

**UFRRJ  
INSTITUTO DE CIÊNCIAS EXATAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM  
MATEMÁTICA E COMPUTACIONAL**

**DISSERTAÇÃO**

**Desenvolvimento de Modelos de Predição de atividade  
Inibitória sobre a DNA Girase de Micobactérias Baseados em  
Estudos de Modelagem Molecular**

**Nathalia de Azevedo**

**2021**



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E  
COMPUTACIONAL**

**NATHALIA DE AZEVEDO**

*Sob a Orientação do Professor*  
**Prof. Dr Carlos Mauricio Rabello de Sant'Anna**  
*e Co-orientação do Professor*  
**Prof. Dr Felipe Leite Coelho Da Silva**

**Dissertação submetida**  
**como** requisito parcial para obtenção  
do grau de **Mestre em Ciências**, no  
Programa de Pós-Graduação em  
Modelagem Matemática e  
Computacional, Área de  
Concentração em Modelagem  
Matemática.

Universidade Federal Rural do Rio de Janeiro  
Biblioteca Central / Seção de Processamento Técnico

Ficha catalográfica elaborada  
com os dados fornecidos pelo(a) autor(a)

D994d De Azevedo, Nathalia, 1989-  
Desenvolvimento de Modelos de Predição de atividade  
Inibitória sobre a DNA Girase de Micobactérias  
Baseados em Estudos de Modelagem Molecular. /  
Nathalia De Azevedo. - Rio de Janeiro, 2021.  
103 f.: il.

Orientador: Carlos Mauricio Sant'Anna.  
Coorientador: Felipe Silva.  
Dissertação(Mestrado). -- Universidade Federal  
Rural do Rio de Janeiro, Programa de Pós-graduação em  
Modelagem Matemática e Computacional PPGMMC-UFRRJ,  
2021.

1. Docagem Molecular. 2. Regressão Linear  
Múltipla. 3. Mycobacterium tuberculosis. 4. M.  
smegmatis. 5. Tuberculose. I. Sant'Anna, Carlos  
Mauricio , 1965-, orient. II. Silva, Felipe, 1981-,  
coorient. III Universidade Federal Rural do Rio de  
Janeiro. Programa de Pós-graduação em Modelagem  
Matemática e Computacional PPGMMC-UFRRJ. IV. Título.



**MINISTÉRIO DA EDUCAÇÃO  
UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO  
PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E  
COMPUTACIONAL**

Folha de homologação Nº 4797/2021 - PPGMMC (12.28.01.00.00.61)

Nº do Protocolo: 23083.075997/2021-17

Seropédica-RJ, 22 de outubro de 2021.

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO**

**INSTITUTO DE CIÊNCIAS EXATAS**

**PROGRAMA DE PÓS-GRADUAÇÃO EM MODELAGEM MATEMÁTICA E COMPUTACIONAL –  
PPGMMC**

**NATHALIA DE AZEVEDO**

Dissertação submetida como requisito parcial para a obtenção de grau de **Mestre**, no Programa de Pós-Graduação em Modelagem Matemática e Computacional - PPGMMC, área de Concentração em Modelagem Matemática e Computacional.

DISSERTAÇÃO APROVADA EM 14/10/2021

**Conforme deliberação número 001/2020 da PROPPG, de 30/06/2020**, tendo em vista a implementação de trabalho remoto e durante a vigência do período de suspensão das atividades acadêmicas presenciais, em virtude das medidas adotadas para reduzir a propagação da pandemia de Covid-19, nas versões finais das teses e dissertações as assinaturas originais dos membros da banca examinadora poderão ser substituídas por documento(s) com assinaturas eletrônicas. Estas devem ser feitas na própria folha de assinaturas, através do SIPAC, ou do Sistema Eletrônico de Informações (SEI) e neste caso a folha com a assinatura deve constar como anexo ao final da tese / dissertação.

Carlos Maurício Rabello de Sant Anna Drº UFRRJ (Orientador, Presidente da Banca)

Ronaldo Malheiros Gregório Drº UFRRJ (membro interno)

Alexandre Sousa da Silva Drº UNIRIO (externo à Instituição)

*(Assinado digitalmente em 22/10/2021 01:21)*  
CARLOS MAURICIO RABELLO DE SANT ANNA  
COORDENADOR CURS/POS-GRADUACAO - TITULAR  
PPGQ (12.28.01.00.00.60)  
Matrícula: 1220404

*(Assinado digitalmente em 27/10/2021 10:21)*  
ALEXANDRE SOUSA DA SILVA  
ASSINANTE EXTERNO  
CPF: 278.613.148-04

*(Assinado digitalmente em 22/10/2021 07:17)*  
RONALDO MALHEIROS GREGORIO  
PROFESSOR DO MAGISTERIO SUPERIOR  
DeptTL/IM (12.28.01.00.00.90)  
Matrícula: 1669607

Para verificar a autenticidade deste documento entre em <https://sipac.ufrj.br/public/documentos/index.jsp> informando seu número: **4797**, ano: **2021**, tipo: **ATA**, data de emissão: **22/10/2021** e o código de verificação: **a3a8e4c9ff**

## **Agradecimentos**

Agradeço, sempre, primeiramente a Deus por tudo que fez, faz e sei que há de fazer em minha vida por Seu grande amor e misericórdia

Aos meus pais, Roberto de Azevedo e Eliane Vieira de Azevedo, pelo tamanho amor e cuidado por mim, sempre me apoiando e me preparando para a vida através de seus gestos e sendo meus espelhos.

A minha amiga e irmã, Fernanda de Azevedo, que é alguém que sei que posso contar em todos os momentos.

Ao meu namorado Rafael de Barros Lima por todo apoio, compreensão, incentivo e por acreditar mais em mim do que eu mesma.

Aos meus orientadores. Ao professor Carlos Maurício R. Sant'Anna, por ter me concedido a honra de ser sua aluna e orientanda e pela sugestão de um projeto pelo qual me apaixonei. Ao professor Felipe Leite Coelho Da Silva por todas as orientações, aulas e exemplo de humildade.

A coordenadora do PPGMMC Claudia Mazza Dias por sempre ouvir minhas angústias a respeito do curso e me tranquilizar. Aos demais professores do corpo docente do Programa, em especial ao professor Ronaldo Gregório por ter lecionado aulas que me fizeram compreender conceitos que eu não havia entendido na graduação.

Aos meus amigos Igor Leite da Silva por ser um anjo prestativo, Helainy De Almeida Torres pelas conversas e trocas de conhecimento. Aos amigos da turma de modelagem molecular que tanto me ajudaram, principalmente o Bruno Henrique Medeiros Mendes.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

## Resumo

Azevedo, Nathalia de. Desenvolvimento de Modelos de Predição de atividade Inibitória sobre a DNA Girase de Micobactérias Baseados em Estudos de Modelagem Molecular. Seropédica, 2019. Dissertação (Mestrado em Modelagem Matemática e Computacional, Matemática). Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2019.

A tuberculose (TB) é uma doença que apresenta elevadas morbidade e mortalidade. Em 2019, aproximadamente 10 milhões de pessoas desenvolveram a doença, das quais cerca de 1,4 milhão de pessoas morreram. Existem antibióticos efetivos contra a bactéria causadora da tuberculose, *Mycobacterium tuberculosis* (MTB), mas o aumento da incidência de casos de TB resistente a esses medicamentos tem exigido esforços para se descobrir novos medicamentos capazes de combater a doença. As topoisomerasas são enzimas que mantêm a topologia do DNA durante a replicação, transcrição e recombinação. A DNA girase é a única topoisomerase tipo II presente no MTB, sendo por isso um alvo interessante a ser explorado para o planejamento de novos medicamentos contra a TB. A DNA girase é composta por duas subunidades, GyrA e GyrB e este projeto tem como objetivo usar grupos de compostos da literatura com dados de atividade registrados sobre a subunidade GyrB da DNA girase de micobactérias para se desenvolver um modelo para a predição da atividade para ser aplicado em futuros procedimentos de triagem virtual. Para atingir esse objetivo, foram combinados métodos de modelagem molecular e regressão linear múltipla para se construir modelos de predição de atividade inibitória ( $pIC_{50}$ ) para séries de compostos presentes na literatura com dados de inibição sobre a GyrB da DNA girase de *M. smegmatis*. Foram obtidos bons modelos, verificados através de validação interna com o método de validação cruzada LOO (Leave One Out). Os resultados da validação cruzada foram expressos pelo coeficiente de correlação da validação cruzada ( $Q^2$ ) e pelo desvio-padrão da validação cruzada ( $S_{PRESS}$ ). As estatísticas de previsão do modelo são expressas pelo coeficiente de correlação múltipla  $R^2_{EXT}$  e pela raiz quadrada média do erro de previsão (RMSEP).

## Abstract

Tuberculosis (TB) is a disease with high morbidity and mortality. In 2019, approximately 10 million people developed the disease, of which about 1.4 million people died. There are effective antibiotics against the bacteria that cause tuberculosis, *Mycobacterium tuberculosis* (MTB), but the increased incidence of TB cases resistant to these drugs has required efforts to discover new drugs capable of fighting the disease. Topoisomerases are enzymes that maintain DNA topology during replication, transcription and recombination. DNA gyrase is the only type II topoisomerase present in MTB, which is why it is an interesting target to be explored for the design of new drugs against TB. DNA gyrase is composed of two subunits, GyrA and GyrB and this project aims to use groups of compounds from the literature with activity data recorded on the GyrB subunit of mycobacterial DNA gyrase, in order to develop a model for the prediction of activity of MTB DNA gyrase inhibitors to be applied in future virtual screening procedures. To achieve this goal, molecular modeling and multiple linear regression methods were combined to build prediction models of inhibitory activity ( $pIC_{50}$ ) for series of compounds present in the literature with inhibitory data on the inhibition of GyrB of *M. smegmatis* DNA gyrase. Good models were obtained, verified through internal validation with the LOO (Leave One Out) cross validation method. The results of the cross-validation were expressed by the correlation coefficient of the cross-validation ( $Q^2$ ) and the standard deviation of the cross-validation ( $S_{PRESS}$ ). The model's prediction statistics are expressed by the multiple correlation coefficient  $R^2_{EXT}$  and the root mean square of the prediction error (RMSEP).



## Lista de Figuras

<b>Figura 1:</b> Modelo para superenrolamento negativo DNA girase. ....	15
<b>Figura 2:</b> Exemplos de heterociclos.....	17
<b>Figura 3:</b> Esquema simplificado do processo de planejamento e desenvolvimento de um novo fármaco.....	19
<b>Figura 4:</b> Identificação, seleção e otimização de moléculas bioativas (fase pré-clínica).....	20
<b>Figura 5:</b> Esquema representando a geração de um modelo de QSAR/QSPR.....	29
<b>Figura 6:</b> Esquema com as etapas pertinentes para a construção de um Modelo QSAR.. ....	41
<b>Figura 7:</b> estrutura geral da série dos compostos do 1-28 usadas neste estudo (vide Tabela 4 para a relação dos substituintes R).....	42
<b>Figura 8:</b> estrutura geral da série de compostos do <b>29-76</b> usadas neste estudo (vide Tabela 5 para a relação dos substituintes X, R <sub>1</sub> e R <sub>2</sub> ). ....	43
<b>Figura 9:</b> estrutura geral da série de compostos <b>77-125</b> usadas neste estudo (vide Tabela 6 para a relação dos substituintes R <sub>1</sub> , R <sub>2</sub> e R <sub>3</sub> ).....	44
<b>Figura 10:</b> imagem do ligante (1) no interior da estrutura 4B6C, gerada com o programa Poseview, em que são mostradas suas principais interações com resíduos de aminoácidos do sítio. ....	50
<b>Figura 11:</b> gráfico resíduos vs valores estimados, que avalia um pressuposto de linearidade, para isso, a linha vermelha deve estar aproximadamente na horizontal. ....	60
<b>Figura 12:</b> gráfico para detectar observações influentes do modelo 2. Resíduos estudantizados serão mais eficazes para detectar observações Y periféricas do que resíduos padronizados.....	61
<b>Figura 13:</b> ligações de hidrogênio que ocorrem no composto <b>91</b> . As ligações de hidrogênio são representadas por linhas tracejadas em verde. Os átomos de C da estrutura do ligante estão em verde e os dos resíduos da enzima estão em cinza claro. ....	68

## Lista de Tabelas

<b>Tabela 1:</b> Mecanismos da resistência de <i>M. tuberculosis</i> aos fármacos. .....	14
<b>Tabela 2:</b> Matriz genérica dos dados .....	30
<b>Tabela 3:</b> métodos apropriados para criação de modelos de acordo com o tipo de x e y.....	35
<b>Tabela 4:</b> Substituintes (vide Figura 7) para a série de compostos do 1-28 usados neste estudo. ....	42
<b>Tabela 5 :</b> Substituintes (vide Figura 8) para a série de compostos do <b>29-76</b> usados neste estudo. ....	44
<b>Tabela 6:</b> Substituintes (vide Figura 9) para a série de compostos <b>77-125</b> usados neste estudo. ....	45
<b>Tabela 7:</b> Resultados do processo de redocagem para as quatro funções disponíveis no programa GOLD 5.7.3.. ....	56
<b>Tabela 8:</b> Comparação dos parâmetros estatísticos dos modelos 2 e 3. .....	62
<b>Tabela 9:</b> parâmetros estatísticos dos modelos 5 e 6 .....	64
<b>Tabela 10:</b> Análise de resíduos dos modelos 5 e 6.....	64
<b>Tabela 11:</b> Comparação dos parâmetros estatísticos dos modelos 3, 4 e 6 .....	65
<b>Tabela A 1:</b> dados da atividade biológica $pIC_{50}$ de acordo com as estruturas. Seguidos dos descritores HOMO, LUMO e Entalpia de interação..	84
<b>Tabela A 2:</b> Descritores obtidos através dos cálculos de docagem com a função ChemScore.....	87
<b>Tabela A 3:</b> Descritores obtidos através dos cálculos de docagem com a função ASP. ....	90
<b>Tabela A 4:</b> Descritores obtidos através dos cálculos de docagem com a função ChemPLP .....	93
<b>Tabela A 5:</b> Descritores obtidos através dos cálculos de docagem com a função GoldScore.....	96
<b>Tabela A 6:</b> Descritores obtidos através dos cálculos de redocagem com a função GooldScore e ChemPLP .....	99

## Lista de abreviaturas e siglas

ANVISA	Agência Nacional de Vigilância Sanitária
ASP	<i>Astex Statistical Potential</i>
CCDC	<i>Cambridge Crystallographic Data Centre</i>
DE.tors	peso do potencial de torção do ligante
EMA	<i>European Medicine Agency</i>
Esolv	energia de solvatação
FDA	<i>Food and Drug Administration</i>
GOLD	<i>Genetic Optimization for Ligand Docking</i>
HIV	Vírus da imunodeficiência humana
HOMO	orbital molecular ocupado de maior energia
IC50	concentração do inibidor necessária para inibir <i>in vitro</i> a atividade enzimática em 50%)
KI	dissociação do complexo enzima-inibidor
LBDD	Ligand Based Drug Design
LUMO	orbital molecular não ocupado de menor energia
MMFF	<i>Merck Molecular Force Field</i>
MNDO	Modified Neglect of Diatomic Overlap
MOPAC	<i>Molecular Orbital Package</i>
OMS	Organização Mundial de Saúde
PCR	regressão das componentes principais
PDB	<i>Protein Data Bank</i>
PLP	<i>Piecewise Linear Potential</i>
PLS	regressão dos mínimos quadrados parciais
PM3	Parametric Method 3
PM6	Parametric Method 6
PM7	Parametric Method 7
Q <sup>2</sup>	coeficiente de correlação da validação cruzada
QSAR	<i>Quantitative Structure-activity Relationship</i>
R <sup>2</sup>	Coeficiente de correlação ao quadrado
RLM	regressão linear múltipla
RMSD	Root-Mean-Squared Deviation
RMSEP	raiz quadrada média do erro de previsão
S.Hbond	termo da ligação de hidrogênio entre a proteína
S.int	soma dos termos de torção interna e de van der Waals interno
SBDD	Structure Based Drug Design
SEP	superfície de energia potencial
Spress	desvio-padrão da validação cruzada
SUS	Sistema Único de Saúde
TB	Tuberculose
TB-MDR	Tuberculose multirresistente
TB-XDR	Tuberculose extensivamente resistente

## Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>11</b>
1.2	Epidemiologia	12
1.3	<i>Mycobacterium tuberculosis</i>	13
1.3.1	DNA girase	14
1.4	Compostos heterocíclicos	16
1.5	Planejamento e desenvolvimento de novos fármacos	18
1.6	Planejamento racional de novos fármacos	21
1.7	Modelagem Molecular	22
1.8	Docagem	25
1.9	Relações quantitativas entre estrutura química e atividade/propriedade (QSAR)	28
1.9.1	Preparo do conjunto de dados	31
1.9.2	Detecção de <i>outliers</i>	33
1.9.3	Métodos de seleção de variáveis	34
1.9.4	Métodos de calibração	35
1.9.4.1	Regressão linear	36
<b>2</b>	<b>Objetivo</b>	<b>39</b>
<b>3</b>	<b>Metodologia</b>	<b>40</b>
3.1	Seleção dos compostos	41
3.2	Construção dos compostos	48
3.3	Docagem molecular	49
3.4	Cálculos das entalpias de interação	51
3.5	Determinação da matriz de dados	52
3.6	Pré-tratamento e análise dos dados	53
3.7	Seleção das variáveis	53

3.8	Estimativa dos modelos .....	54
3.9	Análise dos resíduos.....	54
3.10	Validação dos modelos.....	54
4	Discussão de resultados.....	56
4.1	Docagem molecular.....	56
4.2	Análise da matriz de covariância e Seleção de variáveis.....	57
4.3	Elaboração dos modelos de inibição da subunidade GyrB da DNA girase de <i>M. smegmatis</i> .....	59
4.3.1	Estimativa de modelo com a série de compostos 1-14.....	59
4.3.2	Estimativa de modelo com a série de compostos 77-99.....	60
4.3.3	Estimativa de modelo com a série de compostos 100-124.....	62
4.3.4	Estimativa de modelo com as séries mistas de compostos.....	63
4.4	Análise comparativa dos modelos quanto às avaliações estatísticas ..	64
4.5	Análise dos termos envolvidos nos modelos .....	65
4.5.1	Energia dos orbitais LUMO.....	65
4.5.2	Entalpia de interação $\Delta H_{int}$ e entropia de interação $\Delta S_{int}$ .....	66
4.5.3	Ligação de hidrogênio <i>S.hbond</i> e interação de <i>van der Waals</i> <i>S.vdwext</i> entre ligante e da proteína.....	67
5	Conclusão .....	69
6	Referências Bibliográficas.....	70
7	Anexos .....	84

## 1 INTRODUÇÃO

A tuberculose (TB) é uma doença infectocontagiosa causada principalmente pela bactéria *Mycobacterium tuberculosis*, o bacilo de *Koch*, transmitida de pessoa a pessoa através de gotículas de aerossóis. Os homens são 2 vezes mais susceptíveis à essa doença do que as mulheres e fatores como pobreza, alcoolismo e desnutrição tornam os indivíduos mais vulneráveis (LINDOSO E LINDOSO, 2009).

O tratamento da TB, assegurado de forma gratuita no Brasil pelo Sistema Único de Saúde (SUS), consiste em dois meses de uso de um coquetel de medicamentos contendo os antibióticos rifampicina, isoniazida, pirazinamida e etambutol (HOAGLAND *et al.*, 2016). Após esses dois meses, a doença se torna latente no organismo, sendo necessário continuar o tratamento com a rifampicina e a isoniazida. Em média, o tempo total de tratamento varia de 6 a 9 meses quando a bactéria responde bem aos principais antibióticos e o indivíduo não interrompe o tratamento.

Essa quimioterapia é capaz de curar quase a totalidade dos casos; entretanto, alguns fatores impedem o sucesso desse tratamento. Um deles é a resistência do bacilo de Koch a um ou mais agentes quimioterápicos (CAMPOS,1999). Nesses casos, a doença passa a ser denominada como tuberculose multirresistente (TB-MDR, do inglês *multidrug-resistant*) (CAMPOS,1999). Existem casos em que a TB-MDR apresenta resistência adicional às fluoro-quinolonas e a pelo menos um dos três medicamentos injetáveis de segunda escolha (amicacina, canamicina ou capreomicina), passando a ser considerada tuberculose extensivamente resistente a medicamentos (TB-XDR, do inglês *extensively drug-resistant*) (CAMPOS,1999). Existe ainda registros da TB totalmente resistente a medicamentos (TB-TDR, do inglês *totally drug-resistant*) em países da Europa Oriental, Índia, China e da África (KUMAR E ABUBAKAR, 2015).

A resistência bacteriana é um problema que envolve fatores biológicos como a existência de mutantes resistentes em cepas selvagens e/ou uma provável seleção natural, ou seja, é o resultado da acumulação de diferentes mutações independentes e não é fruto de um único evento (CAMPOS,1999). O surgimento de cepas resistentes de *M. tuberculosis* é um problema grave pois, como existem poucos fármacos efetivos

disponíveis, a TB pode se tornar uma doença intratável. Por isso, sob o ponto de vista terapêutico, é importante que médicos e pacientes sigam o tratamento de forma adequada para evitar possíveis disseminações de bacilos multirresistentes (FERNANDES, *et al.*, 2017).

## **1.2 Epidemiologia**

A Organização Mundial da Saúde (OMS) considera a TB como uma epidemia global e, no seu plano de estratégia para o fim da TB (*End TB*), alerta sobre possibilidade de uma epidemia global também de TB-MDR (WHO, 2015). Por isso, em setembro de 2018, chefes de Estado se reuniram na Assembleia Geral da Organização das Nações Unidas (ONU) para traçar metas globais de prevenção e tratamento da TB.

Segundo o diretor-geral da OMS, Tedros Adhanom Ghebreyesus, “O número de pessoas tratadas para TB cresceu desde a Assembleia Geral da ONU, com mais de 14 milhões de pessoas atendidas com tratamento de TB entre 2018 e 2019. O número de pessoas recebendo tratamento preventivo de TB quadruplicou desde 2015, de 1 milhão em 2015 para mais de 4 milhões em 2019”. O relatório Global OMS de 2020 sobre a TB destaca a redução da incidência e das mortes por tuberculose, mas não o suficiente para atingir as metas globais eliminação da tuberculose até o ano de 2030; mostra que, em 2019, cerca de 10 milhões de pessoas desenvolveram a doença e que aproximadamente 1,4 milhão de pessoas morreram devido à TB. No entanto, estima-se que cerca de 3 milhões não foram diagnosticadas ou não foram oficialmente notificadas às autoridades nacionais.

Este mesmo relatório mostra uma preocupação com o impacto que a pandemia Covid-19 pode ocasionar no progresso da redução global da TB. Isso porque, em muitos países, recursos humanos, financeiros e outros foram realocados da TB para a resposta à Covid-19, afetando o diagnóstico de pessoas com TB, ocasionando uma redução de aproximadamente 25% em países como a Índia, Filipinas e Indonésia nos primeiros 6 meses de 2020 em comparação com o mesmo período em 2019.

A OMS alerta também sobre a incidência de casos de pessoas com TB resistente a medicamentos. Cerca de 465 mil pessoas foram diagnosticadas com TB

resistente aos medicamentos em 2019 e, destas, mais de 60% não conseguiram obter acesso ao tratamento.

O plano de estratégia para combate a TB e suas variantes tem como principais recomendações as intervenções socioeconômicas, assistência e prevenção ao paciente, pesquisa e inovação (WHO, 2015). Um dos objetivos da pesquisa e inovação nessa área é a busca por novos medicamentos capazes de combater a TB-MDR em um período menor, com baixo potencial de interações medicamentosas, baixo custo de produção e baixa toxicidade (HOAGLAND *et al.*,2016).

### **1.3 *Mycobacterium tuberculosis***

A bactéria *M. tuberculosis* é um patógeno intracelular facultativo (se multiplica em meio extracelular produzindo a fagocitose) estritamente aeróbico (necessita de oxigênio para produzir energia e se reproduzir) (SUTORMIN, 2018; BUSH *et al.*, 2014). Esse microrganismo possui na parede celular uma camada de peptidoglicano e lipídeos livres de alto teor lipídico, em geral ácidos micólicos (FORRELAND *et al.*, 2013; LIMA *et al.*, 2001). Esses ácidos formam uma camada cerosa que dificulta o fluxo de nutrientes para o interior da célula e, por isso, a taxa de crescimento do *M. tuberculosis* ocorre de maneira lenta (SUTORMIN, 2018). Além disso, essa camada é ácido-alcool resistente (GROISMAN E OCHMAN 1997).

Esse complexo de ácidos micólicos é covalentemente ligado à infraestrutura, ou núcleo, o heteropolissacarídeo arabinogalactano e peptidoglicano (MIKUSOVÀ,1995). Devido a essa complexidade, a parede celular é um excelente alvo terapêutico (SMITH, 2003). A isoniazida e o etambutol são exemplos de fármacos que afetam a biossíntese do ácido micólico e do arabinano, respectivamente (MIKUSOVÀ,1995). Entretanto, existem cepas de *M. tuberculosis* que são resistentes a esses fármacos, uma decorrência de uma série de fatores biológicos como, por exemplo, a modificação espontânea de cromossomos de genes de alvos dos fármacos, um processo de seleção natural (CAMPOS,1999). O que se sabe até o momento é que, no caso da *M. tuberculosis*, fatores como a redução da disponibilidade relativa do fármaco por superprodução do alvo e alteração do alvo por



mutação fazem parte do mecanismo de resistência aos agentes antimicrobianos (CAMPOS,1999).

A Tabela 1 mostra exemplos e a frequência com que o alvo molecular (gene) em questão apresenta resistência ao medicamento associado.

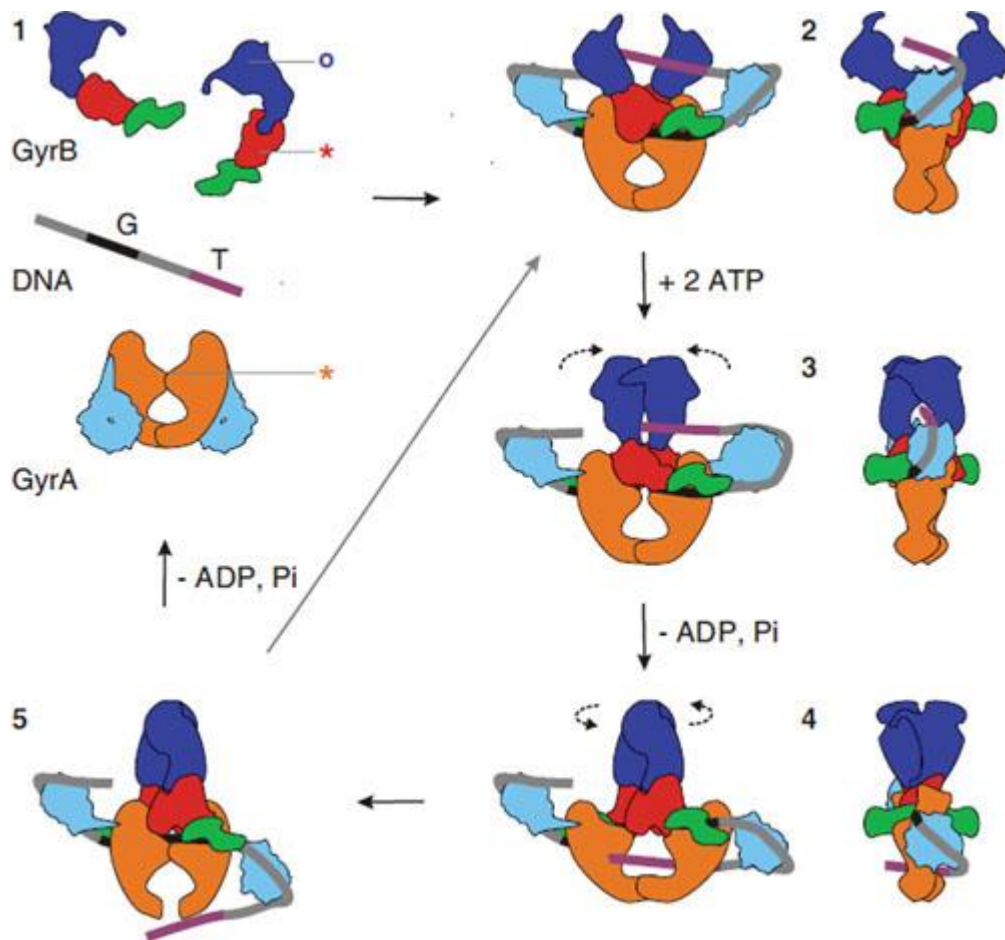
**Tabela 1:** Mecanismos da resistência de *M. tuberculosis* aos fármacos. Fonte: Adaptado de (CAMPOS, 1999; ROSSETI et al., 2002).

<b>Fármacos</b>	<b>Mecanismos da resistência</b>	<b>Genes envolvidos com a resistência</b>	<b>Frequência de mutações em <i>M. tuberculosis</i> resistente (%)</b>
Isoniazida	Inibição da síntese do ácido micólico.	KatG InhA MabA AhpC	42-58 21-34 21-28 10-15
Rifampicina Pirazinamida	Inibição da Transcrição. Inibição da síntese de ácidos graxos	RpobB PnacA	96-100 72-97
Etambutol	Inibição da síntese do arabinogalactano.	EmbCAB	47-65
Estreptomicina	Inibição da síntese proteica.	RpsL Rrs(16S RNA)	52-59 8-21
Fluoraquinolonas	inibição da enzima DNA girase	girA	75-94

Existem casos em que as bactérias apresentam multirresistência (MDR), ou seja, apresenta resistência a mais de um fármaco simultaneamente (ROSSETI *et al.*, 2002). Esses casos de MDR têm feito com que a comunidade científica busque fármacos capazes de atuar em outros domínios dos alvos terapêuticos, como por exemplo, o domínio GyraB da enzima da DNA girase, que desempenha um papel essencial no processo de replicação do DNA bacteriano (BUSH *et al.*, 2014).

### 1.3.1 DNA girase

A enzima DNA girase (Figura 1) pertence à classe das topoisomerases do tipo IIA, presente apenas em células bacterianas (SUTORMIN, 2018; AUBRY, 2006; REECE E MAXWELL,1991). Possui uma única cópia, responsável por replicar o DNA através da quebra e reunião do DNA no domínio girase A (GyrA) e hidrólise de ATP no domínio girase B (GyrB), uma ATPase (KASHYAP *et al.*, 2018; CHAUDHARI *et al.*, 2016).



