



**UNIVERSIDADE FEDERAL DE PERNAMBUCO**  
**CENTRO DE BIOCÊNCIAS**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM INOVAÇÃO TERAPÊUTICA**

**FREDERICO FÁVARO RIBEIRO**

**USO DE FERRAMENTAS DO PLANEJAMENTO RACIONAL DE FÁRMACOS  
AUXILIADO POR COMPUTADOR APLICADO A UMA SÉRIE DE  
TIOSEMICARBAZONAS E SEUS BIOISÓSTEROS FRENTE À DOENÇA DE  
CHAGAS**

**RECIFE - PE**

**2018**

**FREDERICO FÁVARO RIBEIRO**

**USO DE FERRAMENTAS DO PLANEJAMENTO RACIONAL DE FÁRMACOS  
AUXILIADO POR COMPUTADOR APLICADO A UMA SÉRIE DE  
TIOSEMICARBAZONAS E SEUS BIOISÓSTEROS FRENTE À DOENÇA DE  
CHAGAS**

Tese apresentada ao Programa de Pós-Graduação em Inovação Terapêutica, Área de Concentração Fármacos, Medicamentos e Insumos Essenciais para a Saúde, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de doutor em Inovação Terapêutica

Orientador: Prof. Dr. Francisco Jaime Bezerra Mendonça Junior  
Co-Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Scotti

**Recife - PE**

**2018**

**Dados Internacionais de Catalogação na Publicação (CIP) de acordo com ISBD**

Ribeiro, Frederico Fávaro

Uso de ferramentas do planejamento racional de fármacos auxiliado por computador aplicado a uma série de tiossemicarbazonas e seus bioisómeros frente à doença de chagas. – 2018.

153 f. : il.

Orientador: Francisco Jaime Bezerra Mendonça Junior.

Coorientadora: Luciana Scotti.

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Inovação Terapêutica, Recife, 2018.

Inclui referências e apêndices.

1. Chagas, Doença de 2. Farmacologia I. Mendonça Júnior, Francisco Jaime Bezerra (orient.) II. Scotti, Luciana III. Título.

616.9363

CDD (22.ed.)

UFPE/CB – 2018 - 221

**FREDERICO FÁVARO RIBEIRO**

**USO DE FERRAMENTAS DO PLANEJAMENTO RACIONAL DE FÁRMACOS  
AUXILIADO POR COMPUTADOR APLICADO A UMA SÉRIE DE  
TIOSEMICARBAZONAS E SEUS BIOISÓSTEROS FRENTE À DOENÇA DE  
CHAGAS**

Tese apresentada ao Programa de Pós-Graduação em Inovação Terapêutica, Área de Concentração Fármacos, Medicamentos e Insumos Essenciais para a Saúde, da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de doutor em Inovação Terapêutica

Aprovada em 28/02/18

**COMISSÃO EXAMINADORA**

---

Prof. Dr. Francisco Jaime Bezerra Mendonça Júnior / UEPB

---

Profa. Dra. Ana Cristina Lima Leite / UFPE

---

Prof. Dr. Marcus Tullius Scotti / UFPB

---

Prof. Dr. Sebastião José de Melo / UFPE

---

Profa. Dra. Janaina Versiani dos Anjos / UFPE

*A Bruna Lopes da Silva, amiga,  
companheira e esposa.*

## **AGRADECIMENTOS**

- Aos meus orientadores, Prof. Dr. Francisco Jaime Bezerra Mendonça Júnior e Prof<sup>a</sup>. Dr<sup>a</sup>. Luciana Scotti que me guiaram durante essa jornada e compartilharam duas das coisas que mais considero importantes: tempo e conhecimento;
- Ao Prof. Dr. Ricardo Olímpio de Moura e a Prof<sup>a</sup>. Dr<sup>a</sup>. Maria do Carmo Alves de Lima, que também compartilharam seu conhecimento, contribuindo com meu crescimento intelectual, bem como o desenvolvimento dessa tese;
- A Prof<sup>a</sup> Dr<sup>a</sup> Ana Cristina Lima Leite, por ter fornecido as estruturas e os dados de atividade biológica dos compostos apresentados nessa tese;
- A Antoniel Gomes por ter passado horas, muitas vezes durante as madrugadas me ensinando e ajudando a resolver os problemas operacionais relacionados com algumas das técnicas aqui aplicadas;
- A Murilo Morais, Gabriel Lopes e Lúcio Belom, que tiraram centenas de dúvidas quanto a utilização de alguns recursos do Terminal Linux e a elaboração dos scripts desenvolvidos durante a tese;
- Aos colegas e amigos que fiz durante essa jornada, e aqueles que já me acompanham desde o início da carreira acadêmica e que de uma forma ou de outra contribuíram para no meu desenvolvimento como pessoa e como cientista;
- A CAPES pelo aporte financeiro.

“Quem conhece os outros é inteligente.  
Quem conhece a si mesmo é sábio.  
Quem vence os outros é forte.  
Quem vence a si mesmo é poderoso.  
Quem se faz valer tem força de vontade.  
Quem é autossuficiente é rico.  
Quem não perde o seu lugar é estável.  
Quem mesmo na morte não perece, esse vive.”

Lao Tse

## RESUMO

A doença de Chagas é uma doença negligenciada ocasionado pelo *Trypanosoma cruzi*, sendo uma das principais causas de cardiomiopatia e insuficiência cardíaca em regiões endêmicas. Estimativas indicam que 8 milhões de pessoas estão infectadas, ocorrendo 15.000 mortes todos os anos, configurando-se como um grave problema de saúde pública. O tratamento quimioterápico disponível só é eficiente nos estágios iniciais da doença e possui efeitos colaterais graves que podem levar o paciente a interromper o tratamento. O Planejamento Racional de Fármacos Auxiliado por Computador (CADD) se apresenta como uma estratégia para a redução de tempo e custo no processo de busca por novos medicamentos, fornecendo informações estruturais úteis sobre o composto candidato a fármaco, como características estérica, eletrônica e lipofílicas; além de estabelecer as relações qualitativas e quantitativas entre a estrutura química e a atividade biológica de uma série de compostos. Essa tese teve como objetivo utilizar ferramentas do CADD para contribuir no desenvolvimento e otimização de novos candidatos a fármacos com atividade frente ao *Trypanosoma cruzi*. Para alcançar os objetivos propostos foram utilizadas diferentes ferramentas computacionais. 119 moléculas tiveram sua estrutura tridimensional determinada e geometria otimizada por meio da modelagem molecular. Foi realizada a triagem virtual entre os compostos e a enzima cruzaina. Descritores moleculares foram calculados e usados para realização de métodos quimiométricos e análise de QSAR (*Quantitative Structure-Activity Relationships*), como o PCA (Principal Component Analysis), CPCA (*Consensus PCA*) e PLS (*Partial Least Square*). Foram calculadas as superfícies eletrônicas de HOMO, LUMO e Potencial Eletrostático para cada composto. As moléculas foram submetidas ao citocromo P450 para predição dos possíveis metabólitos formados pela ação dessa enzima. De acordo com os resultados do virtual screening, os compostos com menor interação eram quase que exclusivamente de tiosemicarbazonas. Ao todo, 22 resíduos de aminoácidos participaram na formação do complexo ligantes-cruzaina entre os mais frequentes, estavam a Gly23, Cys25, Gly65, Asp161, Gly66 e His162. A análise quimiométrica e de QSAR não conseguiram explicar um percentual satisfatório da variância dos dados, com coeficientes de correlação e predição menores que 80 e 50, respectivamente, refletindo em modelos inadequados para a predição da atividade anti-*T. cruzi*. Não foram observadas diferenças representativas

entre os compostos ativos e não ativos de cada série na análise de superfícies eletrônicas. Foi possível prever os principais metabólitos formados após interação das moléculas com o citocromo P450. Através da docagem molecular, foram identificados os compostos com maior afinidade com o sítio ativo da cruzaina, entretanto não foi observada correlação entre os dados *in vitro* (na enzima) e os scores obtidos *in silico*. A baixa variação dos dados de atividade biológica, foi um dos fatores que afetaram a qualidade dos modelos de QSAR gerados. A predição dos metabólitos permite guiar o planejamento e otimização de novos compostos, incluindo o uso de outros métodos de CADD, visando potencializar a atividade biológica e reduzir a toxicidade dos compostos.

**Palavras-chave:** Doenças de Chagas. *Trypanosoma cruzi*. CADD. Modelagem molecular. QSAR.

## ABSTRACT

Chagas disease is a neglected disease caused by *Trypanosoma cruzi*, being one of the main causes of cardiomyopathy and heart failure in endemic regions. Estimates indicate that 8 million people are infected and 15,000 deaths occurring every year, shaping up as a serious public health problem. Available chemotherapy is effective only in the early stages of the disease and has serious side effects that may lead the patient to discontinue treatment. Computer Aided Drug Design (CADD) is presented as a strategy to reduce time and cost in the process of finding new drugs, providing useful structural information on the candidate drug compound, such as steric, electronic and lipophilic characteristics; besides establishing the qualitative and quantitative relations between the chemical structure and the biological activity of a series of compounds. This thesis aimed to use CADD tools to contribute to the development and optimization of new drug with activity against *Trypanosoma cruzi*. To achieve the proposed objectives, different computational tools were used. 119 molecules had their three-dimensional structure determined and geometry optimized through molecular modeling. Virtual screening between the compounds and the cruzain enzyme was performed. Molecular descriptors were calculated and used to perform chemometric methods and Quantitative Structure-Activity Relationships (QSAR) analysis, such as PCA (Principal Component Analysis), CPCA (Consensus PCA) and PLS (Partial Least Square). The electronic surfaces of HOMO, LUMO and Electrostatic Potential were calculated for each compound. The molecules were submitted to cytochrome P450 to predict the possible metabolites formed by the action of this enzyme. According to the results of virtual screening, the compounds with lower interaction were almost exclusively of thiosemicarbazones. 22 amino acid residues were involved in the formation of the ligands-cruzaine complex. Among the most frequent, were Gly23, Cys25, Gly65, Asp161, Gly66 and His162. The chemometric and QSAR analysis failed to explain a satisfactory percentage of the data variance, with correlation and prediction coefficients lower than 80 and 50, respectively, reflecting in models unsuitable for the prediction of anti-*T. cruzi* activity. No representative differences were observed between the active and non-active compounds of each series in the analysis of electronic surfaces. It was possible to predict the major metabolites formed after interaction of the molecules with cytochrome P450. Through molecular docking, the compounds with greater affinity with the active

site of cruzaina were identified, though no correlation was observed between the *in vitro* data (the enzyme) and the scores obtained *in silico*. The minor variation of the biological activity data was one of the factors that affected the quality of the generated QSAR models. The prediction of the metabolites allows guiding the planning and optimization of new compounds, including the use of other methods of CADD, aiming to potentiate the biological activity and reduce the toxicity of the compounds.

**Key-words:** Chagas disease. Trypanosoma cruzi. CADD. Molecular modeling. QSAR.

## LISTA DE FIGURAS

FIGURA 1 - DISTRIBUIÇÃO DOS CASOS DE INFECÇÃO POR TRYPANOSOMA CRUZI DURANTE O PERÍODO DE 2006 A 2009 .....	21
FIGURA 2 - CICLO DE VIDA DO TRYPANOSOMA CRUZI .....	22
FIGURA 3 - ESTRUTURA QUÍMICA DO BENZNIDAZOL E DO NIFURTIMOX .....	23
FIGURA 4 - ESTRUTURA TRIDIMENSIONAL DA CRUZAÍNA.....	25
FIGURA 5 - TIOSSEMICARBAZONAS COM ATIVIDADE ANTI <i>T. CRUZI</i> .....	26
FIGURA 6 - TIAZOLIDINONAS COM ATIVIDADE ANTI <i>T. CRUZI</i> .....	27
FIGURA 7 - COMPOSTOS 1,3-TIAZOIS COM ATIVIDADE ANTI <i>T. CRUZI</i> .....	27
FIGURA 8 - REPRESENTAÇÃO DOS MODELOS CHAVE/FECHADURA E DO AJUSTE INDUZIDO .....	32
FIGURA 9 - FÓRMULA GERAL DA PCA EM FORMA MATRICIAL .....	35
FIGURA 10 - GRÁFICO DOS LOADINGS DA PCA CONSIDERANDO PC1 E PC2 .....	36
FIGURA 11 - GRÁFICO DOS SCORES DA PCA CONSIDERANDO PC1 E PC2 APRESENTANDO DISTINÇÃO ENTRE COMPOSTOS MAIS E MENOS ATIVOS.	37
FIGURA 12 - GRÁFICO DOS PESOS DOS BLOCOS DO CPCA CONSIDERANDO PC1 E PC2.....	38
FIGURA 13 - ESTRUTURA DAS TIOSEMICARBAZONAS DA SÉRIE AD.....	41
FIGURA 14 - ESTRUTURA DAS TIOSEMICARBAZONAS DA SÉRIE JW.....	42
FIGURA 15 - MAPAS DE DENSIDADE ELETRÔNICA DA TIROSINA 103 EM DIFERENTES RESOLUÇÕES.....	45
FIGURA 16 - REPRESENTAÇÃO TRIDIMENSIONAL DA CRUZAÍNA (PDB: 3I06) .....	46
FIGURA 17 - LIGANTE (QL2) CRISTALIZANDO NO SÍTIO ATIVO DA CRUZAÍNA (PDB: 3I06) .....	46
FIGURA 18 - LOCALIZAÇÃO E POSICIONAMENTO DO GRID BOX NO SÍTIO ATIVO DA CRUZAÍNA .....	49
FIGURA 19 - POSES DE MENOR ENERGIA DE INTERAÇÃO DE TODOS OS COMPOSTOS NO SÍTIO ATIVO DA CRUZAÍNA.....	55

<b>FIGURA 20 - FREQUÊNCIA DE RESÍDUOS ENVOLVIDOS NA FORMAÇÃO DOS COMPLEXOS LIGANTE-RECEPTOR.....</b>	<b>57</b>
<b>FIGURA 21 - BLOCOS DE DESCRITORES OBSERVADOS NO CPCA E SEUS RESPECTIVOS PESOS .....</b>	<b>61</b>
<b>FIGURA 22 - POSIÇÃO DOS COMPOSTOS RESULTADOS DE ATIVIDADE FRENTE A CRUZAÍNA EM RELAÇÃO A PC1 E PC5.....</b>	<b>63</b>
<b>FIGURA 23 - POSIÇÃO DOS COMPOSTOS RESULTADOS DE ATIVIDADE FRENTE A FORMA AMASTIGOTA DE <i>T. CRUZI</i> EM RELAÇÃO A PC1 E PC5 ...</b>	<b>65</b>
<b>FIGURA 24 - POSIÇÃO DOS COMPOSTOS RESULTADOS DE ATIVIDADE FRENTE A FORMA EPIMASTIGOTA DE <i>T. CRUZI</i> EM RELAÇÃO A PC1 E PC5</b>	<b>67</b>
<b>FIGURA 25 - POSIÇÃO DOS COMPOSTOS RESULTADOS DE ATIVIDADE FRENTE A FORMA TRIPOMASTIGOTA DE <i>T. CRUZI</i> EM RELAÇÃO A PC1 E PC5 .....</b>	<b>69</b>
<b>FIGURA 26 - SUPERFÍCIES ELETRÔNICAS DO COMPOSTO MAIS ATIVO (GT-5B) E MENOS ATIVO (GT-06C) DA SÉRIE GT CONSIDERANDO A ATIVIDADE FRENTE A FORMA TRIPOMASTIGOTA DE <i>T. CRUZI</i>.....</b>	<b>73</b>
<b>FIGURA 27 - SUPERFÍCIES ELETRÔNICAS DO COMPOSTO MAIS ATIVO (LA-1A2) E MENOS ATIVO (LA-1B3) DA SÉRIE LA CONSIDERANDO A ATIVIDADE FRENTE A FORMA AMASTIGOTA DE <i>T. CRUZI</i>.....</b>	<b>74</b>
<b>FIGURA 28 - SUPERFÍCIES ELETRÔNICAS DO COMPOSTO MAIS ATIVO (LA-1C1) E MENOS ATIVO (LA-1B1) DA SÉRIE LA CONSIDERANDO A ATIVIDADE FRENTE A FORMA TRIPOMASTIGOTA DE <i>T. CRUZI</i>.....</b>	<b>74</b>
<b>FIGURA 29 - SUPERFÍCIES ELETRÔNICAS DO COMPOSTO MAIS ATIVO (LA-1C1) E MENOS ATIVO (LA-1A5) DA SÉRIE LA CONSIDERANDO A ATIVIDADE FRENTE A CRUZAÍNA.....</b>	<b>75</b>
<b>FIGURA 30 - SUPERFÍCIES ELETRÔNICAS DO COMPOSTO MAIS ATIVO (AD-9V) E MENOS ATIVO (AD-9I) DA SÉRIE AD .....</b>	<b>75</b>
<b>FIGURA 31 -COMPOSTO AD-9I E SEU METABÓLITO APÓS INTERAÇÃO COM O CITOCROMO P450.....</b>	<b>76</b>
<b>FIGURA 32 - COMPOSTO PT03 E SEU METABÓLITO APÓS INTERAÇÃO COM O CITOCROMO P450.....</b>	<b>76</b>
<b>FIGURA 33 - COMPOSTO JW-9E E SEU METABÓLITO FENÓLICO APÓS INTERAÇÃO COM O CITOCROMO P450. ....</b>	<b>77</b>

FIGURA 34 - COMPOSTO MB3-5 E SEU METABÓLITO FENÓLICO APÓS INTERAÇÃO COM O CITOCROMO P450. ....	77
FIGURA 35 - COMPOSTO GT-06A E SEU METABÓLITO APÓS INTERAÇÃO COM O CITOCROMO P450.....	77
FIGURA 36 - COMPOSTO GT-06 E SEU METABÓLITO HIDROXILADO APÓS INTERAÇÃO COM O CITOCROMO P450 .....	78
FIGURA 37 - COMPOSTO MB1-6 E SEU METABÓLITO HIDROXILADO APÓS INTERAÇÃO COM O CITOCROMO P450 .....	78
FIGURA 38 - COMPOSTO MB1-6 E SEU METABÓLITO (BENZOTIOL) APÓS INTERAÇÃO COM O CITOCROMO P450 .....	78
FIGURA 39 - COMPOSTO MB1-6 E SEU METABÓLITO (FTALIMIDA) APÓS INTERAÇÃO COM O CITOCROMO P450 .....	79

## LISTA DE TABELAS

TABELA 1 - ATIVIDADE BIOLÓGICA DOS COMPOSTOS DA SÉRIE AD .....	42
TABELA 2 - ATIVIDADE BIOLÓGICA DOS COMPOSTOS DA SÉRIE JW .....	43
TABELA 3 - PARÂMETROS DE CONFIGURAÇÃO DO GRID BOX.....	48
TABELA 4 - PARÂMETROS MODIFICADOS NO ARQUIVO DE PARÂMETROS DA DOCAGEM.....	50
TABELA 5 - SCORES DE ENERGIA EM ORDEM DECRESCENTE DE AFINIDADE, OBTIDOS PELA DOCAGEM MOLECULAR ENTRE OS LIGANTES E A CRUZAÍNA .....	56
TABELA 6 - COMPARAÇÃO DOS 10 COMPOSTOS MAIS ATIVOS E MENOS ATIVOS CONSIDERANDO AS DUAS ABORDAGENS ( <i>IN VITRO</i> E <i>IN SILICO</i> )...	59
TABELA 7 - RESULTADOS DA ANÁLISE DE CPCA .....	60
TABELA 8 - VARIÂNCIA EXPLICADA PELA ANÁLISE DE PCA CONSIDERANDO OS COMPOSTOS COM RESULTADOS DE ATIVIDADE FRENTE A CRUZAINA .	62
TABELA 9 - VARIÂNCIA EXPLICADA PELA ANÁLISE DE PCA CONSIDERANDO OS COMPOSTOS COM RESULTADOS DE ATIVIDADE FRENTE A FORMA AMASTIGOTA DE <i>T. CRUZI</i> .....	64
TABELA 10 - VARIÂNCIA EXPLICADA PELA ANÁLISE DE PCA CONSIDERANDO OS COMPOSTOS COM RESULTADOS DE ATIVIDADE FRENTE A FORMA EPIMASTIGOTA DE <i>T. CRUZI</i> .....	66
TABELA 11 - VARIÂNCIA EXPLICADA PELA ANÁLISE DE PCA CONSIDERANDO OS COMPOSTOS COM RESULTADOS DE ATIVIDADE FRENTE A FORMA TRIPOMASTIGOTA DE <i>T. CRUZI</i> .....	68
TABELA 12 - RESULTADO DO MODELO PLS GERADO PARA ATIVIDADE FRENTE A CRUZAÍNA.....	70
TABELA 13 - RESULTADO DO MODELO PLS GERADO PARA ATIVIDADE FRENTE A FORMA AMASTIGOTA DE <i>T. CRUZI</i> .....	71
TABELA 14 - RESULTADO DO MODELO PLS GERADO PARA ATIVIDADE FRENTE A FORMA EPIMASTIGOTA DE <i>T. CRUZI</i> .....	71
TABELA 15 - RESULTADO DO MODELO PLS GERADO PARA ATIVIDADE FRENTE A FORMA TRIPOMASTIGOTA DE <i>T. CRUZI</i> .....	72
TABELA 16 - VARIÂNCIA DOS VALORES DE PIC <sub>50</sub> PARA CADA ATIVIDADE...	72

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>17</b>
1.2 OBJETIVOS .....	18
<b>1.2.1 Objetivo Geral</b> .....	<b>18</b>
<b>1.2.2 Objetivos Específicos</b> .....	<b>18</b>
<b>2 REFERENCIAL TEÓRICO</b> .....	<b>20</b>
2.1 DOENÇA DE CHAGAS .....	20
2.2 CRUZAÍNA .....	24
2.3 NOVOS AGENTES COM ATIVIDADE ANTI <i>T. CRUZI</i> .....	25
2.4 MÉTODOS COMPUTACIONAIS COMO FERRAMENTAS PARA O PLANEJAMENTO DE COMPOSTOS BIOATIVOS .....	28
<b>2.4.1 Modelagem molecular</b> .....	<b>29</b>
<b>2.4.2 Docagem Molecular</b> .....	<b>31</b>
<b>2.4.3 Relação quantitativa estrutura-atividade (QSAR)</b> .....	<b>34</b>
<b>2.4.4 Quimiometria</b> .....	<b>34</b>
<b>3 METODOLOGIA</b> .....	<b>41</b>
3.1 OBTENÇÃO E PREPARO DAS ESTRUTURAS DOS COMPOSTOS .....	41
3.2 CONVERSÃO DOS ARQUIVOS .....	43
3.3 MACROMOLÉCULA .....	44
3.4 TRIAGEM VIRTUAL BASEADA NA ESTRUTURA DO RECEPTOR (DOCAGEM MOLECULAR).....	47
<b>3.4.1 Parametrização da Docagem Molecular</b> .....	<b>47</b>
<b>3.4.2 Script de execução da triagem virtual</b> .....	<b>50</b>
<b>3.4.3 Visualização dos resultados</b> .....	<b>50</b>
3.5 TEÓRICO X EXPERIMENTAL .....	51
3.6 QUIMIOMETRIA E QSAR .....	52
3.7 SUPERFÍCIES ELETRÔNICAS .....	53
3.8 PREDIÇÃO DE METABÓLITOS .....	53
<b>4 RESULTADOS E DISCUSSÃO</b> .....	<b>55</b>
4.1 TRIAGEM VIRTUAL BASEADO NA ESTRUTURA DO RECEPTOR (DOCAGEM MOLECULAR).....	55
4.2 QUIMIOMETRIA E QSAR .....	60
<b>4.2.1 CPCA</b> .....	<b>60</b>

<b>4.2.2 PCA.....</b>	<b>61</b>
<b>4.2.3 PLS .....</b>	<b>70</b>
<b>4.3 SUPERFÍCIES ELETRÔNICAS .....</b>	<b>73</b>
<b>4.4 PREDIÇÃO DE METABÓLITOS .....</b>	<b>76</b>
<b>5 CONCLUSÕES .....</b>	<b>80</b>
<b>REFERÊNCIAS.....</b>	<b>82</b>
<b>APÊNDICE A - ESTRUTURA DAS MOLÉCULAS .....</b>	<b>87</b>
<b>APÊNDICE B - ATIVIDADE BIOLÓGICA DOS COMPOSTOS.....</b>	<b>93</b>
<b>APÊNDICE C - INTERAÇÕES FORMADAS ENTRE OS LIGANTES E O SÍTIO ATIVO DA CRUZAINA .....</b>	<b>96</b>
<b>APÊNDICE D - METABÓLITOS OBTIDOS PELO METASITE .....</b>	<b>126</b>
<b>APÊNDICE E - Produção acadêmica DURANTE O DOUTORADO.....</b>	<b>150</b>

## 1 INTRODUÇÃO

Apesar das melhorias e no bom desempenho dos programas de controle dessa parasitose, a doença de Chagas continua sendo um problema de saúde pública, não apenas nos países endêmicos, mas em todo o mundo. Estima-se que 8 milhões de pessoas estão infectadas pelo *Trypanosoma cruzi*, agente causador da doença de Chagas, levando de 10.000 pessoas a óbito todos os anos (BERN, 2015; ILIES, 2014; MAGUIRE, 2015). Atualmente dois medicamentos estão disponíveis para o tratamento da doença de chagas, o Nifurtimox e o Benznidazol, ambos eficientes apenas na fase aguda da doença e apresentam efeitos colaterais graves, que comumente leva o paciente a abandonar o tratamento (MARTINEZ-MAYORGA et al., 2015). Praticamente meio século depois do lançamento desses medicamentos, eles ainda continuam sendo as únicas alternativas disponíveis, demonstrando a urgente necessidade pela busca por novos compostos com atividade anti-*T.cruzi* que sejam eficientes no tratamento na fase crônica e com baixa toxicidade.

O processo de busca por novos fármacos incluindo o tempo para que eles venham a ser comercializados, podia levar aproximadamente 15 anos, demandando um investimento de 1 bilhão de dólares (SHEKHAR, 2008). Visando contornar esse problema, otimizando a busca por novos compostos bioativos, aliado com o avanço tecnológico ocorridos nas últimas décadas, foram desenvolvidos os métodos do Planejamento Racional de Fármacos Auxiliado por Computador, do inglês *Computer-aided drug design* (CADD), que tem como uma das principais vantagens, dirigir a síntese de novas entidades químicas, diminuindo consideravelmente o número de compostos que necessitam ser sintetizados e avaliados, conseqüentemente, reduzindo o tempo e o custo no processo de desenvolvimento de novos candidatos a fármacos (SHEKHAR, 2008).

Os métodos computacionais vêm sendo cada vez mais utilizados tanto na academia como na indústria. Eles permitem obter informações úteis para o estudo da natureza estérica, eletrônica e lipofílica da provável interação do fármaco (ligante) com o alvo biológico (receptor), bem como estabelecer as relações quantitativas entre a estrutura química e a atividade biológica de uma série de compostos, auxiliando na compreensão do mecanismo de ação responsável pelo efeito biológico (COHEN, 1996; SCOTTI, 2006).

Dentre os métodos de CADD os mais utilizados são: a Docagem molecular (*Docking molecular*), que é capaz de simular a interação de um composto candidato a fármaco com o sítio ativo de uma macromolécula alvo, permitindo observar as interações envolvidas na formação do complexo; Os métodos de Relação Quantitativa Estrutura-Atividade, do inglês *Quantitative structure–activity relationship* (QSAR), um conjunto de técnicas capazes de calcular ou estimar propriedades/descriptores moleculares com o intuito de explicar a atividade biológica dos compostos com base em sua estrutura, sendo possível ainda gerar modelos preditivos da atividade para novos compostos.

Com base nessas considerações iniciais, que demonstram a necessidade da busca por novos medicamentos para o tratamento da doença de Chagas, bem como, as vantagens oferecidas pelas ferramentas computacionais para o planejamento racional de fármacos, esta tese visa a aplicação de diferentes ferramentas computacionais frente a uma série de tiossemicarbazonas e seus bioisómeros de modo a contribuir com o planejamento de novas substâncias voltadas ao tratamento das doenças de Chagas.

## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Utilizar ferramentas do Planejamento Racional de Fármacos Auxiliado por Computador em uma série de tiossemicarbazonas e seus bioisómeros, visando contribuir com o desenvolvimento e otimização de novos candidatos a fármacos para o tratamento da doença de Chagas.

### 1.2.2 Objetivos Específicos

- Executar a docagem molecular com emprego dos *softwares* AutoDock, AutoGrid v. 4.2 e AutoDockTools v. 1.5.6;
- Calcular superfícies eletrônicas com auxilia do *software* Spartan v. 14.0 e suas interpretações;

- Calcular vários tipos de descritores utilizando os softwares VolSurf v. 1.0.7.I e Pentacle v. 1.0.7.
- Construir e validar modelos de QSAR nos *softwares* VolSurf v. 1.0.7.I e Pentacle v. 1.0.7;
- Utilizar ferramentas quimiométricas através do VolSurf v. 1.0.7.I, Pentacle v. 1.0.7;
- Predizer metabólitos após ação de enzimas do citocromo P450, utilizando o *software* MetaSite v. 5.1.1.

## 2 REFERENCIAL TEÓRICO

### 2.1 DOENÇA DE CHAGAS

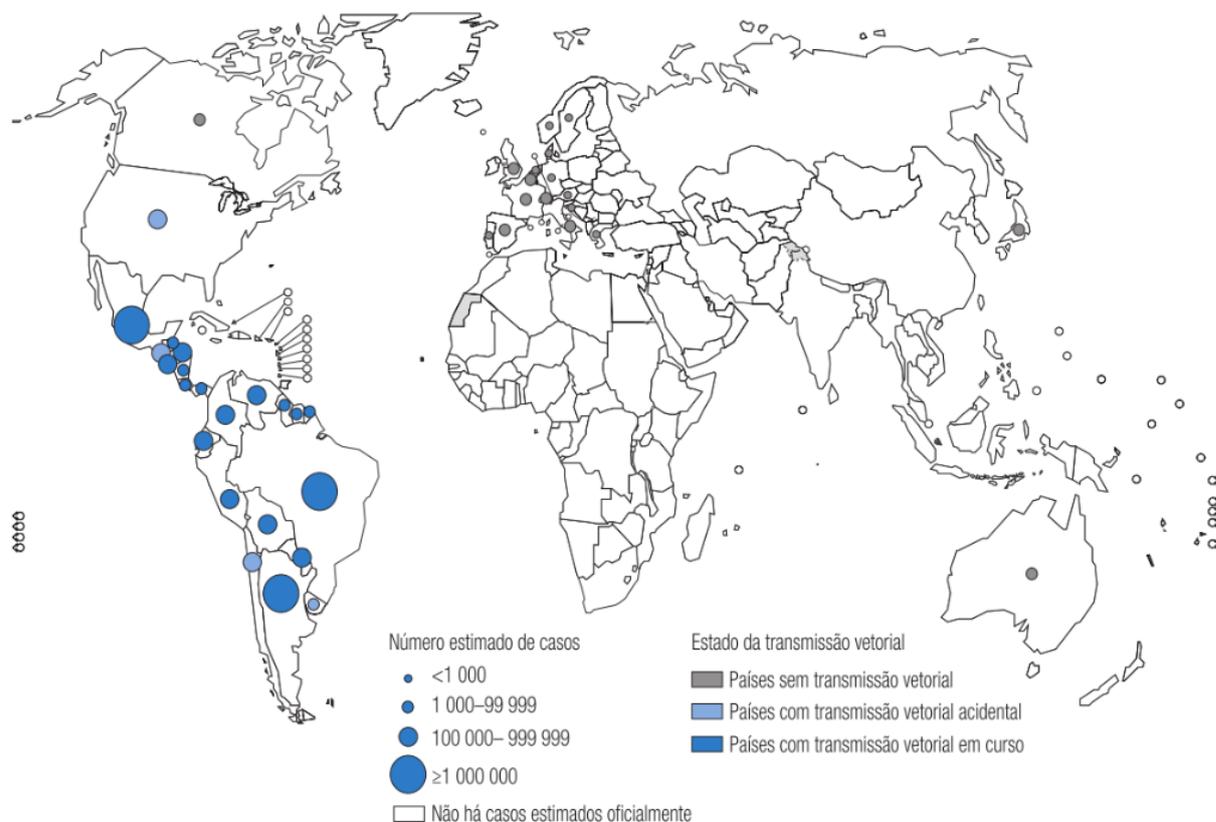
Também conhecida como tripanossomíase americana, a doença de chagas é causada pelo *Trypanosoma cruzi* (*T. cruzi*), protozoário da família Trypanosomatidae que tem como característica, a presença de um flagelo e uma única mitocôndria (SAÚDE, 2010).

Trata-se de um grave problema de saúde pública, estimando-se que 8 milhões de pessoas estão infectadas com o *T. cruzi*, dos quais, dois milhões evoluíram para a fase crônica da doença e ocasionando 14 mil óbitos por ano, sendo classificada como a doença parasitária que mais mata na América Latina (SOUZA, 2010).

É endêmica em 21 países da América Latina: Argentina, Belize, Bolívia, Brasil, Chile, Colômbia, Costa Rica, El Salvador, Equador, Guiana (Britânica), Guiana Francesa, Guatemala, Honduras, México, Nicarágua, Panamá, Paraguai, Peru, Suriname, Uruguai e Venezuela. Sendo o Brasil, responsável por 23% de todos os casos mundiais (ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2010; SOUZA, 2010).

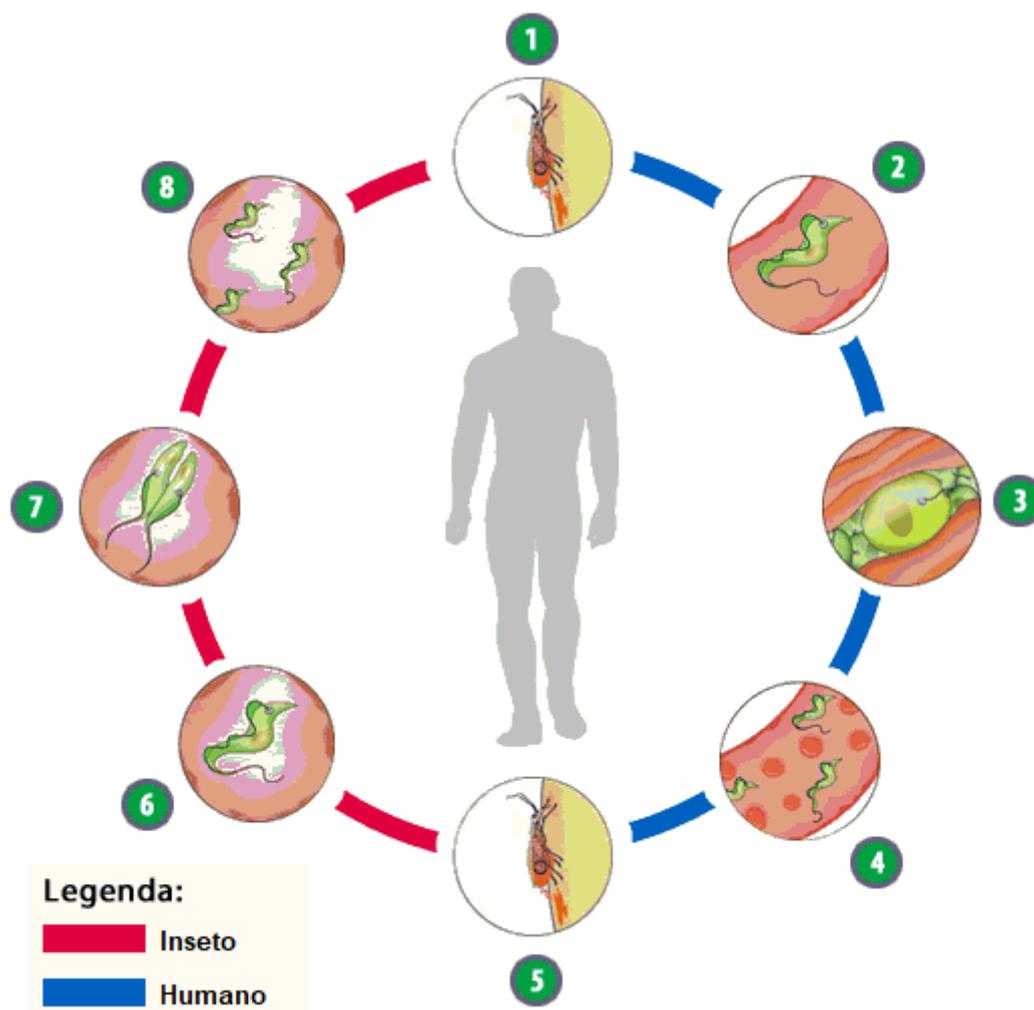
Outros países também começaram a apresentar casos da Doença de Chagas, impulsionado pela mobilidade populacional, seja por pessoas que visitaram as áreas endêmicas e adquiriram a doença, ou doentes que migraram para as áreas não endêmicas(ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2010) (Figura 1).

Figura 1 - Distribuição dos casos de infecção por *Trypanosoma cruzi* durante o período de 2006 a 2009



Fonte: Organização Mundial da Saúde (2010).

A principal forma de transmissão ocorre por intermédio do vetor, popularmente conhecido como barbeiro, insetos hematófagos da subfamília Triatominae infectado pelo *T. cruzi*. (1). Ao picar o seu alvo para se alimentar de sangue, o barbeiro defeca fezes cheias de *T. cruzi* na forma tripomastigota (Flagelado), (2) que penetra a mucosa alcançando a corrente sanguínea através da ferida da picada (Facilitado pelo paciente ao coçar a região da picada). (3) Ocorre a invasão das células do tecido infectado, onde o parasita se transforma em amastigota e multiplica-se. (4) Os amastigotas se transformam novamente em tripomastigotas e saem da célula, ganhando a corrente sanguínea e infectando novas células. (5) O barbeiro não infectado, ao se alimentar do sangue infectado, ingere a forma tripomastigota livre no sangue. (6) Ao alcançar o intestino do inseto ele se transforma em epimastigota e (7) multiplica-se. (8) O parasita na forma epimastigota se transforma novamente em tripomastigota sendo eliminado nas fezes do barbeiro, iniciando-se o ciclo (Figura 2) (RASSI; RASSI; MARIN-NETO, 2010).

Figura 2 - Ciclo de vida do *Trypanosoma cruzi*

Adaptado de: <http://profaerica-ciencias.blogspot.com.br/2016/03/trypanosoma-cruzi.html>.

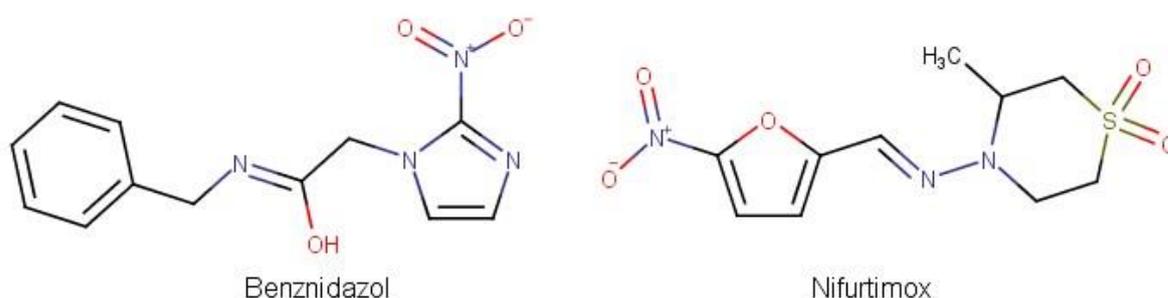
Outras formas de transmissão possíveis são: Oral, que ocorre através da ingestão de alimentos contaminados com o *T. cruzi*; Transfusional e por transplante de órgãos, que ocorrer devido a hemoderivados ou órgãos transplantados de doadores infectados; Vertical, onde mulheres grávidas e infectadas, podem transmitir para o feto durante a gestação ou parto; e por acidente, ao lidar com material contaminado sem os devidos equipamentos de segurança (SAÚDE, 2010).

A infecção se inicia na fase aguda, tendo como principal quadro clínico a miocardite aguda. Após aproximadamente 60 dias e sem tratamento adequado, a doença evolui para a fase crônica, ocorrendo redução do volume parasitário, mas causando insuficiência cardíaca, arritmias graves, aneurismas, acidentes tromboembólicos e morte súbita (COURA; DE CASTRO, 2002; SAÚDE, 2010). Quadro que se agrava em pacientes imunodeprimidos (DOYLE et al., 2011).

Atualmente não existem vacinas disponíveis, e o método preventivo mais efetivo consiste no controle do vetor e na prevenção da transmissão vetorial, transfusional, vertical, oral e acidental (NUSSBAUM et al., 2010; ORGANIZAÇÃO MUNDIAL DA SAÚDE, 2010; RASSI; RASSI; MARIN-NETO, 2010).

Para o tratamento, o medicamento de primeira escolha consiste no Benznidazol (Radanil® – Roche) (Figura 3), distribuído como comprimido de 100mg, com dose diária indicada de 5mg/kg do paciente, para os adultos, e de 5 a 10mg/kg, no caso de crianças, dose essa, que deve ser dividida e administrada de 2 ou 3 vezes ao dia, em um tratamento que pode durar 60 dias. Nos casos de intolerância ao Benznidazol, pode-se administrar o Nifurtimox (Lampit® - Bayer) (Figura 3), também disponível como comprimidos de 120mg, com doses indicadas de 15mg/kg e 8 a 10mg/kg, respectivamente, para crianças e adultos, durante o período de 60 a 90 dias (LO PRESTI et al., 2015; RASSI; RASSI; MARIN-NETO, 2010).

Figura 3 - Estrutura química do Benznidazol e do Nifurtimox



Fonte: RIBEIRO (2018).

Ambos os medicamentos apresentam eficácia boa ou moderada na fase aguda da doença, com uma taxa de cura de 70% para o Benznidazol e 50% para o Nifurtimox, contudo, ao se tratar dos casos crônicos, a taxa de cura é inferior à 20% (COURA; DE CASTRO, 2002). Além disso, esses medicamentos, apresentam graves efeitos colaterais, que incluem: anorexia, náuseas, vômitos, cólicas intestinais, dermatite, edema, mialgia, artralgia, neutropenia, agranulocitose, parestesia e polineurite (SILVA et al., 2009).

O longo período de tratamento juntamente com os efeitos colaterais desses medicamentos acaba por induzir os pacientes a abandonar o tratamento (DAVANÇO, 2015; ROBINSON; DALTON, 2011; SOUZA, 2017). Aliando esses fatores com o fato

de que o Nifurtimox e o Benznidazol foram lançados, respectivamente, em 1967 e 1972 e até então não terem sido substituídos, nota-se a necessidade e urgência na busca por novos compostos com atividade frente ao *T. cruzi*, propiciando um tratamento eficiente frente a fase crônica e com menos efeitos colaterais.

## 2.2 CRUZAÍNA

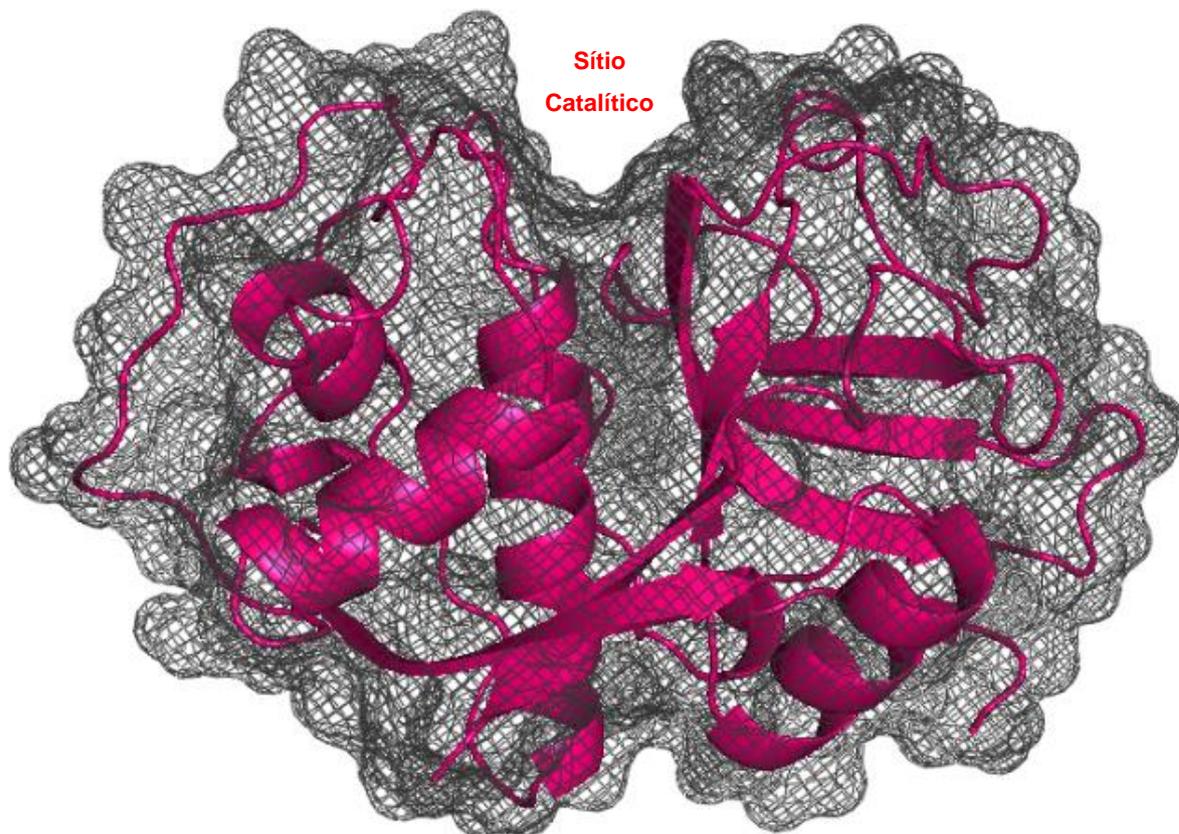
Através do sequenciamento do genoma do *T. cruzi*, novos alvos biológicos em potencial foram identificados, auxiliando o planejamento e desenvolvimento de substâncias bioativas. Dentre eles, destacam-se as Cisteína-proteases, que desempenham papel importante nas interações hospedeiro-parasita, na absorção de nutrientes, evasão do sistema imune e degradação dos tecidos do hospedeiro, sendo indispensável para a sobrevivência do *T. cruzi* (COURA; DE CASTRO, 2002; NUSSBAUM et al., 2010; SILVA et al., 2009).

A cruzaína é a protease mais abundante e mais estudada no *T. cruzi*. Pertence à família papaína e está presente em todas as fases evolutivas do parasita: no caso do epimastigota ele se encontra nos reservosomas, enquanto que no amastigota e tripomastigota, nos lisossomos da superfície celular (NUSSBAUM et al., 2010).

Acredita-se que a inibição da atividade da cruzaína leva ao acúmulo do precursor da protease no complexo de Golgi do parasita, causando lise dessa organela e do retículo endoplasmático, levando o *T. cruzi* a morte (ROBINSON; DALTON, 2011).

A estrutura da cruzaína é formada por dois domínios, um predominantemente  $\alpha$ -helicoidal e o outro formado por extensas interações de folhas  $\beta$ -antiparalelas. O sítio catalítico se encontra em uma cavidade, exatamente entre esses dois domínios (Figura 4).

Figura 4 - Estrutura tridimensional da cruzaina



Legenda: Em roxo, a representação das folhas  $\beta$ -antiparalelas e  $\alpha$ -helicoidais; A malha cinza, representa a superfície da enzima.

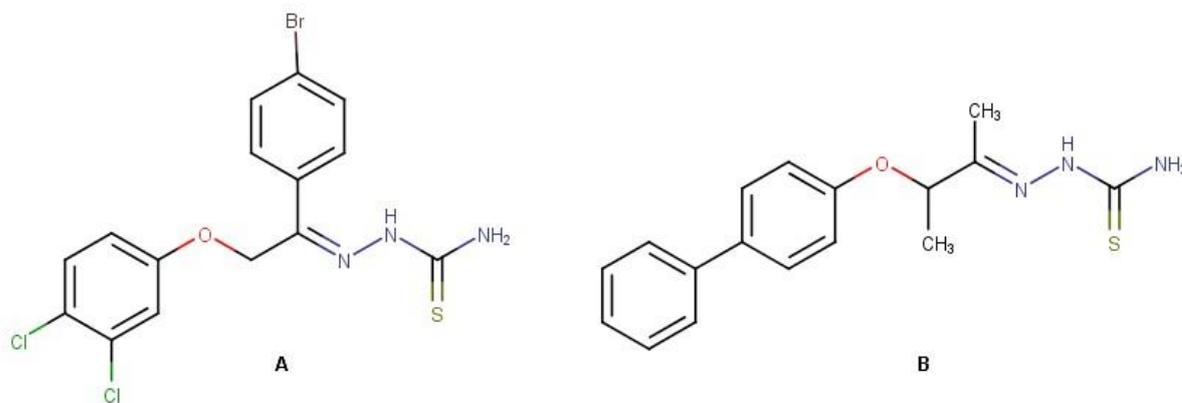
Fonte: RIBEIRO (2018).

### 2.3 NOVOS AGENTES COM ATIVIDADE ANTI *T. CRUZI*

Várias pesquisas estão sendo realizadas visando a obtenção de novos agente ativos frente ao *T.cruzi*. Em um estudo realizado em parceria com o Laboratório de Química Medicinal da Universidade Federal de Alagoas, foram investigados os compostos disponíveis na literatura que tiveram atividade inibitória da cruzaina com valores de  $IC_{50}$  inferiores a  $10 \mu M$ . Foi possível identificar 16 classes químicas com potencial atividade inibitória: piridiltiazol, criptolipinas, aril tiureias, aril ureias, diketopiperazinas, não-diketopiperazinas, peptidomiméticos, epoxi succinatos, epoxi cetonas, tetrahidronaftalenos, benzimidazolas, triazinas, chalconas, oxoindois, iminotiazolidinonas e tiossemicarbazonas (FERREIRA DA SILVA-JÚNIOR et al., 2017).

As tiossemicarbazonas foram amplamente avaliadas para inibição da catepsina L (KUMAR et al., 2010a, 2010b; SAXENA et al., 2016; SONG et al., 2012) e devido sua similaridade com a cruzaina, estudos foram realizados para avaliar a atividade biológica dessa classe química frente ao *T. cruzi*. Espíndola et al. (2015) sintetizaram e testaram a atividade anti *T. cruzi* (formas epimastigota e tripomastigota) de uma série de derivados ariloxi tiossemicarbazona, obtendo resultados promissores. Um dos compostos (Figura 5A), obteve  $IC_{50}$  de 1.8  $\mu M$  e  $CC_{50}$  de 3  $\mu M$  frente a forma epimastigota e tripomastigota respectivamente, resultados melhores que os obtidos pelo benznidazol ( $IC_{50}$  = 48.8  $\mu M$  e  $CC_{50}$  = 6.2  $\mu M$ ). Resultados similares foram obtidos por Moreira e colaboradores (2014) em derivados aril tiossemicarbazonas como o composto apresentado na Figura 5B, que obteve  $IC_{50}$  de 4.5  $\mu M$  frente a forma epimastigota e  $CC_{50}$  de 1.3  $\mu M$  para a forma tripomastigota do *T. cruzi*, resultado melhor que o obtido pelo benznidazol, que teve  $IC_{50}$  de 48.8  $\mu M$  frente a forma epimastigota e  $CC_{50}$  de 6.2  $\mu M$  para a forma tripomastigota; e melhor que o nifurtimox que apresentou  $IC_{50}$  de 5.7  $\mu M$  e  $CC_{50}$  de 2.7  $\mu M$  frente a forma epimastigota e tripomastigota respectivamente.

Figura 5 - Tiossemicarbazonas com atividade anti *T. cruzi*

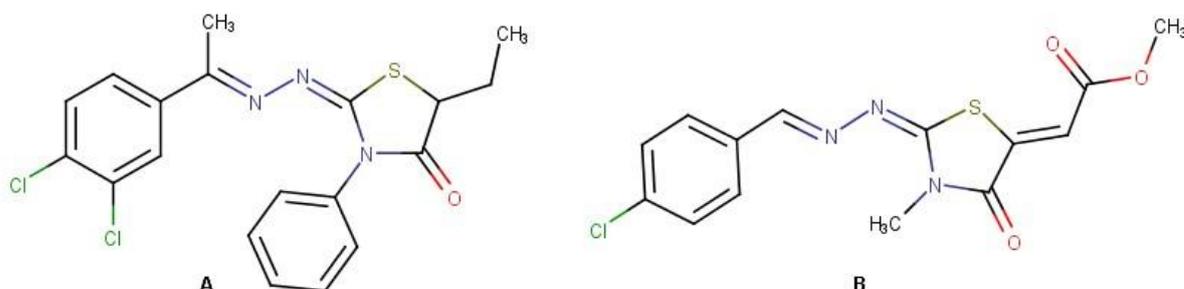


Fonte: RIBEIRO (2018).

Em outro estudo, a tiossemicarbazona foi usada como intermediário para a síntese de novos derivados tiazolidínicos, apresentando atividade frente a forma tripomastigota do *T. cruzi*, melhor que o benznidazol em alguns compostos, como é o caso do composto apresentado na Figura 6A, que obteve um  $IC_{50}$  de 2.9  $\mu M$  frente a forma tripomastigota comparado ao benznidazol com  $IC_{50}$  de 10.6  $\mu M$  (DE OLIVEIRA FILHO et al., 2015). Pizzo et al. (2011) também usaram a tiossemicarbazona como

intermediário para sintetizar uma série de derivados 2-hidrazolil-4-tiazolidinona, testando sua atividade frente a cruzaina e a forma epimastigota do *T. cruzi*, conseguindo, em um de seus compostos (Figura 6B), inibir 75% das cruzainas.

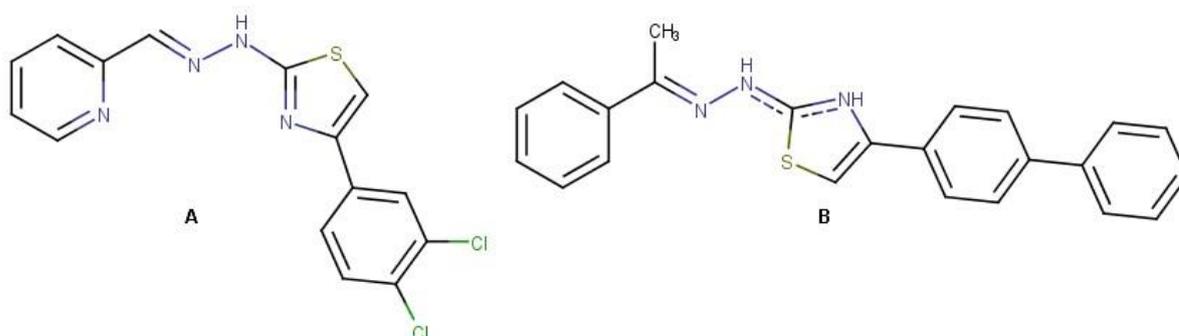
Figura 6 - Tiazolidinonas com atividade anti *T. cruzi*



Fonte: RIBEIRO (2018).

Outros estudos, também usaram a tiossemicarbazona como intermediário para produzir novos compostos 1,3-tiazóis. Cardoso e colaboradores (2014), testaram derivados 2-piridil-tiazóis, frente a forma tripomastigota e epimastigota, onde alguns compostos apresentaram resultados melhores que o Benznidazol ( $IC_{50} = 6.2 \mu M$  frente a forma tripomastigota e  $6.6 \mu M$  frente a forma epimastigota), como pode ser exemplificado pelo composto da Figura 7A, que teve  $IC_{50}$  de  $3.1 \mu M$  e  $5.0 \mu M$  frente as formas tripomastigota e epimastigota do *T. cruzi*. Resultado semelhante ao obtido por Gomes et al. (2016) em derivados 3-bromopropiofenona-hidrazinil-tiazol, como a molécula representada na figura 7B, obtive  $IC_{50}$  de  $3.35 \mu M$  frente a forma tripomastigota, resultado melhor comparado com o do Benznidazol,  $IC_{50}$  de  $6.26 \mu M$ .

