=O)NCc2ncc(o2)- c3cccc3</chem>	79517712	17	0,84	1,99
<chem>CCc1nnc(s1)NC(=O)N(CC)C[C@H](O2)COc(c23))cccc3</chem>	43344165	18	1,26	1,57
<chem>CC(C)(C)Oc(ncc1)cc1CNC(=O)[C@H](CC2=O) [C@H](N2C)c3ccnc3</chem>	92119304	19	0,72	2,11
<chem>CCc1ccc(cc1)CNC(=O)CCc2c(noc2)-c3c(F)cccc3</chem>	75510592	20	1,00	1,76
<chem>c1cc(F)cc(F)c1CC(=O)N[C@H](C)c2nc(cs2)- c3ccncc3</chem>	87335960	21	0,80	1,95
<chem>CCc(o1)ccc1C(=O)Nc2c(cccc2)N(C[C@H]3C) C[C@H](O3)C</chem>	55685381	22	1,45	1,28
<chem>c1ncccc1CN(CCC#N)C(=O)Nc(ccc2)c(c23)nccc3</chem>	80779799	23	1,24	1,49
<chem>c1cc(F)ccc1CNC(=O)[C@H](C(C)C)Sc(n2)nn(c23))ccc3</chem>	30718569	24	0,81	1,88
<chem>c1cccc(c12)CN(CC2)C(=O)[C@H](C(C)C)Sc(o 3)nnc3C(C)(C)C</chem>	77247970	25	1,00	1,69
<chem>CCc(cc1)ccc1[C@H](C)NC(=O)[C@H]2CCC N(C2)c3ccc(Cl)nn3</chem>	40130295	26	0,81	1,87

C1CCCCC1OCC(=O)N(C2)CC[C@H]2Oc(nc3)cc c3C(F)(F)F	93052953	27	0,84	1,84
c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCC OC[C@@H]3CCCO3	67386210	28	0,35	2,34
c1cc(F)c(F)cc1[C@@H]2CN(CCO2)C(=O)NCCC OC[C@H]3CCCO3	67386217	29	0,35	2,34
C[C@@H]1CN(C[C@@H](O1)C)c(cccc2)c2NC(=O) N(C)Cc3ncnn3C	72509269	30	1,04	1,63
c1cc(Cl)ccc1NC(=O)Cn(n2)c(C)cc2C(=O)N3CCO CC3	11927649	31	0,78	1,89
COC[C@@H]1CN(CCO1)C(=O)NCC2(CCCC2)c 3cc(OC)ccc3	74142415	32	1,13	1,51
C1OCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc(n c3)ccc3C(F)(F)F	93053722	33	0,76	1,88
C1OCCC[C@H]1CCC(=O)N(C2)CC[C@H]2Oc (nc3)ccc3C(F)(F)F	93053726	34	0,76	1,88
C1CC1c(n2C)nnc2CN(C)C(=O)Nc3c(F)ccc(c3)- n4cccc4	81067125	35	0,36	2,27
CCc1nc(on1)CN(C)C(=O)NCCc2cnn(c2)- c3cccc3	69762537	36	1,10	1,52
FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)Cn3 nnnc3C(C)C	91788667	37	0,89	1,71
CCCc1nnc(s1)NC(=O)[C@@H](C)c(c2)ccc(c23) OCCO3	68650118	38	1,18	1,42
CC(=O)c1ccc(nc1)N(C2)CCC[C@H]2Cn(nn3)cc3 CCO	23478649	39	1,04	1,55
C1CCn(c12)c(nn2)CNC(=O)CCc3ccc(o3)- c4cccs4	20869699	40	1,01	1,57
Cc(n1)sc(C)c1[C@H](C)N(C)C(=O)CCc2nnc(o2)- c3cccc3	12302511	41	0,96	1,59
C1COCCN1C(=O)Nc2cc(ccc2)OCc3ccc(F)cc3	27343935	42	1,13	1,42
n1cccc1CNC(=O)[C@@H]2CCCN(C2)c3ncnc(c 34)sc(c4)CC	33078299	43	0,80	1,74
c1cc(F)c(F)cc1[C@H](C)N(C)C(=O)NCc(nn2)n(c 23)ccc3	79265262	44	0,63	1,91
C1CCCN1[C@H]2CCN(C2)C(=O)NCc3c(nc3) OC4CCCC4	66178874	45	0,83	1,69
n1nc(C)n(c12)CCC[C@H]2NC(=O)CCc3ncc(o3)- c4cccc4	79594184	46	1,07	1,45
Cc(c1)cc(C)c(c12)ncc(C#N)c2N(CC3)CCC3OC[C @H]4CCCO4	81031040	47	0,59	1,92
Cc(c1)cc(C)c(c12)ncc(C#N)c2N(CC3)CCC3OC[C @H]4CCCO4	81031042	48	0,59	1,92
C1CC[C@H]([C@H]1C)NC(=O)C[C@H](C)NC 2=NS(=O)(=O)c(c23)ccc3	80069368	49	0,75	1,75
C1OCC[C@]1(C(=O)OC)NC(=O)c2cc(ccc2)- c(cs3)nc3C	88220932	50	0,70	1,80
CC(C)(C)c1nc(on1)CCC(=O)N[C@H]2CCOc(c23)ccc3	33159238	51	0,96	1,53
CC(C)(C)c1nc(on1)CCC(=O)N[C@@H]2CCOc(c 23)ccc3	33159237	52	0,96	1,53
O1COc(c12)ccc(c2)NC(=O)[C@H](C)N(C)Cc3c(F)ccc3Cl	3457326	53	0,26	2,22
c1cc(C)ccc1CNC(=O)CCc2nc(no2)-c3cc(F)ccc3	72332141	54	0,78	1,69
CC(C)(O)CN(C)C(=O)Cc1csc(n1)- c(c2)ccc(c23)OCO3	80822767	55	0,81	1,66
c1cccc(OC)c1C[C@H](C)N(C)C(=O)Nc2cncc(c23)ccc3	89357203	56	0,97	1,50
n1cccc(OC(C)C)c1C(=O)N2C[C@H](OC[C@@H]	79531837	57	0,35	2,12