**Figura 1:** Modelo para superenrolamento negativo DNA girase. O domínio GyrB está em vermelho, azul escuro e verde. O domínio GyrA C-terminal azul claro. Os segmentos de DNA G e T são coloridos em preto e roxo, respectivamente. 1 subunidades e DNA em seus estados livres propostos em solução. As estrelas indicam os resíduos do sítio ativo para clivagem de DNA, e o círculo indica a bolsa de ligação de ATP. 2, o segmento G liga-se através de GyrA na interface do dímero, e o domínio GyrA C-terminal envolve o DNA para apresentar o segmento T em um cruzamento positivo. 3, ATP é ligado, o que fecha o grampo GyrB capturando o segmento T, e o segmento G é transitoriamente clivado. 4, a hidrólise de uma molécula de ATP permite que GyrB gire, a porta do DNA se alargue e o segmento T seja transportado através do segmento G clivado. 5, O segmento T sai pela porta C, e o segmento G é religado. A hidrólise da molécula de ATP remanescente redefine a enzima. Os desenhos à direita mostram a vista lateral para as ilustrações 2 a 4. Fonte: Retirado de (BUSH *et al.*, 2014).

Em condições favoráveis a DNA girase (Figura 1) efetua a condensação, manutenção da estrutura e segregação dos cromossomos filhos após a replicação do DNA (BUSH *et al.*, 2014). Para que isso ocorra, uma Seção de DNA denominada porta (gate) ou segmento G liga-se ao dímero superior do domínio GyrA N-terminal (SUTORMIN, 2018; BUSH *et al.*, 2014; CHAMPOUX, 2001). Essa ligação resulta em um segmento de transporte (T) adjacente ao segmento G. No domínio GyrB ocorre o corte nas fitas de DNA (clivagem) que separa o segmento G permitindo que apenas o

segmento T passe pela cavidade formada pelos domínios do domínio GyrA *N*-terminal (SUTORMIN, 2018; BUSH *et al.*, 2014; CHAMPOUX, 2001). O local de clivagem é seguido por religação com hidrólise de ATP para conduzir os ciclos catalíticos e produzir uma superespiral negativa no DNA (CHAUDHARI *et al.*, 2016). O segmento G é religado e o segmento T é liberado através da porta inferior do domínio GyrA *N*-terminal (SUTORMIN, 2018; CHAMPOUX, 2001).

As subunidades GyrA e GyrB, responsáveis pela clivagem e religação do DNA, desempenham um papel importante para a sobrevivência celular pois, sem essas enzimas, as células não são capazes de “desembaraçar” seus cromossomos filhos (SUTORMIN, 2018; BUSH *et al.*, 2014; CHAMPOUX, 2001). Uma inibição enzimática em uma dessas subunidades resulta em um número de mortes celulares considerável, visto que não existem outros mecanismos capazes de desempenhar essa função. Como se trata de uma enzima essencial para a viabilidade bacteriana, é um importante alvo terapêutico de agentes antibióticos (SUTORMIN, 2018; FERNANDES, 2017).

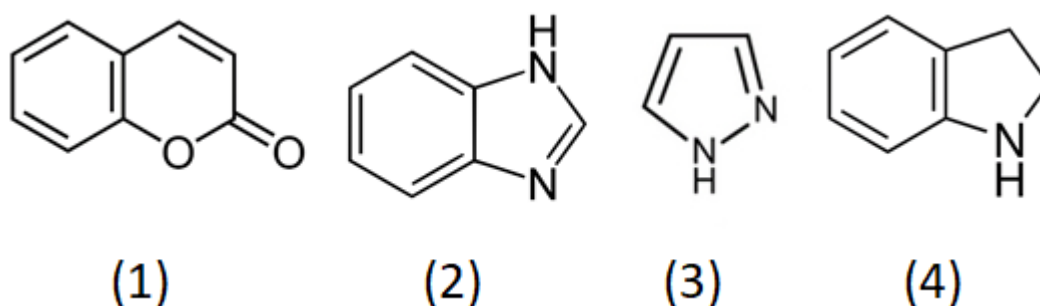
Os derivados de fluoroquinolona têm a capacidade de atuar no processo de replicação na subunidade GyrA. Por isso, têm sido utilizados no tratamento da TB-MDR. Entretanto, devido a mutações no domínio GyrA e a utilização deste tipo de bactericida em outras infecções, cepas de *M. tuberculosis* resistente a esse tipo de fármaco têm surgido (CAMPOS, 1999). Isso tem feito a comunidade científica direcionar as pesquisas para o domínio GyrB (KASHYAP *et al.*, 2018). Compostos como a novobiocina e a coumermicina atuam no domínio GyrB, mas demandam atenção devido às suas propriedades toxicológicas e baixa ação farmacológica (FERNANDES *et al.*, 2017).

#### **1.4 Compostos heterocíclicos**

Compostos heterocíclicos são estruturas cíclicas constituídas de carbono que têm como membros do anel pelo menos dois elementos diferentes (SANTOS, 2017; IUPAC, 1997). Esses compostos são relevantes para indústria farmacêutica, tanto que 62% dos fármacos são compostos com esse tipo de estrutura. A justificativa para o uso desses compostos é ocasionada pelas atividades biológicas que lhes são

atribuídas, assim como a diversidade estrutural e a sua origem que pode de uma fonte natural ou sintética (SANTOS, 2017).

Existem muitos heterociclos conhecidos com ações antifúngica, anti-inflamatória e antibiótica, entre outras (MELO *et al.*, 2006). Especificamente, nesse trabalho, o interesse se dá por alguns compostos que possuem algum relato sobre o domínio GyrB da DNA girase como, por exemplo, os derivados de: cumarina (1, Figura 2); benzimidazol (2, Figura 2); pirazol (3, Figura 2); e indolina (4, Figura 2) (RENUKA *et al.*, 2014; GUPTA *et al.*, 1998).



**Figura 2:** Exemplos de heterociclos.

- Cumarina

As cumarinas costumam estar presentes em vegetais, mas já foram encontradas em fungos e bactérias (RADÜNZ, 2012). Estruturalmente são lactonas do ácido o-hidroxicinâmico (2H-1-benzopirano-2-ona), caracterizadas, na sua forma mais simples, como 1,2-benzopirona (1, Figura 2). (FRANCO *et al.* 2021; RADÜNZ, 2012).

- Benzimidazol

O benzimidazol (2, Figura 2) é considerado um isómero de nucleotídeo, devido a fusão do benzeno com o imidazol. Possui um potencial farmacológico associado a baixa toxicidade (OREN *et al.*, 1998). Está presente em uma das unidades complexadas ao cobalto da vitamina B12 (ANDRES *et al.*, 2004).

- Pirazol

Os pirazóis (3, Figura 2) são compostos aromáticos de anéis de cinco membros com dois átomos de nitrogênio nas posições 1 e 2 (GUPTA *et al.*, 1998). Derivados de pirazol costumam ser atraentes em pesquisas farmacêuticas, pois apresentam um amplo espectro de atividades, incluindo redutora do colesterol, anti-inflamatória, anticâncer e antidepressiva (DESPOTOPOULOU *et al.*, 2009).

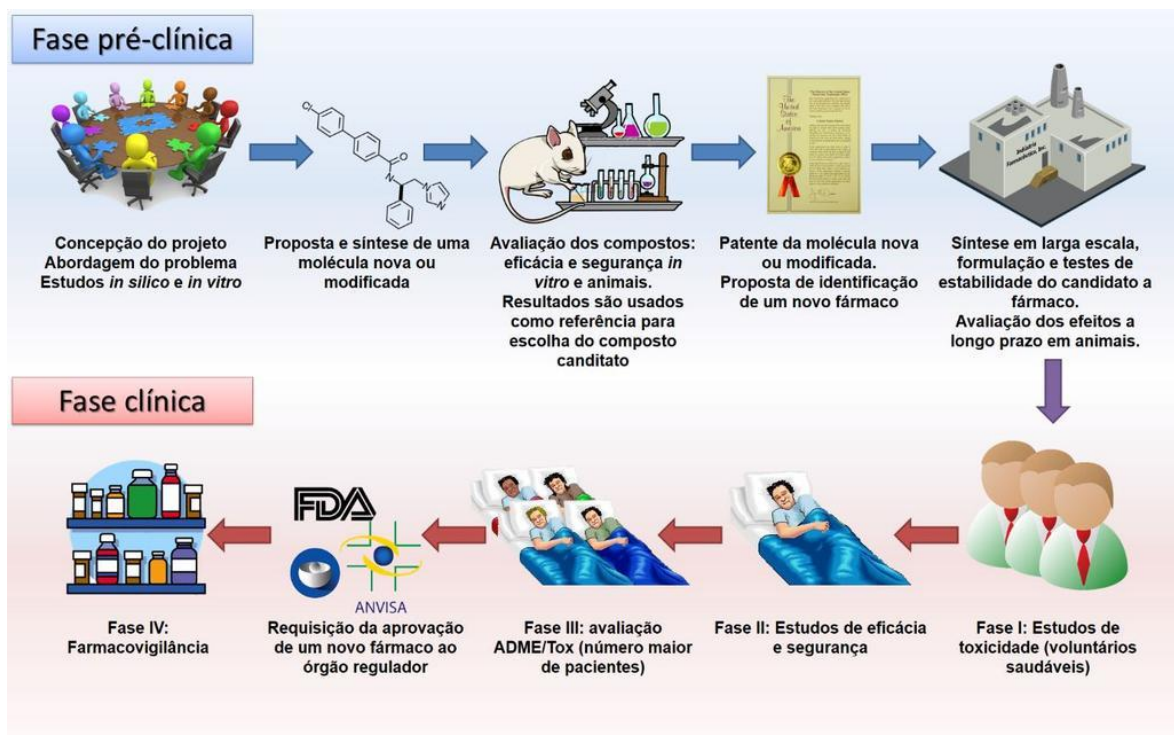
- Indolina

A indolina (4, Figura 2) é constituída pela fusão dos anéis benzeno e pirrol; essa junção traz propriedades características devido a subunidade aromática rica em elétrons. Uma delas é a contribuição para fornecer um ambiente hidrofóbico. Um derivado da indolina é o indol, com uma insaturação entre os carbonos 2 e 3, presente nas proteínas na forma do aminoácido triptofano (DE SA ALVES *et al.*, 2009).

### **1.5 Planejamento e desenvolvimento de novos fármacos**

O processo de introdução de um novo fármaco no mercado é complexo e demanda diversas etapas (Figura 3). Esse processo envolve a integração de várias áreas estratégicas relacionadas à inovação, conhecimento, tecnologia, gerenciamento e altos investimentos em pesquisa, desenvolvimento e inovação (PD&I), (DIMASI *et al.*, 2003; FERREIRA *et al.*, 2011.) Os custos envolvidos em todas as etapas de desenvolvimento de novos fármacos variaram de 1,3 a 1,5 bilhão de dólares. Esse processo requer um tempo médio de dez a doze anos (BHARTI *et al.*, 2019).

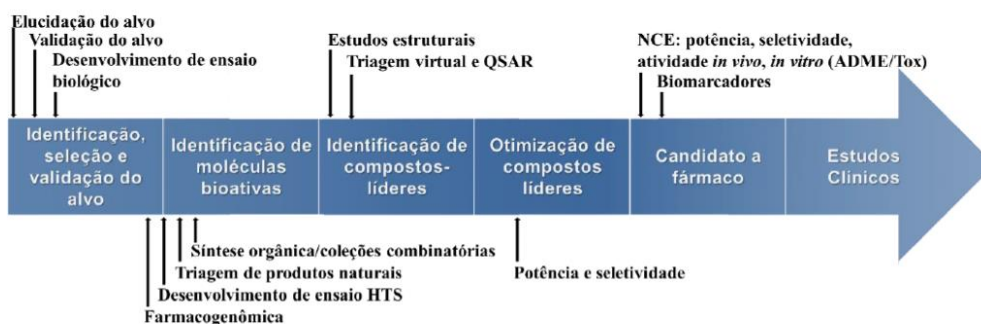
Na Figura 3, o processo está simplificado para fins didáticos. Na prática, uma rede com diferentes fases do processo ocorre concomitantemente, visto que, da concepção do projeto até a introdução do produto no mercado, diferentes compostos são avaliados e modificações estruturais são propostas com o objetivo de se encontrar uma estrutura ideal (BAXTER *et al.*, 2013).



**Figura 3:** Esquema simplificado do processo de planejamento e desenvolvimento de um novo fármaco. Fonte: retirado de (LOMBARDINO E LOWE, 2004).

Nas etapas iniciais, as pesquisas se concentram: na identificação e validação de alvos biológicos, geralmente enzimas importantes em processos fisiopatológicos, no planejamento, identificação e otimização de compostos capazes de modular o alvo biológico estudado em ensaios *in vitro* (LOMBARDINO E LOWE, 2004).

A fase pré-clínica (Figura 4) envolve uma sequência hierárquica de etapas em que diversas estratégias de Química Medicinal são empregadas, desde a identificação e validação de alvos moleculares (proteínas), identificação inicial de substâncias bioativas (hits), identificação e otimização de compostos líderes (ou protótipos), elucidação do mecanismo de ação, avaliação e otimização de propriedades farmacocinéticas até a avaliação da eficácia e segurança (JORGENSEN, 2009).



**Figura 4:** Identificação, seleção e otimização de moléculas bioativas (fase pré-clínica). Fonte: (modificado GUIDO, 2008).

As moléculas aprovadas nos ensaios pré-clínicos são submetidas aos órgãos reguladores, requerendo autorização para iniciar os estudos em seres humanos. Nos Estados Unidos, o órgão responsável é a *Food and Drug Administration* (FDA) (PHARMA, 2013), no Brasil, a Agência Nacional de Vigilância Sanitária (ANVISA) (PIOVESAN E LABRA, 2007) e na Europa, a *European Medicine Agency* (EMA) (REGNSTROM *et al.*, 2010).

Para garantir a segurança e eficácia de um candidato a fármaco devem ocorrer estudos em seres humanos, assim chamada a fase clínica. Inicialmente doses subterapêuticas são administradas em um pequeno número de indivíduos (10 a 15) para coletar informações farmacocinéticas e farmacodinâmicas do composto no organismo humano (MENDOZA, 2011). Em seguida, o candidato a fármaco é administrado em um grupo ainda pequeno de voluntários (20 a 100) para avaliar a toxicidade. A próxima etapa, realiza estudos em pacientes que apresentam a doença (100 a 300 indivíduos) para se avaliar a dosagem, eficácia e segurança da composição. Após esses processos de testagem iniciais, se faz necessário aumentar a credibilidade dos estudos, por isso, testes são aplicados em um grupo maior de indivíduos (300 a  $\geq 3000$ ), aleatorizados e distribuídos em regiões geográficas distintas. Essa é a fase mais onerosa, mais demorada e mais difícil de projetar e desenvolver, especialmente em terapias para doenças crônicas (FRIEDMAN *et al.*, 2010).

Quando os ensaios são satisfatórios, os resultados juntamente com uma série de documentos, contendo uma descrição compreensiva dos métodos e resultados dos

estudos realizados em humanos e animais, processos de produção, forma farmacêutica e prazo de validade, são encaminhados para revisão dos órgãos regulatórios, onde se espera que seja dada ao investidor a autorização de comercializar o novo fármaco (MCGEE, 2006). No mercado, o novo fármaco passa pela fase de farmacovigilância, na qual é realizada a detecção, avaliação e prevenção de efeitos adversos não detectados anteriormente através da notificação espontânea e intensivo monitoramento das agências regulatórias e da própria indústria farmacêutica (HÄRMARK E VAN GROOTHEEST, 2008).

### **1.6 Planejamento racional de novos fármacos**

A busca por um composto bioativo capaz de interagir com uma biomacromolécula e assim causar um efeito biológico ou farmacológico é um processo bastante complexo (BARREIRO, 2001; SANT'ANNA, 2009). Uma das estratégias de busca é o planejamento racional de compostos bioativos, assim chamado porque é orientado por uma hipótese racional sobre o mecanismo de ação destes compostos (BARREIRO, 2001). A ação dessas moléculas bioativas tem seus efeitos associados a interações ou até reações químicas com estruturas macromoleculares presentes no sistema vivo, proteínas, na sua grande maioria (BARREIRO, 2001). Caso estas proteínas sejam receptores celulares, as moléculas bioativas são classificadas como agonistas ou antagonistas de receptores; no caso das enzimas, estas moléculas atuam como inibidores enzimáticos (SANT'ANNA, 2009).

O planejamento racional pode seguir por duas linhas de estratégias: planejamento de fármacos baseado na estrutura do receptor (SBDD – Structure Based Drug Design), quando a estrutura do alvo macromolecular ou do complexo ligante-receptor estão disponíveis (KROEMER, 2007; JORGENSEN, 2009) e planejamento de fármacos baseado na estrutura do ligante (LBDD – Ligand Based Drug Design), quando a estrutura do alvo não é conhecida, mas estruturas de ligantes bioativos são (COHEN, 1996).

Para começar um planejamento racional de um fármaco com a estratégia SBDD é necessário definir um alvo molecular, que pode ser uma proteína presente



em alguma bactéria, vírus, célula etc. Em seguida, deve-se fazer um estudo detalhado de sua estrutura tridimensional, determinada, em geral, por cristalografia de raios X ou espectroscopia de ressonância magnética nuclear (NMR), para identificar as características dos seus sítios de interação com o objetivo de determinar modos de interação de ligantes neles (KUNTZ, 1992; KROEMER, 2007). Essas estruturas são então utilizadas, por exemplo, para fazer um estudo de ancoramento ou docagem molecular (do inglês, *molecular docking*) para a previsão dos modos de interação de cada ligante no alvo. (HUANG *et al.*, 2010).

O planejamento racional de um fármaco baseado na estrutura do ligante (LBDD) utiliza métodos que buscam correlacionar informações obtidas de ligantes com a atividade biológica de compostos com informações estruturais e/ou propriedades físico-químicas destes. É utilizado quando o alvo não é conhecido ou quando a sua estrutura não pode ser obtida por métodos experimentais e nem modelada (COHEN, 1996). Parte do princípio de que moléculas com semelhanças estruturais poderão ter atividades biológicas parecidas (RODRIGUES *et al.*, 2012). Normalmente são escolhidas moléculas que já tenham informações de atividade, como o IC<sub>50</sub> (concentração do inibidor necessária para inibir *in vitro* a atividade enzimática em 50%) conhecidas e com faixas de atividades extensas (MACALINO *et al.*, 2015). Uma das principais abordagens do LBDD é o uso das relações quantitativas estrutura-atividade (QSAR, do inglês *Quantitative Structure-activity Relationship*) que investiga a relação entre as estruturas (ou propriedades associadas às estruturas) dos ligantes e seus efeitos biológicos ou farmacológicos (MACALINO *et al.*, 2015).

Existe também a possibilidade de se combinar os métodos SBDD e LBDD, o que pode garantir uma maior eficiência na busca de novos candidatos a fármacos, pois explora informações estruturais e químicas do alvo e de ligantes, levando à descoberta de protótipos mais promissores (LAVECCHIA E DI GIOVANNI, 2013).

## **1.7 Modelagem Molecular**

Atualmente diversos métodos computacionais podem ser utilizados como ferramentas do planejamento racional de compostos bioativos, sendo a modelagem

molecular uma delas. Trata-se de um conjunto de recursos matemáticos e computacionais que possibilita a análise de sistemas moleculares (SANT'ANNA, 2002). Pode ser usada em todos os estágios do planejamento de compostos candidatos a ensaios biológicos e testes clínicos, desde a descoberta de um protótipo até sua otimização, em termos da previsão de afinidade, especificidade, toxicidade e biodisponibilidade (OPREA, 2005).

Do ponto de vista do planejamento, é importante destacar que, quando se fala do uso da modelagem molecular, não se pretende chegar a uma molécula bioativa simplesmente através do uso de programas de computador. O processo de desenvolvimento dessas moléculas, devido à sua complexidade, envolve necessariamente o trabalho de uma equipe multidisciplinar, que emprega um grande conjunto de métodos computacionais de modo sistemático de forma a facilitar e otimizar o processo de desenvolvimento de compostos bioativos, em uma constante troca de informações com grupos de síntese química e avaliação da atividade destes compostos (SANT'ANNA, 2009).

Os recursos da modelagem molecular proporcionam informações relevantes a respeito de uma estrutura molecular como, por exemplo, a energia que determina sua estabilidade. Com essas informações, é possível identificar dentre uma série quais moléculas interagem melhor com a biomacromolécula e usar esta informação (RODRIGUES *et al.*, 2012).

Os programas de modelagem molecular utilizam abordagens da mecânica clássica, com os métodos mecânica molecular e dinâmica molecular, e da mecânica quântica, com os métodos *ab initio*, semi-empírico e teoria funcional de densidade (SANT'ANNA, 2009). De modo geral, a escolha entre estes métodos depende das propriedades que se deseja avaliar, da precisão desejada e da capacidade computacional disponível para a realização dos cálculos (SANT'ANNA, 2009).

Na sua forma mais simples, a mecânica molecular considera os átomos como se fossem esferas interligadas por uma força elástica. A energia total é determinada por uma série de funções de energia, como funções que dependem das distâncias de ligação, dos ângulos de ligação e dos ângulos diedros, entre outras (LILIAM *et al.*, 1998). A energia total em função dos parâmetros estruturais pode ser representada como uma superfície com máximos e mínimos locais de acordo com as conformações

das moléculas. Assim, utiliza-se um algoritmo de busca de mínimos para se encontrar estruturas de menor energia (SANT'ANNA, 2009). O modelo da mecânica molecular se torna possível pois os parâmetros associados a conjuntos de átomos permanecem razoavelmente constantes entre estruturas diferentes, desde que o tipo e a hibridação dos átomos envolvidos sejam os mesmos (BURKET, 1982).

Nos métodos de dinâmica molecular as moléculas são tratadas como uma coleção de átomos que pode ser descrita por forças newtonianas, em outras palavras, são tratadas como uma coleção de partículas mantidas unidas por forças harmônicas ou elásticas. Assim, o conjunto completo dos potenciais de interação entre as partículas formam um "campo de força" (GUNSTEREN e BERENDSEN, 1990). As equações representam os movimentos moleculares, trajetória, ao longo do tempo que pode ser usada para o estudo de propriedades que dependem do tempo, tais como a difusão, o dobramento de cadeias moleculares (como acontece no enovelamento de proteínas) e a distribuição de moléculas de solvente ao redor de um soluto, entre outros. Em um cálculo de dinâmica molecular, as moléculas, que têm certa energia cinética, podem superar barreiras de energia potencial, o que permite a exploração da superfície de energia potencial (SEP) na busca de outras conformações estáveis (SANT'ANNA, 2009).

A mecânica quântica parte do princípio de que partículas, como o elétron, também possuem comportamento ondulatório. Baseado nisso, Schrödinger desenvolveu a função de onda ( $\Psi$ ), que descreve o comportamento ondulatório dos núcleos e elétrons e possibilita o cálculo da energia  $E$  total através de operadores associados à energia cinética  $T$  e à energia potencial  $V$ , na chamada equação de Schrödinger, Eq. (1). O domínio é dado em função das coordenadas dos núcleos ( $R$ ) e das coordenadas dos elétrons ( $r$ ) (HOUSE, 2004).

$$[T + V]\Psi(R, r) = E \Psi(R, r) \quad (1)$$

Para simplificar o problema, a equação de Schrödinger é resolvida somente em função dos elétrons (SANT'ANNA, 2009). Para que isso seja possível, basta assumir

a imobilidade dos núcleos atômicos. A massa nuclear é várias vezes maior que a massa dos elétrons, fazendo com que os elétrons se ajustem quase que instantaneamente a qualquer mudança de posição nuclear. Esta manipulação é conhecida como aproximação de Born-Oppenheimer, que permite separar a equação de Schrödinger em duas: uma para a parte eletrônica e outra para a nuclear (BORN E OPPENHEIMER, 1927).

O método semi-empírico utiliza conhecimento prévio de parâmetros obtidos em dados experimentais para auxiliar na solução da equação de Schrödinger. Além disso, a equação de Schrödinger é aplicada apenas para os elétrons de valência, pois são os que mais participam das ligações químicas (ALCÁCER, 2007). Essa técnica ajuda a reduzir o tempo e o custo computacional, principalmente quando se faz necessário o uso de cálculos quânticos para obter propriedades termodinâmicas e eletrônicas de sistemas moleculares com muitos elétrons. Entretanto, esse método é restrito a sistemas que possuem apenas os elementos pertencentes ao pacote para quais foram estabelecidos os parâmetros (SANT'ANNA, 2009).

O programa MOPAC (do inglês *Molecular Orbital Package*) (Stewart Computational Chemistry) (MOPAC,2016) de domínio gratuito, abrange grande parte dos métodos semi-empíricos contendo parâmetros, desde os mais antigos, como o MNDO (*Modified Neglect of Diatomic Overlap*) (DEWAR e THIEL, 1977), MINDO/3, AM1 (Austin Model 1) (DEWAR, et al., 1985) e PM3 (Parametric Method 3) (STEWART, 1989; STEWART, 2004; STEWART, 2007), até os mais recentes, PM6 (e suas modificações) e PM7, (STEWART, 2007) (STEWART, 2013), que possuem parâmetros para 70 elementos da Tabela, inclusive metais de transição (SANT'ANNA, 2009).

## **1.8 Docagem**

A metodologia de docagem molecular (KUNTZ *et al.*, 1982) consiste em prever a orientação e a conformação preferenciais de uma molécula inserida em outra quando estas formam um complexo estável (MORGON E COUNTINHO, 2007). Em outras palavras, é o acoplamento molecular de compostos nos sítios de ligação de

macromoléculas, determinando os modos de ligação mais prováveis e estimando a afinidade de ligação.

Esses modos de ligações podem ocorrer de duas maneiras: primeiro, a molécula do ligante é considerada como uma estrutura flexível, que se liga à estrutura rígida da proteína, o que corresponde ao antigo modelo chave-fechadura (VERLI E BARREIRO, 2005); segundo, se admite que a proteína tem pelo menos alguns graus de liberdade conformacionais, para se adaptar ao ligante de modo a estabelecer uma baixa energia para o complexo (RODRIGUES *et al.*, 2012). A maioria dos programas de docagem molecular considera somente o ligante flexível, devido ao custo computacional; alguns, no entanto, levam em conta a flexibilidade ao menos parcial da proteína, o que torna o processo mais preciso, mas também mais lento (KROEMER, 2007).

Cada conformação e orientação do ligante, obtidas por algoritmos de busca, permite um encaixe diferenciado no sítio de ligação da biomacromolécula, os quais são classificados através de funções de pontuação. Com isso, cada modo de interação é quantificado para assim se estabelecer um ranqueamento destes (DEL CISTIA, 2010; RODRIGUES *et al.*, 2012).

Os algoritmos podem, por exemplo, realizar alterações aleatórias nos graus de liberdade do ligante, explorando diversas conformações e orientações possíveis, como é o caso dos algoritmos genéticos, embasados na teoria de seleção natural de Charles Darwin (TCHOUGRÉEFF, 2008). As informações geradas pelas conformações das moléculas são codificadas; assim, cada grau de liberdade do ligante é considerado um “gene” dos “cromossomos”. Uma série de operações genéticas, como mutações e cruzamentos, realizam diversas combinações de orientações/conformações iniciais dos ligantes, “cromossomos pais”, para formar novas orientações/conformações adotadas pelos ligantes, “cromossomos filhos”. Os “novos cromossomos” gerados são avaliados e, se forem mais favoráveis de acordo com a pontuação calculada por funções específicas, os indivíduos que os possuem substituem os indivíduos menos aptos da população inicial. O processo é então repetido até que se encontrem orientações/conformações com a melhor aptidão possível dentro de um número pré-determinado de gerações (BROOIJMANS E KUNTZ, 2003; DIAS E AZEVEDO Jr., 2008).

As informações das orientações/conformações, geradas pelos algoritmos de busca são então avaliadas pelas funções de pontuação, utilizadas para ordenar os ligantes para determinar a ordem de aptidão, dando uma estimativa da afinidade dos ligantes pelo alvo molecular. No entanto, deve-se considerar que muitas funções de pontuação não levam em conta alguns termos importantes para o processo de interação, como entropia e participação da água; assim, nem sempre se obtém uma correlação direta entre a pontuação e a afinidade dos compostos pela proteína (KITCHEN *et al.*, 2004).

Essas funções de pontuação podem ser estabelecidas com base em conhecimento (*knowledge-based scoring functions*), onde as energias de interação são obtidas através de análises estatísticas de interações de complexos conhecidos (GOHLKE *et al.*, 2000); campo de força (*force-field-based scoring functions*), nas quais se efetua o somatório da energia externa, gerada pela interação ligante/proteína, com a energia interna do ligante (aqui as energias são descritas em função das interações de Van der Waals e interações eletrostáticas (KITCHEN, *et al.*, 2004); e, por fim, empíricas (*empirical scoring functions*), quando se assume que a energia de interação pode ser calculada por uma soma de termos individuais baseados em dados experimentais (KITCHEN, *et al.*, 2004).

Existem vários programas disponíveis para realizar docagem e, nesse trabalho, foi utilizado o programa GOLD (*Genetic Optimization for Ligand Docking*) (JONES *et al.*, 1997), disponibilizado pelo *Cambridge Crystallographic Data Centre – CCDC* <https://www.ccdc.cam.ac.uk>. O GOLD utiliza o Algoritmo Genético para propor diversos modos de interação, ou poses, para os quais se estabelecem as pontuações com as funções escolhidas. O GOLD possui quatro funções de pontuação pré-definidas:

ASP (*Astex Statistical Potential*) baseada no conhecimento e gera potenciais estatísticos a partir da frequência de interação entre pares de átomos da proteína e do ligante que foram observadas experimentalmente em uma coleção de complexos ligante-proteína (MOOJI & VERDONK, 2005);

ChemPLP modela a complementariedade estérica entre ligante e proteína, além de levar em conta os ângulos e distâncias de interações de hidrogênio e interações envolvendo metais. A sigla PLP (*Piecewise Linear Potential*) representa um

potencial específico que modela a atração e repulsão de átomos diferentes de hidrogênio entre proteína e ligante e potenciais internos para os ligantes, que incluem um termo para colisões entre átomos e um termo torcional derivado do campo de força *Tripos* (KORB, 2009);

A função *Chemscore* estima a variação da energia livre total do sistema, com base em observações experimentais de 82 complexos proteína-ligante, quando ocorre a ligação do ligante com a proteína através da soma ponderada de uma série de termos de energia livre, que incluem interações lipofílicas, interações com metais e ligações de hidrogênio (ELDRIDGE *et al.*, 1997).