Figura 7 - Compostos 1,3-tiazóis com atividade anti *T. Cruzii*



Fonte: RIBEIRO (2018).

## 2.4 MÉTODOS COMPUTACIONAIS COMO FERRAMENTAS PARA O PLANEJAMENTO DE COMPOSTOS BIOATIVOS

O Planejamento Racional de Fármacos Auxiliado por Computador é uma estratégia para a redução dos gastos bem como otimiza a pesquisa na busca de novos medicamentos para o tratamento de diversas doenças, inclusive das negligenciadas. Os métodos usados para identificação de alvos para potenciais candidatos a fármacos e avaliar a interações patógeno-hospedeiro, bem como a química combinatória para a descoberta de novos compostos bioativos e a tecnologia recombinante se tornaram algumas das dez biotecnologias de maior importância para a melhora da saúde em países em desenvolvimento (DAAR et al., 2002).

Estima-se que processo de descoberta e desenvolvimento de um novo fármaco até que ele chegue ao mercado leve ao menos 10 anos, custando algo em torno de 2.6 bilhões de dólares. Além disso, dos candidatos a fármacos que chegam até os testes clínicos, menos de 12% são aprovados pela FDA (*Food and Drug Administration*), o que representa um ou dois medicamentos aprovados (PHRMA, 2015; SHEKHAR, 2008).

O método de ensaio *in vitro* de rastreio de alto rendimento, testa milhões de compostos frente ao alvo de interesse. Este procedimento, além de ser caro e demorado, obtém taxas de acerto de aproximadamente 0.02%. Em paralelo, quando utilizado os métodos de planejamento racional de fármacos auxiliado por computador, do inglês *Computer-aided drug design* (CADD), as taxas de acerto podem chegar a 35% sem nem mesmo ter sido necessário realizar a síntese dos compostos, gerando uma redução significativa de tempo e de custo (SHEKHAR, 2008). Ao realizar um estudo teórico prévio, o número de compostos a serem sintetizados e testados na procura por melhores substâncias bioativas reduz sensivelmente, otimizando as investigações de novas entidades químicas candidatas a fármacos.

As técnicas de CADD são capazes de descrever e explicar os resultados experimentais, sendo possível testá-los antes mesmo da realização dos experimentos de bancada (MONTANARI, 2000). Os resultados preditivos obtidos destas ferramentas permitem identificar características estruturais essenciais das moléculas investigadas que são responsáveis pela atividade biológica, ou seja, o farmacóforo. Devido a essas vantagens, as metodologias CADD estão cada vez mais presentes na academia e nas indústrias farmacêuticas.

Vale salientar que mesmo com os futuros avanços computacionais, como hardwares mais poderosos, novos *softwares* e atualizações dos já existentes, surgimento e evolução de novas técnicas e algoritmos voltados para o CADD, os experimentos tradicionais *in vitro*, *ex vivo* e *in vivo* serão sempre necessários, uma vez que os sistemas biológicos são extremamente complexos e de natureza não linear, onde pequenas variações em suas condições (pH, temperatura, concentração iônica, etc) podem fazer como que se comportem de maneira completamente diferente, o que não pode ser completamente simulado em computadores (SHEKHAR, 2008).

#### **2.4.1 Modelagem molecular**

A Modelagem Molecular (MM) é um conjunto de técnicas computacionais baseadas nos métodos da química teórica e dados experimentais, possibilitando a compreensão das interações moleculares de modo quantitativo e qualitativo.

Essas técnicas permitem observar e analisar detalhes relacionados a maquinaria molecular, compreender o funcionamento dos sistemas biológicos, além de fornecer ferramentas que possibilitam prever os potenciais dos compostos candidatos a fármacos (SANT'ANNA, 2009). Isso é possível, uma vez que fornecem informações quanto às configurações moleculares, como geometria molecular, aspectos eletrônicos, moleculares e atômicos além de forças hidrofóbicas (COHEN, 1996; RIBEIRO et al., 2015; SCOTTI, 2006).

Durante os anos 80, com o avanço da tecnologia computacional, química estrutural teórica e experimental, da biotecnologia, cristalografia de raio X e RMN, o desenvolvimento e utilização das técnicas de MM foram intensificadas (COHEN, 1996).

A MM representa a relação estrutura atividade com base na estrutura molecular tridimensional, aproximando-se mais da representação real dos sistemas biológicos. Essa representação é obtida por meio de cálculos de parâmetros ou propriedades moleculares, relativos à estrutura molecular, gerando informações sobre suas propriedades, como ponto de fusão, massa molecular, propriedades termodinâmicas, estado físico, reatividade, solubilidade, inibição enzimática, antagonismo e metabolismo (RIBEIRO et al., 2015; SCOTTI, 2006).

Propriedades moleculares como ponto de fusão podem ser computadas ou estimadas com base nas estruturas e conformações de moléculas individuais que são obtidas através de algoritmos de interpolação, extrapolação e cálculos computacionais. No primeiro, é obtida uma correlação entre a propriedade de interesse e outra propriedade de moléculas relacionadas. Na extrapolação, a correlação não inclui a molécula de interesse, a correlação é estendida de modo a possibilitar estimar a propriedade. No caso dos cálculos computacionais, as propriedades são obtidas em diferentes níveis de precisão podendo em alguns casos, se equiparar a precisão obtida experimentalmente, sendo possível substituir os experimentos quando eles são de difícil execução ou impossíveis de serem realizados, podendo ser obtidos por métodos teóricos de mecânica molecular, métodos semi-empíricos e cálculos *ab initio* de mecânica quântica (LEACH, 2001; VASILESCU, D.; GIRMA, 2002).

A mecânica molecular consiste no cálculo das geometrias e energias conformacionais das moléculas, fazendo uso de campos de força empíricos, sendo os mais comuns os campos de Allingers, MM2 e MM3. Os campos possuem parâmetros que podem ser ajustados para obter o melhor ajuste possível das propriedades calculadas e experimentais. Os métodos se baseiam na suposição de que os parâmetros correspondentes e as constantes de força são transferíveis de uma molécula para outra e usa os princípios da espectroscopia vibracional. É feita uma suposição quanto ao comprimento e ângulos “naturais” de ligação, os desvios resultam em tensões de ligação e tensões angulares. A energia é minimizada selecionando os diedros flexíveis da conformação inicial e submetendo-os a sucessivos cálculos onde métodos de busca que envolvem rotações em ligações moleculares são usados, procurando explorar diferentes conformações, distâncias geométricas, dentre outras variáveis otimizando a geometria. Os parâmetros são modificados por pequenos incrementos até que a energia estrutural alcance o mínimo local, o que não necessariamente representa o mínimo global. Normalmente, são métodos mais simples e rápidos, sendo possível lidar com moléculas grandes contendo muitos átomos (COELHO et al., 1999; SANT’ANNA, 2009; SCOTTI, 2006).

Os métodos Semi-empíricos além de calcular a geometria da molécula, também consideram os orbitais moleculares. Baseiam-se na mecânica quântica, contudo, fazendo uso de dados experimentais. Formulações matemáticas de funções de ondas são empregadas para descrever os orbitais, sendo as principais funções os

orbitais Slater e gaussianos. Os orbitais interagem com um campo de força de auto-consistência, minimizando a energia do sistema (núcleo-elétron). Os diferentes métodos semi-empíricos se diferenciam pela aproximação realizada considerando a repulsão entre elétrons em diferentes orbitais. Os primeiros métodos foram baseados na teoria de Huckel estendida e CNDO (*complete neglect of differential overlap*), contudo, métodos como o MINDO-3 (*Modified Intermediate Neglect of Differential Overlap*) (BINGHAM; DEWAR; LO, 1975), MNDO (*Modified Neglect of Diatomic Overlap*)(DEWAR; THIEL, 1977) e AM1 (*Austin Model 1*) (DEWAR et al., 1985), foram desenvolvidos, gerando resultados mais confiáveis. O MOPAC, por exemplo, inclui os métodos MINDO-3 e MNDO, e fornece informações quanto a cargas atômicas, momento dipolo, potencial de ionização, dentre outras(COHEN, 1996; SCOTTI, 2006).

Os cálculos *Ab-initio* fazem uso da equação exata de Schrodinger, sem utilizar nenhum tipo de aproximação ou dados experimentais, apenas em parâmetros e conceitos da química teórica. Baseia-se nos métodos de campo auto-consistente (Método de *Hartree-Fock*) que considera as funções de onda. Envolvem a população eletrônica total da molécula, tratando um elétron de cada vez e interagindo-os com os demais elétrons. Assume-se um conjunto de orbitais e então é calculada a energia de repulsão elétron-elétron. A energia obtida é utilizada para calcular um novo conjunto de orbitais, repetindo o processo sucessivamente até se obter a auto-consistência. Quanto mais orbitais forem considerados, mais preciso será o cálculo e mais tempo será necessário para sua realização. Os métodos *Ab-initio* fornecem resultados quanto à geometria do estado fundamental molecular, energia, distribuição de cargas, além de possibilitar a descrição dos processos de reação e cálculo dos estados de transição e intermediários. São recomendados quando não existem dados experimentais, contudo exige uma grande demanda computacional (COHEN, 1996; SCOTTI, 2006).

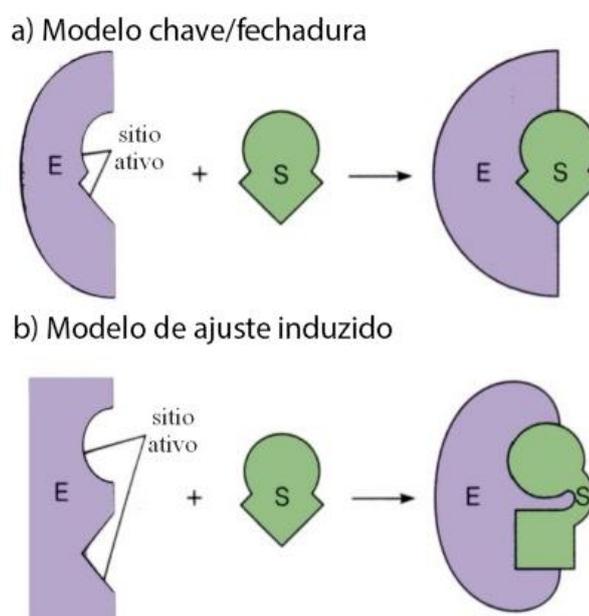
#### **2.4.2 Docagem Molecular**

A docagem molecular é um dos métodos de modelagem molecular mais utilizados. Ele permite determinar a orientação e conformação (pose) do ligante candidato a fármaco junto ao sítio ativo da macromolécula alvo, possibilitando verificar a afinidade entre eles, bem como observar quais resíduos e tipos de interação estão participando na formação do complexo ligante-receptor. Desta forma, ele tenta

reproduzir o processo natural da interação ligante-receptor regido por forças intermoleculares, como a hidrofobicidade, forças de *Van der Waals*, ligações de hidrogênio, interações eletrostáticas e ligações iônicas (COHEN, 1996; RIBEIRO et al., 2015).

Ele se baseia no modelo inicialmente proposto por Fisher em 1894, onde um composto se ligaria especificamente ao sítio ativo rígido de uma macromolécula, sendo popularmente conhecido como modelo “chave-fechadura” (Figura 8a). Estudos posteriores verificaram que a macromolécula não era uma estrutura rígida, podendo se apresentar em diferentes conformações, sendo proposto, em 1958 por Koshland, o modelo de ajuste induzido, onde o ligante promove pequenas mudanças conformacionais no sítio ativo da macromolécula permitindo uma melhor ligação (Figura 8b)(DURRANT; MCCAMMON, 2010).

Figura 8 - Representação dos modelos chave/fechadura e do ajuste induzido



Fonte: <http://docentes.esalq.usp.br/luagallo/image005.png>.

O método pode ser resumido em duas etapas: A primeira, consiste na amostragem de conformações, gerando diferentes conformações do ligante no sítio ativo da macromolécula, sendo obtidos por algoritmos de amostragem como os algoritmos de correspondência, construção incremental, busca simultânea de cópias múltiplas, LUDI, Monte Carlo, algoritmos genéticos e dinâmica molecular; Na segunda etapa as conformações são ranqueadas a partir de uma função de escores que estimam a afinidade de ligação, permitindo selecionar as melhores conformações e

poses. As funções de escores podem ser baseadas: no campo de força, empíricos ou baseado no conhecimento (MENG et al., 2011).

Para a realização da docagem molecular, é essencial ter conhecimento da estrutura tridimensional dos ligantes, bem como da macromolécula. Esta última, pode ser obtida por métodos experimentais como a cristalografia de raio x, a espectroscopia de ressonância magnética (RMN) e microscopia eletrônica (BERMAN et al., 2000). Atualmente o pesquisador possuem a disposição bancos de dados online como o *Protein Data Bank* (PDB) que armazena, gerencia e disponibiliza as estruturas tridimensionais das macromoléculas de forma gratuita, sendo acessado via internet, além de estar implementado em vários *softwares*. O PDB já conta com aproximadamente 140591 estruturas cadastradas, sendo 44165 sequências de proteínas distintas.

É desejável que o ligante natural do sítio ativo da macromolécula esteja presente na estrutura 3D, desse modo facilitando a identificação do sítio ativo e ainda permitindo comparar as afinidades de ligação entre o ligante natural e as moléculas candidatas a fármacos. Nos casos em que isso não for possível, podem ser utilizados métodos de identificação de cavidades na superfície da macromolécula, ou ainda, considerar toda a macromolécula como área de busca do sítio ativo, processo que é denominado de *blind docking* (MENG et al., 2011).

Com a docagem molecular é possível realizar triagem virtual com uma grande quantidade de ligantes, permitindo selecionar aqueles que apresentam maiores possibilidades de formar o complexo ligante-receptor, baseando-se no ranking de energias na formação do complexo para cada escore (RIBEIRO, 2014).

Apesar do seu potencial algumas limitações podem ser observadas: a) A necessidade de se possuir a estrutura 3D da macromolécula em boa resolução, caso contrário não é possível realizar a docagem ou seu resultado não será preciso; b) A maioria dos softwares não consideram a flexibilidade do receptor, e quando consideram, apenas as cadeias laterais que são consideradas flexíveis ou exigem um grande poder computacional; c) Não ter conhecimento da localização do sítio ativo pode diminuir a precisão ou inviabilizar a análise.

### 2.4.3 Relação quantitativa estrutura-atividade (QSAR)

O QSAR é um método que considera a variação da atividade biológica observada em um conjunto de compostos com base em suas diferenças estruturais gerando um modelo matemático formado por parâmetros ou descritores, avaliado por índices estatísticos, que tem o objetivo de estimar/prever a atividade de novas moléculas (ALCAIDE, 2009; GARCIA-DOMENECH et al., 2015).

Os descritores moleculares consistem na informação química referente a estrutura molecular dos compostos, transformada por um procedimento matemático em números úteis que podem ser analisados por métodos de análise multivariada. Para realização desse método, não é necessário conhecer a estrutura do alvo biológico, contudo é necessário ter conhecimento da estrutura química do ligante e ter dados de atividade biológica dos mesmos ( $IC_{50}$ , MIC,  $ED_{50}$ , etc). Nos modelos clássicos de QSAR (2D-QSAR), são utilizados descritores físico-químicos, como hidrofobicidade, refratividade molar e parâmetros eletrônicos. O desenvolvimento de métodos de 3D-QSAR, o qual considera a estrutura tridimensional das moléculas, possibilitou a aplicação de cálculos de campo de força que gera informações quanto às características estéricas, regiões hidrofóbicas e campos eletrostáticos (SPECK-PLANCHE; CORDEIRO, 2013).

Quando encontrado uma correlação entre a estrutura química e a atividade biológica, é possível obter informações sobre propriedades que estão mais relacionadas com a atividade em questão, permitindo assim o planejamento de novas estruturas químicas com potencial biológico mais pronunciado (KARELSON; LOBANOV; KATRITZKY, 1996). Além disso, pode-se gerar um modelo matemático preditivo, permitindo que outros compostos candidatos a fármacos sejam testados antes mesmo de serem sintetizados e estimando seu potencial biológico através do direcionamento da síntese. Ou ainda propor um farmacóforo para a síntese, para posteriormente ter a atividade biológica investigada e se tornar um potencial candidato a fármaco.

### 2.4.4 Quimiometria

A quimiometria é a área da química que extrai informações importantes de dados multivariados, através de ferramentas matemáticas e estatísticas. As técnicas

mais empregadas nos estudos quimiométricos são a Análise dos Componentes Principais (PCA) e a Regressão de Mínimos Quadrados Parciais (PLS).

### *Análise dos Componentes Principais (PCA)*

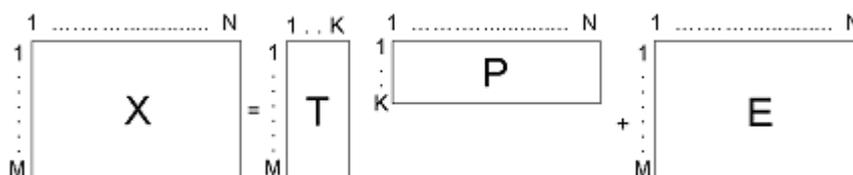
O PCA consiste em um método de análise multivariada que permite verificar tendências em um determinado grupo de objetos com base em diversas variáveis. Neste método os dados são organizados em uma matriz  $X$ , e reduzidos a duas matrizes menores, *loading* ( $P$ ) e *score* ( $T$ ) e o resultado é um sumário descrito por Componentes Principais (PC) (ALCAIDE, 2009; SANT'ANNA, 2002; SCOTTI, 2006).

A matriz  $X$  é formada por  $M$  linhas, que representam os objetos e  $N$  colunas, referentes às variáveis. A matriz *loading* ( $P$ ) é uma matriz  $K \times N$ , formada por  $K$  vetores, denominadas componentes principais (PC), que contém informações sobre as variáveis, enquanto que a matriz *score* ( $T$ ) contém informações sobre os objetos e possui dimensão  $M \times K$ . A fórmula geral da PCA é dada pela equação 1,

$$X = TP' + E \quad (1)$$

onde  $E$  é uma matriz de resíduo  $M \times N$ , formada pelas informações que não foram explicadas pelo modelo. Essa mesma equação pode ser demonstrada de forma matricial como demonstrado na figura 9.

Figura 9 - Fórmula geral da PCA em forma matricial

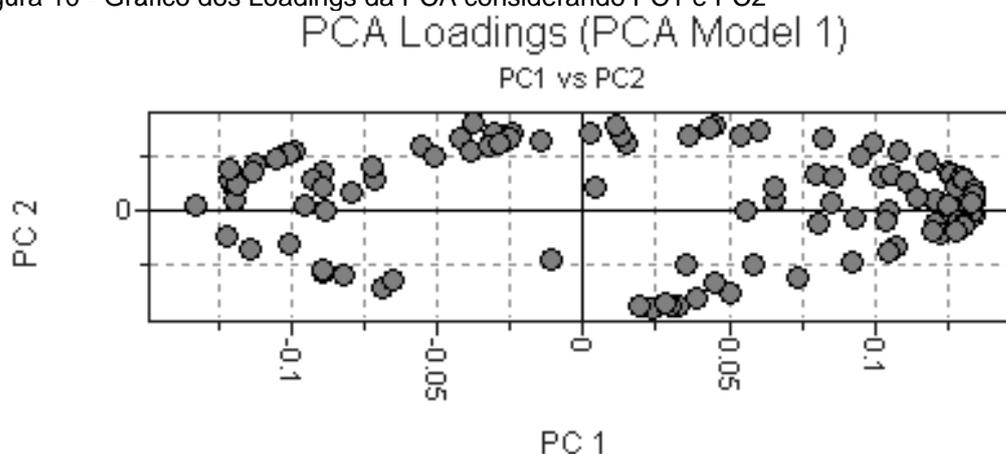


Fonte: ALCAIDE (2009).

As PCs são ortogonais em relação às outras, e com isso não apresentam correlação entre si e são totalmente independentes. Cada PC extraída, contém uma maior quantidade de informações e conseqüentemente busca explicar melhor a variâncias do modelo (ALCAIDE, 2009; SCOTTI, 2006). Os resultados obtidos podem ser plotados em gráficos 2D e 3D demonstrando a posição relativa das componentes

principais tanto para os *loadings* (variáveis), como para os *scores* (objetos), sendo possível observar com facilidade a presença de *outliers*. No primeiro caso, se pode observar a influência das variáveis, aquelas com pesos menores, tem pouca contribuição para o modelo e se localizam mais ao centro, enquanto que as variáveis de maior peso, contribuem mais para o modelo e localizam-se na região periférica do gráfico (Figura 10). No plot scores é possível verificar se o modelo foi capaz de distinguir os compostos mais ativos dos menos ativos, como demonstrado na Figura 11 (ALCAIDE, 2009; SCOTTI, 2006). Os círculos menores representam os compostos menos ativos, e nesse caso, ficaram concentrados mais à esquerda do gráfico, enquanto os círculos maiores, compostos mais ativos, ficaram concentrados a direita.

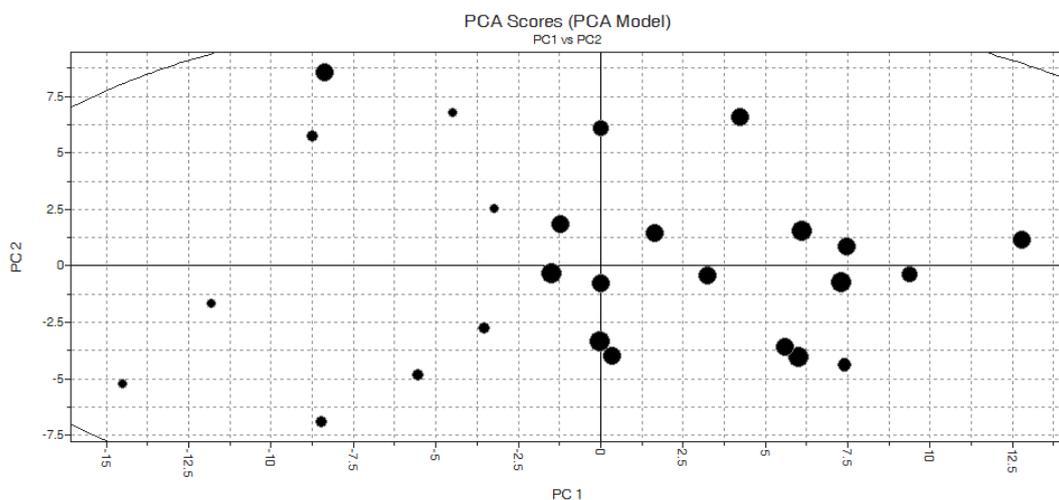
Figura 10 - Gráfico dos Loadings da PCA considerando PC1 e PC2



Legenda: Variáveis com maior contribuição localizados na periferia, e várias de menor contribuição no centro do gráfico.

Fonte: RIBEIRO (2018).

Figura 11 - Gráfico dos Scores da PCA considerando PC1 e PC2 apresentando distinção entre compostos mais e menos ativos

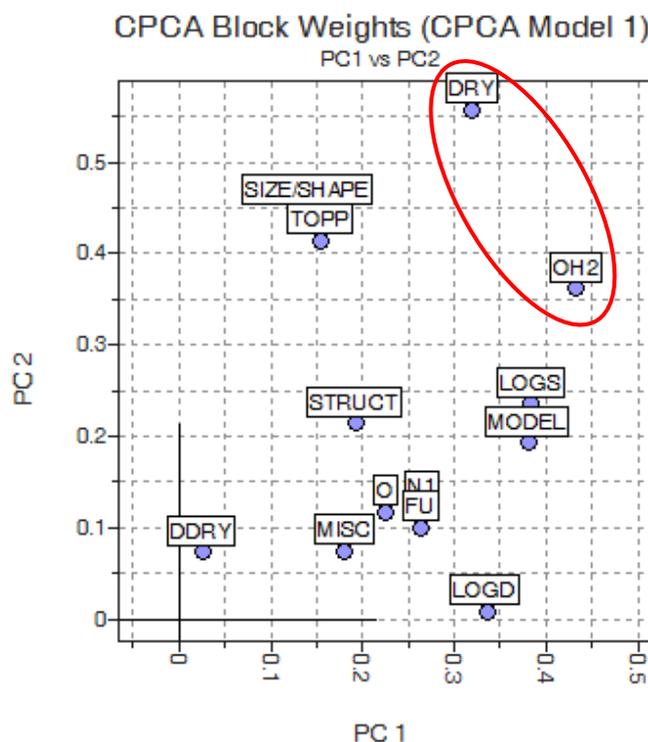


Legenda: Círculos menores representam compostos mais ativos e círculos maiores representam compostos menos ativos.

Fonte: RIBEIRO (2018).

Com o intuito de melhorar a interpretação dos modelos, foi desenvolvido uma derivação da PCA que introduz o conceito de blocos múltiplos de descritores, ou seja, as variáveis são agrupadas em blocos de acordo com um perfil comum. Esse método foi denominado de Consensus PCA (CPCA) e é usado nos casos em que há uma grande quantidade de variáveis disponíveis no modelo (WESTERHUIS; KOURTI; MACGREGOR, 1998). Esta metodologia permite uma seleção prévia dos descritores que influenciam o modelo. Exemplificando, na figura 12 tem-se um gráfico dos pesos dos blocos do CPCA. No caso, os blocos que apresentam maior influência no modelo, explicam a maior variância em PC1 e PC2, no caso DRY e OH2, que agrupam respectivamente, variáveis relacionadas com a hidrofobicidade, e hidrofiliidade. Dessa forma, acreditaríamos que o equilíbrio do parâmetro logP (solubilidade em octanol/água) deve ser importante no composto biologicamente ativo; e partindo dessa suposição poderíamos realizar outras análises, como a regressão de mínimos quadrados parciais (PLS), considerando apenas as variáveis pertencentes a esses dois blocos, que confirmem nossa hipótese, ou não.

Figura 12 - Gráfico dos pesos dos blocos do CPCA considerando PC1 e PC2



Fonte: RIBEIRO (2018).

### *Regressão de mínimos quadrados parciais (PLS)*

Consiste em uma análise de regressão que tenta explicar a variação de uma variável dependente  $Y$  (ex: atividade biológica) em função de uma diversidade de variáveis independentes  $X$ , ou no caso, os descritores calculados. O PLS tem a vantagem de ser capaz de lidar com situações onde a quantidade de variáveis é muito superior a quantidade de observações. Esta ferramenta é considerada um estudo de QSAR, pois descritores são calculados com base na estrutura molecular e a atividade biológica é considerada. Um diferencial deste método para a PCA é a possibilidade de gerar um modelo preditivo capaz de estimar valores da atividade biológica de moléculas (ALCAIDE, 2009; MORELLATO, 2010; WOLD; SJÖSTRÖM; ERIKSSON, 2001).

A equação geral do PLS segue o mesmo princípio da regressão linear múltipla, como descrito na equação 2:

$$Y = XB + E \quad (2)$$

onde  $Y$  é uma matriz  $M \times N$ , sendo  $M$  as observações (no caso, os compostos) e  $N$ , as variáveis dependentes;  $X$ , uma matriz  $M \times K$ , sendo  $K$  os valores dos preditores das  $M$  observações;  $B$  é uma matriz  $K \times N$  dos coeficientes de regressão e  $E$ , a matriz de resíduo com mesma dimensão da matriz  $Y$  (MAITRA; YAN, 2008; MORELLATO, 2010; WOLD; SJÖSTRÖM; ERIKSSON, 2001).

O PLS busca um conjunto de fatores ortogonais como as PCs, aqui chamadas de variáveis latentes (VL), que decompõem as matrizes  $X$  e  $Y$ , como demonstrado nas equações 3 e 4 (MAITRA; YAN, 2008; MORELLATO, 2010).

$$X = TP^T + E \quad (3)$$

Onde  $T$  é a matriz *score* de  $X$ , com dimensão  $M \times A$ , sendo  $A$  as VLs;  $P$  é a matriz *loading* de  $X$ , com dimensão  $A \times N$  e  $E$  a matriz dos resíduos de  $X$ .

$$Y = UQ^T + F \quad (4)$$

Sendo  $U$  a matriz *score* de  $Y$ , com dimensão  $M \times A$ ;  $Q$  é a matriz *loading* de  $Y$ , com dimensão  $A \times N$  e  $F$  a matriz dos resíduos de  $Y$ .

Como foi dito anteriormente, as VLs são equivalentes as PCs geradas no PCA. Ambas são extraídas em ordem decrescente de importância, ou seja, a primeira VL possui mais informações que o segundo e assim sucessivamente. Assim como as PCs, cada VL é ortogonal em relação a outra, não existindo correlação entre elas. Apesar das semelhanças, as VLs se diferem das PCs, pois o vetor ortogonal do PLS é obtido representando tanto a estrutura da matriz  $X$  quanto da matriz  $Y$ , maximizando os ajustes em ambas, enquanto que as PCs representam apenas a melhor estrutura da matriz  $X$  (SCOTTI, 2006).

Para verificar a robustez e a capacidade de predição do modelo realiza-se o teste de validação cruzada (*cross-validation* (CV)), que gera vários modelos com objetos reduzidos, usados para prever os valores das variáveis  $Y$ , que são comparadas aos valores observados. Por fim é calculado o desvio padrão do erro de predição (DPEP) (Equação 5) e o coeficiente de correlação preditivo ( $Q^2$ ) (Equação 6) (ALCAIDE, 2009; SCOTTI, 2006).

$$DPEP = \sqrt{\sum \frac{(Y - \hat{Y})^2}{M}} \quad (5)$$

Onde  $Y$  é o valor experimental,  $\hat{Y}$  o valor predito e  $M$  o número de objetos.

$$Q^2 = 1 - \left[ \frac{\sum (Y - \hat{Y})^2}{\sum (Y - \bar{Y})^2} \right] \quad (6)$$

Sendo  $\bar{Y}$  o valor médio de  $Y$ .

Este método de validação é considerado uma técnica de predição interna, permitindo estimar a capacidade preditiva do modelo sem a utilização de valores externos, extremamente importante quando o número de objetos/compostos for pequeno (SCOTTI, 2006).

Outra forma de verificar a capacidade preditiva do modelo é a predição externa. Nesse método, os objetos são divididos em dois grupos, o primeiro é o grupo de treinamento, que é usado para construir o modelo PLS, enquanto o segundo, grupo de teste, é usado para verificar se o modelo é capaz de prever os valores de  $Y$ , comparando os valores experimentais com os valores preditos pelo modelo (SCOTTI, 2006).

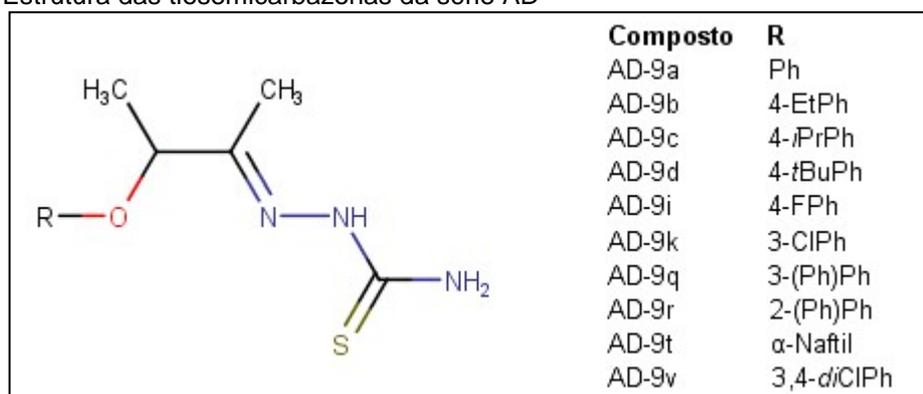
### 3 METODOLOGIA

#### 3.1 OBTENÇÃO E PREPARO DAS ESTRUTURAS DOS COMPOSTOS

As moléculas usadas nesse trabalho foram cedidas pelo Laboratório de Planejamento em Química Medicinal (LpQM) da Universidade Federal de Pernambuco, que forneceu a estrutura no padrão SMILES, bem como os respectivos valores de IC<sub>50</sub> em µM frente a enzima cruzaina e aos três estágios de desenvolvimento do *Trypanosoma cruzi* (amastigota, tripomastigota e epimastigota).

O conjunto é composto por 119 moléculas, pertencentes a 14 séries de compostos, divididos em 3 classes químicas distintas: 1,3-Tiazóis, 4-Tiazolidinonas e Tiossemicarbazonas. As estruturas das moléculas e os valores de atividade biológica estão dispostas nas figuras 13 e 14 e tabelas 1 e 2 (de forma a exemplificar), e nos apêndices.

Figura 13 - Estrutura das tiosemicarbazonas da série AD



Legenda: Ph = fenil; Et = etil; *i*Pr = isopropil; *t*Bu = terc-butil.

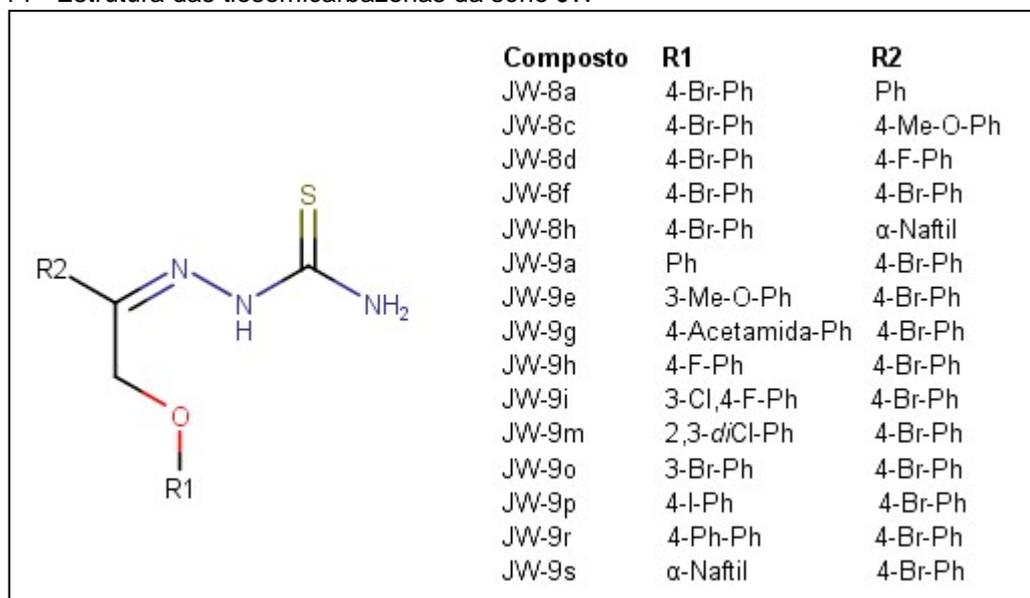
Fonte: RIBEIRO (2018).

Tabela 1 - Atividade biológica dos compostos da série AD

Composto	Amastigota	Tripomastigota	Epimastigota	Cruzaina
AD-9a	210.6	27.3	26.1	-
AD-9b	18.8	4.6	11.6	-
AD-9c	17.9	1.5	5.9	-
AD-9d	17	1.4	4.6	-
AD-9i	>391.8	23.2	86.4	-
AD-9k	36.8	5.5	45.8	-
AD-9q	31.9	1.3	2.2	-
AD-9r	31.9	-	2.2	-
AD-9t	34.8	1.5	4.9	-
AD-9v	16.3	4.4	4.4	-

Fonte: RIBEIRO (2018).

Figura 14 - Estrutura das tiosemicarbazonas da série JW



Legenda: Ph = fenil; Me = metil.

Fonte: RIBEIRO (2018).

Tabela 2 - Atividade biológica dos compostos da série JW

Composto	Amastigota	Tripomastigota	Epimastigota	Cruzaina
JW-8a	68.6	2.4	14.5	67.6
JW-8c	2.5	5	3.1	38.7
JW-8d	2.6	3.6	3.8	53.1
JW-8f	22.6	3.9	11.9	11.8
JW-8h	2.4	14.7	7.5	73.6
JW-9a	137.3	6.6	3	61.8
JW-9e	2.5	1.2	3.7	62.8
JW-9g	2.4	3.5	15.6	80.7
JW-9h	2.6	3.1	1.8	84
JW-9i	2.4	1.7	3.4	93.7
JW-9m	23.1	5.7	4.4	81.8
JW-9o	2.7	6.2	3.9	81.3
JW-9p	2.1	2.1	2.9	63.3
JW-9r	22.7	12.3	4.1	63.7
JW-9s	2.4	3.4	3	-

Fonte: RIBEIRO (2018).

A estrutura dos compostos em SMILES foram convertidas para 2D através do uso do *software* MarvinSketch versão 17.24.0 para Linux. Cada composto foi salvo individualmente em um arquivo no formato .mol e nomeado de acordo com o código do composto.

O *software* HyperChem versão 8.0.7 para Windows, foi utilizado para converter as estruturas 2D para suas representações 3D e posteriormente foi realizado a otimização geométrica a partir do método de mecânica molecular MM+ e o método semi-empírico AM1. As estruturas otimizadas, foram salvas no formato .mol.

### 3.2 CONVERSÃO DOS ARQUIVOS

Alguns dos softwares usados na execução desse trabalho requerem que os arquivos estejam em extensões de arquivos específico sendo necessário a conversão dos arquivos originais para o formato mais adequado.

Nesses casos, foi utilizado o *software* OpenBabel versão 2.4 para Linux. Ele é executado a partir de comandos via Terminal. O comando pode ser generalizado da seguinte forma:

```
babel -imol *.mol -opdb *.pdb
```

onde **babel** executa o *software*, **-imol** é a indicação da extensão do arquivo de entrada (**-i**: input; **mol**: extensão do arquivo original), **\*.mol** indica que todos os arquivos .mol do diretório atual serão convertidos, **-opdb** indica que a extensão de saída será .pdb (**-o**: output; **pdb**: formato de saída) e **\*.pdb** indica que todos os arquivos devem ser salvos mantendo o nome original e o formato .pdb.

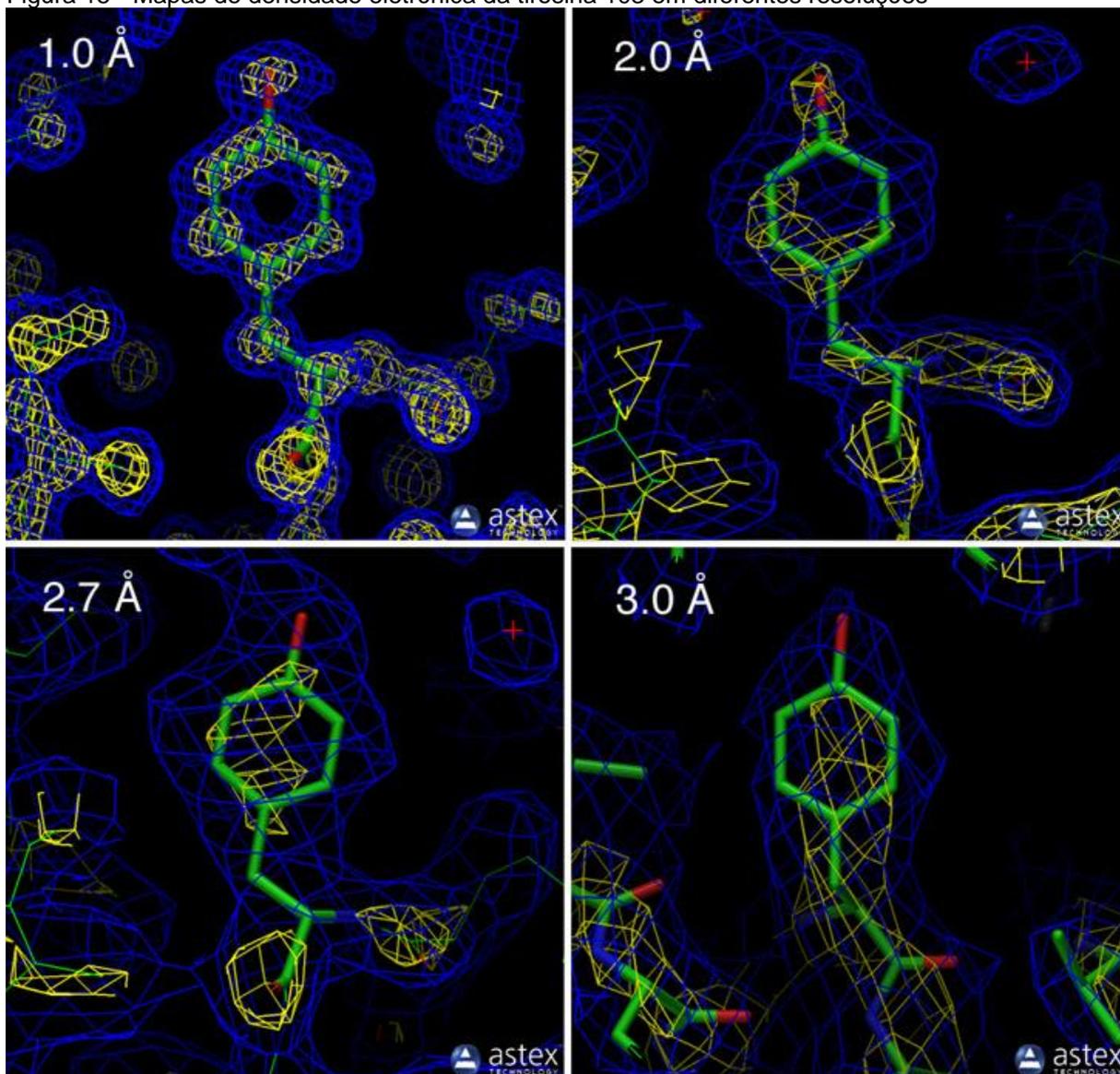
### 3.3 MACROMOLÉCULA

A estrutura tridimensional da cruzaina foi obtida diretamente do Protein Data Bank (PDB). Ao buscar pela macromolécula, foram encontradas 25 estruturas registradas para o *Trypanosoma cruzi*.

*A priori*, a seleção da melhor estrutura cristalográfica, levou em consideração 2 critérios:

1. A resolução do arquivo: Segundo o Guia de Compreensão dos Dados PDB (RCSB PDB, 2017), mede o nível de detalhes contidos no padrão de difração do raio-X através do cristal contendo a proteína, bem como o nível de detalhes observados no mapa de densidade eletrônica calculado para a proteína cristalizada. Essa resolução é medida em Ångström (Å) e quanto menor o valor, melhor é a qualidade do arquivo, devido ao detalhamento de informação contido nele (Figura 15).

Figura 15 - Mapas de densidade eletrônica da tirosina 103 em diferentes resoluções



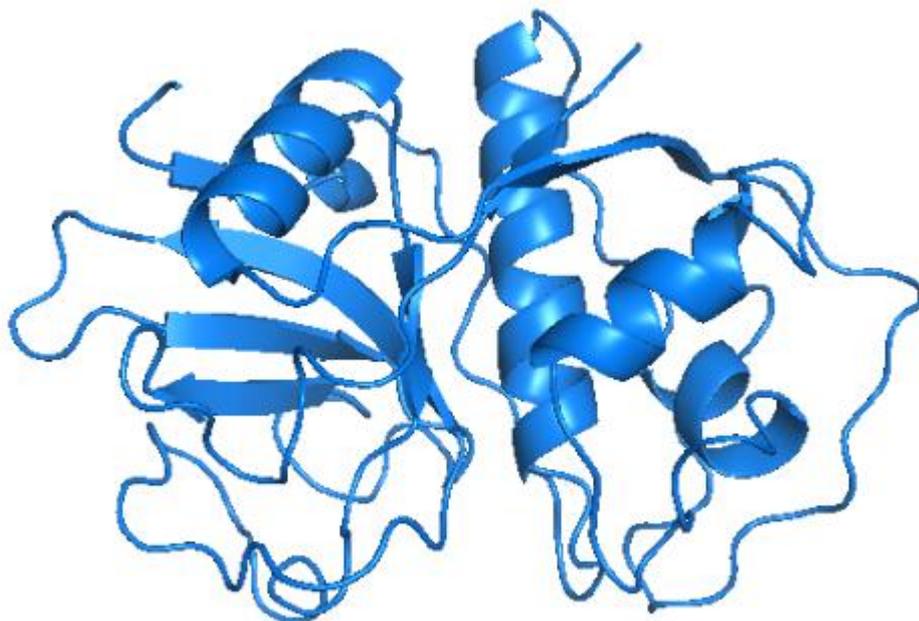
Legenda: As malhas azuis e amarelas representam as regiões de alta densidade eletrônica e os bastões a estrutura atômica do resíduo.

Fonte: <http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/resolution>.

2. Presença de um ligante co-cristalizado no sítio ativo: Isso facilita a identificação das coordenadas do sítio ativo nos *softwares* e permite ainda a validação dos dados, podendo ser o ligante usado como controle da docagem.

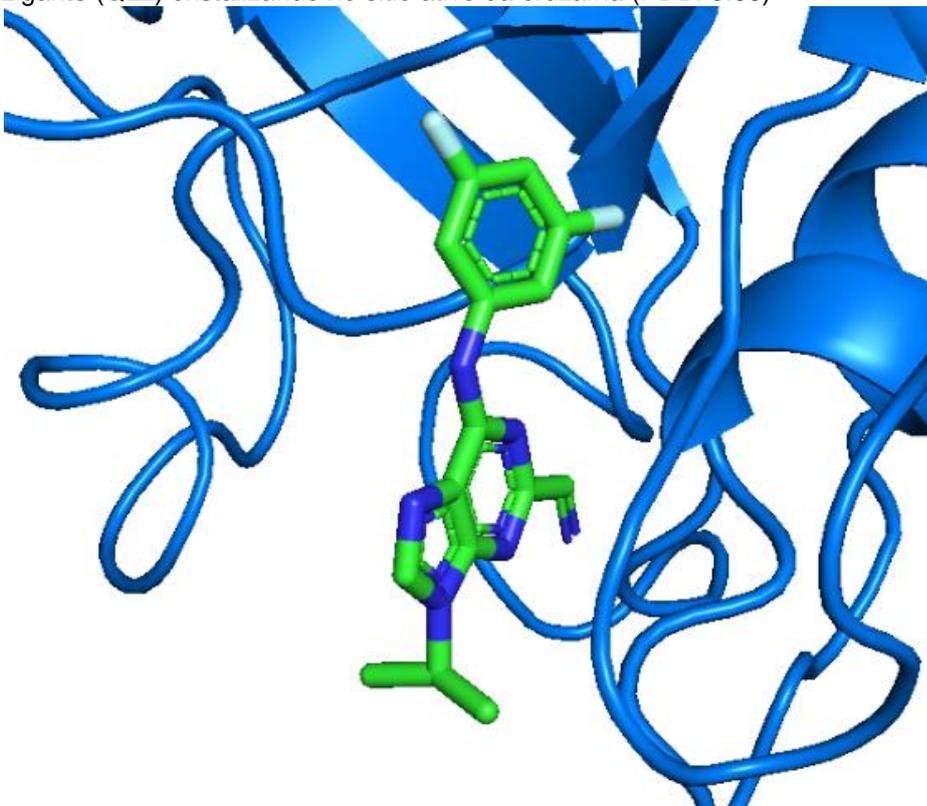
A estrutura registrada sob o código PDB **3I06** (Figura 16) foi selecionada, apresentando resolução de 1.1 Å, melhor entre as estruturas encontradas no PDB, e com um derivado purina nitrila (Figura 17) cristalizado no sítio ativo da cruzaina, que será identificado nesse trabalho sob código original do arquivo PDB, **QL2**.

Figura 16 - Representação tridimensional da cruzaina (PDB: 3I06)



Fonte: RIBEIRO (2018).

Figura 17 - Ligante (QL2) cristalizando no sitio ativo da cruzaina (PDB: 3I06)



Fonte: RIBEIRO (2018).

O *software* PyMOL versão 1.7.4.5 para Linux, foi utilizado para realizar o primeiro tratamento do arquivo da macromolécula, removendo as moléculas de água, cofatores e átomos não ligados, além de também isolar o ligante cristalizado em um arquivo separado. Ambos os arquivos, da macromolécula e do ligante, foram salvos na extensão .pdb.

### 3.4 TRIAGEM VIRTUAL BASEADA NA ESTRUTURA DO RECEPTOR (DOCAGEM MOLECULAR)

A triagem virtual baseada na estrutura do receptor, foi tratado aqui como a execução de várias docagens moleculares sequenciadas, sendo cada docagem relativa a um ligante e a enzima alvo.

Nesse tópico serão descritos a metodologia e parametrização usada para a execução de cada docagem. Em seguida será explicado como todo o processo foi automatizado para a execução da triagem virtual.

#### 3.4.1 Parametrização da Docagem Molecular

Um conjunto de *softwares* foi usado na execução da triagem virtual, sendo eles: AutoDock e AutoGrid versão 4.2.6 e o MGLTools versão 1.5.6 para Linux.

Usando o *software* AutoDockTools, foi realizado outro tratamento da macromolécula, adicionando os hidrogênios à estrutura da cruzaina e calculado as cargas parciais Gasteiger. Nesse processo, o software funde os hidrogênios não polares com os átomos de carbono, uma vez que os considera como hidrogênios ligados ao carbono (THE SCRIPPS RESEARCH INSTITUTE, 2017a) e por fim o arquivo foi salvo no formato .pdbqt. Este é um formato de arquivo padrão do AutoDock 4. Sua diferença principal para o formato .pdb é a adição de informações quanto as cargas parciais e os tipos de átomos do AutoDock, que para alguns deles é feito uma categorização especial: OA, NA e SA, respectivamente, para oxigênio, nitrogênio e enxofre aceptores de ligações de hidrogênio; HD para hidrogênio doador de ligações de hidrogênio; A para carbono de ciclo planares (THE SCRIPPS RESEARCH INSTITUTE, 2017b).

Ainda no AutoDockTools, foram configurados a área de busca, ou *grid box*, e os parâmetros para a execução do AutoGrid. Logo, foram definidos os tipos de átomos

presentes nos ligantes (A, C, Cl, I, NA, F, OA, N, Br, SA, HD), para que sejam calculados os *grid maps* de cada um desses tipos de átomos. Os *grid maps* são arquivos que contêm as informações relativas as energias de interação dos vários átomos presentes no ligante, com a macromolécula. Atualmente o *software* inclui parâmetros para os átomos de H, C, N, O, F, Mg, P, S, Cl, Ca, Mn, Fe, Zn, Br, que são os mais frequentemente observados nos sistemas biológicos.

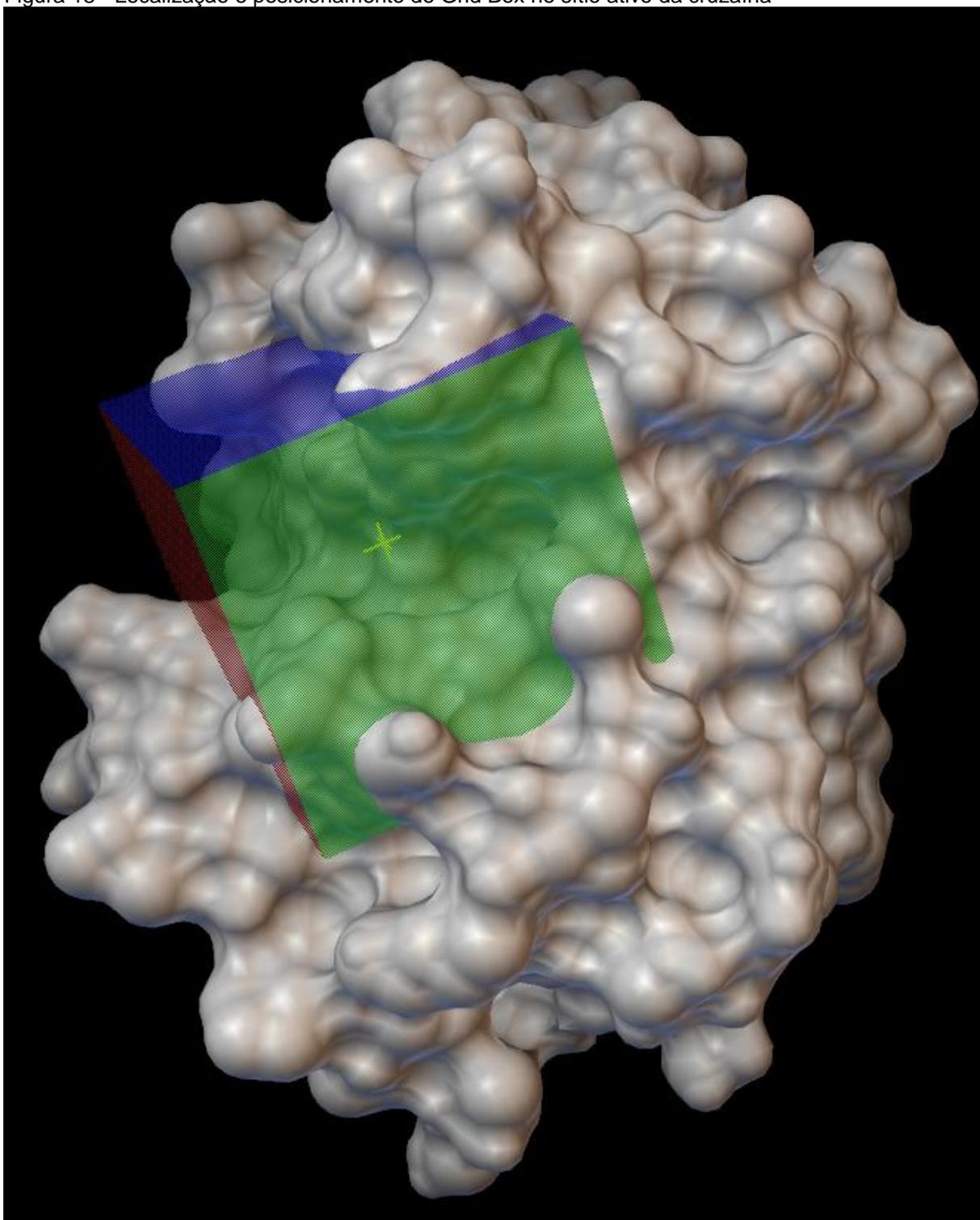
As coordenadas do *grid box*, foram determinadas tendo como base o ligante co-cristalizado com a macromolécula. O *grid box* foi centralizado nas coordenadas do ligante e o tamanho ajustado de modo a comportar as outras moléculas. Dados detalhados e a visualização *grid box* podem ser observados na Tabela 3 e Figura 18.

Tabela 3 - Parâmetros de configuração do *grid box*

<b>Parâmetros</b>	<b>Valores</b>
Nº de pontos em X	45
Nº de pontos em Y	45
Nº de pontos em Z	45
Espaçamento	0.375
Centro do <i>grid box</i> em X	-5.628
Centro do <i>grid box</i> em Y	-34.07
Centro do <i>grid box</i> em Z	2.781

Fonte: RIBEIRO (2018).

Figura 18 - Localização e posicionamento do Grid Box no sítio ativo da cruzaina



Fonte: RIBEIRO (2018).

Antes de executar o AutoDock, foi preciso gerar o arquivo de parâmetros da docagem (\*.dpf). Foi utilizado o Algoritmo Genético Lamarckiano, mantendo a maioria dos parâmetros com os valores padrões, modificando apenas os valores dos

seguintes parâmetros (Tabela 4). Com os arquivos .dpf de cada molécula gerado, o AutoDock foi finalmente executado.

Tabela 4 - Parâmetros modificados no arquivo de parâmetros da docagem

Parâmetro	Valor
Nº de indivíduos na população	150
Nº máximo de avaliações de energia	1750000
Nº máximo de gerações	27000
Estado do ligando não ligado	unbound_model bound

Fonte: RIBEIRO (2018).