2C)c(c3C)cc(C)cc3				
c1cccc(c1CC)NC(=O)CCC(=O)N2CCC[C@H]2 c3nccs3	19146445	58	0,92	1,53
O1CCN(C1=O)CCC(=O)N(CC2)CCC2OCc3ccc(F) cc3	92076090	59	0,65	1,79
CC(C)c1nc(on1)CCC(=O)N2CCN(CC2)c3ccncc3	48105992	60	0,96	1,47
c1ccc(F)c(F)c1[C@H](C2)[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956312	61	0,58	1,85
n1oc(C)nc1CNC(=O)N(CCc(c23)cccc2)[C@H]3C c4cccc4	90593151	62	0,84	1,58
c1cc(Cl)ccc1CCNC(=O)Cc2csc(n2)N3CCCC3=O	8133311	63	0,83	1,58
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(cc(F)c3)-c4cncnc4	11974390	64	0,58	1,82
O1CCC[C@H]1C(=O)NC[C@H](C2)Oc(c23)c(c c(F)c3)-c4cncnc4	11974396	65	0,58	1,82
c1cccc(Cl)c1CNC(=O)CCn(c2=O)cnc(c23)sc(C)c 3C	40024984	66	0,66	1,75
c1cc(Cl)cc(Cl)c1[C@H](C)NC(=O)CCn(c2=O)c nc(c23)scc3	12749371	67	0,39	2,01
COc1nccc(c1C#N)N[C@H](CC2=O)CN2CCCc3c cccc3	72153248	68	0,88	1,52
CC(C)n(nc1)c(C)c1C(=O)OCC(=O)N[C@H]2C CCc(c23)cccc3	92631569	69	0,87	1,53
n1[nH]c(C)nc1CCC(=O)N(C)Cc2csc(n2)- c3cccc3	72454569	70	0,88	1,52
c1c(F)ccc(F)c1[C@H](C2CC2)NC(=O)N(C)Cc3 cnc(s3)CC	74893708	71	0,71	1,69
FC(F)(F)c1c(Cl)ccc(c1)NC(=O)[C@H](CC2=O) CN2Cc3ccc3	7306435	72	0,43	1,96
Cc(n1)scc1CC(=O)N(C2)CCC[C@H]2c(c3)nc(C)cc3Cc4c(F)cccc4	72354119	73	0,41	1,99
CCS(=O)(=O)N(CC1)CCC1C(=O)NCCc(c2)ccc(c 23)OCC3	74616895	74	0,73	1,65
CC(C)(C)NC(=O)CCSc(o1)nnc1- c2cccc(c23)cccc3	59188606	75	1,04	1,34
C1OCC[C@H]1CN(CC)C(=O)CCc2nc(no2)- c3cccc3	48263350	76	0,80	1,57
CC(C)(C)c1ccc(cc1)OC[C@H](O)Cn2c(C)cc(= O)n(c23)ncn3	45097288	77	0,87	1,49
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c2 3)cc(Cl)cc3	3448391	78	0,76	1,61
O1CCC[C@H]1CNC(=O)COC(=O)c(c2Cl)sc(c23) cc(Cl)cc3	3448394	79	0,76	1,61
CCOC(=O)[C@H]1CCCN(C1)C(=O)Nc2c(cccc2) OC3CCCC3	79682507	80	1,00	1,35
c1ccnc(OC)c1CNS(=O)(=O)c2cccc(c23)cncc3	77615954	81	1,08	1,27
CC(=O)c1cnc(s1)NC(=O)C2(CCCC2)c3c(F)cc(F) cc3	89811381	82	0,89	1,45
Cc(n1)scc1CCNC(=O)N(C)Cc(c2)ccc(c23)OCO3	44967753	83	0,83	1,51
c1cccc(OC)c1COC[C@H](O)CN2CCOc(c23)cc cc3	60857051	84	1,17	1,16
C1OCC[C@H]1CN(C)C(=O)NC[C@H](C)c2nc(no 2)-c3cccc3	89717833	85	0,91	1,41
C1OCC[C@H]1CN(C)C(=O)NC[C@H](C)c2nc(no2)-c3cccc3	89717835	86	0,91	1,41
C1CCCCC1(C#N)NC(=O)[C@H](C)Sc(cc2)nc(c23)cccc3	5627281	87	0,97	1,32
c1cccc1OCC(=O)N(C2)CCC[C@H]2C(=O)Nc(cc3C)cnc3	79577108	88	0,58	1,71

c1cc(F)ccc1[C@@H]2CN(CCO2)C(=O)NCCCN3c cnc3	42896953	89	0,75	1,54
Cc1cn(nn1)CCNC(=O)CCSc(c2)ccc(c23)OCCO3	65528008	90	0,64	1,66
s1ccn(c12)cc(n2)CC(=O)NCc3cc(ccc3)OCc4cccc c4	59411307	91	0,75	1,54
C1CCCCC1NC(=O)COC(=O)c2cc(ccc2)Oc3cccc c3	3393368	92	0,82	1,46
c1cnc(OC)cc1CNC(=O)N(C2)CCC[C@@H]2c(n3)oc(c34)cccc4	45494692	93	0,65	1,63
s1ccc(C)c1C(=O)[C@@H]2CCCN(C2)C(=O)Nc(c (cc3)OC)cc3C	39393772	94	0,58	1,69
s1ccc(C)c1C(=O)[C@@H]2CCCN(C2)C(=O)Nc(c(c c3)OC)cc3C	39393770	95	0,58	1,69
FC(F)COc(nc1)cc1CNC(=O)Cc2cccc(c23)cccn3	91046021	96	0,31	1,96
c1ccc(Cl)cc1[C@@H](C)NC(=O)[C@@H](CC2=O) CN2Cc3cccn3	31601820	97	0,35	1,91
c1cccc(c1C(F)(F)F)CN(C2CC2)C(=O)NCc3ncn(n 3)C	92377367	98	0,67	1,58
CO[C@@H]1CCCN(C1)C(=O)c2cc(ccc2)S(=O)(=O)NCC3CC3	76245437	99	0,62	1,62
CCCOc(nc1)cc1CNC(=O)N2CCC[C@@H]2CN3 CCOCC3	73875207	100	0,77	1,48
CCCCNC(=O)[C@@H](C#N)c(n1)c(nc(c12)cccc2)N (C3)CCC[C@@H]3C	4837255	101	0,66	1,58
c1cc(Cl)cc(Br)c1OCC(=O)NCc2ccc(o2)- c3csc(n3)C	12904725	102	0,90	1,34
C1CCCC1NC(=O)Cc2nc(on2)Cc3ccc(cc3)OCC	72552497	103	0,96	1,27
c1cccc(c12)N(C(=O)CC2)CC(=O)NCCc(n3)sc(c3 4)cccc4	74718345	104	0,73	1,50
COC[C@@H](C)NC(=O)CCn(c(c12)cccc2)c(c1)- c3ccncc3	19456616	105	0,83	1,39
c1cccc(c12)oc(c2)[C@@H](C)C(=O)NCc3cc(ncc 3)OCC(F)F	91045925	106	0,46	1,76
c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@@H](C)Oc 3ccc(Cl)cc3	81348380	107	0,36	1,86
Cc1onc(CC)c1C(=O)NCCCO[C@@H]2CCCc(c2 3)cccc3	70874454	108	0,99	1,23
c1cncc(c12)cccc2NC(=O)CCC(=O)N(C3)CCc(c3 4)sc4	54849997	109	0,64	1,57
c1cccc(c12)N(CC2)C(=O)NCCOCc3ccc(cc3)OC	93201872	110	1,11	1,10
c1ccnc(OC)c1C(=O)N(C2)CCC[C@@H]2CCc3cc c(F)cc3	71749519	111	0,75	1,47
O1CCOC(c12)ccc(c2)CN(C)C(=O)NCCCOc(c3C) cccc3	78518847	112	0,47	1,73
CCOC(=O)c1cnn(c1)[C@@H]2CCCN(C2)C(=O) NCC3(CC)CCC3	91701517	113	0,96	1,25
Cc1c(Cl)c(C)n(n1)CC(=O)NCc2ncc(o2)-c3cccc3	80338882	114	0,80	1,40
s1cccc1[C@@H](CC)NC(=O)N(CCO2)C[C@@H]2c3 cc(F)c(F)cc3	67406599	115	0,36	1,84
s1cccc1[C@@H](CC)NC(=O)N(CCO2)C[C@@H]2 c3cc(F)c(F)cc3	67406602	116	0,36	1,84
n1cccc1CN(CC)C(=O)NCc2cn(nn2)-c3cccc3	87166849	117	0,99	1,20
N#CC[C@@H](C)N(C)C(=O)c1csc(n1)- c(c2)ccc(c23)OCO3	78420281	118	1,00	1,19
n1cccn1CC2(CC2)C(=O)N(C3)CCC[C@@H]3c(n 4)[nH]c(c45)cccc5C	91321262	119	0,90	1,29
c1sccc1[C@@H](C)C(=O)N[C@@H](C)c2cc(ccc2) N(C3=O)CCO3	78090111	120	0,73	1,46
COCCOc(ccc1)c1NC(=O)N(C)[C@@H](C)c(c2)	66403521	121	0,69	1,50