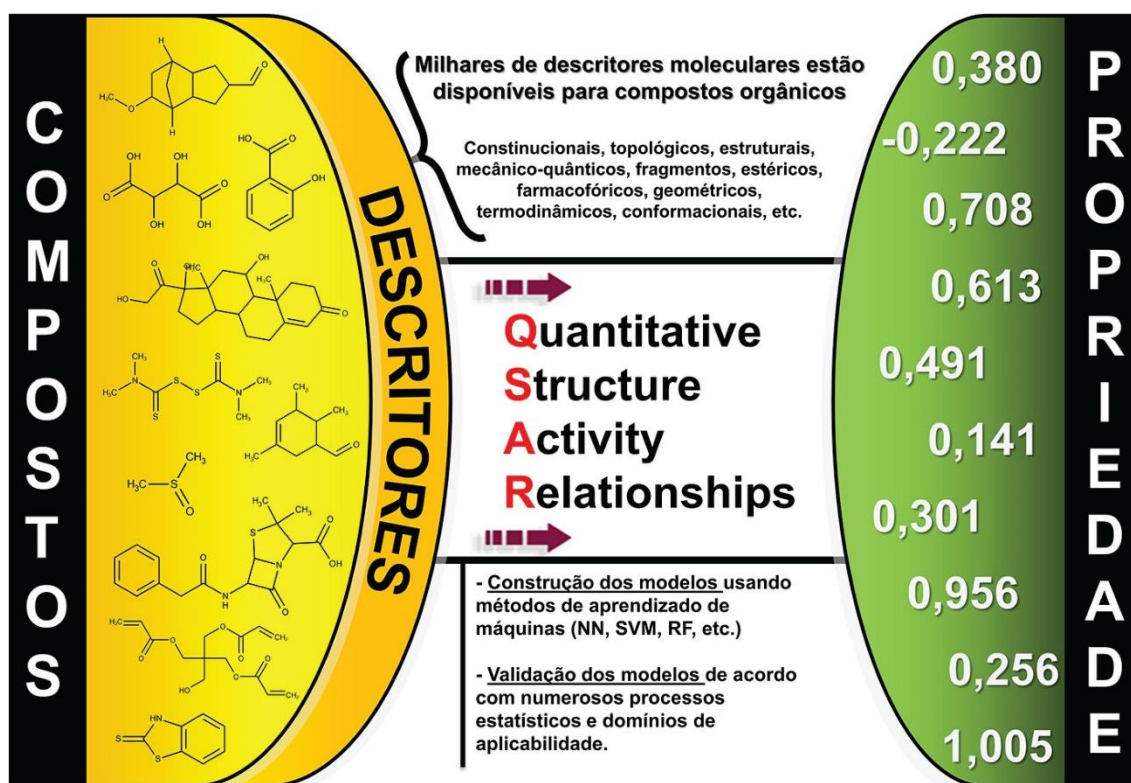
A função *Goldscore* considera parâmetros empíricos, por isso, calcula a soma de termos de energia para as interações de Van der Waals e de ligações de hidrogênio entre a proteína e o ligante, juntamente com termos de energia torcional e de interações de Van der Waals e de ligações de hidrogênio internas para os ligantes (JONES *et al.*, 1997).

Para validar a função mais apropriada para a docagem é necessário realizar a redocagem, um processo que usa uma ou mais estruturas cristalográficas de complexos proteína-ligante, de onde se extrai o ligante cocrystalizado e se insere novamente através da docagem. Com isso, calcula-se o desvio médio quadrático (RMSD, do inglês *Root Mean Square Deviation*) a partir da sobreposição dos dados experimentais com as poses obtidas. O resultado é considerado de boa qualidade quando o RMSD é menor que 2,0 Å (GOHLKE *et al.*, 2000).

### **1.9 Relações quantitativas entre estrutura química e atividade/propriedade (QSAR)**

A abordagem QSAR, do inglês *Quantitative structure–activity relationship*, pode ser aplicada na identificação racional de novos ligantes/protótipos com atividade/propriedade desejada, na otimização da atividade/propriedade e na identificação de compostos potencialmente perigosos em estágios preliminares do desenvolvimento (TROP SHA, 2010). Além disso, o QSAR, como as demais metodologias teóricas aplicadas no planejamento de fármacos, contribui para redução do custo do desenvolvimento do fármaco, reduz o número de animais utilizados em

ensaios experimentais e diminui resíduos que seriam descartados por compostos improváveis de terem sucesso (CRONIN, 2010).



**Figura 5:** Esquema representando a geração de um modelo de QSAR/QSPR. Fonte: retirado (ALVES et al., 2018).

A abordagem QSAR (Figura 5) pode ser descrita como um método matemático e estatístico de análise de dados para desenvolver modelos que estabelecem uma relação entre as atividades biológicas e as estruturas químicas de uma série de compostos semelhantes (FERREIRA *et al.*, 2002; ALVES *et al.*, 2018). O que diferencia esses compostos entre si, em geral, é a presença de um ou mais grupos substituintes em posições estabelecidas na estrutura química comum à série (FERREIRA *et al.*, 2002). Assim, dados experimentais da atividade biológica de compostos químicos são correlacionados com descritores moleculares (ALVES *et al.*, 2018). Essa análise deve ser capaz de explicar o fenômeno observado (atividade) e fazer projeções a partir de um modelo que compara a medida quantitativa da atividade Y com dados físico-químicos e estruturais de n variáveis descritivas X (FERREIRA *et al.*, 2002).



Os descritores são números obtidos através de procedimentos matemáticos e lógicos que utilizam informações químicas codificadas em uma representação simbólica de uma molécula em um número útil, ou o resultado de algum experimento padronizado (TODESCHINI E CONSONNI, 2000). Esses descritores podem representar: propriedades físico-químicas como massa molecular, refratividade molar, logP; representações bidimensionais (2D), como número de átomos, número de ligações, índices de conectividade; e propriedades tridimensionais (3D), que dependem da conformação das moléculas como volume de Van der Waals, área de superfície acessível ao solvente, entre outras (XUE E BAJORATH, 2000).

Os descritores e os dados atividade/propriedade são arranjados em uma matriz conforme a Tabela 2, onde os dados dos objetos estão dispostos nas linhas, a atividade/propriedade é armazenada na coluna *Y* e os descritores na matriz *X*. Transformações de dados podem ser aplicadas para alterar as distribuições dos valores das variáveis, por exemplo, para aproximá-los de uma distribuição normal (VARMUZA E FILZMOSER, 2009). A partir dessas informações, é possível construir um modelo categórico (ex., ativos/inativos, tóxico/não tóxico) ou quantitativo, na qual a propriedade estudada (*Y*) é descrita por uma função de um ou mais descritores (*X*), como por exemplo, o modelo linear de Hansch-Fujita Eq. (2), que foi um marco nos estudos de QSAR (KUBINY, 1993; HASCH e FUJITA, 1964), como a equação (2):

**Tabela 2:** Matriz genérica dos dados

<b>Identificador químico</b>	<b>Atividade/ Propriedade</b>	<b>Descritor 1</b>	<b>Descritor 2</b>	<b>Descritor 3</b>	<b>...</b>	<b>Descritor n</b>
Molécula 1	Y1	X11	X12	X13	...	X1n
Molécula 2	Y2	X21	X22	X23	...	X2n
Molécula 3	Y3	X31	X32	X33	...	X3n
Molécula 4	...	...	...	...	...	...
Molécula n	Yn	Xn1	Xn2	Xn3	...	Xnn

$$\log \frac{1}{C} = aX_{Lipofílico} + bX_{eletrônico} + cX_{estérico} + dX_{polar} + e \quad (2)$$

Nessa equação, *C* é a concentração molar de cada composto capaz de produzir resposta biológica definida, tais como: IC<sub>50</sub>, concentração necessária para inibir 50% da atividade da enzima; LD<sub>100</sub>, concentração molar do fármaco capaz de matar 100%

dos indivíduos (como parasitas ou células tumorais) em que é administrado;  $ED_{50}$ , concentração molar do fármaco capaz de produzir 50% de seu efeito máximo; entre outras. Os símbolos  $X$ 's representam os descritores,  $a, b, c, d$  são coeficientes de ajuste e  $e$  representa um erro residual de previsão (CRONIN, 2010).

Segundo a OECD (*Organization for Economic Co-operation and Development*) um estudo QSAR/QSPR (QSRP do inglês *quantitative structure–property relationships*) deve seguir uma série de orientações durante o desenvolvimento e validação de modelos, como, por exemplo: definir e preparar o conjunto de dados; detecção e/ou remoção de *outliers* e pontos de alavancagem; seleção de variáveis (descritores); construção de modelos de calibração; validação dos modelos; e definição do domínio de aplicabilidade (TROPSHA, 2010; DEARDEN *et. al*, 2009; OECD, 2004).

### 1.9.1 Preparo do conjunto de dados

O entendimento do conjunto de dados disponível para o estudo QSAR é importante para a escolha do método mais adequado, pois, em alguns modelos, como por exemplo a regressão linear (simples RLS ou múltipla RLM), é essencial que os dados  $X$  tenham uma relação forte com os dados  $Y$  a serem modelados. No entanto, nem sempre isso é possível, devido às propriedades físico-químicas dos compostos (VARMUZA E FILZMOSE, 2009).

A covariância é uma medida que indica o grau de associação das variáveis  $X$  e  $Y$  Eq. (3). A covariância padronizada é chamada correlação Eq. (4) denotada por (JOHNSON E WICHERN, 1988). A mais utilizada é a correlação Pearson indicada para dados normalmente distribuídos; no entanto, é bastante sensível à presença de *outliers*.

$$Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] \quad (3)$$

$$\rho_{xy} = \left( \frac{Cov XY}{\sqrt{V(X)V(Y)}} \right) = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} \quad (4)$$

onde:  $\rho_{xy}$  é o coeficiente de correlação entre as variáveis  $X$  e  $Y$ ;  $\mathbb{E}$ , Esperança matemática;  $V$ , variância;  $\sigma$ , desvio padrão.

O intervalo da correlação, independentemente do método adotado, varia de  $-1 \leq \rho_{xy} \leq 1$ , um valor próximo de 1 indica perfeita relação linear, um valor próximo de -1 perfeita relação linear inversa (BASTOS, 2015). Valores menores que 0,3 indicam baixa relação linear.

Tal medida é importante no processo de análise dos dados. De acordo com a literatura, descritores com coeficiente de correlação entre si maiores que 0,95 devem ser retirados do modelo (VARMUZA E FILZMOSEER, 2009). Além disso, variáveis altamente correlacionadas (colineares) tornam a matriz de covariância singular e, conseqüentemente, o inverso não pode ser calculado (VARMUZA E FILZMOSEER, 2009). Nesse caso, bastante comum para dados químicos, a opção é trabalhar regressão das componentes principais (PCR), regressão dos mínimos quadrados parciais (PLS), entre outras (FERREIRA, 2015).

Uma matriz de covariância é singular quando ao menos uma variável é proporcional a outra, ao menos uma variável é combinação linear de outras variáveis, ao menos uma variável é constante, ou o número de variáveis é maior que o número de objetos (VARMUZA E FILZMOSEER, 2009). Ou seja, saber o grau de relacionamento entre as variáveis é importante tanto para saber se a variável deve ser retirada da análise quanto para avaliar que tipo de método (vide 1.9.4, Tabela 3) é mais propício para desenvolver o modelo.

Outra razão para se fazer um pré-tratamento nos dados é remover as fontes de variações da atividade biológica antes da análise. Nesses casos de dados de origem experimental recomenda-se a transformação logarítmica para linearizar os dados sem perda da interpretabilidade. No entanto, se os dados experimentais são séries temporais, dados digitalizados de um espectro contínuo, indica-se transformações de primeira ou segunda derivada (VARMUZA E FILZMOSEER, 2009).

A recomendação mais usual é aplicar a centralização e escala após a transformação dos dados (VARMUZA E FILZMOSEER, 2009). O objetivo é que as colunas tenham média zero (centralização) e a mesma variância (escala) Eq. (5).

$$X_{ij}(\text{média centralizada}) = X_{ij}(\text{original}) - \bar{X}_{ij} \quad (5)$$

Outra transformação bastante usual em quimiometria é o escalonamento automático Eq. (6).

$$X_{ij}(\text{auto escala}) = \frac{X_{ij}(\text{original}) - \bar{X}_{ij}}{s_j} \quad (6)$$

Deve se salientar que a necessidade de tratamento deve ser analisada e realizada para cada variável separadamente, pois nem todas as variáveis utilizaram o mesmo processo de construção (VARMUZA E FILZMOSEER, 2009).

### 1.9.2 Detecção de *outliers*

*Outlier* é um dado que pode ser considerado uma anomalia ou ruído. Esses valores se desviam do valor esperado com base em um modelo específico. Sua identificação e possível remoção é um importante problema para fins de mineração (AGGARWAL, 2017). O problema da detecção de *outlier* encontra aplicações em vários outros domínios e sua detecção torna-se um desafio quando existe um relacionamento significativo entre os diferentes pontos de dados, como no caso das séries temporais (AGGARWAL, 2017).

O *outlier* pode, por exemplo, arruinar a estimativa clássica dos componentes principais em uma análise PCA, pois o *outlier* se torna um ponto de alavancagem e influencia fortemente o resultado (VARMUZA E FILZMOSEER, 2009). Neste caso, ao identificar a presença de *outliers*, o analista deve optar por um método de análise robusto que tolera desvios da distribuição normal multivariada (VARMUZA E FILZMOSEER, 2009).

Existem diversos métodos para detecção de *outliers*, sendo grande parte oriundos de técnicas que usam o conceito da distância euclidiana ou distância de

Mahalanobis, bastante difundido na detecção de *outliers* multivariados (AGGARWAL, 2017; VARMUZA E FILZMOSER, 2009). Essas técnicas que utilizam o conceito da proximidade definem um ponto de dados como um *outlier* quando a sua localidade é pouco povoada (AGGARWAL, 2017).

Dentre as técnicas que se utilizam desse conceito estão análise de *cluster* e K-vizinhos. Entretanto, quando a análise for feita em um número pequeno de amostras e testes com uma única variável, a recomendação é uso de estratégias mais simples como: teste-t; soma dos quadrados dos desvios; *boxplot*; e técnicas baseadas em modelos de regressão linear, onde a principal suposição é a de que dados normalmente distribuídos e que não se enquadram naturalmente no modelo, são considerados *outliers* (AGGARWAL, 2017; VARMUZA E FILZMOSER, 2009).

### 1.9.3 Métodos de seleção de variáveis

O número de variáveis (descritores) para estudos QSAR deve ser devidamente avaliado, pois o uso excessivo de variáveis no modelo causa um ajuste forçado ou sobreajuste, do inglês *overfitting*, que consiste na obtenção de um valor elevado do coeficiente de correlação decorrente do número excessivo de variáveis incluídas no modelo e não de seu ajuste natural aos valores observados da atividade (SILVER, 2013).

Por esse motivo, apenas um subconjunto  $K$  deve ser utilizado na construção de cada modelo (FERREIRA *et al.*, 2002). Existe um limite para o valor de  $K$ ; do ponto de vista matemático, o valor máximo de  $K$  é igual a  $n - 1$ , sendo  $n$  o número total de observações. No entanto, alguns pesquisadores, como (KUBNYI, 1993 e UNGER, 1973), sugerem que para cada grupo de cinco ou seis compostos seja incluído apenas uma variável no modelo, para reduzir a possibilidade de *sobreajuste*. Isso se torna um problema nos estudos de QSAR, pois na maioria dos casos, o número de compostos disponíveis é menor que o número de descritores (VARMUZA E FILZMOSER, 2009). Nesses casos há necessidade de traçar alguma estratégia para encontrar combinações de  $K$  variáveis, dentre as  $m$  disponíveis, capazes de produzir modelos matemáticos que descrevam adequadamente os valores observados da atividade biológica (FERREIRA *et al.*, 2002).

Uma sugestão inicial é eliminar as variáveis que possuem uma porcentagem considerável de valores ausentes e/ou valores iguais. Por outro lado, as variáveis que devem ser preservadas são aquelas com alta variância e com alto coeficiente de correlação com a variável  $Y$ . No entanto, deve se salientar que algumas variáveis com correlação fraca são capazes de explicar a variabilidade de  $y$  (VARMUZA E FILZMOSER, 2009).

Após essa triagem inicial podem ser usadas estratégias do tipo *Stepwise*, regressão de melhores conjuntos, PCA, Algoritmo genético e análise de *cluster* (HOLFMANN *et al.*, 2007; FRALEY E RAFTERY, 1998; BROADHURST *et al.*, 1997). Independente da estratégia utilizada, as variáveis sugeridas precisam ser investigadas por critérios de ajuste como AIC (*Akaike's Information Criterion*), BIC (*Bayes Information Criterion*),  $C_p$  de Mallows e  $R^2$  (MASSART *et al.* 1997; FRANK AND TODESCHINI 1994; MALLOWS, 1973). Espera-se que esses critérios leve um bom modelo, contudo, vale a pena ressaltar um modelo final precisa passar por uma avaliação cuidadosa (VARMUZA E FILZMOSER, 2009).

#### 1.9.4 Métodos de calibração

As investigações QSAR requerem métodos de calibração multivariada, na qual, o objetivo é estimar os valores apropriados dos coeficientes de regressão e prever uma estimativa realista dos erros de previsão (FERREIRA *et al.*, 2002). Esses parâmetros são estimados a partir de um conjunto de calibração contendo valores das variáveis  $X$  e  $Y$  para  $n$  amostras. Dependendo de quão definida é a relação  $X$  e  $Y$ , diferentes níveis de criação e aplicação de modelos que prevejam  $Y$  a partir de  $X$  podem ser distinguidos (Tabela 3). O conhecimento do relacionamento entre as variáveis, pode determinar a escolha do método mais adequado.

**Tabela 3:** métodos apropriados para criação de modelos de acordo com o tipo de  $x$  e  $y$ . Fonte: adaptado de (VARMUZA E FILZMOSER, 2009).

n° de variáveis $y$	n° de variáveis $x$	Método
1	1	RLS
1	Várias	RLM, PLS, PCR, Ridge regression, Lasso
Várias	Várias	PLS2, CCA

Métodos de Regressão Linear Simples (RLS) (JOHNSON E WICHERN, 1988) relacionam uma propriedade  $Y$  a uma única variável  $X$ ; é um método sensível a *outliers* e raramente são utilizados em quimiometria. Por outro lado, a Regressão Linear Múltipla (RLM) que relaciona a propriedade  $Y$  com várias variáveis  $X$ , costuma ser mais usual; entretanto, possui algumas limitações: não é recomendada em casos de variáveis altamente colineares e/ou em casos que o número de objetos é menor que o número de variáveis (TROPISHA, 2010).

A regressão por componentes principais (PCR) (STONE E BROOKS, 1990) é indicada em casos de colinearidade entre as variáveis. O método dos mínimos quadrados parciais (PLS) (HOESKULDSSOM,1988) é o mais usado em quimiometria por não ser sensível a variáveis colineares e por aceitar uma grande quantidade de variáveis. Tanto o PCR quanto o PLS são modelos de regressão linear, embora existam versões não lineares (STONE E BROOKS, 1990). A regressão RIDGE (HOER e KENHARD,1970) e LASSO (TIBSHIRANI,1996) são alternativas para o PCR. A regressão RIDGE usa todas as variáveis originais  $X$  no modelo final, enquanto a regressão LASSO usa apenas um subconjunto de variáveis  $X$ . Em suma, os métodos de regressão linear são convenientes porque geralmente permitem uma interpretação e métodos como PLS e PCR evitam o problema de sobreajuste.

Apesar de se escolher um método mais adequado para tentar descrever um conjunto de observações experimentais, isso não garante que essas observações são bem descritas através do modelo. A descrição de um conjunto de observações por um modelo é feita por hipótese. Sendo assim, uma vez construído o modelo, é preciso submetê-lo a testes para verificar a veracidade da hipótese em que ele está fundamentado. A avaliação consiste em verificar se a especificação do modelo se adapta convenientemente aos dados observados (VARMUZA E FILZMOSER, 2009).

#### **1.9.4.1 Regressão linear**

A regressão linear múltipla (RLM) é uma combinação linear de variáveis independentes, também chamadas explicativas,  $X_1 + X_2 + \dots + X_k$  capaz de reproduzir da melhor forma possível os valores experimentais de um grupo de  $n$  observações do fenômeno  $Y$  (NETER *et al.*,1996).

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (7)$$

Na equação 7,  $Y_i$  é a variável aleatória a ser modelada,  $\beta_0$  é o termo da constante de ajuste,  $\beta_1 + \beta_2 + \dots + \beta_k$  são os coeficientes das variáveis independentes (parâmetros do modelo),  $X_1 + X_2 + \dots + X_k$  são valores conhecidos e  $\varepsilon$  é o erro associado devido à natureza estocástica do modelo (NETER *et al.*,1996).

A regressão linear simples (RLS) Eq. (8) é um caso especial de regressão linear múltipla pois relaciona variável dependente  $Y_i$  a uma única variável de resposta  $x$  (NETER *et al.*,1996).

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (8)$$

Os parâmetros  $\beta_1 + \beta_2 + \dots + \beta_k$  são estimados através do método dos mínimos quadrados parciais que consiste em encontrar o conjunto de valores capaz de minimizar os desvios (ao quadrado) entre cada um dos valores observados  $Y_i$  e os respectivos valores previstos,  $\hat{Y}_i$ , ou seja, o somatório  $\sum_{i=1}^n (Y_i - \hat{Y}_i)$  (GAUDIO E ZADONADE, 2001; DRAPER E SMITH,1981).

A determinação do erro  $\varepsilon_i$  é uma tarefa árdua porque os fatores que contribuem para o seu valor são irregulares, tais como possíveis erros aleatórios inerentes ao fenômeno observado, erros experimentais na medida de  $Y_i$  e a própria qualidade do ajuste do modelo. Isso implica que na prática o que se obtém é uma estimativa do modelo Eq. (9), na qual o erro  $\varepsilon_i$  não participa da equação (GAUDIO E ZADONADE, 2001), pois em QSAR as análises são realizadas um pequeno subconjunto de compostos, dentre os vários compostos, conhecidos e desconhecidos, que apresentam alguma atividade sobre o sistema em estudo. Dessa forma, estabelece-se em QSAR a designação de modelo ou estimativa de modelo para as equações de regressão representam alguma relação entre estrutura e atividade na qual as regras de proposição de modelos de Unger e Hansch tenham sido observadas.

$$\hat{Y} = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k \quad (9)$$



Uma vez construído a estimativa de modelo é preciso verificar o grau de ajuste, capacidade do modelo em reproduzir os valores observados dos objetos, através da análise dos resíduos  $ei = (y_i - \hat{y}_i)$  e da análise dos coeficientes significativos (GAUDIO E ZADONADE, 2001; BUSSAB, 1988).

A análise de distribuição dos resíduos da regressão verifica falta de ajuste do modelo, pois, nos casos em que o modelo é bem ajustado, o conjunto de resíduos  $ei$  contém apenas erros aleatórios, característicos da natureza dos dados. Quando o modelo apresenta falta de ajuste, além dos erros aleatórios, os resíduos contêm erros sistemáticos devidos a especificações incorretas do modelo. Espera-se que os resíduos de um bom modelo apresentem distribuição normal em torno de zero (GAUDIO E ZADONADE, 2001).

Existem diferentes métodos para se avaliar os coeficientes significativos; o mais usual é a análise de variância (ANOVA), que é um conjunto de modelos estatísticos fundamentados nos valores observados ( $y_i$ ), nos valores previsto ( $\hat{y}_i$ ) e na média global dos valores ( $\bar{y}_i$ ). Decompondo a variação total das observações do experimento ( $\sum(y_i - \hat{y}_i)^2$ ) em partes que podem ser atribuídas a causas controladas ( $\sum(\hat{y}_i - \bar{y}_i)^2$ ) e em partes atribuídas a causas não controladas ( $\sum(y_i - \hat{y}_i)^2$ ).

$$\sum(y_i - \hat{y}_i)^2 = \sum(\hat{y}_i - \bar{y}_i)^2 + \sum(y_i - \hat{y}_i)^2 \quad (10)$$

A demonstração da equação 10 pode ser encontrada em (DRAPER E SMITH,1981).

A análise da variância (ANOVA) ajuda a compreender o significado de alguns critérios de qualidade do ajuste do modelo, tais como:

Coeficiente de correlação ao quadrado ( $R^2$ ) Eq. (11): é a razão da variabilidade explicada pela variabilidade total da regressão. Quanto mais próximo de 1 for o seu valor, melhor a capacidade desse modelo em explicar a variabilidade dos valores observados da atividade biológica.

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y}_i)^2}{\sum(y_i - \hat{y}_i)^2} \quad (11)$$

Desvio padrão (s) Eq. (12): é dado pela razão da variabilidade que o modelo não consegue explicar pelo grau de liberdade do modelo. Um bom modelo deve apresentar o desvio padrão em torno de zero.

$$s = \frac{\sum(y_i - \hat{y}_i)^2}{(i - 1)} \quad (12)$$

Teste F: é um teste estatístico de hipótese, que avalia o grau de significância do  $R^2$  associada à regressão, em suma, verifica o quanto da variabilidade de  $Y$  pode ser explicada pelas variáveis  $X_1 + X_2 + \dots + X_k$  e o quanto pode ser atribuída ao efeito do erro aleatório  $e$ . Para validar  $R^2$  através do teste F, compara-se o valor de F, obtido no modelo, com o valor de referência de uma Tabela apropriada (JOHNSON E WICHERN, 1988).

## 2 Objetivo

Este estudo nasce da proposta de trabalhar com doenças negligenciadas sob uma ótica de desenvolvimento de novos fármacos. O objetivo deste projeto é a criação de um modelo de correlação entre os dados experimentais da concentração necessária para inibir 50% da atividade ( $IC_{50}$ ) (BARREIRO E FRAGA, 2015) da subunidade GyrB da DNA girase de micobactérias, com descritores moleculares dos compostos ativos, que possa ser usado para futuras previsões de atividade de novos compostos.

### 2.1 Objetivos específicos

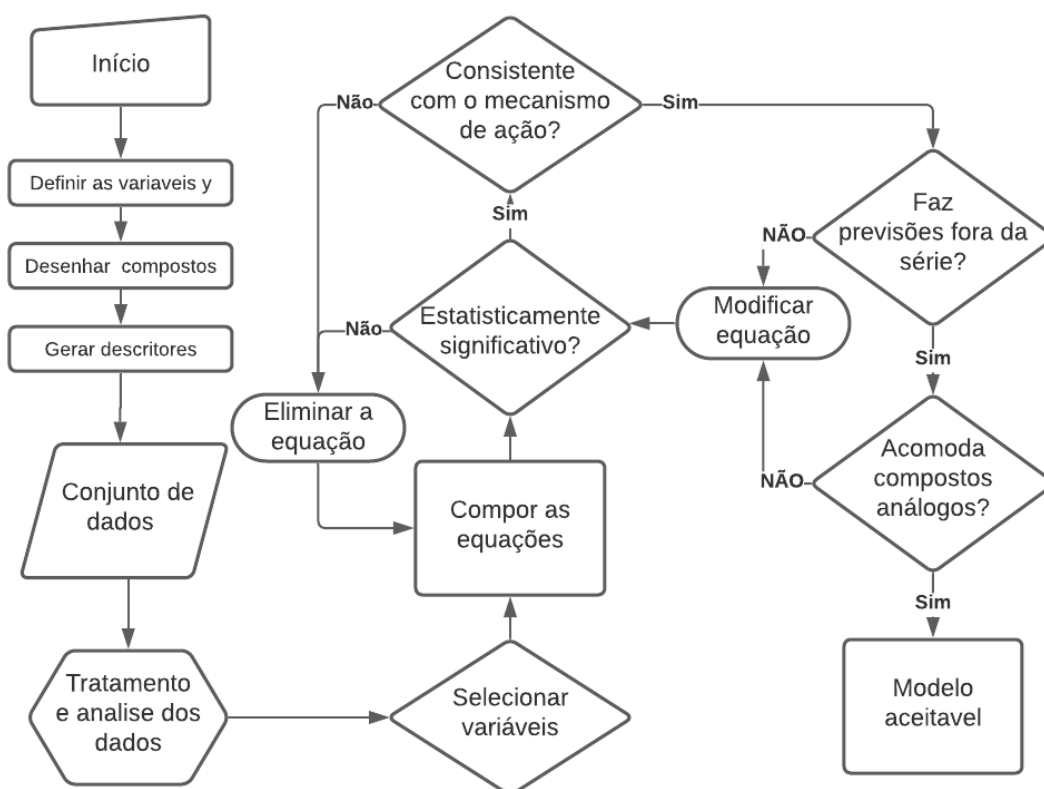
Os objetivos específicos desse trabalho foram:

- Calcular propriedades físico-químicas dos compostos obtidos na literatura usando cálculos quânticos semiempíricos, como a energia dos orbitais de fronteira;
- Utilizar métodos de docagem molecular para se investigar o modo de interação dos compostos com o sítio ativo da GyrB de *M. smegmatis*;

- Gerar descritores relacionados à formação dos complexos enzima-inibidor;
- Desenvolver e validar modelos de rQSAR empregando-se técnicas de regressão linear múltipla.

### 3 Metodologia

Neste trabalho pretende-se combinar métodos de modelagem molecular e de regressão linear múltipla para se construir modelos de predição de atividade inibitória sobre a subunidade B da DNA girase ( $pI_{C_{50}}$ ) para uma série de compostos obtidos na literatura (Vide 3.1) com ação inibitória sobre a subunidade B da DNA girase de *M. smegmatis*, espécie do mesmo gênero de *M. tuberculosis*. A identidade dos resíduos entre as enzimas das duas espécies é superior a 82%. O uso da enzima de *M. smegmatis* nos trabalhos experimentais se deve a razões práticas. Deve se salientar que uma equação de regressão linear difere de um modelo QSAR que depende de uma série de etapas (Figura 6) (GAUDIO E ZADONADE, 2001).