### 3.4.2 Script de execução da triagem virtual

O Autodock permite a utilização de scripts para automatizar operações o que inclui a execução das triagens virtuais. Nesse trabalho, foram elaborados *scripts* baseados nos *scripts* genéricos que o  *Scripps Research Institute* fornece no tutorial “*Using AutoDock 4 for Virtual Screening*” sendo modificados de modo a tornar a execução da triagem virtual mais interativa, fácil e eficiente.

É importante destacar que não se trata de um novo *software*, mas sim, uma sequência de códigos que para automatizar a execução de várias docagens no AutoDock 4.

Os *scripts* atualizados podem ser obtidos gratuitamente no GitHub através do endereço:

[https://github.com/FredericoFavaro/AutoDock\\_VirtualScreening\\_Script](https://github.com/FredericoFavaro/AutoDock_VirtualScreening_Script)

### 3.4.3 Visualização dos resultados

O *script* gera uma lista com os *scores* de menor energia de cada composto em ordem crescente, permitindo observar rapidamente, os compostos com melhor afinidade.

O AutoDockTools foi utilizado para gerar um arquivo .pdb da pose de menor energia para cada composto, sendo eles exportados para o Pymol junto com a estrutura da cruzaina, possibilitando observar no espaço tridimensional como ficaram posicionados os compostos no sítio ativo.

Para observar os resíduos de aminoácidos e as interações que eles formaram com os compostos, foi utilizado o *software* LigPlot+ versão 1.4 para Linux. Ele gera automaticamente diagramas de interação (Ligações de hidrogênio e interações hidrofóbicas) do complexo ligante-receptor em 2D e ainda permite a visualização 3D através do Pymol.

### 3.5 TEÓRICO X EXPERIMENTAL

Para comparar os valores dos *scores* com os dados dos testes biológicos frente a cruzaína, foi realizado uma análise de correlação.

Antes de determinar qual método seria usado, foi realizado o teste de normalidade de Lilliefors, para verificar se os dados de IC<sub>50</sub> e os *scores* da docagem seguiam uma distribuição que se aproximava da normal.

O teste avalia o grau de concordância entre a distribuição de um conjunto de valores amostrais e a estatística do teste baseado no nível de significância determinado (SIEGEL, 2006). Dessa forma as hipóteses testadas são:

$H_0$ : *Dados seguem uma distribuição normal*

$H_1$ : *Dados não seguem uma distribuição normal*

Determinando um nível de significância em 5% (0.05), rejeita-se a hipótese nula de que os dados seguem uma distribuição que se aproxima da normal, caso o p-valor seja menor que 0.05. Os *scores* obtiveram um p-valor de 0.9685 e o IC<sub>50</sub> da atividade frente a cruzaína, o p-valor foi de 0.04084.

Como apenas os *scores* seguiam uma distribuição próxima da normal, foi escolhido o coeficiente de correlação de Spearman. Trata-se de um método não-paramétrico análogo ao método de correlação de Pearson, para medir o grau de associação linear entre duas variáveis que devem ser ao menos de escala de mensuração ordinal, de forma que os elementos das variáveis possam ser ordenados (SIEGEL, 2006). A diferença entre os dois métodos consiste no fato de que o coeficiente de correlação de Spearman, ao invés de usar os valores das variáveis em análise, faz uso de seus postos, por esse motivo existe a exigência quanto a escala de mensuração das variáveis. O coeficiente de correlação de Spearman ( $\rho_s$ ) pode assumir, valores entre -1 e 1, sendo que quanto mais próximo esse valor for de -1 e

1, maior será a associação entre as variáveis, enquanto que quanto mais próximo de zero, menor será a associação.

Um teste de hipóteses é realizado para verificar se o coeficiente de correlação obtido é significativo. Para isso as hipóteses testadas são:

$H_0$ : Não existe associação entre  $a$  e  $b$

$H_1$ : Existe associação entre  $a$  e  $b$

onde  $a$  e  $b$  são as duas variáveis avaliadas.

Rejeita-se a hipótese nula, caso o p-valor obtido seja menor que o nível de significância fixado (5%). Essas análises foram realizadas no *software* R versão 3.0.1, fazendo uso do pacote ISWR, para a correlação de Spearman.

### 3.6 QUIMIOMETRIA E QSAR

Antes da execução das análises, as estruturas dos compostos no formato .mol foram convertidas para o formato .sdf, que é compatível com os softwares usados.

A CPCA foi utilizada como análise exploratória e executada no *software* VolSurf+ versão 1.0.7 para Windows. As estruturas dos compostos em formato .sdf foram importadas para o VolSurf+ e os descritores foram gerados. O VolSurf+ calcula descritores químicos através dos mapas tridimensionais de energia de interação gerados pelo GRID (Campos de interação molecular baseado no GRID), comprimindo essa informação em descritores numéricos quantitativos. Ao todo, são 128 descritores: 28 relativos ao *probe* água (OH2), 24 ao *probe* hidrofóbico (DRY), 6 do *probe* acceptor de ligações de hidrogênio (O), 6 do *probe* doador de ligações de hidrogênio (N1) e 40 descritores não-baseados nos campos de interação molecular (MOLECULAR DISCOVERY LTD, 2009). Em seguida, o modelo CPCA foi gerado usando o auto escalonamento como pré-processamento.

A PCA e PLS, foram realizados no *software* Pentacle versão 1.0.7 para Windows. Inicialmente, foram importados para o Pentacle, os compostos e um arquivo formato .txt com a lista de valores da atividade biológica. Usando os parâmetros padrões, os descritores foram calculados. Em ambas as análises foram gerados um modelo para cada atividade biológica. Apesar da PCA não ser um método de QSAR, o Pentacle permite incluir os valores de atividade biológica ao modelo, desta forma,

pode-se identificar por meio de gráficos de dispersão, se o modelo foi capaz de separar os compostos ativos dos compostos inativos com base nas suas similaridades e diferenças. Os modelos de PCA foram gerados considerando 5 PCs e para os modelos de PLS, foram considerados 5 LVs e a validação cruzada foi realizada pelo método *Leave-One-Out*.

O número de compostos variou em cada modelo, uma vez que nem todos os compostos possuíam dados para as 4 atividades biológicas: Os modelos com atividade para cruzaína incluíram 76 compostos; para forma amastigota 85 compostos, para forma tripomastigota 108 compostos; e para forma epimastigota 80 compostos.

### 3.7 SUPERFÍCIES ELETRÔNICAS

As superfícies eletrônicas foram calculadas através do *software* Spartan versão 1.1.4 para Windows. Para os cálculos, foi preciso importar as moléculas no Spartan e realizar novamente a otimização do composto, no caso, foi utilizado o método semi-empírico, mantendo as configurações padrões. Com as moléculas otimizadas, foram calculadas as superfícies de HOMO, LUMO e o Mapa de potencial eletrostático.

Os resultados foram avaliados comparando visualmente as superfícies entre os compostos.

### 3.8 PREDIÇÃO DE METABÓLITOS

O citocromo p450 é uma família de enzimas capaz de mediar a biotransformação de diversas substâncias orgânicas, catalisando uma grande variedade de reações oxidativas e redutivas, com a intenção de melhorar a solubilidade das substâncias (MOLECULAR DISCOVERY LTD, 2014; OGU; MAXA, 2000). Logo, prever a formação de metabólitos é de fundamental importância para o planejamento de novos candidatos a fármacos, visto que a estrutura planejada pode sofrer alterações através da ação dessas enzimas, modificando sua ação ou sua toxicidade.

A predição foi realizada no programa MetaSite versão 5.1.1 para Windows, que é capaz de prever questões relacionadas ao metabolismo mediados pelo citocromo p450. Todo o processo é realizado automaticamente pelo programa, bastando

importar os compostos em sua estrutura 3D. O software basicamente compara os padrões de interação entre a proteína e o ligante, identificando os possíveis metabólitos gerados após a interação com a enzima (MOLECULAR DISCOVERY LTD, 2014). O MetaSite possui representações pré-definidas dos citocromos 1A1, 1A2, 2B6, 2C9, 2C19, 2D6, 2E1, 3A4, 3A5, que podem ser selecionados pelo usuário.

Para essa pesquisa, os compostos foram importados mantendo os parâmetros padrões do *software*. Em seguida, foi executada a ferramenta de identificação de metabólitos, considerando todas as 119 moléculas e selecionando os citocromos do fígado, 2C9, 2D6, 3A4, para obter a predição dos metabólitos exclusivamente por meio da metabolização dos compostos nesse órgão.

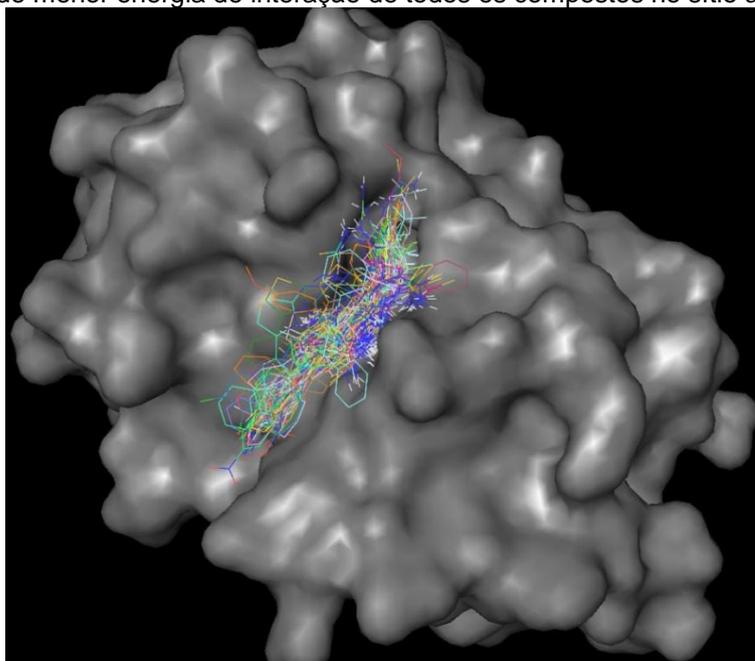
## 4 RESULTADOS E DISCUSSÃO

### 4.1 TRIAGEM VIRTUAL BASEADO NA ESTRUTURA DO RECEPTOR (DOCAGEM MOLECULAR)

Todas as docagens, incluindo o controle (ligante co-cristalizado com a enzima do arquivo PDB) foram bem-sucedidos. A única exceção foi o composto JW-9i que, segundo o log do AutoDock, possui átomos cujo os parâmetros de campo de força não foram implementados no AutoDock, o que não confere, visto que a molécula é composta apenas por átomos suportados. Na tentativa de solucionar o problema, o arquivo da molécula foi vistoriado, mas não foi encontrado a causa. Outra solução testada, consistiu em criar um arquivo para o composto em outro *software* (Spartan '14 para Windows). Ainda assim, o problema se perpetuou. Por se tratar de uma única molécula pertencente a série de compostos mais numerosa, o composto foi excluído do experimento. Logo, a triagem virtual foi realizada considerando 118 moléculas e o composto co-cristalizado.

Na Figura 21 é exposto a pose e conformação de menor energia de interação de cada um dos ligantes sobrepostos no sítio ativo da cruzaina, demonstrando como os compostos ficaram acomodados no sítio ativo.

Figura 19 - Poses de menor energia de interação de todos os compostos no sítio ativo da cruzaina



Fonte: RIBEIRO (2018).

Os scores energéticos do complexo ligante-enzima de cada composto estão organizados no Tabela 5 em ordem decrescente, do composto com maior afinidade para o composto de menor afinidade. De acordo com os resultados, os scores variaram de -8.68 kcal/mol a -4.29 kcal/mol. O controle obteve um score de -5.83 kcal/mol, ficando em 92<sup>o</sup> lugar comparado as outras moléculas.

Pode-se ainda observar que os melhores ligantes pertencem a classe dos tiazóis e tiazolidinonas enquanto que as moléculas com menor afinidade com o sítio ativo, são quase que exclusivamente tiosemicarbazonas.

Tabela 5 - Scores de energia em ordem decrescente de afinidade, obtidos pela docagem molecular entre os ligantes e a cruzaina

Ligante	Score	Ligante	Score	Ligante	Score	Ligante	Score
PT01.16	-8.68	AR-2c	-7.04	MB1.7	-6.46	TP-11	-5.83
GT-09	-8.27	PA-6k	-6.97	TZ-11	-6.46	QL2	-5.83
PT01.13	-8.27	AR-2k	-6.94	LA-1a1	-6.45	JW-9e	-5.79
GT-18	-8.26	PT02.1	-6.94	PT01.7	-6.44	LA-INT-3	-5.71
TZBP-05	-8.10	PM-Int3B	-6.93	PA-6a	-6.42	MB1.9	-5.70
PT01.19	-8.04	PA-6b	-6.91	TP-02	-6.41	AD-9v	-5.65
GT-04D	-8.03	GT-04A	-6.85	TAP-02	-6.39	JW-8c	-5.53
GT-17	-8.00	JW-9g	-6.82	JW-9p	-6.35	TS-05	-5.52
AR-2l	-7.99	LA-1a2	-6.81	LA-INT-1	-6.35	TS-10	-5.48
PT01.3	-7.94	MB1.8	-6.79	GT-13	-6.33	TS-06	-5.43
GT-11	-7.93	LA-1b2	-6.78	TAP-04	-6.33	LA-INT-2	-5.41
TZBP-04	-7.83	AR-2d	-6.77	JW-8f	-6.31	MB1.12	-5.38
PT01.18	-7.82	GT-05D	-6.77	MB1.6	-6.31	AD-9b	-5.34
JW-9s	-7.66	TZ-05	-6.76	JW-9a	-6.28	MB1.3	-5.34
PT01.9	-7.65	TP-05	-6.74	AD-9q	-6.27	MB2.2	-5.34
PT01.2	-7.64	JW-9o	-6.68	JW-8a	-6.22	TS-08	-5.34
PT01.5	-7.60	TZ-12	-6.67	TAP-10	-6.19	AD-9k	-5.30
TZ-04	-7.55	LA-1b1	-6.66	GT-06C	-6.16	AD-9i	-5.24
JW-9r	-7.49	TP-09	-6.65	JW-8d	-6.16	AD-9 <sup>a</sup>	-5.14
JW-8h	-7.47	TAP-06	-6.63	GT-06A	-6.14	TS-01	-5.14

Continua

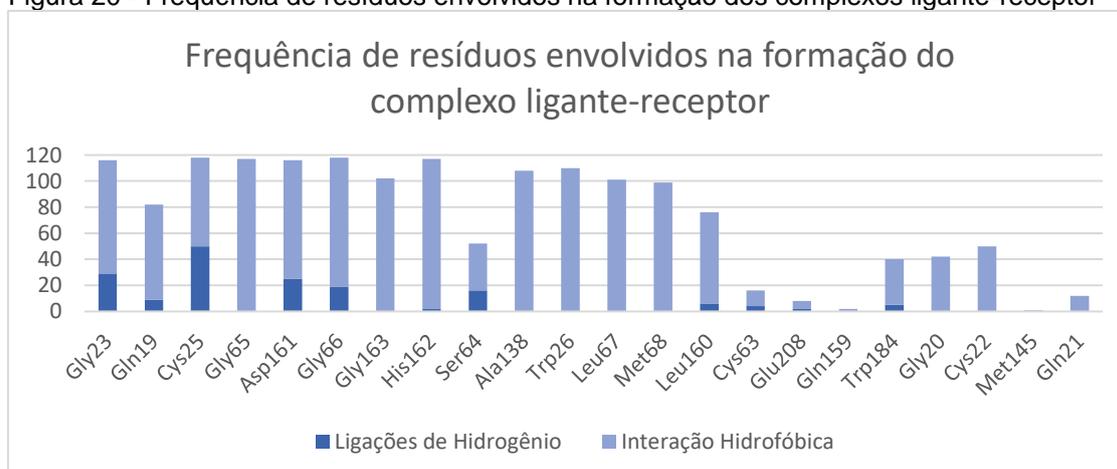
Continuação

Ligante	Score	Ligante	Score	Ligante	Score	Ligante	Score
LA-1a5	-7.45	MB3.5	-6.62	GT-03	-6.11	AD-9c	-5.03
PT01.6	-7.38	LA-1b3	-6.58	JW-9h	-6.10	TS-02	-5.03
AR-2m	-7.30	TAP-08	-6.58	LA-1c4	-6.10	MB1.2	-4.97
PT01.8	-7.29	TZ-06	-6.57	AD-9t	-5.98	TS-12	-4.93
PT01.14	-7.28	AR-2h	-6.56	TAP-11	-5.98	AD-9d	-4.90
PA-6d	-7.23	AD-9r	-6.55	TP-04	-5.98	AP	-4.60
TZ-07	-7.20	GT-05B	-6.55	PT03	-5.96	MB2.1	-4.39
PT01.4	-7.14	GT-06	-6.52	TP-06	-5.92	P	-4.38
PA-6h	-7.08	JW-9m	-6.51	TP-10	-5.91	MB02	-4.29
TZBP-06	-7.05	TAP-07	-6.47	LA-1c1	-5.88		

Fonte: RIBEIRO (2018).

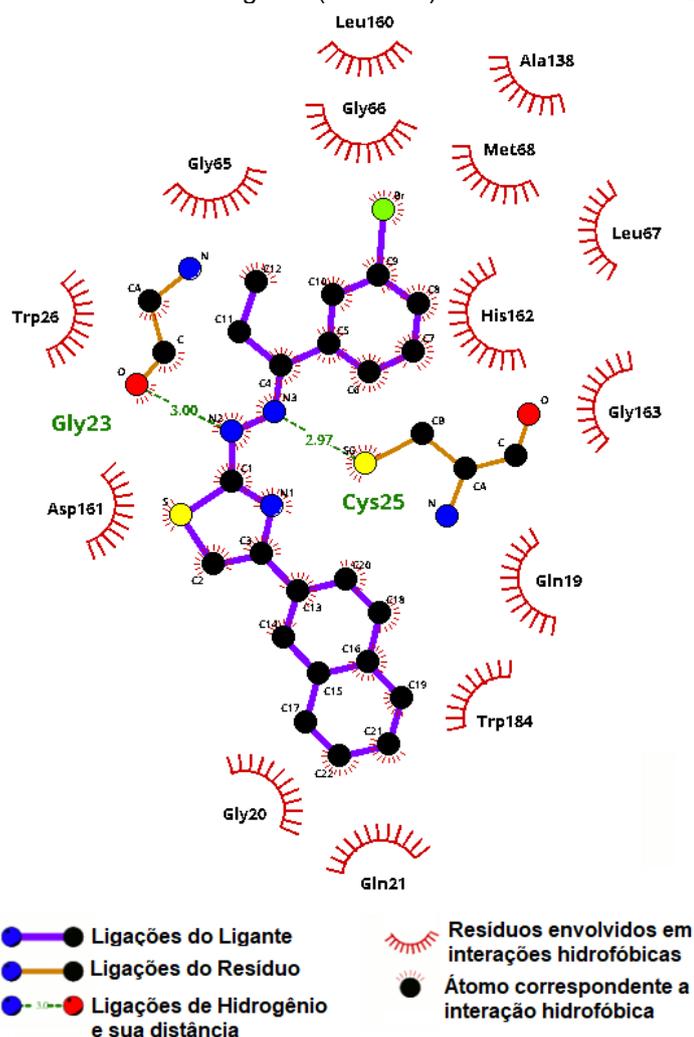
Com relação aos resíduos, 22 aminoácidos estão envolvidos na formação dos complexos ligante-cruzaina. Os resíduos que ocorreram com maior frequência foram Gly23, Cys25, Gly65, Asp161, Gly66 e His162, correspondendo de 97,5% (116) a 99,2% (118) dos ligantes, enquanto que os aminoácidos com menor frequência foram Glu208, Gln159, Met145 (Figura 22). Estudos apontam que os resíduos Cys25, His159 e Asn175 formam uma tríade catalítica que seria responsável pela atividade inibitória da cruzaina (MCGRATH et al., 1995; ROBINSON; DALTON, 2011; WIGGERS et al., 2013). Dos compostos aqui avaliados, 118 interagiram com a Cys25, como pode ser observado na figura 23.

Figura 20 - Frequência de resíduos envolvidos na formação dos complexos ligante-receptor



Fonte: RIBEIRO (2018).

Figura 23 - Interações formadas entre o ligante (PT01-16) e o sítio ativo da cruzaina



Fonte: RIBEIRO (2018).

Outra informação importante, se refere à frequência e os tipos de interações. Ligações de hidrogênio e interações hidrofóbicas foram identificadas, contudo essa última ocorreu em maior frequência, o que acaba por justificar a maior afinidade dos derivados tiazóis e tiazolidinas, quando comparados com suas respectivas tiosemicarbazonas. Os mesmos resíduos que ocorreram em maior frequência, foram os que mais interagiram com o sítio ativo através de ligações de hidrogênio, com exceção da Gly65 e His162. Em contrapartida, a Gln19 e Ser64, apesar não serem frequentes, tiveram uma frequência elevada de ligações de hidrogênio (Figura 22).

Ao realizar a análise de correlação de Spearman entre os scores e os valores de  $IC_{50}$  frente a cruzaina, o p-valor obtido foi de 0.1524, não rejeitando a hipótese nula de que não há correlação entre as duas variáveis, no caso, entre os valores de  $IC_{50}$  da atividade biológica frente a cruzaina e os scores obtidos na triagem virtual. Resultado que é reafirmado pelo coeficiente de correlação ( $r^2$ ) de -0.1668943, valor

muito próximo de zero, indicando fraca correlação entre as duas variáveis. Isso significa que os scores não foram capazes de predizer a atividade dos compostos, não havendo correlação entre os resultados da docagem e os dados de inibição da enzima nos estudos *in vitro*, conforme pode facilmente ser observado na tabela a seguir, onde são apresentados os 10 compostos mais ativos e menos ativos considerando as duas abordagens (*in vitro* e *in silico*), onde se evidencia a ausência de correlação.

Tabela 6 - Comparação dos 10 compostos mais ativos e menos ativos considerando as duas abordagens (*in vitro* e *in silico*)

Mais ativos				Menos ativos			
<i>in vitro</i>		<i>in silico</i>		<i>in vitro</i>		<i>in silico</i>	
Composto	IC <sub>50</sub>	Composto	Score	Composto	IC <sub>50</sub>	Composto	Score
MB3.5	1.35	PT01.16	-8.68	TP-10	85	LA-INT-3	-5.71
TZ-04	7	PT01.13	-8.27	PT01.18	86.3	JW-8c	-5.53
TP-05	10	PT01.19	-8.04	TS-08	87.6	TS-05	-5.52
TAP-10	10	AR-2I	-7.99	TAP-04	89	TS-10	-5.48
TAP-11	11	PT01.3	-7.94	TAP-02	90	TS-06	-5.43
PT03	11.7	PT01.18	-7.82	AR-2I	93.3	LA-INT-2	-5.41
JW-8f	11.8	PT01.9	-7.65	TAP-06	95	TS-08	-5.34
TP-11	12	PT01.2	-7.64	PT01.14	95.2	TS-01	-5.14
TS-05	13.9	PT01.5	-7.6	PT01.6	97.3	TS-02	-5.03
TZ-07	17.6	TZ-04	-7.55	AR-2c	97.5	TS-12	-4.93

Fonte: RIBEIRO (2018).

Leite e colaboradores (2006) compararam os scores obtidos em derivados tiossemicarbazonas, aminoacil-tiossemicarbazidas e acil-thiazolidonas frente a enzima cruzaina com os seus respectivos valores de pIC<sub>50</sub> (-log do IC<sub>50</sub>) através de gráfico de dispersão, observando aparente correlação entre as variáveis, contudo nenhum parâmetro estatístico foi utilizado como validação. Usando a mesma abordagem, Filho et al. (2009) obteve resultados semelhantes ao comparar os scores da docagem molecular entre derivados 3-(4-aril-substituídos)-1,2,4-oxadiazol-N-acilhidrazonas e a cruzaina com seus respectivos pIC<sub>50</sub>. Contudo, um estudo mais recente, realizou a docagem molecular com anti-inflamatórios não esteroides com as ciclooxigenases 1 e 2 correlacionando os scores obtidos, com os dados de IC<sub>50</sub>, mas não foram observados relações entre as duas variáveis (MIKRA et al., 2017), semelhante ao resultado obtido nessa tese.

Apesar de ser possível encontrar correlação entre as duas variáveis, os scores obtidos através da docagem molecular não refletem diretamente na IC<sub>50</sub> de uma substância, eles mensuram a força da interação do ligante com o sítio ativo o que não necessariamente leva a inibição da enzima. Outros fatores estão relacionados com a atividade do composto, os quais não são levados em consideração na docagem molecular, como a própria concentração do composto.

## 4.2 QUIMIOMETRIA E QSAR

### 4.2.1 CPCA

A análise exploratória foi realizada considerando os 119 compostos, 128 variáveis independentes (descritores) agrupados em 13 blocos. De acordo com a tabela 7, as componentes principais explicaram 76.732% da variância total dos dados.

Tabela 7 - Resultados da análise de CPCA

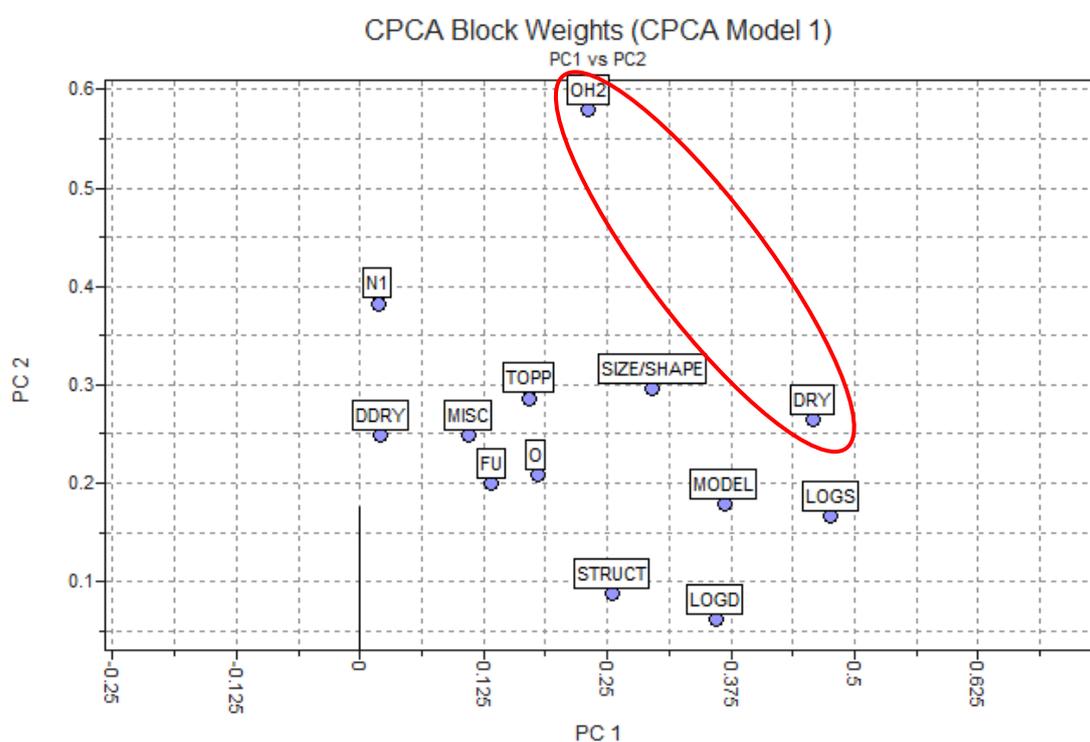
PC	% da variância explicada	% da variância explicada acumulada
1	34.317	34.317
2	20.300	54.617
3	9.196	63.812
4	8.063	71.960
5	4.674	76.732

Fonte: RIBEIRO (2018).

Ao plotar o gráfico de pesos (Figura 24), os blocos de descritores com maior importancia foram o DRY e OH2. O bloco DRY envolve descritores relativos as regiões hidrofóbicas como os **Volumes hidrofóbicos (descritores D1-D8)** que geram um campo lipofílico, determinando as regiões que atraem interações hidrofóbicas; **Fatores de capacidade do DRY (CD1 – CD8)** que representam a proporção do volume hidrofóbico sobre o total da superfície molecular; e **Momentos de energia de interação do DRY (ID1-ID4)** que representa o desequilíbrio entre o centro da massa de uma molécula e o baricentro da região hidrofóbica. O Bloco OH2 envolve descritores opostos ao bloco DRY, ou seja, **Volumes hidrofílicos (descritores W1 – W8)** calcula campos de força que atraem interações com moléculas de água, os

descritores com campos com -0.2 a -1.0 kcal/mol são relativos a força de polarização e dispersão, enquanto os descritores com campos moleculares de -2.0 a -6.0 kcal/mol são relativos a regiões polares e doador-receptor de ligações de hidrogênio fortes; **Fatores de capacidade do OH2 (CD1 – CD8)** que representam a proporção do volume hidrofílico sobre o total da superfície molecular; e **Momentos de energia de interação do DRY (IW1-IW4)** que representa o desequilíbrio entre o centro da massa de uma molécula e o baricentro da região hidrofílica (MOLECULAR DISCOVERY LTD, 2009). Logo, essas características foram as consideradas de maior importância para os compostos.

Figura 21 - Blocos de descritores observados no CPCA e seus respectivos pesos



Fonte: RIBEIRO (2018).

#### 4.2.2 PCA

A PCA foi realizada 4 vezes, uma para cada atividade biológica (atividade frente a cruzáina, frente a forma amastigota, frente a forma epimastigota e frente a forma tripomastigota de *T. cruzi*). O intuito foi verificar se o modelo do PCA conseguiria distinguir os compostos ativos dos inativos, uma vez que o *software* (Pentacle) permite incluir os valores das atividades biológicas aos compostos, possibilitando observar nos gráficos se houve alguma tendência relativa à atividade dos compostos.

Inicialmente a Análise de Componentes Principais levou em consideração apenas os descritores pertencentes aos blocos de maior peso de acordo com a CPCA, contudo os resultados não foram satisfatórios. Logo a análise foi realizada considerando todos os descritores disponíveis no software, como será descrito a seguir.

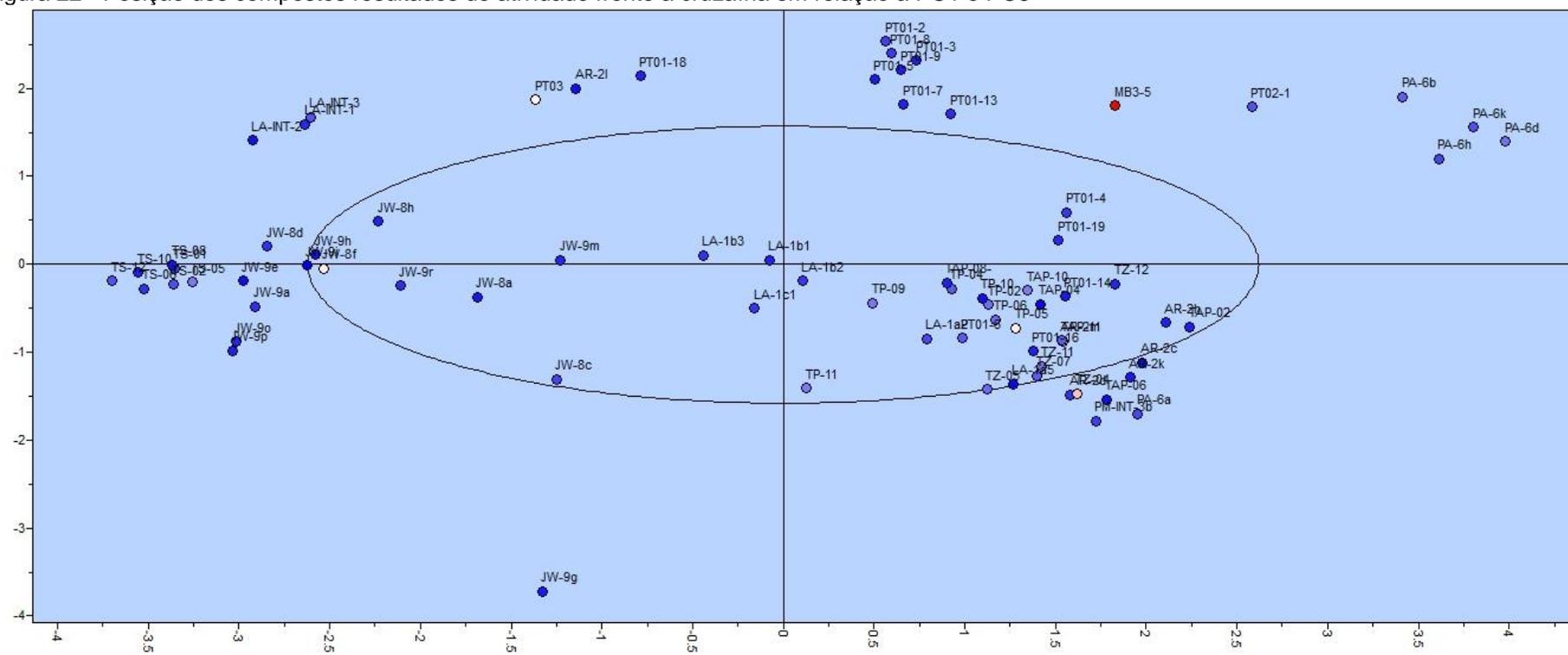
Avaliando a atividade frente a cruzaina, a PCA foi realizada considerando 76 moléculas e 427 descritores. O percentual de variância explicada acumulada pelo modelo, mesmo considerando os 5 PCs chegou ao máximo de 59.39% (Tabela 8). Um percentual baixo, incapaz de gerar um bom modelo, como pode ser visto na figura 25, que representa a posição relativa dos compostos no espaço dos componentes principais. No gráfico, cada composto é representado por um círculo e a cor do círculo representa a escala da atividade biológica, sendo o Vermelho o mais ativo e o azul escuro menos ativo. Nota-se que não houve separação entre os compostos mais ativos e menos ativos.

Tabela 8 - Variância explicada pela análise de PCA considerando os compostos com resultados de atividade frente a cruzaina

PC	% da variância explicada	% da variância explicada acumulada
1	29.31	29.31
2	10.26	39.57
3	8.65	48.22
4	6	54.22
5	5.16	59.39

Fonte: RIBEIRO (2018).

Figura 22 - Posição dos compostos resultados de atividade frente a cruzaina em relação a PC1 e PC5



Legenda: Círculos vermelhos representam os compostos mais ativos, e círculos azuis escuro representam os compostos menos ativos.

Fonte: RIBEIRO (2018).

Quando realizada a PCA levando em consideração a atividade frente a forma amastigota do *T. cruzi*, também foi obtido um percentual de variância baixo, 55.94% considerando os 5 PCs (Tabela 9). No gráfico da posição relativa dos compostos no espaço dos componentes principais, pode-se observar mais claramente a dispersão desordenada dos compostos ativos (em vermelho) dos menos ativos (em azul) (Figura 26), demonstrando que não foi possível separá-los com base nos PCs formados com os descritores.

Tabela 9 - Variância explicada pela análise de PCA considerando os compostos com resultados de atividade frente a forma amastigota de *T. cruzi*

PC	% da variância explicada	% da variância explicada acumulada
1	27.71	21.71
2	9.66	37.36
3	8.3	45.67
4	5.87	51.53
5	4.41	55.94

Fonte: RIBEIRO (2018).



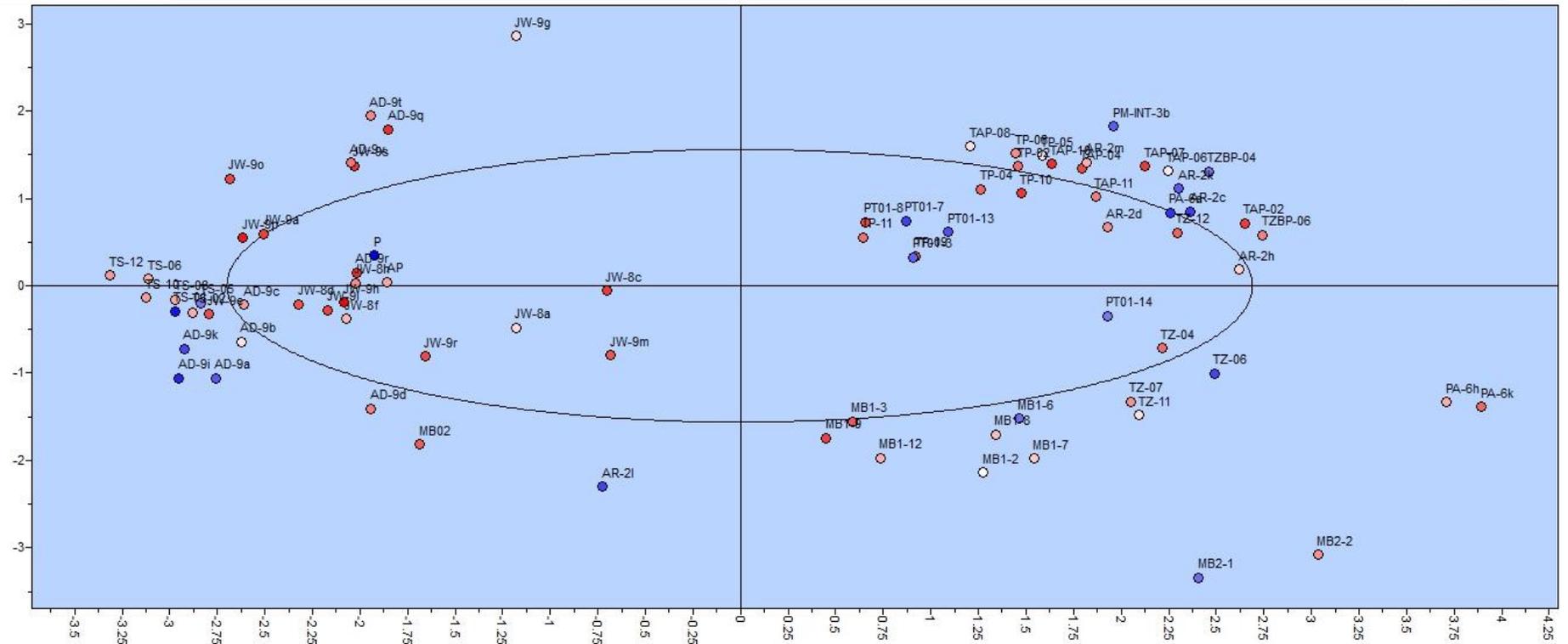
Levando em consideração a atividade frente a forma epimastigota, os resultados não foram muito diferentes das análises anteriores, apresentando um baixo percentual da variância total explicada, mesmo considerando os 5 PCs (57.18%) (Tabela 10) o que refletiu na incapacidade do modelo de separar os compostos ativos dos inativos, como pode ser verificado na figura 27.

Tabela 10 - Variância explicada pela análise de PCA considerando os compostos com resultados de atividade frente a forma epimastigota de *T. cruzi*

PC	% da variância explicada	% da variância explicada acumulada
1	30.86	30.86
2	10.05	40.93
3	6.87	47.80
4	5.37	53.17
5	4.01	57.18

Fonte: RIBEIRO (2018).

Figura 24 - Posição dos compostos resultados de atividade frente a forma epimastigota de *T. cruzi* em relação a PC1 e PC5



Legenda: Cada ponto representa um composto. Círculos vermelho representam os compostos mais ativos, e círculos azuis escuro representam os compostos menos ativos.

Fonte: RIBEIRO (2018).

Para a atividade frente a forma tripomastigota do *T. cruzi* os resultados não foram diferentes. O percentual de variância explicada foi de 57.11% incluindo os 5 PCs (Tabela 11). O que mais uma vez contribuiu em um modelo incapaz de separar os compostos ativos dos não ativos com base nos PCs obtidos. No gráfico de scores (Figura 28) observa-se claramente a dispersão dos ligantes no espaço dos PCs e a falta de distinção entre os compostos ativos (vermelho) dos compostos com baixa atividade (azul).

Tabela 11 - Variância explicada pela análise de PCA considerando os compostos com resultados de atividade frente a forma tripomastigota de *T. cruzi*

PC	% da variância explicada	% da variância explicada acumulada
1	28.23	28.23
2	10.90	39.13
3	8.07	47.21
4	5.56	52.77
5	4.43	57.11

Fonte: RIBEIRO (2018).



### 4.2.3 PLS

Assim como a PCA, foi realizado um PLS para cada conjunto de moléculas com resultados de atividade biológica visto que a intenção era gerar um modelo preditivo da respectiva atividade biológica.

Iniciando pela atividade biológica para cruzaina, o modelo final foi gerado incluindo 76 compostos e 176 descritores. O percentual da variância explicada foi baixo (57.14%). O  $r^2$ , índice representativo da qualidade do modelo, obteve um valor máximo de 0.75 na última variável latente, contudo, e o  $q^2$ , que representa habilidade preditiva do modelo obtido a partir do teste de validação cruzada, foi de 0.15, o mais baixo entre as VLs. Por se tratar de um modelo preditivo, o ideal é escolher a VL com maior valor de  $q^2$ , no caso a VL2 ( $q^2 = 0.36$ ) (Tabela 12). Apesar de ser o melhor resultado, ainda está muito distante de um modelo aceitável, que deve apresentar valores de  $r^2$  e  $q^2$  superiores a 0.8 e 0.5, respectivamente.

Tabela 12 - Resultado do modelo PLS gerado para atividade frente a cruzaina

VL	% da variância	% da variância	$r^2$	$q^2$
	explicada	explicada acumulada		
1	22.33	22.33	0.35	0.19
2	20.65	42.98	0.53	0.36
3	5.99	48.98	0.6	0.33
4	4.74	53.71	0.68	0.26
5	3.43	57.14	0.75	0.15

Fonte: RIBEIRO (2018).

O modelo gerado para a atividade frente a forma amastigota de *T. cruzi*, incluiu 86 moléculas e 219 descritores. Esse modelo foi o que apresentou os melhores resultados de  $r^2$  e  $q^2$ . Nota-se na Tabela 13 que a 5ª VL obteve um  $r^2$  de 0.81 e um  $q^2$  de 0.46 valores bem próximos do aceitável, contudo o modelo explicou um percentual muito baixo da variância dos dados (51.05%) indicando que os valores dos coeficientes de correlação e predição não são confiáveis.

Tabela 13 - Resultado do modelo PLS gerado para atividade frente a forma amastigota de *T. cruzi*

VL	% da variância		r <sup>2</sup>	q <sup>2</sup>
	explicada	explicada acumulada		
1	20.19	20.19	0.38	0.26
2	11.12	31.31	0.65	0.49
3	5.21	36.51	0.73	0.5
4	5.5	42.02	0.79	0.45
5	9.03	51.05	0.81	0.46

Fonte: RIBEIRO (2018).

O modelo gerado para a atividade frente a forma epimastigota de *T. cruzi*, incluiu 80 moléculas e 190 descritores. Como pode ser observado na tabela 14, o modelo explicou 57.04% da variância dos dados na VL 5, e apesar de apresentar um r<sup>2</sup> de 0.75 (< 0.8) o fator de predição foi muito baixo (0.15) o que aliado ao baixo percentual da variância explicada (57.04%) o torna um modelo inadequado para realizar uma boa predição da atividade biológica.

Tabela 14 - Resultado do modelo PLS gerado para atividade frente a forma epimastigota de *T. cruzi*

VL	% da variância		r <sup>2</sup>	q <sup>2</sup>
	explicada	explicada acumulada		
1	17.71	17.71	0.4	0.19
2	23.82	41.53	0.57	0.37
3	4.41	45.94	0.65	0.3
4	4.06	50	0.72	0.14
5	7.04	57.04	0.75	0.15

Fonte: RIBEIRO (2018).

O modelo gerado para atividade frente a forma tripomastigota de *T. cruzi*, foi composto por 108 moléculas e 141 variáveis e obteve os piores resultados. Percentual de variância total explicada de 59.32% e r<sup>2</sup> de 0.53, incluindo valores negativos no fator de predição nas VLs 4 e 5, indicando um fraco ajuste do modelo.

Tabela 15 - Resultado do modelo PLS gerado para atividade frente a forma tripomastigota de *T. cruzi*

VL	% da variância explicada	% da variância explicada acumulada	r <sup>2</sup>	q <sup>2</sup>
1	25.82	25.82	0.2	0.12
2	21.13	46.94	0.32	0.20
3	5.23	52.17	0.38	0.14
4	1.97	54.15	0.49	-0.10
5	5.17	59.32	0.53	-0.13

Fonte: RIBEIRO (2018).

Infelizmente, os resultados obtidos tanto com a PCA como PLS não foram suficientes para gerar um modelo preditivo para a atividade biológica. Algumas suposições podem ser feitas no intuito de justificar os resultados obtidos na PCA e nos modelos de PLS: Variação dos valores de pIC<sub>50</sub>: Os compostos apresentaram uma variação muito pequena nos valores de pIC<sub>50</sub>, afetando diretamente o modelo. Analisando a variância dos dados (Quadro 16), fica evidente a pequena faixa de variação dos dados. Uma observação que comprova essa afirmação, é que o grupo com maior variância dos valores de atividade (atividade frente às formas amastigotas de *T. cruzi*) também foi o que gerou os melhores modelos.

Tabela 16 - Variância dos valores de pIC<sub>50</sub> para cada atividade

Atividade	Variância
Cruzaína	0.14
Amastigota	0.48
Tripomastigota	0.35
Epimastigota	0.25

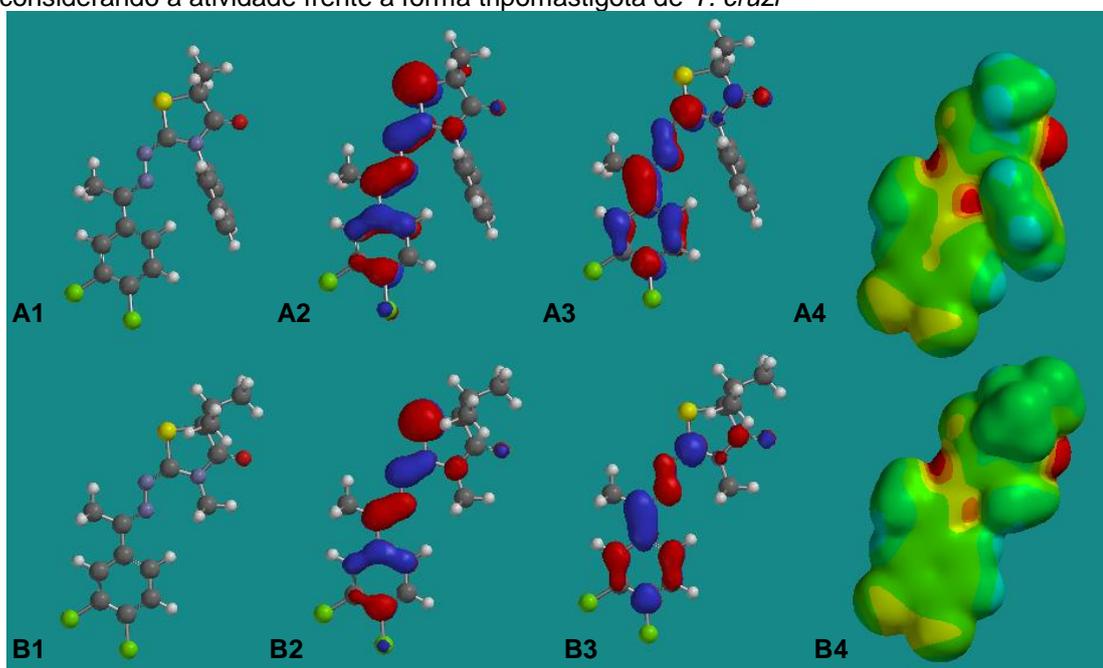
Fonte: RIBEIRO (2018).

Quantidade de compostos: Como mencionado na metodologia, nem todas as moléculas tinham resultados quanto a atividade biológica, ficando no caso da cruzaína, por exemplo, com 78 moléculas. Possivelmente, um banco de dados maior e com uma distribuição melhor dos valores de atividade biológica permitirá a obtenção de modelos com resultados mais preditivos.

### 4.3 SUPERFÍCIES ELETRÔNICAS

Visando complementar as análises quimiométricas e de QSAR apresentadas no tópico anterior, foram calculadas as superfícies eletrônicas dos compostos, na tentativa de identificar padrões e diferenças que podem estar associados aos valores de atividade. Por questões didáticas, serão apresentadas as superfícies eletrônicas dos compostos mais ativos e menos ativos de algumas séries frente a cada atividade biológica (Figura 29 - 33).

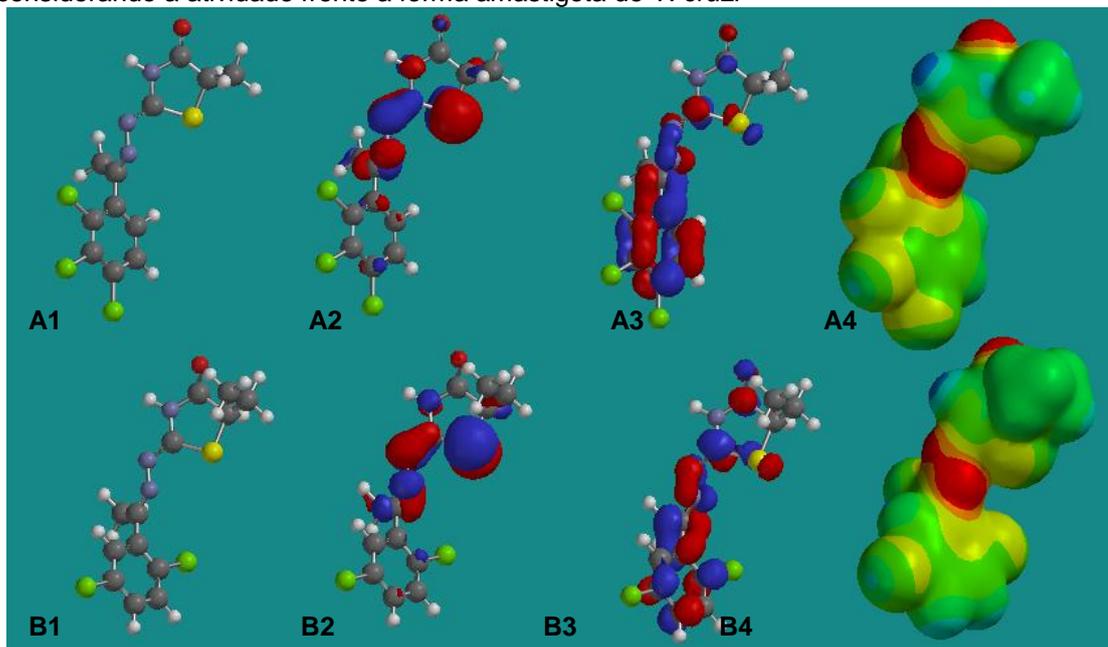
Figura 26 - Superfícies eletrônicas do composto mais ativo (GT-5b) e menos ativo (GT-06c) da série GT considerando a atividade frente a forma tripomastigota de *T. cruzi*



Legenda: A: Mais ativo; B: Menos ativo; 1: Estrutura de referência; 2: Homo; 3: Lumo; 4: Mapa de potencial eletrostático.

Fonte: RIBEIRO (2018)

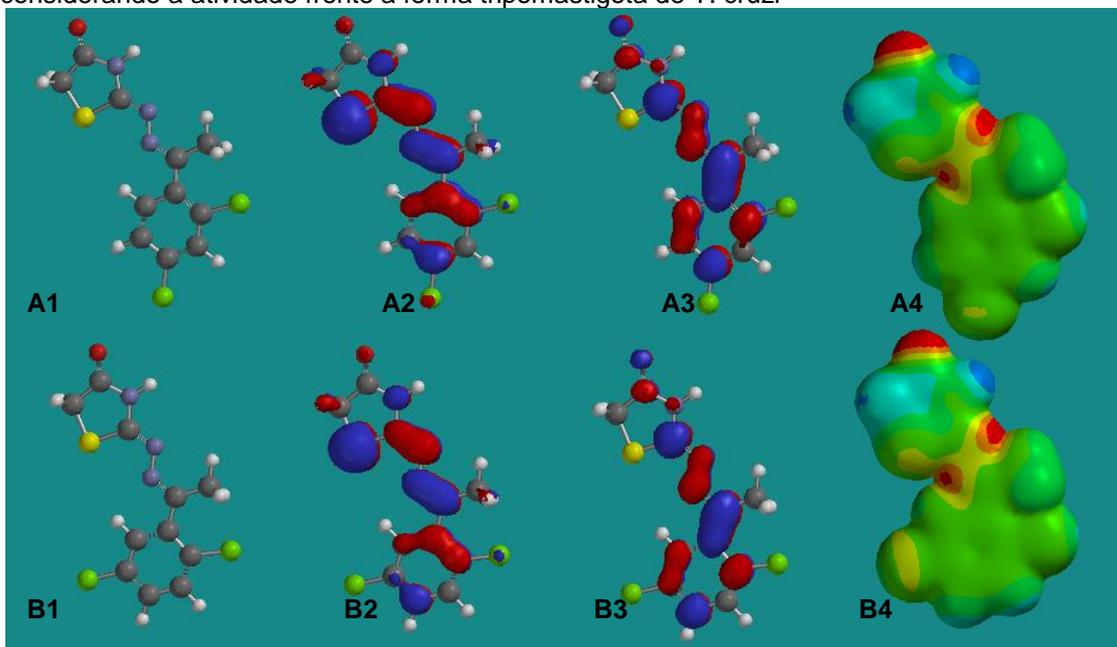
Figura 27 - Superfícies eletrônicas do composto mais ativo (LA-1a2) e menos ativo (LA-1b3) da série LA considerando a atividade frente a forma amastigota de *T. cruzi*



Legenda: A: Mais ativo; B: Menos ativo; 1: Estrutura de referência; 2: Homo; 3: Lumo; 4: Mapa de potencial eletrostático.

Fonte: RIBEIRO (2018).

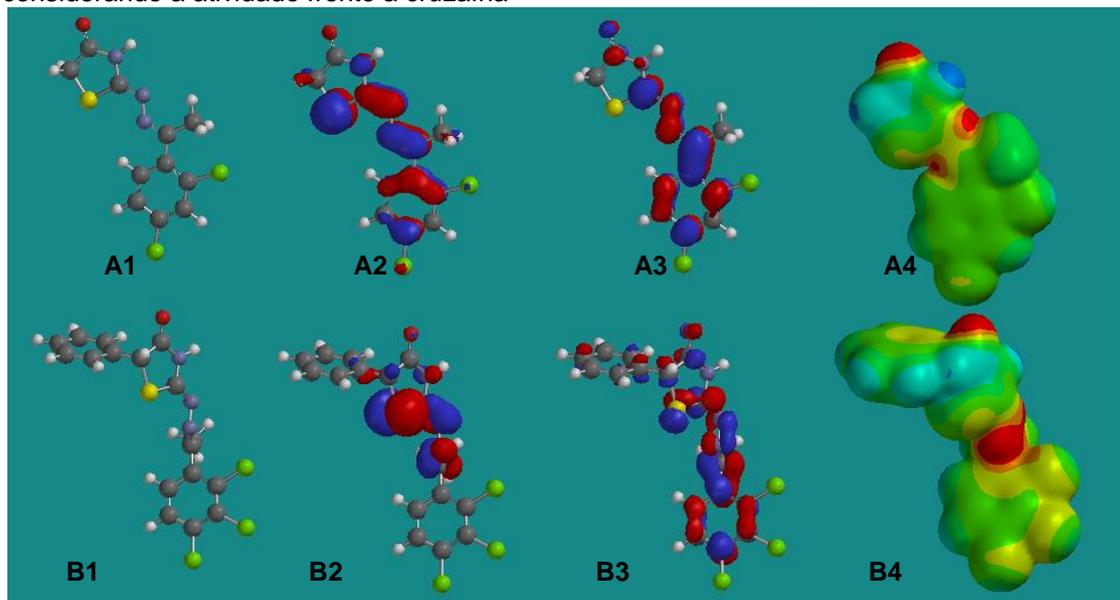
Figura 28 - Superfícies eletrônicas do composto mais ativo (LA-1c1) e menos ativo (LA-1b1) da série LA considerando a atividade frente a forma tripomastigota de *T. cruzi*



Legenda: A: Mais ativo; B: Menos ativo; 1: Estrutura de referência; 2: Homo; 3: Lumo; 4: Mapa de potencial eletrostático.