oc(c23)cccc3				
COCCOc(cccc1)c1NC(=O)N(C)[C@H](C)c(c2)oc(c23)cccc3	66403524	122	0,69	1,50
c1cc(C)ccc1CNC(=O)CCc2ncc(o2)-c3ccc(F)cc3	12929356	123	0,95	1,24
FC(F)(F)[C@@]1(O)CCN(C1)C(=O)CCc2ncc(o2)-c3cccc3	77989783	124	0,58	1,60
FC(F)(F)[C@]1(O)CCN(C1)C(=O)CCc2ncc(o2)-c3cccc3	77989787	125	0,58	1,60
CCOC(=O)CSc(c1C#N)nc(c(c12)CCCC2)N3CCOCC3	903744	126	0,71	1,46
CC(C)NC(=O)COC(=O)c1nn(c(c12)CCC2)-c3c(F)cccc3	46509084	127	0,65	1,52
c1cnccc1[C@H](C)N(C)C(=O)NCCCO[C@@H]2CCCc(c23)cccc3	73221057	128	0,57	1,59
c1cccc1CCC(=O)N(C2)CC2c3nc(no3)-c4ccsc4	75156512	129	0,60	1,56
Cn(c1)nc(CC)c1CNC(=O)N2C[C@@H](OC[C@H]2C)c(c3C)cccc3	90824539	130	0,49	1,67
Cc1onc(C)c1CN(C)C(=O)N[C@@H](C)c(c2)oc(c23)c(OC)cccc3	79157214	131	0,58	1,58
c1ccc(C)c(c1C)NC(=O)CNC(C)(C)c(c2)ccc(c23)OCO3	64344145	132	0,84	1,31
CCc1nc(on1)CN(C)C(=O)NCc2csc(c23)cccc3	69772036	133	0,95	1,21
c1cccn(c12)c(nn2)[C@H](C)NC(=O)N(C)Cc(s3)nc3CC	83175027	134	0,43	1,73
CC(C)c1nc([nH]n1)SCC(=O)N(C)C[C@@H](O2)COc(c23)cccc3	14110875	135	0,54	1,61
CC(C)CCO[C@H]1CCN(C1)C(=O)NCc2cn(nn2)-c3cccc3	87166632	136	0,84	1,31
CCc(n1)sc1CC(=O)NCCc2ccc(cc2)-n3cccn3	65520577	137	0,72	1,43
c1nc(C)nc(c12)CCC[C@@H]2NC(=O)COc(c3)ccc(c34)cccc4	72242521	138	0,64	1,51
o1cccc1C(=O)C[C@H]2COCCN2C(=O)c3n[nH]c(c34)cccc4	95439902	139	0,58	1,57
c1cccc(c12)OCC[C@@H]2C(=O)N[C@H](C)COc3ccc(F)cc3	76798231	140	0,88	1,26
c1cc(F)c(F)cc1[C@@H](C(=O)OC)NC(=O)[C@@H](C2)Sc(c23)cccc3	89364439	141	0,26	1,88
c1ncccc1C(=O)NCc2c(cc(C)cc2)OC[C@@H]3CCOC3	87455925	142	0,76	1,38
C1CCC[C@@H]([C@H]12)N([C@@H](C2)C(=O)OC)C(=O)CCNC(=O)c3cccc3	55448395	143	0,65	1,48
Cc1onc(C)c1CN(C)C(=O)NCCOc2cccc(c23)cccc3	56960730	144	0,90	1,23
COCCc1cc(ccc1)NC(=O)N2CCC[C@@H]2c(n3)sc(c34)cccc4	76128493	145	0,70	1,43
c1nccn1CCCNC(=O)N2CCC[C@@H]2c(n3)sc(c34)cccc4	42896682	146	0,57	1,56
c1cccc(c12)[C@H](CCC2(C)C)CNS(=O)(=O)CC3CCOCC3	93939051	147	0,80	1,32
s1ccn(c12)cc(n2)CNC(=O)N(C)CCOc(c3C)cccc3	78984628	148	0,87	1,25
c1ccc(F)c(F)c1CCC(=O)N[C@H]2CCCN(c23)nc(n3)CC	88197817	149	0,24	1,88
COc(cc1)cc(Cl)c1NC(=O)N(C)Cc(n2)sc(c23)cccc3	68561409	150	0,62	1,50
O1COc(c12)ccc(c2)CN(C)C(=O)NCCCc(n3)sc(c34)CCCC4	66204557	151	0,65	1,47
c1cc(C)ccc1CNC(=O)CCc2nnc(o2)-c3c(C)occ3	40756727	152	0,65	1,47
O1CCC[C@H]1Cn2c(SCC(=O)NC(C)C)ncc2-c3cccc3	8709934	153	0,83	1,28