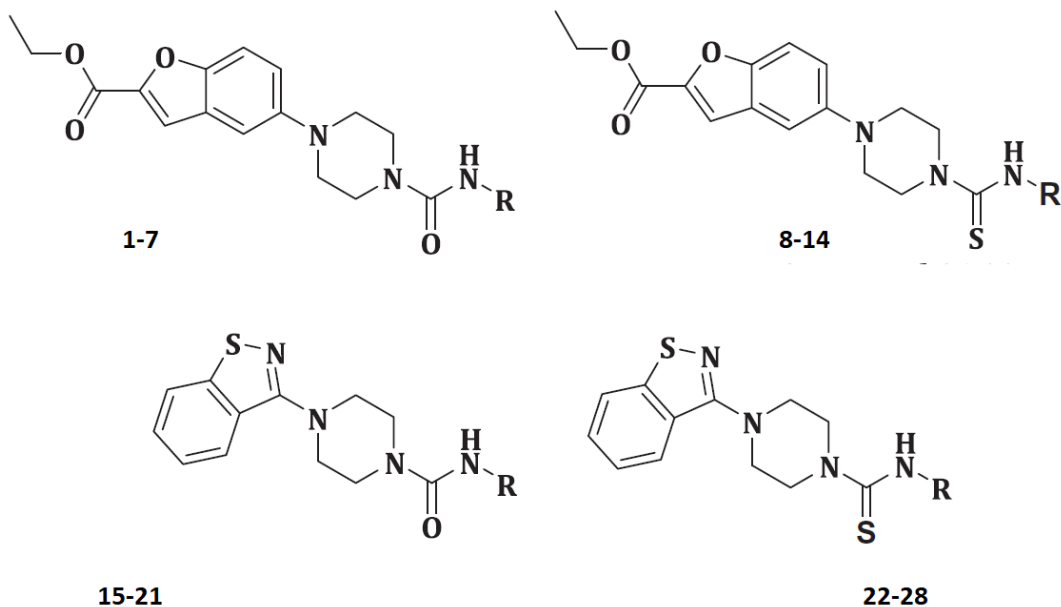
**Figura 6:** Esquema com as etapas pertinentes para a construção de um Modelo QSAR. **Fonte:** Adaptado de Gaudio e Zadonade, 2001.

Todas as análises estatísticas, gráficas e desenvolvimento das estimativas de modelo desse trabalho foram realizadas no programa R (R CORE TEAM, 2020). Disponível em edições comerciais e de código aberto e é executado na área de trabalho (Windows, Mac e Linux).

### 3.1 Seleção dos compostos

A fim de obter uma diversidade estrutural e uma variação dos valores da atividade biológica, variável y a ser modelada, a série de ligantes propostos neste trabalho advêm de três publicações:

- 1) Série de compostos **1 - 28** (Figura 7, Tabela 4) (REDDY, 2014).

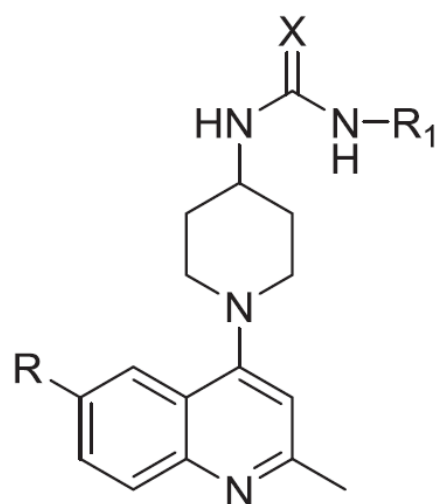


**Figura 7:** estrutura geral da série dos compostos do 1-28 usadas neste estudo (vide Tabela 4 para a relação dos substituintes R).

**Tabela 4:** Substituintes (vide Figura 7) para a série de compostos do 1-28 usados neste estudo.

Composto	R	Composto	R
1	Fenil	15	Fenil
2	4-Clorofenil	16	4-Clorofenil
3	4-Acetilfenil	17	4-Acetilfenil
4	4-Nitrofenil	18	4-Nitrofenil
5	4-Metoxifenil	19	4-Metoxifenil
6	Benzil	20	Benzil
7	4-Clorobenzil	21	4-Clorobenzil
8	Fenil	22	Fenil
9	4-Clorofenil	23	4-Clorofenil
10	4-Acetilfenil	24	4-Acetilfenil
11	4-Nitrofenil	25	4-Nitrofenil
12	4-Metoxifenil	26	4-Metoxifenil
13	Benzil	27	Benzil
14	4-Clorobenzil	28	4-Clorobenzil

2) Série de compostos **29 - 76** (Figura 8, Tabela 5) (MEDAPI, 2015).



**29-52 :X=O**

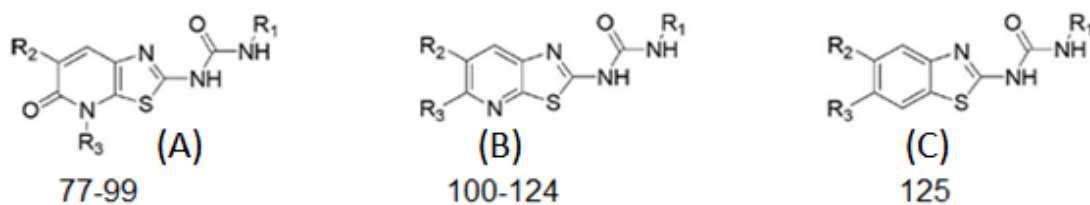
**53-76 :X=S**

**Figura 8:** estrutura geral da série de compostos do **29-76** usadas neste estudo (vide Tabela 5 para a relação dos substituintes X, R<sub>1</sub> e R<sub>2</sub>).

**Tabela 5 :** Substituintes (vide Figura 8) para a série de compostos do **29-76** usados neste estudo.

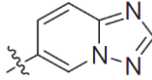
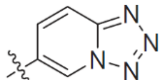
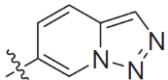
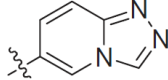
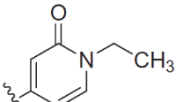
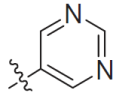
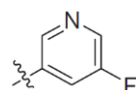
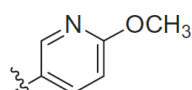
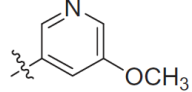
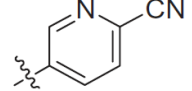
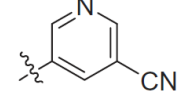
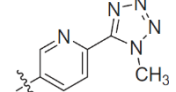
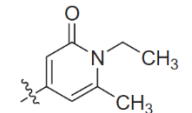
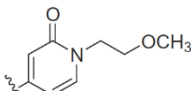
Composto	X	R <sub>1</sub>	R <sub>2</sub>	Composto	X	R <sub>1</sub>	R <sub>2</sub>
29	O	H	Fenil	53	S	H	Fenil
30	O	H	4-Fluorfenil	54	S	H	4-Fluorfenil
31	O	H	4-Clorofenil	55	S	H	4-Clorofenil
32	O	H	4-Nitrofenil	56	S	H	4-Nitrofenil
33	O	H	p-tolil	57	S	H	p-tolil
34	O	H	4-Metoxifenil	58	S	H	4-Metoxifenil
35	O	OCH <sub>3</sub>	Fenil	59	S	OCH <sub>3</sub>	Fenil
36	O	OCH <sub>3</sub>	4-Fluorfenil	60	S	OCH <sub>3</sub>	4-Fluorfenil
37	O	OCH <sub>3</sub>	4-Clorofenil	61	S	OCH <sub>3</sub>	4-Clorofenil
38	O	OCH <sub>3</sub>	4-Nitrofenil	62	S	OCH <sub>3</sub>	4-Nitrofenil
39	O	OCH <sub>3</sub>	p-tolil	63	S	OCH <sub>3</sub>	p-tolil
40	O	OCH <sub>3</sub>	4-Metoxifenil	64	S	OCH <sub>3</sub>	4-Metoxifenil
41	O	F	Fenil	65	S	F	Fenil
42	O	F	4-Fluorfenil	66	S	F	4-Fluorfenil
43	O	F	4-Clorofenil	67	S	F	4-Clorofenil
44	O	F	4-Nitrofenil	68	S	F	4-Nitrofenil
45	O	F	p-tolil	69	S	F	p-tolil
46	O	F	4-Metoxifenil	70	S	F	4-Metoxifenil
47	O	CF <sub>3</sub>	Fenil	71	S	CF <sub>3</sub>	Fenil
48	O	CF <sub>3</sub>	4-Fluorfenil	72	S	CF <sub>3</sub>	4-Fluorfenil
49	O	CF <sub>3</sub>	4-Clorofenil	73	S	CF <sub>3</sub>	4-Clorofenil
50	O	CF <sub>3</sub>	4-Nitrofenil	74	S	CF <sub>3</sub>	4-Nitrofenil
51	O	CF <sub>3</sub>	p-tolil	75	S	CF <sub>3</sub>	p-tolil
52	O	CF <sub>3</sub>	4-Metoxifenil	76	S	CF <sub>3</sub>	4-Metoxifenil

3) Série de compostos **77 - 125** (Figura 9, Tabela 6) (KAMSRI et. al., 2019).

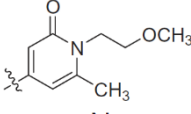
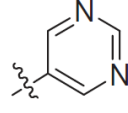
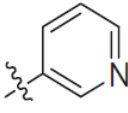
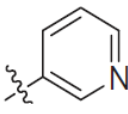
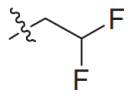
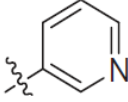
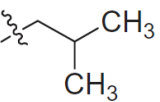
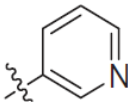
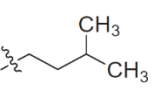
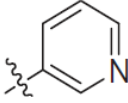
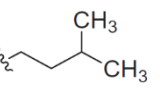
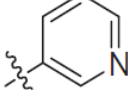
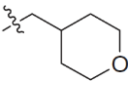
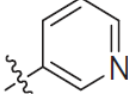
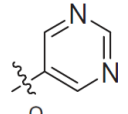
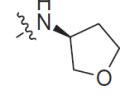
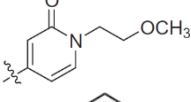
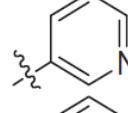
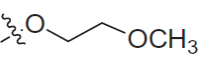
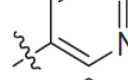
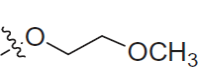
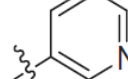
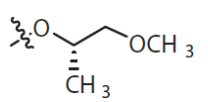
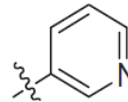


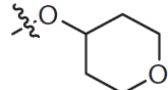
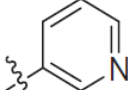
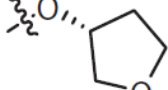
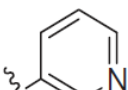
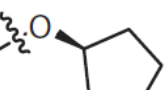
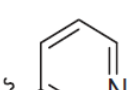
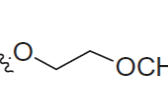
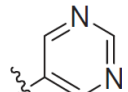
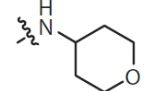
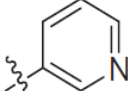
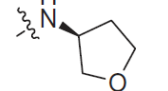
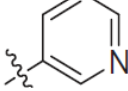
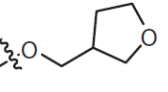
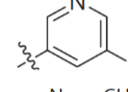
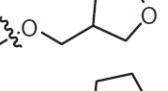
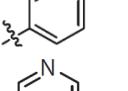
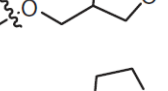
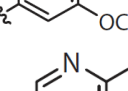
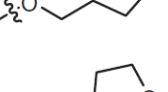
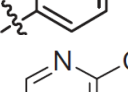
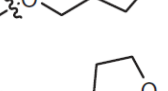
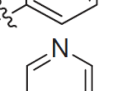
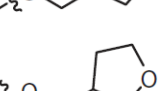
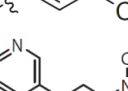
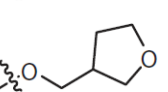
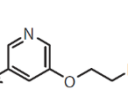
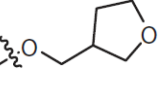
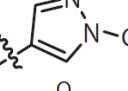
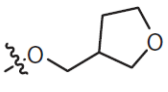
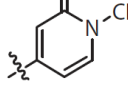


**Figura 9:** estrutura geral da série de compostos **77-125** usadas neste estudo (vide Tabela 6 para a relação dos substituintes R<sub>1</sub>, R<sub>2</sub> e R<sub>3</sub>).

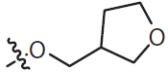
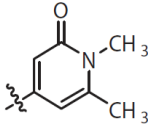
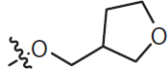
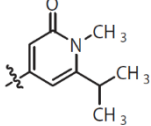
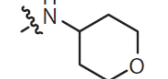
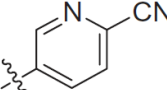
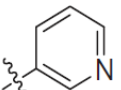
**Tabela 6:** Substituintes (vide Figura 9) para a série de compostos **77-125** usados neste estudo.

Composto	Estrutura Geral	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>
77	A	Etil	Isopropil	
78	A	Etil	Isopropil	
79	A	Etil	Isopropil	
80	A	Etil	Isopropil	
81	A	Etil	Isopropil	
82	A	Etil	Isopropil	
83	A	Etil	Isopropil	
84	A	Etil	Isopropil	
85	A	Etil	Isopropil	
86	A	Etil	Isopropil	
87	A	Etil	Isopropil	
88	A	Etil	Isopropil	
89	A	Etil	Isopropil	
90	A	Etil	Isopropil	



91	A	Etil	Isopropil	
92	A	Etil	H	
93	A	Etil	Etil	
94	A	Etil	Isopropil	
95	A	Etil		
96	A	Etil		
97	A	Alil		
98	A	Etil		
99	A	Etil		
100	B	Etil	Metoxil	
101	B	Etil		
102	B	Alil	H	
103	B	Alil		
104	B	Etil		
105	B	Etil		

106	B	Etil		
107	B	Etil		
108	B	Etil		
109	B	Etil		
110	B	Etil		
111	B	Etil		
112	B	Etil		
113	B	Etil		
114	B	Etil		
115	B	Etil		
116	B	Etil		
117	B	Etil		
118	B	Etil		
119	B	Etil		
120	B	Etil		
121	B	Etil		

122	B	Etil		
123	B	Etil		
124	B	Etil		
125	C	Alil		H

**Nota:** A relação dos compostos com os respectivos valores da atividade biológica pode ser encontrada na Tabela A1 sessão Anexos.

### 3.2 Construção dos compostos

As estruturas tridimensionais dos compostos foram construídas no programa Spartan'14 (WAVEFUNCTION, Inc.) disponibilizado no Laboratório de Modelagem Molecular da Universidade Federal Rural do Rio de Janeiro (IQ-UFRRJ). Este programa otimiza a geometria molecular da estrutura através da realização de cálculos quânticos e de mecânica molecular. As estruturas aqui estudadas tiveram suas energias minimizadas usando o campo de força MMFF (*Merck Molecular Force Field*). Os mínimos obtidos foram refinados através de nova minimização de energia com o método quântico semi-empírico PM6 (STEWART, 2007), sendo gravados em formato Sybyl (mol2).

No programa Spartan'14 é possível tanto construir as estruturas quanto obter informações e imagens 3D a respeito das propriedades estruturais e eletrônicas, (as últimas chamados descritores químico-quânticos), que podem ser exploradas na tentativa de obter modelos de correlação, no que se chama QSAR.

Nesse trabalho, investigou-se os descritores químico-quânticos da energia dos orbitais HOMO e LUMO. O orbital molecular ocupado de maior energia (HOMO-*Highest Occupied Molecular Orbital*) está associado ao caráter elétron-doador de um composto; assim, quanto maior a energia, maior a capacidade elétron-doadora (CARVALHO, 2003). O orbital molecular não ocupado de menor energia (LUMO-

*Lowest Unoccupied Molecular Orbital*) está associado ao caráter elétron-aceitador; quanto menor a energia, menor será a resistência para aceitar elétrons (CARVALHO, 2003). Tais energias são usadas como índices de reatividade química e são comumente correlacionadas com outros índices, tais como afinidade eletrônica e potencial de ionização (CARVALHO, 2003).

Outro descritor obtido pelo programa foi a energia de solvatação ( $E_{\text{solv}}$ ), calculada pela diferença das energias da molécula no vácuo e em água. No programa Spartan'14 há 2 modelos para o cálculo da energia de solvatação, sendo o modelo SM5.4 aplicado para cálculos no nível semiempírico (GIESEN *et al.*, 1997). A eletrostática e a polarização soluto-solvente são incluídas no modelo de forma autoconsistente pela equação de Born generalizada com cargas de classe IV, e os efeitos de primeira camada de solvatação são modelados em termos de áreas de superfície acessíveis a solvente que dependem de geometrias de soluto e quatro descritores de solvente. A inclusão de propriedades de solvente no termo de primeira camada de solvatação fornece um modelo que prevê energias livres de solvatação para qualquer solvente para o qual essas propriedades sejam conhecidas (GIESEN *et al.*, 1997).

### 3.3 Docagem molecular

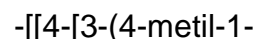
Uma das variáveis independentes  $X$  (descritores moleculares) incluídas nesse estudo foi a pontuação de ajuste obtida nos estudos de docagem realizados no programa GOLD 5.7.3 (CCDC).

Inicialmente, foi realizada uma busca na página do *Protein Data Bank* (PDB), um banco de dados de acesso público que contém estruturas 3D experimentalmente determinadas de proteínas e ácidos nucleicos (BERMAN *et al.*, 2000), para se selecionar a estrutura cristalográfica com o código 4B6C do domínio GyrB da DNA girase de *M. smegmatis*, adequada ao estudo (CHAUDHARI *et al.*, 2016).

O programa GOLD confere flexibilidade total aos ligantes e flexibilidade parcial às proteínas durante o processo de docagem. Para isso, o programa estabelece um ponto central e um raio de busca para a determinação da região onde os ligantes devem interagir com a proteína, ou seja, o sítio de interação. O ponto central pode ser

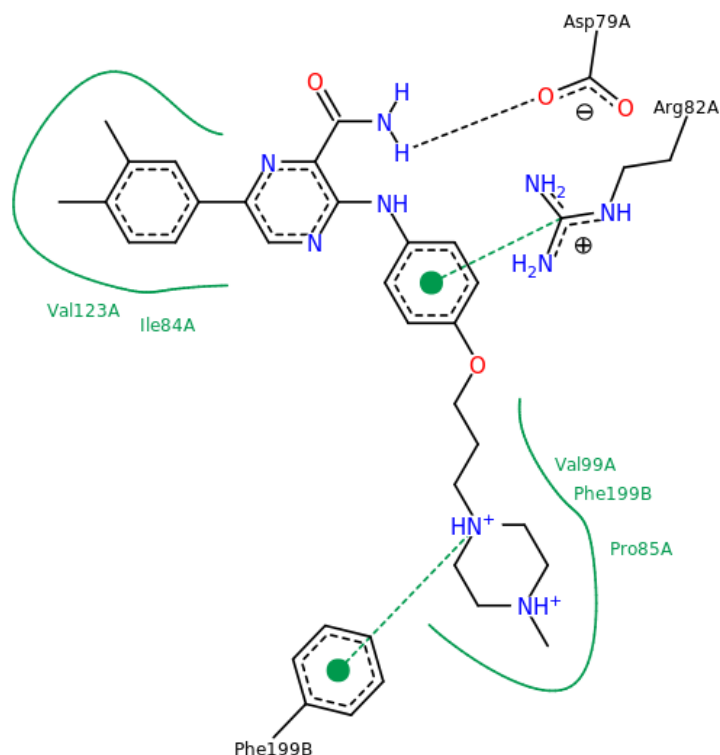
um átomo de um aminoácido da enzima, do próprio ligante co-cristalizado ou simplesmente as coordenadas de um ponto específico. O raio de busca determina o tamanho total do sítio de interação que será investigado e ele deve ser grande o suficiente para acomodar todos os ligantes de interesse (GOLD User Guide, 2019). Nesse trabalho o ponto central foi definido como o carbono gama do resíduo Asp79, e o raio de busca de 15 Å.

Para validar o uso do programa para a docagem nesta estrutura, foi feito um processo de redocagem, no qual se avaliou, através dos valores de RMSD, qual das funções de pontuação disponíveis no programa GOLD (ASP, ChemPLP, Chemscore e Goldscore) seria a mais apropriada para prever os modos de interação do ligante co-cristalizado



piperazinil)propoxi]fenil]amino]pirazina-2-carboxamida (1) (Figura 10) na enzima.

Todas as análises foram feitas em triplicata.



**Figura 10:** imagem do ligante (1) no interior da estrutura 4B6C, gerada com o programa Poseview, em que são mostradas suas principais interações com resíduos de aminoácidos do sítio.  
**Fonte:** <https://www.rcsb.org/structure/4B6C>

### 3.4 Cálculos das entalpias de interação

Neste estudo também foram utilizados os valores de entalpia de interação ( $\Delta H_{int}$ ) como variáveis independentes  $X$ . Relacionar esses parâmetros com dados estruturais possibilita entendimento do comportamento de biomoléculas (JELESAROV E BOSSHARD, 1999). Uma maneira de obter valores de  $\Delta H_{int}$  do complexo enzima-ligante é realizando cálculos quânticos.

Para realizar esses cálculos é necessário preparar as estruturas, em três etapas I-III:

I- Efetuar o recorte do complexo: os arquivos gravados na etapa de docagem, contendo os complexos de maiores pontuações, foram abertos no programa PyMOL (DELANO, 2002). O objetivo dessa etapa diminuir o tamanho da estrutura com a qual seriam efetuados os cálculos quânticos. Para isso, foi feito um recorte de raio de 5 Å ao redor do ligante. Feito isso, os recortes foram salvos em formato PDB, extensão que descreve a estrutura molecular por uma matriz de coordenadas cartesianas; em seguida; foi feita a adição de átomos de hidrogênio nas ligações truncadas para fechar a valência, com o programa GOLD;

II- Definir a carga total dos complexos, calculada contando-se o número total de aminoácidos carregados que existem na estrutura que são: Lys (carga positiva), Arg (carga positiva), Asp (carga negativa) e Glu (carga negativa);

III- Editar e incluir no arquivo PDB palavras chaves necessárias para os cálculos (<http://openmopac.net/manual/index.html>), tais como: MOZYME, que utiliza o método de Orbitais Moleculares Localizados para acelerar os cálculos executados; MMOK, que inclui a correção da mecânica molecular para ligações peptídicas; EPS, que define a constante dielétrica a ser usada no cálculo, dentro da aproximação que representa o meio externo ao sistema como um contínuo; NOOPT e OPT-H, que permite a

otimização somente da geometria dos hidrogênios; CHARGE, que define a carga do sistema; PDBOUT, que gera um arquivo “.pdb” ao final do procedimento.

Os cálculos quânticos para minimização da energia foram realizados com o programa MOPAC (*Stewart Computational Chemistry*) (MOPAC, 2016), através do método semi-empírico PM7 (STEWART, 2013). A representação do meio externo ao sistema como um contínuo é uma aproximação e, com objetivo de analisar o comportamento das estruturas sob diferentes condições, foi utilizado primeiro a constante dielétrica para o interior de proteínas (EPS=4.0) de modo a compensar parcialmente o efeito os aminoácidos mais distantes que foram excluídos no momento do recorte. Em um segundo momento, os cálculos foram realizados em meio aquoso (EPS=78.4), considerando possibilidade de existência de água próxima ao sítio.

Os valores de entalpia de interação  $\Delta H_{int}$ , ou seja, os valores das entalpias envolvidas no equilíbrio da formação do complexo enzima-inibidor, foram calculadas com base na equação 13:

$$\Delta H_{int} = \Delta H_{f_{complexo}} - (\Delta H_{f_{proteína}} + \Delta H_{f_{ligante}}) \quad (13)$$

Onde  $\Delta H_{int}$  é a entalpia de interação;  $\Delta H_{f_{complexo}}$  é a entalpia de formação do complexo;  $\Delta H_{f_{proteína}}$  é a entalpia de formação do sítio vazio e  $\Delta H_{f_{ligante}}$  corresponde à entalpia de formação do ligante.

### 3.5 Determinação da matriz de dados

A determinação da matriz de dados (Tabela 2, vide sessão 1.9) é um dos fatores mais importantes nos estudos QSAR. Os valores da variável dependente  $Y$ , 125 observações ao todo, foram extraídos da literatura (KAMSRI *et. al.*, 2019; MEDAPI, 2015; REDDY, 2014). As atividades biológicas são mensuradas pelo parâmetro  $IC_{50}$  sobre a subunidade GyrB da DNA girase de *M. smegmatis*. Por se tratar de um dado experimental, foi aplicada a conversão logaritmo Eq. (14) aos

valores do  $IC_{50}$  para linearizar os dados (FERREIRA E KIRAL, 2011; POPE E LIU, 2020).

$$pIC50 = -\log IC50 \quad (14)$$

As variáveis independentes  $X$  (descritores moleculares), são os dados referentes aos descritores químico-quânticos, obtido através do programa Spartan'14; valores gerados nos estudos de docagem molecular; e cálculos de entalpia de interação  $\Delta H_{int}$  em kJ/mol. Ao todo foram 33 descritores, que podem ser visualizados nas Tabelas presentes nos Anexos, sessão 7.

### 3.6 Pré-tratamento e análise dos dados

O pré-tratamento e análise dos dados é uma etapa importante em qualquer trabalho dessa natureza, principalmente quando existe uma quantidade considerável de observações e variáveis. O bom entendimento dos dados possibilita a escolha do método mais apropriado para aquele conjunto de dados. Algumas características importantes devem ser avaliadas, como por exemplo, se o número de variáveis é maior que o número de objetos, ou se as variáveis são altamente correlacionadas. Ambos problemas são comuns em quimiometria, e o uso de métodos como PCR e PLS são os indicados nesses casos.

O primeiro tratamento foi remover descritores com valores constantes ou nulos. Depois foi feita a análise da matriz de correlação entre as variáveis e o  $pIC_{50}$  para verificar o nível de correlação entre as variáveis e a existência de variáveis autocorrelacionadas.

### 3.7 Seleção das variáveis

Dentre os 33 descritores deve se definir quais devem ser incorporados ao modelo, de modo a conseguir parâmetros de ajuste adequado sem que haja o sobreajuste do modelo. A abordagem utilizada para definir essas variáveis foi o método de regressão para os melhores Subconjuntos (*Best subsets*) (HOLFMANN *et*



*al.*; 2007). Esse método realiza inúmeras combinações entre os possíveis subconjuntos das variáveis, cada possibilidade foi avaliada pelos parâmetros de ajuste  $R^2$  e  $S^2$  (HOLFMANN *et al.*; 2007).

### **3.8 Estimativa dos modelos**

Os modelos foram gerados no programa R utilizando equações de regressão linear múltipla. Na primeira etapa houve a tentativa de se construir modelos com todas as observações. Em seguida, com a perspectiva de se obter melhores resultados, desmembrou-se as observações de acordo com a série dos compostos.

Técnicas de regressão linear são bastantes sensíveis aos *outliers* e pontos de alavancagem. Por isso, após o desenvolvimento de alguns modelos iniciais, foi feita a interpretação gráfica para identificar possíveis *outlier* e pontos de alavancagem sobre o modelo.

### **3.9 Análise dos resíduos**

Todos os modelos para os inibidores do domínio GyrB de *M. smegmatis* passaram por testes para analisar se os resíduos assumem os pressupostos: seguem uma distribuição normal; são independentes; e têm variância  $\sigma^2$  constante (homocedasticidade).

A normalidade dos resíduos foi verificada pelo teste Shapiro-Wilk, cujo P-valor deve ser superior à 0,05 para indicar que possuem distribuição normal (STEINSKOG *et al.*, 2007). Diante do pressuposto que os resíduos seguem uma distribuição normal. Foi realizado o teste de homoscedasticidade Brush-Pagan, na qual, o P-valor também deve ser superior à 0,05 para indicar a homoscedasticidade (BAUM *et al.*, 1999). O teste da independência, ou seja, os resíduos não são autocorrelacionados foi realizado através do teste de Ddurbin-Watson, os valores desse teste devem estar entre 1,5 e 2,5 (TILLMAN, 1975).

### **3.10 Validação dos modelos**

A validação interna de todos os modelos de QSAR foi realizada utilizando os compostos do conjunto treinamento, através do método de validação cruzada LOO (do inglês *Leave One Out*) (WONG, 2015). Esse método treina o modelo preditivo com  $(n - 1)$  observações de uma amostra de tamanho  $n$ . Esse procedimento é repetido várias vezes até que todos os compostos do conjunto tenham sido retirados e preditos ao menos uma vez (WONG, 2015). Os resultados da validação cruzada foram expressos pelo coeficiente de correlação da validação cruzada ( $Q^2$ ) Eq. (15) e pelo desvio-padrão da validação cruzada ( $S_{PRESS}$ ) Eq. (16)

$$Q^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y}_i)^2} \quad (15)$$

$$S_{PRESS} = \frac{\sum(y_i - \hat{y}_i)^2}{n - k - 1} \quad (16)$$

Onde,  $(y_i)$  são os valores observados,  $(\hat{y}_i)$  valores previstos e  $(\bar{y}_i)$  a média global dos valores.  $(n)$  representa o número de amostras e  $(k)$  se refere ao número de descritores do modelo (GAUDIO E ZADONADE, 2001).

A validação externa, dos modelos foi desenvolvida utilizando o conjunto teste contendo 30% dos compostos que não foram utilizados na geração dos modelos. As estatísticas de previsão do modelo são expressas pelo coeficiente de determinação múltipla  $R^2_{EXT}$  Eq. (17), também conhecido como coeficiente de correlação múltipla ou variância explicada (KIRALJ E FERREIRA, 2009), e a raiz quadrada média do erro de previsão (RMSEP) Eq. (18) (KIRALJ E FERREIRA, 2009).

$$R^2_{EXT} = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - (\hat{y}_i))^2} \quad (17)$$

$$RMSEP = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{n}} \quad (18)$$

A avaliação estatística do modelo final deve considerar a relação em que o coeficiente de correlação  $R^2$  Eq. (11), calculado para o conjunto de treinamento, seja maior que a correlação da validação cruzada  $Q^2$ . Nos modelos de QSAR o valor mínimo aceito para  $Q^2$  é 0,5, enquanto  $R^2 > 0,6$ . Caso a diferença entre  $R^2$  e  $Q^2$  exceda 0,2-0,3 isso é um forte indicio de que o modelo teve um sobreajuste (KIRALJ E FERREIRA, 2009).

## 4 Discussão de resultados

### 4.1 Docagem molecular

Os resultados do processo de redocagem foram avaliados em função do valor de RMSD (Tabela 7). A função de pontuação que gerou o melhor RMSD para a enzima (vide item 3.3, Figura 10) no sítio ativo da enzima (PDB: 4BC6), foi GoldScore. No entanto, as outras funções apresentaram valores dentro do limite aceitável,  $RMSD < 2$ . Sendo assim, foram realizados quatro estudos de docagem molecular, um para cada função disponível.

**Tabela 7:** Resultados do processo de redocagem para as quatro funções disponíveis no programa GOLD 5.7.3. Estão apresentados os melhores valores de RMSD obtidos para cada função.