Fonte: RIBEIRO (2018).

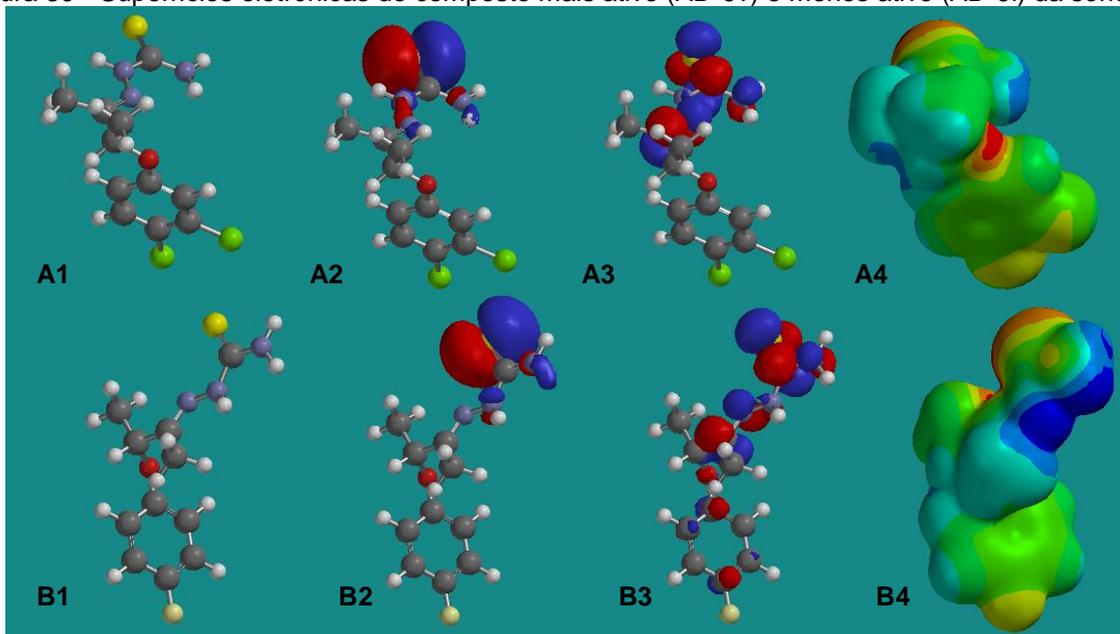
Figura 29 - Superfícies eletrônicas do composto mais ativo (LA-1c1) e menos ativo (LA-1a5) da série LA considerando a atividade frente a cruzaina



Legenda: A: Mais ativo; B: Menos ativo; 1: Estrutura de referência; 2: Homo; 3: Lumo; 4: Mapa de potencial eletrostático.

Fonte: RIBEIRO (2018).

Figura 30 - Superfícies eletrônicas do composto mais ativo (AD-9v) e menos ativo (AD-9i) da série AD



Legenda: A: Mais ativo; B: Menos ativo; 1: Estrutura de referência; 2: Homo; 3: Lumo; 4: Mapa de potencial eletrostático.

Fonte: RIBEIRO (2018).

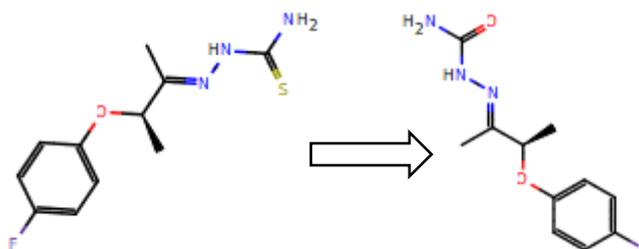
As superfícies eletrônicas se comportaram de forma muito semelhante dentro das séries, não sendo identificado diferenças significativas que pudessem justificar a diferença de atividade dos compostos, corroborando assim com os resultados obtidos através das análises quimiométricas e de QSAR.

#### 4.4 PREDIÇÃO DE METABÓLITOS

Os metabólitos preditos foram similares para cada uma das classes químicas analisadas devido a semelhança estrutural que compartilham. Por esse motivo, serão expostos aqui alguns casos específicos dos metabólitos com maior probabilidade de ocorrerem. A lista completa com todos os metabólitos preditos pode ser observada nos apêndices.

Para as tiosemicarbazonas, na maioria dos casos ocorreu a substituição do enxofre pelo oxigênio, como exemplificado na figura 34. Modificação essa que promove aumento da hidrofília do composto.

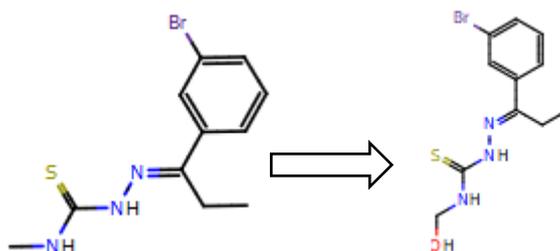
Figura 31 -Composto AD-9i e seu metabólito após interação com o citocromo P450



Fonte: RIBEIRO (2018).

Tiosemicarbazonas que apresentam grupos metil ligado ao nitrogênio, sofrem hidroxilação gerando os respectivos álcoois metílico, conforme pode ser exemplificado pelo intermediário PT03 (Figura 34)

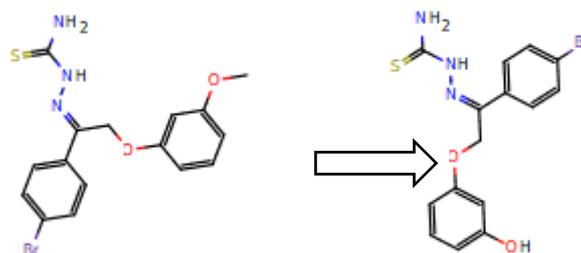
Figura 32 - Composto PT03 e seu metabólito após interação com o citocromo P450



Fonte: RIBEIRO (2018).

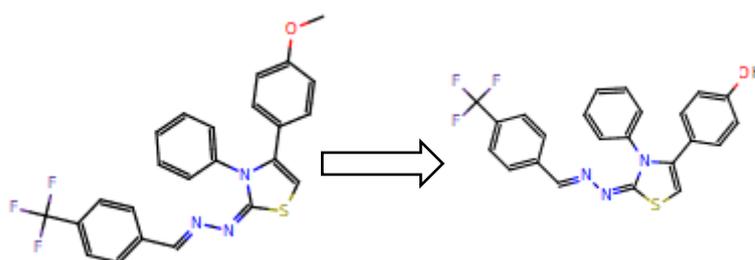
Da mesma forma, compostos com metoxilas ligadas à anéis benzênicos são oxidados aos respectivos fenóis, conforme pode ser observado nos exemplos apresentados nas figuras 36 e 37.

Figura 33 - Composto JW-9e e seu metabólito fenólico após interação com o citocromo P450.



Fonte: RIBEIRO (2018).

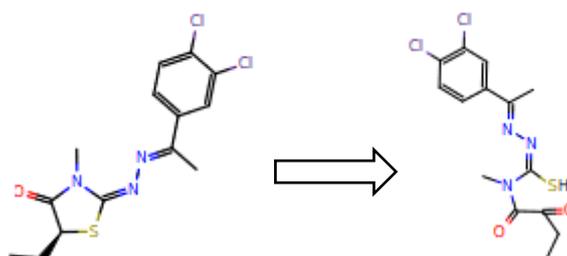
Figura 34 - Composto MB3-5 e seu metabólito fenólico após interação com o citocromo P450.



Fonte: RIBEIRO (2018).

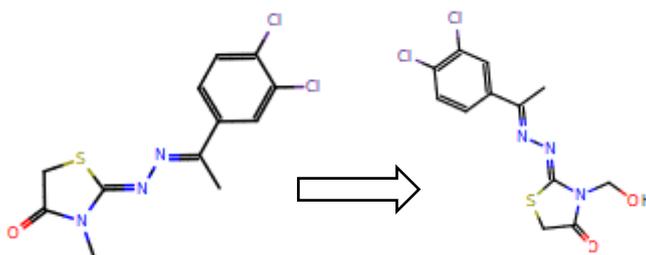
As tiazolidinonas também apresentaram padrão constante de metabolização, sendo a quebra do anel entre o enxofre e o C-5, a quebra mais frequente, levando a formação de um tiol e de uma carbonila (Figura 38) A única exceção foi o composto GT-06, em que o principal metabólito formado consistiu da reação de hidroxilação do grupo metila ligado ao Nitrogênio-3, assim como observado para as tiazemicarbazonas (Figura 39)

Figura 35 - Composto GT-06a e seu metabólito após interação com o citocromo P450



Fonte: RIBEIRO (2018).

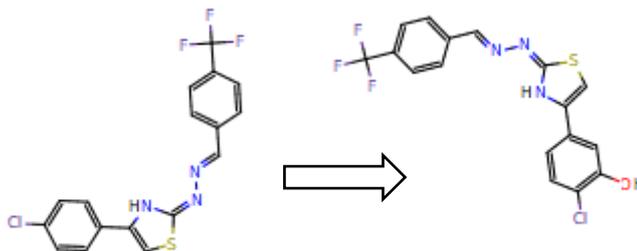
Figura 36 - Composto GT-06 e seu metabólito hidroxilado após interação com o citocromo P450



Fonte: RIBEIRO (2018).

Os derivados 1,3-tiazólicos tiveram como padrão a adição de uma hidroxila fenólica no benzeno ligado ao Carbono-4 do tiazol. (Figura 40).

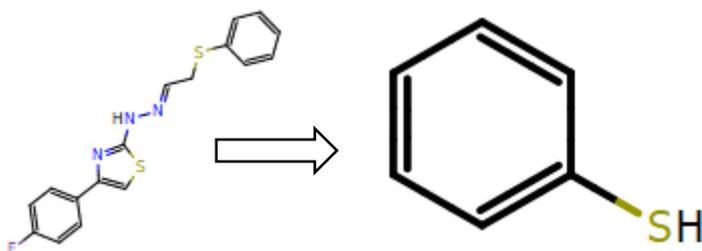
Figura 37 - Composto MB1-6 e seu metabólito hidroxilado após interação com o citocromo P450



Fonte: RIBEIRO (2018).

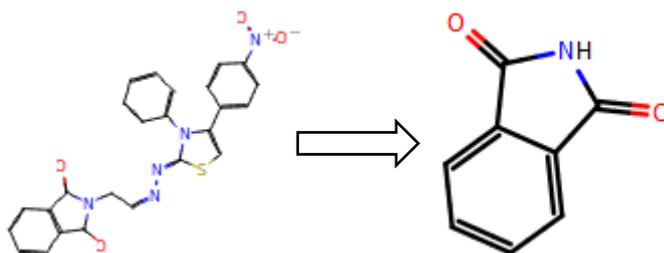
Exceções ocorreram com os compostos TZ-05, TZ-07, PA-6b, PA-6d e PA-6k onde houve a quebra da molécula gerando como metabólitos pequenos fragmentos como o benzotiol (Série TZ) (Figura 41) e a ftalimida (Série PA) (Figura 42).

Figura 38 - Composto MB1-6 e seu metabólito (benzotiol) após interação com o citocromo P450



Fonte: RIBEIRO (2018).

Figura 39 - Composto MB1-6 e seu metabólito (ftalimida) após interação com o citocromo P450



Fonte: RIBEIRO (2018).

## 5 CONCLUSÕES

- Através da docagem molecular, foram identificados os compostos com maior afinidade com o sítio ativo da cruzaina bem como os principais resíduos de aminoácidos envolvidos na formação do complexo ligante-receptor. Entretanto não foi observada correlação entre os dados *in vitro* (na enzima) e os scores obtidos *in silico*, indicando a impossibilidade de predição da atividade utilizando essa técnica de análise.
- Quanto aos métodos quimiométricos (CPCA e PCA para cada uma das atividades avaliadas *in vitro*), apenas o CPCA explicou um percentual da variância aceitável, superior a 60%, onde foi evidenciado que os blocos de descritores DRY e OH2 foram os de maior peso. Porém nenhuma das análises de PCA permitiu uma separação eficiente e aceitável dos compostos ativos e inativos.
- As análises de QSAR (PLS) não geraram modelos adequados, apresentando fatores de correlação e predição baixos, não sendo, portanto confiáveis para predição das atividades biológicas frente a cruzaina e as 3 formas evolutivas do *T. cruzi*.
- As análises qualitativas das superfícies eletrônicas (HOMO, LUMO e potencial eletrostático) não foram capazes de identificar padrões que diferenciem os compostos ativos dos inativos.
- Através do software MetaSite, foram preditos os principais metabólitos formados após a interação dos compostos com as enzimas citocromo P450 do fígado.

Durante o trabalho, foram desenvolvidos scripts para automatizar a execução da docagem molecular através do AutoDock, possibilitando a realização da triagem virtual, sendo eficaz para automatização e velocidade desse procedimento. Contudo, os scripts ainda estão em desenvolvimento com a proposta de incluir novas funções e deixar sua execução mais amigável para o usuário.

Quando os scores obtidos na docagem molecular foram comparados com os valores de atividade biológica frente a cruzaina, através da correlação de Spearman, não foi identificado correlação estatística entre as duas variáveis. Modificações no

protocolo da docagem usada nessa pesquisa podem ser realizadas na tentativa de identificar alguma correlação entre as duas variáveis. A primeira, consiste em modificar os valores correspondentes aos parâmetros do algoritmo genético Lamarkiano, melhorando a estatística e robustez dos resultados. Outra possibilidade é considerar os resíduos de aminoácidos do sítio ativo como móveis, permitindo a mudança de conformação dos resíduos, de acordo com a interação deles com os ligantes. Pode-se ainda implementar a inclusão de moléculas de água a docagem. Sabe-se que as moléculas de água podem intermediar a formação de ligações de hidrogênio entre o ligante e o sítio ativo. Logo, pode-se investigar a importância das moléculas de água na formação do complexo com a cruzaina e incluí-las na análise de docagem. Esse processo é controverso pois as moléculas de água podem prejudicar a busca por poses dos ligantes, mas é uma possibilidade viável.

Com relação a quimiometria e a análise de QSAR, os dados não foram suficientes para gerar bons modelos preditivos da atividade frente as diferentes formas evolutivas de *T. cruzi*. Uma das formas possíveis para obter um melhor modelo é o aumento do número de compostos de modo a ter uma maior variação dos valores da atividade biológica. Esses novos compostos, podem ser obtidos através de dados de outras pesquisas publicadas, bem como, de bancos de dados online como o ZINC e o ChEMBL.

A predição dos metabólitos, forneceu abertura para que sejam realizadas novas análises *in silico* para prever a atividade desses metabólitos frente ao *T. cruzi*, por meio da docagem molecular e análise de QSAR, assim como auxiliar na predição da toxicidade dos metabólitos a partir de ferramentas computacionais de ADME-Tox (absorção, distribuição, metabolismo, excreção e toxicidade) e PAINS (*pan-assay interference compounds*).

## REFERÊNCIAS

- ALCAIDE, Á. D. **Development of high-performance algorithms for a new generation of versatile molecular descriptors. The Pentacle software.** [s.l.] Universitat Pompeu Fabra, 2009.
- BERMAN, H. M. et al. The Protein Data Bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 2000.
- BERN, C. Chagas' Disease. **New England Journal of Medicine**, v. 373, n. 5, p. 456–466, 2015.
- BINGHAM, R. C.; DEWAR, M. J. S.; LO, D. H. Ground states of molecules. XXV. MINDO/3. Improved version of the MINDO semiempirical SCF-MO method. **Journal of the American Chemical Society**, v. 97, n. 6, p. 1285–1293, mar. 1975.
- COELHO, L. W. et al. Aplicação de mecânica molecular em química inorgânica. **Química Nova**, v. 22, n. 3, p. 396–404, 1999.
- COHEN, N. C. **Guidebook on Molecular Modeling in Drug Design.** 1. ed. California: Academic Press, 1996.
- COURA, J. R.; DE CASTRO, S. L. A critical review on chagas disease chemotherapy. **Memórias do Instituto Oswaldo Cruz**, v. 97, n. 1, p. 3–24, 2002.
- DAAR, A S. et al. Top ten biotechnologies for improving health in developing countries. **Nat Genet**, v. 32, n. 2, p. 229–232, 2002.
- DAVANÇO, M. G. Farmacocinética Do Benznidazol Administrado Em Coelhos Na Forma De Comprimidos De Liberação Imediata E Comprimidos De Liberação Prolongada. 2015.
- DE MORAES GOMES, P. A. T. et al. New 1,3-thiazole derivatives and their biological and ultrastructural effects on Trypanosoma cruzi. **European Journal of Medicinal Chemistry**, v. 121, p. 387–398, 2016.
- DE OLIVEIRA CARDOSO, M. V. et al. 2-Pyridyl thiazoles as novel anti-Trypanosoma cruzi agents: Structural design, synthesis and pharmacological evaluation. **European Journal of Medicinal Chemistry**, v. 86, p. 48–59, 2014.
- DE OLIVEIRA FILHO, G. B. et al. Structural design, synthesis and pharmacological evaluation of 4-thiazolidinones against Trypanosoma cruzi. **Bioorganic & Medicinal Chemistry**, v. 23, n. 23, p. 7478–7486, 2015.
- DEWAR, M. J. S. et al. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. **Journal of the American Chemical Society**, v. 107, n. 13, p. 3902–3909, jun. 1985.
- DEWAR, M. J. S.; THIEL, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. **Journal of the American Chemical Society**, v. 99,

n. 15, p. 4899–4907, jun. 1977.

DOYLE, P. S. et al. The trypanosoma cruzi protease cruzain mediates immune evasion. **PLoS Pathogens**, v. 7, n. 9, p. 1–11, 2011.

DURRANT, J. D.; MCCAMMON, J. A. Computer-aided drug-discovery techniques that account for receptor flexibility. **Current Opinion in Pharmacology**, v. 10, n. 6, p. 770–774, 2010.

ESPÍNDOLA, J. W. P. et al. Synthesis and structure-activity relationship study of a new series of antiparasitic aryloxy thiosemicarbazones inhibiting Trypanosoma cruzi cruzain. **European Journal of Medicinal Chemistry**, v. 101, p. 818–835, 2015.

FERREIRA DA SILVA-JÚNIOR, E. et al. Molecular Docking Studies Applied to a Dataset of Cruzain Inhibitors. **Current Computer-Aided Drug Design**, v. 13, p. 1–12, 19 maio 2017.

FILHO, J. M. DOS S. et al. Design, synthesis and cruzain docking of 3-(4-substituted-aryl)-1,2,4-oxadiazole-N-acylhydrazones as anti- Trypanosoma cruzi agents. **Bioorganic & Medicinal Chemistry**, v. 17, n. 18, p. 6682–6691, 2009.

GARCIA-DOMENECH, R. et al. Predicting antiprotozoal activity of benzyl phenyl ether diamine derivatives through QSAR multi-target and molecular topology. **Molecular Diversity**, v. 19, n. 2, p. 357–366, 2015.

ILIES, M. A. New synthetic strategies for the management of chagas disease (American trypanosomiasis). **Journal of Medicinal Chemistry**, v. 57, n. 2, p. 296–297, 2014.

KARELSON, M.; LOBANOV, V. S.; KATRITZKY, A. R. Quantum-Chemical Descriptors in QSAR/QSPR Studies. **Chemical reviews**, v. 96, n. 3, p. 1027–1044, 9 maio 1996.

KUMAR, G. D. K. et al. Design, synthesis, and biological evaluation of potent thiosemicarbazone based cathepsin L inhibitors. **Bioorganic & Medicinal Chemistry Letters**, v. 20, n. 4, p. 1415–1419, 2010a.

KUMAR, G. D. K. et al. Functionalized benzophenone, thiophene, pyridine, and fluorene thiosemicarbazone derivatives as inhibitors of cathepsin L. **Bioorganic & Medicinal Chemistry Letters**, v. 20, n. 22, p. 6610–6615, 2010b.

LEACH, A. R. **Molecular Modeling: Principles and Applications**. [s.l.: s.n.].

LEITE, A. C. L. et al. Synthesis, docking, and in vitro activity of thiosemicarbazones, aminoacyl-thiosemicarbazides and acyl-thiazolidones against Trypanosoma cruzi. **Bioorganic and Medicinal Chemistry**, v. 14, n. 11, p. 3749–3757, 2006.

LO PRESTI, M. S. et al. Trypanothione reductase inhibitors: Overview of the action of thioridazine in different stages of Chagas disease. **Acta Tropica**, v. 145, p. 79–87, 2015.

MAGUIRE, J. H. Treatment of Chagas' Disease — Time Is Running Out. **New England Journal of Medicine**, v. 373, n. 14, p. 1369–1370, 2015.

MAITRA, S.; YAN, J. Principle Component Analysis and Partial Least Squares : **Casualty Actuarial Society**, p. 79–90, 2008.

MARTINEZ-MAYORGA, K. et al. Cruzain inhibitors: efforts made, current leads and a structural outlook of new hits. **Drug Discovery Today**, v. 00, n. 00, p. 1–9, 2015.

MCGRATH, M. E. et al. The crystal structure of cruzain: a therapeutic target for Chagas' disease. **Journal of molecular biology**, v. 247, n. 2, p. 251–9, 1995.

MENG, X.-Y. et al. Molecular docking: a powerful approach for structure-based drug discovery. **Current computer-aided drug design**, v. 7, n. 2, p. 146–57, 2011.

MIKRA, C. et al. Molecular Docking and Structure Activity Relationship Studies of NSAIDs. What do they Reveal about IC50? **Letters in Drug Design & Discovery**, v. 14, n. 8, p. 949–958, 2017.

MOLECULAR DISCOVERY LTD. **VolSurf+ user manual**. [s.l: s.n.].

MOLECULAR DISCOVERY LTD. **MetaSite user manual**. [s.l: s.n.].

MONTANARI, C. A. A química medicinal na próxima década. **Química Nova**, v. 23, n. 1, p. 134–137, 2000.

MOREIRA, D. R. M. et al. Conformational restriction of aryl thiosemicarbazones produces potent and selective anti-Trypanosoma cruzi compounds which induce apoptotic parasite death. **European Journal of Medicinal Chemistry**, v. 75, p. 467–478, 2014.

MORELLATO, S. A. **Modelos de regressão PLS com erros heteroscedásticos**. [s.l.] Universidade Federal e São Carlos, 2010.

NUSSBAUM, K. et al. Trypanosomatid parasites causing neglected diseases. **Current medicinal chemistry**, v. 17, n. 15, p. 1594–1617, 2010.

OGU, C. C.; MAXA, J. L. Drug interactions due to cytochrome P450. **Proceedings (Baylor University. Medical Center)**, v. 13, n. 4, p. 421–3, 2000.

ORGANIZAÇÃO MUNDIAL DA SAÚDE. **Avanços para superar o impacto global de doenças. Primeiro relatório da OMS sobre doenças tropicais negligenciadas**. [s.l: s.n.].

PHRMA. **2015 Biopharmaceutical Research Industry Profile**. Washington: PhRMA, 2015.

PIZZO, C. et al. Synthesis of 2-hydrazolyl-4-thiazolidinones based on multicomponent reactions and biological evaluation against Trypanosoma Cruzi.

**Chemical biology & drug design**, v. 77, n. 3, p. 166–72, mar. 2011.

RASSI, A.; RASSI, A.; MARIN-NETO, J. A. Chagas disease. **The Lancet**, v. 375, n. 9723, p. 1388–1402, 2010.

RIBEIRO, F. F. et al. Computational and Investigative Study of Flavonoids Active Against *Trypanosoma cruzi* and *Leishmania* spp. **Natural Product Communications**, v. 10, n. 6, p. 917–920, 2015.

RIBEIRO, J. R. V. **Computer-aided Drug Design: Lead Discovery**. [s.l.] Faculdade de Ciências da Universidade do Porto, 2014.

ROBINSON, M. W.; DALTON, J. P. **Cysteine Proteases of Pathogenic Organisms**. 1. ed. New York: Springer, 2011. v. 712

SANT'ANNA, C. M. R. Glossário de termos usados no planejamento de fármacos (Recomendações da IUPAC para 1997). **Química Nova**, v. 25, n. 3, p. 505–512, 2002.

SANT'ANNA, C. M. R. Molecular modeling methods in the study and design of bioactive compounds: An introduction. **Revista Virtual de Química**, v. 1, n. 1, 2009.

SAÚDE, M. DA. **Doenças infecciosas e parasitárias**. 8. ed. Brasília: Ministério da Saúde, 2010.

SAXENA, A. et al. HHS Public Access. v. 74, n. 1, p. 27–34, 2016.

SCOTTI, L. **Modelagem Molecular Aplicada À Cosmetologia : Planejamento De Compostos Antienvhecimento**. [s.l.] Universidade de São Paulo, 2006.

SHEKHAR, C. In Silico Pharmacology: Computer-Aided Methods Could Transform Drug Development. **Chemistry and Biology**, v. 15, p. 413–414, 2008.

SILVA, J. J. N. et al. Quimioterapia da doença de Chagas: estado da arte e perspectivas no desenvolvimento de novos fármacos. **Quim. Nova**, v. 32, n. 9, p. 2444–2457, 2009.

SONG, J. et al. Synthesis and biochemical evaluation of thiochromanone thiosemicarbazone analogues as inhibitors of cathepsin L. **ACS Medicinal Chemistry Letters**, v. 3, n. 6, p. 450–453, 2012.

SOUZA, M. L. DE. **Planejamento de inibidores da enzima cruzáina de *Trypanosoma cruzi* candidatos a fármacos contra a doença de Chagas**. [s.l.: s.n.].

SOUZA, W. DE. **Doenças negligenciadas**. Rio de Janeiro: Academia Brasileira de Ciências, 2010. v. 1

SPECK-PLANCHE, A.; CORDEIRO, M. N. D. S. Evolution of Graph Theory-Based QSAR Methods and Their Applications to the Search for New Antibacterial Agents.

**Current Topics in Medicinal Chemistry**, v. 13, p. 3101–3117, 2013.

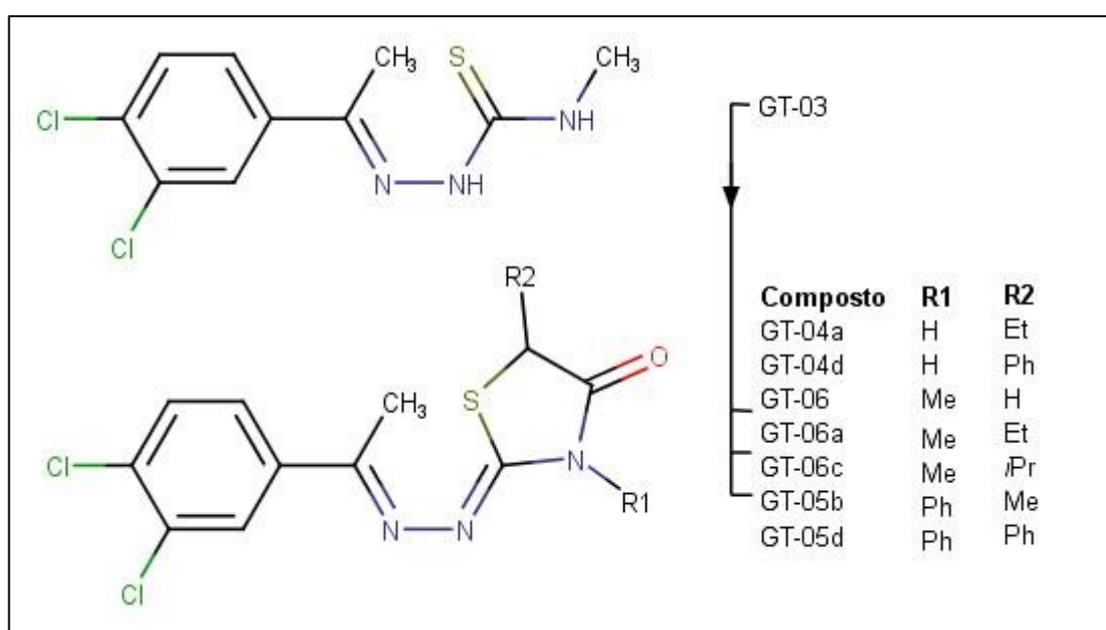
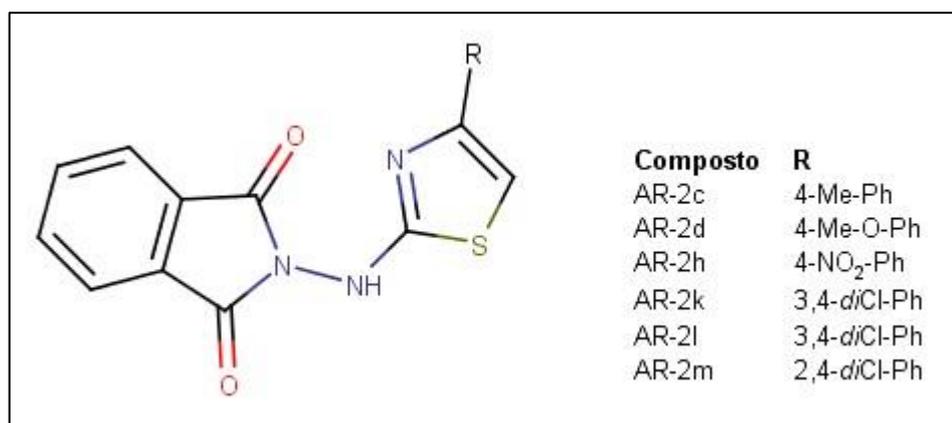
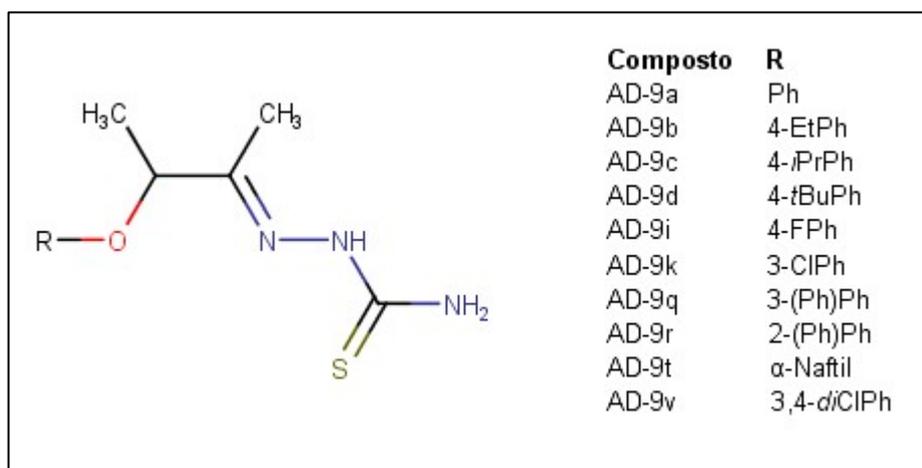
VASILESCU, D.; GIRMA, R. Quantum molecular modeling of quercetin simulation of the interaction with the free radical t-BuOO. **International Journal of Quantum Chemistry**, v. 90, n. 2, p. 888–90, 2002.

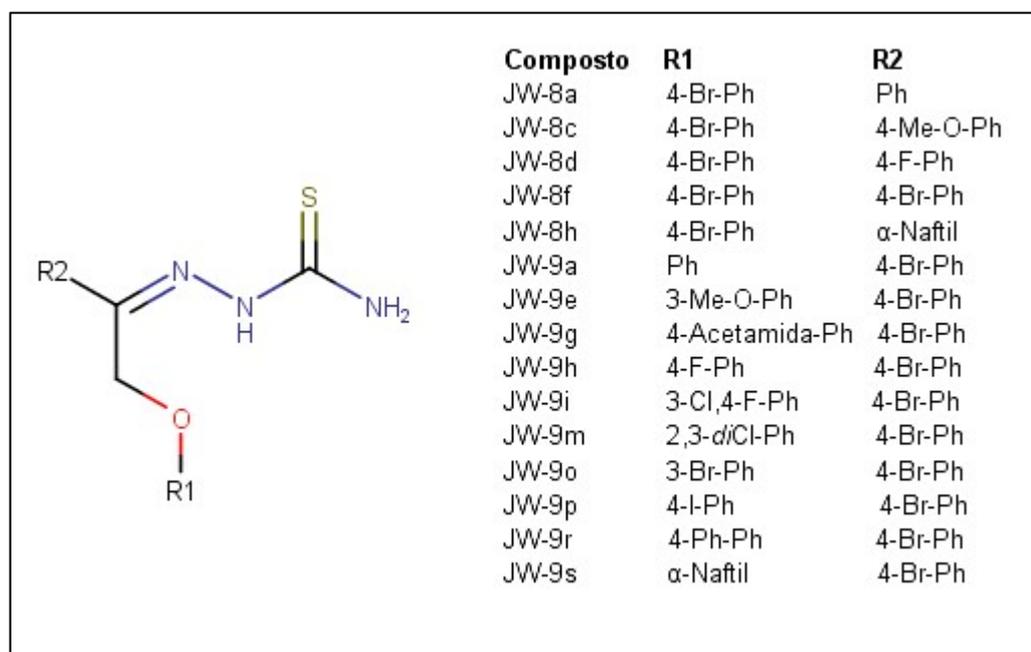
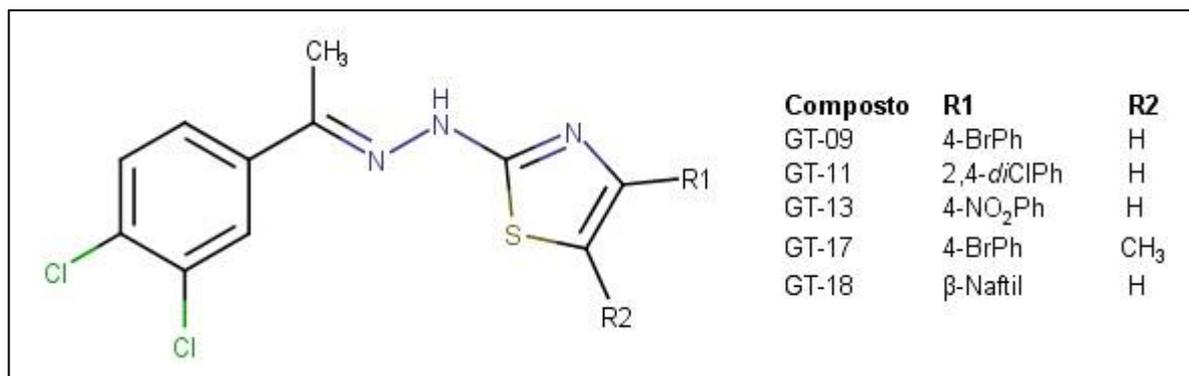
WESTERHUIS, J. A; KOURTI, T.; MACGREGOR, J. F. Analysis of multiblock and hierarchical PCA and PLS models. **J. Chemometrics**, v. 12, n. 5, p. 301–321, 1998.

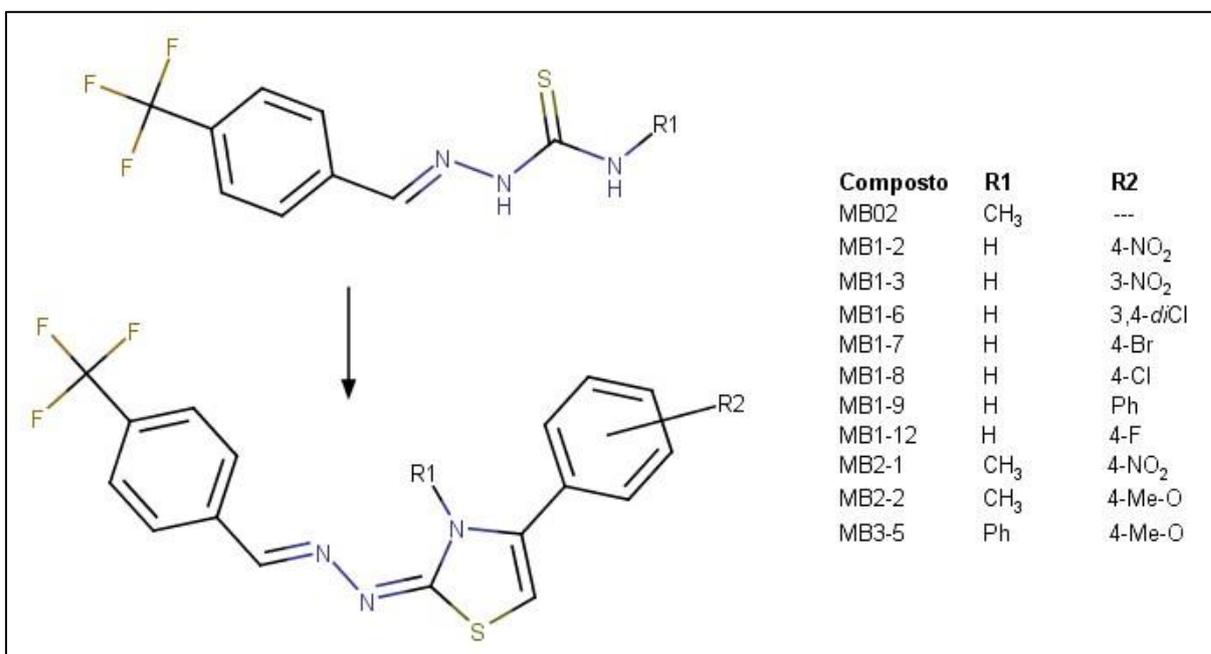
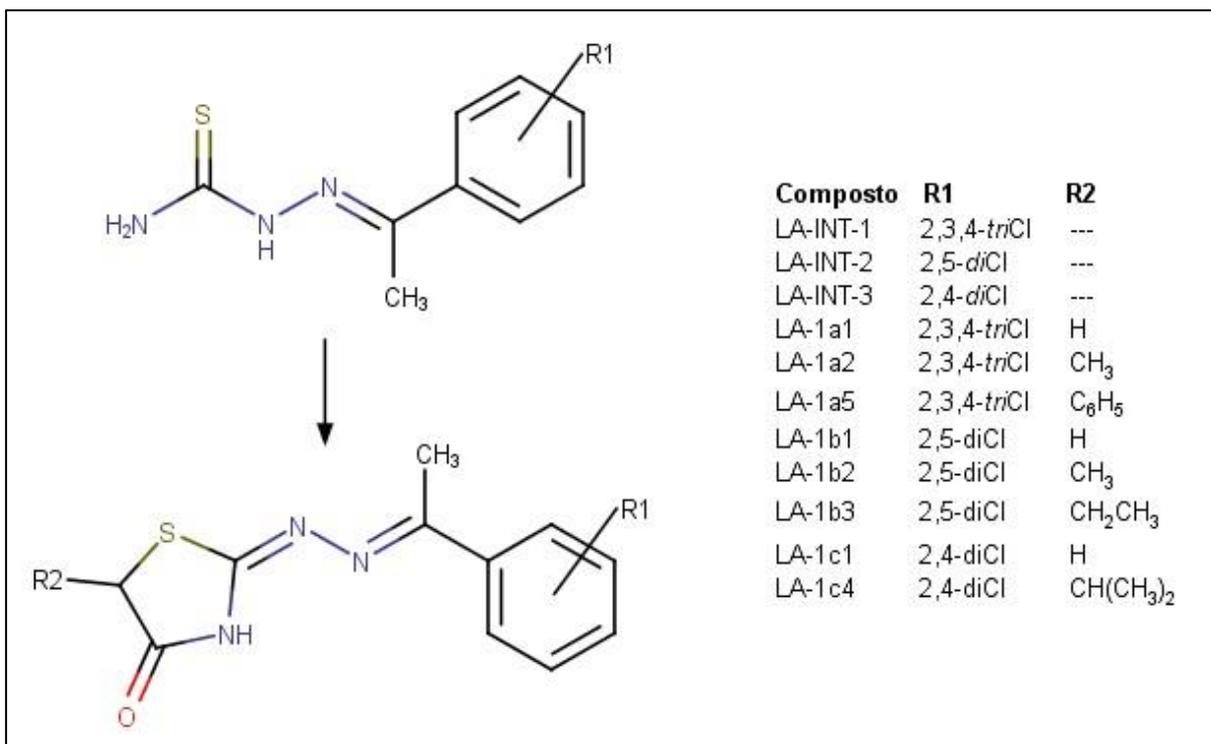
WIGGERS, H. J. et al. Non-peptidic cruzain inhibitors with trypanocidal activity discovered by virtual screening and in vitro assay. **PLoS neglected tropical diseases**, v. 7, n. 8, p. e2370, jan. 2013.

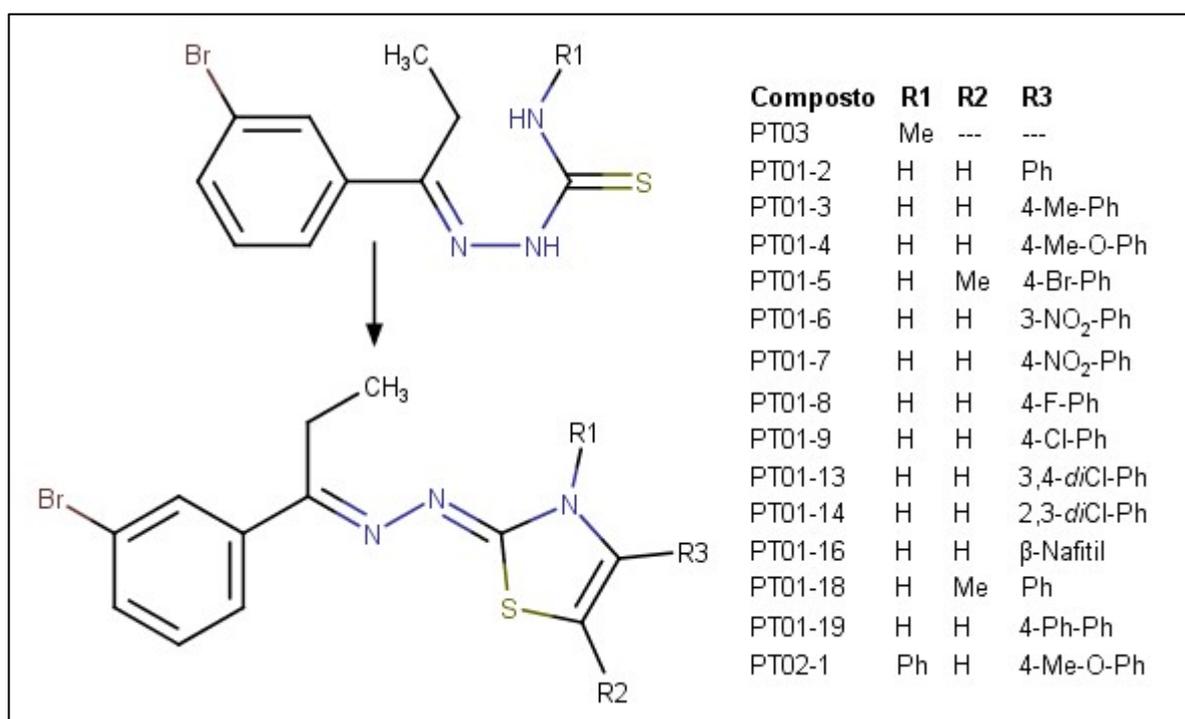
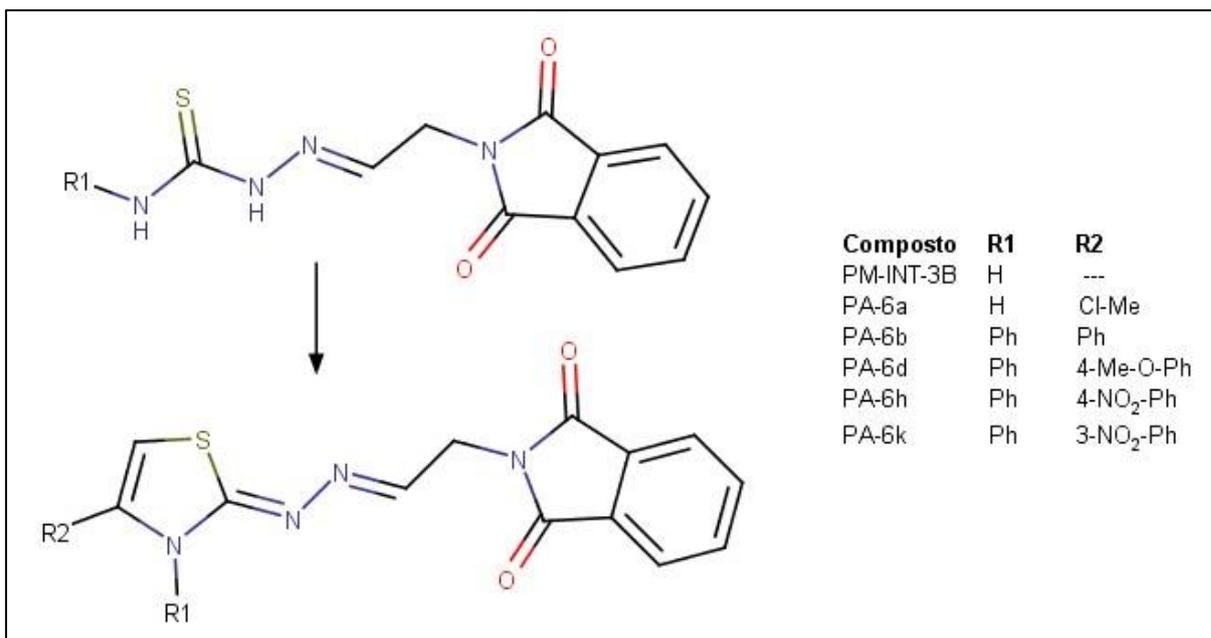
WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

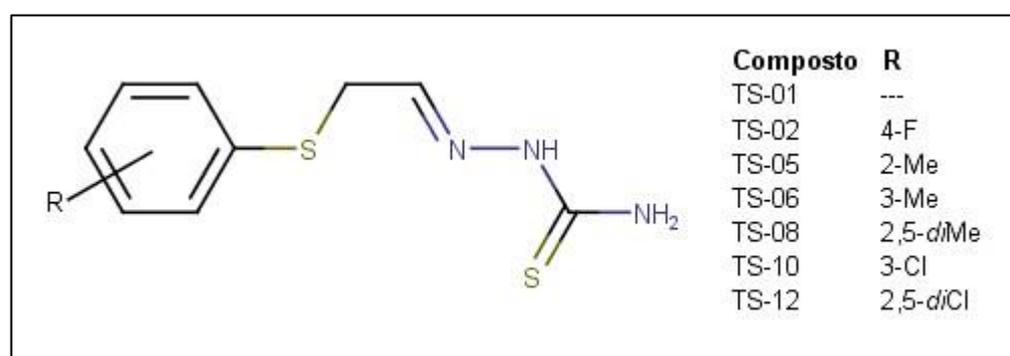
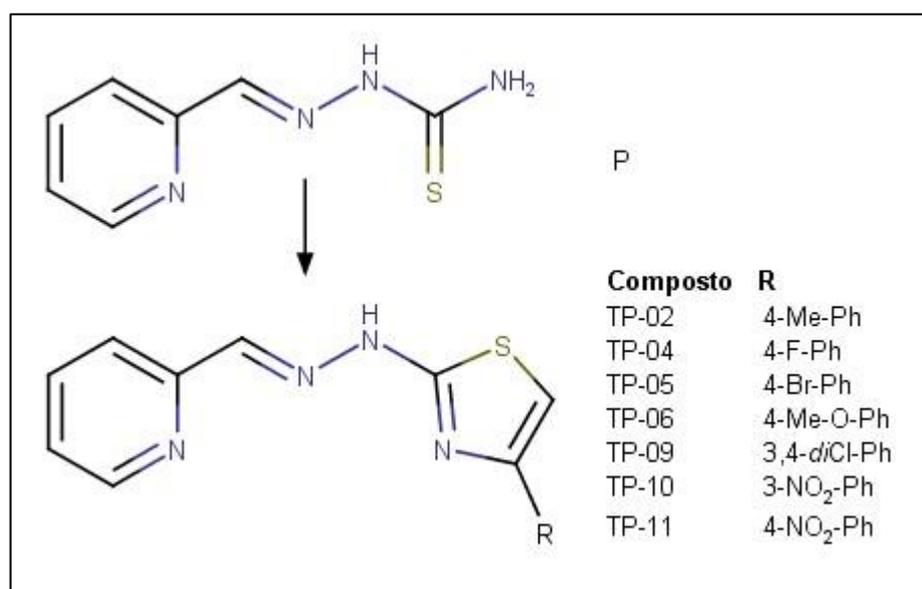
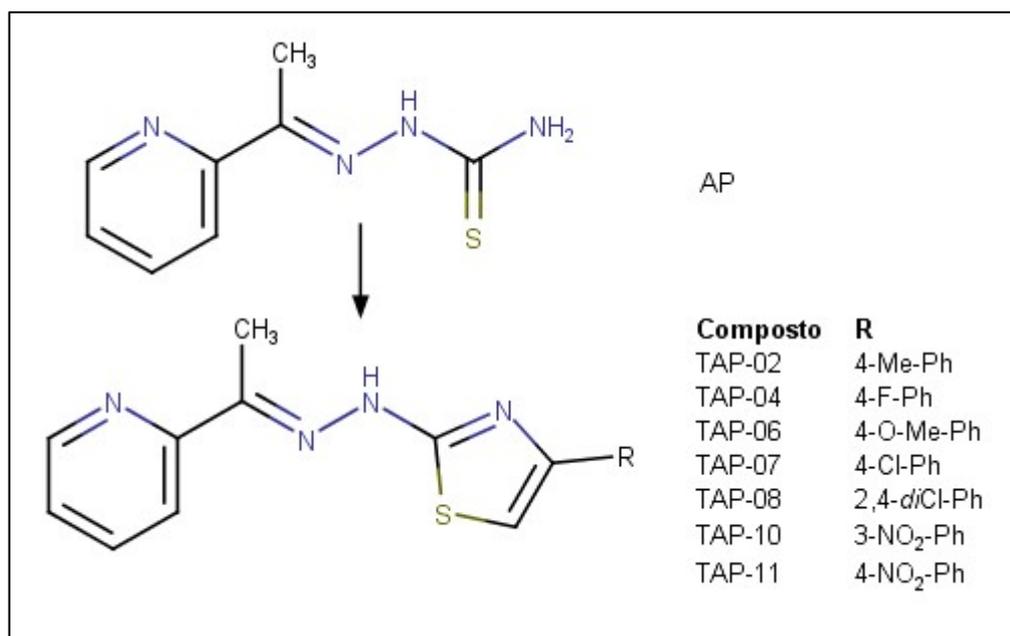
## APÊNDICE A - ESTRUTURA DAS MOLÉCULAS

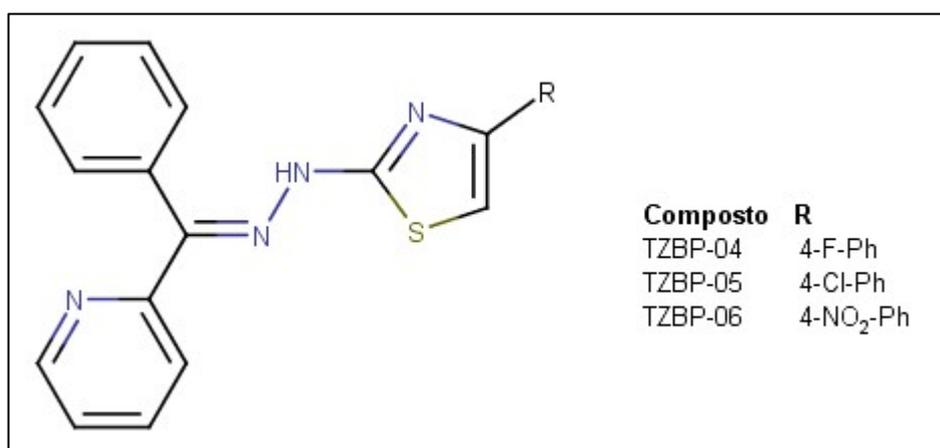
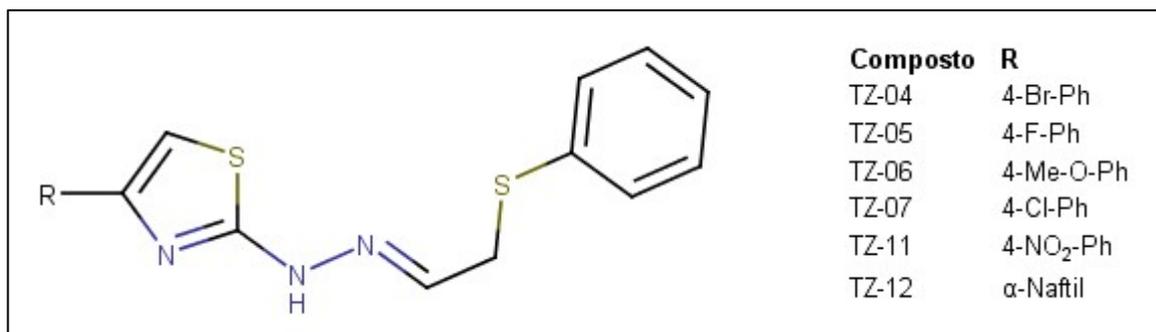












## APÊNDICE B - ATIVIDADE BIOLÓGICA DOS COMPOSTOS

Composto	Epimastigota	Amastigota	Tripomastigota	Cruzaina
AD-9a	26.1	210.6	27.3	
AD-9b	11.6	18.8	4.6	
AD-9c	5.9	17.9	1.5	
AD-9d	4.6	17	1.4	
AD-9i	86.4	>391.8	23.2	
AD-9k	45.8	36.8	5.5	
AD-9q	2.2	31.9	1.3	
AD-9r	2.2	31.9		
AD-9t	4.9	34.8	1.5	
AD-9v	4.4	16.3	4.4	
AP	4.5		2.8	
AR-2c	70.38	5	106.98	97.5
AR-2d	6.45	5	50.28	61.1
AR-2h	13.27	1	84.94	62.8
AR-2k	43.16	1	91.06	75.3
AR-2l	69.28	1	88.82	93.3
AR-2m	10.13	5	22.74	27.2
GT-03			11.29	
GT-04A			>50	
GT-04D			>50	
GT-05B			4.25	
GT-05D			17.76	
GT-06			>50	
GT-06A			48.56	
GT-06C			>50	
GT-09			>50	
GT-11			>50	
GT-13			>50	
GT-17			19.46	
GT-18			>50	
JW-8a	14.5	68.6	2.4	67.6
JW-8c	3.1	2.5	5	38.7
JW-8d	3.8	2.6	3.6	53.1
JW-8f	11.9	22.6	3.9	11.8
JW-8h	7.5	2.4	14.7	73.6
JW-9a	3	137.3	6.6	61.8
JW-9e	3.7	2.5	1.2	62.8
JW-9g	15.6	2.4	3.5	80.7
JW-9h	1.8	2.6	3.1	84
JW-9i	3.4	2.4	1.7	93.7