O1CCC[C@ @H]1CNC(=O)Cc2csc(n2)SCc3c(F)c ccc3	8939348	154	0,61	1,50
c1cccc1C(=O)C[C@H]2COCCN2C(=O)CCc3csc n3	95412325	155	0,58	1,52
c1cccc1C(=O)C[C@ @H]2COCCN2C(=O)CCc3 cscn3	95412324	156	0,58	1,52
O1COc(c12)ccc(c2)CN(CC)C(=O)Nc3cccc(c34)c ncc4	43162190	157	0,82	1,28
c1cccc(c12)OC[C@ @H](C)N2C(=O)c3ccc(=O)n(n3)CCCC	72532639	158	0,59	1,51
OCCOC(=O)c1cc(-c2cccc2)c(s1)N3CCOCC3	26413486	159	1,03	1,06
c1cccc(F)c1C(=O)C[C@H](C)C(=O)NCCc(n2)cn(c23)cccc3	79400264	160	0,23	1,86
c1cccc(c12)onc2CC(=O)NCc3cc(ncc3)OC4CCC4	90013735	161	0,91	1,18
c1ccnc(c12)c(ccc2)CC(=O)NCCCO[C@ @H]3CC Cc(c34)cccc4	72821379	162	0,46	1,63
c1cccc1[C@ @H](O2)CN[C@H]2C(=O)NC CCOC3CCOCC3	74542092	163	0,40	1,68
n1cccc1CCNC(=O)N(CCO2)C[C@H]2c3ccc(Br) cc3	67589176	164	0,65	1,44
s1cccc1CN(CC=C)C(=O)N[C@H](C)c2nc(no2)C3 CCCC3	80567565	165	0,74	1,34
c1ccn(C)c(=O)c1NC(=O)CCC(=O)N(C2)CCc(c23)scc3	74073007	166	0,25	1,82
Cc1ccc(cc1)-c(nc(s2)C)c2CC(=O)Nc3ccnn3CCC	75598283	167	0,82	1,26
Cc1csc(=O)n1CC(=O)N(C2)CC2OCc3cccc(c34)c ccc4	80941753	168	0,53	1,55
C1CC1c2ncc(cn2)NC(=O)N(C3)CC[C@ @H]3Oc 4cc(Cl)ccc4	90479947	169	0,69	1,39
C1CCCCC1OCC(=O)NC[C@ @H](C)c2nc(no2)- c3cccc3	89955773	170	0,86	1,22
c1ncccc1OCCCC(=O)N[C@ @H]2CCCOc(c23)cc cc3	79880561	171	0,71	1,35
c1cccc(c12)OC[C@H](C2)C(=O)N[C@H](C)c(c3) oc(c34)c(OC)ccc4	74509023	172	0,54	1,52
c1ccc(F)cc1CN(CC)C(=O)c(c2)cnc(c23)[nH]nc3C	47382594	173	0,66	1,40
C1CCCN1C(=O)c2cc(ccc2)NC(=O)N(C)Cc3ncn n3C	72509551	174	0,36	1,70
n1ccncc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3)CCO C	92841199	175	0,67	1,38
o1nc(C)nc1COCC(=O)N[C@H](C2CCC2)c3ccc(Cl)cc3	74566760	176	0,67	1,38
c1cccc(c12)OCC[C@H]2CC(=O)NCc(o3)ccc3C(= O)OCC	79586889	177	0,92	1,13
c1cc(C)c(C)cc1CNC(=O)Cn(cn2)cc2- c(n3)scc3CC	65250562	178	0,83	1,22
c1cccc(COC)c1CNC(=O)N(C)Cc(n2)sc(c23)CCC C3	83177089	179	0,73	1,32
c1cccc(c12)OCC[C@ @H]2NC(=O)Cc3coc(n3)- c4cccc4	32207533	180	0,78	1,27
C1COCCN1C(=O)CSc(n2C[C@ @H]3CCCCO3)nc c2-c4cccc4	11641712	181	0,41	1,63
C1COCCN1C(=O)CSc(n2C[C@H]3CCCCO3)ncc2 -c4cccc4	11641718	182	0,41	1,63
c1cc(Cl)ccc1C[C@ @H](C)CC(=O)N[C@ @H]2CC Cn(c23)nc(n3)C(C)C	92958887	183	0,43	1,61
N#Cc1cc(c(F)cc1)CNC(=O)N(CCO2)C[C@H]2c3 cccc3	75435037	184	0,79	1,25
c1cccc(c12)oc(=O)n2CCC(=O)N(C)CC(=O)Nc3cc	7312811	185	-0,10	2,13

cc(C)c3C				
C1OCC[C@H]1CN(C)C(=O)Nc(ccc2)cc2COCc3c cco3	78493097	186	0,73	1,30
c1cccn(c12)c(=O)n(n2)CC(=O)NC[C@H](C)Oc 3ccc(F)cc3	80760513	187	0,36	1,66
c1c(F)ccc(c12)CCC[C@H]2NC(=O)CCc3nnc(o 3)C4CCCC4	93061622	188	0,38	1,65
C1CCCC=C1CCNC(=O)Cn(c2=O)c(C)nc(c23)on c3C	35507406	189	0,55	1,47
n1nc(Cl)ccc1SCC(=O)N[C@H](C)c(c2)ccc(c23)cc cc3	10653606	190	0,61	1,40
c1nccn1C[C@H](C)NC(=O)N(C[C@H]2C)C[C@ H](O2)c3ccc(F)cc3	80570599	191	0,24	1,78
c1ccnc(c12)c(ccc2)CC(=O)NCc3cc(ncc3)OC(C)(C)C	92531571	192	0,72	1,29
O1CCC[C@H]1CNC(=O)CCCc2nc(no2)- c(cc3)ccc3C	4433832	193	0,61	1,40
n1oc(C)cc1CN(C)C(=O)CCc2ccc(o2)- c3c(F)cccc3	25137544	194	0,30	1,71
c1cccc(OC)c1CNC(=O)CN2C(=O)CSc(c23)nccc3	6802489	195	0,42	1,59
O1CCC[C@H]1C(=O)NCc(n2)n(c(c23)cccc3)Cc4 c(C)ccc(C)c4	40197222	196	0,20	1,80
C1COCCC1OCCC(=O)Nc(n2)sc(c23)cccc3OC	84520887	197	0,69	1,31
CN1C[C@H](CC1=O)c(n2)n(c(c23)cccc3)CCO c4cccc(c45)cccc5	40250743	198	0,46	1,55
CCc(n1)sc1CC(=O)N(CCO2)C[C@H]2c3ccc(F)c c3	55196763	199	0,66	1,34
c1ccccc1C[C@H](C(=O)C)NC(=O)CSc(n2)n(C) c(c23)ccc(Cl)c3	44049212	200	0,15	1,85
c1cc(Cl)ccc1CCNC(=O)COc2ncnc(c2c34)sc3CC C4	3599992	201	0,70	1,29
OCC[C@H]1COCCN1C(=O)c2c(OCC)cn(n2)- c3cccc3	92452672	202	0,64	1,36
O1CCC[C@H]1CNC(=O)CCc2ncc(o2)- c(c3)ccc(C)c3C	24924387	203	0,73	1,26
O1CCC[C@H]1CNC(=O)CCc2ncc(o2)- c(c3)ccc(C)c3C	24924391	204	0,73	1,26
Cc(o1)nc(C)c1C(=O)N[C@H](CC2=O)CN2CC Oc3cccc3	73959717	205	0,68	1,32
C1COCCC1OCCC(=O)Nc(ccc2)cc2- c(n3)oc(C)c3C	84530501	206	0,67	1,32
c1cccc(c12)CCC[C@H]2NC(=O)COC(=O)c3cccc (n3)OC	83939506	207	0,73	1,26
O1CCOc(c12)ccc(c2)CN(C)C(=O)N[C@H]3CC Cn(c34)nc(n4)CC	87249506	208	0,29	1,70
COC[C@H]1CN(CCO1)c2c(C(=O)OCC)cnc(c23) ccc(C)c3	78342925	209	0,63	1,36
O1CCC[C@H]1C(=O)NCCc(n2)n(c(c23)cccc3) Cc4c(C)cccc4	40198730	210	0,41	1,57
C1COCCN1C(=O)[C@H](CC)NC(=O)Cc2coc(c23))cc4c(c3)CCC4	84223098	211	0,12	1,86
O1COc(c12)ccc(c2)CN[C@H](C)C(=O)N[C@H]3 CCCc(c34)cccc4	59818114	212	0,43	1,55
s1ccnc1CN(C)C(=O)Nc(cc2)cc(c23)n(cc3)CCOC	92841424	213	0,48	1,49
C1CC1[C@H](O)CN(C)C(=O)c2nc(C(C)C)n(n2))c3c(F)cccc3	87561073	214	0,15	1,82
CCOc(cc1)ccc1CC(=O)N[C@H](C)c(c2)oc(c23))c(OC)ccc3	25137832	215	0,58	1,39
COc(cc1)ccc1CC(=O)N[C@H](C)c(c2)oc(c23)c	57908442	216	0,58	1,39