Função	Melhor RMSD (Å)
ASP	1,299
ChemPLP	1,096
ChemScore	1,270
GoldScore	0,528

Ao finalizar a docagem molecular, o programa GOLD 5.7.3 disponibiliza um arquivo externo chamado *bestranking.lst*, uma lista com a melhor pontuação de score de cada estrutura analisada. Nessa lista, além da pontuação há também outras propriedades listadas de acordo com a função escolhida para fazer o processo de docagem. Como exemplo, no *bestranking.lst* da função ChemScore existem dados referentes ao termo da ligações de hidrogênio entre a proteína e o ligante definido pela sigla *S(h bond)*. Nesse trabalho, os dados do arquivo *bestranking.lst* são parte

dos descritores moleculares utilizados como  $m$  variáveis descritivas  $X$  do modelo. Os cálculos foram feitos em triplicata; dessa forma, foi feita a média aritmética dos três arquivos de saídas *bestranking.lst* antes de começar as análises.

No programa GOLD 5.7.3 é possível aplicar a técnica de *reescore* (repontuação), no processo de docagem. Essa técnica recalcula a pontuação de uma pose gerada por uma função com uma outra função. Diferentes funções de pontuação podem ter um desempenho melhor em casos específicos. Por exemplo, uma função pode superar a outra na classificação de compostos com um tipo de atividade, enquanto o inverso pode ocorrer com outras classes (GOLD USER GUIDE, 2019). Nesse trabalho utilizou-se essa estratégia na tentativa de se obter pontuações de saídas, que nesse estudo são parte das variáveis independentes, que apresentassem uma melhor correlação com a variável dependente  $pIC_{50}$ . A docagem foi realizada com a função GoldScore e o *reescore* com a função ChemPLP.

As pontuações resultantes do processo de docagem molecular (sessão Anexos) não foram muito adequadas para a previsão do perfil de atividade. Isso já era esperado, pois as funções de pontuação não levam em conta alguns termos importantes para o processo de interação, como entropia e participação da água (KITCHEN *et al.*, 2004). Por isso, outros métodos devem ser utilizados na busca da correlação, como cálculos semiempíricos dos complexos e dados obtidos a partir de informações eletrônicas calculadas dos compostos, tais como valores da Energia do HOMO, Energia do LUMO e  $E_{solv}$ .

## 4.2 Análise da matriz de covariância e Seleção de variáveis

Antes de aplicar alguma estratégia para selecionar quais variáveis devem ser incorporadas ao modelo, foi necessário determinar a matriz de covariância para avaliar a correlação entre as variáveis independentes e a variável dependente. Observou-se que as variáveis independentes *Fitness* e *Svdw\_ext*, que são pontuações de saída da função GoldScore são autocorrelacionadas. Isso implica que tais variáveis não devem estar presentes em uma mesma equação de regressão linear múltipla. As variáveis *Score* e *DG*, do arquivo de saída da função ChemPLP e as

variáveis *Smap* e *De.clash* do arquivo de saída da função ASP também são autocorrelacionadas.

Outra observação a ser feita é a correlação entre as variáveis independentes e a variável dependente. Em geral, os valores de correlação foram inferiores a 0,32, índices de 0,3 a 0,5 indicam uma correlação fraca (MUKAKA, 2012). Mesmo as 33 variáveis apresentando uma correlação fraca com as 125 observações em questão, foi verificada a possibilidade de se obter um modelo de regressão linear múltipla com um  $R^2$  moderado, índices de 0,5 a 0,7 (MUKAKA, 2012). Para isso, foi utilizado o método de seleção dos melhores subconjuntos (*Best subsets*) (HOLFMANN *et al.*; 2007) para avaliar quais das 33 variáveis seriam mais adequadas. Esse método investiga os melhores modelos, segundo o critério de menor soma de quadrados residual, para subgrupos de variáveis preditoras de todos os tamanhos. Além disso, mostra uma estimativa das melhores regressões e a função também retorna os valores de  $R^2$  e  $S^2$  de cada regressão.

Os modelos sugeridos pelo *Best subsets* apresentaram coeficientes de  $R^2$  inferiores à 0,2. Diante disso, investigou-se a possibilidade de haver uma melhora na correlação e, conseqüentemente, modelos com coeficientes de correlação mais adequados, caso os compostos fossem divididos de acordo com a série. Por dois motivos: primeiro porque, os 125 compostos selecionados foram obtidos a partir de três referências diferentes, com diferenças experimentais no método de obtenção dos valores de  $IC_{50}$ ; e segundo porque, o método QSAR é mais adequado para estabelecer uma correlação entre as atividades biológicas e as estruturas químicas de uma série de compostos semelhantes, especialmente compostos que se diferenciam somente pela presença de um ou mais grupos substituintes diferentes em posições estabelecidas na estrutura química comum à série (FERREIRA *et al.*, 2002), o que não é o caso para as séries estudadas. Como já foi dito (vide item 3.1), existem as séries **1-14**; **15-28**; **29-76**; **77-99**; **100-124**; e o composto **125** e estas não são estruturalmente relacionadas. Feita a separação dos dados de acordo com essas séries, foi feita a matriz de covariância.

A estratégia de separar os dados de acordo com a série dos compostos rendeu alguns resultados satisfatórios. Na série **1-14**, por exemplo, a variável *S.vdw\_ext* tem um coeficiente de correlação de -0,67 com o  $pIC_{50}$ . Nas séries **15-28** e **29-76** as

correlações apresentaram índices inferiores a 0,2, isso indica que a correlação é depressível (MUKAKA, 2012). Sendo assim, os dados referentes à essas duas séries, **15-28** e **29-76**, foram descartados para a construção de um possível modelo neste estudo. Já com os dados das séries **77-99** e **100-124** foi possível obter coeficientes de correlação fraco ou moderado, em determinadas variáveis. Já o composto **125**, não foi possível associa-lo a nenhuma série, sendo assim descartado desse projeto.

Cada conjunto de dados relativo à série mencionada anteriormente, passou pelo método de seleção de variáveis *Best subsets*. Esse método faz combinações entre os possíveis subconjuntos das variáveis. Entretanto, é preciso avaliar se as variáveis sugeridas são estatisticamente significativas, ou seja, se os coeficientes são diferentes de zero para aquela equação. Cada possibilidade de modelo sugerida pelo método foi avaliada, mas só serão discutidas no item 4.3 as estimativas de modelos que apresentaram variáveis significativas e coeficientes de ajuste razoáveis.

### **4.3 Elaboração dos modelos de inibição da subunidade GyrB da DNA girase de *M. smegmatis***

#### **4.3.1 Estimativa de modelo com a série de compostos 1-14**

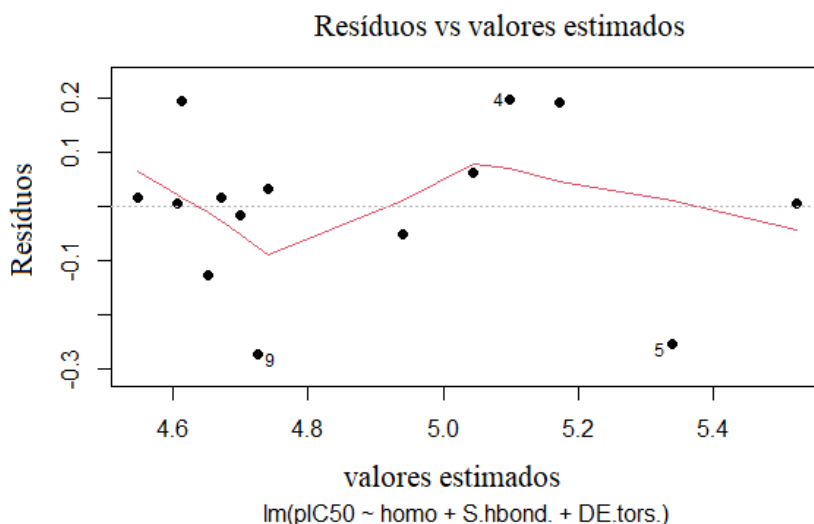
O conjunto de dados, com as 14 observações da série 1-14, resultou no modelo 1 Eq. (19).

$$pIC50 = 35,808 + 3,185HOMO - 2,446S.hbond - 1,059DE.tors \quad (19)$$

Onde *HOMO* é o descritor da energia do HOMO obtido com o método PM6 do programa Spartan'14; *S.hbond* é o descritor que representa o termo da ligação de hidrogênio entre a proteína e o ligante obtido nos cálculos de docagem com a função ChemScore, e *DE.tors* é o descritor obtido nos cálculos de docagem com a função ChemPLP, representa o peso do potencial de torção do ligante.

Antes de efetuar as validações interna e externa no modelo 1, foi realizada a análise dos resíduos. O teste de Shapiro-Wilk teve um valor  $P=0,14$ , logo os resíduos seguem uma distribuição normal. No entanto, esse modelo 1 não é válido pois falha no teste de Ddurbin-Watson, cujo valor foi 2,99 que é superior ao limite estabelecido.

Outra forma de verificar a independência e homocedasticidade dos resíduos é através do gráfico dos resíduos em função dos valores estimados, (Figura 11). Os dados deveriam estar distribuídos de maneira razoavelmente aleatória e como amplitude semelhante em torno de zero.



**Figura 11:** gráfico resíduos vs valores estimados, que avalia um pressuposto de linearidade, para isso, a linha vermelha deve estar aproximadamente na horizontal.

#### 4.3.2 Estimativa de modelo com a série de compostos 77-99

As 23 observações da série de compostos **77-99** resultaram no modelo 2 Eq. (20), com um  $R^2 = 0,74$ . A estatística de avaliação interna desse modelo foi verificada pelo coeficiente de correlação da validação cruzada,  $Q^2 = 0,56$  que apesar de não ser tão próximo do número 1, está dentro do limite sugerido por (KIRALJ E FERREIRA, 2009). O desvio-padrão da validação cruzada modelo 1,  $S_{PRESS} = 0,082$ , um valor satisfatório, pois o desejado é um valor próximo de zero.

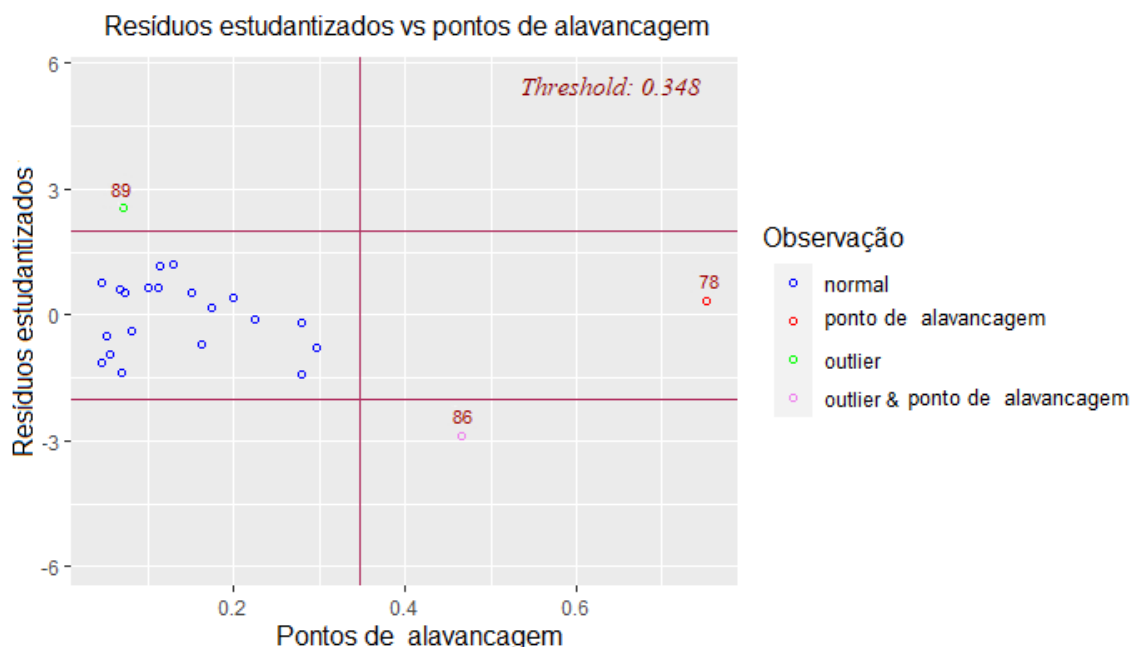
$$pIC50 = 3,681 - 0,146S.int + 0,453S.hbond + 0,00191\Delta Hint \quad (20)$$

Onde  $S.int$ , obtido nos cálculos de docagem com a função de pontuação GoldScore, representa a soma dos termos de torção interna e de van der Waals

interno (ou seja, do ligante);  $S.hbond$  representa o termo da ligação de hidrogênio entre a proteína e o ligante obtido nos cálculos de docagem com a função de pontuação ChemScore e  $\Delta H_{int}$  é a entalpia de interação em meio aquoso, em kJ/mol, obtido através dos cálculos semiempíricos com o método PM7 do programa Mopac2016 com a constante dielétrica d' água,  $EPS= 78.4$

Os pressupostos da análise dos resíduos foram atendidos. Os resíduos seguem uma distribuição normal, o teste de Shapiro-Wilk teve valor de 0,90. São independentes, o teste de Ddurbin-Watson teve valor de 2,15. São homocedásticos, com o teste de Brush-Pagan com valor de 0,36.

A estatística da avaliação externa do modelo 2 foi verificada pelo coeficiente de correlação para os dados de teste,  $R^2_{Ext} = 0,58$ , apresentou um resultado inferior ao  $R^2 = 0,74$  do conjunto de treinamento. Essa divergência de valores entre esses coeficientes são um indicativo de que o modelo 2, apesar de apresentar coeficientes de ajustes adequados para o conjunto de treinamento, não foi capaz de obter um bom resultado quando exposto a dados do conjunto de teste. Como a técnica de regressão linear é bastante sensível a pontos de alavancagem e *outliers*, foi realizada a análise gráfica (Figura 12) do modelo 2, para identificar as observações que estariam influenciando o modelo.



**Figura 12:** gráfico para detectar observações influentes do modelo 2. Resíduos estudantizados serão mais eficazes para detectar observações Y periféricas do que resíduos padronizados.



As observações indicadas no gráfico foram retiradas uma de cada vez, e em seguida elaborou-se os possíveis modelos. Após diversas investigações, retirou-se da série **77-99** os compostos **86, 89 e 96**. Com esse novo conjunto de dados contendo 20 observações, chegou-se ao modelo 3 Eq. (21).

$$pIC50 = 3,333 - 0,138S.int + 0,649S.hbond + 0,0020\Delta Hint \quad (21)$$

Onde os descritores são os mesmos descritos no modelo 2.

O modelo 3 difere do modelo 2 no tamanho ( $n$ ) da amostra, pois foram retirados os pontos influentes do conjunto de dados. Na Tabela 8, é possível comparar que a retirada de pontos influentes de um conjunto de dados é capaz de melhorar significativamente os coeficientes de ajustes, o que torna o modelo 3 mais adequado.

**Tabela 8:** Comparação dos parâmetros estatísticos dos modelos 2 e 3

Modelo	$R^2$	$R^2_{Ext}$	$Q^2$	$S_{PRESS}$	RMSEP	$k$	$n$
2	0,74	0,58	0,56	0,082	0,064	3	23
3	0,88	0,84	0,84	0,056	0,042	3	20

Os pressupostos da análise dos resíduos foram atendidos no modelo 3. Os erros têm distribuição normal, o teste de Shapiro-Wilk teve valor de 0,91. Os erros são independentes, o valor do teste de Ddurbin-Watson foi igual a 2,31. Os erros são homocedásticos, o teste Brush-Pagan obteve um P-valor de 0,86.

#### 4.3.3 Estimativa de modelo com a série de compostos 100-124

O método de seleção de variáveis *best subset* indicou que o conjunto de dados com as 25 observações da série de compostos **100-124** resultariam em modelos com  $R^2$  inferior a 0,5. No entanto, com a remoção das observações **106, 120 e 121** foi possível construir o modelo 4 Eq. (22).

$$pIC50 = -1,058 - 1,551LUMO + 0,070S.vdw_{ext} - 0,0911S.hb_{ext}.1 \quad (22)$$

Onde *LUMO* é o descritor da energia do LUMO, obtido com o método PM6 do programa Spartan'14; *S.vdw<sub>ext</sub>*, descritor obtido nos cálculos de docagem com a função de pontuação GoldScore, representa energia de *van der Waals* do complexo entre o ligante e a proteína e *S.hb<sub>ext</sub>.1*, descritor obtido nos cálculos de docagem utilizando a técnica de reescore, onde foi combinado a função GoldScore e ChemPLP, representa a energia das ligações de hidrogênio do complexo entre o ligante e a proteína.

O modelo 4 obteve  $R^2 = 0,82$  para o conjunto de treinamento. O valor de  $Q^2$  foi 0,78 e o de  $S_{PRESS}$  foi 0,063. Na validação externa, obteve-se  $R^2_{Ext} = 0,73$  e RMSEP= 0,042. O modelo 4 também passou pelos testes de análise de resíduos. O teste de Shapiro-Wilk teve valor de 0,13. O teste de Durbin-Watson teve valor de 2,49 e o teste de Brush-Pagan teve P-valor de 0,69.

#### 4.3.4 Estimativa de modelo com as séries mistas de compostos

Com o intuito de criar um modelo com uma diversidade estrutural mais abrangente, chegou-se ao conjunto de dados com uma mescla de observações de cada série. É importante salientar que se chegou a esse banco de dados após diversos procedimentos: avaliações de *boxplot*, experimentações de modelos e, em seguida, retirar pontos de alavancagem.

Dessa forma chegou se ao banco de dados com as 23 observações: 8, 9, 11, 13 e 14 (série **1-14**); 84, 88, 93, 94, 96 e 99 (série **77-99**); 100, 105, 106, 107, 109, 112, 113, 115, 117, 118 e 123 (série **100-124**).

Com esses dados foi possível construir o modelo 5 Eq. (23) e o modelo 6 Eq. (24).

$$pIC50 = 4,324 + 0,288 S.hbond + 0,0009 \Delta Hint \quad (23)$$

$$pIC50 = 4,212 + 0,229 S.hbond + 0,0009 \Delta Hint - 0,202 LUMO \quad (24)$$

Onde  $S_{\text{bond}}$ , descritor obtido nos cálculos de docagem, representa o termo da ligação de hidrogênio entre a proteína e o ligante,  $\Delta H_{\text{int}}$  é a entalpia de interação em meio aquoso, dado em kJ/mol, obtido através dos cálculos semiempíricos com o método PM7 do programa Mopac2016 com a constante dielétrica da água, e  $LUMO$  é o descritor da energia do LUMO, obtido com o método PM6 do programa Spartan'14.

O modelo 6 é uma espécie de aprimoramento do modelo 5, pois foi adicionada a variável  $LUMO$ . A inclusão dessa variável trouxe melhorias significativas nos coeficientes de ajuste, como pode ser visto na Tabela 9 que compara os modelos 5 e 6.

**Tabela 9:** parâmetros estatísticos dos modelos 5 e 6

Modelo	$R^2$	$R^2_{Ext}$	$Q^2$	$S_{PRESS}$	RMSEP	$k$	$n$
5	0,80	0,74	0,73	0,029	0,0225	2	23
6	0,88	0,83	0,83	0,023	0,0179	3	23

Ambos modelos passaram pelos testes de análise de resíduos, vide a Tabela 10.

**Tabela 10:** Análise de resíduos dos modelos 5 e 6

Modelo	Shapiro-Wilk	Brush-Pagan	Ddurbin-Watson
5	0,94	0,76	1,76
6	0,75	0,93	2,45

Tanto o modelo 5 quanto o modelo 6, apresentaram coeficientes de ajustes dentro do limite aceitável. De acordo com o princípio da parcimônia deve se optar pelo modelo mais simples, que nesse caso é o modelo 5, pois possui uma variável a menos. No entanto, seria razoável escolher um modelo que apresenta uma avaliação melhor, visto que ambos modelos possuem o número de variáveis dentro da recomendação que é 1 variável a cada 5 compostos.

A maneira correta de comparar modelos com número diferentes de variáveis é através do  $R_{ajust}$  e do p-valor; o modelo mais preditivo terá o maior valor de  $R_{ajust}$  e o menor para o p-valor (GAUDIO E ZADONADE, 2001). O modelo 6, que apresenta um  $R_{ajust} = 0,86$  e p-valor =  $1,68 \times 10^{-8}$ , tem uma capacidade preditiva superior ao modelo 5, que possui um  $R_{ajust} = 0,78$  e p-valor =  $1,99 \times 10^{-7}$ .

#### 4.4 Análise comparativa dos modelos quanto às avaliações estatísticas

Dentre os modelos discutidos na sessão 4.3, é preciso compará-los para definir qual o modelo mais apropriado. Já foram comparados anteriormente os modelos 2 com 3 e 5 com 6. Os modelos 3, 4 e 6 possuem 3 variáveis, sendo assim não precisam ser comparados em relação ao  $R_{ajust}$  e p-valor. A comparação mais adequada é com relação ao grau de previsibilidade. Observe, Tabela 11, que os valores de  $Q^2$  dos modelos 3 e 6 diferem apenas em 1%. Já com os valores de  $S_{PRESS}$  o modelo 6, apresenta um valor menor.

**Tabela 11:** Comparação dos parâmetros estatísticos dos modelos 3, 4 e 6

<b>Modelo</b>	<b><math>R^2</math></b>	<b><math>R^2_{Ext}</math></b>	<b><math>R_{ajust}</math></b>	<b><math>Q^2</math></b>	<b><math>S_{PRESS}</math></b>	<b>RMSEP</b>	<b><math>k</math></b>	<b><math>n</math></b>
3	0,88	0,84	0,86	0,84	0,056	0,042	3	20
4	0,82	0,73	0,79	0,78	0,063	0,042	3	22
6	0,88	0,83	0,86	0,83	0,023	0,017	3	23

Além de toda essa avaliação, deve se ressaltar que o conjunto de dados do modelo 3 advém de uma série de estruturas semelhantes que diferem entre si apenas nos substituintes, enquanto o conjunto de dados do modelo 6 provém de 3 séries distintas. Essa diversidade estrutural confere ao modelo 6 uma vantagem a mais, por ser mais abrangente do ponto de vista estrutural.

## 4.5 Análise dos termos envolvidos nos modelos

### 4.5.1 Energia dos orbitais LUMO

A utilização das energias dos orbitais de fronteira HOMO e LUMO como descritores moleculares em QSAR se baseia na teoria dos Orbitais de Fronteira, desenvolvida por Fukui e colaboradores na década de 1950. Fukui observou a reatividade de compostos aromáticos e percebeu que é possível prever o local preferencial de ataque de um reagente nucleofílico, eletrofílico ou radicalar de uma molécula através da densidade eletrônica dos orbitais HOMO e LUMO sobre cada átomo das moléculas envolvidas na reação química (FUKUI *et al.*, 1954; FUKUI *et al.*, 1952).

O LUMO é um orbital vazio e o fato da energia dele surgir como um descritor no modelo 4 (série dos compostos **100-124**) e no modelo 6 (séries mistas) é um indicativo de que ao menos para parte dos compostos selecionados a inibição da enzima pode envolver algum caráter eletrofílico deles, possivelmente atuando como inibidores que se ligam covalentemente à enzima. Todos os compostos estudados neste trabalho possuem grupos ureia ou tioureia. Há registros na literatura de inibidores enzimáticos da classe das uréias que causam a carbamoilação de resíduos do sítio ativo de enzimas, inativando-as (JOHNSON *et al.*, 2011).

Assim como no trabalho de LI e colaboradores os modelos 4 e 6 apresentam sinal negativo no coeficiente do LUMO (LI, *et al.*, 2010). Isso indica que a atividade inibitória pode aumentar (maior pIC<sub>50</sub>) à medida que a energia do orbital LUMO diminui. Isso está de acordo com a teoria dos orbitais de fronteira, porque quanto menor for a energia do LUMO do inibidor, mais fácil seria um ataque nucleofílico da enzima sobre esse inibidor, que levaria finalmente à carbamoilação e inativação da enzima.

#### 4.5.2 Entalpia de interação $\Delta H_{int}$ e entropia de interação $\Delta S_{int}$

A afinidade com que um ligante interage com a proteína pode ser definida pela energia livre de Gibbs ( $\Delta G$ ). Apesar de  $\Delta G$  não aparecer nos modelos, o entendimento desse termo permite uma análise da participação de algumas variáveis dos modelos. Existe uma relação Eq. (25) entre a energia de livre de Gibbs e as constantes de dissociação do complexo enzima-inibidor ( $K_i$ ), que em muitos casos é proporcional ao IC<sub>50</sub>:

$$\Delta G = RT \ln K_i \quad (25)$$

Onde,  $R$  é a constante universal dos gases (8,314 J.K<sup>-1</sup>.mol<sup>-1</sup>),  $T$  é a temperatura em Kelvin,  $K_i$  é constante de inibição, equivalente à constante de dissociação do complexo enzima-inibidor. Os termos e os princípios físico-químicos básicos do reconhecimento molecular através da termodinâmica podem ser

encontrados na literatura (DE RUITER e OOSTENBRINK, 2011; KROGSGAARD-LARSEN *et al.*, 2002).

A energia livre de interação entre uma enzima e um ligante ( $\Delta G_{int}$ ) envolve uma combinação de um termo entálpico e um entrópico envolvidos no processo de interação entre enzimas e ligantes para formar o complexo proteína-ligante Eq. (26) (BROOIJMANS E KUNTZ, 2003).

$$\Delta G_{int} = \Delta H_{int} - T\Delta S_{int} \quad (26)$$

Onde  $\Delta H_{int}$  é a entalpia de interação,  $T$  é a temperatura absoluta do sistema e  $\Delta S_{int}$  é a entropia de interação.

O termo  $\Delta H_{int}$  está em boa parte representado pelo descritor  $\Delta H_{int}$  presente nos modelos 3, 5 e 6 e é resultado da estabilização entálpica proveniente das interações intermoleculares.

O termo  $\Delta S_{int}$  descreve as variações nas entropias translacional, rotacional e conformacional dos ligantes pela interação com a cavidade do sítio ativo. Assumindo que a perda das entropias translacional e rotacional por interação com a enzima seria semelhante para todos os ligantes, já que são moléculas de massas parecidas, a parte de  $\Delta S_{int}$  mais importante para diferenciar o  $\Delta G_{int}$  de diferentes ligantes seria a entropia conformacional, que diz respeito às perdas nas torções das ligações do ligante ao se inserir no sítio ativo da enzima, que em solução aquosa giram livremente.

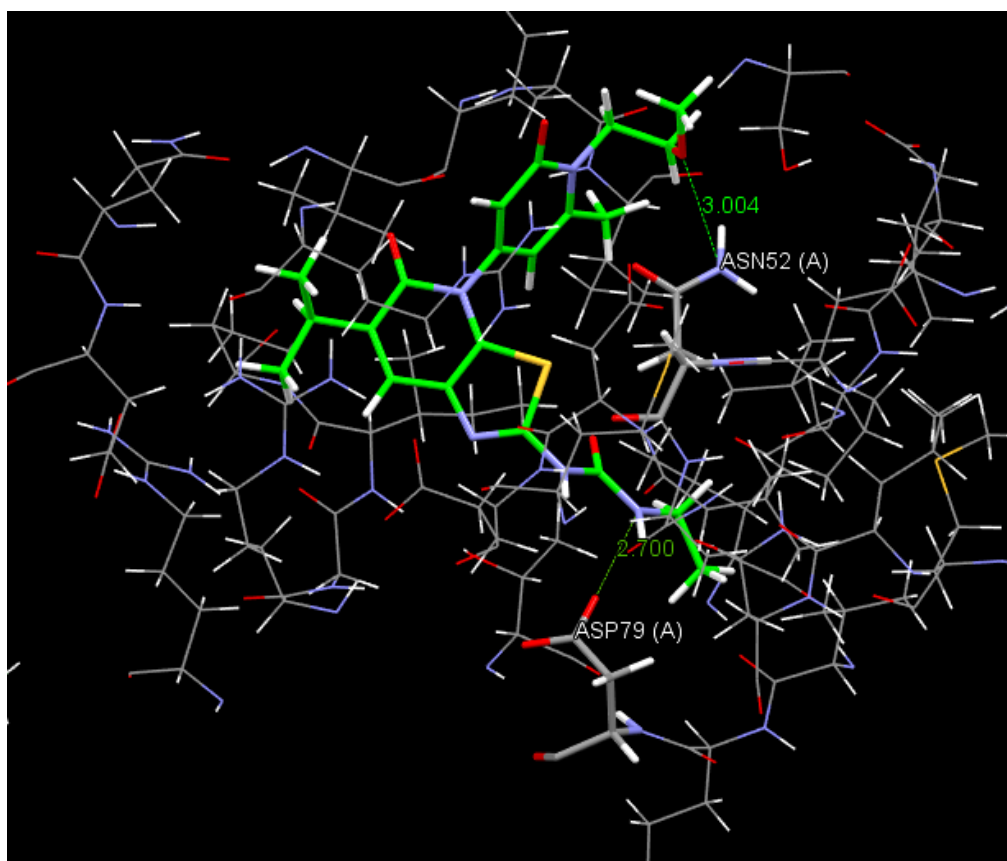
Essa perda de movimento resulta em uma diminuição da entropia, resultando em um  $\Delta S_{int}$  negativo e, conseqüentemente, um  $\Delta G_{int}$  menos favorável. O descritor  $S_{int}$ , obtido nos cálculos de docagem, contém um termo associado à torção interna do ligante; logo esse termo reflete em parte a entropia torcional do ligante que é perdida na interação com o sítio.

#### **4.5.3 Ligação de hidrogênio *S.hbond* e interação de *van der Waals S.vdw<sub>ext</sub>* entre ligante e da proteína**

A variável *S.hbond* representa o termo da ligação de hidrogênio entre a proteína e o ligante. As ligações de hidrogênio são interações intermoleculares de natureza

essencialmente eletrostática observadas entre um átomo receptor de ligação de hidrogênio e um doador de ligação de hidrogênio (BUGG, 2012; KROGSGAARD-LARSEN *et al.*, 2002). As ligações de hidrogênio ditas clássicas acontecem quando um átomo de hidrogênio ligado a nitrogênio, flúor ou oxigênio interage com um desses mesmos átomos em outra molécula, geralmente. A distância entre os átomos em uma ligação de hidrogênio deve ter um comprimento de ligação entre 2,5 e 3,0 Å (KROGSGAARD-LARSEN *et al.*, 2002).

Apenas como exemplo da participação de ligações de hidrogênio na interação ligante-enzima, na Figura 13 é possível perceber duas ligações de hidrogênio que ocorrem com o composto **91**, cuja atividade biológica  $pIC_{50}=5,60$ . Uma ligação de hidrogênio ocorre entre o oxigênio do resíduo Asp79 com um dos grupos NH do grupamento ureia e a outra entre o oxigênio do grupamento metoxila e a cadeia lateral do resíduo Asn52.