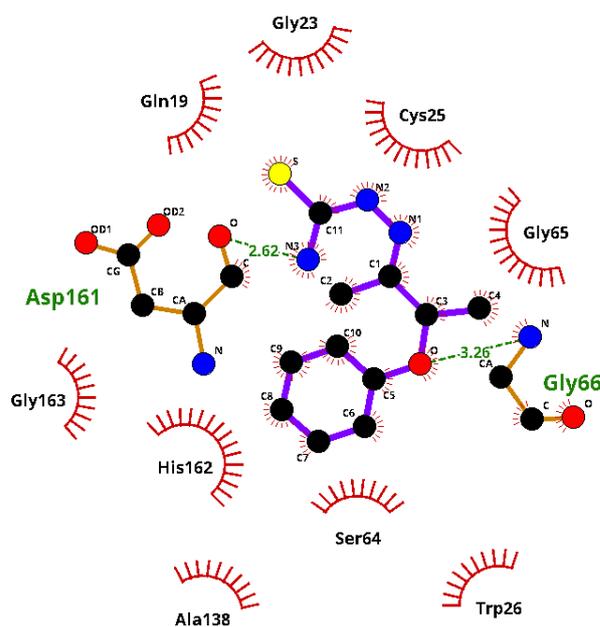
Continua

<b>Composto</b>	<b>Epimastigota</b>	<b>Amastigota</b>	<b>Tripomastigota</b>	<b>Cruzaina</b>
JW-9m	4.4	23.1	5.7	81.8
JW-9o	3.9	2.7	6.2	81.3
JW-9p	2.9	2.1	2.1	63.3
JW-9r	4.1	22.7	12.3	63.7
JW-9s	3	2.4	3.4	
LA-1a1		74.26	9.2	
LA-1a2		71.29	65.47	40.4
LA-1a5			30.13	76.7
LA-1b1		82.73	78.49	65.3
LA-1b2		79.06	41.5	33.9
LA-1b3		≥100	4.99	32.9
LA-1c1		82.73	1.1	32.2
LA-1c4				
LA-INT-1		84.28	41.8	67.3
LA-INT-2			2.97	49.8
LA-INT-3			3.66	30.6
MB02	14.44	50	6.37	
MB1.12	3.877	5	12.05	
MB1.2	31.03	10	28.15	
MB1.3	12.83	>100	31.73	
MB1.6	11.31	50	15.56	
MB1.7	3.639	>100	34.24	
MB1.8	10.41	5	17.8	
MB1.9	3.937	25	28.73	
MB2.1	32.68	>100	66.24	
MB2.2	7.357		22.69	
MB3.5		10	4.33	1.35
P	119.5		17.3	
PA-6a	85.67	73.67	2.21	29.6
PA-6b		228.05	8.76	25.9
PA-6d		213.43	10.03	20.9
PA-6h	12.18	206.82	3.18	31
PA-6k	5.99	103.41	0.89	37.4
PM-Int3B	36.94	147.76		46.6
PT01.13		54.92	129.72	33.6
PT01.14	46.31	54.92	59.2	95.2
PT01.16		229.16	50.98	52.7
PT01.18		12.49	4.79	86.3
PT01.19		216.26	3.35	24.5
PT01.2	73.8	64.71	5.51	79.2
PT01.3	3.41	125.30	16.64	58.5
PT01.4		120.10	3.84	55.2
PT01.5	69.3	20.87	5.8	57.1
PT01.6	28.98	23.18	15.05	97.3

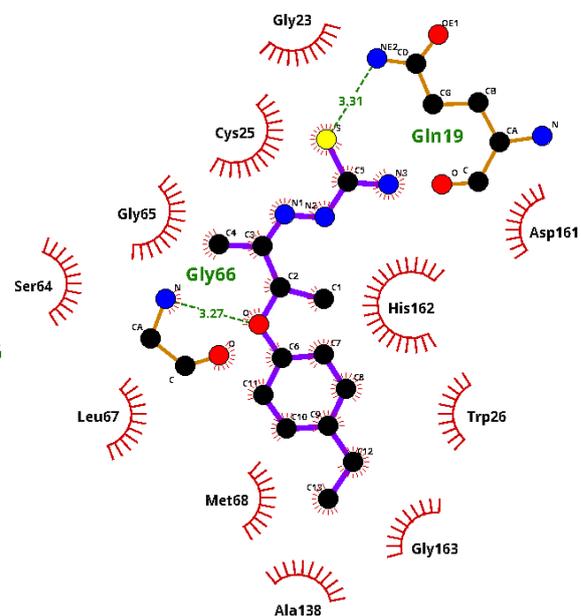
Continua

<b>Composto</b>		<b>Epimastigota</b>	<b>Amastigota</b>	<b>Tripomastigota</b>	<b>Cruzaina</b>
PT01.7		231.85	60	61.3	
PT01.8		123.67	146.05	72.8	
PT01.9		118.83	140.34	51.3	
PT02.1		50.77	53.1	34.3	
PT03		33.31	36.31	11.7	
TAP-02	2.5	100	1.7	90	
TAP-04	2.8	91.2	3.8	89	
TAP-06	14.7	100	2.1	95	
TAP-07	2.5	100	2.1	-3	
TAP-08	14.9	100	1.9	81	
TAP-10	2.4	100	3.9	10	
TAP-11	5.2	100	36.7	11	
TP-02	3.8	100	16.8	23	
TP-04	3.5	18.2	72.7	22	
TP-05	18.3	100	2.6	10	
TP-06	5.6	100	4.8	20	
TP-09	4	100	1.2	21	
TP-10	2.5	100	74.2	85	
TP-11	4.4	100	152.9	12	
TS-01	101.89		88.98	42.1	
TS-02	6.08	20.55	5.75	27.5	
TS-05	15.33	41.78	6.1	13.9	
TS-06	6.26	<4.17	1.8	43.4	
TS-08	6.16	<3.95	5.92	87.6	
TS-10	5.47	96.24	14.13	75.9	
TS-12	6.08	84.96	8.16	37.4	
TZ-04	5.43	>239.02	14.34	7	
TZ-05		>291.17	41.67	22.9	
TZ-06	62.73	>281.31	18.51		
TZ-07	6.81	>277.85	4.06	17.6	
TZ-11	16.03	67.49	8.45	18.9	
TZ-12	4.39	>266.30	2	61.8	
TZBP-04	23.18	2.67	2.41		
TZBP-05		12.82	5.23		
TZBP-06	6.51	2.49	5.46		

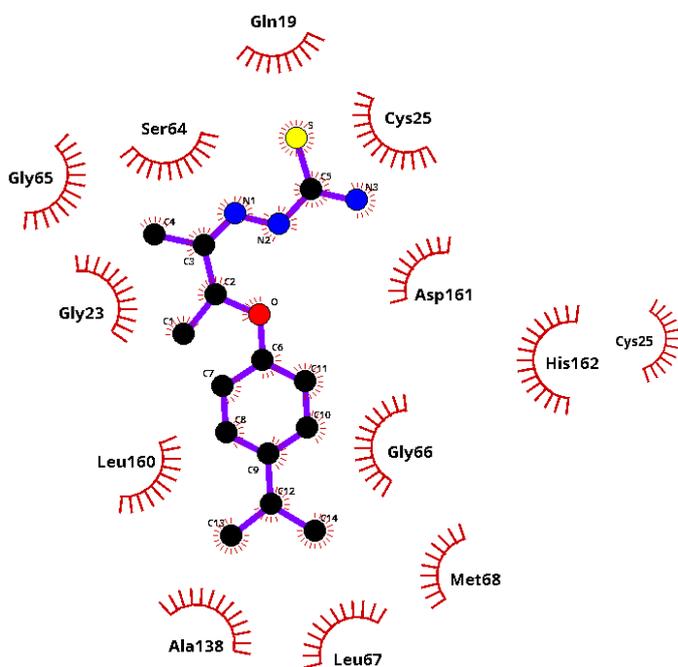
## APÊNDICE C - INTERAÇÕES FORMADAS ENTRE OS LIGANTES E O SÍTIO ATIVO DA CRUZAINA