(ccc3)OCC				
Cc1nc([nH]n1)SCC(=O)N2CCOc(c23)cc(cc3)N(C)C	74191425	217	0,50	1,47
n1ccncc1CN(C)C(=O)N[C@@H](C)c2cc(ccc2)-n3ccnc3	80551568	218	0,53	1,43
c1cc(F)ccc1[C@@H](C)[C@@H](2C)CN2C(=O)Nc3ccn(n3)CCOC	77440429	219	0,32	1,65
n1occc1CC(=O)N[C@@H](c2ccccc2)Cc3ccc(Cl)c3	81282781	220	0,82	1,14
Clc1cccc(F)c1CC(=O)NCCn(n2)c(=O)n(c23)cccc3	58244146	221	0,17	1,79
C1CO[C@@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3	92414547	222	0,54	1,41
C1CO[C@H](C)[C@H]1C(=O)N[C@H](CC2=O)CN2CCOc3ccccc3	92414544	223	0,54	1,41
c1cccc(c12)CCC[C@H]2C(=O)NCc3ccc(nc3)N(C)CC	40121323	224	0,92	1,03
N#Cc1cc(ccc1)-c2ncnn2CCc3ccc(Cl)cc3	88521299	225	0,81	1,14
o1cccc1-c(no2)nc2CCC(=O)N[C@H](CCC3)[C@@H]3Cc4cccc4	78150926	226	0,53	1,42
n1cccc1S[C@@H](C)C(=O)NCc2cc(ccc2)-n3ccnc3	84977059	227	0,53	1,42
CC1(C)CN(CCO1)C(=O)Nc2cc(ccc2)OC[C@@H]3CCCO3	78270310	228	0,72	1,22
n1cccc(O)c1C(=O)N(C2)CCC[C@H]2CCc3ccc(F)cc3	71735509	229	0,49	1,45
c1cc(Br)ccc1O[C@H](C)C(=O)NCc2c(nccc2)N3COCC3	31326556	230	0,49	1,45
c1cccc(c12)oc(c2)[C@H](C)NC(=O)COc3c(OC)c3cc3OC	22314745	231	0,55	1,39
c1cc(F)ccc1OCC(=O)NCCc(n2)sc(c23)cccc3	40323351	232	1,01	0,93
c1cccn(c12)cc(n2)CNC(=O)CCc3c(=O)oc(c34)c3cc4	7944046	233	0,07	1,86
CC(C)(C)OC(=O)N(C1)C[C@H](CO)c(c12)nn(c2)CC3CCCC3	91302842	234	0,64	1,29
n1cccc1CNC(=O)[C@@H](CC2=O)CN2c(c3)cc3(C)c3C	9668971	235	0,44	1,50
Cc(o1)ccc1[C@@H](C)NC(=O)CN2C(=O)[C@H](C)Oc(c23)cccc3	23306977	236	0,59	1,34
c1cccc1[C@@H]2CN(CCO2)C(=O)NCc3c(CC)n3(C)c3OC	93149538	237	0,51	1,43
COCc1cc(ncn1)NC(=O)N2C[C@@H](C)[C@H]2C)c3ccc(F)cc3	93316218	238	0,40	1,53
Cn1nnnc1S[C@H](C)C(=O)N[C@@H](C)c(c2)cc3c(c23)CCCC3	22784584	239	0,58	1,35
COCc1cc(ncn1)NC(=O)N2C[C@H](C)[C@H]2C)c3ccc(F)cc3	93316214	240	0,40	1,53
n1oc(C)cc1CN(C)C(=O)CCc2cnn(c2)-c3ccccc3	31657942	241	0,48	1,45
CCCN(n1)cc(Cl)c1C(=O)N(C2)CCC[C@@H]2Cc3cccn3	72384932	242	0,58	1,35
s1ccnc1CN(C)C(=O)NCc2cc(ccc2)-n3cccn3	80696288	243	0,48	1,45
FC(F)(F)c1cc(ccc1)[C@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3	92036660	244	0,79	1,13
FC(F)(F)c1cc(ccc1)[C@@H](C2CC2)NC(=O)N(C)C[C@@H]3CCOC3	92036658	245	0,79	1,13
c1cccc(c12)sc(n2)CN(C)C(=O)N[C@@H]3CCCN(c34)nc(n4)CC	87251107	246	0,19	1,73
CC(C)Cn(c1=O)c(=O)n(C)c(c12)nc3n2c(C)c(C)n3	4314324	247	0,37	1,55