**Figura 13:** ligações de hidrogênio que ocorrem no composto **91**. As ligações de hidrogênio são representadas por linhas tracejadas em verde. Os átomos de C da estrutura do ligante estão em verde e os dos resíduos da enzima estão em cinza claro.

Em boa parte da literatura, as interações de van der Waals são descritas como interações que ocorrem entre os átomos de duas moléculas apolares (BUGG, 2012). A distância de ligação entre os núcleos dos átomos é o que determina se essas forças são atrativas ou repulsivas. Quando a distância ligação é muito curta há uma sobreposição das nuvens eletrônicas que ocasiona no aumento de energia. Essa área estericamente repulsiva tende a diminuir a afinidade de ligação receptor-ligante, enquanto nas distâncias mais longas a tendência é que exista uma região de atração entre os átomos devido as forças de dispersão ocasionadas por interações favoráveis entre dipolos instantâneos e induzidos em dois átomos adjacentes. Essa força de interação é pequena, em média 0,2 kJ/mol, se for considerar um único contato átomo-átomo (KROGSGAARD-LARSEN *et al.*, 2002). Entretanto, caso haja uma alta complementaridade geométrica no complexo proteína-ligante o somatório das interações de van der Waals pode ser significativo. A losartana, por exemplo, é um fármaco que faz importantes interações de van der Waals entre suas subunidades e os resíduos de aminoácidos hidrofóbicos (BARREIRA E FRAGA, 2014).

O método PM7 é capaz de calcular bem tanto ligações de hidrogênio quanto as forças de dispersão entre moléculas e, portanto, o efeito de ambas está incluído no termo  $\Delta H_{int}$ . No entanto, embora seu desempenho seja significativamente melhor do que dos métodos semiempíricos anteriores na descrição dessas interações, o método ainda apresentou um erro médio entre 0,74 e 0,78 kcal/mol para um conjunto de moléculas usadas na sua validação (STEWART, 2013). A inclusão de termos explícitos  $S.hbond$  e  $S.vdw_{ext}$  em modelos gerados neste trabalho pode ser resultado dessa limitação, permitindo a obtenção de melhores correlações com os dados experimentais.

## 5 Conclusão

Neste trabalho foram combinados métodos de modelagem molecular e de regressão linear múltipla para se construir modelos de predição de atividade inibitória da subunidade B da DNA girase ( $pIC_{50}$ ) para uma série de compostos da literatura com ação inibitória sobre a subunidade B da DNA girase de *M. smegmatis*, micobactéria do mesmo gênero do bacilo causador da TB, *M. tuberculosis*.



Diferente de estudos de QSAR tradicionais, em que se usam apenas descritores dos ligantes como variáveis independentes, aqui usamos os métodos de docagem molecular e cálculos quânticos semiempíricos para se gerar também descritores relacionados à formação dos complexos enzima-inibidor. A docagem molecular foi usada para se prever o modo de interação mais provável entre cada inibidor e a enzima, enquanto que os cálculos semiempíricos foram usados para se calcular a entalpia resultante da interação entre eles. Combinando-se dados dos dois métodos, foi possível se obter correlações estatisticamente significativas entre os resultados de atividade inibitória da enzima e os descritores selecionados.

Do ponto de vista estatístico os melhores modelos foram 3 e 6, que são praticamente semelhantes a níveis estatísticos, pois o coeficiente de ajuste  $R^2$  teve valor de 0,88 para ambos os modelos e o coeficiente de validação cruzada  $Q^2$  para o modelo 3 é 0,84 enquanto para o modelo 6 é 0,83.

Esses modelos, que foram calibrados por meio dos dados experimentais de atividade inibitória da enzima, podem servir como ferramentas para a predição de atividade de novas moléculas planejadas para agirem como inibidores da subunidade B da DNA girase ou em estudos de triagem virtual de novos candidatos a fármacos com ação antimicobacteriana, incluindo sobre o bacilo causador da TB, *M. tuberculosis*.

## 6 Referências Bibliográficas

ALCÁCER, L. Introdução à Química quântica computacional. Rio de Janeiro. Editora: IST Press, Lisboa, 2007. 305p. (Ensino da Ciência e da Tecnologia, 20)

AGGARWAL, C. C.; Análise Outlier. Springer New York, 2017, doi: [http://dx.doi.org/10.1007/978-3-319-47578-3\\_1](http://dx.doi.org/10.1007/978-3-319-47578-3_1)

ALVES, VINICIUS M.; BRAGA, RODOLPHO C.; MURATOV, EUGENE N. E ANDRADE, HORTA C. QUIMIOINFORMÁTICA: UMA INTRODUÇÃO. Quím. Nova, São Paulo, v. 41, n.

2, pág. 202-212, fevereiro de 2018. <https://doi.org/10.21577/0100-4042.20170145>

ANDRES, E. et al. Vitamin B12 (Cobalamin) Deficiency in Elderly Patients. *Canadian Medical Association Journal*, v. 171, p. 251-259, 2004.

AUBRY A, FISHER LM, JARLIER V, CAMBAU E. First functional characterization of a singly expressed bacterial type II topoisomerase: the enzyme from *Mycobacterium tuberculosis*. *Biochem Biophys Res Commun*. 2006 Sep 15;348(1):158-65. doi: 10.1016/j.bbrc.2006.07.017. Epub 2006 Jul 13. PMID: 16876125.

BARREIRO, E. J.; FRAGA, C. A. M.; RODRIGUES, C. R.; MIRANDA, A.L.P. Estratégias em Química medicinal para o planejamento de fármacos. *Revista Brasileira de Ciências Farmacêuticas*, v. 37, n. 3, 2001.

BARREIRO, E. J.; FRAGA, FRAGA, C. A. M. Química Medicinal-: As bases moleculares da ação dos fármacos. Artmed Editora, 2014.

BARREIRO, E. J.; FRAGA, C. A. M. Química Medicinal–As Bases Farmacológicas da Ação dos Fármacos. (3ªEd.). Artmed, 2015.

BASTOS, Francisco de Assis Amaral. Estatística e Probabilidade. 1ª Edição, Fortaleza, 2015.

BAUM, C.; WIGGINS, V. BPAGAN: Stata module to perform Breusch-Pagan test for heteroskedasticity. 1999.

BAXTER, K.; HORN, E.; GAL-EDD, N.; ZONNO, K.; O'LEARY, J.; TERRY, P. F.; TERRY, S. F. An end to the myth: there is no drug development pipeline. *Science translational medicine*, v. 5, n. 171, p. 171cm1, 6 fev. 2013.

BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N. & BOURNE, P. E. The Protein Data Bank. *Nucleic Acids Res.*, v. 28, p. 235-242, 2000. ([www.rcsb.org](http://www.rcsb.org)).

BHARTI, D. R; HEMROM, A.J.; LYNN, A.M. GCAC: sistema de fluxo de trabalho da galáxia para construção de modelos preditivos para triagem virtual. *BMC Bioinformatics* 19, 550 (2019) doi: 10.1186 / s12859-018-2492-8]

BORN, M. E OPPENHEIMER, J. R. ZUR Quantentheorie der Molekeln. *Ann. Phys.*, v.84, p.457-484, 1927.

BROADHURST, D., GOODACRE, R., JONES, A., ROWLAND, J. J., KELL, D. B.: Genetic algorithms as a method for variable selection in multiple linear regression and partial least squares regression, with applications to pyrolysis mass spectrometry. *Anal. Chim. Acta* 348, 1997.

BROOIJMANS, N. E KUNTZ, I. D. Molecular recognition and docking algorithms. *Annu. Rev. Biophys. Biomol. Struct.*, v. 32, p. 335-373, 2003.

BUGG, T. D. H. *Introduction to Enzyme and Coenzyme Chemistry: Second Edition*, 3o ed; John Wiley & Sons: West Sussex, 2012.

BURKERT, U.; Allinger, N. L.; *Molecular Mechanics*; ACS Monograph No.177; Washington, D. C. 1982. @article{doi:10.1128/ecosalplus.ESP-0010-2014.

BUSH G. N., EVANS-ROBERTS K., MAXWELL A. LOVETT S. T. DNA Topoisomerases. *EcoSal Plus*. volume 6, number 2. 2015. doi 10.1128/ecosalplus.ESP 0010-2014

BUSSAB, W. O. – *Análise de Variância e de Regressão*. Atual Editora, São Paulo, 1988.

CAMPOS, H. S. *Mycobacterium tuberculosis* resistente: de onde vem a resistência? *Bol. Pneumol. Sanit.*, Rio de Janeiro, v. 7, n. 1, p. 51-64, jun. 1999.

CARVALHO, I. *Introdução à modelagem molecular de fármacos no curso experimental de química farmacêutica*. *Quim. Nova*, 2003.

CCDC GOLD-Protein Ligand Docking Software. Available at: <http://www.ccdc.cam.ac.uk/solutions/csd-discovery/components/gold/> (Accessed January 22, 2020).

CHAMPOUX J. J. DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem.* 2001. doi: 10.1146/annurev.biochem.70.1.369. PMID: 11395412

CHAUDHARI, K.; SURANA, S.; JAIN, P.; PATEL, H. M. *Mycobacterium Tuberculosis (MTB) GyrB inhibitors: An attractive approach for developing novel drugs against TB*. *European Journal of Medicinal Chemistry* 2016. [<https://doi.org/10.1016/j.ejmech.2016.08.034>]

COHEN, N. C. *Guidebook on Molecular Modeling in Drug Design*. 1. ed. California: Academic Press, 1996.

CONSONNI, V.; TODESCHINI, R. Molecular descriptors. In: PUZYN, T.; LESZCZYNSKI, J.; CRONIN, M. T. (Eds.). *Recent Advances in QSAR Studies*. Dordrecht: Springer, 2010. p. 29.

CRONIN, M. T. Quantitative structure-activity relationships (QSARs) -- applications and methodology. In: PUZYN, T.; LESZCZYNSKI, J.; CRONIN, M. T. (Eds.). *Recent Advances in QSAR Studies*. Dordrecht: Springer, 2010. p. 3–11.

CUNHA, U. S.; MACHADO, S. A.; FIGUEIREDO, A. F. Uso de análise exploratória de dados e de regressão robusta na avaliação do crescimento de espécies comerciais de terra firme da Amazônia. *Revista Árvore* [online]. 2002, v. 26, n. 4, pp. 391-402. Disponível em: <<https://doi.org/10.1590/S0100-67622002000400001>>

DEARDEN, J. C.; CRONIN, M. T. D.; KAISER, K. L. E. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR and QSAR in environmental research*, v. 20, n. 3-4, p. 241–66, jan. 2009.

DELANO, Warren Lyford et al. *PyMOL*. 2002.

DELICISTIA, C. N. Modelagem molecular aplicada ao estudo de reações de inibição enzimática com aplicação potencial no controle de *Leishmania amazonensis*. Tese de Doutorado. UFRRJ. 2010.

De Ruiter, A.; Oostenbrink, C. Free energy calculations of protein-ligand interactions. *Curr. Opin. Chem. Biol.* 2011, 15 (4), 547–552 DOI: 10.1016/j.cbpa.2011.05.021.

DE SA ALVES, F., BARREIRO, E., & MANSSOUR FRAGA, C. (2009). From Nature to Drug Discovery: The Indole Scaffold as a 'Privileged Structure' Mini-Reviews in Medicinal Chemistry, 9 (7), 782-793. doi: 10.2174 / 138955709788452649

DESPOTOPOULOU, C., KLIER, L., & KNOCHEL, P. Synthesis of Fully Substituted Pyrazoles via Regio- and Chemoselective Metalations. *Organic Letters*, 11(15), 3326–3329. (2009) doi:10.1021/ol901208d

DEWAR, M. J. S. et al. Development and use of quantum mechanical molecular model.76.AM1: a new general purpose quantum mechanical molecular model. *Journal of the American Chemical Society*, v. 107, n. 13, p. 3902–3909, 1985.

DEWAR, M. J. S.; THIEL, W. Ground states of molecules. 38. The MNDO method. Approximations and parameters. *Journal of the American Society of Chemistry*, v. 99, n. 15,

p. 4899–4907, 1977.

DIAS, R. & AZEVEDO Jr., W. F. Molecular docking algorithms. *Curr. Drug Targets*, v. 9, p. 1040-1047, 2008.

DIMASI JA, Hansen RW, Grabowski HG. The price of innovation: new estimates of drug development costs. *J Health Econ.* 2003 Mar;22(2):151-85. doi: 10.1016/S0167-6296(02)00126-1. PMID: 12606142.

DRAPER, N. R.; Smith, H.; *Applied Regression Analysis*. John Wiley & Sons; New York, 1981.

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

FERNANDES, G. F. S.; CHIN, C. M.; SANTOS, J. L.; potenciais alvos moleculares para o desenvolvimento de novos fármacos antituberculose. *Quím. Nova*, São Paulo, v. 40, n. 5, 2017.

FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Seleção de variável em QSAR. *Química Nova*, v. 25, n. 3, pág. 439-448, 2002.

FERREIRA, R. S; GLAUCIUS, O.; ANDRICOPULO, A. D. Integração das técnicas de triagem virtual e triagem biológica automatizada em alta escala: oportunidades e desafios em P&D de fármacos. *Quím. Nova*, São Paulo, v. 34, n. 10, p. 1770-1778, 2011.

FORINA, M., LANTERI, S., CERRATO OLIVEROS, M.C., PIZARRO MILLAN, C.: Selection of useful predictors in multivariate calibration. *Anal. Bioanal. Chem.* 380, 2004.

FORRELAND, M., KLEPP, L., GIOFFRÉ, A., GARCÍA, J., MORBIDONI, H., SANTANGELO, M., CATALDI, A. E BIGI, F. (2013). Virulence factors of the *Mycobacterium tuberculosis* complex. *Virulence*, 4, pp. 2-49.

FRAGOSO T.P.; CARNEIRO J.W.; VARGAS M.D. Aminequinone-hydroxylquinoneimine tautomeric equilibrium revisited: molecular modeling study of the tautomeric equilibrium and substituent effects in 4-(4-R-phenylamino)naphthalene-1,2-diones. *J Mol Model.* 2010 May;16(5):825-30. doi: 10.1007/s00894-009-0579-x. Epub 2009 Sep 13. PMID: 19756783.

FRALEY, C., RAFTERY, A. How many clusters? Which clustering method? Answers via

model-based cluster analysis. *Comp. J.* 41, 1998.

FRANCO, DAIANA P. ET AL. A importância das cumarinas para a química medicinal e o desenvolvimento de compostos bioativos nos últimos anos. *química nova* [online]. 2021, v. 44, n. pp. 180-197. Epub 07 Maio 2021. ISSN 1678-7064. <https://doi.org/10.21577/0100-4042.20170654>.

FRANK, I. E., TODESCHINI, R.: *The Data Analysis Handbook*. Elsevier, Amsterdam, the Netherlands, 1994.

FRIEDMAN, L. M.; FURBERG, C. D.; DEMETS, D. L. *Fundamentals of Clinical Trials*. 4. ed. New York, NY: Springer, 2010.

FUKUI, K., YONEZAWA, T., & SHINGU, H. A Molecular Orbital Theory of Reactivity in Aromatic Hydrocarbons. *The Journal of Chemical Physics*, 20(4), 722–725. (1952). doi:10.1063/1.1700523

FUKUI, K., YONEZAWA, T., NAGATA, C., & SHINGU, H. Molecular Orbital Theory of Orientation in Aromatic, Heteroaromatic, and Other Conjugated Molecules. *The Journal of Chemical Physics*, 22(8), 1433–1442. (1954). doi:10.1063/1.1740412

FURNIVAL, G. M., WILSON, R. W. Regressions by leaps and bounds. *Technometrics* 16, 1974.

GAUDIO, A. C.; ZANDONADE, E. Proposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. *Quím. Nova*, São Paulo, v. 24, n. 5, p. 658-671, Oct. 2001. Available from <https://doi.org/10.1590/S0100-40422001000500013>.

GOHLKE, H.; HENDLICH, M.; KLEBE, G. Knowledge-based scoring function to predict protein-ligand interactions. *Journal of Molecular Biology*, v. 295, n. 2, p. 337-356, 2000.

GROISMAN, E., OCHMAN, H. (1997). How to become a pathogen. *Trends in Microbiology*. 1997, 2, pp. 289-93.

GUHA, R.; VAN DRIE, J. H. Structure-activity landscape index: identifying and quantifying activity cliffs. *Journal of chemical information and modeling*, v. 48, n. 3, p. 646–58, mar. 2008.

GUIDO, R. V. C. Planejamento de inibidores da enzima gliceraldeído-3-fosfato desidrogenase de *Trypanosoma cruzi*: biologia estrutural e química medicinal. Universidade de São Paulo,

2008.

GIESEN, D., HAWKINS, G., LIOTARD, D. et al. A universal model for the quantum mechanical calculation of free energies of solvation in non-aqueous solvents. *Theor. Chem. Acc.* 98, 85–109 (1997). <https://doi.org/10.1007/s002140050283>

GUNSTEREN, V. W. F. .; BERENDSEN, H. JC. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angewandte Chemie International Edition in English*, v. 29, n. 9, p. 992-1023, 1990.

GUPTA, RR, KUMAR, M., & GUPTA, V. *Heterocyclic Chemistry*. (1998). doi: 10.1007 / 978-3-642-72276-9

HANSCH, C.; FUJITA, T.; *J. Am. Chem. Soc.* 1964, 86, 1616.

HÄRMARK, L.; VAN GROOTHEEST, A. C. Pharmacovigilance: methods, recent developments, and future perspectives. *European journal of clinical pharmacology*, v. 64, n. 8, p. 743–52, ago. 2008.

HOAGLAND, D. T.; Liu, J.; Lee, R. B.; Lee, Richard E. New agents for the treatment of drug-resistant *Mycobacterium tuberculosis*. *Advanced Drug Delivery Reviews* 2016.

HOERL, A. E., KENNARD, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12, 1970.

HOESKULDSSON, A. PLS regression methods. *Journal Chemometrics*. V. 2, 1988.

HOFMANN, M., GATU, C., KONTOGHORGHES, E. J. Efficient algorithms for computing the best subset regression models for large-scale problems. *Computat. Stat. Data Anal.* 52, 2007, 16–29

HOFFMANN, R. *Análise de regressão: uma introdução à Econometria*. 4ª ed. São Paulo: Hucitec, 2006.

HOUSE, J. E. *Fundamentals of Quantum Chemistry*, Elsevier, San Diego, 2004.

HUANG, H.; YU, H.; CHEN, C.; HSU, C.; CHEN, H. LEE, K.; TSAI, F. & CHEN, C. Y. Current developments of computer-aided drug design. *J. Taiwan Inst Chem. Eng.*, v. 41, n. 6, p. 623-635, 2010.

IUPAC. Compendium of Chemical Terminology, 2<sup>a</sup> ed. (o "Livro de Ouro"). Compilado por AD McNaught e A. Wilkinson. Blackwell Scientific Publications, Oxford (1997). Versão online (2019-) criada por SJ Chalk. ISBN 0-9678550-9-8. <https://doi.org/10.1351/goldbook>.

JELESAROV, I.; BOSSHARD, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of biomolecular recognition. *J Mol Recognit.* 1999 Jan-Feb;12(1):3-18. doi: 10.1002/(SICI)1099-1352(199901/02)12:1<3::AID-JMR441>3.0.CO;2-6. PMID: 10398392.

JOHNSON, R. A. e WICHERN, Dean W. *Applied Multivariate Statistical Analysis*. Englewood Cliffs, NJ: Prentice Hall, 1988.

JOHNSON, D.S. et al. Discovery of PF-04457845: A Highly Potent, Orally Bioavailable, and Selective Urea FAAH Inhibitor. *ACS Med. Chem. Lett.*; v. 2, p. 91-96, 2011. doi: 10.1021/ml100190t. PMID: 21666860; PMCID: PMC3109749.

JONES, G. et al. Development and Validation of a Genetic Algorithm for Flexible Docking. *Journal of Molecular Biology*, v. 267, n. 3, p. 727-748, 1997.

JORGENSEN, W.; The many roles of computation in drug discovery. *Science*. 2004 Mar 19;303(5665):1813-8. doi: 10.1126/science.1096361. PMID: 15031495.

JORGENSEN, W.; *Efficient Drug Lead Discovery and Optimization*. *Accounts of chemical research*, (2009).

KASHYAP, A.; Kumar, P. S.; Silakari, O. In silico designing of domain B selective gyrase inhibitors for effective treatment of resistant tuberculosis. *Tuberculosis*, v. 112, P. 83-88, 2018. ISSN 1472-9792.

KAMSRI, P.; Punkvang, A.; Hannongbua, S.; Suttisintong, K.; Kittakoo, P.; Spencer, P.; Mulholland, A.J.; Pungpo, P. In silico study directed towards identification of the key structural features of GyrB inhibitors targeting MTB DNA gyrase: HQSAR, CoMSIA and molecular dynamics simulations, SAR and QSAR in *Environmental Research*, 30:11, 775-800, (2019). DOI: 10.1080/1062936X.2019.1658218

KIRALJ, Rudolf; FERREIRA, Márcia. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. *Journal of the Brazilian Chemical Society*, v. 20, n. 4, p. 770-787, 2009.



KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nature Reviews in Drug Discovery*, v. 3, n. 11, p. 935-949, 2004.

KORB, O.; STÜTZLE, T.; EXNER, T. E. Empirical Scoring Functions for Advanced Protein-Ligand Docking with PLANTS. *Journal of chemical information and modeling*, v. 49, n. 1, p. 84–96, 2009.

KRAMER, R.: *Chemometric Techniques for Quantitative Analysis*. Marcel Dekker, New York, 1998.

KROEMER, R.T. *Structure-based drug design: Docking and scoring*. Current protein and peptide science, 2007.

KROGSGAARD-LARSEN, P.; Strømgaard, K.; Madsen, U. *Textbook of drug design and discovery*, 4oed; CRC Press: Boca Raton, 2002.

KUBINYI, H.; QSAR: Hansch Analysis and Related Approaches. In: *Methods and Principles in Medicinal Chemistry*; R. Mannhold, P. Krogsgaard-Larsen e H. Timmerman Eds.; Vol. 1; VCH; Weinheim, 1993.

KUMAR, K.; Abubakar I. Clinical implications of the global multidrug-resistant tuberculosis epidemic. *Clinical Clinic, Journal of Royal College of Physicians of London*. 2015. *Clinical Medicine* 2015 Vol 15, No 6: s37–s42. DOI: <https://doi.org/10.7861/clinmedicine.15-6-s37>

KUNTZ, I.D. et al. Geometric approach to macromolecule-ligand interactions. *Journal of Molecular Biology*, v.161, p.269-288. 1982.

KUNTZ, I. D. Structure-based strategies for drug design and discovery. *Science*, v. 257, n. 5073, p. 1078-1082, 1992.

LAVECCHIA, A. & DI GIOVANNI, C. Virtual screening strategies in drug discovery: a critical review. *Curr. Med. Chem.*, v. 20, n. 23, p. 2839-2860, 2013.

LI, Bo-Jian; CHIANG, Chih-Chia; HSU, Ling-Yih. QSAR Studies of 3, 3'-(Substituted-Benzylidene)-Bis-4-Hydroxycoumarin, Potential HIV-1 Integrase Inhibitor. *Journal of the Chinese Chemical Society*, v. 57, n. 4A, p. 742-749, 2010.

LIMA, V., BONATO, V., LIMA, K., SANTOS, S., SANTOS, R., GONÇALVES, E., BRANDÃO, L. RODRIGUES, J. E SILVA, C. (2001). Role of trehalose dimycolate in recruitment of cells

and modulation of production of cytokines and no in tuberculosis. *Infection and Immunity*, 69, p. 5305.

LINDOSO, J.A.L.; Lindoso, A.A.B.P. - Neglected tropical diseases in Brazil. *Rev. Inst. Med. trop. S. Paulo*, 51(5): 247-253, 2009.

LOMBARDINO, J. G.; LOWE, J. A. The role of the medicinal chemist in drug discovery--then and now. *Nature reviews. Drug discovery*, v. 3, n. 10, p. 853–62, out. 2004.

MACALINO, S. J. Y.; GOSU, V.; HONG, S. & CHOI, S. Role of computer-aided drug design in modern drug discovery. *Arch. Pharm. Res.*, v. 38, n. 9, p. 1686-1701, 2015

MAGGIORA, G. M. On outliers and activity cliffs--why QSAR often disappoints. *Journal of chemical information and modeling*, v. 46, n. 4, p. 1535, 2006.

MALLOWS, C. L.: *Technometrics* 15, 1973, 661–675. Some comments on Cp

MCGEE, P. Clinical trials on the move. *Drug discovery & development*, v. 9, n. 6, p. 16–22, 2006.

MEDAPI, B.; Renuka, J.; Saxena, S.; Sridevi, J.P; Medishetti, R.; Kulkarni, P.; YOGEE SWARI, P.; SRIRAM, D.; Design and synthesis of novel quinoline–aminopiperidine hybrid analogues as *Mycobacterium tuberculosis* DNA gyraseB inhibitors, *Bioorganic & Medicinal Chemistry*, Volume 23, Issue 9, Pages 2062-2078, ISSN 0968-0896. 2015. <https://doi.org/10.1016/j.bmc.2015.03.004>.

MELO, JÚLIO O. F. et al. Heterociclos 1,2,3-triazólicos: histórico, métodos de preparação, aplicações e atividades farmacológicas. *Química Nova* [online]. 2006, v. 29, n. 3, pp. 569-579. Disponível em: <<https://doi.org/10.1590/S0100-40422006000300028>>. Epub 25 Maio 2006. ISSN 1678-7064. <https://doi.org/10.1590/S0100-40422006000300028>.

MENDOZA, L. Phase 0 clinical trials will overcome stagnation of anticancer drug development? *Klinická onkologie : casopis České a Slovenské onkologické společnosti*, v. 24, n. 2, p. 143–5, jan. 2011.

MIKUSOVÁ K, SLAYDEN RA, BESRA GS, BRENNAN PJ. Biogenesis of the mycobacterial cell wall and the site of action of ethambutol. *Antimicrob Agents Chemother*. 1995 Nov;39(11):2484-9. doi: 10.1128/AAC.39.11.2484.

MOOJI, W. T. M.; VERDONK, M. L. General and Targeted Statistical Potentials for Protein-Ligand Interactions. *PROTEINS: Structure, Function, and Bioinformatics*, v. 61, n. 2, p. 272-287, 2005.

MOPAC2016, JAMES J. P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA. 2016.

MORGON, N. H.; COUTINHO, K. Métodos de química teórica e modelagem molecular. São Paulo. Livraria da física, 2007.

MUKAKA, Mavuto M. A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal*, v. 24, n. 3, p. 69-71, 2012.

NETER, J., Kutner, M. H., Nachtsheim, C. J., and Wasserman, W. *Applied Linear Statistical Models* (4th edition). Boston: McGraw-Hill. (1996).

OECD. Principles for the validation, for regulatory purposes, of (Quantitative) Structure-Activity Relationship models. (2004). Disponível em: <<http://www.oecd.org/chemicalsafety/risk-assessment/37849783.pdf>>

OPREA, T. I. *Cheminformatics in drug discovery*. Weinheim: Wiley-VCH, p.493. 2005.

OREN, I. et al. Synthesis and Antimicrobial Activity of Some Novel 2,5-and /or 6-Substituted Benzoxazole and Benzimidazole Derivatives. *European Journal of Pharmaceutical Sciences*, v. 7, p. 153-160, 1998

PEREIRA, D. A. & WILLIAMS, J. A. Origin and evolution of high throughput screening. *Br. J. Pharmacol.*, v. 152, n. 1, p. 53-61, 2007.

PEREIRA, D. H.; LA PORTA, F. A.; SANTIAGO, R. T.; GARCIA, D. R.; RAMALHO, T. C. Novas Perspectivas sobre o Papel dos Orbitais Moleculares de Fronteira no Estudo Da Reatividade Química: Uma Revisão. *Rev. Virtual Quim.*, 2016, 8 (2), 425-453.

PIOVESAN, M. F.; LABRA, M. E. Institutional change and political decision-making in the creation of the Brazilian National Health Surveillance Agency. *Cadernos de Saúde Pública*, v. 23, n. 6, p. 1373–1382, jun. 2007.

PHARMA. Pharmaceutical research and manufacturers of America, 2013 biopharmaceutical research industry profile. Washington, DC: [s.n.]. Disponível em: <<http://www.phrma.org/profiles-reports>>.

R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.

RADÜNZ, L.L. et al. Rendimento extrativo de cumarina de folhas de guaco (*Mikania glomerata* Sprengel) submetidas a diferentes temperaturas de secagem. *Revista Brasileira de Plantas Mediciniais* [online]. 2012, v. 14, n. 3, pp. 453-457. Disponível em: <<https://doi.org/10.1590/S1516-05722012000300005>>. Epub 03 Dez 2012. ISSN 1983-084X. <https://doi.org/10.1590/S1516-05722012000300005>.

REECE, R. J.; MAXWELL, A. DNA gyrase: structure and function. *Critical Reviews in Biochemistry and Molecular Biology*, v. 26, n. 3-4, p. 335-375, 1991.

REDDY, K. I.; Srihari, K.; Renuka, J.; Sree, K. S.; Chuppala, A.; Jeankumar, V. U.; Sridevi, J. P.; Babu, K. S.; Yogeeswari, P.; Sriram, D.; An efficient synthesis and biological screening of benzofuran and benzo[d]isothiazole derivatives for *Mycobacterium tuberculosis* DNA Gyrase inhibition, *Bioorganic & Medicinal Chemistry*, Volume 22, Issue 23, Pages 6552-6563, ISSN 0968-0896, 2014. <https://doi.org/10.1016/j.bmc.2014.10.016>.

REGNSTROM, J.; KOENIG, F.; ARONSSON, B.; REIMER, T.; SVENDSEN, K.; TSIGKOS, S.; FLAMION, B.; EICHLER, H.-G.; VAMVAKAS, S. Factors associated with success of market authorisation applications for pharmaceutical drugs submitted 113 to the European Medicines Agency. *European journal of clinical pharmacology*, v. 66, n. 1, p. 39–48, janeiro. 2010.

RENUKA, J.; INDRASENA, K.; SRIHARI, K.; ULLAS, V.; SHRAVAN, M.; PADMA, J.; YOGEEESWARI, P.; SUDHAKAR, K.; SRIRAM, D.; *Bioorg. Med. Chem.* 2014, 22, 4924

RODRIGUES, R. P.; MANTOANI, S. P.; DE ALMEIDA, J. R.; PINSETTA, F. R. SEMIGUINI, E. P.; DA SILVA, V. B. & DA SILVA, C. H. P. Estratégias de triagem virtual no planejamento de fármacos. *Rev. Virtual Química*, v. 4, n. 6, p. 739-776, 2012.

