

UNIVERSIDADE DE SÃO PAULO  
INSTITUTO DE QUÍMICA DE SÃO CARLOS

CÉLIO FERNANDO LIPINSKI

Planejamento de inibidores da polimerização de microtúbulos: predição da atividade biológica de um grupo de chalconas e proposição de novas estruturas

São Carlos  
2021

CÉLIO FERNANDO LIPINSKI

Planejamento de inibidores da polimerização de microtúbulos: predição da atividade biológica de um grupo de chalconas e proposição de novas estruturas

Tese apresentada ao Instituto de Química de São Carlos da Universidade de São Paulo como parte dos requisitos para a obtenção do título de Doutor em Ciências.

Área de concentração: Físico-Química

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

**Exemplar revisado**

O exemplar original encontra-se em acervo reservado na Biblioteca do IQSC-USP

São Carlos

2021

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico para fins de estudo e pesquisa, desde que citada a fonte.

Assinatura:

Célio F. Lipinski

Data:

15/02/22.

*Ficha Catalográfica elaborada pela Seção de Referência e Atendimento ao Usuário do SBI/IQSC*

Lipinski, Célio Fernando

Planejamento de inibidores da polimerização de microtúbulos: predição da atividade biológica de um grupo de chalconas e proposição de novas estruturas / Célio Fernando Lipinski. — São Carlos, 2021.

142 f.

Tese (Doutorado em Físico-Química) — Instituto de Química de São Carlos / Universidade de São Paulo, 2021.

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

1. Câncer. 2. Chalconas. 3. Predição de atividade biológica. 4. Proposição de estruturas. I. Título.



*À Bernadete e ao Leônidas*

## **AGRADECIMENTOS**

Aos meus pais, Luiz e Terezinha Lipinski, que foram capazes de formar uma família da qual sinto muita satisfação por fazer parte.

À minha amada esposa Bernadete Okopna Lipinski, à qual reafirmo aqui, o meu mais sincero desejo de lhe fazer companhia até o último dia da minha vida.

Ao meu filhinho Leônidas Lipinski, que já nos seus dois primeiros anos de existência nesse mundo, me ensinou, na prática, o significado da palavra transcendência.

Ao professor Albérico Borges Ferreira da Silva que, muito acima das atribuições de um orientador, tem a capacidade de preparar seus alunos para desafios que vão além da vida acadêmica.

A todos os meus familiares e amigos, pelo apoio e pelas alegrias que me deram.

A todos os colegas do grupo de Química Quântica, por toda a ajuda e pela ótima companhia ao longo desses anos. Em especial, à Aline Alves Oliveira, pela substancial contribuição a este trabalho, e também pelo seu exemplo de força.

Aos funcionários do DQFM, do serviço de informática, da biblioteca e do serviço de pós-graduação do IQSC, pelo excelente atendimento.

Ao Instituto de Química de São Carlos e à Universidade de São Paulo, pelo ótimo acolhimento.

À CAPES, pelo suporte financeiro.

## RESUMO

Dentre os diversos agentes anticancer estudados atualmente, as chalconas são alternativas bastante promissoras. A atuação dessas se dá a partir da inibição da polimerização dos microtúbulos, os quais são fundamentais nos processos de mitose e manutenção estrutural da célula. Visando aprimorar a funcionalidade dessa classe de compostos, foi realizado um estudo com um grupo de 87 chalconas, onde foram utilizados métodos computacionais de predição de atividade biológica e de análise de interações entre ligante e receptor. Para a predição da atividade biológica, foram gerados modelos de Relações Quantitativas Estrutura-Atividade (QSAR) utilizando métodos de Mínimos Quadrados Parciais (PLS) e Redes Neurais Artificiais (ANN), além da geração de modelos de Holograma QSAR (HQSAR) e QSAR tridimensional (QSAR 3D). No estudo das interações, foram realizadas análises de docagem molecular, buscando encontrar características nos complexos que influenciem na atividade biológica de tais compostos. Nas análises de QSAR, os métodos de PLS e ANN apresentaram resultados bastante semelhantes e igualmente satisfatórios. Nas análises de HQSAR, a escolha adequada da melhor combinação de distinção dos fragmentos possibilitou a geração de um modelo com capacidade preditiva. E nas análises de QSAR 3D, o método de Análise Comparativa dos Campos Moleculares (CoMFA) apresentou resultados ligeiramente superiores aos do método de Análise Comparativa dos Índices de Similaridade (CoMSIA). As análises de docagem molecular sugerem que a interação desses compostos se dê numa região ligeiramente diferente que a colchicina (ligante de referência) nesse sítio de ligação pois, apesar do sítio de ligação desses compostos estar localizado na  $\beta$ -tubulina, os ligantes interagem numa região de interface, entre a  $\alpha$  e a  $\beta$ -tubulina. Após todas as análises, foi possível propor um grupo de 36 novas estruturas, e prever a atividade biológica das mesmas a partir dos modelos gerados, a fim de que possam ser uma alternativa para a modulação da dinâmica dos microtúbulos e para o tratamento do câncer.

Palavras-chave: Câncer. Chalconas. Predição de atividade biológica. Proposição de estruturas.

## ABSTRACT

Chalcones are naturally occurring compounds reported as promising anticancer agents. They act as a microtubule polymerization inhibitor, which are fundamental in the mitosis process and structural cell. In order to improve the functionality of this class of compounds, a group of 87 chalcones were studied by computational methods capable of predicting biological activities and ligand-receptor interactions. For the prediction of biological activity, Quantitative Structure-Activity Relationships (QSAR) models were generated using Partial Least Squares (PLS) and Artificial Neural Networks (ANN) methods, hologram QSAR (HQSAR) and three-dimensional QSAR (3D QSAR) models were also built. In the study of interactions, molecular docking analyses were performed seeking to find features in the complexes that influence the biological activity of such compounds. In the QSAR analyses, the PLS and ANN methods showed very similar and equally satisfactory results. In the HQSAR analyses, the right choice of the best distinguishing combination of fragments enabled the generation of a model with predictive capacity. In the QSAR 3D analyses, the Comparative Molecular Field Analysis (CoMFA) method presented results slightly superior to those of the Comparative Molecular Similarity Indices Analysis (CoMSIA) method. Molecular docking analyses showed that the interaction of these compounds occurs in a slightly different region than colchicine (reference ligand) at that binding site, because, although the binding site of these compounds is located in  $\beta$ -tubulin, the ligands interact in an interface region between  $\alpha$  and  $\beta$ -tubulin. After all, it was possible to propose a group of 36 new structures, and to predict their biological activity from the built models. The new proposed compounds can possibly be used as an alternative for modulating microtubule dynamics and for cancer treatment.