CCCCC				
c1cccc1C(=O)N(C2)CCC[C@H]2C(=O)OCCc3cccc3	73299335	248	0,72	1,20
COc1n(C)nc(CC)c1CN[C@H]2CCOc(c23)c(Cl)cc3Cl	89211609	249	0,43	1,48
Clc1cnn(c1)CCNC(=O)CCn(nc2)c(c23)ccc(Cl)c3	40760451	250	0,68	1,23
CCN(CC)C(=O)CSc(o1)nnc1-c(c2)ccc(c23)cccc3	6416097	251	0,55	1,36
Cn(n1)cnc1CNC(=O)N(C2)CC[C@H]2Oc3c(Cl)cccc3	91775455	252	0,66	1,25
C1COCCC1OCCC(=O)Nc2ccc(OC)c(c23)nccc3	84532176	253	0,50	1,40
C1COCCN1CCNC(=O)Cc2coc(c23)ccc4c3cccc4	21874011	254	0,52	1,38
c1c(Cl)ccc(Cl)c1NC(=O)[C@H](C)Sc(n2CC)nnc2C3CC3	7618357	255	0,56	1,33
C[C@H](C1)[C@H]1C(=O)N2CCN(CC2)C(=O)NCCc(c3)oc(c34)cccc4	74380190	256	0,49	1,40
Cc(n1)scc1CCCC(=O)N(C2)CC[C@H]2Oc3cc(Cl)ccc3	89624334	257	0,54	1,35
Cc(n1)scc1CCCC(=O)N(C2)CC[C@H]2Oc3cc(Cl)ccc3	89624337	258	0,54	1,35
c1cccc(c12)OCC[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956457	259	0,70	1,19
c1cccc(c12)OCC[C@H]2C(=O)NCc3cc(ncc3)OC4CCC4	89956458	260	0,70	1,19

RF = random forest

Tabela C4 – Predição das propriedades farmacocinéticas ADME-Tox.

ZINC	Estrutura	AOH	Caco2	logBB	HERG	LAH
92964267	1	100,00	1325,58	-0,20	-4,66	0,38
20423620	2	93,89	589,78	0,46	-4,54	-0,16
72087376	3	100,00	578,95	-0,52	-3,90	0,48
12449018	4	100,00	956,12	-0,28	-4,70	0,14
6932434	5	100,00	1659,92	-0,36	-4,17	-0,19
67280610	6	100,00	797,92	-0,19	-4,81	0,21
63492379	7	96,24	953,08	-0,57	-4,82	-0,21
646627	8	100,00	517,93	-0,38	-4,21	0,17
66956124	9	100,00	666,48	-0,63	-4,37	0,32
44567226	10	94,09	1206,38	-0,56	-4,55	-0,43
71901309	11	100,00	6448,24	0,29	-3,87	0,17
75581011	12	100,00	1976,41	-0,13	-4,39	-0,45
4906567	13	100,00	2334,47	-0,15	-4,84	0,69
956542	14	100,00	1281,57	-0,46	-4,89	0,30
3459546	15	93,82	555,30	-1,06	-4,92	0,14
56919886	16	100,00	2274,18	-0,17	-4,86	0,36
79517712	17	100,00	1896,11	-0,15	-4,80	0,07
43344165	18	93,29	649,65	-0,53	-3,28	-0,06
92119304	19	89,15	556,80	-0,51	-1,84	-0,38
75510592	20	100,00	1480,90	-0,54	-4,78	0,43
87335960	21	100,00	1645,75	0,00	-4,82	0,06
55685381	22	100,00	3484,94	-0,04	-4,74	0,31
80779799	23	90,90	567,19	-0,86	-4,39	-0,22
30718569	24	93,85	680,11	-0,60	-3,92	-0,19
77247970	25	100,00	1890,31	-0,21	-3,34	0,11
40130295	26	100,00	692,72	-0,44	-4,12	0,45
93052953	27	100,00	2545,82	0,15	-3,73	-0,15
67386210	28	100,00	2056,86	-0,08	-4,15	-0,05

67386217	29	100,00	2555,65	-0,03	-4,36	-0,01
72509269	30	92,41	782,80	-0,38	-3,24	-0,17
11927649	31	100,00	1257,47	-0,39	-4,85	-0,17
74142415	32	100,00	1740,57	-0,21	-3,23	0,13
93053722	33	100,00	2064,40	0,03	-3,58	-0,16
93053726	34	100,00	2091,71	0,08	-3,79	-0,14
81067125	35	100,00	1386,20	-0,33	-4,72	0,65
69762537	36	94,10	664,12	-0,70	-3,24	-0,02
91788667	37	94,83	631,50	-0,54	-2,84	-0,02
68650118	38	92,24	781,57	-0,70	-3,72	-0,04
23478649	39	89,28	608,00	-1,00	-4,65	-0,18
20869699	40	94,45	504,32	-0,83	-4,70	0,15
12302511	41	100,00	1655,02	-0,40	-4,72	-0,20
27343935	42	100,00	2633,70	0,05	-4,19	0,10
33078299	43	95,21	521,06	-0,58	-3,87	0,21
79265262	44	100,00	1022,29	-0,25	-3,92	0,24
66178874	45	100,00	549,62	0,20	-4,84	0,52
79594184	46	94,29	707,05	-0,83	-4,77	0,01
81031040	47	100,00	2466,03	-0,31	-4,97	0,16
81031042	48	100,00	2021,52	-0,33	-4,38	0,09
80069368	49	93,52	1022,18	-0,78	-3,95	-0,27
88220932	50	95,96	1000,81	-0,51	-5,00	0,00
33159238	51	100,00	1746,20	-0,35	-2,87	-0,21
33159237	52	95,19	971,16	-0,55	-3,85	-0,17
3457326	53	100,00	1409,37	0,60	-4,88	0,03
72332141	54	100,00	957,62	-0,62	-4,95	0,05
80822767	55	93,60	1077,08	-0,27	-2,74	-0,41
89357203	56	100,00	1616,75	-0,34	-4,73	0,37
79531837	57	100,00	4164,58	0,04	-4,51	0,29
19146445	58	100,00	2581,81	-0,14	-4,50	-0,08
92076090	59	86,93	665,40	-0,67	-3,67	-0,91
48105992	60	86,31	753,11	-0,72	-3,94	-0,88
89956312	61	100,00	3864,61	0,11	-4,37	0,51
90593151	62	100,00	668,59	-0,67	-4,41	0,33
8133311	63	95,11	963,18	-0,20	-2,92	-0,32
11974390	64	88,68	787,58	-0,55	-4,01	-0,56
11974396	65	88,28	689,26	-0,55	-4,02	-0,51
40024984	66	100,00	1654,80	-0,22	-4,10	-0,10
12749371	67	100,00	1864,30	0,05	-3,87	-0,13
72153248	68	92,81	543,13	-1,05	-4,93	-0,15
92631569	69	100,00	786,62	-0,76	-4,47	0,30
72454569	70	92,47	718,49	-0,66	-4,33	-0,26
74893708	71	100,00	2200,69	-0,05	-3,88	0,54
7306435	72	100,00	1002,70	-0,01	-4,12	-0,13
72354119	73	100,00	2067,20	-0,04	-4,83	0,67
74616895	74	93,91	1046,28	-0,75	-3,71	-0,32
59188606	75	100,00	1730,56	-0,31	-4,87	0,29
48263350	76	93,42	1414,37	-0,55	-3,91	-0,86
45097288	77	89,59	563,21	-1,03	-4,48	-0,17
3448391	78	100,00	790,62	-0,21	-3,91	-0,14
3448394	79	95,78	767,15	-0,35	-3,82	-0,18
79682507	80	100,00	1856,30	-0,38	-3,95	0,48
77615954	81	89,01	743,39	-0,72	-4,73	-0,26
89811381	82	100,00	1446,31	-0,21	-4,96	0,13