ROSSETTI, Maria Lúcia Rosa et al. Tuberculose resistente: revisão molecular. *Revista de Saúde Pública* [online]. 2002, v. 36, n. 4. Epub 30 Set 2002. ISSN 1518-8787. <https://doi.org/10.1590/S0034-89102002000400021>.

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., Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. 2009. *Rev. Virtual Quim.*, 2009, 1 (1), 49-57.

SANTOS, V., GONSALVES, A., & ARAÚJO, C. Abordagem didática para o desenvolvimento de moléculas bioativas: regra dos cinco de lipinski e preparação de heterociclo 1,3,4-oxadiazol em forno de micro-ondas doméstico. *química nova*. (2017).. <https://doi.org/10.21577/0100-4042.20170135>

SCOTTI, L.; SCOTTI, M.; Cardoso, C.; Pauletti, P.; Gamboa, I. C.; Bolzani, V. S.; Velasco, M. V. R.; Menezes, C. M. S.; FERREIRA, E. I. Modelagem molecular aplicada ao desenvolvimento de moléculas com atividade antioxidante visando ao uso cosmético. *Revista Brasileira de Ciências Farmacêuticas*, v. 43, n. 2, 2007.

SCHENONE, M.; Dančík, V.; Wagner, BK.; Clemons, PA.; Target identification and mechanism of action in chemical biology and drug discovery. *Nat Chem Biol*. 2013 Apr;9(4):232-40. doi: 10.1038/nchembio.1199. PMID: 23508189; PMCID: PMC5543995.

SILVER, N. O Sinal e o Ruído. Editora Intrinseca, 2013.

SMITH, I. Mycobacterium tuberculosis pathogenesis and molecular determinants of virulence. *Clinical Microbiology*, 16, pp. 463-496, (2003).

STEINSKOG, D. J.; TJØSTHEIM, D. B.; KVAMSTØ, N. G. A cautionary note on the use of the Kolmogorov–Smirnov test for normality. *Monthly Weather Review*, v. 135, n. 3, p. 1151-1157, 2007.

STEWART, J. J. P. Optimization of parameters for semi-empirical methods. *Journal of Computational Chemistry*, Colorado Springs, Colorado , v. 10, n. 2, p. 221-264, 1989.

STEWART, J. J. P. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements. *Journal of Molecular Modeling*, v. 10, n. 2, p. 155, 2004. <https://doi.org/10.1007/s00894-004-0183-z>

STEWART, J. J. P. Optimization of parameters for semiempirical methods V: Modification of NDDO approximations and applications to 70 elements. *Journal of Molecular Modeling*, v. 13, n. 12, p. 1173-1213, 2007.

STEWART, J. J. P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters. *Journal of Molecular Modeling*, v. 19, n. 1, p. 1-32, 2013.

STONE, M., BROOKS, R. J. Continuum regression: Cross-validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal component regression. *J. R. Statist. Soc. B.* 52, 1990.

SUTORMIN D., RUBANOVA N., LOGACHEVA M., GHILAROV D., SEVERINOV K. Single-nucleotide-resolution mapping of DNA gyrase cleavage sites across the *Escherichia coli* genome. *Nucleic acids research*. Volume 47, 2018 <https://doi.org/10.1093/nar/gky1222>

TCHOUGRÉEFF, A.L. *Hybrid Methods of Molecular Modeling*. Ed. New York: Springer, 2008.

TIBSHIRANI, R. J. Regression shrinkage and selection via the lasso *J. Royal Stat. Soc. B.* 58, 1996.

TILLMAN, John A. The power of the Durbin-Watson test. *Econometrica: Journal of the Econometric Society*, p. 959-974, 1975.

The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

TODESCHINI, R.; CONSONNI, V. *Handbook of Molecular Descriptors*, Wiley. 2000.

TROPSHA, A. Best practices for QSAR model development, validation, and exploitation. *Molecular informatics*, v. 29, n. 6-7, p. 476–488, 6 jul. 2010.

UNGER, S. H.; HANSCH, C. Model building in structure-activity relations. Reexamination of adrenergic blocking activity of. beta.-halo-. beta.-arylalkylamines. *Journal of medicinal chemistry*, v. 16, n. 7, p. 745-749, 1973.

VARMUZA, K.; FILZMOSER, P. *Introduction to Multivariate Statistical Analysis in Chemometrics*, New York: CRC Press, 2009. ISBN-10: 1420059475 e ISBN-13: 978-1420059472.

VERLI, H.; Barreiro, E. J. Um paradigma da química medicinal: a flexibilidade dos ligantes e receptores. *Quím. Nova* v.28, n.1, 2005.

WONG, Tzu-Tsung. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognition*, v. 48, n. 9, p. 2839-2846, 2015.

WORLD HEALTH ORGANIZATION. Global tuberculosis report 2020. Geneva; 2020. Licence: CC BY-NC-SA 3.0 IGO.

WORLD HEALTH ORGANIZATION. Estimates of TB and MDR-TB burden are produced by WHO in consultation with countries 2019.

WORLD HEALTH ORGANIZATION. The end TB strategy. Global strategy and targets for tuberculosis prevention, care and control after 2015. Geneva: WHO, 2015.

XUE, L.; BAJORATH, J. Molecular descriptors in chemoinformatics, computational combinatorial chemistry, and virtual screening. *Combinatorial chemistry & high throughput screening*, v. 3, n. 5, p. 363–72, out. 2000.

YAO, L.; Evans, J.A; Rzhetsky, A. Novel opportunities for computational biology and sociology in drug discovery. *Trends Biotechnol.* V. 27, 2009.

## 7 Anexos

**Tabela A 1:** dados da atividade biológica pIC<sub>50</sub> de acordo com as estruturas. Seguidos dos descritores HOMO, LUMO e Entalpia de interação.

Composto	pIC <sub>50</sub> na GyrB de <i>M. smegmatis</i>	HOMO em eV	LUMO em eV	Esolv em kJ/mol	Entalpia EPS=4 em kJ/mol	Entalpia Eps=78 em kJ/mol
1	4,77	-8,15	-0,94	-60,71	-43,51	-204,30
2	4,68	-8,22	-0,99	-60,59	-488,89	-405,69
3	5,10	-8,26	-1,00	-82,81	-68,19	25,68
4	5,29	-8,39	-1,27	-81,68	-250,29	-152,82
5	5,09	-8,22	-0,76	-83,61	-266,70	-157,73
6	5,36	-8,17	-0,75	-69,89	85,64	163,78
7	5,53	-8,10	-0,91	-62,72	-173,93	-132,01
8	4,81	-8,21	-0,96	-87,49	-43,31	7,71
9	4,45	-8,28	-1,03	-87,40	-342,75	-320,51
10	4,56	-8,31	-1,03	107,20	-300,19	-343,32
11	4,89	-8,44	-1,45	106,06	74,32	44,56
12	4,61	-8,18	-0,96	-99,22	114,21	137,53
13	4,69	-8,28	-1,04	-83,65	-74,83	23,73
14	4,52	-8,19	-0,95	-83,65	334,19	-298,64

15	4,70	-8,15	-0,94	-60,71	-188,74	-47,71
16	5,75	-8,17	-0,71	-114,03	-331,46	-109,18
17	4,53	-8,70	-0,88	-109,17	-244,48	-117,67
18	4,57	-8,33	-1,23	-134,04	-84,21	340,67
19	4,58	-7,98	-0,55	-123,81	-696,73	-534,25
20	5,63	-8,02	-0,58	-115,48	-269,46	-134,16
21	5,16	-8,06	-0,62	-125,59	-390,39	-509,18
22	5,04	-8,16	-0,71	-142,48	28,00	-132,92
23	5,03	-8,23	-0,76	-141,63	-248,16	-127,49
24	4,96	-8,57	-0,93	-137,09	-413,48	-322,23
25	5,74	-8,38	1,40	-85,99	-328,33	-269,39
26	5,54	-8,07	-0,62	-85,43	-444,13	-259,48
27	4,34	-8,12	-0,67	-78,70	9,71	475,09
28	4,95	-8,17	-0,71	-88,28	-444,19	-427,10
29	4,76	-8,31	-0,33	-43,65	-295,42	-168,58
30	4,91	-8,34	-0,53	-61,26	-1141,77	-979,73
31	4,46	-8,36	-0,37	-40,01	-354,52	-235,66
32	5,42	-8,48	-1,50	-65,76	-349,30	-151,39
33	4,89	-8,37	-0,38	-44,20	-479,63	-308,28
34	5,14	-8,30	-0,46	-57,23	-354,67	-177,84
35	4,77	-8,50	-1,30	-51,08	-274,68	-125,09
36	4,71	-8,26	-0,54	-62,64	-405,69	-209,89
37	4,69	-8,26	-0,29	-78,08	-375,87	-247,42
38	5,67	-8,09	-0,40	-63,82	-188,26	-64,40
39	4,66	-8,26	-0,35	-44,84	-619,06	-483,28
40	5,06	-8,13	-0,44	-59,76	-122,71	-98,58
41	4,91	-8,12	-0,42	-46,89	-287,34	-110,34
42	6,02	-8,14	-0,56	-63,97	-291,77	-202,85
43	4,73	-8,16	-0,46	-43,31	-281,35	-139,08
44	5,60	-8,27	-1,52	-68,76	-626,36	-168,88
45	4,79	-8,17	0,47	-47,53	-461,27	-332,12
46	4,95	-8,08	-0,50	-58,73	-309,44	-368,56
47	5,16	-8,29	1,32	-54,63	-284,96	-129,10
48	5,00	-8,10	-0,57	-66,51	-377,70	-157,76
49	4,94	-8,07	-0,38	-42,68	-298,48	-238,25
50	4,17	-8,42	-0,64	-53,05	-310,21	-244,38
51	5,03	-8,10	-0,41	-53,57	-220,14	-77,03
52	4,34	-8,47	-0,68	-48,79	-267,34	-250,82
53	4,69	-8,43	-0,64	-35,79	-337,35	-248,37
54	4,39	-8,48	-0,69	-52,98	-796,58	-836,35
55	5,75	-8,47	-0,68	-32,08	-368,50	-172,30
56	4,25	-8,61	-1,49	-57,40	-726,80	-699,23
57	4,31	-8,49	-0,69	-36,32	-424,41	-294,32
58	4,23	-8,40	-0,62	-52,44	-285,79	-181,71
59	5,17	-8,62	-1,29	-43,02	-223,25	-120,49
60	4,40	-8,38	-0,60	-54,88	-383,70	-204,37



61	4,82	-8,40	-0,62	-35,08	-361,42	-182,75
62	4,41	-8,66	-0,87	-50,32	-299,60	-79,94
63	4,99	-8,35	-0,62	-42,83	-79,93	14,91
64	4,31	-8,74	-0,91	-46,06	-416,71	-389,45
65	4,83	-8,71	-0,86	-33,83	-210,91	-95,52
66	4,78	-8,35	-0,92	-50,35	-330,66	-253,78
67	4,70	-8,76	-0,90	-30,21	-266,49	-37,55
68	4,38	-8,89	-1,52	-54,92	-725,56	-166,54
69	4,47	-8,77	-0,91	-34,46	-293,38	-117,78
70	4,61	-8,59	-0,85	-49,72	-386,53	-186,18
71	4,32	-8,90	-1,32	-41,43	-236,72	-139,23
72	4,49	-8,55	-0,84	-56,98	-304,19	-67,20
73	4,78	-8,65	-0,84	-33,10	-222,24	-141,69
74	4,39	-8,38	-0,84	-40,98	-267,65	-141,25
75	4,90	-8,29	-0,30	-61,22	-295,83	-179,46
76	4,53	-8,33	-0,42	-57,03	-257,72	-275,87
77	4,96	-8,53	-0,96	-110,28	-219,22	-176,12
78	4,70	-8,71	-1,29	0,00	-597,03	-521,80
79	4,80	-8,54	-0,97	-89,21	-171,67	-121,54
80	3,66	-8,63	-1,03	-11,46	-235,75	-307,88
81	5,60	-8,38	-0,56	-91,74	-94,18	-117,24
82	4,64	-8,49	-0,87	-102,86	-231,62	-134,47
83	5,00	-8,48	-0,79	-79,37	-151,52	-84,80
84	4,64	-8,24	-0,45	-89,04	-184,57	-101,80
85	5,22	-8,33	-0,58	-98,01	-116,54	-52,90
86	4,82	-8,38	-0,63	-106,90	-112,25	-129,84
87	5,40	-8,42	-0,65	-112,30	-176,49	-145,08
88	4,70	-8,54	-1,26	-101,43	-208,97	-246,22
89	5,30	-8,33	-0,49	-92,21	-159,41	-109,35
90	5,30	-8,44	-0,66	-99,41	-286,13	-78,73
91	5,60	-8,39	-0,58	-104,00	-117,68	-93,01
92	4,08	-8,63	-0,95	-110,85	-196,26	-191,10
93	4,52	-8,36	-0,58	-100,64	-144,83	-180,61
94	4,40	-8,36	-0,57	-97,99	-50,54	-164,04
95	4,40	-8,59	-0,79	-97,14	-221,61	-157,66
96	4,60	-8,32	-0,55	-94,19	-191,72	-200,07
97	4,30	-8,36	-0,59	-88,49	-123,23	-81,21
98	4,40	-8,33	-0,56	-92,97	-238,35	-139,16
99	4,85	-8,39	-0,61	-117,59	-187,17	-12,13
100	4,66	-9,01	-1,50	-103,81	-265,24	-205,78
101	5,43	-8,67	-1,18	-109,17	-332,87	-291,43
102	4,06	-9,22	-1,36	-89,07	-158,28	-203,97
103	4,34	-8,92	-1,37	-88,87	-77,94	-122,83
104	4,60	-8,90	-1,34	-93,22	-259,49	-275,21

continua

**Tabela A**      **continuação**

Composto	pIC <sub>50</sub> na GyrB	HOMO	LUMO	Esolv	Entalpia	Entalpia
105	4,92	-8,86	-1,29	-90,01	-390,44	-244,08
106	5,05	-8,87	-1,31	-103,92	-233,47	-186,30
107	4,89	-8,94	-1,36	-104,95	-363,93	-206,72
108	4,41	-9,14	-1,43	-116,16	-281,28	-189,48
109	4,77	-9,06	-1,54	-107,92	-248,83	-294,40
110	4,52	-8,51	-1,01	-106,00	-37,03	-38,73
111	4,52	-8,61	-1,07	-109,59	-199,85	-185,16
112	5,15	-9,02	-1,49	-98,56	-223,96	-174,61
113	4,96	-8,77	-1,25	-117,04	-269,97	-233,76
114	5,10	8,87	-1,38	-117,73	-423,10	-257,41
115	5,05	-8,99	-1,44	-119,08	-314,16	-238,50
116	4,85	-8,89	-1,36	-124,70	-258,15	-231,86
117	4,92	-8,89	-1,33	-142,14	-318,20	-298,12
118	5,00	-8,85	-1,29	129,20	-202,52	-188,62
119	5,40	-8,89	-1,32	-144,45	15,63	177,24
120	5,00	-8,48	-1,09	-105,24	-228,73	-126,18
121	6,00	-8,97	-1,32	-113,30	-276,52	-276,58
122	5,34	-8,83	-1,33	-101,33	-425,20	-235,59
123	5,10	-8,78	-1,31	-90,94	-417,13	-286,01
124	4,33	-8,47	-0,94	-117,19	-273,25	-183,58
125	3,80	-9,00	-0,88	-109,87	-306,83	-197,93

**Tabela A 2:** Descritores obtidos através dos cálculos de docagem com a função ChemScore

composto	Score	DG	S(hbond)	S(lipo)	H(rot)	DE(clash)	DE(int)	intcor
1	32,15	-34,33	1,50	228,73	1,14	1,98	0,20	1,36
2	31,55	-34,65	1,49	231,71	1,14	2,33	0,77	1,40
3	29,91	-30,81	1,51	198,43	1,15	0,14	0,76	1,38
4	27,36	-28,44	1,11	189,93	1,17	0,75	0,32	1,38
5	31,06	-33,81	1,48	228,79	1,32	1,81	0,93	1,39
6	35,71	-37,35	1,57	257,64	1,37	0,99	0,65	0,04
7	33,91	-40,36	1,60	282,45	1,37	6,08	0,37	0,07
8	31,53	-34,20	1,63	223,91	1,14	2,27	0,41	0,46
9	30,52	-34,75	1,38	235,77	1,14	2,51	1,72	0,50
10	27,88	-30,14	1,50	192,94	1,15	0,91	1,35	0,40
11	25,55	-27,53	1,24	178,60	1,17	0,56	1,42	0,36
12	29,50	-33,35	1,42	226,67	1,32	2,27	1,59	0,56
13	33,49	-36,88	1,51	255,31	1,37	2,16	1,23	0,02

continua

Tabela A 2 continuação

composto	Score	DG	S(hbond)	S(lipo)	H(rot)	DE(clash)	DE(int)	intcor
14	34,55	-39,50	1,61	274,76	1,37	4,08	0,87	0,07
15	32,15	-34,33	1,50	228,73	1,14	1,98	0,20	1,36
16	28,86	-29,49	0,00	205,20	0,00	0,33	0,30	1,35
17	26,85	-27,50	1,00	181,60	1,00	0,57	0,07	1,45
18	31,06	-31,47	0,98	194,08	0,00	0,19	0,22	1,37
19	27,24	-27,84	0,00	213,02	1,00	0,23	0,37	1,49
20	27,69	-28,31	1,00	191,17	1,12	0,31	0,32	0,01
21	29,34	-30,79	0,99	212,73	1,12	1,10	0,35	0,00
22	28,75	-29,42	0,00	204,63	0,00	0,39	0,28	0,46
23	27,93	-28,28	0,00	194,84	0,00	0,18	0,17	0,50
24	27,69	-31,12	0,99	212,89	1,00	2,86	0,56	0,45
25	30,09	-31,79	1,54	180,97	0,00	0,39	1,31	0,32
26	25,71	-28,64	0,53	204,68	1,00	1,59	1,34	0,55
27	28,89	-30,40	0,97	209,79	1,12	1,38	0,14	0,00
28	28,30	-30,99	1,00	214,09	1,12	1,50	1,19	0,00
29	29,07	-30,26	0,00	233,65	1,00	0,08	1,10	1,79
30	27,52	-30,71	0,00	237,56	1,00	2,19	1,01	0,96
31	29,26	-33,26	0,71	238,96	1,00	1,11	2,88	1,86
32	27,57	-28,76	1,71	171,97	1,00	0,22	0,98	0,81
33	28,65	-37,65	0,91	270,93	1,00	6,93	2,06	1,82
34	27,15	-28,05	0,89	189,39	1,00	0,00	0,90	1,03
35	27,33	-30,32	1,00	205,63	1,00	1,93	1,06	1,88
36	25,56	-25,93	0,00	200,79	1,19	0,34	0,03	1,07
37	27,77	-31,76	0,00	246,48	1,00	3,45	0,54	1,99
38	29,08	-30,36	0,00	238,84	1,20	0,06	1,22	1,00
39	27,58	-30,85	0,00	242,86	1,19	2,20	1,07	2,04
40	27,61	-30,65	1,00	213,25	1,22	2,10	0,94	1,07
41	28,71	-30,18	0,00	237,33	1,20	1,02	0,45	1,86
42	27,84	-30,68	1,00	213,07	1,20	1,87	0,97	1,03
43	30,84	-36,27	0,88	264,80	1,22	2,95	2,48	1,94
44	25,78	-27,46	0,37	204,36	1,24	0,20	1,48	0,87
45	28,83	-30,14	0,00	237,00	1,20	0,20	1,11	1,91
46	28,60	-37,96	0,88	278,36	1,18	6,69	2,67	1,10
47	26,20	-31,66	0,88	225,80	1,24	4,03	1,43	1,90
48	28,21	-30,16	0,00	241,13	1,38	0,86	1,08	1,13
49	27,77	-32,47	0,00	256,52	1,18	3,24	1,46	2,04
50	30,30	-34,55	0,92	244,07	1,00	0,95	3,29	1,02
51	30,80	-35,78	0,77	267,28	1,38	3,23	1,74	1,98
52	26,48	-28,27	0,00	216,64	1,00	0,60	1,19	1,06
53	29,30	-35,13	0,93	248,77	1,00	3,27	2,55	1,82
54	26,15	-27,92	1,21	179,22	1,00	0,91	0,86	1,04
55	24,98	-26,92	0,98	177,15	1,00	1,00	0,93	1,88

continua

Tabela A 2 continuação

composto	Score	DG	S(hbond)	S(lipo)	H(rot)	DE(clash)	DE(int)	intcor
56	24,08	-24,99	0,00	188,66	1,00	0,52	0,40	0,90
57	26,81	-32,20	0,00	250,24	1,00	5,02	0,36	1,86
58	28,26	-30,19	0,00	233,10	1,00	0,83	1,10	1,02
59	27,15	-31,37	0,99	215,04	1,00	3,07	1,15	1,91
60	28,38	-29,89	0,00	234,89	1,20	0,34	1,17	1,06
61	28,02	-31,34	0,00	242,91	1,00	2,18	1,14	1,82
62	27,24	-29,38	0,00	226,12	1,00	0,91	1,23	0,96
63	26,25	-28,30	0,00	221,27	1,20	0,83	1,22	1,91
64	26,69	-29,02	0,00	223,11	1,00	1,14	1,19	1,02
65	29,70	-32,19	0,00	250,20	1,00	1,03	1,47	1,76
66	28,52	-31,42	0,00	243,58	1,00	1,75	1,14	0,97
67	26,45	-30,62	0,00	236,76	1,00	2,71	1,45	1,80
68	26,43	-30,21	0,90	207,62	1,00	2,67	1,11	0,86
69	26,30	-33,30	0,00	259,62	1,00	5,78	1,22	1,84
70	28,86	-30,80	0,00	238,33	1,00	0,90	1,04	1,04
71	27,94	-33,69	0,97	235,38	1,00	4,28	1,46	1,87
72	31,60	-36,64	0,84	269,11	1,22	2,83	2,21	1,04
73	27,78	-33,05	0,00	257,55	1,00	4,03	1,24	1,75
74	30,49	-36,84	0,87	269,90	1,22	3,89	2,46	1,79
75	28,85	-30,38	0,00	234,72	1,00	0,18	1,36	0,89
76	27,14	-28,45	0,00	218,20	1,00	0,25	1,06	0,96
77	20,04	-22,68	1,92	118,84	1,22	0,51	2,13	1,99
78	20,00	-21,72	2,32	99,12	1,21	0,47	1,24	1,97
79	21,39	-22,04	1,92	112,63	1,18	0,37	0,27	1,96
80	20,34	-22,14	1,12	137,11	1,22	0,16	1,64	1,97
81	23,06	-25,61	2,29	135,72	1,33	0,31	2,24	2,00
82	19,41	-19,46	1,76	95,92	1,22	0,03	0,01	1,94
83	21,32	-21,45	1,93	107,28	1,18	0,11	0,02	1,98
84	20,11	-22,54	1,88	122,77	1,40	0,04	2,39	1,97
85	20,89	-23,37	1,70	134,68	1,38	0,17	2,31	1,96
86	26,44	-32,00	3,36	163,76	1,51	1,50	4,06	1,96
87	23,31	-25,73	2,79	126,38	1,51	0,23	2,18	1,98
88	22,10	-24,11	2,32	119,11	1,20	0,69	1,31	2,24
89	22,80	-24,37	1,96	134,21	1,30	0,04	1,54	2,10
90	21,87	-24,23	1,96	143,45	1,78	0,03	2,33	2,04
91	20,78	-23,19	2,47	119,19	1,75	0,81	1,61	2,06
92	19,85	-19,96	1,76	95,50	1,00	0,10	0,02	1,92
93	23,03	-24,07	1,32	146,92	1,18	0,92	0,11	1,94
94	21,69	-25,84	1,46	158,01	1,17	1,95	2,20	1,96
95	21,19	-21,56	1,81	113,52	1,26	0,14	0,24	1,94
96	21,71	-22,47	1,73	124,91	1,33	0,16	0,60	1,98
97	24,29	-24,98	0,93	177,77	1,72	0,35	0,34	2,07

continua

**Tabela A 2 continuação**

<b>composto</b>	<b>Score</b>	<b>DG</b>	<b>S(hbond)</b>	<b>S(lipo)</b>	<b>H(rot)</b>	<b>DE(clash)</b>	<b>DE(int)</b>	<b>intcor</b>
98	26,04	-26,42	1,81	162,64	1,61	0,10	0,28	2,02
99	23,45	-26,02	1,94	149,89	1,36	0,32	2,25	1,95
100	21,25	-21,95	1,65	121,39	1,27	0,16	0,55	1,69
101	31,46	-33,09	3,51	175,57	1,82	0,33	1,29	1,63
102	25,96	-27,42	1,79	162,13	1,18	0,41	1,06	0,45
103	26,75	-28,69	2,57	168,26	1,98	0,14	1,80	1,94
104	25,46	-26,47	2,51	148,05	1,84	0,30	0,71	1,87
105	24,79	-26,13	2,34	149,05	1,79	0,11	1,23	1,80
106	25,88	-27,73	2,89	137,76	1,37	0,48	1,36	1,90
107	26,36	-28,38	2,52	154,20	1,39	0,19	1,83	1,82
108	26,08	-28,88	3,19	139,21	1,39	0,74	2,06	1,60
109	23,25	-24,04	2,55	130,16	2,03	0,32	0,47	1,72
110	29,10	-30,51	3,38	143,19	1,18	0,17	1,24	1,61
111	26,41	-29,70	2,65	157,46	1,19	1,32	1,97	1,61
112	27,98	-29,94	3,47	145,11	1,60	0,23	1,73	1,81
113	28,86	-30,03	2,59	169,28	1,52	0,16	1,02	1,88
114	27,09	-29,08	2,75	161,83	1,77	1,03	0,95	1,89
115	27,16	-28,15	2,59	154,86	1,60	0,17	0,83	1,78
116	29,32	-32,32	3,68	167,19	1,95	1,00	2,00	1,92
117	28,26	-30,15	3,36	157,80	1,95	0,61	1,28	1,84
118	28,41	-31,03	2,76	189,84	2,30	0,54	2,09	1,90
119	27,45	-28,03	0,99	215,84	2,35	0,27	0,31	2,23
120	27,19	-28,69	3,56	131,54	1,58	0,21	1,29	1,52
121	30,17	-31,70	3,72	152,50	1,59	0,67	0,85	1,92
122	29,99	-33,67	4,02	160,20	1,55	2,22	1,46	1,83
123	33,08	-36,11	3,87	187,05	1,64	1,56	1,47	3,18
124	28,93	-29,98	2,55	169,70	1,51	0,18	0,87	1,67
125	27,47	-28,14	1,59	173,56	1,15	0,45	0,21	0,41

**Tabela A 3:** Descritores obtidos através dos cálculos de docagem com a função ASP.

<b>composto</b>	<b>Score</b>	<b>ASP</b>	<b>S(Map)</b>	<b>DE(clash)</b>	<b>DE(int)</b>	<b>intcor</b>
1	38,66	39,95	-199,73	0,55	0,74	1,41
2	37,71	38,98	-194,92	0,63	0,65	1,42
3	38,73	39,95	-199,73	0,20	1,02	1,37
4	37,29	38,91	-194,56	0,16	1,46	1,40
5	37,77	39,29	-196,45	0,04	1,48	1,42
6	41,75	44,32	-221,61	1,65	0,93	0,11
7	38,57	39,74	-198,68	0,74	0,42	0,07
8	37,31	37,91	-189,55	0,24	0,36	0,57
9	35,26	39,24	-196,18	2,42	1,55	0,55
10	36,87	38,78	-193,89	0,06	1,85	0,48
11	35,53	38,58	-192,90	1,49	1,56	0,39

continua

**Tabela A 3 continuação**

<b>composto</b>	<b>Score</b>	<b>ASP</b>	<b>S(Map)</b>	<b>DE(clash)</b>	<b>DE(int)</b>	<b>intcor</b>
12	35,53	36,52	-182,60	0,60	0,39	0,58
13	39,53	41,35	-206,73	1,19	0,62	0,04
14	38,03	39,25	-196,27	0,17	1,06	0,02
15	38,66	39,95	-199,73	0,55	0,74	1,41
16	31,23	32,41	-162,03	0,89	0,28	1,35
17	30,23	32,19	-160,96	1,58	0,39	1,46
18	35,01	39,67	-198,36	4,28	0,38	1,37
19	34,17	37,48	-187,41	2,96	0,35	1,49
20	31,04	31,86	-159,31	0,02	0,81	0,00
21	32,70	33,45	-167,26	0,02	0,73	0,00
22	27,11	29,24	-146,22	0,92	1,22	0,46
23	32,18	34,15	-170,73	0,60	1,36	0,50
24	28,43	30,43	-152,16	0,83	1,17	0,45
25	33,80	36,95	-184,75	1,83	1,32	0,32
26	33,86	36,01	-180,06	1,82	0,33	0,55
27	32,16	32,74	-163,69	0,05	0,53	0,00
28	32,87	33,66	-168,31	0,02	0,78	0,00
29	33,62	34,81	-174,07	0,01	1,19	1,82
30	28,56	33,52	-167,61	3,62	1,34	0,95
31	33,53	35,08	-175,42	0,36	1,20	1,91
32	29,26	32,64	-163,18	0,71	2,67	0,82
33	29,13	32,67	-163,37	0,70	2,85	1,82
34	32,80	34,56	-172,80	0,78	0,99	1,04
35	28,61	32,40	-162,01	1,11	2,68	1,85
36	34,02	38,84	-194,21	0,98	3,85	1,08
37	31,88	33,21	-166,04	0,00	1,33	1,99
38	33,37	35,04	-175,22	0,44	1,24	1,04
39	33,56	39,09	-195,45	2,92	2,61	2,04
40	35,05	36,27	-181,34	0,02	1,20	1,07
41	34,48	35,75	-178,77	0,15	1,13	1,88
42	30,74	32,30	-161,51	0,16	1,40	1,06
43	32,81	35,25	-176,23	0,11	2,33	1,93
44	32,66	36,25	-181,23	1,12	2,46	0,94
45	30,84	33,99	-169,94	0,07	3,08	1,95
46	28,68	29,92	-149,59	0,09	1,15	1,11
47	34,24	36,33	-181,65	0,39	1,71	1,94
48	30,25	30,59	-152,97	0,04	0,31	1,16
49	34,40	39,05	-195,24	1,61	3,04	2,05
50	34,30	35,53	-177,63	0,03	1,20	0,99
51	35,72	38,97	-194,84	2,09	1,16	2,04
52	33,48	35,88	-179,39	1,31	1,09	1,06
53	33,74	36,16	-180,82	0,86	1,56	1,86