Keywords: Cancer. Chalcones. Biological activity prediction. Structures proposition.

## LISTA DE FIGURAS

Figura 1 – Esquema simplificado do ciclo celular. O processo se inicia na fase G <sub>1</sub> . .....	18
Figura 2 – Estrutura geral dos microtúbulos e representação dos processos de polimerização e despolimerização dos mesmos, a partir das estruturas diméricas de $\alpha$ -tubulina e $\beta$ -tubulina. A $\gamma$ -tubulina indicada é responsável por nuclear os microtúbulos a partir de sua extremidade (-), dando início ao processo de polimerização. ....	19
Figura 3 – Estrutura da combretastatina A4.....	24
Figura 4 – Estrutura base das chalconas. ....	24
Figura 5 – Método da roleta. A probabilidade de escolha de cada um dos indivíduos está diretamente relacionada à sua aptidão. ....	33
Figura 6 – Representação simplificada do algoritmo genético. ....	35
Figura 7 – Representação do funcionamento das Redes Neurais Artificiais do tipo MLP.....	40
Figura 8 – Processo de geração de um holograma molecular. ....	41
Figura 9 – Estruturas base dos compostos analisados. A maioria das moléculas é oriunda das estruturas base 2, 3 e 4. ....	47
Figura 10 – Representação estrutural da tubulina. Código PDB 1SA0. ....	52
Figura 11 – Ligante cristalográfico CN2 do sítio de ligação da colchicina, fornecido no PDB.....	53
Figura 12 – Detecção de outliers. ....	56
Figura 13 – Teste de <i>leave-N-out</i> . ....	57
Figura 14 – Teste de randomização da atividade biológica. ....	58
Figura 15 – Contribuição de cada descritor para o vetor de regressão do modelo PLS. .....	59
Figura 16 – Predição da atividade biológica do modelo de PLS. ....	60

Figura 17 – Predição da atividade biológica do modelo <i>Consensus</i> ANN – conjunto de teste. ....	63
Figura 18 – Verificação da distribuição dos valores de atividade biológica na separação dos conjuntos de treinamento e de teste para o modelo de HQSAR. ....	66
Figura 19 – Mapa de contribuições individuais da molécula 2a para o modelo HQSAR. A cor amarela indica contribuição favorável à atividade biológica. ....	68
Figura 20 – Mapa de contribuições individuais da molécula 3a para o modelo HQSAR. As cores verde e amarelo indicam contribuições favoráveis à atividade biológica. ....	68
Figura 21 – Mapa de contribuições individuais da molécula 4a para o modelo HQSAR. As cores verde e amarelo indicam contribuições favoráveis à atividade biológica. ....	69
Figura 22 – Subestrutura comum a todos os compostos, selecionada para o alinhamento dos mesmos nas análises de QSAR 3D. ....	70
Figura 23 – Verificação da distribuição dos valores de atividade biológica na separação dos conjuntos de treinamento e de teste para os modelos de QSAR 3D. ....	71
Figura 24 – Mapas de contorno estéricos para a molécula 3a, gerados a partir da plataforma CoMFA para o modelo de QSAR 3D. ....	72
Figura 25 – Mapas de contorno eletrostáticos para a molécula 3a, gerados a partir da plataforma CoMFA para o modelo de QSAR 3D. ....	73
Figura 26 – Mapas de contorno estéricos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D. ....	73
Figura 27 – Mapas de contorno eletrostáticos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D. ....	74
Figura 28 – Mapas de contorno hidrofóbicos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D. ....	74

Figura 29 – Mapas de contorno de doadores e aceptores de ligação de hidrogênio para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D. ....	75
Figura 30 – Estrutura remanescente utilizada para os cálculos de docagem molecular, sendo mantidas a $\alpha$ -tubulina (direita) e a $\beta$ -tubulina (esquerda). ....	78
Figura 31 – Sítio de ligação da colchicina, no qual encontra-se o ligante cristalográfico CN2. ....	78
Figura 32 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 2 $\beta$ no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1. ....	81
Figura 33 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 2d no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1. ....	82
Figura 34 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 3a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1. ....	83
Figura 35 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 4a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1. ....	84
Figura 36 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e a molécula de GTP no seu respectivo sítio de ligação. Os resíduos de aminoácidos foram mantidos rígidos para fins de comparação. ....	86
Figura 37 – Plotagem dos valores de pontuação <i>versus</i> pontuação dos complexos formados entre os ligantes selecionados e a $\alpha\beta$ -tubulina quando os resíduos flexíveis foram os mesmos da flexibilidade 1. Nota-se a baixa dispersão dos valores no gráfico. ....	88
Figura 38 – Plotagem dos valores de pontuação <i>versus</i> pontuação dos complexos formados entre os ligantes selecionados e a $\alpha\beta$ -tubulina quando os resíduos flexíveis foram os mesmos da flexibilidade 2 nos cálculos de docagem. Nota-se uma maior dispersão dos pontos se comparados aos da Figura tal, e também, dois agrupamentos	

principais: um para valores de pontuação abaixo de 40, e outro acima de 40.  
.....89

Figura 39 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 2 $\beta$  no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....90

Figura 40 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 2d no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....91

Figura 41 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....92

Figura 42 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 4a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....93