44967753	83	100,00	3849,06	0,14	-2,68	-0,09
60857051	84	100,00	4401,59	-0,24	-4,72	0,01
89717833	85	100,00	2024,62	-0,29	-3,93	-0,13
89717835	86	100,00	1586,62	-0,36	-3,82	-0,12
5627281	87	100,00	1814,40	-0,33	-3,98	0,14
79577108	88	93,88	802,11	-0,72	-4,94	-0,24
42896953	89	100,00	1798,51	-0,19	-4,47	-0,03
65528008	90	91,61	759,84	-0,76	-3,97	-0,40
59411307	91	100,00	1366,30	-0,30	-4,31	0,25
3393368	92	100,00	1039,29	-0,62	-4,98	0,25
45494692	93	100,00	1289,26	-0,43	-4,69	0,40
39393772	94	100,00	750,99	-0,50	-3,92	0,40
39393770	95	100,00	1874,39	-0,20	-4,14	0,47
91046021	96	100,00	2880,12	0,00	-4,78	0,19
31601820	97	94,01	847,94	-0,34	-3,11	-0,43
92377367	98	100,00	1865,93	-0,07	-3,67	0,04
76245437	99	85,94	575,61	-0,99	-4,46	-0,41
73875207	100	94,59	730,77	0,14	-4,60	-0,03
4837255	101	94,28	504,11	-0,91	-3,94	0,17
12904725	102	100,00	1303,13	0,01	-4,71	0,18
72552497	103	93,21	663,04	-0,91	-4,18	-0,12
74718345	104	100,00	2334,60	-0,26	-4,34	-0,25
19456616	105	100,00	1433,70	-0,45	-4,20	-0,27
91045925	106	100,00	820,76	-0,44	-4,90	0,33
81348380	107	93,17	548,56	-0,63	-4,56	-0,18
70874454	108	100,00	1794,59	-0,61	-4,93	0,42
54849997	109	93,60	755,19	-0,60	-4,53	-0,20
93201872	110	100,00	2494,10	-0,24	-4,41	0,14
71749519	111	100,00	3227,70	-0,10	-4,95	0,49
78518847	112	100,00	2761,73	-0,22	-4,33	0,51
91701517	113	100,00	1204,84	-0,60	-3,63	0,33
80338882	114	100,00	1058,10	-0,41	-4,91	0,06
67406599	115	100,00	2891,91	0,34	-4,31	0,52
67406602	116	100,00	2836,33	0,34	-4,18	0,51
87166849	117	100,00	1470,28	-0,55	-4,97	0,15
78420281	118	93,42	1338,28	-0,38	-4,48	-0,70
91321262	119	100,00	1574,23	-0,26	-4,43	0,28
78090111	120	96,49	874,21	-0,39	-3,67	-0,04
66403521	121	100,00	5606,68	-0,03	-4,99	0,35
66403524	122	100,00	4927,16	-0,04	-4,40	0,29
12929356	123	100,00	740,98	-0,51	-4,83	0,23
77989783	124	100,00	917,24	-0,43	-4,63	0,17
77989787	125	100,00	1395,35	-0,27	-4,26	0,13
903744	126	100,00	1085,06	-0,60	-4,10	-0,12
46509084	127	96,58	747,31	-0,63	-4,18	0,13
73221057	128	100,00	2715,58	-0,31	-4,22	0,37
75156512	129	96,95	699,19	-0,44	-4,57	-0,07
90824539	130	100,00	3660,98	0,08	-2,65	0,31
79157214	131	100,00	1403,45	-0,43	-3,73	0,39
64344145	132	100,00	949,54	0,30	-4,74	0,31
69772036	133	100,00	1222,80	-0,38	-3,87	0,00
83175027	134	100,00	959,26	-0,47	-3,36	-0,01
14110875	135	91,85	589,22	-0,70	-3,98	-0,20
87166632	136	100,00	918,10	-0,77	-4,83	0,29
65520577	137	100,00	1063,36	-0,49	-4,93	0,20

72242521	138	100,00	1040,37	-0,40	-4,34	0,18
95439902	139	90,90	865,05	-0,63	-4,76	-0,55
76798231	140	100,00	5120,85	0,19	-4,65	0,29
89364439	141	100,00	1103,69	-0,17	-4,77	0,29
87455925	142	100,00	2890,52	-0,31	-4,99	0,02
55448395	143	88,27	641,02	-0,69	-2,92	-0,57
56960730	144	100,00	1101,69	-0,68	-4,83	0,40
76128493	145	100,00	2221,95	0,02	-3,63	0,27
42896682	146	100,00	1521,68	-0,26	-4,57	0,18
93939051	147	100,00	3064,29	-0,20	-4,06	0,21
78984628	148	100,00	5433,18	0,09	-4,75	0,22
88197817	149	100,00	1381,89	-0,26	-4,03	0,10
68561409	150	100,00	1628,78	0,09	-3,73	0,14
66204557	151	100,00	1993,47	-0,18	-3,48	0,34
40756727	152	95,31	758,26	-0,69	-4,36	0,00
8709934	153	100,00	2201,72	-0,27	-4,23	0,12
8939348	154	100,00	867,80	-0,46	-4,57	-0,04
95412325	155	100,00	2194,24	-0,17	-3,98	-1,02
95412324	156	94,40	1539,47	-0,32	-3,89	-1,01
43162190	157	100,00	1873,00	-0,25	-4,51	0,10
72532639	158	100,00	1919,52	-0,30	-4,10	-0,15
26413486	159	96,16	988,68	-0,56	-4,81	0,00
79400264	160	100,00	878,78	-0,66	-4,86	-0,12
90013735	161	100,00	717,00	-0,79	-4,76	0,11
72821379	162	100,00	2108,09	-0,28	-4,58	0,33
74542092	163	100,00	4011,71	-0,01	-3,72	-0,06
67589176	164	100,00	1813,29	-0,02	-4,93	0,32
80567565	165	100,00	1342,64	-0,40	-2,97	0,29
74073007	166	91,19	894,78	-0,56	-4,18	-0,56
75598283	167	100,00	2074,26	-0,30	-4,84	0,61
80941753	168	100,00	858,33	-0,53	-4,50	-0,31
90479947	169	100,00	1286,82	-0,23	-4,67	0,34
89955773	170	89,79	526,10	-0,86	-4,09	-0,23
79880561	171	100,00	1417,55	-0,47	-4,73	-0,06
74509023	172	100,00	6003,16	0,15	-4,46	0,28
47382594	173	96,45	697,11	-0,69	-4,98	0,22
72509551	174	88,40	500,10	-0,85	-4,00	-0,18
92841199	175	100,00	1501,69	-0,48	-4,79	-0,08
74566760	176	89,58	530,61	-0,91	-4,19	-0,36
79586889	177	100,00	1096,85	-0,77	-4,67	-0,04
65250562	178	100,00	1514,52	-0,38	-4,46	0,18
83177089	179	100,00	3187,34	-0,12	-4,01	0,25
32207533	180	100,00	1339,38	-0,22	-4,83	0,01
11641712	181	100,00	2572,29	-0,11	-4,07	-0,50
11641718	182	100,00	1702,95	-0,25	-4,25	-0,49
92958887	183	100,00	2104,30	-0,22	-4,15	0,47
75435037	184	95,18	601,07	-0,59	-4,71	0,14
7312811	185	95,07	1170,10	-0,59	-3,52	-0,38
78493097	186	100,00	3339,45	-0,20	-4,61	-0,03
80760513	187	96,94	1009,75	-0,57	-4,71	-0,24
93061622	188	95,65	524,76	-0,57	-3,43	0,30
35507406	189	83,44	574,70	-0,88	-2,87	-0,57
10653606	190	100,00	2087,70	-0,14	-4,60	0,21
80570599	191	100,00	2258,99	0,01	-3,67	0,06