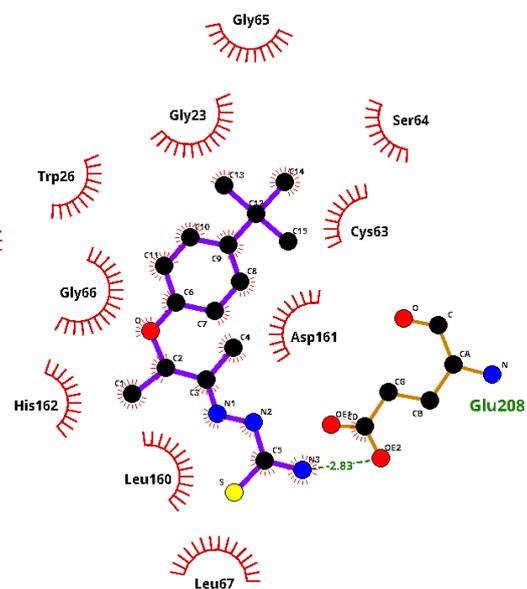
AD\_9a\_3i06



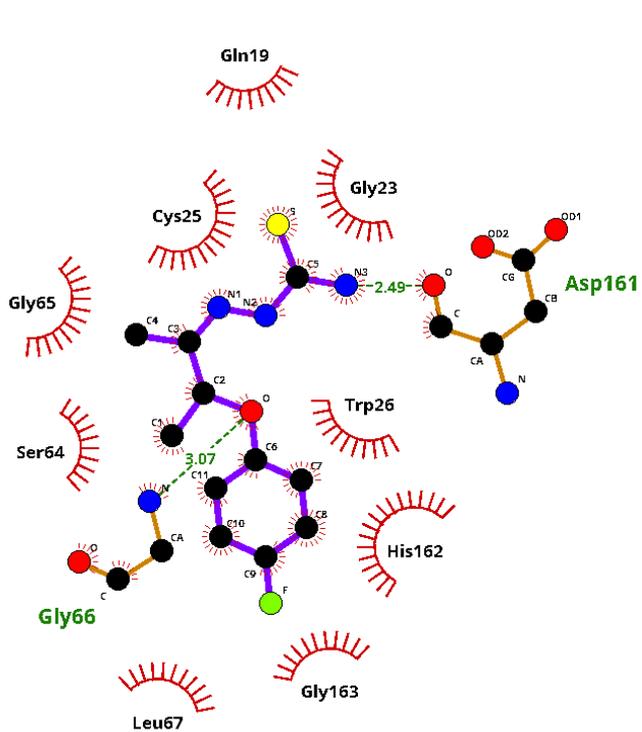
AD\_9b\_3i06



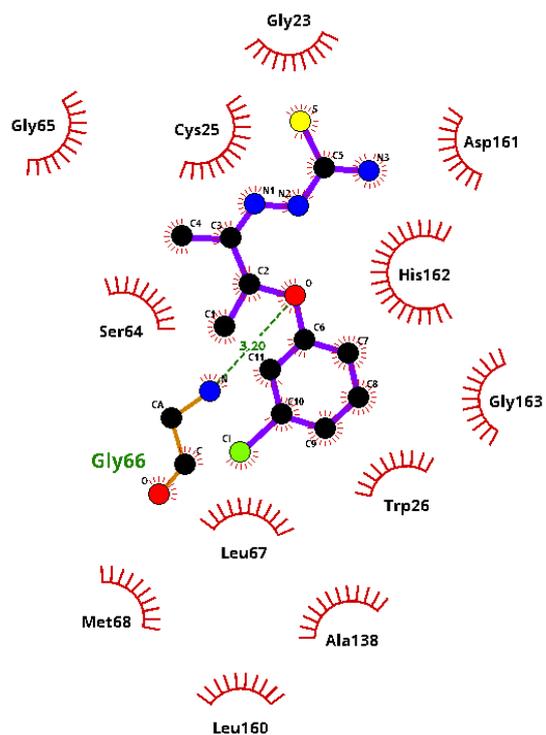
AD\_9c\_3i06



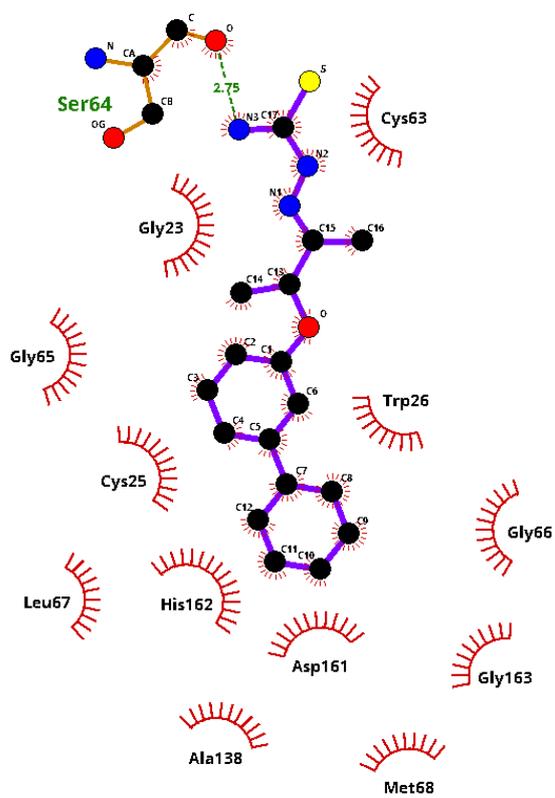
AD\_9d\_3i06



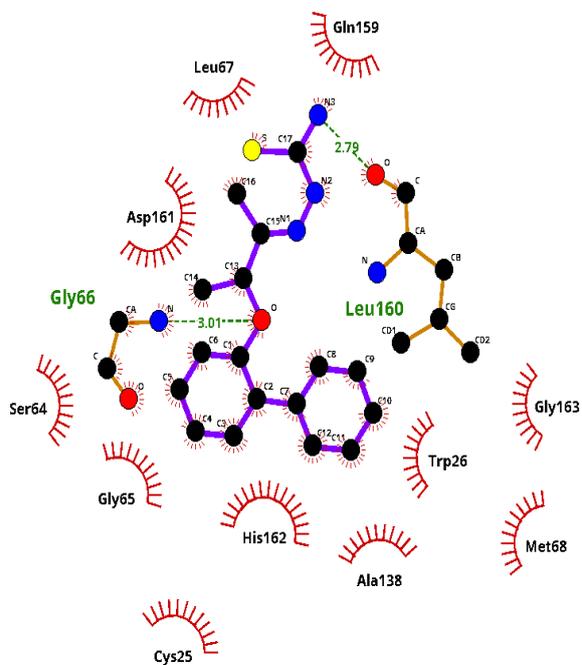
AD\_9i\_3i06



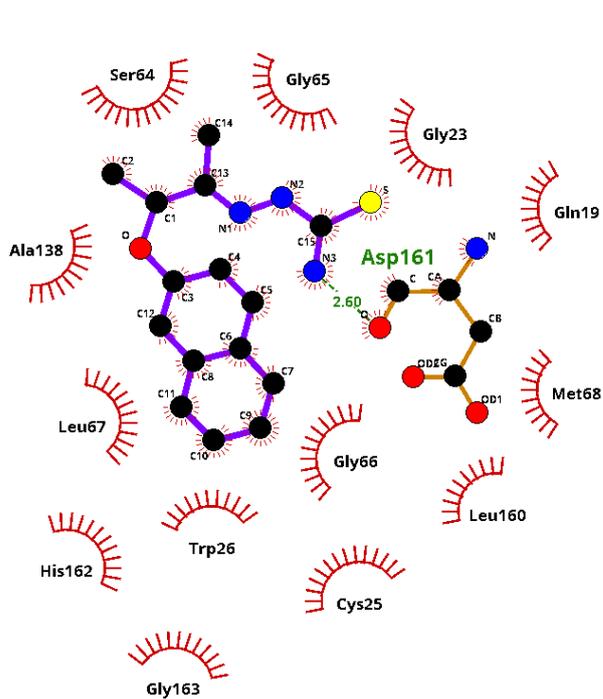
AD\_9k\_3i06



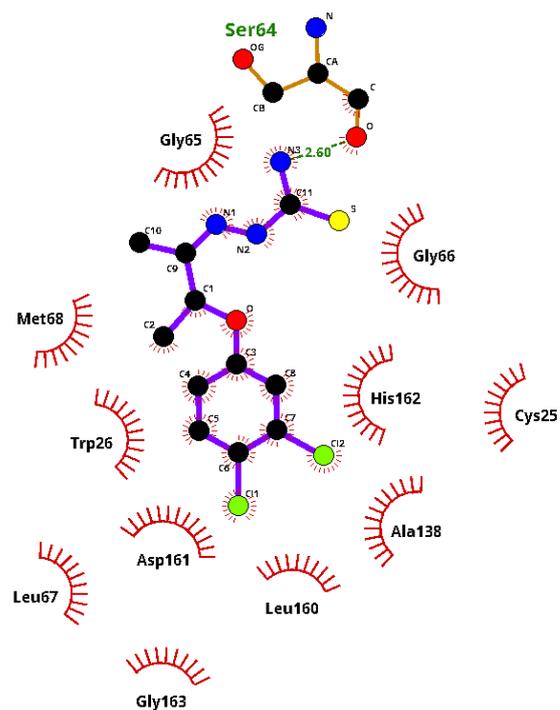
AD\_9q\_3i06



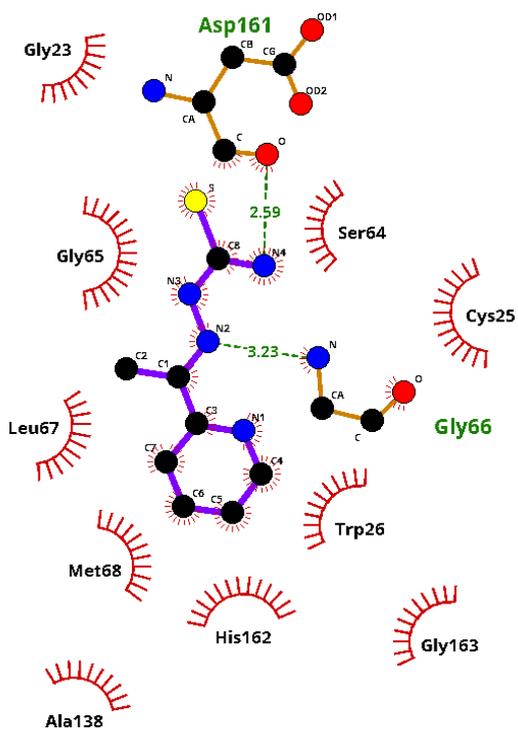
AD\_9r\_3i06



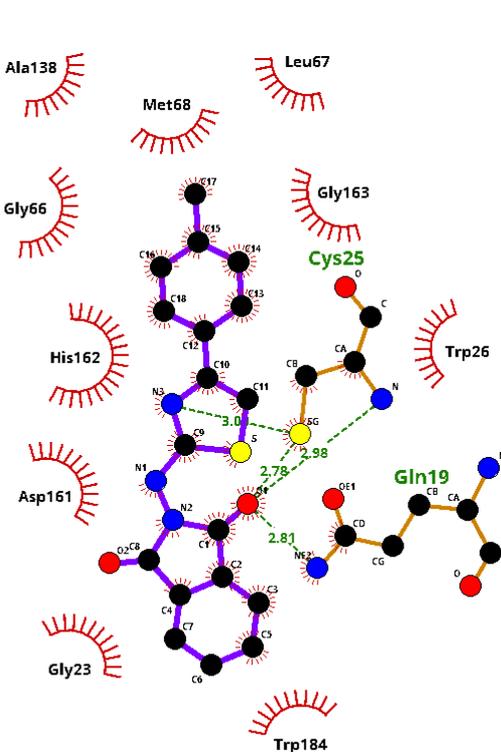
AD\_9t\_3i06



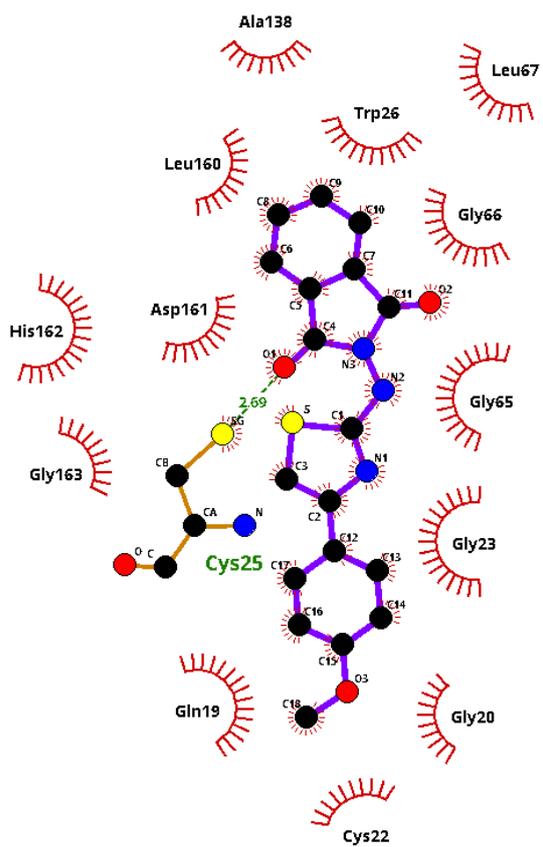
AD\_9v\_3i06



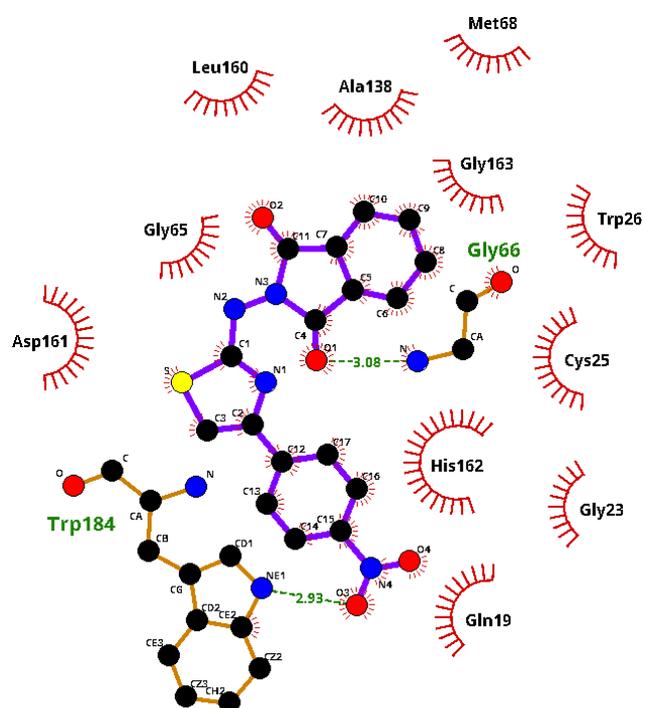
AP\_3i06



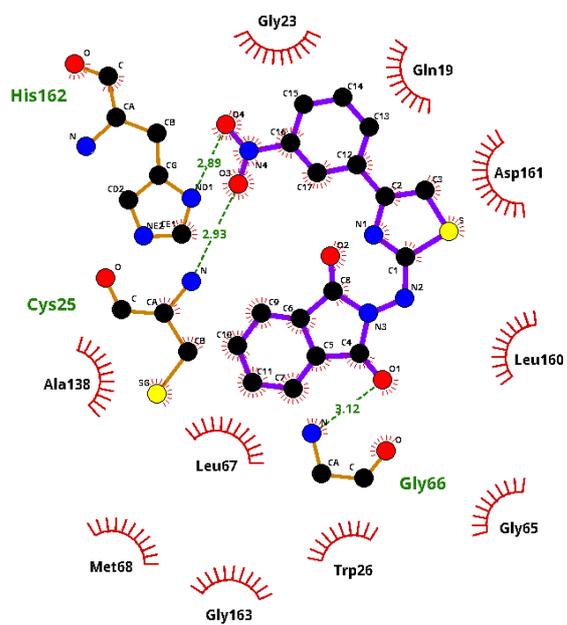
AR\_2c\_3i06



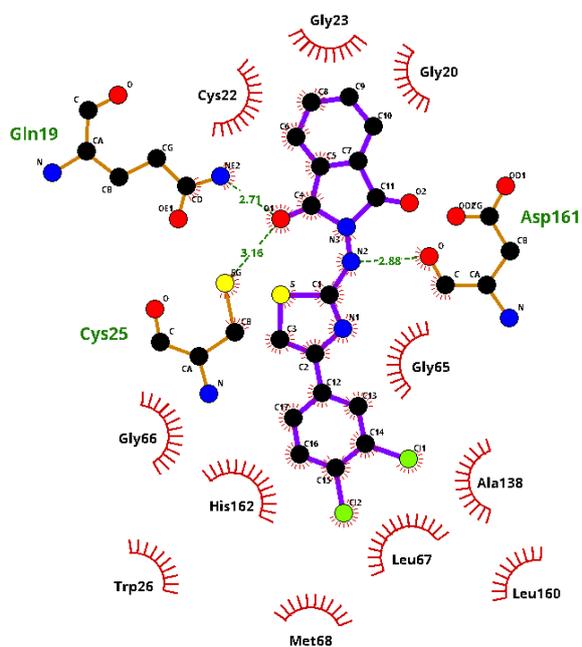
AR\_2d\_3i06



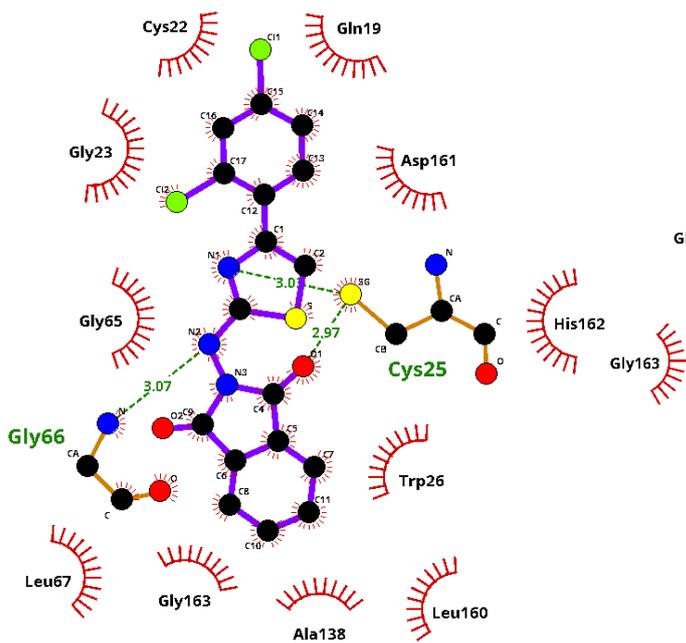
AR\_2h\_3i06



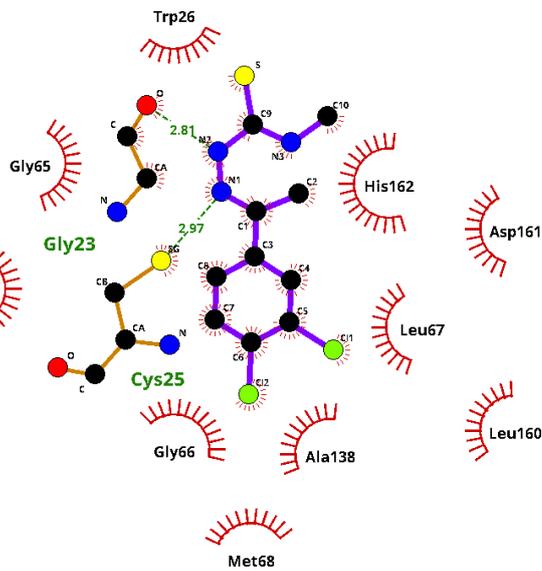
AR\_2k\_3i06



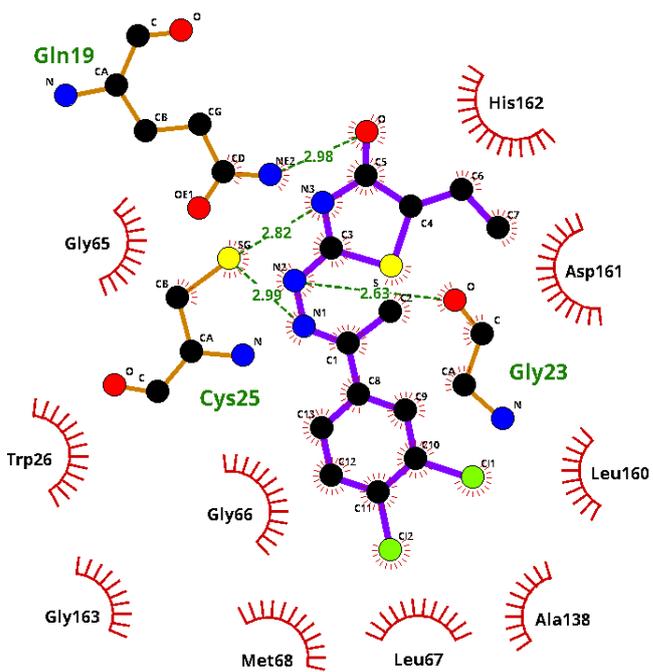
AR\_2l\_3i06



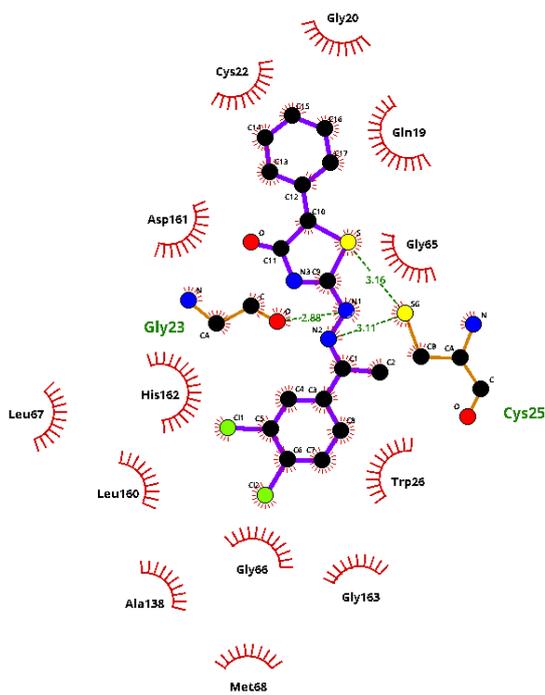
AR\_2m\_3i06



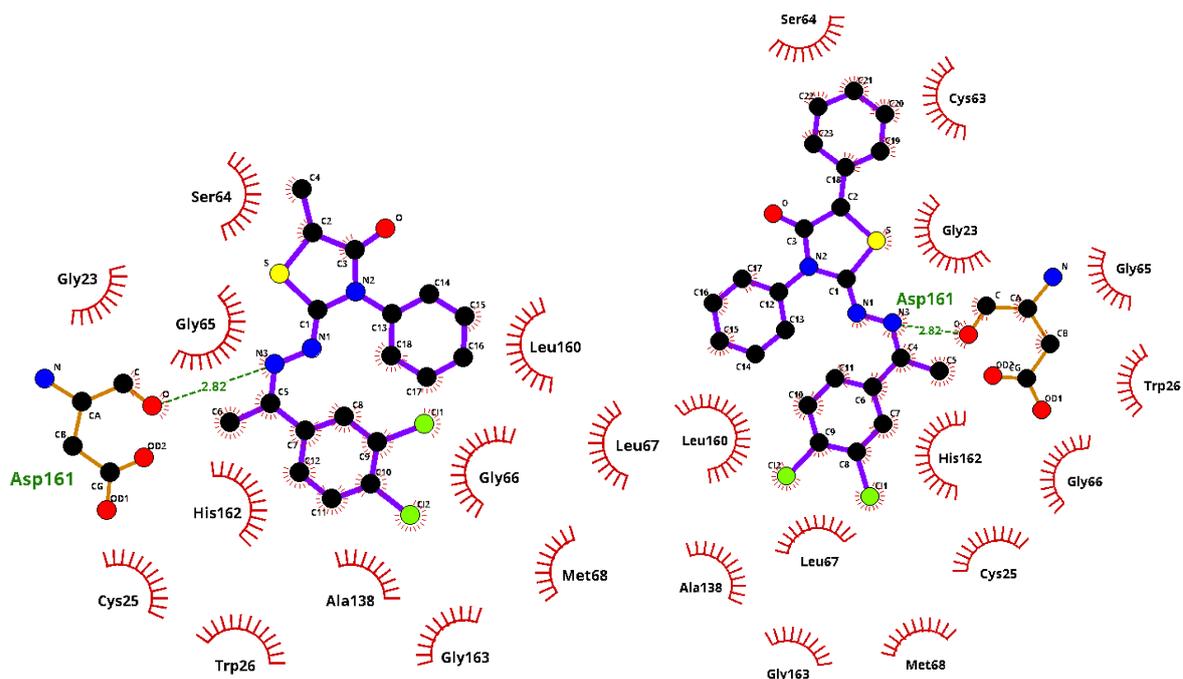
GT\_03\_3i06



GT\_04a\_3i06

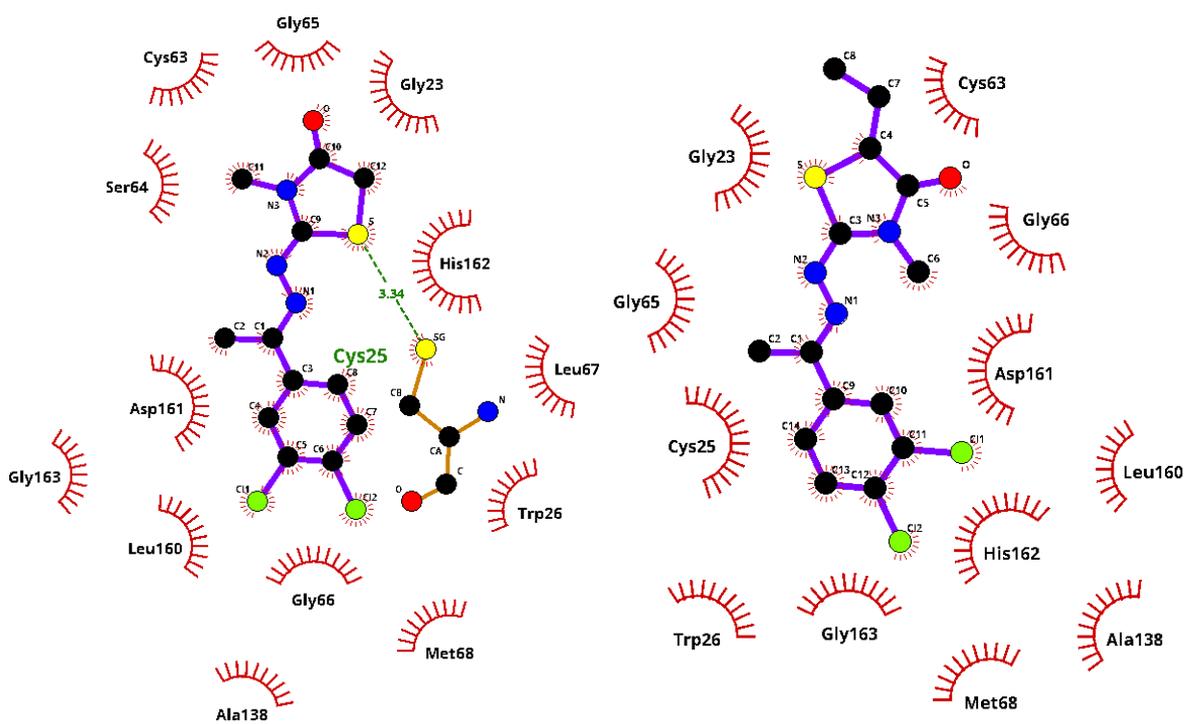


GT\_04d\_3i06



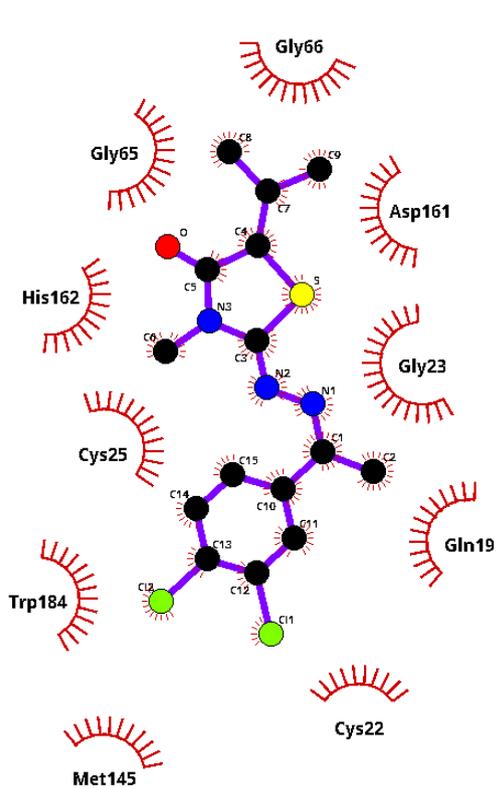
GT\_05b\_3i06

GT\_05d\_3i06

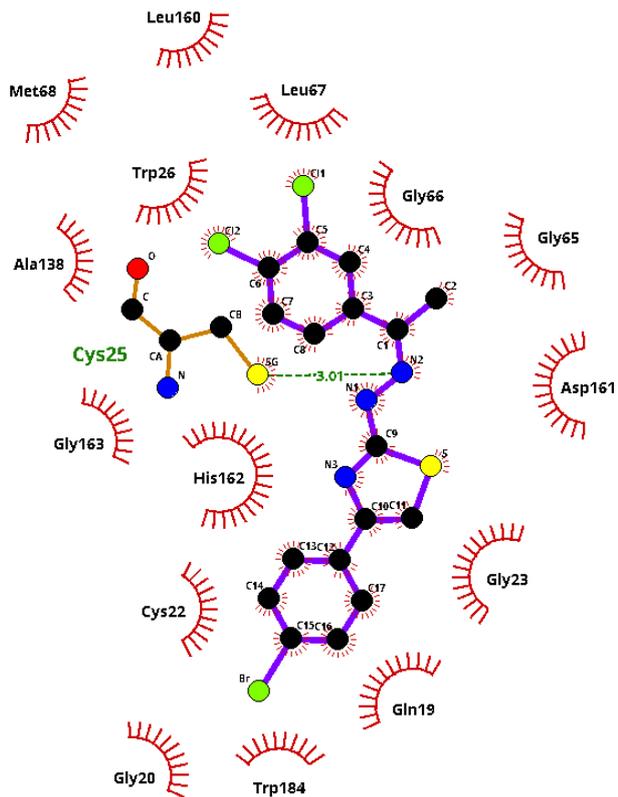


GT\_06\_3i06

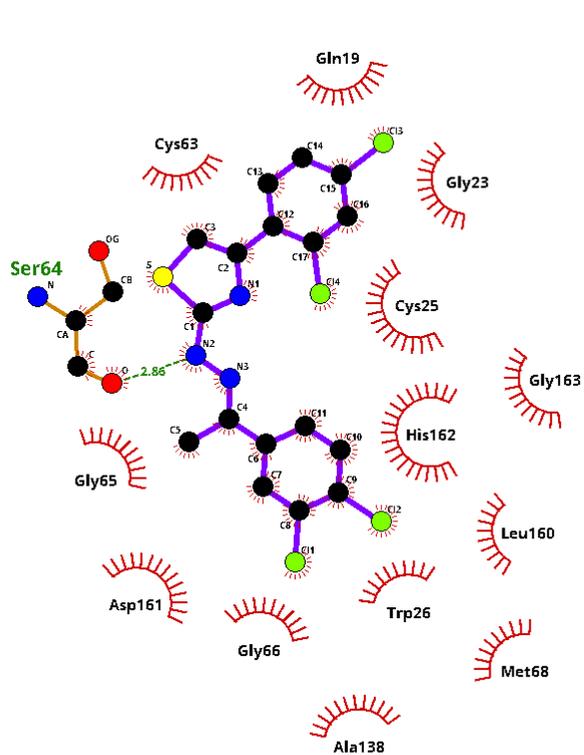
GT\_06a\_3i06



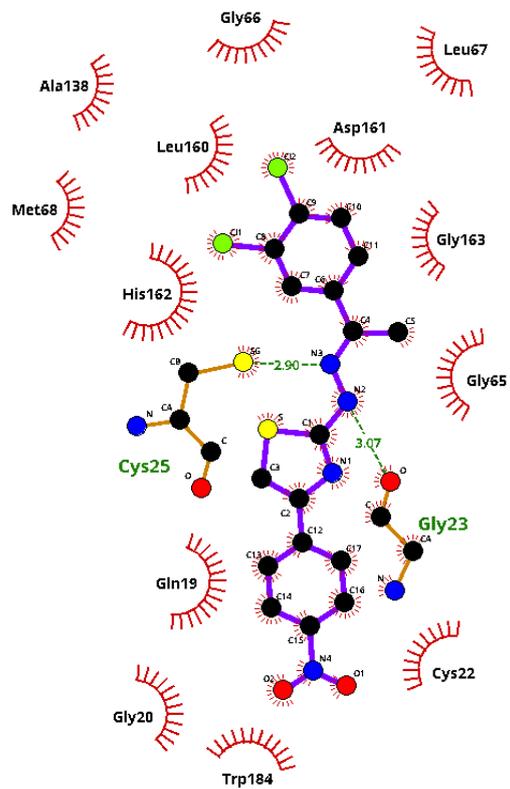
GT\_06c\_3i06



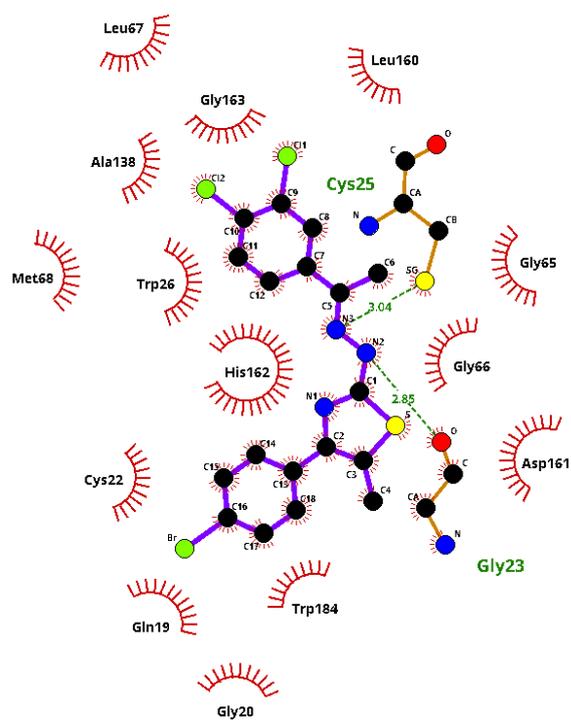
GT\_09\_3i06



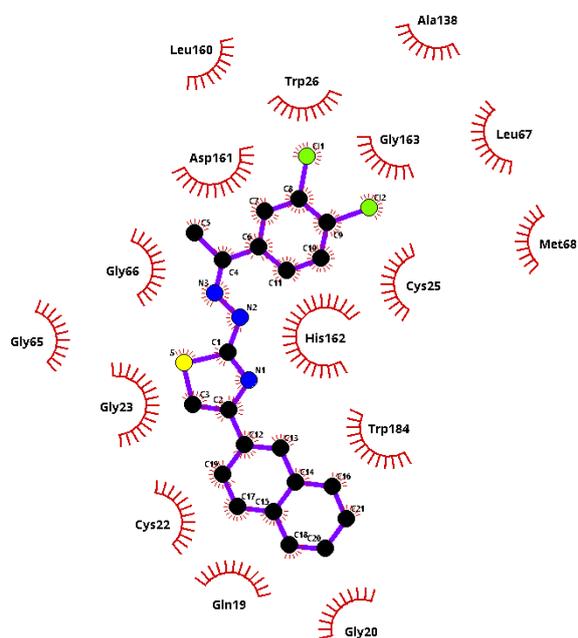
GT\_11\_3i06



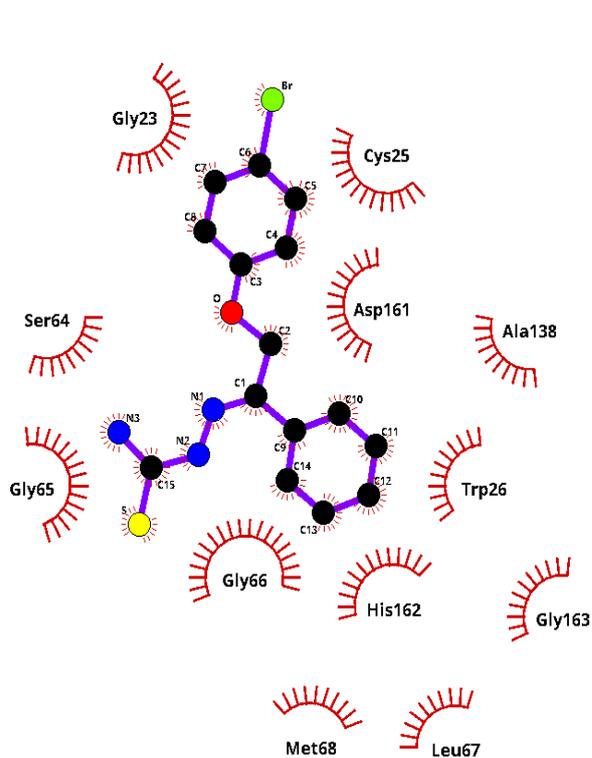
GT\_13\_3i06



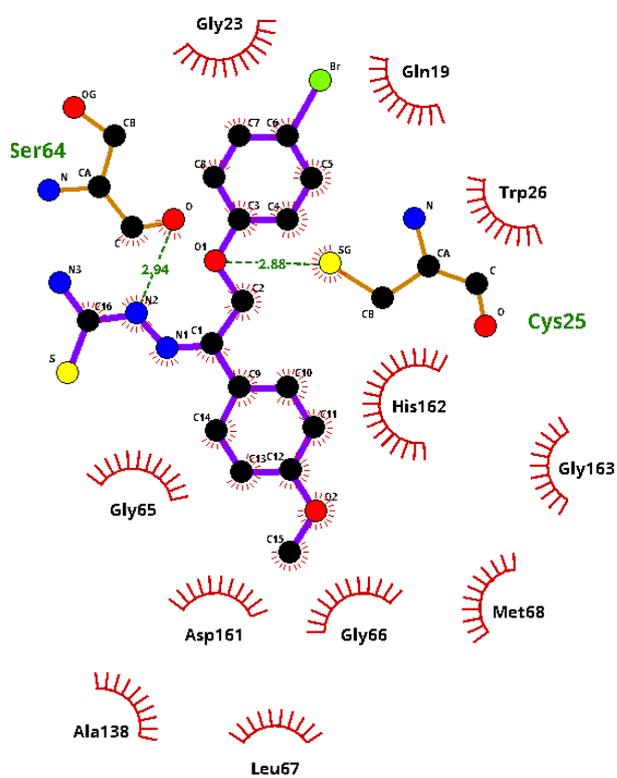
GT\_17\_3i06



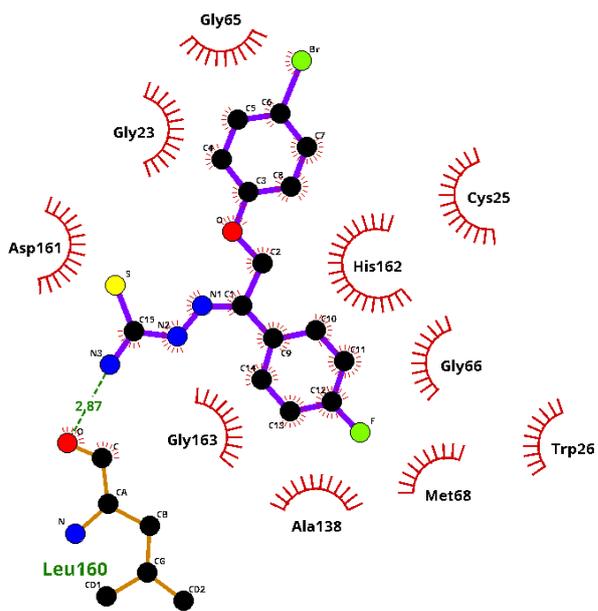
GT\_18\_3i06



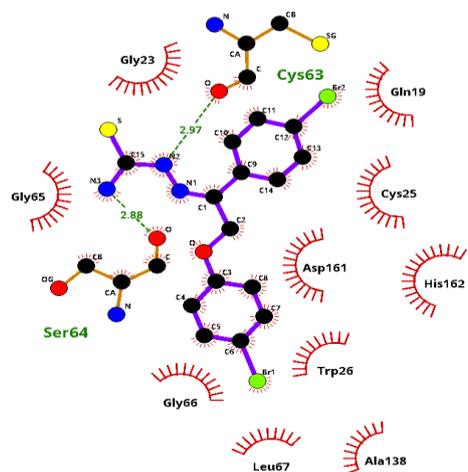
JW\_8a\_3i06



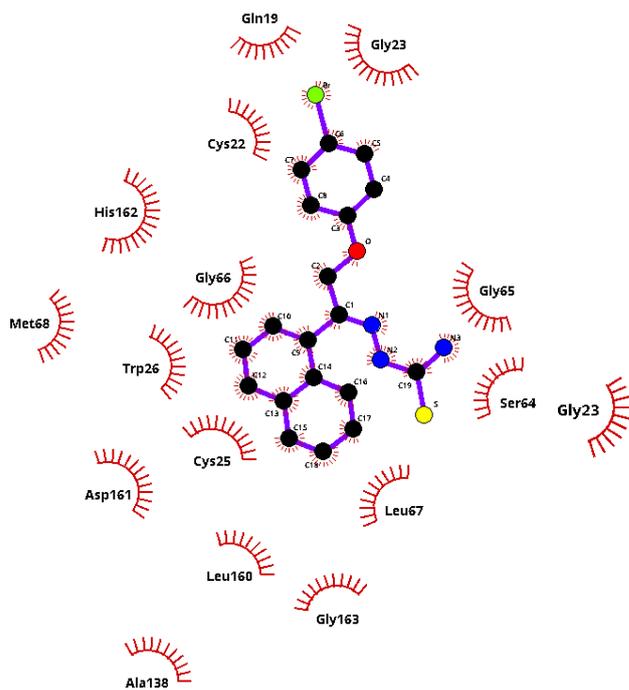
JW\_8c\_3i06



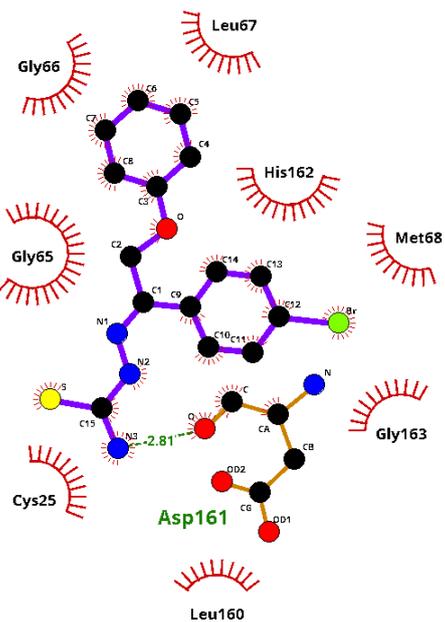
JW\_8d\_3i06



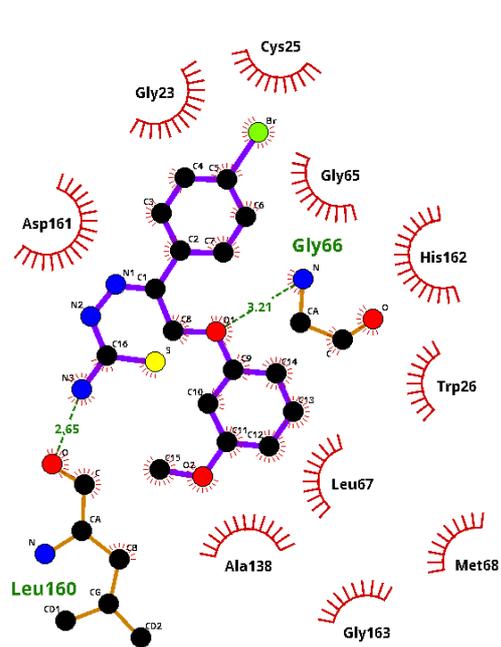
JW\_8f\_3i06



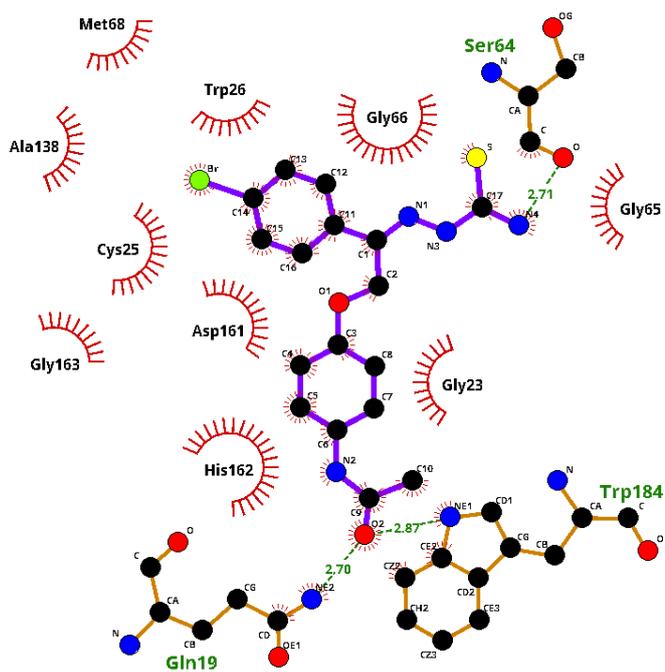
JW\_8h\_3i06



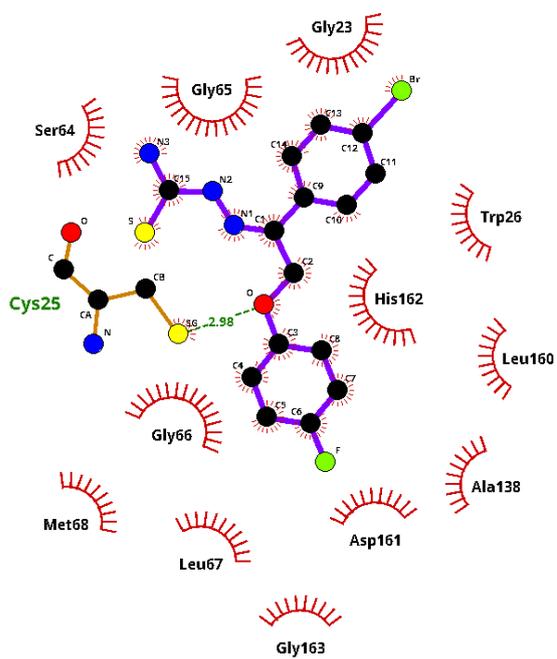
JW\_9a\_3i06



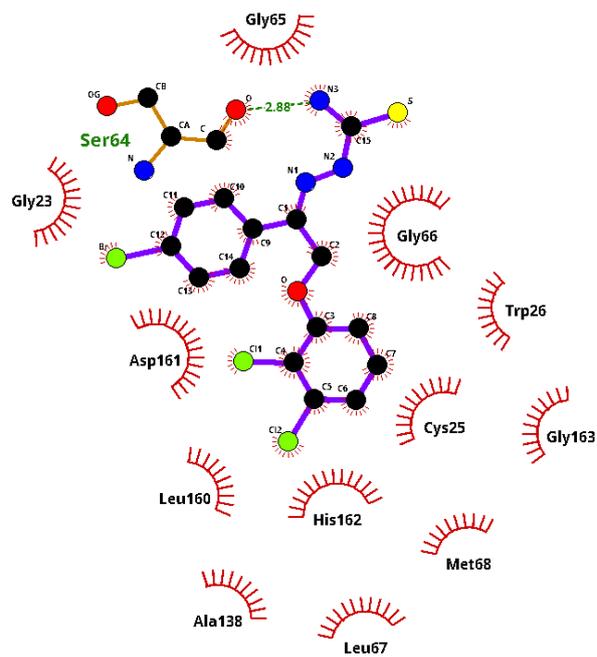
JW\_9e\_3i06



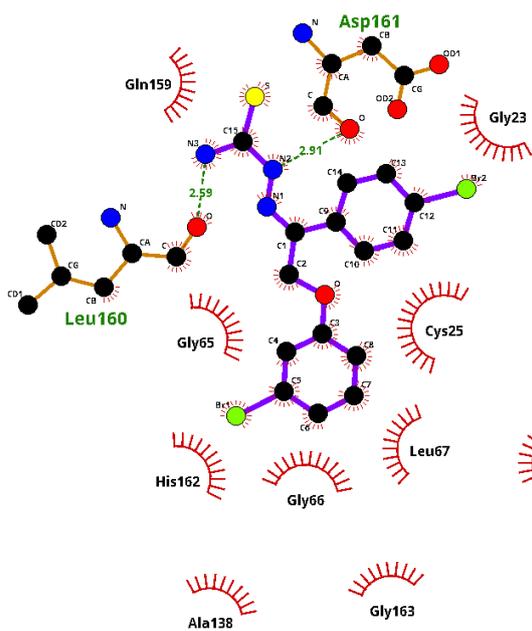
JW\_9e\_3i06



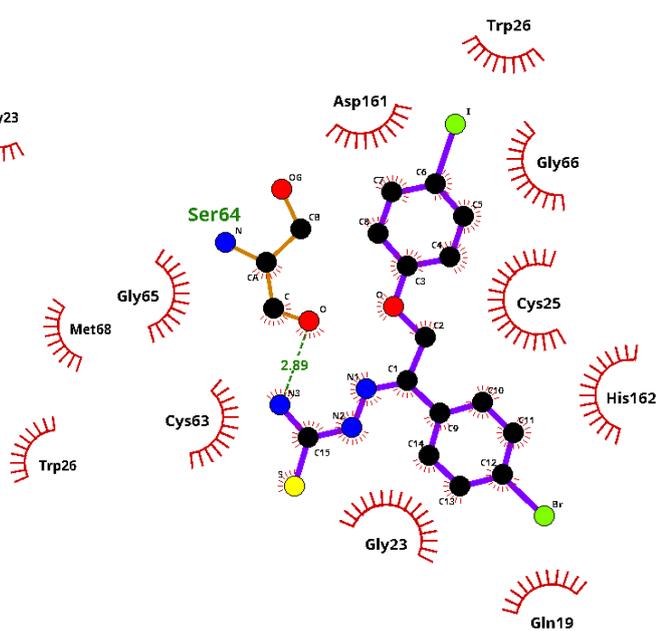
JW\_9h\_3i06



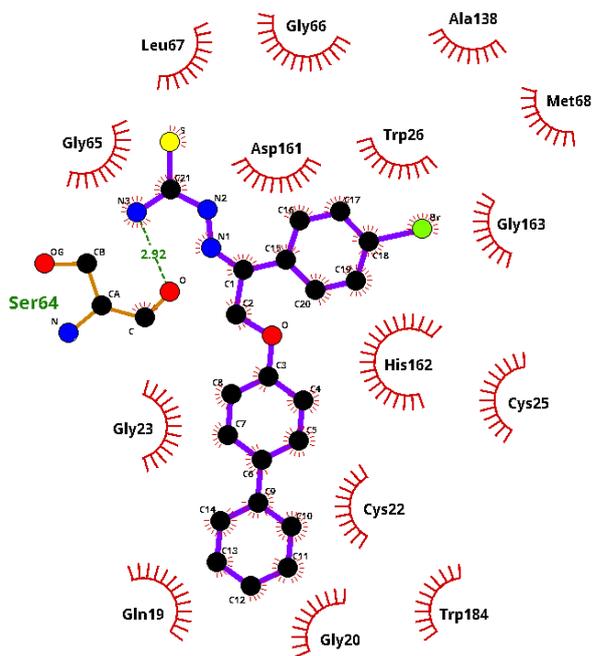
JW\_9m\_3i06



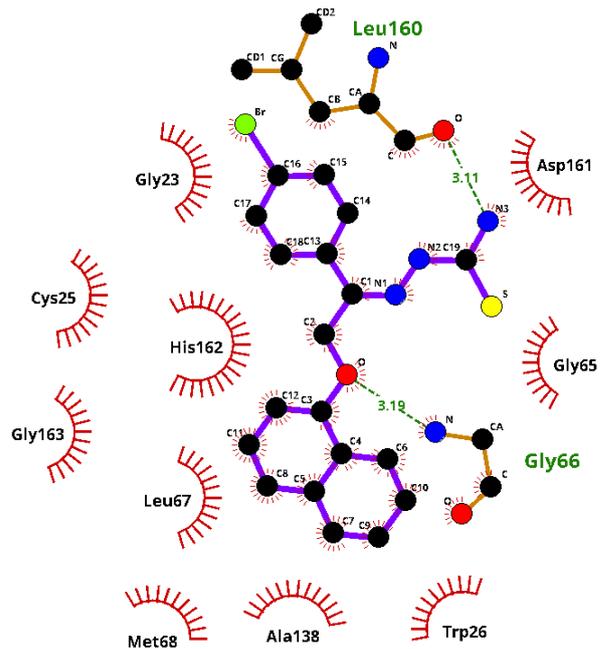
JW\_9o\_3i06



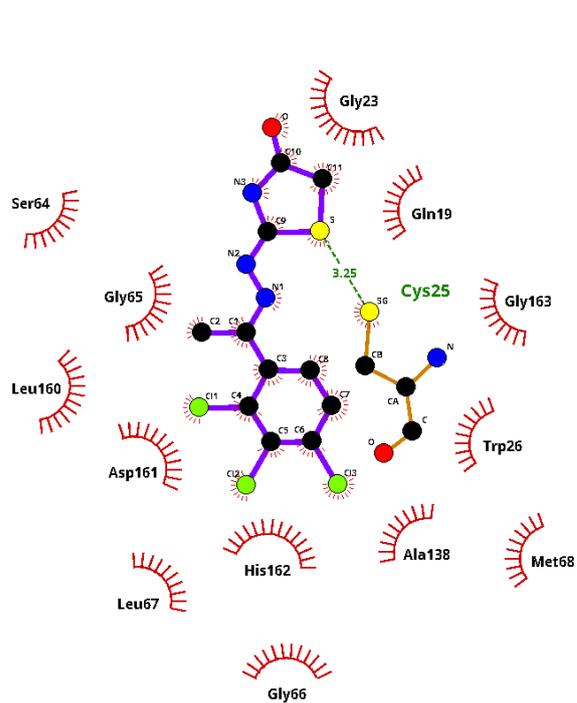
JW\_9p\_3i06



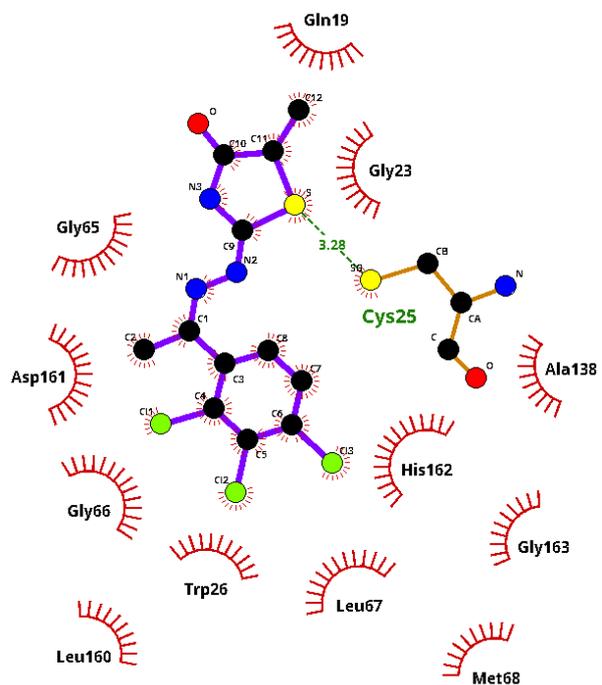
JW\_9r\_3i06



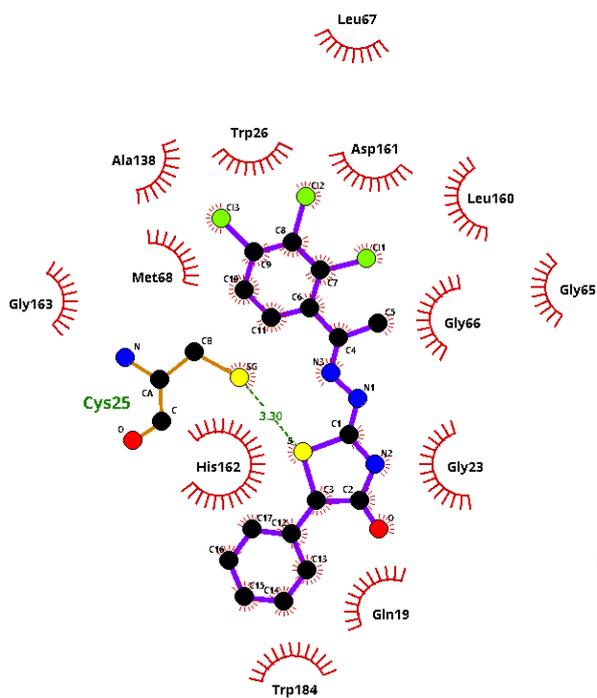
JW\_9s\_3i06



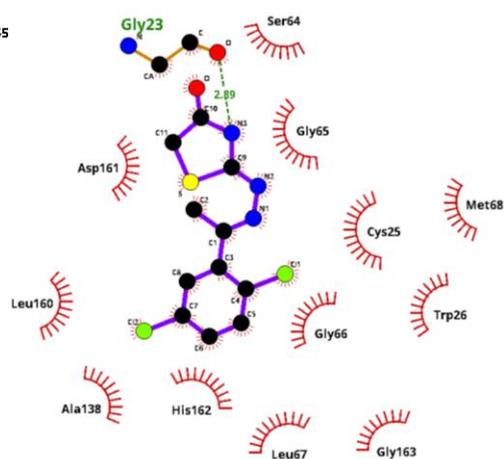
LA\_1a1\_3i06



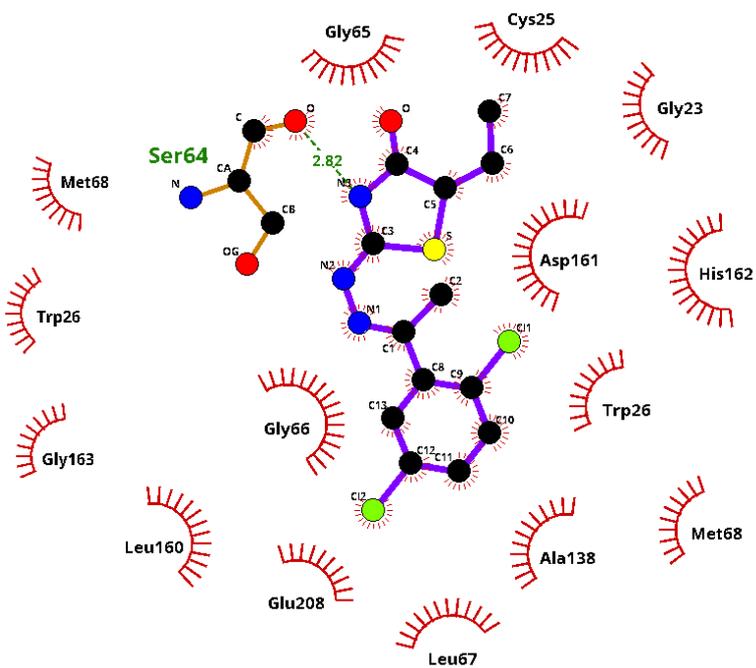
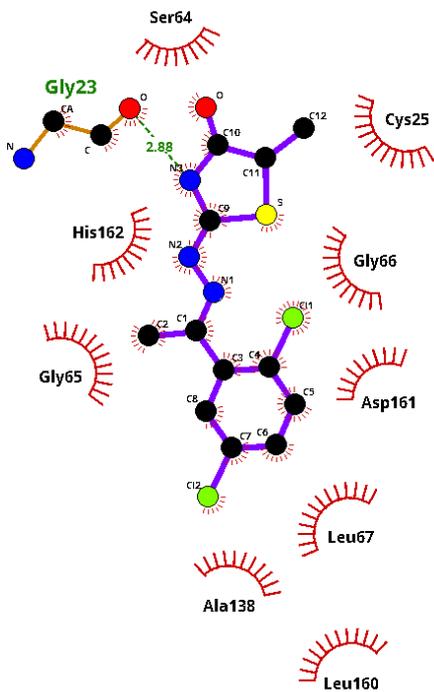
LA\_1a2\_3i06



LA\_1a5\_3i06

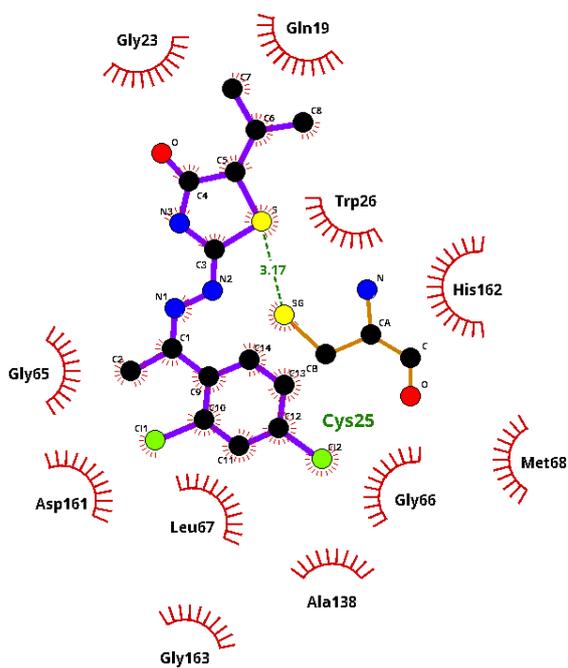
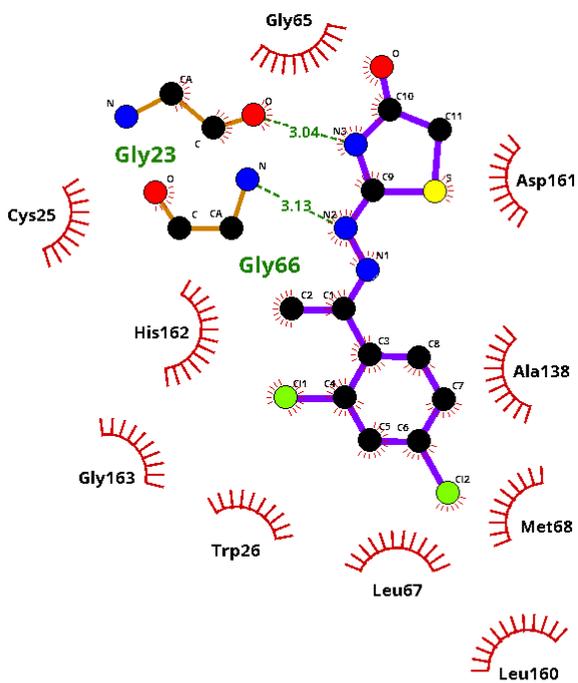


LA\_1b1\_3i06



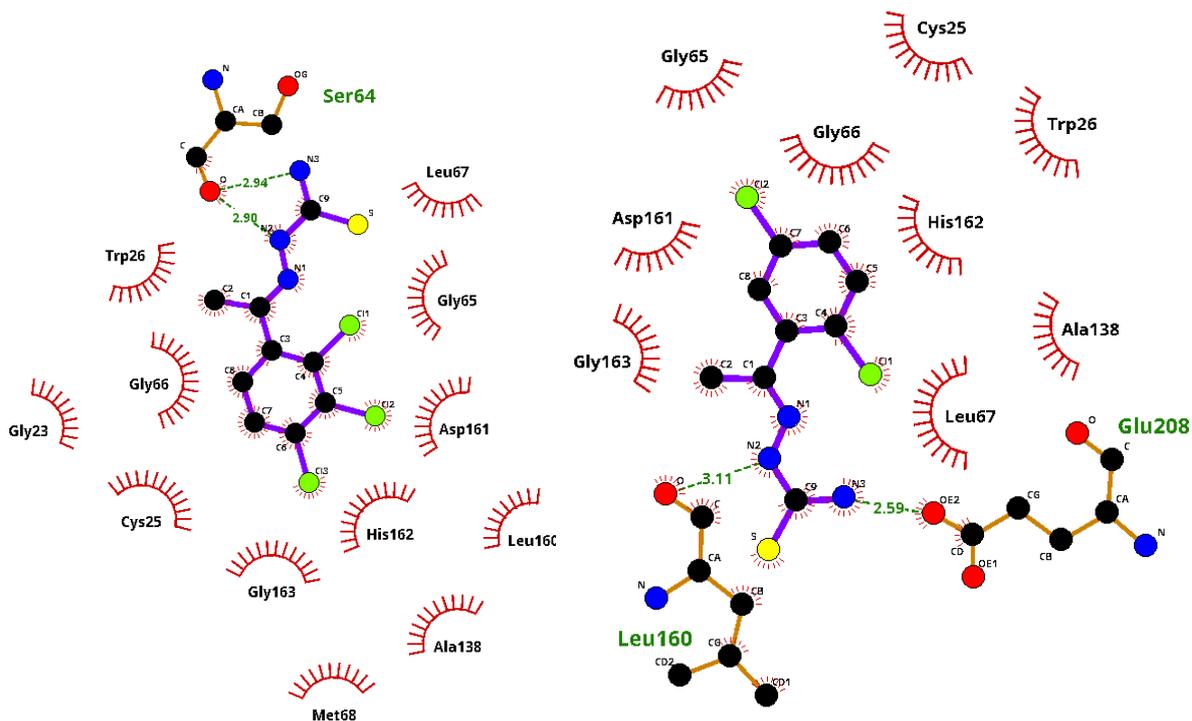
LA\_1b2\_3i06

LA\_1b3\_3i06



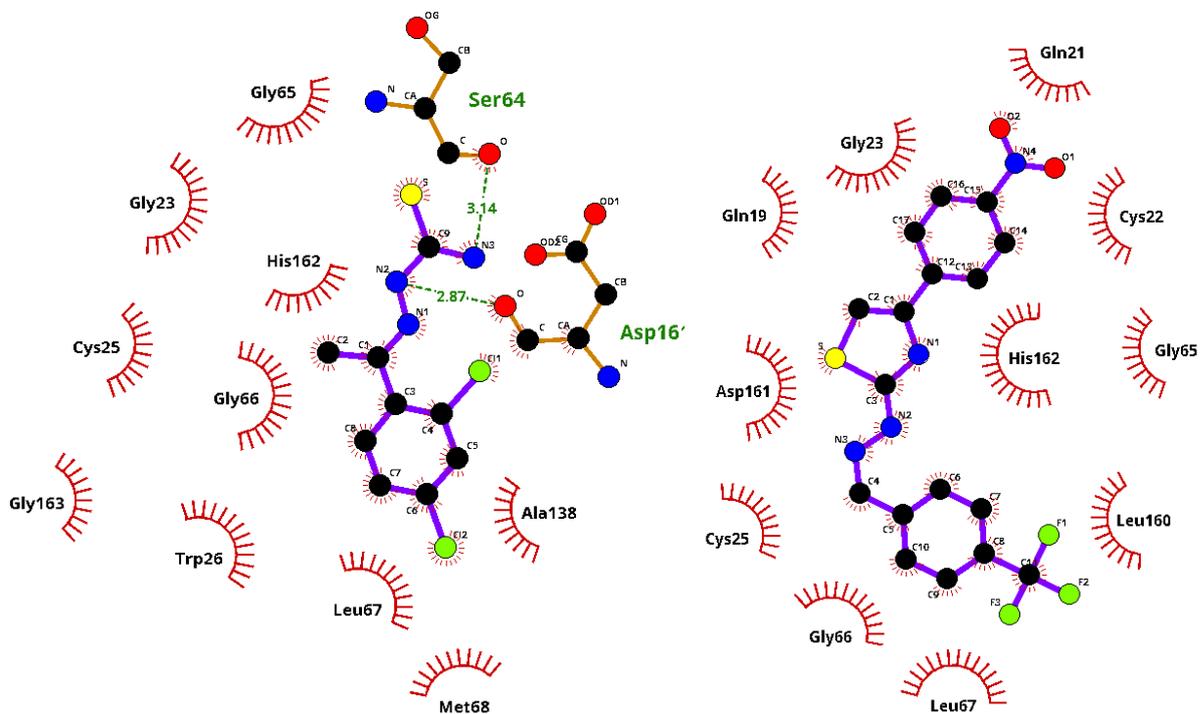
LA\_1c1\_3i06

LA\_1c4\_3i06



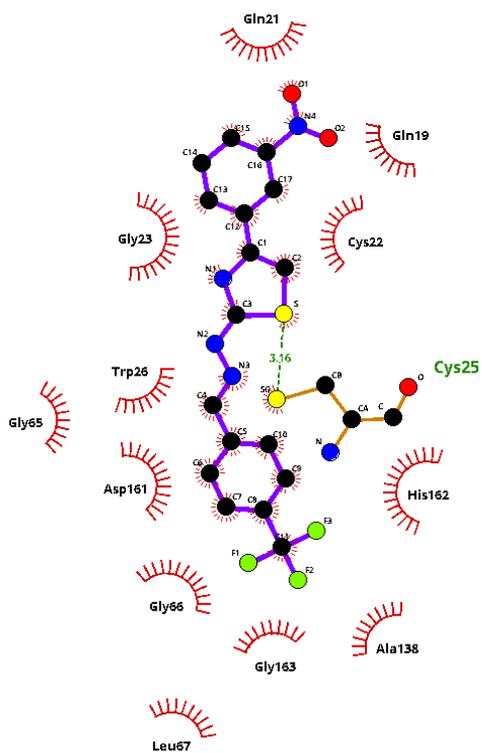
LA INT 1 3i06

LA INT 2 3i06

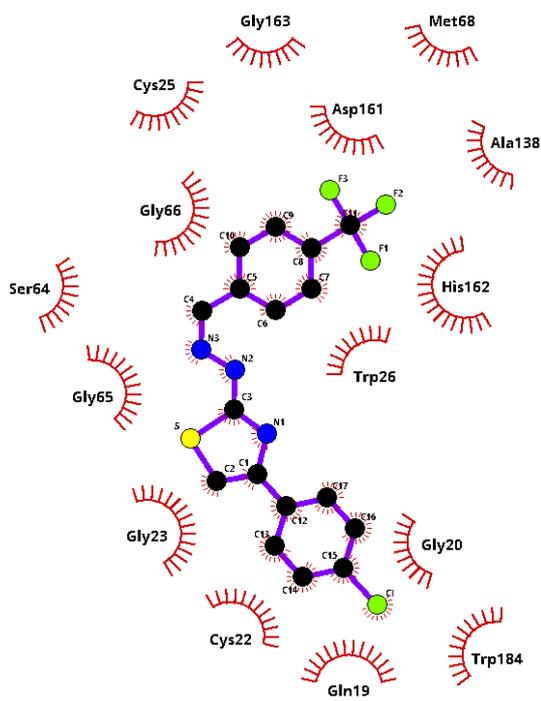


LA\_INT\_3\_3i06

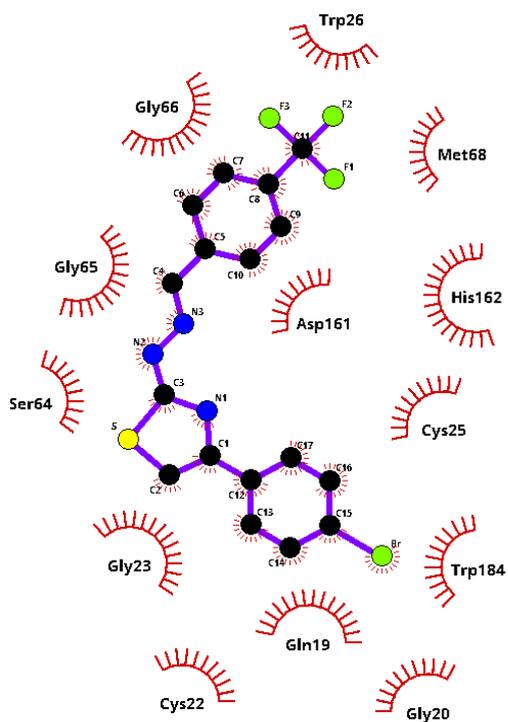
MB1\_2\_3i06



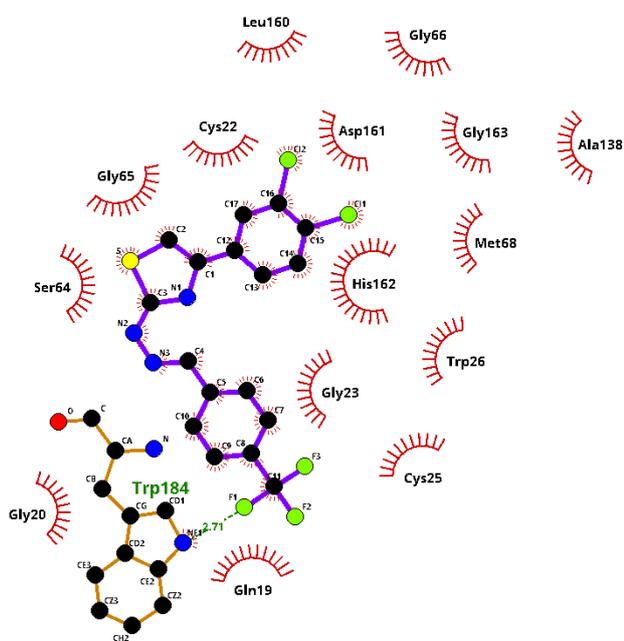
MB1\_3\_3i06



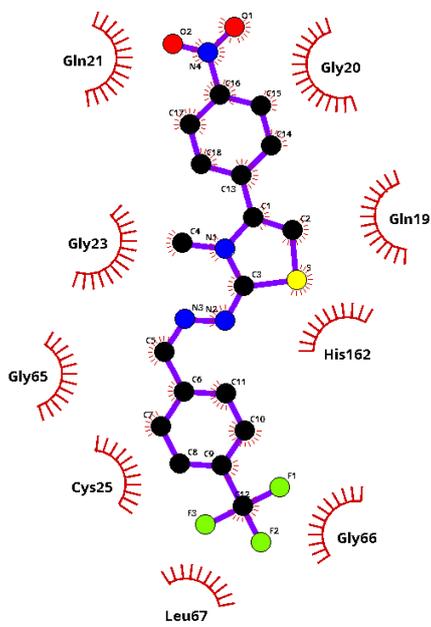
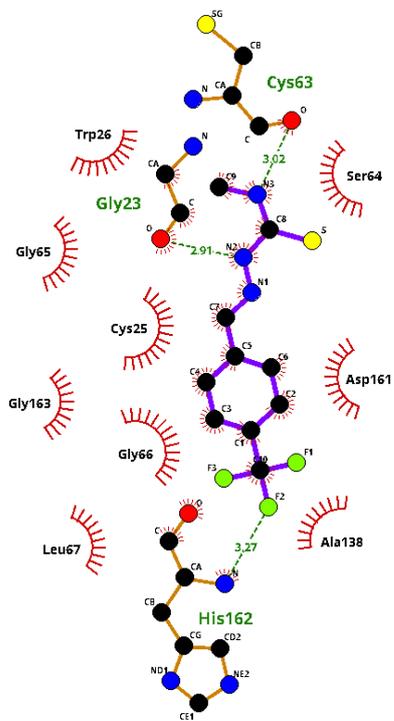
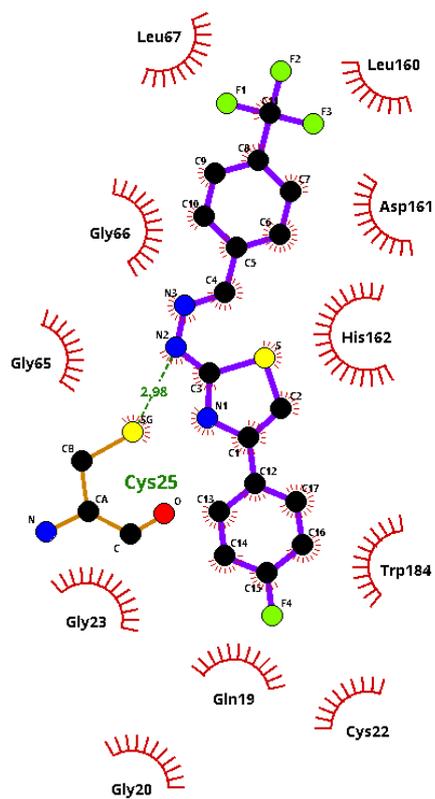
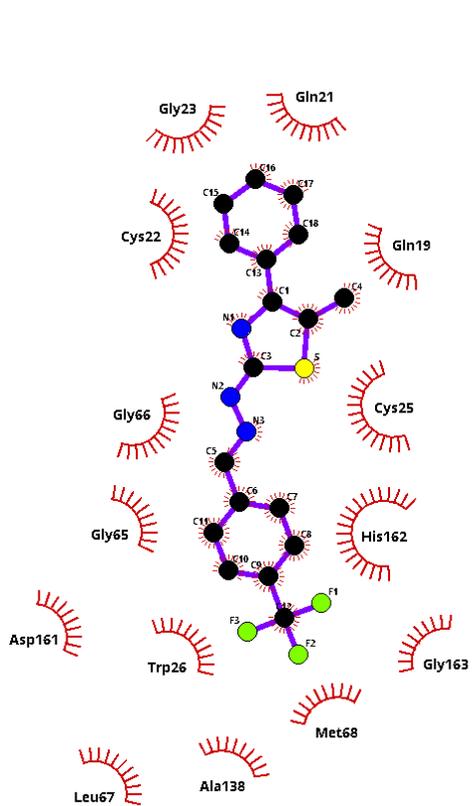
MB1\_6\_3i06

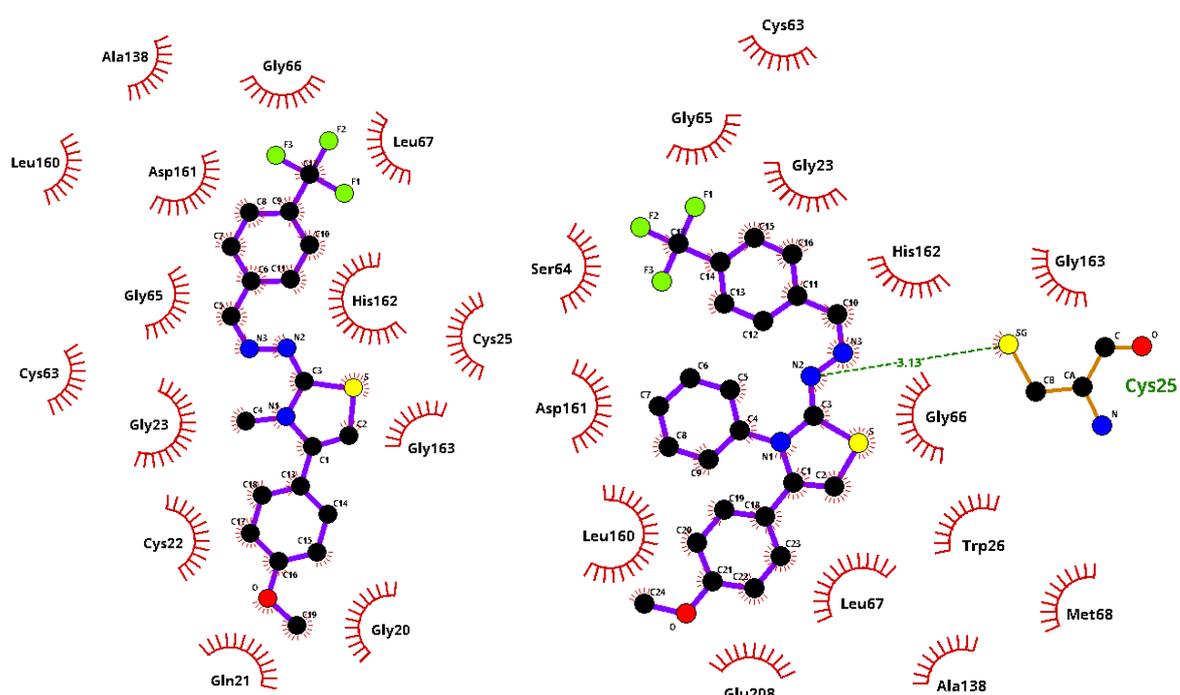


MB1\_7\_3i06



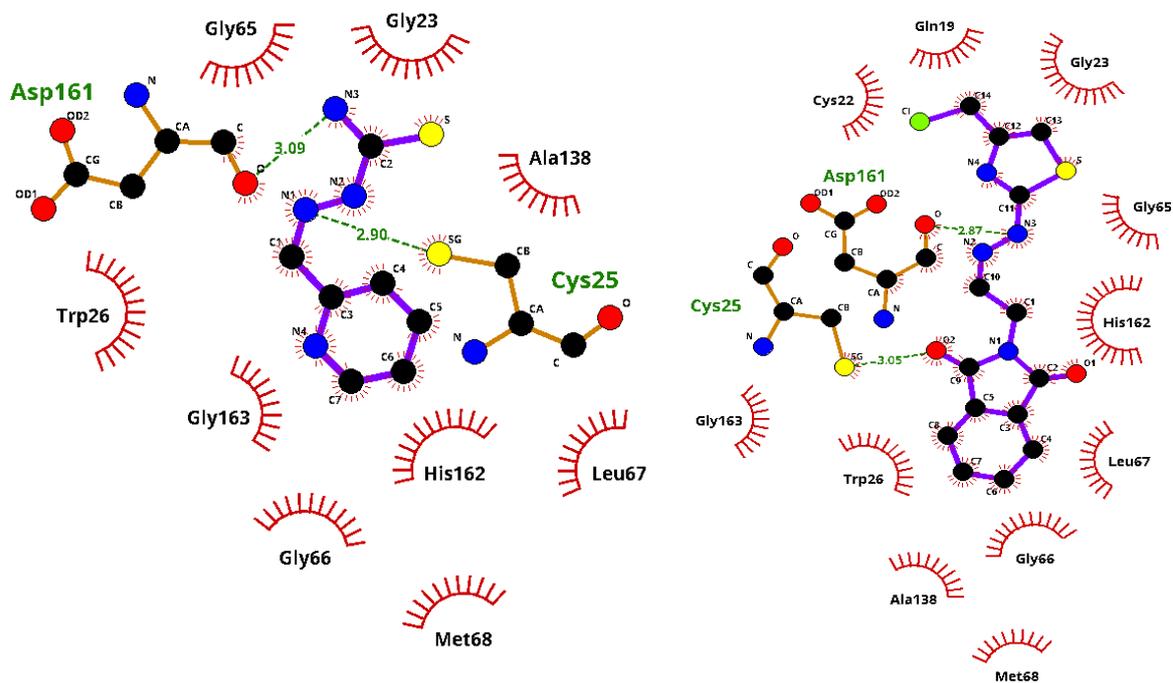
MB1\_8\_3i06





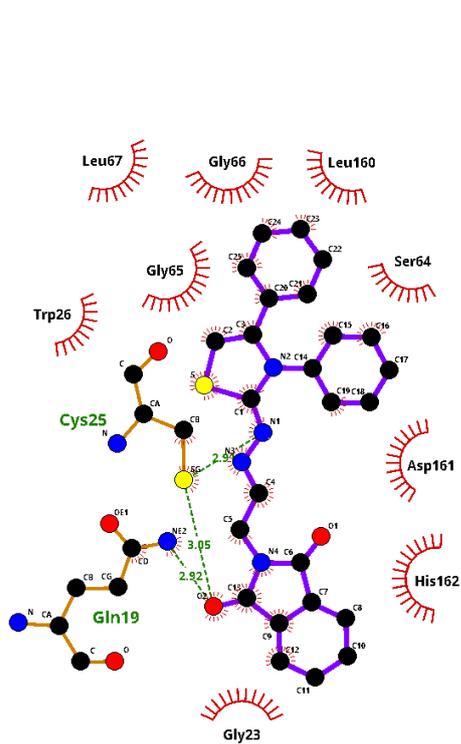
MB2\_2\_3i06

MB3\_5\_3i06

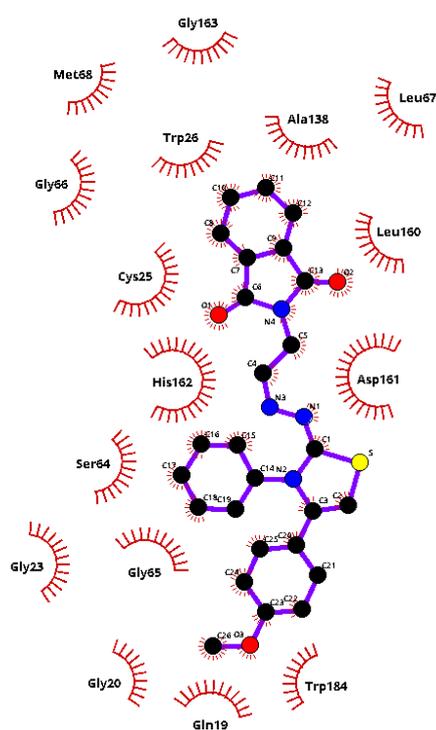


P\_3i06

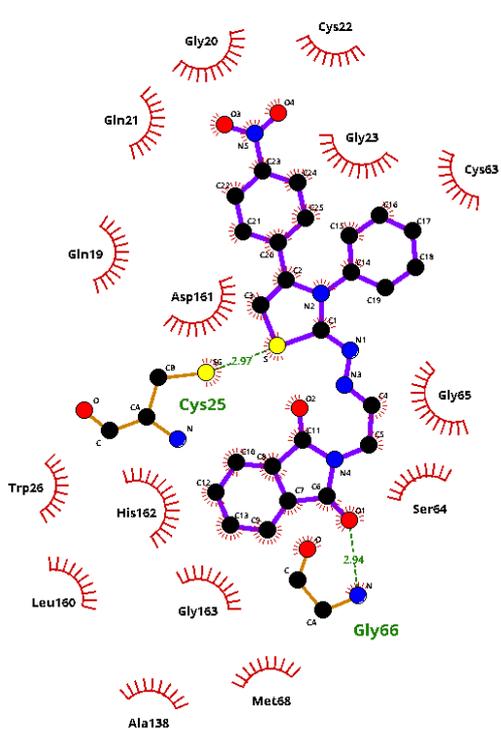
PA\_6a\_3i06



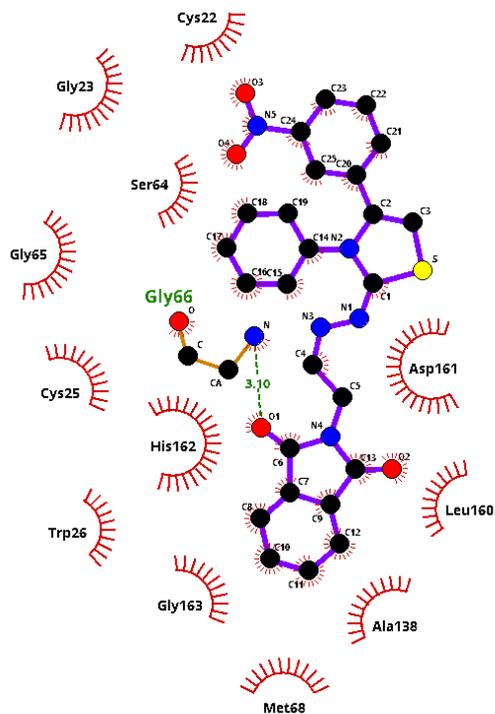
PA\_6b\_3i06



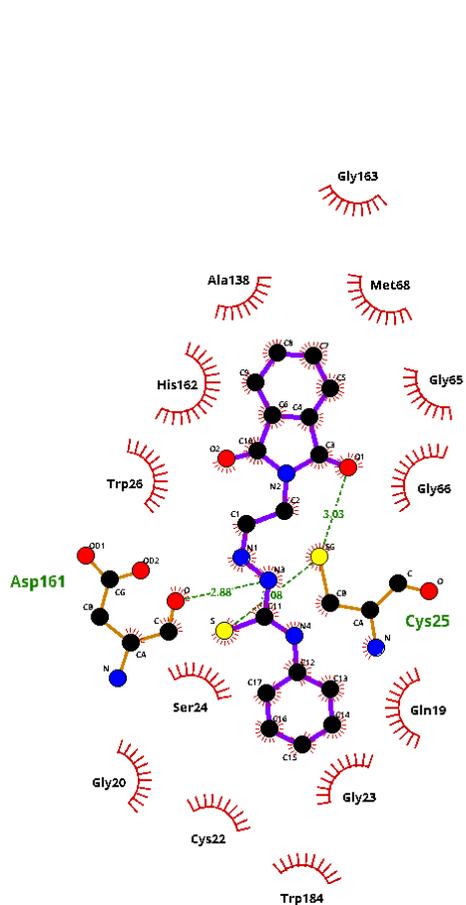
PA\_6d\_3i06



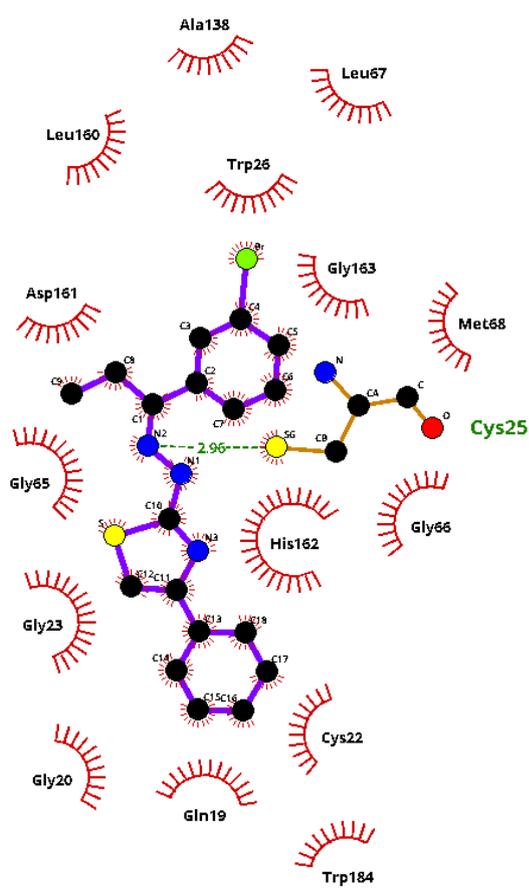
PA\_6h\_3i06



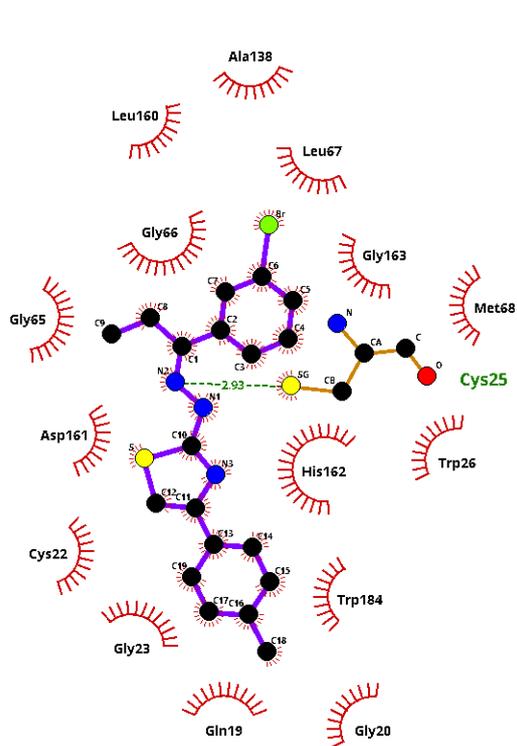
PA\_6k\_3i06



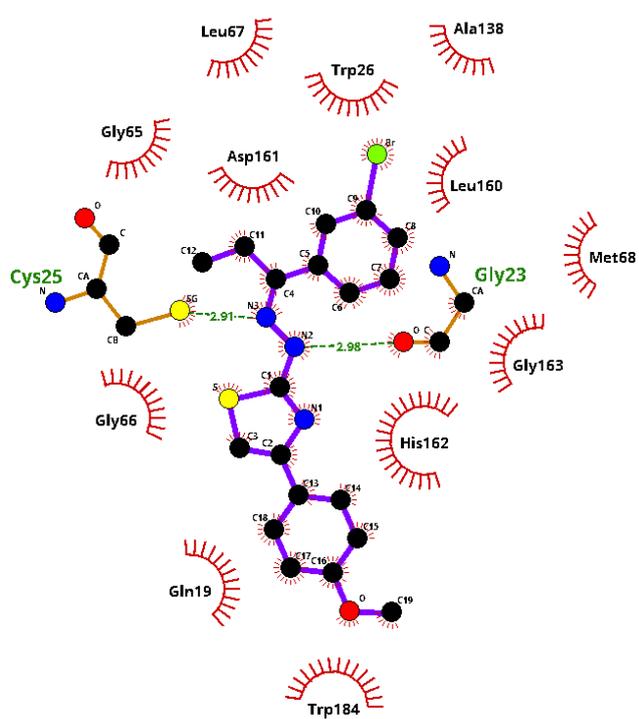
PM\_INT\_3b\_3i06



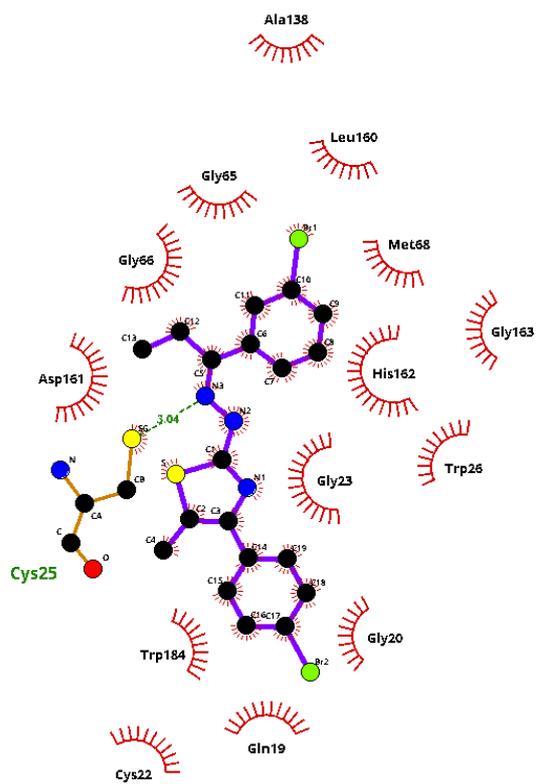
PT01\_2\_3i06



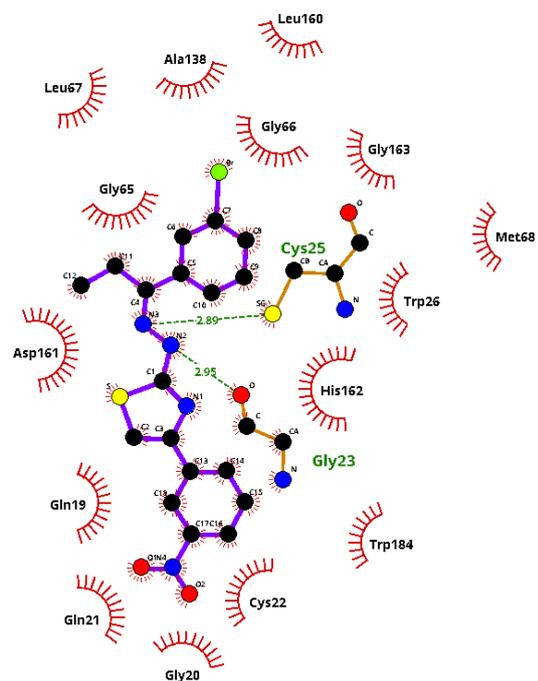
PT01\_3\_3i06



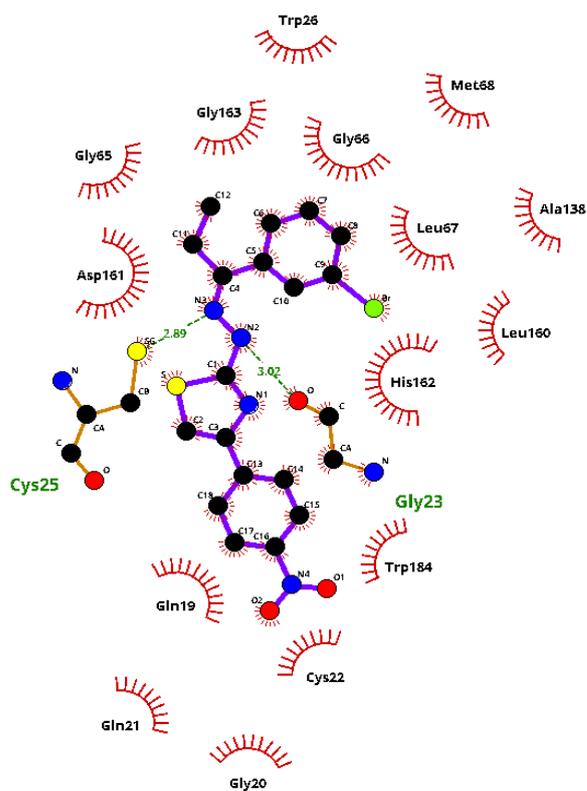
PT01\_4\_3i06



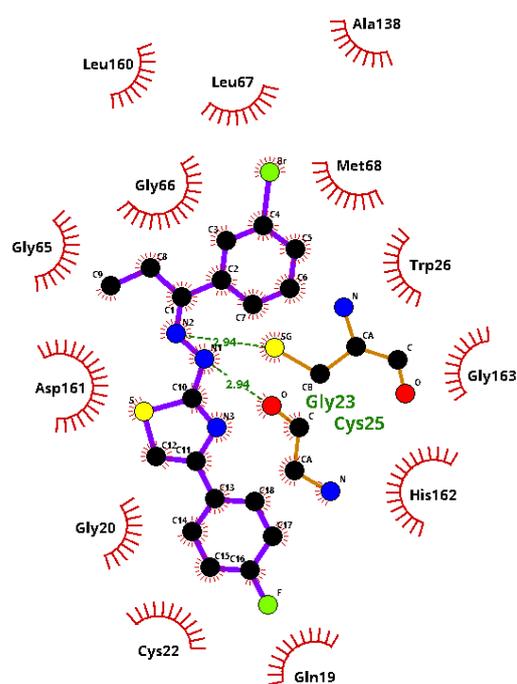
PT01\_5\_3i06



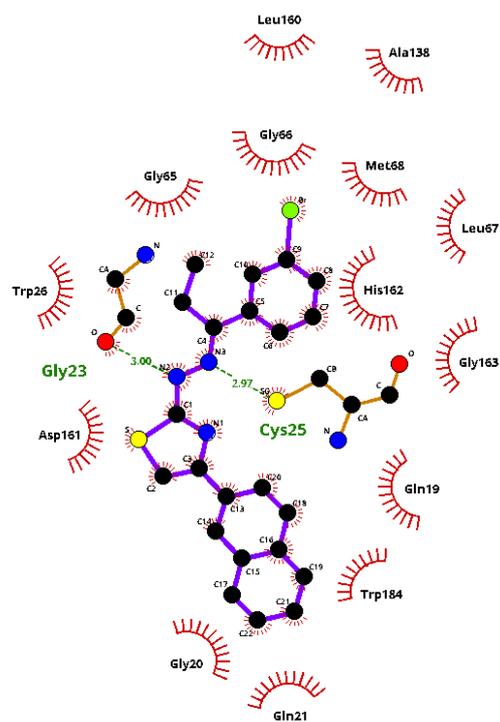
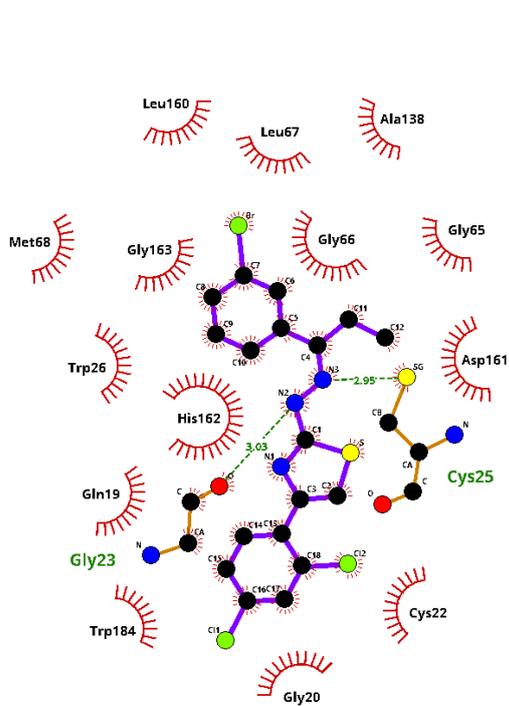
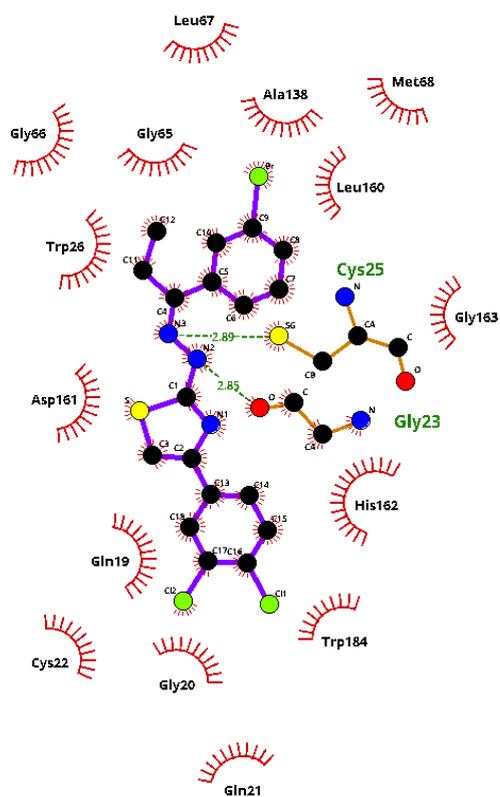
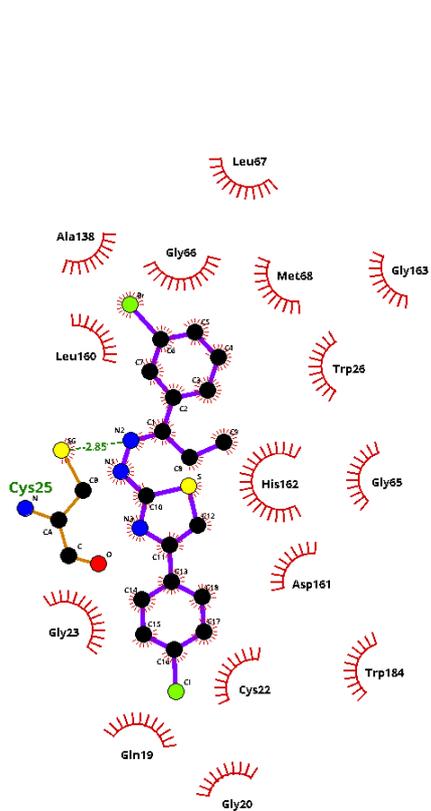
PT01\_6\_3i06