Figura 43 – Representação da LYS254, com os átomos de hidrogênio indicados nas posições Z2 e Z3, bem como os átomos de carbono das posições G e D. ....94

Figura 44 – Representação da LYS352, com os átomos de hidrogênio indicados nas posições Z2 e Z3, assim como os átomos de carbono das posições G e D. ....95

Figura 45 – Representação da LEU248, com indicação dos átomos de carbono das posições G e D (D1 e D2). ....96

Figura 46 – Representação dos resíduos SER178 e THR179, os quais foram separados para uma melhor visualização. Conforme a linha tracejada, a ligação entre esses resíduos ocorre a partir do átomo de carbono de SER178 e o átomo de nitrogênio da THR179. ....97

Figura 47 – Representação da GLN11, com indicação do átomo de hidrogênio da posição E21, responsável pelas ligações de hidrogênio no complexo formado.  
.....97

Figura 48 – Modo de ligação obtido para as chalconas no sítio de ligação da colchicina (2a, 2 $\beta$ , 2c, 2d, 2g, 2nn, 2rr, 2vv, 2xx, 2zz, 3a, 3b, 3c, 4a, 4b, 4c, 4d, 5d, 6a, 8a e 8d), sendo omitidos os átomos de hidrogênio. Estão apresentados os principais resíduos envolvidos na formação desses complexos, assim como a molécula de GTP. ....	98
Figura 49 – Representação da fosbretabulina dissódica. Fármaco que já está disponível comercialmente. ....	102
Figura 50 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 3a13 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....	105
Figura 51 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 4a1 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....	106
Figura 52 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 4a9 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....	107
Figura 53 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 3a1 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....	112
Figura 54 – Diagrama bidimensional do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 3a2 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2. ....	113
Figura 55 – Representação dos resíduos GLY144, TYR224 e ASN101, todos pertencentes à $\alpha$ -tubulina. Para fins de uma melhor compreensão do comportamento, seus respectivos posicionamentos em um dos complexos foram mantidos. Além, de estarem na formação do complexo dos ligantes 3a1 e 3a2, também se fizeram presentes em diversas análises anteriores, porém com uma menor frequência que os resíduos já citados anteriormente.....	114
Figura 56 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 2a e 2d, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	134

Figura 57 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 2g e 2nn, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	135
Figura 58 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 2rr e 2vv, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	136
Figura 59 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 2xx e 2zz, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	137
Figura 60 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 3b e 3c, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	138
Figura 61 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 4b e 4c, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	139
Figura 62 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 4d e 5d, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	140
Figura 63 – Diagramas 2D dos complexos formados entre a $\alpha\beta$ -tubulina e os ligantes 6a e 8a, selecionados para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	141
Figura 64 – Diagramas 2D do complexo formado entre a $\alpha\beta$ -tubulina e o ligante 8a, selecionado para as análises de docagem molecular, após ser fornecida a flexibilidade 2 a alguns dos resíduos. ....	142

## LISTA DE TABELAS

Tabela 1 – Parâmetros de distinção dos fragmentos utilizados no método HQSAR. ...	42
Tabela 2 – Parâmetros de evolução aplicados no cálculo de docagem molecular. ....	53
Tabela 3 – Parâmetros estatísticos obtidos para o modelo de PLS. ....	58
Tabela 4 – Parâmetros estatísticos dos 18 melhores modelos de MLP-ANN, e valores de $r^2_{\text{teste}}$ e EQM do ANN-CM, gerados a partir do fator de normalização (FN). ....	61
Tabela 5 – Valores estatísticos de predição dos modelos PLS e ANN-CM. ....	61
Tabela 6 – Predição da atividade biológica dos modelos PLS e ANN-CM. ....	62
Tabela 7 – Diferença entre os valores de $r^2_{\text{teste}}$ e resíduo médio das metodologias aplicadas na pesquisa anterior e na ANN-CM. ....	63
Tabela 8 – Resultados das combinações de distinção de fragmentos quando utilizado o tamanho padrão de fragmento (4-7). ....	64
Tabela 9 – Resultados das análises da influência do tamanho dos fragmentos a partir da melhor combinação de fragmentos – A.B.H.DA. ....	65
Tabela 10 – Predição dos valores da atividade biológica ( $pIC_{50}$ ) para os conjuntos de treinamento e de teste a partir das análises de HQSAR. ....	67
Tabela 11 – Resultados dos parâmetros estatísticos dos melhores modelos de QSAR 3D (CoMFA e CoMSIA). ....	71
Tabela 12 – Predição dos valores da atividade biológica ( $pIC_{50}$ ) para os conjuntos de treinamento e de teste a partir das análises de QSAR 3D (CoMFA e CoMSIA). ....	76
Tabela 13 – Comparação entre as funções de pontuação testadas para a $\beta$ -tubulina e para a $\alpha\beta$ -tubulina, baseada nos valores de pontuação e RMSD para as funções ChemPLP, ChemScore e GoldScore. ....	77
Tabela 14 – Valores de pontuação e RMSD para as funções GoldScore, ChemPLP e ChemScore. ....	79
Tabela 15 – Valores de pontuação para a melhor conformação de cada um dos compostos selecionados para as análises de docagem molecular (flexibilidade 1). ...	80

Tabela 16 – Comparação dos valores de pontuação entre o primeiro e segundo testes de flexibilidade fornecida a alguns dos resíduos de aminoácidos da $\alpha$ e $\beta$ -tubulina. Pode ser percebido o contraste entre as escolhas de flexibilidade. ....	87
Tabela 17 – Estruturas propostas para o sítio de ligação da colchicina. ....	102
Tabela 18 – Valores preditos de atividade biológica ( $pIC_{50}$ ) pelos métodos de QSAR (PLS), HQSAR e QSAR 3D (CoMFA e CoMSIA) para as 36 moléculas propostas. ...	109
Tabela 19 – Moléculas propostas que apresentaram os melhores valores de atividade biológica ( $pIC_{50}$ ) nas predições de todos os modelos gerados. ....	110
Tabela 20 – Compostos e respectivos valores de atividade biológica ( $pIC_{50}$ ) das 87 chalconas analisadas neste trabalho. ....	127
Tabela 21 – Valores dos descritores para cada um dos compostos.....	131