92531571	192	100,00	2351,04	-0,34	-4,82	0,35
4433832	193	91,03	656,63	-0,94	-4,52	-0,26
25137544	194	100,00	966,90	-0,57	-4,85	0,05
6802489	195	92,64	895,74	-0,52	-4,32	-0,44
40197222	196	100,00	1071,75	-0,42	-4,12	0,21
84520887	197	94,06	1307,73	-0,63	-4,74	-0,41
40250743	198	100,00	1365,72	-0,25	-5,00	0,24
55196763	199	100,00	2184,51	0,20	-3,81	-0,40
44049212	200	100,00	798,76	-0,56	-4,69	0,25
3599992	201	100,00	973,60	-0,32	-4,38	0,37
92452672	202	100,00	1547,09	-0,56	-4,77	-0,22
24924387	203	93,18	610,53	-0,77	-4,39	-0,02
24924391	204	100,00	2664,74	-0,27	-3,98	-0,02
73959717	205	91,82	978,93	-0,71	-4,46	-0,52
84530501	206	100,00	1909,07	-0,43	-4,69	-0,11
83939506	207	94,92	602,88	-0,80	-4,85	0,09
87249506	208	100,00	3351,73	-0,11	-3,69	0,17
78342925	209	100,00	5189,68	0,07	-4,16	-0,31
40198730	210	100,00	1486,32	-0,36	-4,12	0,10
84223098	211	100,00	1473,18	-0,26	-2,29	-0,47
59818114	212	96,14	874,21	0,34	-4,79	0,04
92841424	213	100,00	2401,48	-0,21	-4,56	0,06
87561073	214	100,00	1744,71	-0,42	-4,47	-0,06
25137832	215	100,00	4312,37	-0,14	-4,48	0,28
57908442	216	100,00	2402,56	-0,28	-4,65	0,32
74191425	217	91,61	778,06	-0,75	-4,89	-0,11
80551568	218	93,29	585,56	-0,75	-4,73	0,01
77440429	219	100,00	1210,98	-0,18	-3,23	0,05
81282781	220	100,00	651,37	-0,52	-4,43	0,17
58244146	221	94,35	668,27	-0,56	-4,65	-0,15
92414547	222	91,71	1635,86	-0,31	-2,23	-0,93
92414544	223	92,44	1714,24	-0,30	-2,32	-0,91
40121323	224	100,00	1321,75	-0,39	-3,70	0,44
88521299	225	100,00	2079,47	-0,20	-4,76	0,17
78150926	226	100,00	1100,04	-0,63	-4,41	0,11
84977059	227	100,00	2284,84	-0,23	-4,93	0,10
78270310	228	100,00	3434,38	-0,01	-3,78	0,12
71735509	229	100,00	1329,89	-0,40	-4,38	0,45
31326556	230	100,00	1011,94	-0,19	-3,10	-0,15
22314745	231	100,00	5601,34	-0,04	-4,80	0,09
40323351	232	96,19	654,35	-0,40	-4,24	-0,08
7944046	233	93,99	929,03	-0,74	-4,55	-0,33
91302842	234	100,00	1769,93	-0,54	-4,76	0,49
9668971	235	94,76	907,91	-0,56	-4,47	-0,10
23306977	236	100,00	2093,49	-0,22	-4,25	-0,26
93149538	237	100,00	1121,87	-0,30	-3,61	0,41
93316218	238	100,00	1517,33	-0,15	-4,15	0,04
22784584	239	94,08	623,23	-0,61	-3,14	0,07
93316214	240	93,35	563,29	-0,44	-3,92	0,05
31657942	241	94,09	634,39	-0,84	-4,83	-0,14
72384932	242	100,00	2562,59	-0,16	-4,66	0,20
80696288	243	100,00	1904,70	-0,14	-4,46	0,11
92036660	244	100,00	6909,41	0,44	-3,34	0,39
92036658	245	100,00	5810,99	0,42	-2,88	0,34
87251107	246	100,00	1677,17	-0,27	-4,27	0,20

4314324	247	100,00	3037,93	-0,31	-4,31	-0,05
73299335	248	100,00	885,70	-0,71	-4,97	-0,05
89211609	249	100,00	1722,35	0,77	-4,85	0,72
40760451	250	95,26	518,99	-0,48	-4,44	-0,05
6416097	251	95,02	667,03	-0,62	-4,39	-0,17
91775455	252	94,35	682,49	-0,53	-4,46	-0,07
84532176	253	100,00	2830,82	-0,32	-4,80	-0,27
21874011	254	87,47	527,39	0,29	-4,13	-0,31
7618357	255	100,00	2759,67	0,06	-4,53	0,56
74380190	256	94,59	623,17	-0,50	-3,03	0,00
89624334	257	100,00	2740,58	0,04	-4,58	0,05
89624337	258	100,00	3212,35	0,16	-4,20	-0,02
89956457	259	100,00	1580,70	-0,31	-3,59	0,18
89956458	260	100,00	1510,75	-0,28	-4,14	0,24

ANEXO D – Conjunto de dados de inibidores da cruzaina utilizado na construção dos modelos de QSAR 2D e 3D,

OXADIAZOL

GLICOLAMIDA

TIOAMIDA