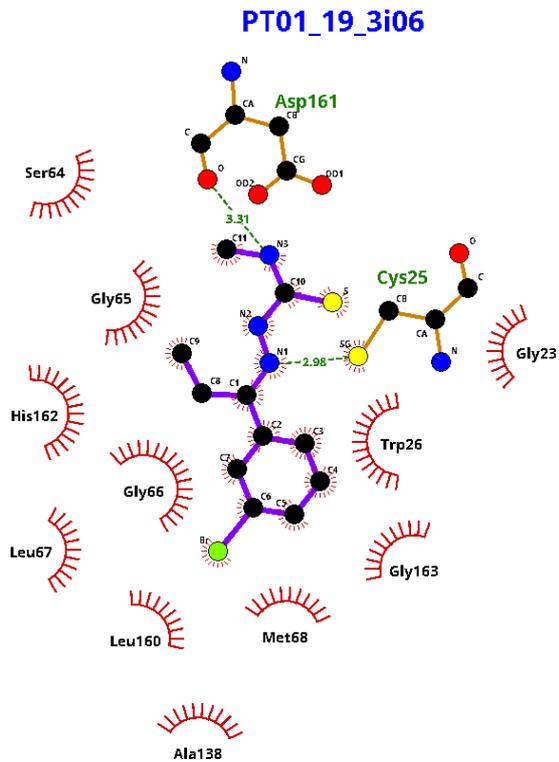
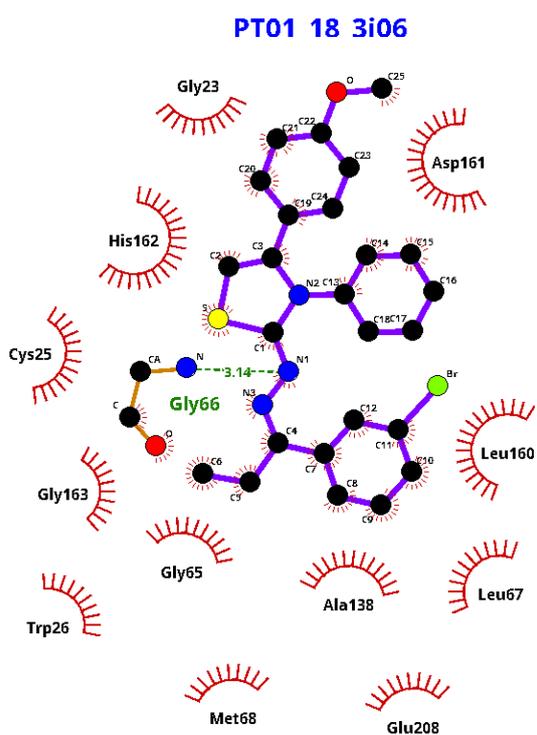
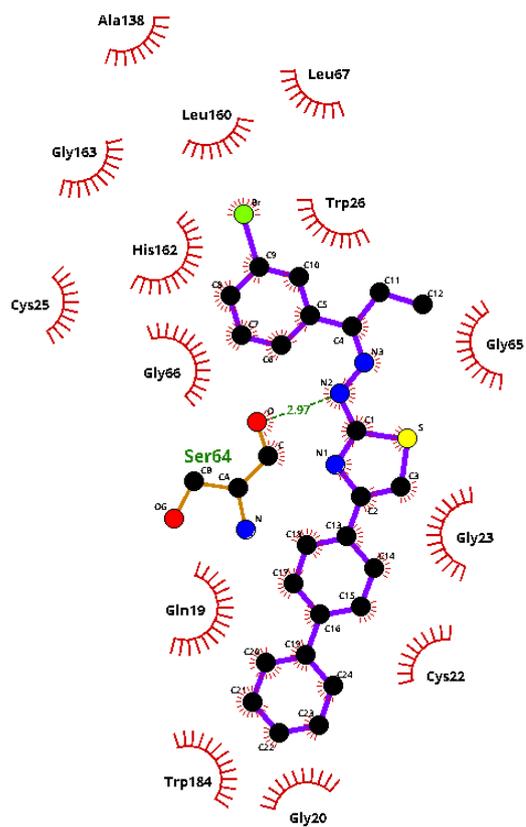
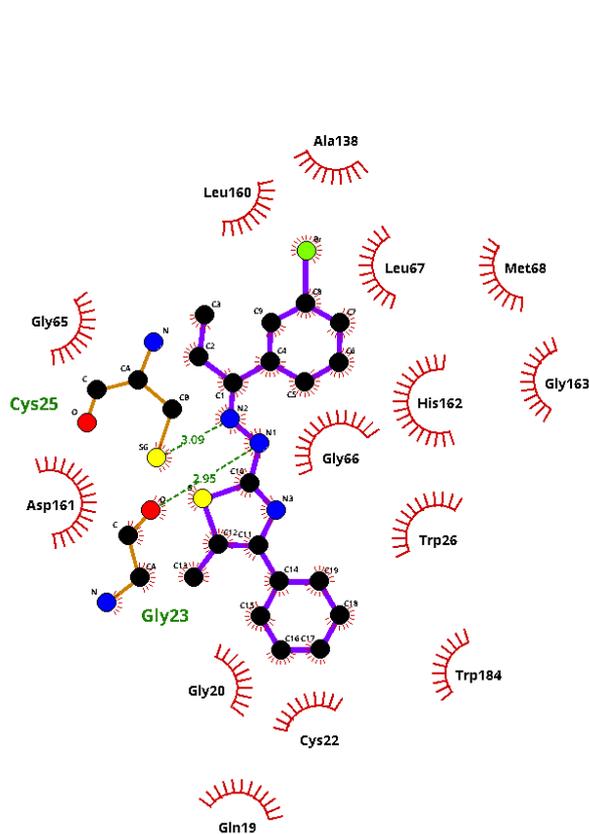


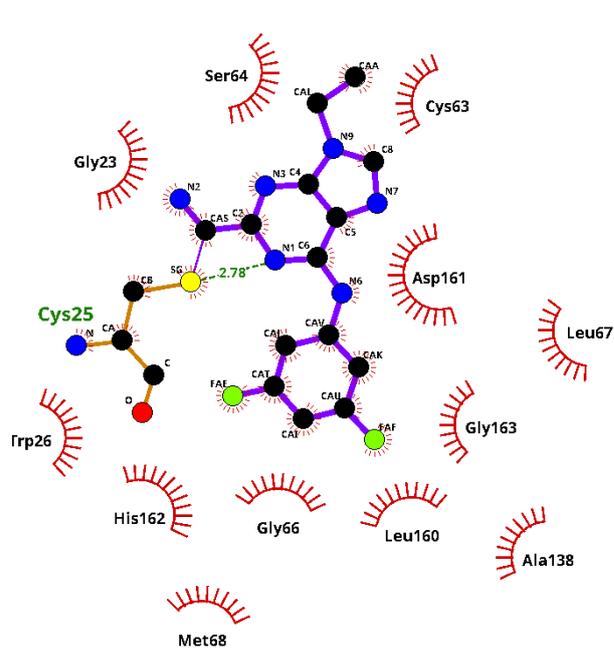
PT01\_7\_3i06



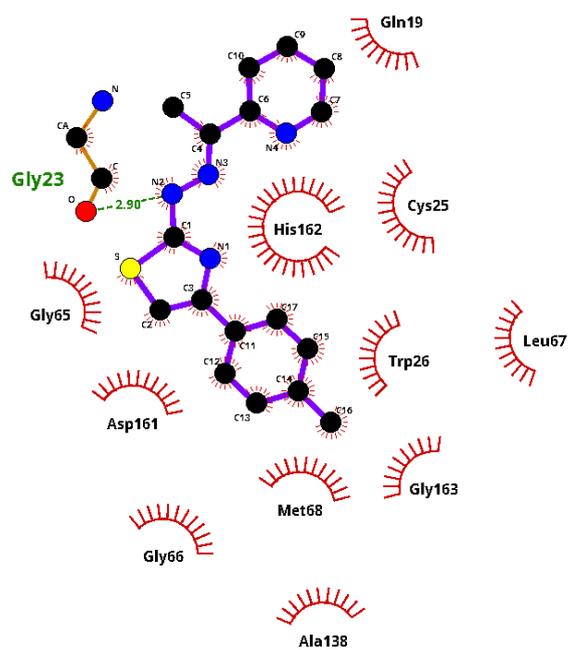
PT01\_8\_3i06



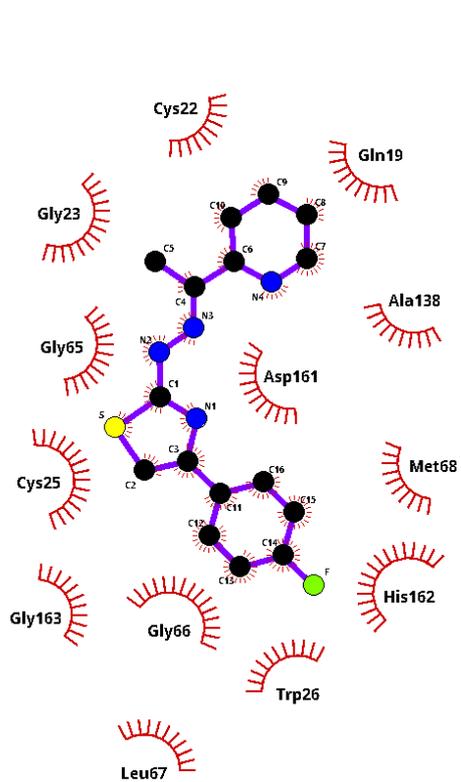




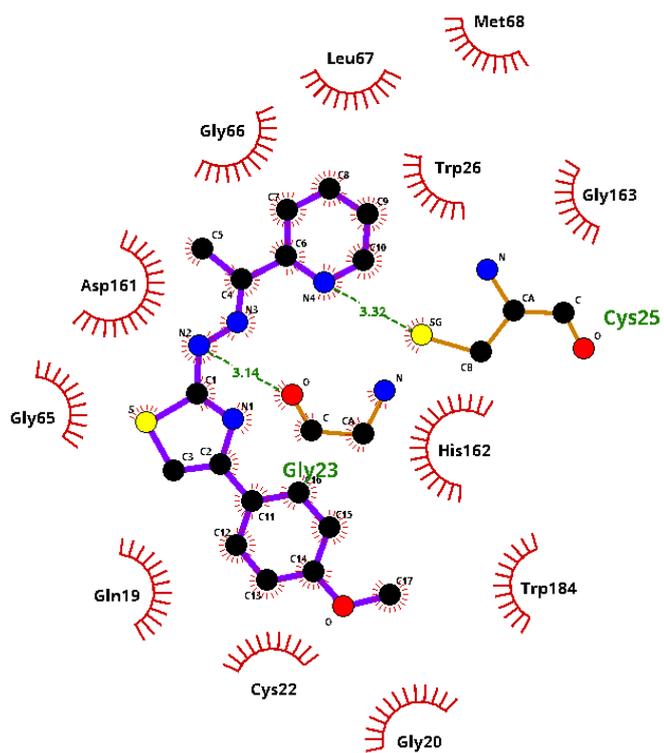
3i06\_QL2\_3i06



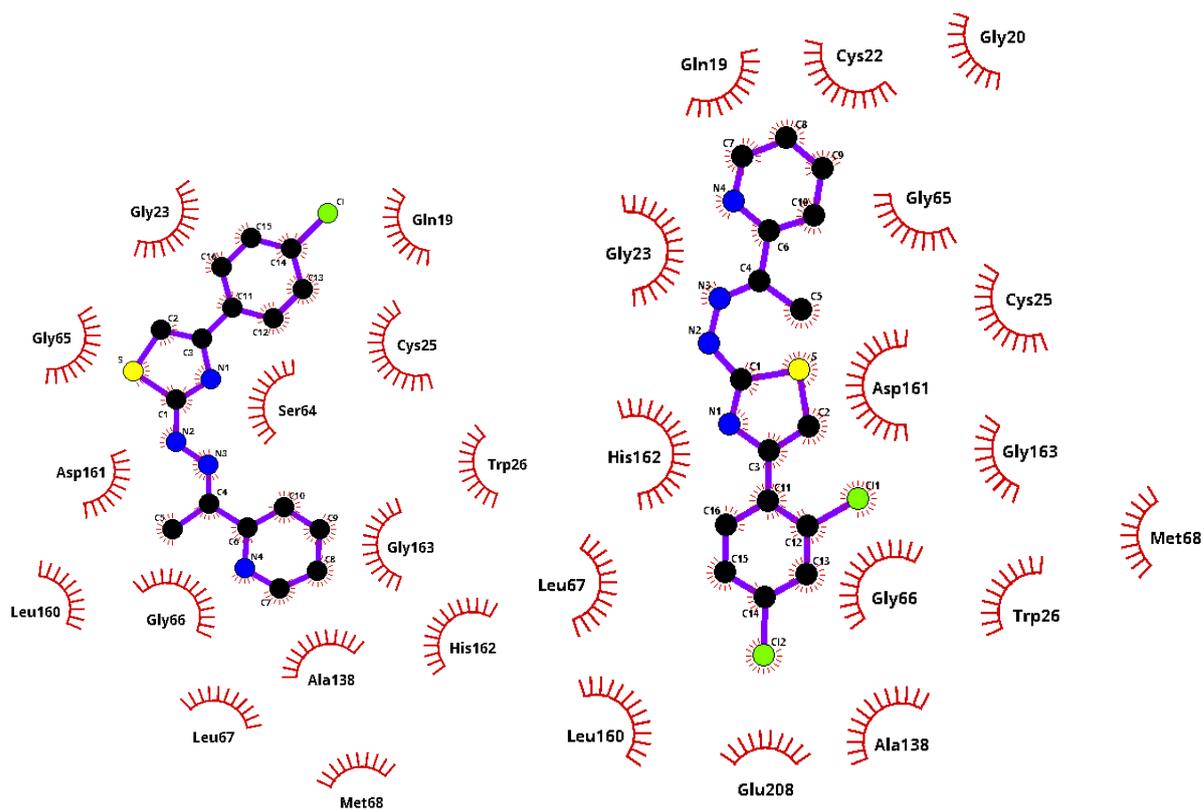
TAP\_02\_3i06



TAP\_04\_3i06

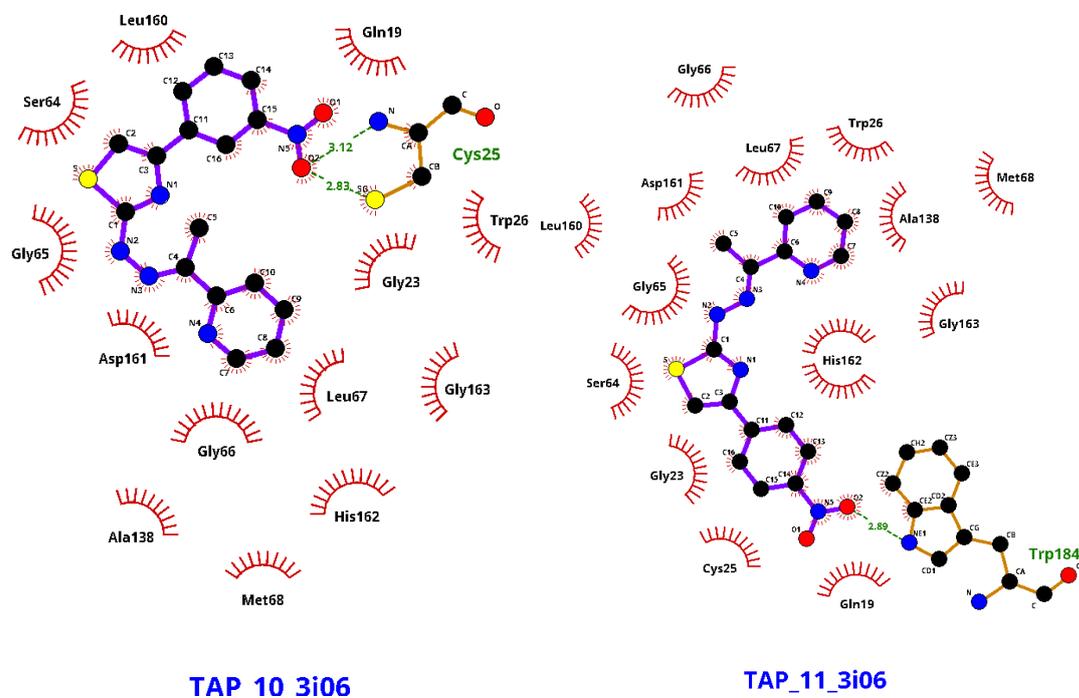


TAP\_06\_3i06



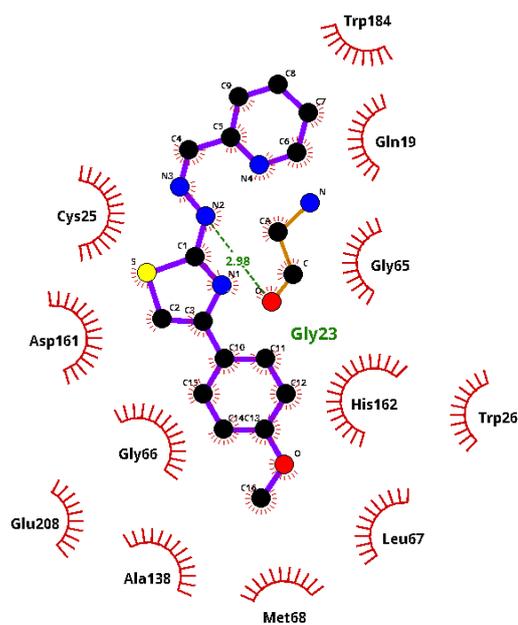
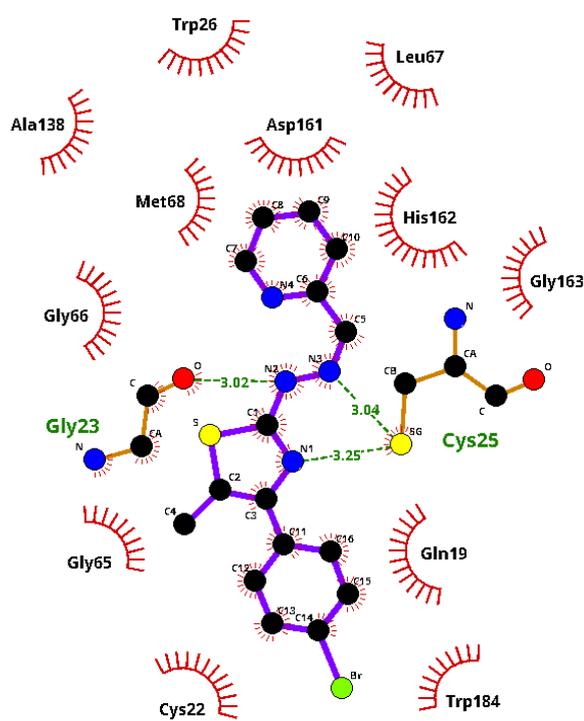
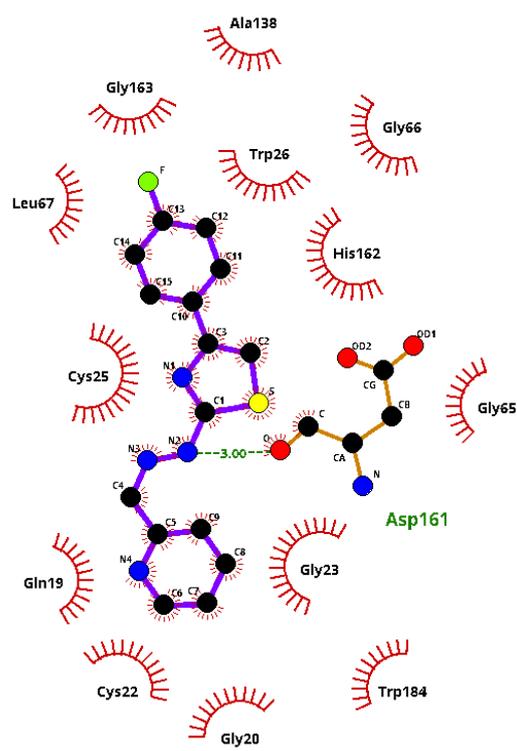
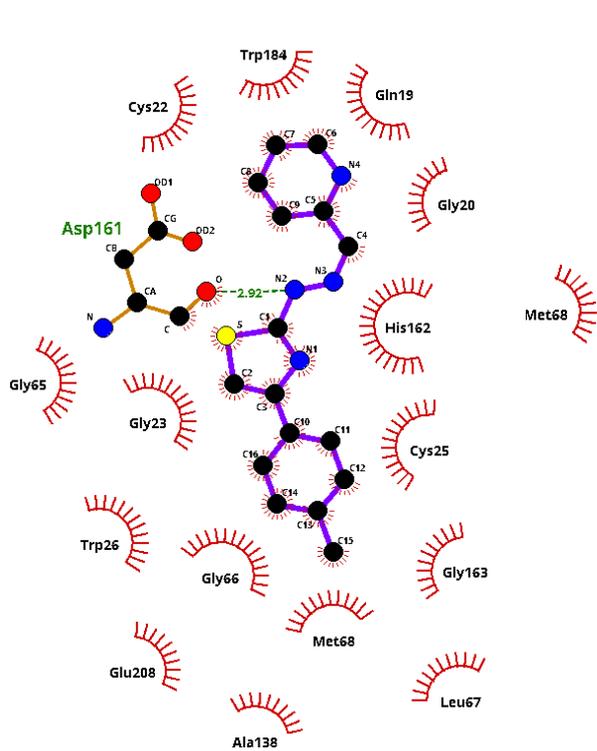
TAP\_07\_3i06

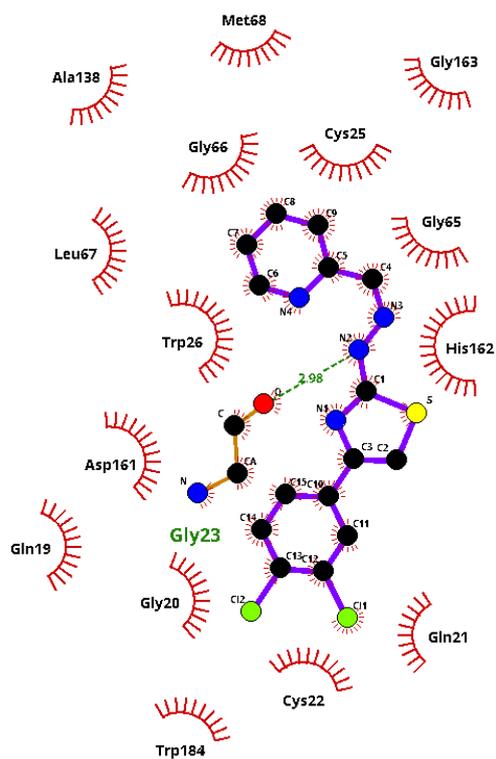
TAP\_08\_3i06



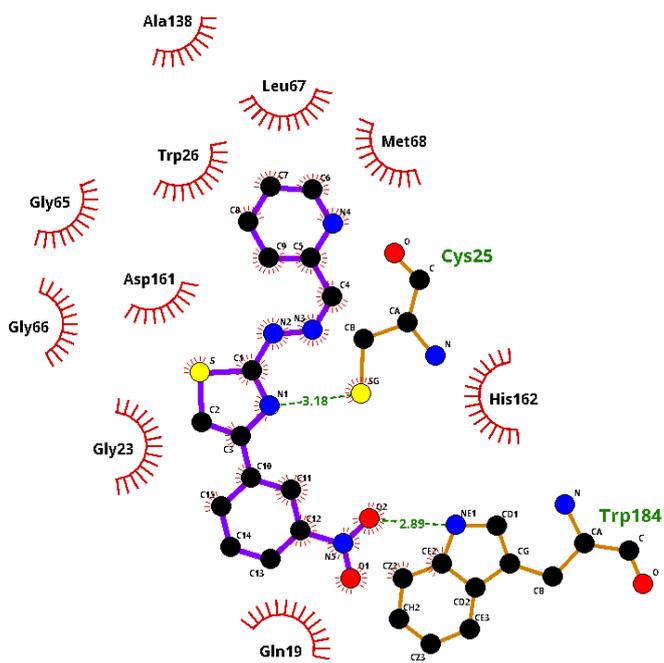
TAP\_10\_3i06

TAP\_11\_3i06

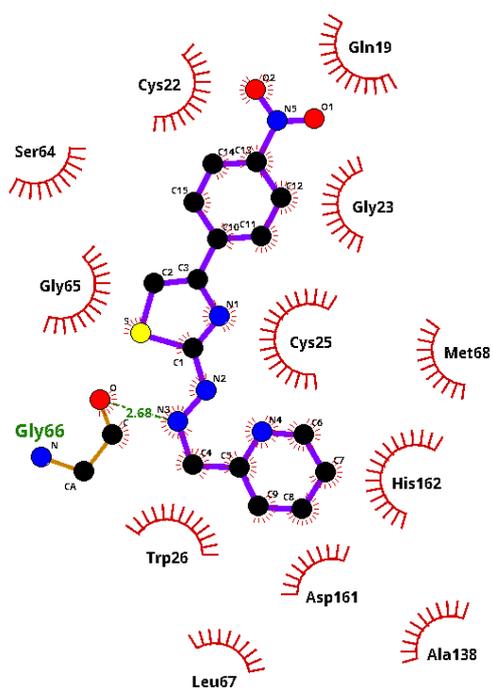




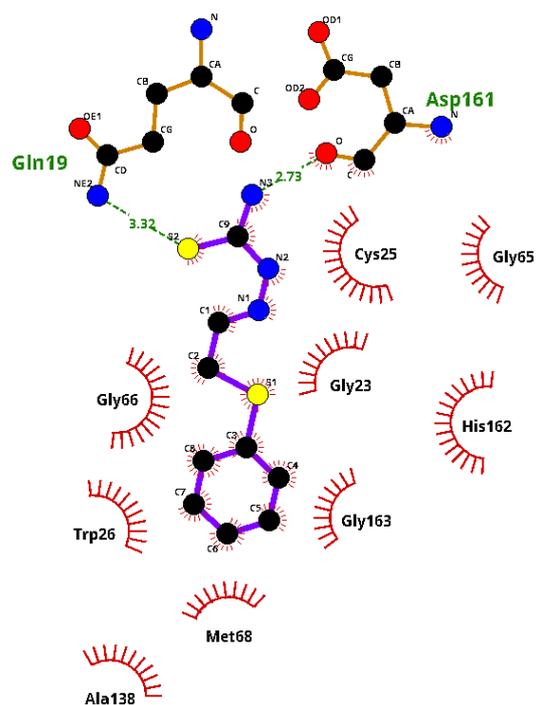
TP\_09\_3i06



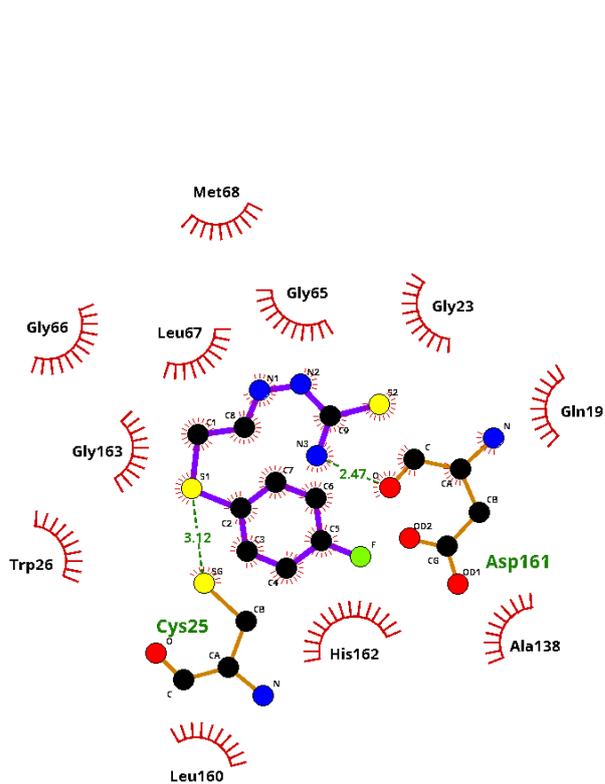
TP\_10\_3i06



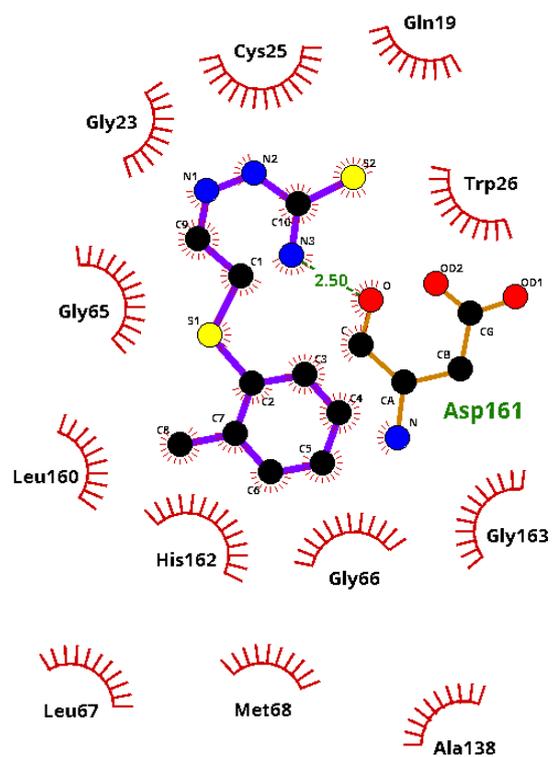
TP\_11\_3i06



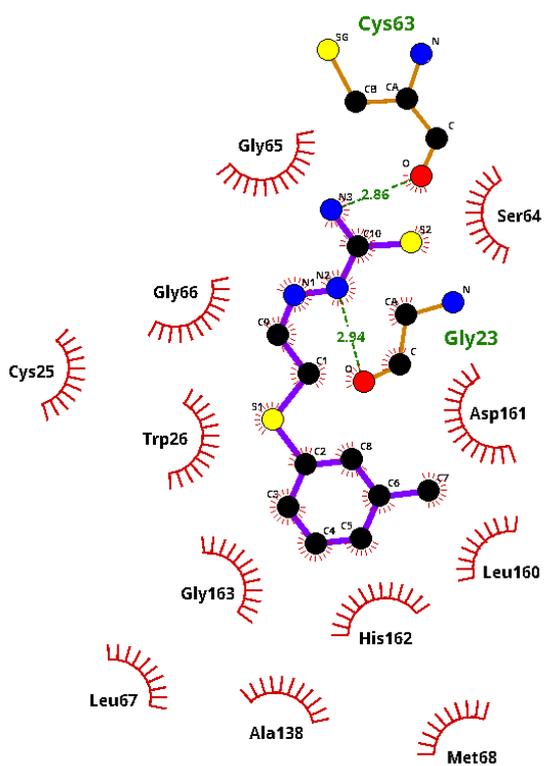
TS\_01\_3i06



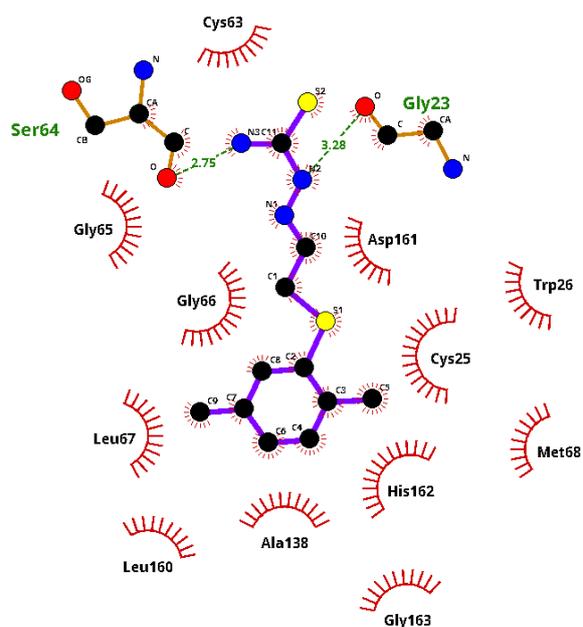
TS\_02\_3i06



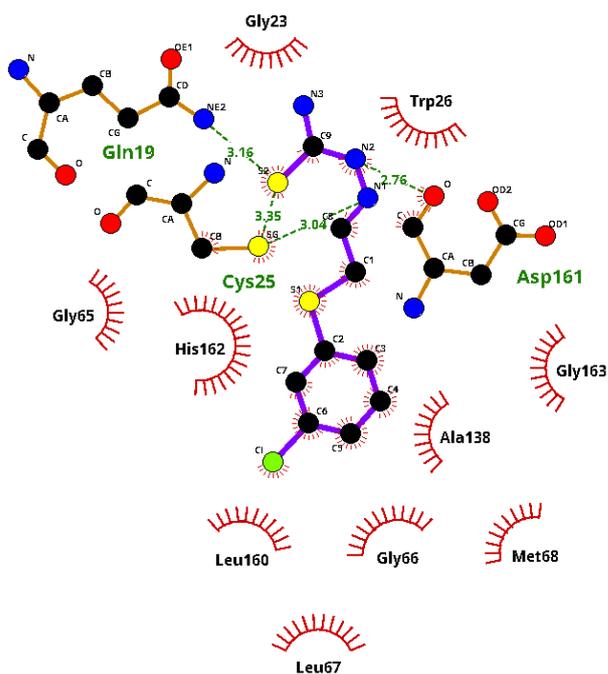
TS\_05\_3i06



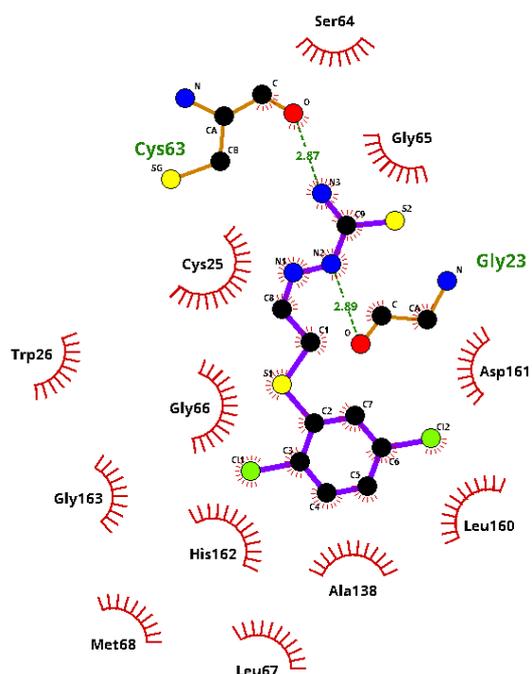
TS\_06\_3i06



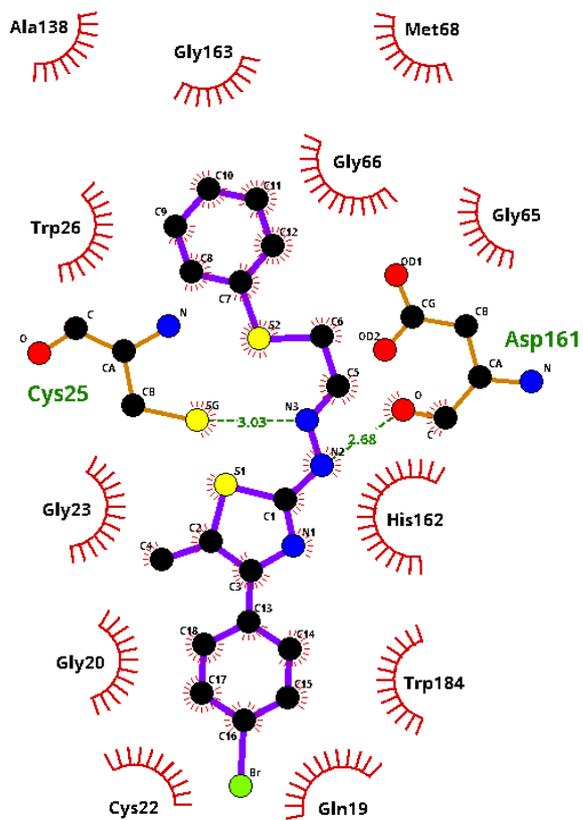
TS\_08\_3i06



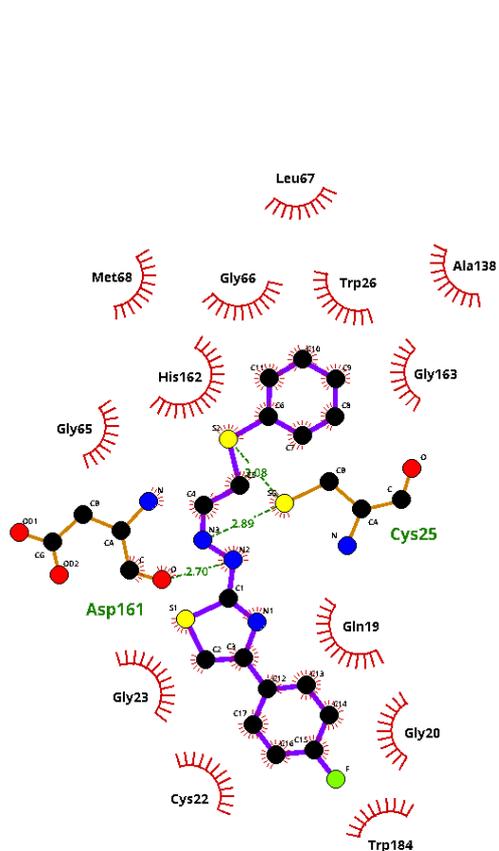
TS\_10\_3i06



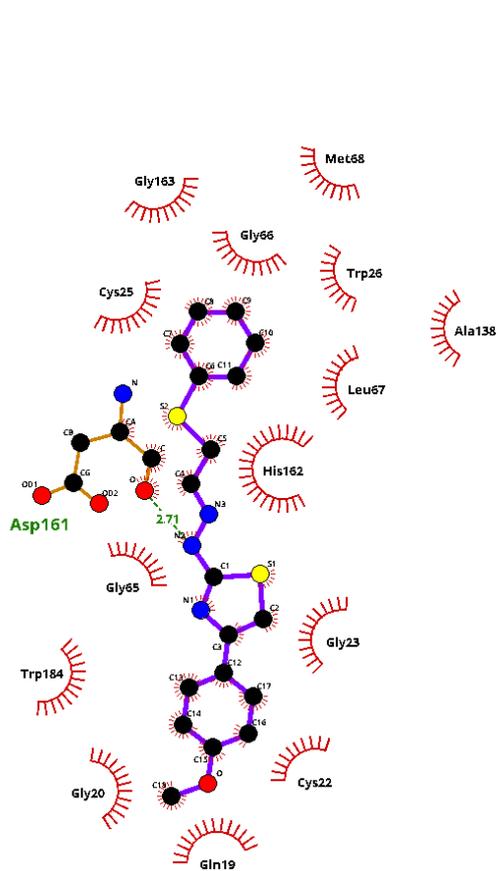
TS\_12\_3i06



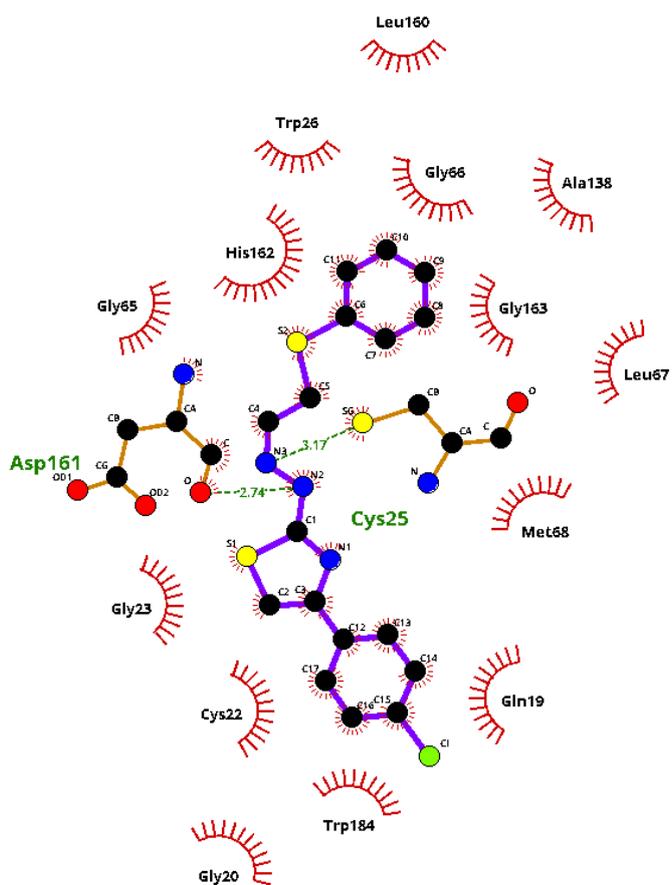
TZ\_04\_3i06



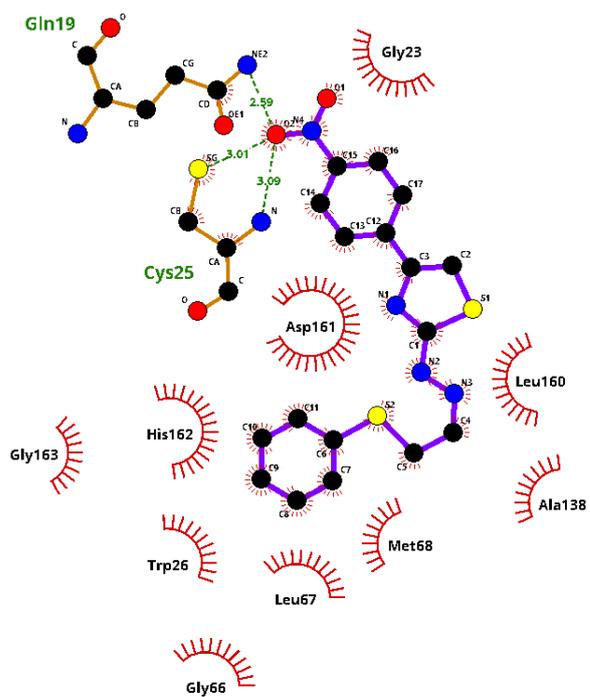
TZ\_05\_3i06



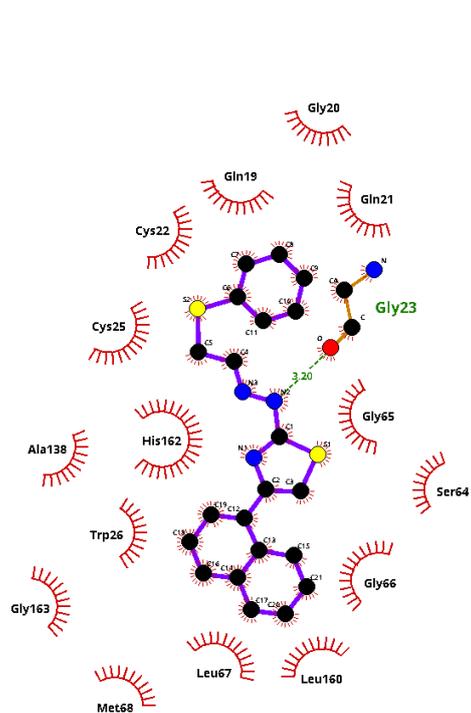
TZ\_06\_3i06



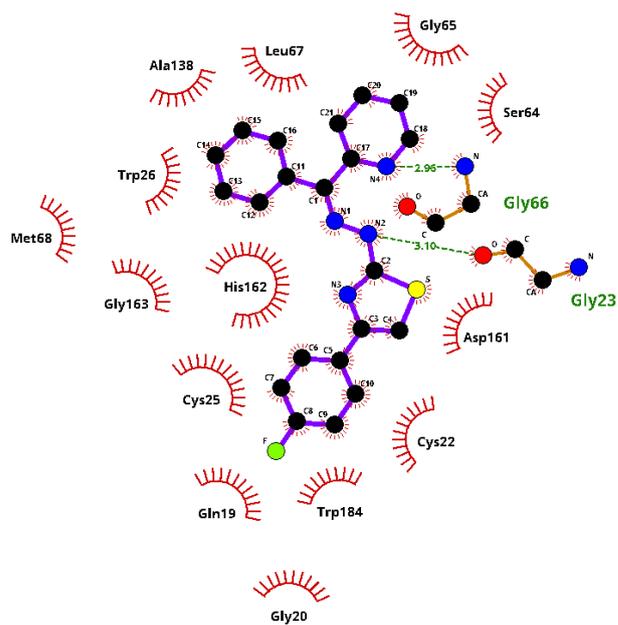
TZ\_07\_3i06



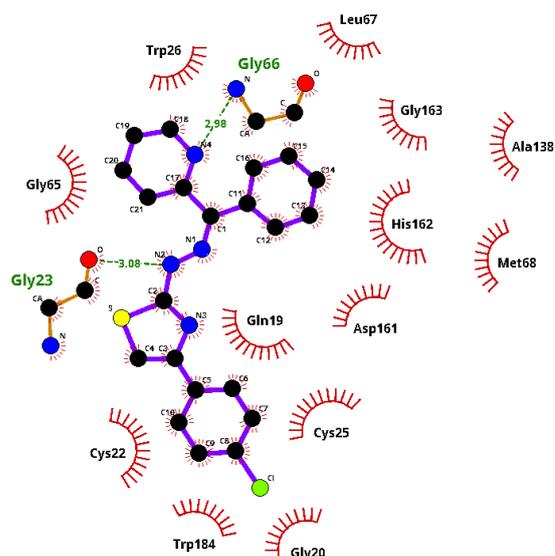
TZ\_11\_3i06



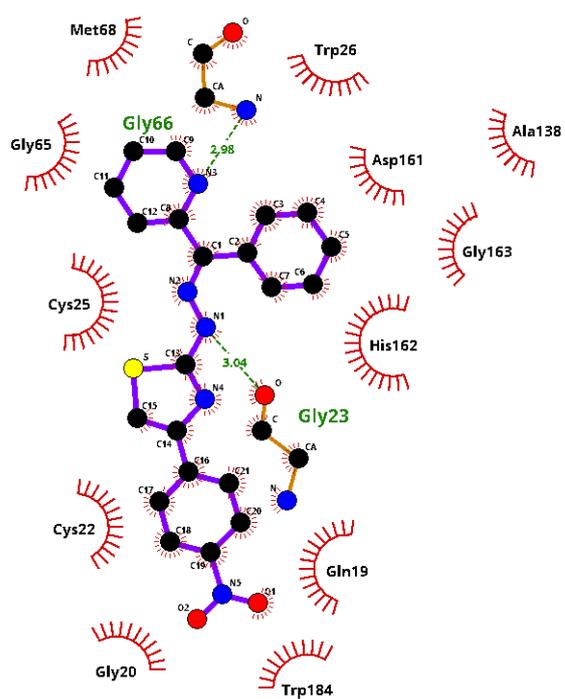
TZ\_12\_3i06



TZBP\_04\_3i06



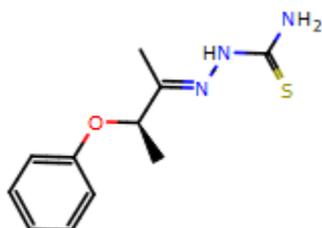
TZBP\_05\_3i06



TZBP\_06\_3i06

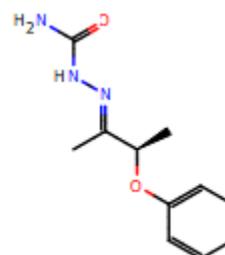
## APÊNDICE D - METABÓLITOS OBTIDOS PELO METASITE