## LISTA DE ABREVIATURAS E SIGLAS

ADMET	Absorção, Distribuição, Metabolismo, Excreção e Toxicidade
AG	Algoritmo genético
ANN	<i>Artificial neural network</i>
CA4	Combretastatina A4
CA4DP	Combretastatina A4 disódio fosfato
CM	<i>Consensus model</i>
CoMFA	<i>Comparative molecular field analysis</i>
CoMSIA	<i>Comparative molecular similarity indices analysis</i>
DFT	<i>Density functional theory</i>
EQM	Erro quadrático médio
FN	Fator de normalização
GDP	<i>Guanosine diphosphate</i>
GOLD	<i>Genetic Optimization for Ligand Docking</i>
GTP	<i>Guanosine triphosphate</i>
HQSAR	<i>Hologram QSAR</i>
LLA	Leucemia linfóide aguda
LLC	Leucemia linfóide crônica
LMA	Leucemia mielóide aguda
LMC	Leucemia mielóide crônica
MIFs	<i>Molecular interaction fields</i>
MLP	<i>Multilayer Perceptron</i>
PDB	<i>Protein Data Bank</i>
Ph <sup>+</sup>	Cromossomo Philadelphia anormal
PLS	<i>Partial least squares</i>
PRESS	<i>Prediction Residues Error Squares Sum</i>
QSAR	<i>Quantitative Structure-Activity Relationships</i>
RMN	Ressonância magnética nuclear
RMSD	<i>Root-Mean-Square deviation</i>
SEC	<i>Standard Error of Calibration</i>
SEV	<i>Standard Error of Validation</i>

## LISTA DE ABREVIATURAS E SIGLAS (II)

ALA	Alanina
ARG	Arginina
ASP	Aspartato
ASN	Asparagina
CYS	Cisteína
PHE	Fenilalanina
GLY	Glicina
GLU	Glutamato
GLN	Glutamina
HIS	Histidina
ILE	Isoleucina
LEU	Leucina
LYS	Lisina
MET	Metionina
PRO	Prolina
SER	Serina
TYR	Tirosina
THR	Treonina
TRP	Triptofano
VAL	Valina

## SUMÁRIO

<b>1. INTRODUÇÃO</b> .....	17
<b>1.1 Câncer</b> .....	17
<b>1.2 Leucemia</b> .....	20
1.2.1 Leucemia mielóide crônica.....	21
<b>1.3 Células K562</b> .....	22
<b>1.4 Os agentes antitubulina</b> .....	23
<b>1.5 Planejamento de fármacos</b> .....	25
<b>1.6 Docagem molecular</b> .....	26
1.6.1 O programa GOLD.....	29
<b>1.7 Algoritmo genético</b> .....	31
<b>1.8 Relação Quantitativa Estrutura-Atividade (QSAR)</b> .....	35
1.8.1 Mínimos Quadrados Parciais (PLS) .....	37
1.8.2 Redes Neurais Artificiais.....	39
<b>1.9 HQSAR</b> .....	40
<b>1.10 QSAR 3D</b> .....	42
<b>2. OBJETIVOS</b> .....	45
<b>3. METODOLOGIA</b> .....	46
<b>3.1 Modelos de QSAR</b> .....	46
3.1.1 Metodologia de PLS.....	48
3.1.2 Metodologia de ANN.....	48
<b>3.2 Modelos de HQSAR</b> .....	50
<b>3.3 Modelos de QSAR 3D</b> .....	50
<b>3.4 Docagem molecular</b> .....	51

<b>4. RESULTADOS E DISSCUSSÃO</b> .....	55
<b>4.1 Modelos de QSAR</b> .....	55
4.1.1 Análises de PLS.....	55
4.1.2 Análises de ANN.....	60
<b>4.2 Análises de HQSAR</b> .....	64
<b>4.3 Análises de QSAR-3D</b> .....	69
4.3.1 Modelos de CoMFA.....	72
4.3.2 Modelos de CoMSIA.....	73
<b>4.4 Análises de docagem molecular</b> .....	77
4.4.1 LYZ254.....	94
4.4.2 LYS352.....	94
4.4.3 LEU248.....	96
4.4.4 THR179 e SER178.....	96
4.4.5 GLN11.....	97
4.4.6 GTP.....	98
<b>4.5 Proposição de estruturas</b> .....	99
<b>5. CONCLUSÃO E PERSPECTIVAS</b> .....	115
<b>REFERÊNCIAS</b> .....	116
<b>APÊNDICES</b> .....	127

## 1. INTRODUÇÃO

### 1.1 Câncer

O câncer é um conjunto de diferentes doenças que se caracterizam por um descontrole no processo de mitose no organismo, a partir do qual, a célula passa a se reproduzir indiscriminadamente, não respondendo mais aos mecanismos de controle fisiológicos, passando a se alimentar dos nutrientes que eram destinados, até então, às células saudáveis.<sup>1,2</sup> Esse descontrole se deve a alterações que ocorrem no DNA celular. Sua origem se dá por condições multifatoriais, que podem agir em conjunto ou em sequência para iniciar ou evoluir um quadro de câncer. Uma série de procedimentos e combinações de métodos são necessários para o tratamento dos diversos tipos de câncer, dependendo do seu tamanho, estágio e localização, além do tipo de célula.<sup>1,2</sup>