**continua**

**Tabela A 3 continuação**

<b>composto</b>	<b>Score</b>	<b>ASP</b>	<b>S(Map)</b>	<b>DE(clash)</b>	<b>DE(int)</b>	<b>intcor</b>
54	31,52	37,78	-188,91	1,91	4,36	1,03
55	34,24	37,41	-187,05	1,67	1,50	1,88
56	34,03	37,82	-189,12	0,65	3,14	0,92
57	31,73	33,61	-168,03	1,00	0,88	1,86
58	30,04	33,36	-166,82	0,10	3,23	1,02
59	34,26	37,09	-185,46	0,20	2,63	1,96
60	34,35	37,29	-186,47	0,41	2,53	1,07
61	33,57	35,01	-175,03	0,06	1,38	1,84
62	38,20	43,41	-217,07	1,24	3,98	0,98
63	35,65	39,58	-197,89	1,15	2,77	1,88
64	35,52	37,31	-186,57	0,27	1,52	1,01
65	37,04	39,67	-198,33	1,07	1,56	1,80
66	33,94	35,08	-175,40	0,08	1,06	0,98
67	33,59	36,67	-183,35	0,36	2,72	1,83
68	35,78	38,31	-191,53	0,08	2,45	0,86
69	37,51	44,81	-224,05	4,11	3,19	1,80
70	34,94	37,37	-186,84	0,52	1,91	1,01
71	37,25	40,27	-201,33	0,70	2,32	1,95
72	40,36	42,30	-211,52	0,95	0,99	1,08
73	38,52	43,63	-218,13	2,15	2,96	1,77
74	37,84	40,99	-204,95	2,03	1,13	1,94
75	33,78	34,81	-174,06	0,10	0,93	0,91
76	34,21	35,51	-177,54	0,16	1,14	0,96
77	31,40	34,68	-173,42	1,02	2,27	2,04
78	30,54	33,97	-169,85	0,10	3,33	2,00
79	30,68	33,50	-167,48	0,73	2,09	2,04
80	31,43	35,35	-176,77	1,71	2,22	2,01
81	29,51	33,31	-166,53	0,12	3,68	2,01
82	30,11	31,60	-157,99	0,00	1,48	1,99
83	30,58	31,99	-159,97	0,03	1,38	1,97
84	30,03	32,54	-162,68	0,26	2,26	2,02
85	32,25	33,85	-169,25	0,49	1,11	1,98
86	30,94	34,41	-172,07	1,26	2,22	2,08
87	29,69	32,58	-162,89	0,30	2,58	1,99
88	34,94	41,69	-208,43	4,54	2,21	2,31
89	30,38	31,97	-159,86	0,11	1,48	2,07
90	31,93	35,03	-175,13	0,91	2,19	2,15
91	31,05	32,85	-164,27	0,38	1,42	2,15
92	29,62	31,06	-155,28	0,23	1,20	1,93
93	27,89	31,04	-155,22	1,20	1,95	1,95
94	29,82	31,52	-157,61	0,05	1,66	2,01
95	31,65	34,12	-170,60	0,32	2,15	2,12

**continua**

**Tabela A 3 continuação**

<b>composto</b>	<b>Score</b>	<b>ASP</b>	<b>S(Map)</b>	<b>DE(clash)</b>	<b>DE(int)</b>	<b>intcor</b>
96	29,50	31,99	-159,97	0,02	2,48	2,04
97	31,38	36,68	-183,40	1,42	3,87	2,15
98	30,53	33,77	-168,83	1,02	2,22	2,05
99	32,50	34,47	-172,33	0,19	1,78	2,07
100	34,53	35,14	-175,69	0,12	0,49	1,73
101	34,74	37,49	-187,47	0,42	2,33	1,86
102	32,91	35,26	-176,31	1,12	1,23	0,46
103	32,59	33,46	-167,32	0,64	0,24	1,85
104	35,59	36,66	-183,30	0,17	0,90	1,89
105	35,62	37,22	-186,09	0,32	1,27	1,88
106	34,77	36,55	-182,74	0,01	1,77	1,80
107	35,32	35,90	-179,48	0,06	0,52	1,93
108	35,05	36,88	-184,41	0,09	1,74	1,64
109	35,95	37,49	-187,47	0,87	0,68	1,80
110	33,93	34,60	-173,00	0,02	0,65	1,57
111	36,18	36,78	-183,88	0,01	0,59	1,63
112	37,33	38,34	-191,72	0,16	0,86	1,82
113	37,26	37,96	-189,82	0,17	0,53	1,92
114	38,99	42,60	-212,99	1,77	1,84	1,88
115	38,27	40,57	-202,85	0,33	1,97	1,86
116	38,17	40,05	-200,25	0,14	1,74	1,95
117	38,48	42,14	-210,72	1,87	1,80	1,99
118	36,48	38,31	-191,57	0,10	1,73	1,99
119	42,10	45,19	-225,93	1,75	1,33	2,19
120	36,00	38,13	-190,63	0,56	1,57	1,60
121	37,74	39,33	-196,64	0,06	1,53	1,92
122	38,46	40,90	-204,52	1,46	0,99	1,93
123	42,29	44,09	-220,46	1,06	0,74	3,18
124	38,18	39,38	-196,92	0,50	0,70	1,73
125	30,59	34,37	-171,83	2,36	1,42	0,42

**Tabela A 4: Descritores obtidos através dos cálculos de docagem com a função ChemPLP**

<b>composto</b>	<b>Score</b>	<b>S(PLP)</b>	<b>S(hbond)</b>	<b>S(cho)</b>	<b>DE(clash)</b>	<b>DE(tors)</b>	<b>intcor</b>
1	77,90	-75,90	1,56	0,00	0,00	1,36	0,03
2	76,95	-74,95	1,47	0,00	0,00	1,21	0,03
3	77,17	-74,27	1,44	0,00	0,00	0,72	0,04
4	73,87	-71,88	1,45	0,00	0,00	1,20	0,04
5	74,22	-72,40	1,00	0,00	0,00	0,63	0,09
6	90,16	-91,46	0,00	0,00	0,00	0,73	0,17
7	88,47	-86,46	1,00	0,00	0,01	0,54	0,07
8	82,18	-80,39	1,00	0,00	0,03	1,00	0,83
9	79,50	-77,17	1,30	0,00	0,01	1,26	0,84

continua



**Tabela A 4**      **continuação**

<b>composto</b>	<b>Score</b>	<b>S(PLP)</b>	<b>S(hbond)</b>	<b>S(cho)</b>	<b>DE(clash)</b>	<b>DE(tors)</b>	<b>intcor</b>
10	78,02	-76,47	1,00	0,00	0,01	1,06	0,68
11	72,21	-70,31	1,00	0,00	0,01	0,90	0,54
12	76,28	-74,79	1,24	0,00	0,00	1,58	0,91
13	90,98	-92,75	0,00	0,00	0,00	1,01	0,13
14	86,72	-85,71	1,00	0,00	0,00	1,07	0,13
15	77,90	-75,90	1,56	0,00	0,00	1,36	0,03
16	75,20	-75,82	0,00	0,00	0,00	0,31	0,00
17	65,34	-62,50	1,00	0,00	0,00	0,18	0,20
18	75,36	-67,94	2,77	0,00	0,00	0,44	0,00
19	74,80	-75,20	0,00	0,00	0,00	0,27	0,12
20	73,94	-71,84	1,00	0,00	0,03	0,44	0,01
21	73,19	-73,84	0,00	0,00	0,00	0,33	0,01
22	69,34	-69,48	0,00	0,00	0,00	0,42	0,72
23	72,73	-73,04	0,00	0,00	0,05	0,53	0,80
24	65,66	-63,42	1,00	0,00	0,00	0,71	0,67
25	72,83	-72,22	3,01	0,00	8,72	0,06	0,43
26	75,38	-76,24	0,00	0,00	0,08	0,85	0,91
27	76,64	-75,22	0,75	0,00	0,36	0,24	0,01
28	78,38	-75,85	1,00	0,00	0,00	0,24	0,01
29	71,26	-69,48	1,00	0,00	0,03	0,88	0,57
30	62,80	-64,20	0,00	0,00	0,00	1,40	1,25
31	71,42	-69,51	1,00	0,00	0,00	0,83	0,58
32	62,50	-63,54	0,00	0,00	0,00	1,00	0,96
33	62,21	-64,14	0,00	0,00	1,31	0,61	0,60
34	68,13	-70,89	0,00	0,00	0,00	2,09	1,43
35	61,07	-59,98	1,00	0,00	0,00	1,26	0,59
36	67,94	-70,73	0,00	0,00	0,00	2,14	1,48
37	71,32	-70,05	1,09	0,00	1,52	0,56	0,63
38	74,51	-77,17	0,00	0,00	0,00	2,03	1,31
39	66,72	-69,65	0,00	0,00	0,00	1,86	0,75
40	73,40	-75,93	0,00	0,00	0,00	2,04	1,46
41	73,20	-76,15	0,00	0,00	0,00	1,82	0,69
42	69,20	-71,78	0,00	0,00	0,00	1,99	1,39
43	70,77	-73,89	0,00	0,00	0,00	1,91	0,70
44	63,21	-63,90	0,15	0,00	0,00	1,15	1,15
45	70,02	-70,00	0,92	0,00	1,52	0,97	0,70
46	71,42	-73,74	0,00	0,00	0,00	1,93	1,53
47	63,95	-66,53	0,00	0,00	0,00	1,66	0,73
48	70,60	-73,59	0,00	0,00	0,00	2,32	1,65
49	61,97	-62,87	0,00	0,00	0,00	0,82	0,72
50	73,82	-76,95	0,00	0,00	0,00	2,23	1,32
51	73,14	-75,93	0,00	0,00	0,00	1,84	0,77

**continua**

Tabela A 4 continuação

composto	Score	S(PLP)	S(hbond)	S(cho)	DE(clash)	DE(tors)	intcor
52	64,29	-64,30	0,00	0,00	0,00	0,74	1,46
53	72,08	-75,45	0,00	0,00	0,00	2,01	0,64
54	70,85	-73,62	0,00	0,00	0,02	2,07	1,39
55	64,78	-65,25	0,00	0,00	0,01	0,56	0,66
56	67,22	-69,97	0,00	0,00	0,00	1,93	1,11
57	69,38	-68,51	0,83	0,00	0,40	0,94	0,66
58	72,18	-74,72	0,00	0,00	0,00	1,97	1,40
59	63,71	-66,25	0,00	0,00	0,00	1,62	0,71
60	75,86	-78,26	0,00	0,00	0,02	1,96	1,53
61	70,82	-73,78	0,00	0,00	0,00	1,80	0,63
62	71,42	-74,78	0,00	0,00	0,02	2,31	1,28
63	70,12	-72,97	0,00	0,00	0,00	1,74	0,62
64	76,05	-78,76	0,00	0,00	0,00	2,14	1,44
65	76,02	-79,36	0,00	0,00	0,00	1,95	0,56
66	71,46	-73,98	0,00	0,00	0,00	1,93	1,35
67	77,87	-79,74	0,41	0,00	1,23	1,20	0,55
68	68,05	-68,42	0,00	0,00	0,04	0,70	1,04
69	72,78	-75,97	0,00	0,00	0,00	1,89	0,56
70	73,98	-76,75	0,00	0,00	0,00	2,04	1,31
71	68,63	-66,47	1,09	0,00	0,02	0,84	0,58
72	77,94	-80,75	0,00	0,00	0,00	2,14	1,46
73	76,02	-79,27	0,00	0,00	1,02	1,38	0,53
74	73,21	-76,77	0,00	0,00	0,00	2,04	0,53
75	74,28	-77,14	0,00	0,00	0,00	1,99	1,12
76	69,37	-72,17	0,00	0,00	0,00	2,03	1,26
77	56,49	-54,27	0,00	1,00	0,01	1,28	1,80
78	55,27	-48,31	2,85	0,00	0,14	1,66	1,78
79	58,39	-53,35	1,00	1,00	0,80	1,03	1,74
80	56,90	-55,28	1,00	0,00	0,00	1,58	1,74
81	60,28	-52,93	2,65	0,00	0,19	1,05	1,60
82	55,28	-47,42	2,00	1,06	0,00	1,51	1,60
83	60,95	-56,14	1,11	1,00	0,00	1,59	1,68
84	57,13	-52,51	1,00	0,82	0,83	0,84	1,68
85	59,89	-54,48	1,35	1,00	0,00	1,66	1,61
86	58,45	-50,48	1,98	1,00	0,00	1,35	1,61
87	56,78	-51,89	1,00	1,00	0,00	1,36	1,61
88	59,94	-52,75	3,00	0,00	0,01	1,72	1,62
89	59,90	-59,75	1,27	0,00	0,86	2,28	1,65
90	62,45	-63,26	0,00	0,00	0,00	1,29	1,78
91	62,88	-60,50	1,33	0,00	0,00	1,74	1,68
92	56,40	-48,15	2,14	1,00	0,00	1,37	1,55
93	57,51	-53,08	1,84	0,00	0,13	1,26	1,56

continua

**Tabela A 4** continuação

composto	Score	S(PLP)	S(hbond)	S(cho)	DE(clash)	DE(tors)	intcor
94	57,40	-48,36	2,16	0,99	0,00	1,01	1,61
95	55,41	-46,76	2,00	0,89	0,01	0,90	1,72
96	60,49	-56,31	2,00	0,00	0,00	1,73	1,63
97	75,02	-71,81	2,00	0,06	0,00	2,41	1,74
98	73,32	-68,46	2,00	0,00	0,00	1,47	1,71
99	66,65	-58,60	3,00	0,06	0,00	1,43	1,72
100	62,44	-55,77	1,53	1,06	0,07	1,07	1,11
101	67,21	-74,01	1,00	0,00	11,33	0,41	2,04
102	68,80	-64,30	0,33	1,60	0,00	0,64	0,01
103	69,38	-66,82	0,98	0,00	0,00	0,91	1,46
104	68,73	-61,12	2,79	0,00	0,03	1,07	1,41
105	66,80	-63,03	1,67	0,80	4,16	0,48	1,46
106	66,62	-62,36	1,76	0,00	0,68	0,81	1,29
107	66,57	-63,17	1,69	0,68	3,87	0,63	1,40
108	69,04	-62,83	2,83	1,14	4,78	1,00	0,79
109	65,28	-59,56	1,52	1,04	2,23	0,55	1,27
110	61,76	-57,78	1,71	0,00	0,62	1,24	1,94
111	62,28	-63,24	1,67	0,77	9,14	0,71	2,01
112	68,56	-67,35	2,11	0,98	7,64	0,87	1,31
113	69,63	-69,02	2,03	1,00	7,68	1,06	1,34
114	63,66	-58,85	1,01	1,00	0,00	1,50	1,78
115	68,46	-65,77	2,13	1,00	6,14	0,94	1,27
116	73,39	-73,44	0,99	0,00	1,17	1,71	1,48
117	75,96	-65,83	4,52	0,00	0,00	2,51	1,41
118	78,25	-77,65	1,27	0,00	0,50	2,49	2,27
119	79,79	-73,66	1,69	1,00	0,00	2,05	2,17
120	74,55	-67,40	3,13	0,92	2,95	1,29	0,51
121	71,67	-64,29	1,99	0,98	0,00	1,52	1,49
122	73,48	-67,90	3,43	0,00	2,84	1,64	1,42
123	80,09	-72,77	3,52	0,00	2,16	1,39	1,38
124	64,61	-62,73	0,98	0,00	0,00	1,55	2,05
125	71,95	-67,00	1,86	0,00	0,00	0,33	0,01

**Tabela A 5:** Descritores obtidos através dos cálculos de docagem com a função GoldScore

composto	Fitness	S(hb_ext)	S(vdw_ext)	S(int)	intc
1	74,78	0,00	55,07	-0,94	-16,06
2	72,51	1,72	54,66	-4,36	-16,24
3	69,61	2,06	50,30	-1,61	-18,96
4	72,14	3,65	52,65	-3,90	-17,55
5	76,11	2,30	55,69	-2,76	-16,72
6	70,97	0,00	53,35	-2,38	-10,85
7	72,62	0,34	54,06	-2,05	-11,02

continua

**Tabela A 5 continuação**

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
8	77,73	0,06	58,05	-2,15	-27,00
9	77,17	0,00	58,84	-3,73	-25,26
10	72,49	1,35	55,97	-5,82	-30,39
11	71,22	2,01	54,58	-5,85	-27,50
12	73,16	2,84	55,09	-5,43	-27,47
13	75,48	1,12	56,96	-3,97	-18,59
14	74,23	0,33	57,44	-5,08	-18,72
15	74,78	0,00	55,07	-0,94	-16,06
16	73,10	0,00	53,48	-0,44	-29,97
17	61,89	0,61	47,61	-4,18	-13,05
18	73,27	0,00	54,02	-1,00	-31,29
19	72,12	0,00	52,60	-0,20	-25,27
20	67,87	0,00	50,07	-0,98	-24,57
21	68,83	0,00	51,25	-1,65	-20,53
22	68,96	0,00	51,28	-1,55	-39,92
23	74,74	0,00	55,98	-2,23	-39,35
24	66,60	0,21	50,61	-3,19	-29,16
25	72,98	0,00	55,09	-2,76	-39,17
26	76,82	0,00	56,13	-0,35	-30,48
27	69,39	0,00	51,63	-1,60	-36,21
28	71,38	0,46	53,75	-2,99	-20,82
29	67,18	0,00	50,02	-1,60	-5,26
30	64,32	0,00	50,46	-5,06	-6,76
31	66,13	0,10	53,93	-8,13	-5,33
32	64,74	0,00	51,45	-6,01	-7,00
33	66,53	0,11	54,57	-8,61	-5,34
34	65,22	0,00	50,76	-4,57	-6,77
35	63,82	5,68	46,25	-5,45	-6,27
36	63,55	0,00	49,24	-4,15	-7,61
37	61,34	0,00	46,95	-3,22	-5,22
38	66,21	0,00	52,16	-5,50	-7,47
39	61,99	0,00	46,65	-2,16	-6,16
40	66,00	0,00	51,57	-4,90	-7,71
41	65,63	0,00	48,79	-1,45	-5,96
42	68,22	0,00	53,32	-5,09	-7,55
43	65,29	0,00	48,40	-1,27	-6,01
44	68,21	0,00	53,93	-5,94	-7,73
45	65,03	0,00	48,21	-1,25	-6,01
46	69,37	0,00	54,44	-5,48	-7,65
47	68,21	5,66	49,83	-5,96	-6,92
48	66,03	0,00	51,34	-4,56	-8,22
49	64,64	0,00	48,21	-1,65	-5,91

**continua**

**Tabela A 5 continuação**

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
50	67,49	0,00	51,05	-2,71	-6,73
51	67,38	0,66	51,63	-4,27	-6,68
52	65,56	0,00	50,01	-3,21	-7,00
53	65,99	0,00	49,19	-1,64	-5,30
54	64,56	0,00	50,66	-5,10	-6,82
55	65,47	0,11	54,39	-9,43	-5,38
56	65,29	0,00	50,29	-3,85	-7,08
57	62,00	0,00	46,56	-2,01	-5,35
58	65,19	0,00	50,35	-4,04	-6,88
59	63,61	5,26	46,53	-5,62	-6,34
60	67,07	0,11	53,55	-6,68	-7,68
61	61,53	0,00	47,02	-3,13	-5,29
62	66,04	0,00	48,28	-0,34	-6,33
63	60,80	0,12	45,19	-1,45	-6,04
64	66,07	0,00	48,22	-0,23	-6,65
65	65,26	0,00	48,06	-0,82	-4,87
66	66,73	0,00	48,82	-0,39	-6,54
67	64,87	0,00	47,84	-0,91	-4,96
68	66,89	0,00	50,96	-3,19	-6,74
69	63,70	0,00	47,55	-1,68	-4,95
70	66,96	0,00	51,47	-3,82	-6,55
71	66,34	5,63	48,25	-5,63	-5,89
72	68,57	0,00	53,20	-4,57	-7,53
73	63,34	0,00	48,87	-3,86	-4,91
74	65,22	0,00	48,65	-1,67	-5,69
75	68,71	0,00	51,65	-2,30	-6,70
76	67,26	0,00	51,04	-2,93	-6,88
77	65,78	0,00	50,58	-3,77	-0,94
78	63,57	7,25	45,68	-6,49	-0,76
79	70,55	0,10	52,95	-2,36	-0,86
80	69,06	0,77	50,94	-1,76	-0,83
81	65,75	2,55	50,04	-5,61	-2,06
82	64,71	2,77	47,79	-3,77	-5,34
83	65,37	2,89	47,70	-3,11	-5,14
84	70,51	4,56	51,17	-4,40	-5,61
85	68,97	8,60	48,35	-6,10	-6,15
86	72,46	4,54	51,12	-2,38	-5,86
87	68,20	7,98	47,25	-4,75	-5,95
88	71,61	0,33	53,46	-2,22	-5,84
89	66,49	0,00	50,21	-2,55	-4,33
90	70,96	8,54	48,95	-4,89	-2,11
91	69,14	8,76	49,01	-7,01	-5,31

**continua**

**Tabela A 5** continuação

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
92	62,07	2,58	43,99	-1,00	-3,46
93	66,15	3,09	48,01	-2,96	-4,36
94	65,52	1,86	48,68	-3,27	-5,10
95	65,51	3,03	47,90	-3,39	-3,68
96	67,80	4,24	47,47	-1,70	-4,05
97	76,61	4,74	55,57	-4,54	-6,76
98	76,81	3,42	55,40	-2,79	-6,61
99	65,27	6,00	44,20	-1,51	-3,36
100	73,17	2,94	51,91	-1,14	-1,81
101	90,61	7,96	63,46	-4,61	-7,55
102	72,97	1,56	52,92	-1,36	-0,71
103	79,08	0,87	59,53	-3,64	-2,17
104	84,52	8,44	57,38	-2,82	-2,59
105	76,50	0,62	57,18	-2,74	-1,29
106	75,60	0,43	57,07	-3,30	-0,55
107	78,16	2,85	56,23	-2,00	-0,41
108	76,80	2,81	54,91	-1,52	-2,04
109	83,70	8,57	56,44	-2,48	-2,66
110	79,90	0,71	58,95	-1,87	-3,38
111	77,32	3,44	55,11	-1,90	-3,29
112	79,02	2,85	57,30	-2,62	0,00
113	83,51	3,34	60,04	-2,39	-0,43
114	82,31	2,87	58,73	-1,31	-0,96
115	79,58	3,44	56,87	-2,06	0,02
116	86,81	4,56	62,18	-3,24	-0,91
117	83,90	3,85	59,45	-1,69	-0,21
118	87,85	4,70	64,76	-5,90	-2,57
119	90,76	3,37	67,10	-4,88	-4,04
120	81,11	3,64	56,79	-0,61	-3,17
121	84,35	3,10	59,95	-1,19	-4,04
122	87,21	2,39	62,12	-0,59	-5,78
123	89,95	2,03	65,71	-2,44	-8,70
124	90,75	5,61	64,66	-3,77	-4,25
125	74,42	4,09	53,56	-3,32	-0,67

**Tabela A 6:** Descritores obtidos através dos cálculos de redocagem com a função GoldScore e ChemPLP

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
1	55,60	2,79	39,89	-2,04	-15,68
2	61,46	0,00	44,98	-0,39	-15,13
3	66,27	0,00	49,14	-1,30	-16,82
4	60,51	0,00	44,15	-0,20	-16,28

continua

**Tabela A 6 continuação**

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
5	58,57	0,00	43,08	-0,66	-15,75
6	63,37	8,40	43,28	-4,55	-10,31
7	65,60	2,73	46,42	-0,96	-10,94
8	61,89	0,77	45,01	-0,77	-26,75
9	63,98	0,86	46,68	-1,06	-24,85
10	62,61	0,39	46,72	-2,02	-28,53
11	62,64	0,15	47,73	-3,14	-27,28
12	65,75	0,30	48,36	-1,05	-26,61
13	65,67	1,15	48,34	-1,95	-17,70
14	76,85	0,00	57,78	-2,60	-18,48
15	68,25	0,00	50,04	-0,55	-29,88
16	73,28	0,00	53,69	-0,54	-29,90
17	64,94	1,27	47,09	-1,07	-14,09
18	73,09	0,00	53,49	-0,46	-31,31
19	72,00	0,00	52,55	-0,26	-25,44
20	67,19	0,00	49,78	-1,26	-24,56
21	69,32	0,00	51,45	-1,42	-20,56
22	68,86	0,00	51,34	-1,74	-39,92
23	74,60	0,00	55,72	-2,01	-39,35
24	65,58	0,00	52,25	-6,27	-29,98
25	72,91	0,00	54,70	-2,31	-39,17
26	76,60	0,00	55,95	-0,33	-30,65
27	69,60	0,00	52,06	-1,98	-36,23
28	72,83	0,00	53,38	-0,57	-20,82
29	66,59	0,00	49,62	-1,63	-5,44
30	67,21	0,00	52,37	-4,79	-6,97
31	66,26	0,08	53,64	-7,57	-5,53
32	66,32	0,00	51,85	-4,97	-7,20
33	65,40	0,15	54,91	-10,26	-5,52
34	67,75	0,00	52,58	-4,55	-7,49
35	63,95	5,74	46,23	-5,36	-6,45
36	65,63	0,11	51,19	-4,86	-8,04
37	62,78	0,00	47,76	-2,90	-5,91
38	67,98	0,28	50,16	-1,27	-7,87
39	62,78	0,00	47,03	-1,89	-6,50
40	67,64	0,30	50,28	-1,80	-8,04
41	66,41	0,00	48,88	-0,80	-6,38
42	71,05	0,00	55,08	-4,68	-7,92
43	65,46	0,15	48,66	-1,60	-6,42
44	71,23	0,00	55,41	-4,96	-8,36
45	66,83	0,00	50,56	-2,70	-6,41
46	71,84	0,00	55,66	-4,69	-8,52

**continua**

**Tabela A 6 continuação**

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
47	70,82	5,89	51,62	-6,04	-7,43
48	68,67	0,10	53,49	-4,97	-8,70
49	67,08	0,00	50,68	-2,61	-6,90
50	67,88	0,00	51,57	-3,03	-6,95
51	67,99	0,00	53,26	-5,23	-7,20
52	65,53	0,00	49,52	-2,56	-7,20
53	66,16	0,10	54,14	-8,37	-5,51
54	66,31	0,00	51,62	-4,66	-7,03
55	66,00	0,09	53,85	-8,13	-5,58
56	66,60	0,00	51,38	-4,04	-7,28
57	66,77	0,10	55,16	-9,17	-5,56
58	67,49	0,00	52,30	-4,42	-7,60
59	63,50	5,53	46,62	-6,14	-6,51
60	67,34	0,00	54,11	-7,05	-8,17
61	63,39	0,00	48,24	-2,94	-5,96
62	67,74	0,07	50,05	-1,14	-6,67
63	64,21	0,17	47,89	-1,81	-6,39
64	67,09	0,27	49,77	-1,62	-6,77
65	65,53	0,13	49,12	-2,14	-5,07
66	68,42	0,00	53,11	-4,61	-6,63
67	65,66	0,11	49,12	-1,99	-5,15
68	68,46	0,00	52,65	-3,93	-7,02
69	66,15	0,12	49,38	-1,87	-5,21
70	69,00	0,00	53,19	-4,14	-7,17
71	66,08	5,83	48,16	-5,96	-6,10
72	69,74	0,00	54,00	-4,50	-8,15
73	65,83	0,00	50,34	-3,40	-5,58
74	67,07	0,09	50,15	-1,97	-6,07
75	68,81	0,00	52,09	-2,82	-6,85
76	66,88	0,00	50,23	-2,18	-7,08
77	63,40	2,32	47,62	-4,39	0,02
78	62,24	1,84	47,83	-5,36	0,01
79	62,67	2,02	47,96	-5,30	-0,02
80	62,51	2,29	47,64	-5,30	-0,01
81	63,66	2,35	49,36	-6,57	-0,93
82	62,97	2,91	46,63	-4,05	-4,19
83	63,03	2,67	47,34	-4,72	-4,23
84	62,98	2,05	48,11	-5,22	-4,13
85	63,50	2,58	47,68	-4,63	-4,84
86	67,62	3,24	50,28	-4,75	-4,90
87	65,01	2,06	48,58	-3,86	-4,70
88	65,98	2,56	49,17	-4,19	-3,65

**continua**



**Tabela A 6 continuação**

<b>composto</b>	<b>Fitness</b>	<b>S(hb_ext)</b>	<b>S(vdw_ext)</b>	<b>S(int)</b>	<b>intc</b>
89	61,09	2,83	46,86	-6,17	-2,90
90	67,77	2,99	51,30	-5,76	-2,39
91	66,94	3,05	50,21	-5,16	-4,30
92	62,05	2,49	43,85	-0,73	-3,28
93	63,85	2,96	46,86	-3,55	-3,56
94	63,16	2,65	47,33	-4,57	-4,19
95	64,58	2,65	47,73	-3,70	-3,34
96	65,32	3,76	46,27	-2,05	-3,43
97	77,94	3,64	56,55	-3,45	-5,87
98	76,37	3,15	55,55	-3,16	-5,68
99	64,70	0,00	48,86	-2,48	-3,21
100	71,88	1,99	51,39	-0,76	-1,45
101	81,32	0,36	61,92	-4,18	-6,66
102	72,29	4,08	51,54	-2,67	-0,69
103	76,89	7,61	54,73	-5,97	-2,09
104	81,67	7,83	55,71	-2,77	-1,82
105	76,00	0,41	57,20	-3,06	-0,94
106	74,03	3,10	54,74	-4,34	-0,32
107	76,98	2,08	55,83	-1,87	-0,17
108	75,15	3,30	54,55	-3,16	-1,42
109	81,05	7,89	55,10	-2,60	-1,63
110	77,84	0,36	57,94	-2,18	-3,12
111	76,47	0,85	56,34	-1,85	-2,88
112	77,79	2,88	56,18	-2,33	0,34
113	82,06	2,76	59,37	-2,33	0,26
114	79,32	2,21	57,98	-2,61	-0,18
115	77,74	2,97	56,52	-2,95	0,55
116	78,64	6,24	61,88	-12,69	-0,85
117	82,73	2,61	59,69	-1,95	-0,59
118	80,53	0,27	66,73	-11,49	-3,36
119	75,20	3,19	52,79	-0,58	-2,70
120	80,58	3,77	57,26	-1,92	-1,72
121	82,25	2,47	59,77	-2,41	-1,74
122	85,23	2,39	62,46	-3,04	-1,47
123	88,02	1,89	63,56	-1,26	-28,11
124	90,86	5,50	65,63	-4,88	-3,90
125	74,52	3,83	53,14	-2,38	-0,65