Substrato

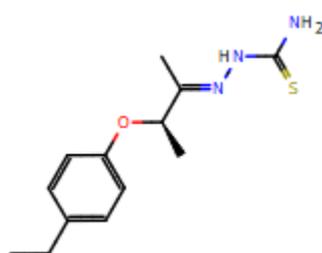


AD-9a.sdf

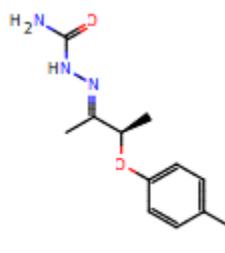
Metabólito



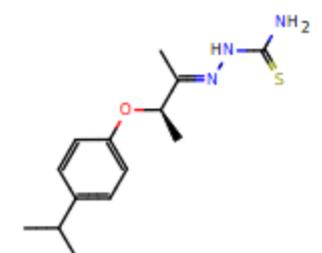
MIM = 221.116427



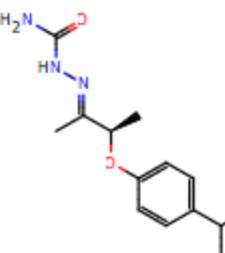
AD-9b.sdf



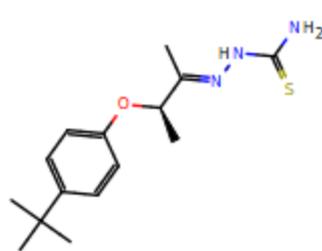
MIM = 249.147727



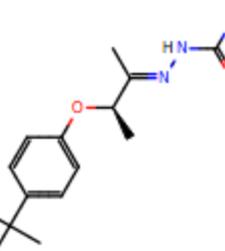
AD-9c.sdf



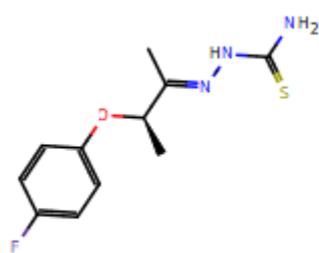
MIM = 263.163377



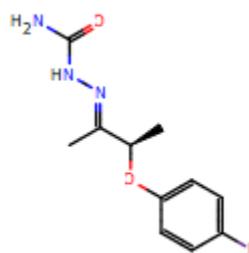
AD-9d.sdf



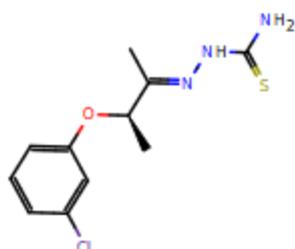
MIM = 277.179027



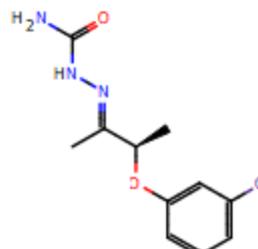
AD-9i.sdf



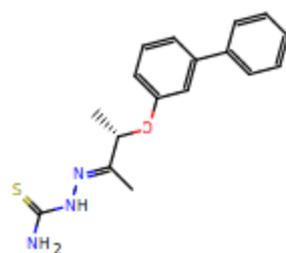
MIM = 239.107005



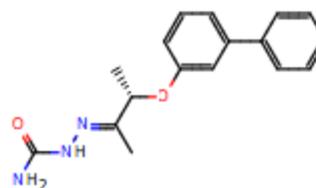
AD-9k.sdf



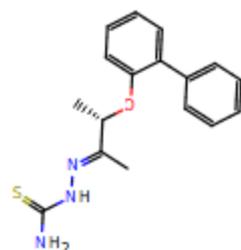
MIM = 255.077454



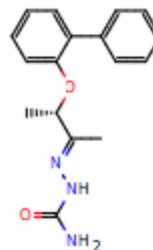
AD-9q.sdf



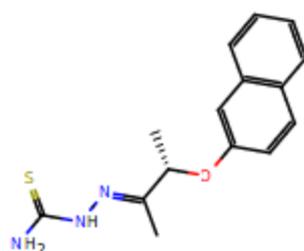
MIM = 297.147727



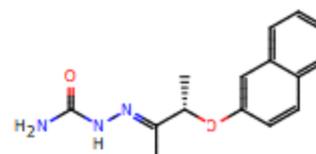
AD-9r.sdf



MIM = 297.147727



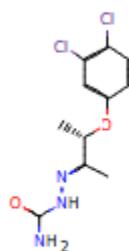
AD-9t.sdf



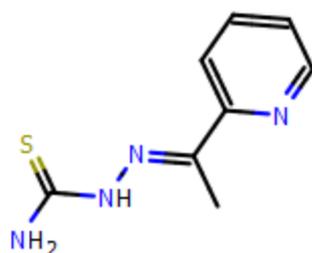
MIM = 271.132077



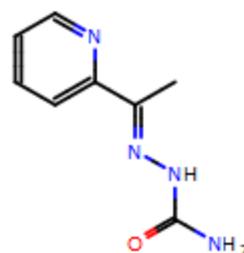
AD-9v.sdf



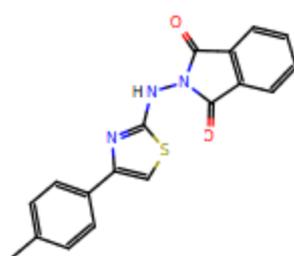
MIM = 289.038482



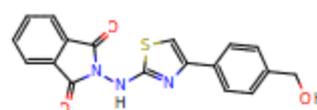
AP.sdf



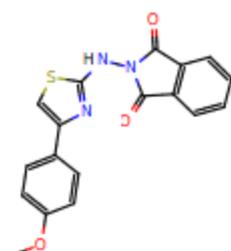
MIM = 178.085461



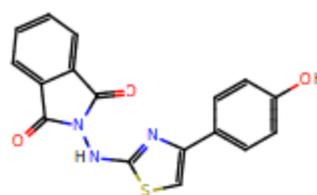
AR-2c.sdf



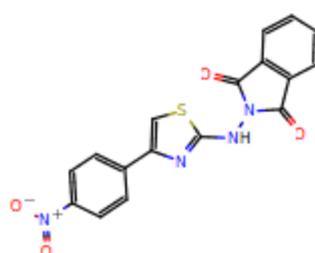
MIM = 351.067762



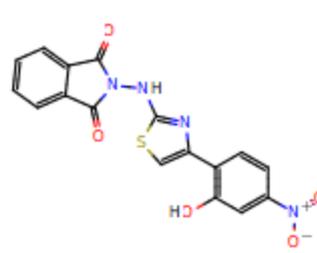
AR-2d.sdf



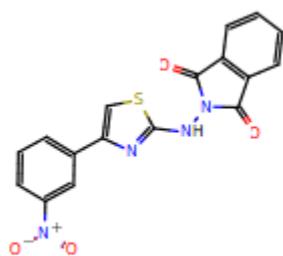
MIM = 337.052112



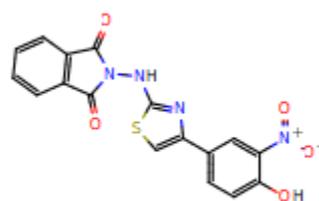
AR-2h.sdf



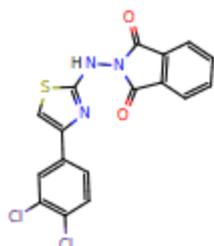
MIM = 382.037190



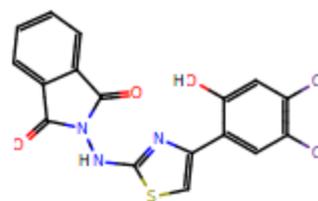
AR-2k.sdf



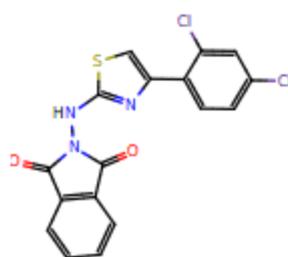
MIM = 382.037190



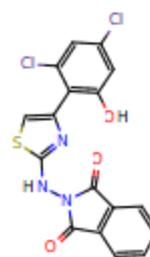
AR-2l.sdf



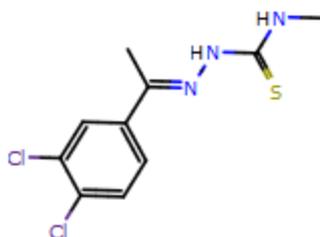
MIM = 404.974168



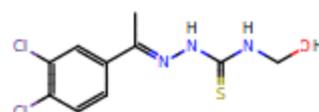
AR-2m.sdf



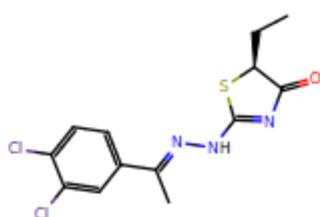
MIM = 404.974168



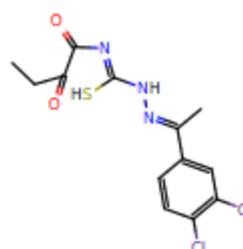
GT-03.sdf



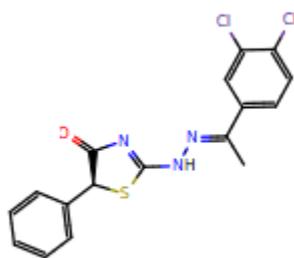
MIM = 290.999988



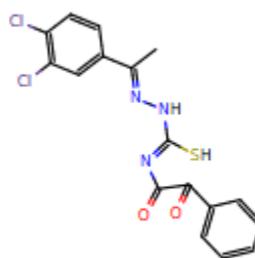
GT-04a.sdf



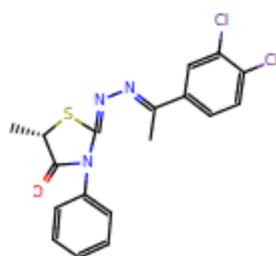
MIM = 345.010553



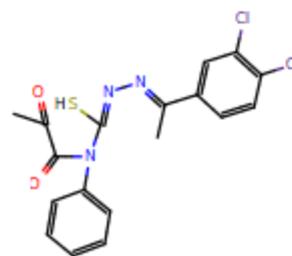
GT-04d.sdf



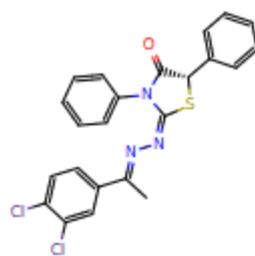
MIM = 393.010553



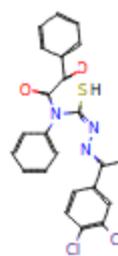
GT-05b.sdf



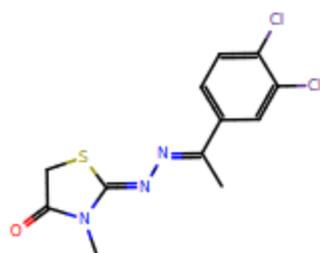
MIM = 407.026203



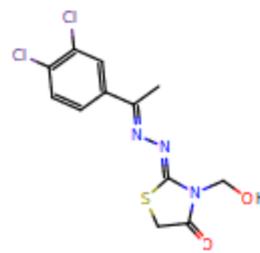
GT-05d.sdf



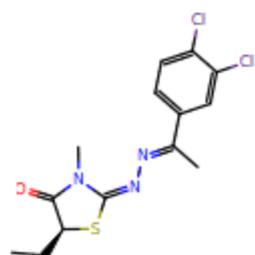
MIM = 469.041853



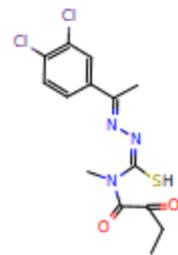
GT-06.sdf



MIM = 330.994903



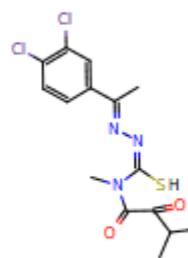
GT-06a.sdf



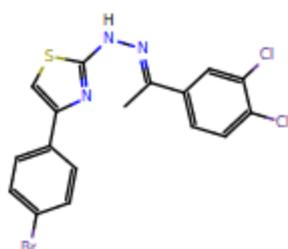
MIM = 359.026203



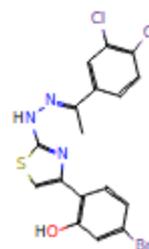
GT-06c.sdf



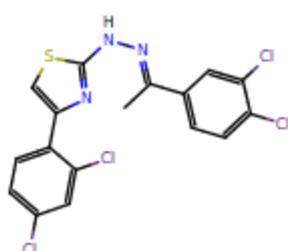
MIM = 373.041853



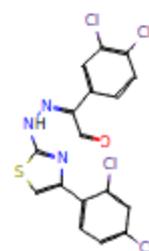
GT-09.sdf



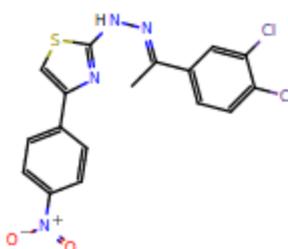
MIM = 454.926150



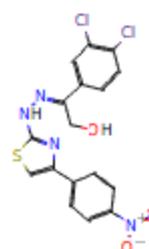
GT-11.sdf



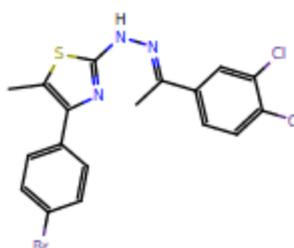
MIM = 442.922044



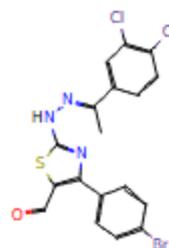
GT-13.sdf



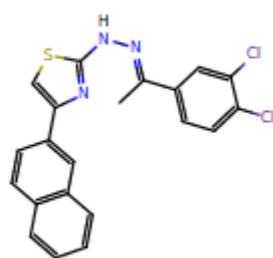
MIM = 422.000717



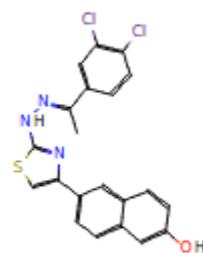
GT-17.sdf



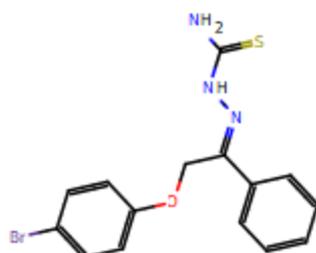
MIM = 466.926150



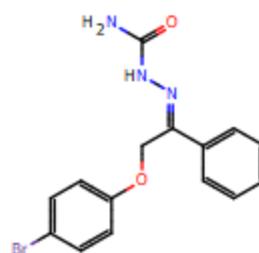
GT-18.sdf



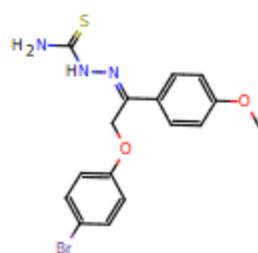
MIM = 427.031288



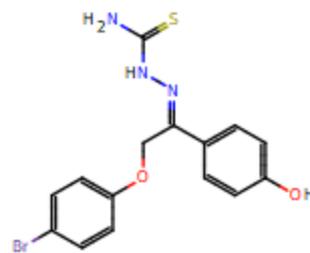
JW-8a.sdf



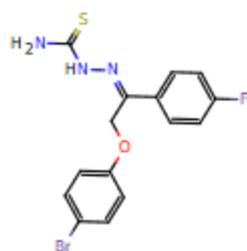
MIM = 347.026939



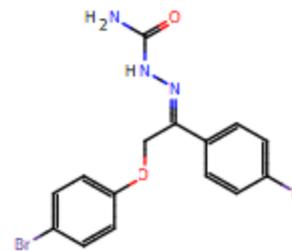
JW-8c.sdf



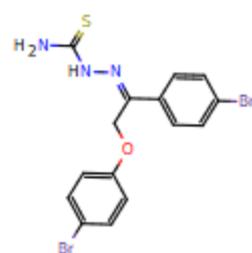
MIM = 378.999010



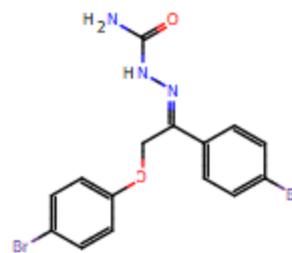
JW-8d.sdf



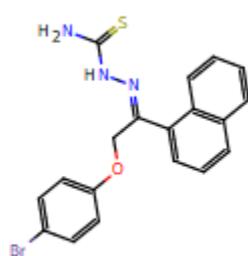
MIM = 365.017517



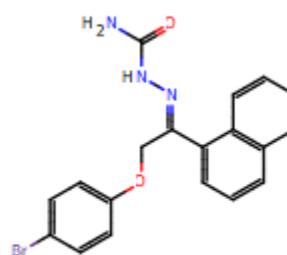
JW-8f.sdf



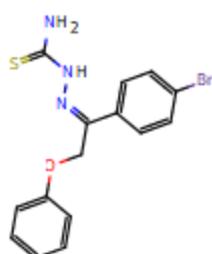
MIM = 424.937451



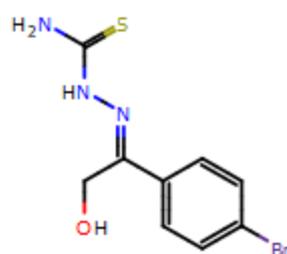
JW-8h.sdf



MIM = 397.042589



JW-9a.sdf



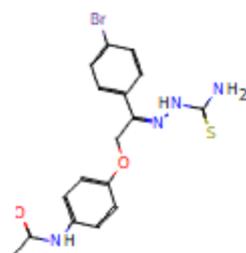
MIM = 286.972795



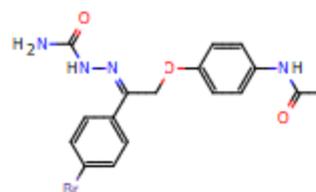
JW-9e.sdf



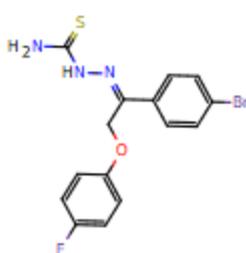
MIM = 378.999010



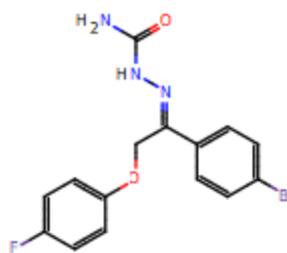
JW-9g.sdf



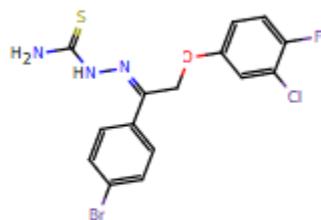
MIM = 404.048403



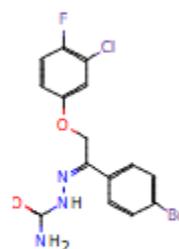
JW-9h.sdf



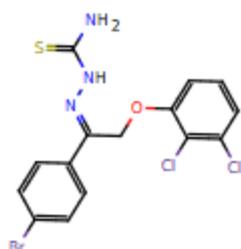
MIM = 365.017517



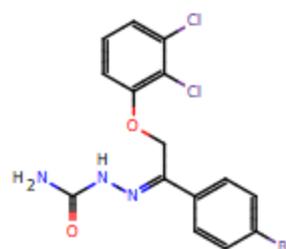
JW-9i.sdf



MIM = 398.978545



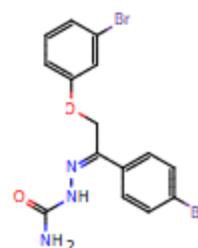
JW-9m.sdf



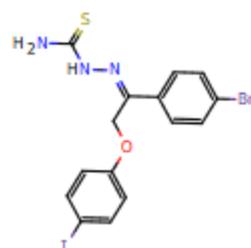
MIM = 414.948994



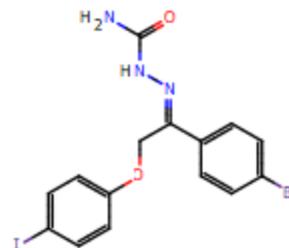
JW-9o.sdf



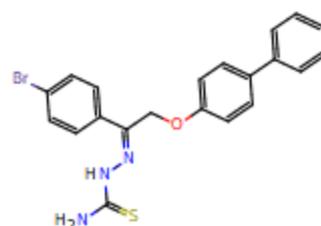
MIM = 424.937451



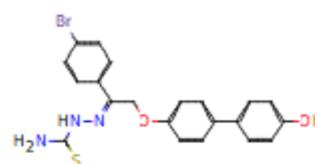
JW-9p.sdf



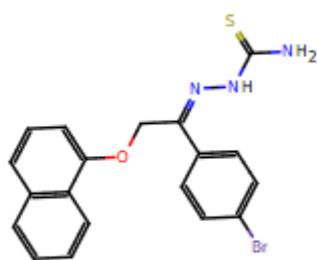
MIM = 472.923587



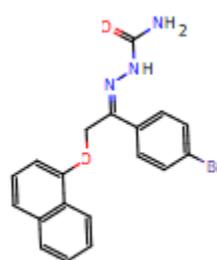
JW-9r.sdf



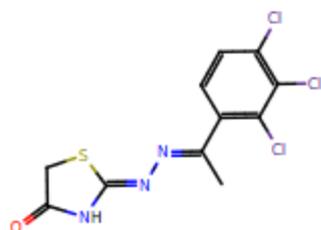
MIM = 455.030310



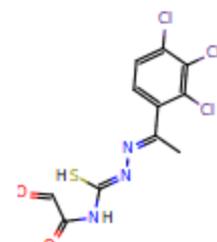
JW-9s.sdf



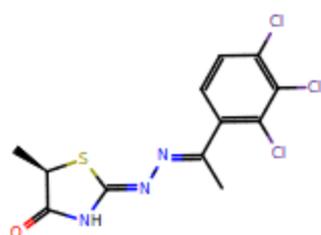
MIM = 397.042589



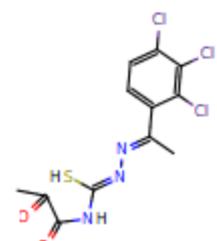
LA-1a1.sdf



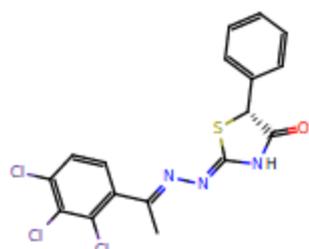
MIM = 350.940281



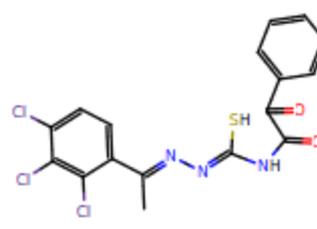
LA-1a2.sdf



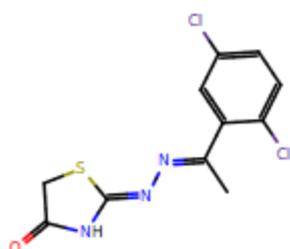
MIM = 364.955931



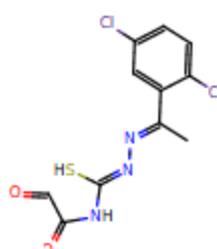
LA-1a5.sdf



MIM = 426.971581



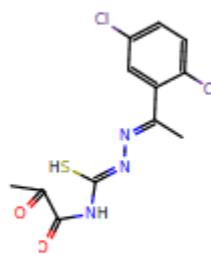
LA-1b1.sdf



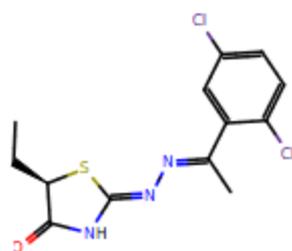
MIM = 316.979253



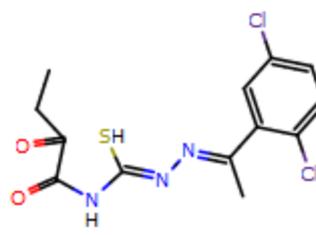
LA-1b2.sdf



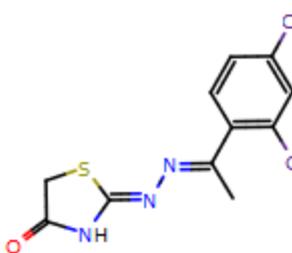
MIM = 330.994903



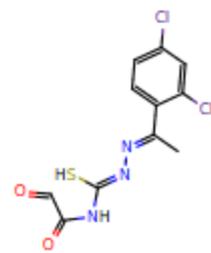
LA-1b3.sdf



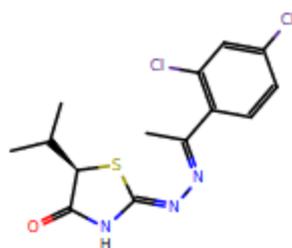
MIM = 345.010553



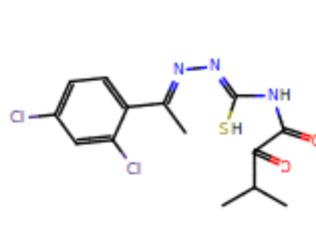
LA-1c1.sdf



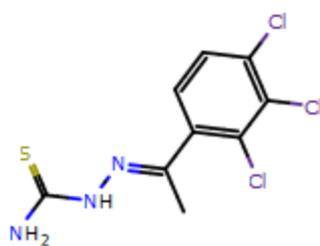
MIM = 316.979253



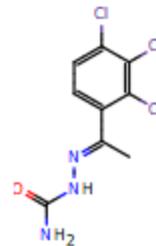
LA-1c4.sdf



MIM = 359.026203



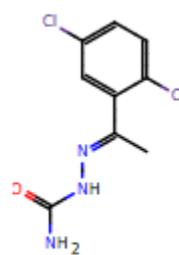
LA-INT-1.sdf



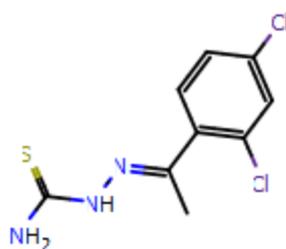
MIM = 278.973295



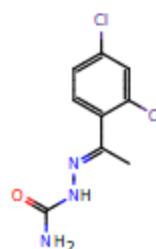
LA-INT-2.sdf



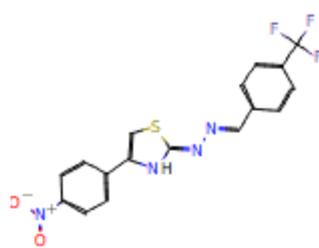
MIM = 245.012267



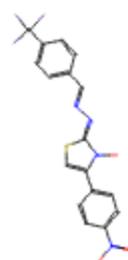
LA-INT-3.sdf



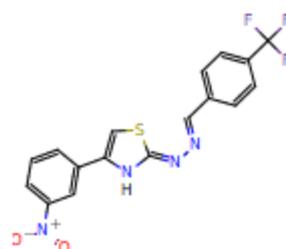
MIM = 245.012267



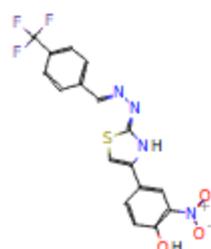
MB1-2.sdf



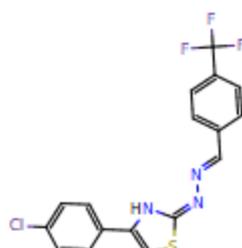
MIM = 408.050396



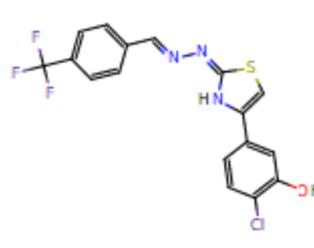
MB1-3.sdf



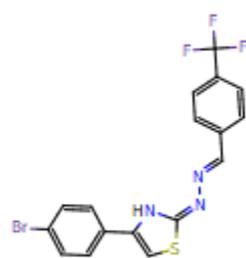
MIM = 408.050396



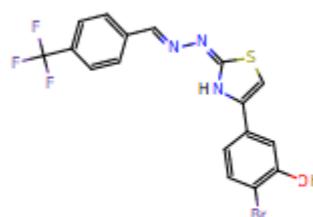
MB1-6.sdf



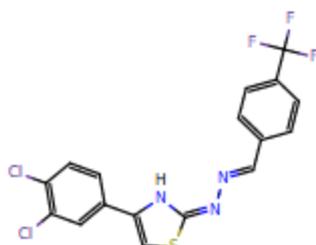
MIM = 397.026345



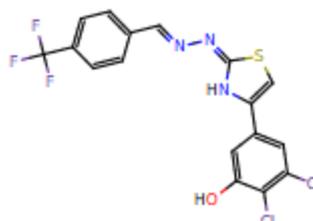
MB1-7.sdf



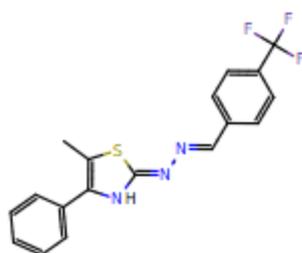
MIM = 440.975830



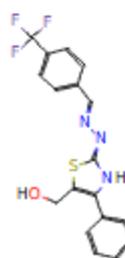
MB1-8.sdf



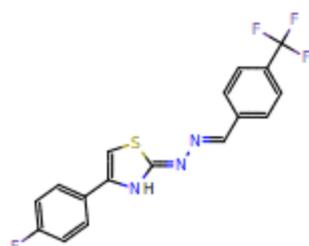
MIM = 430.987373



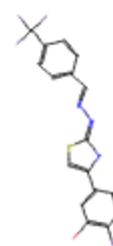
MB1-9.sdf



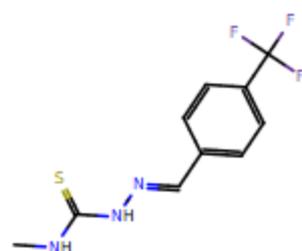
MIM = 377.080968



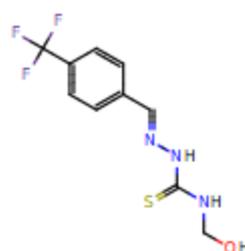
MB1-12.sdf



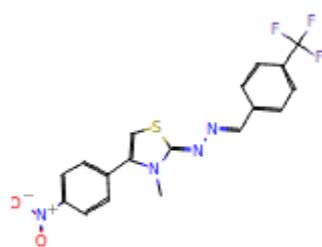
MIM = 381.055896



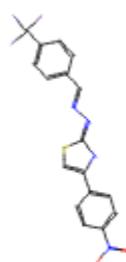
MB02.sdf



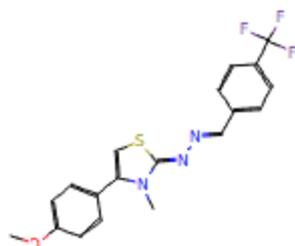
MIM = 277.049668



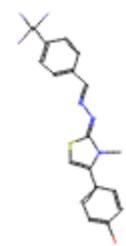
MB2-1.sdf



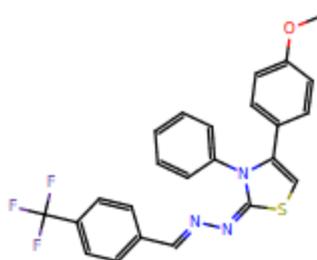
MIM = 392.055481



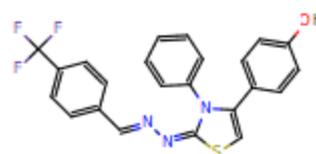
MB2-2.sdf



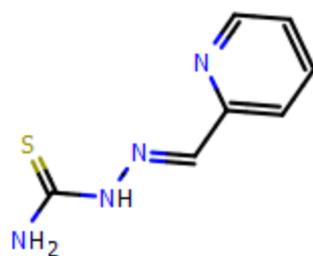
MIM = 377.080968



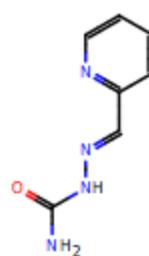
MB3-5.sdf



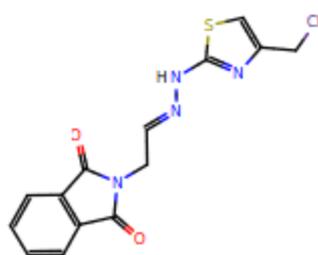
MIM = 439.096618



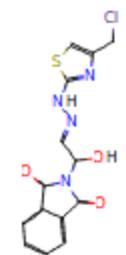
P.sdf



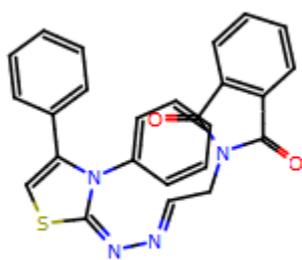
MIM = 164.069811



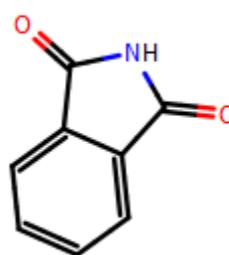
PA-6a.sdf



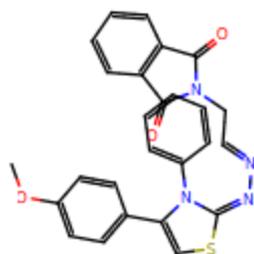
MIM = 350.024039



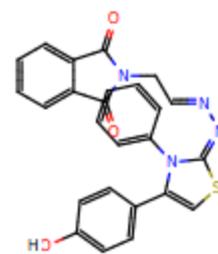
PA-6b.sdf



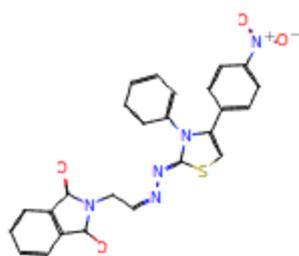
MIM = 147.032028



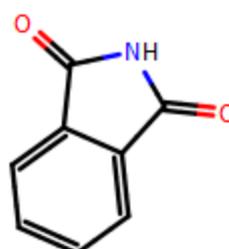
PA-6d.sdf



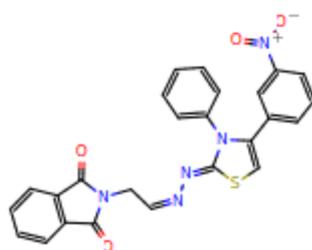
MIM = 454.109961



PA-6h.sdf



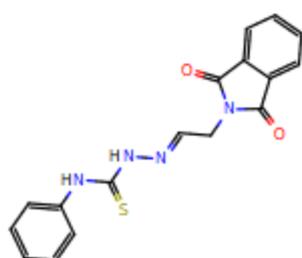
MIM = 147.032028



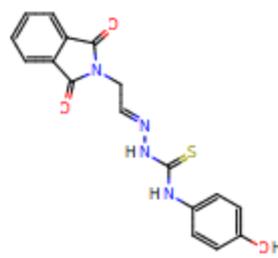
PA-6k.sdf



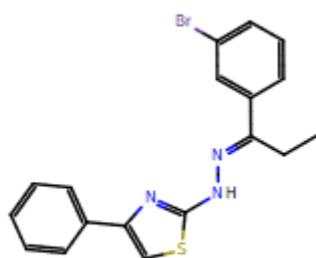
MIM = 147.032028



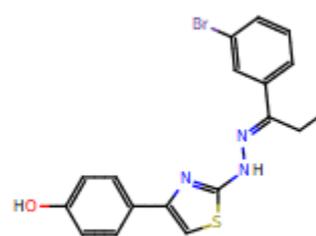
PM-INT-3b.sdf



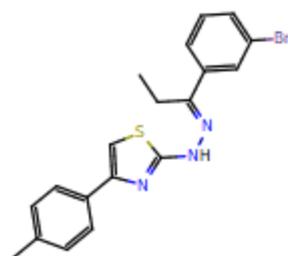
MIM = 354.078661



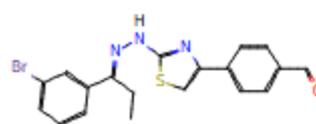
PT01-2.sdf



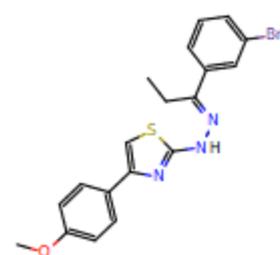
MIM = 401.019745



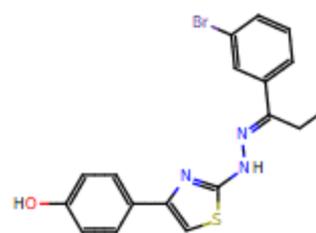
PT01-3.sdf



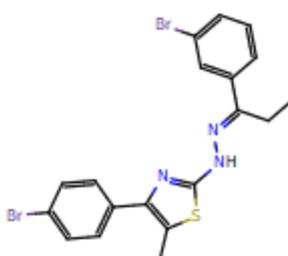
MIM = 413.019745



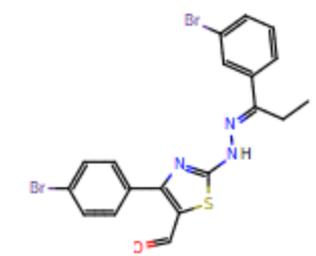
PT01-4.sdf



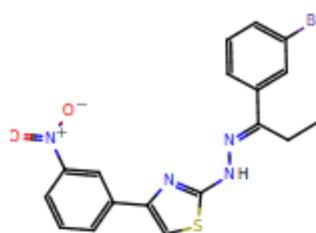
MIM = 401.019745



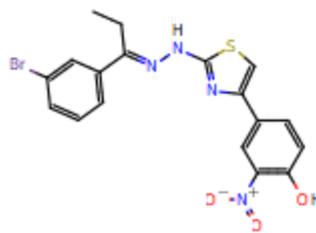
PT01-5.sdf



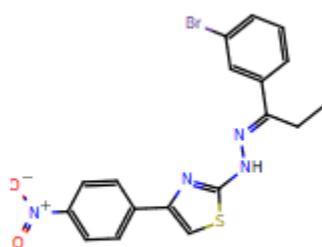
MIM = 490.930257



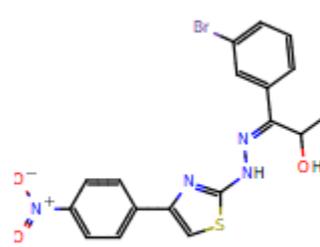
PT01-6.sdf



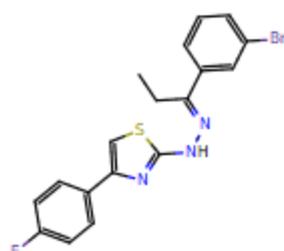
MIM = 446.004823



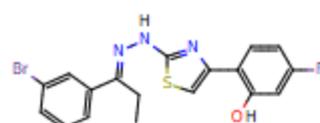
PT01-7.sdf



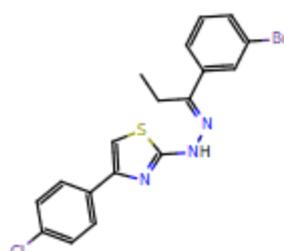
MIM = 446.004823



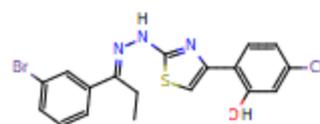
PT01-8.sdf



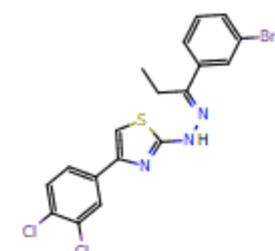
MIM = 419.010323



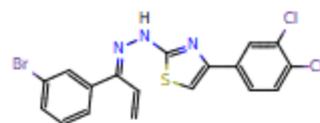
PT01-9.sdf



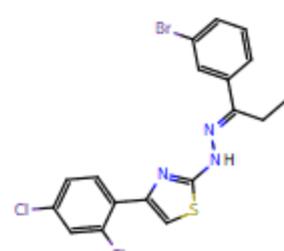
MIM = 434.980773



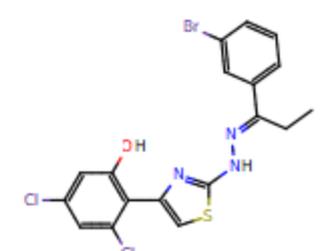
PT01-13.sdf



MIM = 450.931236



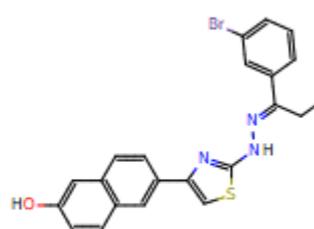
PT01-14.sdf



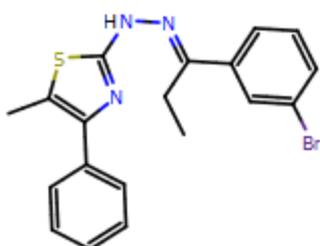
MIM = 468.941801



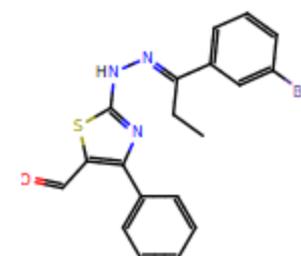
PT01-16.sdf



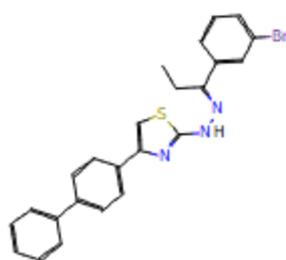
MIM = 451.035395



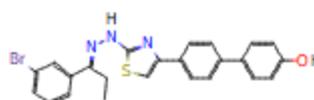
PT01-18.sdf



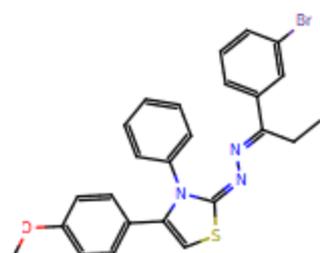
MIM = 413.019745



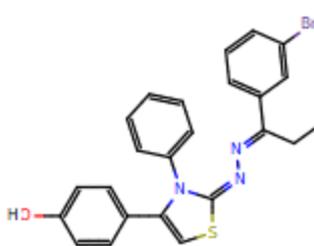
PT01-19.sdf



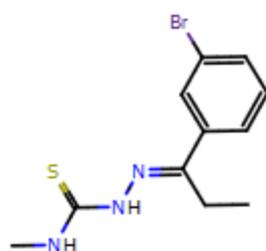
MIM = 477.051045



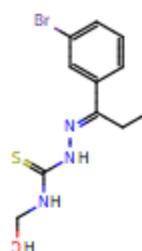
PT02-1.sdf



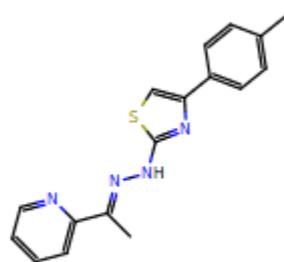
MIM = 477.051045



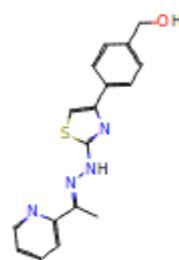
PT03.sdf



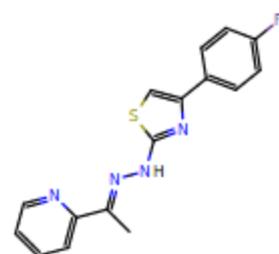
MIM = 315.004095



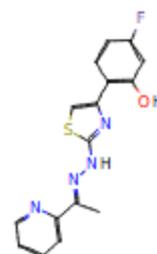
TAP-02.sdf



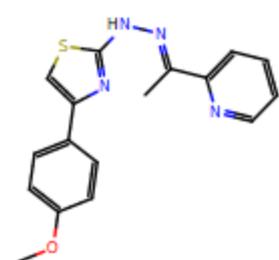
MIM = 324.104482



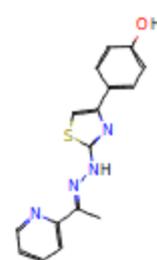
TAP-04.sdf



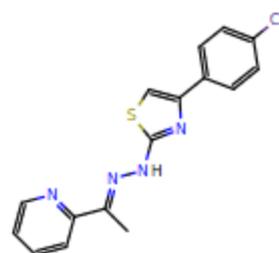
MIM = 328.079410



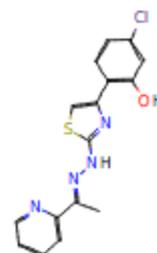
TAP-06.sdf



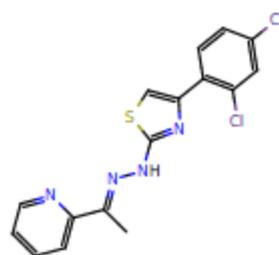
MIM = 310.088832



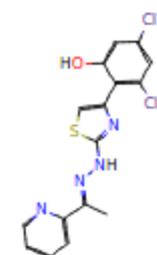
TAP-07.sdf



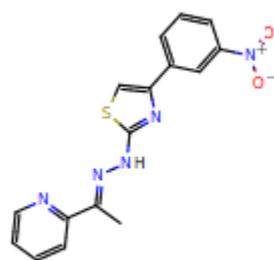
MIM = 344.049860



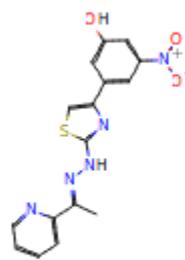
TAP-08.sdf



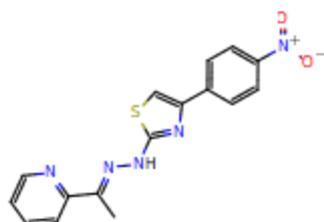
MIM = 378.010887



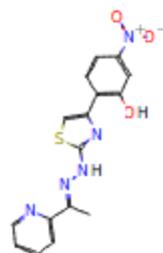
TAP-10.sdf



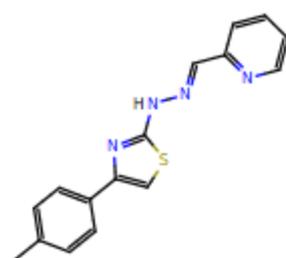
MIM = 355.073910



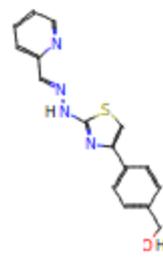
TAP-11.sdf



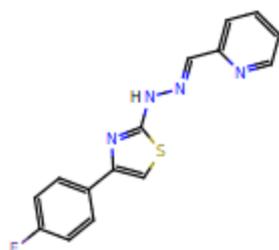
MIM = 355.073910



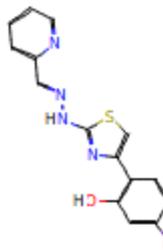
TP-02.sdf



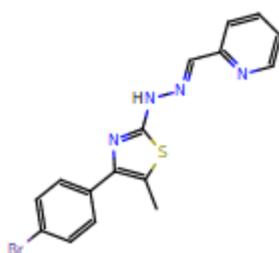
MIM = 310.088832



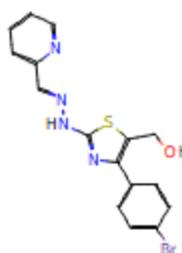
TP-04.sdf



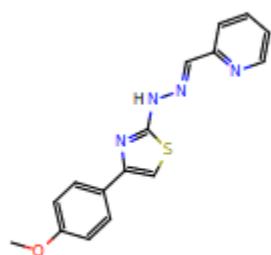
MIM = 314.063760



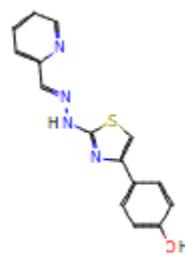
TP-05.sdf



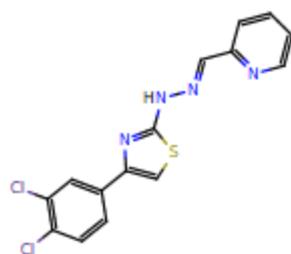
MIM = 387.999344



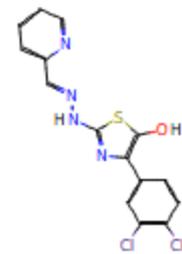
TP-06.sdf



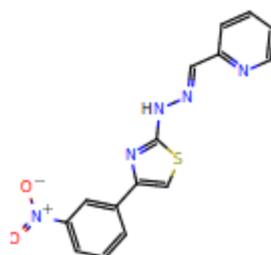
MIM = 296.073182



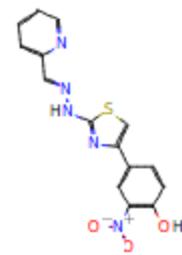
TP-09.sdf



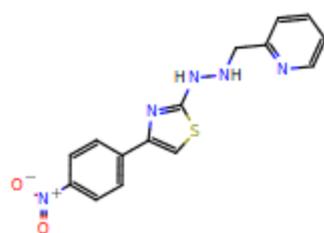
MIM = 363.995237



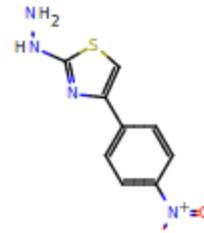
TP-10.sdf



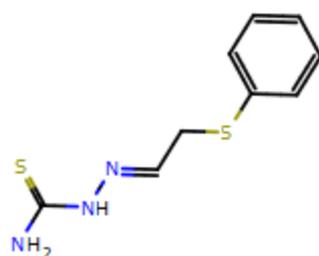
MIM = 341.058260



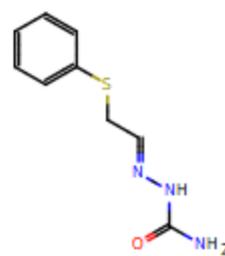
TP-11.sdf



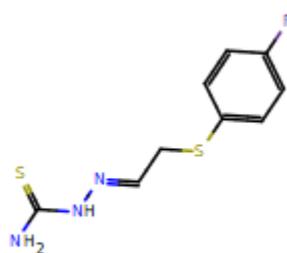
MIM = 236.036797



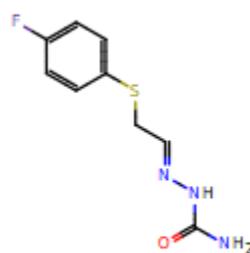
TS-01.sdf



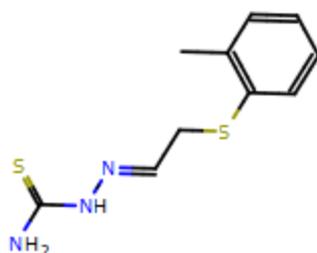
MIM = 209.062283



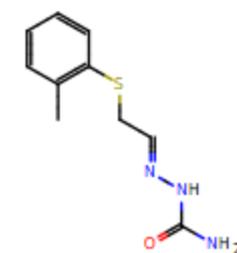
TS-02.sdf



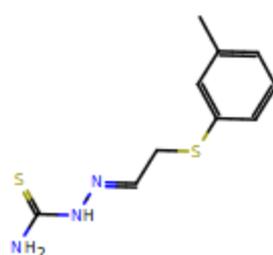
MIM = 227.052861



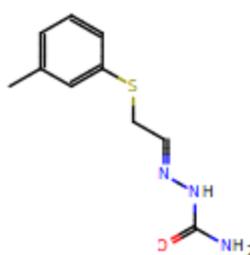
TS-05.sdf



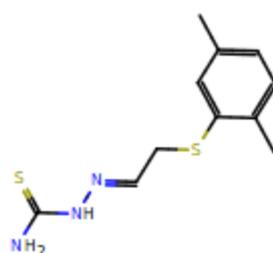
MIM = 223.077933



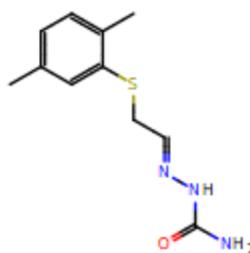
TS-06.sdf



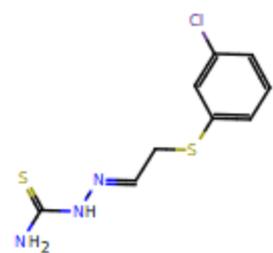
MIM = 223.077933



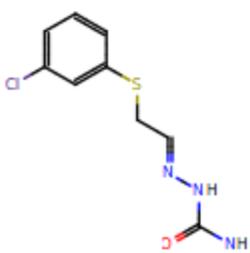
TS-08.sdf



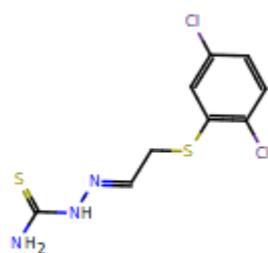
MIM = 237.093583



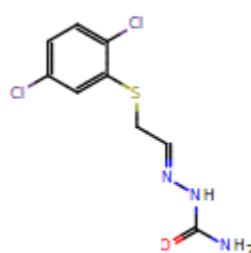
TS-10.sdf



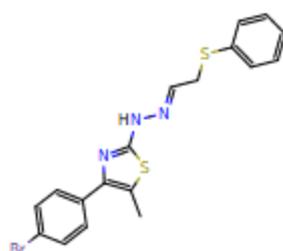
MIM = 243.023311



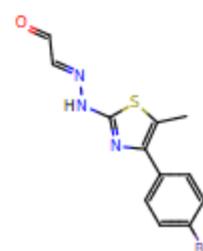
TS-12.sdf



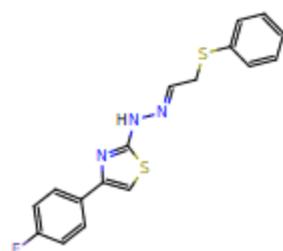
MIM = 276.984338



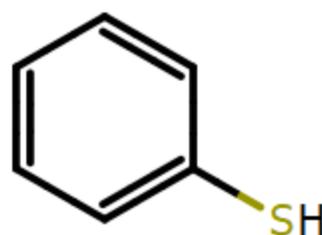
TZ-04.sdf



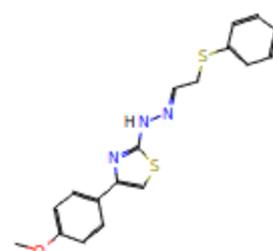
MIM = 322.972795



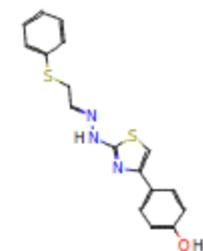
TZ-05.sdf



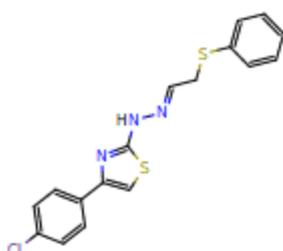
MIM = 110.019021



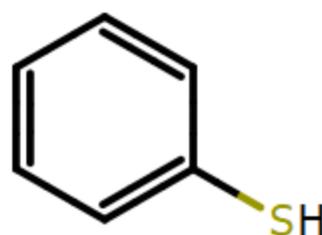
TZ-06.sdf



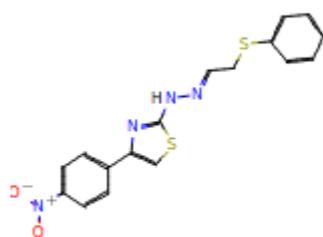
MIM = 341.065654



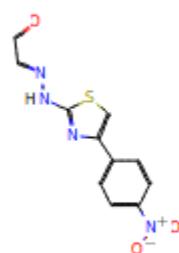
TZ-07.sdf



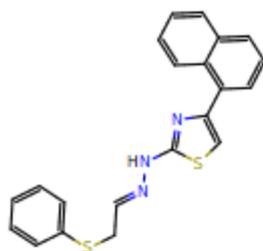
MIM = 110.019021



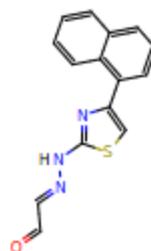
TZ-11.sdf



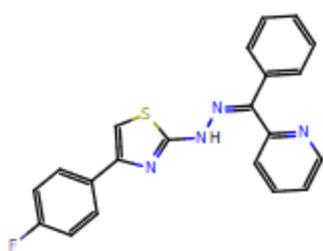
MIM = 276.031711



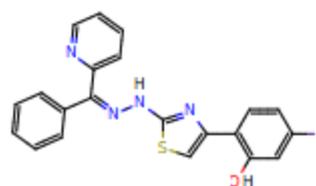
TZ-12.sdf



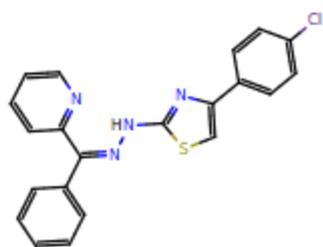
MIM = 281.062283



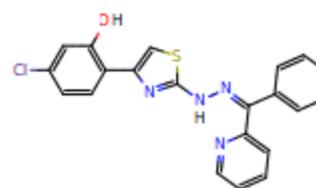
TZBP-04.sdf



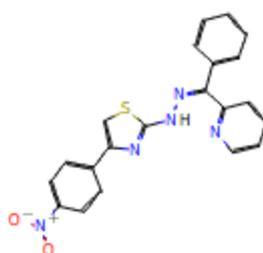
MIM = 390.095060



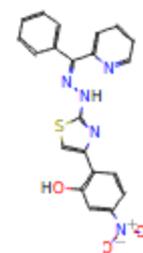
TZBP-05.sdf



MIM = 406.065510



TZBP-06.sdf



MIM = 417.089560

## APÊNDICE E - PRODUÇÃO ACADÊMICA DURANTE O DOUTORADO

### Artigos Publicados



MENDONÇA JUNIOR, F.J.B.; **RIBEIRO, F.F.**; TAVARES, J.F.; SILVA, M.S.; BARBOSA-FILHO, J. M.; SCOTTI, M.T. Natural Product Inhibitors of Topoisomerases: Review and Docking Study. *Current Protein & Peptide Science*, v. 19, p. 275-291, 2018.

Qualis: **B1** Fator de impacto: **2,576**



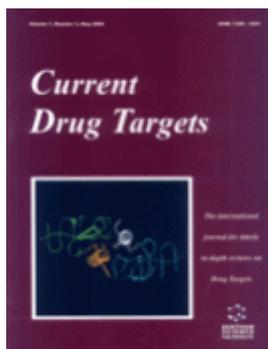
ALVES, M.F.; SCOTTI, M.T.; FÉLIX, MAYARA B.; ISHIKI, H.; MENDONÇA-JÚNIOR, F.J.B.; **RIBEIRO, F.F.**; TAVARES, J.F.; BARBOSA-FILHO, J.M.; SCOTTI, L.; DINIZ, M. Toxicological evaluation in silico and in vivo of secondary metabolites of *Cissampelos sympodialis* in *Rattus norvegicus*. *Current Drug Metabolism*, v. 18, n. 6, p. 566-576, 2017.

Qualis: **B1** Fator de impacto: **2,659**



SILVA-JÚNIOR, E.F.; FRANÇA, P.H.B.; **RIBEIRO, F.F.**; MENDONÇA-JÚNIOR, F.J.B.; SCOTTI, L.; SCOTTI, M.T.; AQUINO, T.M.; ARAÚJO-JÚNIOR, J.X. Molecular Docking Studies Applied to a Dataset of Cruzain Inhibitors. *Current Computer-Aided Drug Design*, v. 13, p. 1-8, 2017.

Qualis: **B2** Fator de impacto: **1,517**



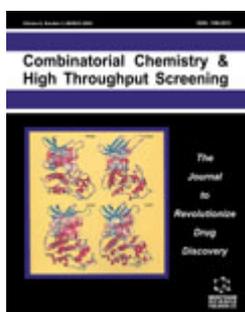
SCOTTI, L.; MENDONCA-JUNIOR, F.J.B.; ISHIKI, H.; RIBEIRO, F.F.; SINGLA, R.; FILHO, J.; SILVA, M.; SCOTTI, M.T. Docking Studies for Multi-Target Drugs. *CURRENT DRUG TARGETS*, v. 18, p. 592-604, 2017.

Qualis: **A2** Fator de impacto: **3,236**



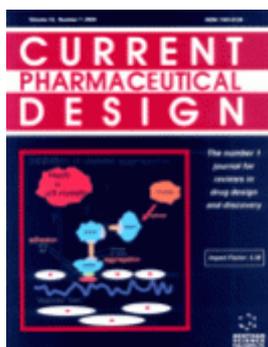
FÉLIX, M.B.; SOUZA, E.R.; LIMA, M.C.A.; FRADE, D.K.G.; SERAFIM, V.L.; RODRIGUES, K.A.F.; NÉRIS, P.L.N.; RIBEIRO, F.F.; SCOTTI, L.; SCOTII, M.T.; AQUINO, T.M.; MENDONÇA-JUNIOR, F.J.B.; OLIVEIRA, M.R. Antileishmanial Activity of new Thiophene-indole Hybrids: Design, Synthesis, Biological and Cytotoxic Evaluation, and Chemometric Studies. *Bioorganic & Medicinal Chemistry* (Print), v. 2, p. 001, 2016.

Qualis: **B1** Fator de impacto: **2,930**



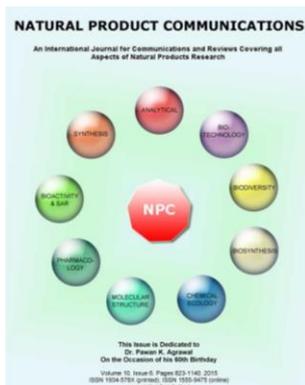
SCOTTI, M.T.; SCOTTI, L.; ISHIKI, H.; RIBEIRO, F.F.; CRUZ, R.M.; OLIVEIRA, M.; MENDONÇA-JUNIOR, F.J.B. Natural Products as a Source for antileishmanial and antitrypanosomal Agents. *Combinatorial Chemistry & High Throughput Screening*, v. 19, p. 1-1, 2016.

Qualis: **B2** Fator de impacto: **0,952**



SCOTTI, L.; MENDONÇA-JUNIOR, F.J.B.; MOURA, R.O.; ISHIKI, H.; RIBEIRO, F.F.; SILVA, M.S.; BARBOSA-FILHO, J.M.; SCOTTI, M.T. Multi-target drugs for Neglected Diseases. *Current Pharmaceutical Design*, v. 22, p. 1-29, 2016.

Qualis: **A2** Fator de impacto: **2,611**



RIBEIRO, F.F.; MENDONÇA-JUNIOR, F.J.B.; SILVA, M.S.; SCOTTI, M. T.; SCOTTI, L. Computational and Investigative Study of Flavonoids Active Against *Trypanosoma cruzi* and *Leishmania* spp. *Natural Product Communications*, v. 10, p. 917-920, 2015.

Qualis: **B3** Fator de impacto: **0,884**

## Artigos Aceitos

- SCOTTI, L.; ISHIKI, H.; MENDONÇA-JÚNIOR, F.J.B.; RIBEIRO, F.F.; SASTRYARLA, N.; SILVA, M.S.; BARBOSA-FILHO, J.M.; SCOTTI, M. T. Computational and Metabolic Studies on a Set of N-Myristoyltransferase Inhibitors Against *Trypanosoma Brucei*. *International Journal of Quantitative Structure-Property Relationships*, 2018.  
Qualis: **C** Fator de impacto: ?
- RIBEIRO, F. F. ; Mendonça Junior F. J. B. ; SILVA, M. S. ; SCOTTI, M. T. ; SCOTTI, L. . Docking of Natural Products against Neurodegenerative Diseases: General Concepts. *Letters in Drug Design & Discovery*, 2017.  
Qualis: **B3** Fator de impacto: **1,17**
- MENDONCA-JUNIOR, F. ; FAVARO, F. ; Scotti, Luciana ; SCOTTI, Marcus Tullius . The Genetic Mechanisms of Antibiotic Resistance and the Role of Antibiotic Adjuvants. *CURRENT TOPICS IN MEDICINAL CHEMISTRY*, 2018.  
Qualis: **B1** Fator de impacto: **2,561**
- Scotti, Luciana; Hamilton M. Ishiki ; FAVARO, F. ; ARAUJO, R. S. A. ; OLIVEIRA, R. B. ; Tiago B. de OLIVEIRA ; BARBOSA-FILHO, J. M. ; SCOTTI, Marcus Tullius . IN SILICO STUDIES IN ANTI-INFLAMMATORY DRUG DISCOVERY; A MINI-REVIEW. *CURRENT ORGANIC CHEMISTRY*, 2018.  
Qualis: **B2** Fator de impacto: **1,924**

- PUTTA, S.; SASTRYARLA, N.; PELUSO, I.; TIWARI, D.K.; REDDY, G.; GIRI, P.V.; KUMAR, N.; MALLA, R.; CHARI, V.B.; D, R.S.R.; BADE, R.; MANNARAPU, M.; BARRETTO, G.; LU, D.; TARASOV, V.; CHUBAREV, V.; FAVARO, F.F.; SCOTTI, L.; SCOTTI, M.T.; KAMAL, M.A.; ALIEV, G.; RAO, C.; PERRY, G.; BISHAYEE, A. Anthocyanins: Possible role as Multitarget therapeutic agents for prevention and therapy of chronic diseases. *Current Pharmaceutical Design*, 2017.
- Qualis: **A2** Fator de impacto: **2,611**

#### Capítulo de livro publicado

- SCOTTI, L.; MENDONÇA-JÚNIOR, F.J.B.; ISHIKI, H.M.; RIBEIRO, F.F.; DUARTE, M.C.; SANTANA, G.S.; OLIVEIRA, T.B.; DINIZ, M.F.F.M.; QUINTANS-JÚNIOR, L.J.; SCOTTI, M.T. Computer-Aided Drug Design Studies in Food Chemistry. In: Alexandru Mihai Grumezescu; Alina Maria Holban. (Org.). *Natural and Artificial Flavoring Agents and Food Dyes*. 1ed. London: Elsevier, 2018, v. 7, p. 261-297.