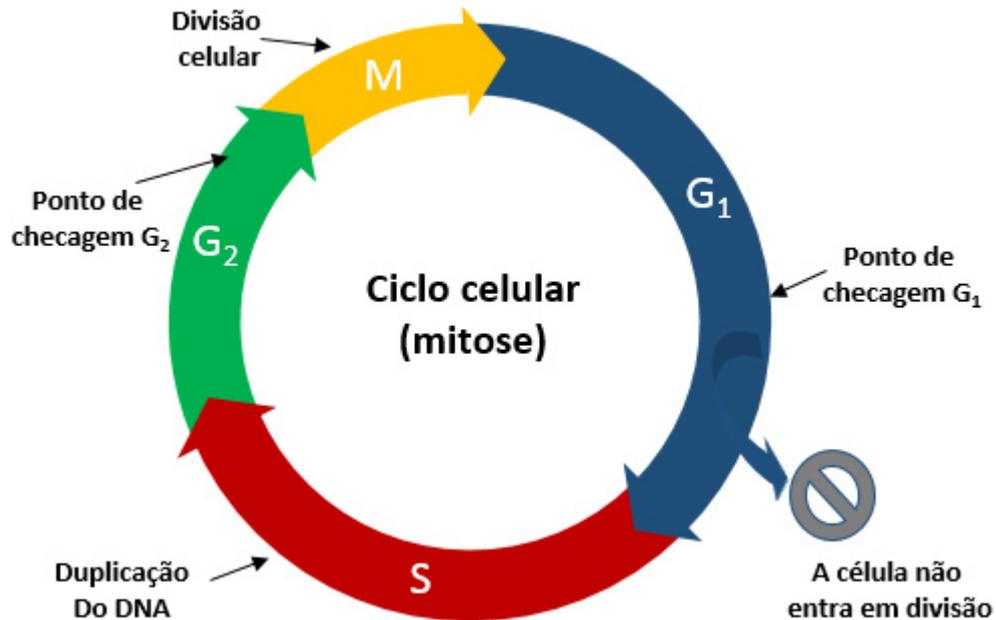
Segundo informações da Organização Mundial da Saúde, o câncer causou a morte de 9,6 milhões de pessoas em todo o mundo no ano de 2018. Isso significa que, a cada 6 mortes, 1 é relacionada ao câncer. Além das mortes, seu impacto econômico também é bastante significativo, com estimativas que ultrapassam a casa dos trilhões de dólares. No Brasil, há a estimativa de que o biênio 2019-2020 se encerre com cerca de 686 mil novos casos.<sup>3,4</sup>

O processo de divisão celular, conforme a Figura 1, pode ser resumidamente descrito iniciando no estágio de gap1 (G1), no qual são emitidos sinais para que o processo continue ou seja interrompido. A partir dessas manifestações por determinadas proteínas, a célula demonstra sua aptidão para passar à etapa seguinte, a fase de síntese (S), onde ocorre a duplicação do DNA. Como garantia de não haverem danos de duplicação, a célula passa pela fase de gap2 (G2), onde é feita a análise de todo o material genético. Se a cópia não for exata, ocorre a liberação de proteínas supressoras que bloqueiam a passagem ao estágio final, a mitose, fase em que duas células idênticas são formadas.<sup>5,6,7</sup>

Quando as etapas de verificação não ocorrem, células defeituosas podem surgir do processo de divisão celular. Caso esse processo se repita de maneira sistemática, essas células passam a reproduzir desordenadamente, o que acabará por resultar numa estrutura massiva composta de células aberrantes - o tumor. Por

isso, se faz necessária a utilização de substâncias que interrompam tal processo de mitose.<sup>5,6,7</sup>

Figura 1 – Esquema simplificado do ciclo celular. O processo se inicia na fase G<sub>1</sub>.

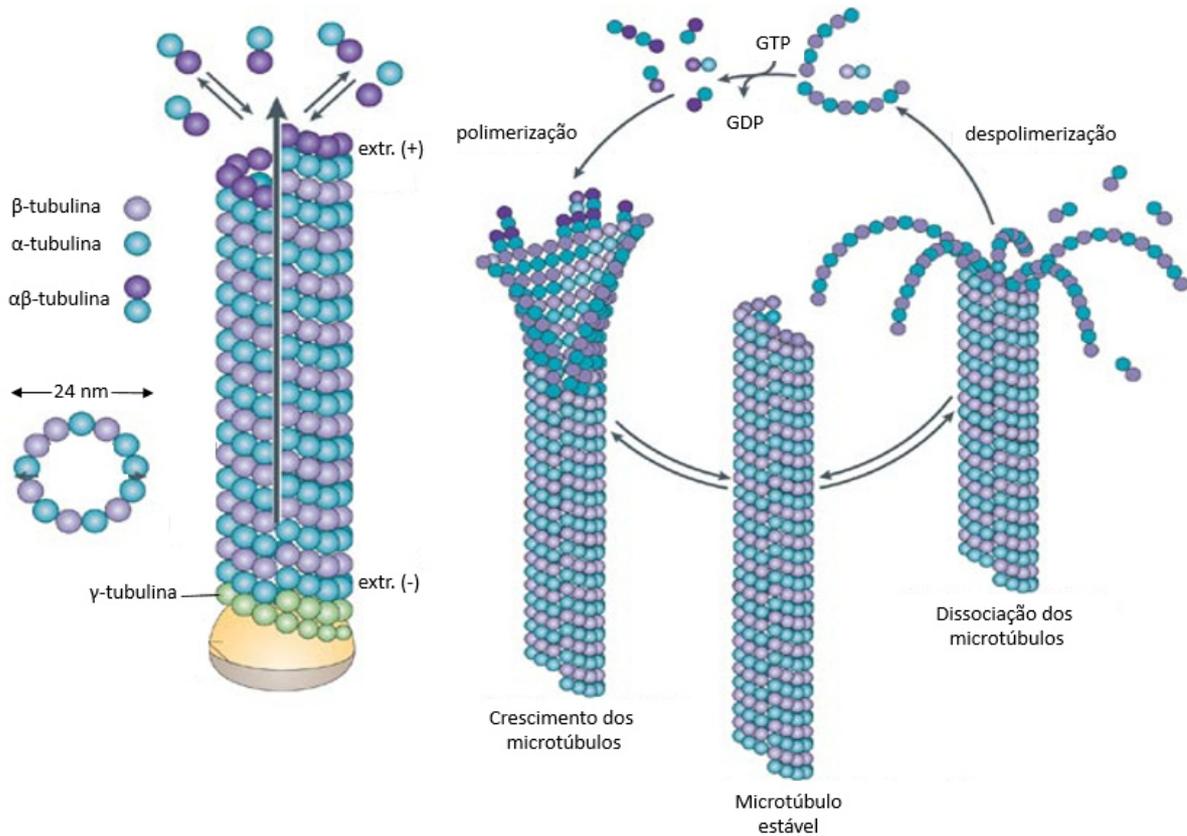


Fonte: Elaborada pelo autor.

Para o tratamento, uma grande quantidade de fármacos encontra-se disponível, atuando a partir de diversos mecanismos de ação. Entretanto, alguns fatores causam grande limitação quanto ao uso desses medicamentos, como elevada toxicidade e resistência. Por isso, a busca por novas alternativas aos medicamentos já existentes ainda se mostra altamente válida.

Algumas substâncias interagem com os componentes dos microtúbulos, os quais, no processo de divisão celular, são responsáveis por diversos tipos de interações celulares, como absorção de alimentos, descarte de organelas e a separação dos cromossomos destinados à célula-filha durante o fuso mitótico. Uma proteína denominada tubulina é formada por dois heterodímeros:  $\alpha$ -tubulina e  $\beta$ -tubulina, e a polimerização dessa estrutura dimérica irá formar os microtúbulos, como ilustra a Figura 2.<sup>8,9</sup>

Figura 2 – Estrutura geral dos microtúbulos e representação dos processos de polimerização e despolimerização dos mesmos, a partir das estruturas diméricas de  $\alpha$ -tubulina e  $\beta$ -tubulina. A  $\gamma$ -tubulina indicada é responsável por nuclear os microtúbulos a partir de sua extremidade (-), dando início ao processo de polimerização.



Fonte: Adaptada de CONDE *et al.*<sup>10</sup>

O trifosfato de guanosina (GTP) interage tanto com a  $\alpha$  quanto com a  $\beta$ -tubulina. Sua concentração em solução está diretamente relacionada com a formação dos microtúbulos. Quando em baixas concentrações, o GTP tende a se ligar ao seu sítio na  $\alpha$ -tubulina, onde estabiliza a proteína e, por se posicionar numa região de interface entre a  $\alpha$  e a  $\beta$ -tubulina, acaba por favorecer a associação de ambas, formando as estruturas diméricas da tubulina. Quando em altas concentrações, o GTP além de se posicionar no seu sítio da  $\alpha$ -tubulina, acaba por ocupar também o sítio na  $\beta$ -tubulina em uma das extremidades. Esse sítio não confere a mesma estabilidade ao GTP, que pode se hidrolisar para GDP (difosfato de guanosina). Quando a estrutura dimérica está associada ao GDP, ela apresentará uma curvatura, desfavorecendo a formação do protofilamento, ao passo que quando associada ao GTP será reta, formato esse que favorece a formação do protofilamento. Apesar da menor estabilidade quando associada ao GDP, ainda fica garantida a continuidade de formação dos filamentos,

desde que haja quantidade suficiente de GTP em solução, para garantir a estabilidade do microtúbulo.<sup>11</sup>

No processo de formação dos filamentos podem ser consideradas duas extremidades: (+) e (-). A extremidade (+) indica o direcionamento de formação de um microtúbulo, e a extremidade (-) indica a extremidade a partir da qual ocorre a dissociação do mesmo. Quando a dinâmica dos microtúbulos é suprimida, a formação do eixo mitótico fica comprometida, desencadeando mecanismos que resultam em apoptose. E compostos que interferem nessa dinâmica são agentes anticâncer de grande relevância no tratamento.<sup>12</sup>

Dentre os mais variados tipos de câncer, a leucemia se diferencia pelo fato de estar diretamente relacionada com o sangue e com a medula óssea. No Brasil, estima-se que ocorram cerca de 11 mil novos casos em 2020.<sup>13</sup>

## **1.2 Leucemia**

A neoplasia do tecido hematopoiético oriundo da medula óssea recebe o nome de leucemia. Essa doença envolve o sangue periférico e pode infiltrar o baço, o fígado e os nódulos linfáticos. Sua classificação poderá ser como aguda ou crônica, dependendo do grau de maturidade e da linhagem das células afetadas.

A leucemia aguda é caracterizada pelo acúmulo e proliferação de células imaturas na medula óssea, sangue periférico e outros órgãos, substituindo as células sanguíneas normais. Caso não haja tratamento adequado em tempo hábil, o tempo de vida do paciente é de aproximadamente 6 meses.<sup>14</sup>

Já na leucemia crônica, a proliferação se dará por células maduras também na medula, sangue periférico e outros órgãos. Contudo, é de menor intensidade que a anterior, podendo levar de 2 a 6 anos para atingir a fase acelerada, variando conforme a linhagem da célula que se prolifera.

Além da classificação entre as leucemias aguda e crônica, também ocorre a distinção entre as linhagens celulares mielóide e linfóide. O que acaba por gerar 4 tipos de leucemia: leucemia mielóide aguda (LMA), leucemia mielóide crônica (LMC), leucemia linfóide aguda (LLA) e leucemia linfóide crônica (LLC).<sup>14</sup>

A LMC apresenta uma linhagem celular característica, a K562, a qual está diretamente ligada aos interesses desta pesquisa, pois é nessa linhagem que os compostos deste estudo foram testados.

### 1.2.1 Leucemia mielóide crônica

A leucemia mielóide crônica se origina a partir de um descontrole no processo de mitose de células tronco hematopoiéticas, havendo aumento de células mielóide no sangue periférico e na medula óssea. Pessoas de qualquer faixa etária podem ser acometidas por essa doença, ainda que a idade média esteja em torno de 53 anos. Ainda que muitos pacientes sejam assintomáticos, alguns sintomas que podem caracterizar essa doença são perda de peso, fadiga, anorexia e esplenomegalia. Para os assintomáticos, o diagnóstico se baseia diretamente na contagem anormal de células.<sup>14</sup>

A LMC evolui de uma fase crônica, tida como benigna, até a fase acelerada. A etapa posterior à fase acelerada, é a crise blástica, onde as células não atingem a maturação, acabando por se assemelharem aos mieloblastos ou linfoblastos presentes em pacientes com quadro leucemia aguda.<sup>14</sup>

A primeira fase da doença é conhecida como a fase crônica, e é caracterizada por excessiva proliferação da linhagem mielóide. Nessa primeira etapa, muitos pacientes são assintomáticos, principalmente no início. Ao longo do tempo, para alguns em meses, para outros em anos, começam a aparecer sintomas como perda de peso, fadiga, fraqueza, irritabilidade, dores de cabeça, febre e suor noturno.

O diagnóstico acidental da doença é bastante comum, onde o paciente busca atendimento médico por outros motivos ou até mesmo em exames de rotina. Assim nota-se o caráter silencioso dessa primeira fase.<sup>14,15</sup>

A fase crônica é muito bem definida clinicamente. Apresenta como características a leucocitose, a hepatomegalia e a esplenomegalia volumosa. Células granulocíticas imaturas como o mieloblastos, mielócitos, promielócitos e metamielócitos surgem nesse caso. A maioria dos pacientes apresenta casos de basofilia e eosinofilia. Contudo, ainda com uma boa definição dos fatores relacionados com a doença, ainda é passível de um diagnóstico equivocado na fase inicial, onde pode ser confundida com uma infecção desconhecida ou então outros tipos de leucemia.<sup>16</sup>

Na segunda fase dessa doença, ocorre a perda progressiva da diferenciação celular. Para essa fase, há uma grande variação dos critérios de definição. O critério mais utilizado é o da Organização Mundial da Saúde. Essa fase é bastante significativa pelo fato de sinalizar a evolução da doença para a crise blástica.<sup>17,18</sup>

A crise blástica apresenta a grande maioria dos sintomas e características da leucemia mielóide aguda ou da leucemia linfóide aguda. E, dos pacientes que chegam nessa fase, aproximadamente 75% falecem antes de 6 meses.<sup>19</sup>

Dentre todas as fases, essa é a de mais fácil identificação. Pois, segundo a Organização Mundial da Saúde, considera-se crise blástica quando o número de blastos estiver acima de 20%, independentemente de serem transformações mielóides ou linfóides. As mielóides ocorrem em aproximadamente 70% dos casos de crise blástica.

Quando a crise blástica é do tipo mielóide, pode ocorrer a predominância de transformação de apenas uma linhagem, como mieloblástica, eosinofílica, basofílica, monocítica, megacarioblástica, eritoblástica, ou então podem ocorrer múltiplas linhagens. No caso da crise blástica linfóide, pode haver uma súbita ocorrência, mesmo sem ter passado pela fase aguda.<sup>17</sup>

### **1.3 Células K562**

A linhagem celular K562 foi cultivada a partir das células de um paciente com leucemia mielóide crônica, nas quais se encontrava o cromossomo Philadelphia (Ph<sup>+</sup>). Essa linhagem se mostra como um bom modelo experimental para estudos relacionados com a chamada terapia de diferenciação do câncer, devido ao fato de apresentar alta plasticidade, resultando em alta capacidade de diferenciação quando submetida a diferentes agentes químicos. Essa terapia de diferenciação estuda, basicamente, o tratamento de células neoplásicas com agentes indutores de diferenciação celular.<sup>20</sup>

Células da linhagem K562 não apresentam grande resistência a antraciclinas ou outras drogas antileucêmicas. Por isso, são amplamente utilizadas para testar a ação de diferentes compostos. Contudo, mesmo as antraciclinas sendo drogas bastantes potentes, com aplicações em tumores sólidos e outras leucemias, por conta das características desse tipo de célula, sua aplicação na LMC Ph<sup>+</sup> está limitada apenas à fase aguda da doença, dificultando assim o tratamento na fase inicial.<sup>21</sup>

Uma alternativa de tratamento que tem como alvo essa linhagem e busca se antecipar à fase aguda da doença comprometendo a polimerização dos microtúbulos, e conseqüentemente, o funcionamento das células anormais, é realizada com os chamados agentes antitubulina.

## 1.4 Os agentes antitubulina

A indústria farmacêutica há muito tempo se dedica a estudar compostos que têm como alvo os microtúbulos. Eles podem atuar em diversos sítios na tubulina. Os três mais importantes são: o dos taxóides, o dos alcalóides da vinca e o da colchicina. Todos possuem a característica de atuarem interrompendo o ciclo de polimerização da tubulina. Quando os compostos alcançam os seus respectivos sítios de interação, muitos apresentam pouca eficácia em células que não estão se dividindo. Conseqüentemente, muitas dessas células podem sobreviver e adquirir resistência à mesma. Por isso, uma alternativa conveniente é a interrupção do suprimento de sangue dos tumores, impedindo a polimerização dos microtúbulos, resultando na morte da célula. A colchicina tem como diferencial, em relação aos taxóides e os alcalóides da vinca, essa capacidade de agir também na vasculatura que irriga o tumor.<sup>22</sup>

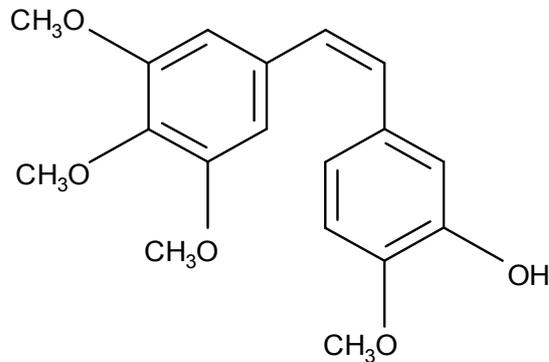
A colchicina é um produto natural e foi o primeiro agente desestabilizador de microtúbulos a ser identificado. Apesar de sua importância histórica, ela não está incluída entre os fármacos comercialmente disponíveis, em razão da sua elevada neurotoxicidade, já que essa droga possui índice terapêutico baixo, o que a levava a ser administrada em concentrações próximas dos limites aceitáveis para o organismo.<sup>23,24</sup>

A seleção de conjuntos de novos compostos que apresentem similaridade estrutural e que possam atuar no mesmo sítio de interação da colchicina é uma alternativa para os esforços intensivos direcionados à identificação de novos agentes antitubulina. As combretastatinas e seus derivados atuam nesse mesmo sítio e estão entre os mais estudados para essa finalidade. Elas constituem um grupo de agentes antimetabólicos isolados a partir da casca da árvore sul-africana *Combretum caffrum*. Também podem ser encontradas na árvore *Combretum leprosum*, popularmente conhecida como Mufumbo, cuja ocorrência se dá na caatinga e pantanal brasileiro.<sup>25-27</sup> Sua estrutura básica é constituída por dois anéis benzênicos polioxigenados, ligados por uma ponte de etileno.

Um grupo de pesquisadores, liderados por Petit, isolam, sintetizam e avaliam diversos tipos de combretastatinas desde o início da década de 80, verificando suas elevadas atividades antimetabólicas, e conseqüentemente, suas aplicações no tratamento do câncer.<sup>28-33</sup> Nos últimos anos, a combretastatina A4 (CA4) vem sendo

amplamente estudada e apontada com grande potencial anticancer.<sup>34-38</sup> Sua estrutura é apresentada na Figura 3:

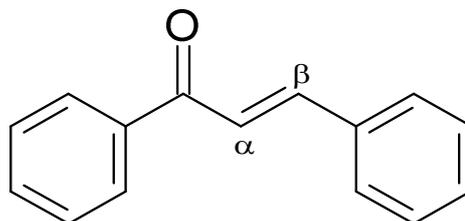
Figura 3 – Estrutura da combretastatina A4.



Fonte: Elaborada pelo autor.

O estudo de ligantes do sítio da colchicina tem sido bastante recomendado, e um grande número de pequenas moléculas sintéticas ligantes desse sítio tem sido identificado. As chalconas que serão objeto de estudo neste trabalho, estão nesse grupo. Elas inibem a polimerização dos microtúbulos a partir do mesmo sítio de interação que o alcalóide colchicina, na  $\beta$ -tubulina. São moléculas de cadeia aberta que contêm dois anéis aromáticos ligados por um fragmento enona de três carbonos, conforme Figura 4.<sup>39</sup>

Figura 4 – Estrutura base das chalconas.



Fonte: Elaborada pelo autor.

De maneira análoga às combretastatinas, as chalconas têm apresentado elevados valores de atividade biológica quando testadas suas respectivas capacidades em processos antimitóticos.<sup>39</sup> Também são utilizadas em muitas outras pesquisas que visam o desenvolvimento de novos fármacos, por apresentarem uma

extensa gama de aplicações, como antimitótica, anti-inflamatória, antiviral, bactericida e fungicida.

Atualmente, os principais medicamentos comercialmente disponíveis para o tratamento da LMC são inibidores da tirosina quinase, como por exemplo, o Imatinibe, o Dasatinibe, Nilotinibe e o Bosutinibe.<sup>40</sup>

### 1.5 Planejamento de fármacos

Desde as primeiras tentativas do ser humano de curar suas doenças, a metodologia utilizada no desenvolvimento de medicamentos foi a de tentativa e erro. E isso acabava por gerar, além dos riscos à saúde, elevadas demandas de mão de obra, tempo e custo de produção. Mas, ao longo do tempo, a racionalidade empregada no processo possibilitou estudos muitos mais direcionados e eficientes, fazendo com que o processo de tentativa e erro fosse substancialmente melhorado.

A evolução dos métodos *in vitro* a partir dos anos 50 possibilitou grandes avanços no processo de teste de um candidato a fármaco, principalmente na diminuição de riscos à saúde humana, pois apenas compostos com o efeito biológico desejado passariam a ser testados *in vivo*, reduzindo significativamente a quantidade de substâncias utilizadas num processo de tentativa e erro.<sup>41</sup> A partir dos anos 80, com o aumento da capacidade computacional e da caracterização estrutural de macromoléculas, os métodos computacionais acabaram por alavancar o planejamento racional de fármacos, pois foram capazes de aprimorá-lo, reduzindo custos e dando mais agilidade no processo de descoberta de um novo fármaco, já que muitos dos compostos menos promissores podem ser identificados e excluídos a partir de tais análises prévias.<sup>42,43</sup> Contudo, vale ressaltar que as três etapas citadas (*in vivo*, *in vitro* e *in silico*) se complementam. Nenhuma delas substitui a outra. Todas são fundamentais no planejamento racional de fármacos.

Em 2016, o custo médio para a produção de um novo medicamento era de aproximadamente 2,8 bilhões de dólares. Esses custos envolvem grande diversidade de material biológico nos testes dos compostos, como também na própria síntese e caracterização desses compostos. Vale ressaltar que o custo da produção de um novo medicamento abarca os custos dos testes realizados com os todos compostos nas fases clínica e pré-clínica.<sup>44</sup> Além dos custos de pesquisa e desenvolvimento, o tempo

empregado é outro fator de relevância, onde, em média, o tempo de pesquisa de um novo fármaco até a sua comercialização pode ultrapassar os 10 anos.

E uma alternativa para reduzir a onerosidade de tal processo é, a partir da estrutura tridimensional de macromoléculas e de diversos recursos computacionais, buscar um melhor entendimento das interações entre um fármaco e o seu respectivo alvo terapêutico, ou seja, as interações ligante-receptor.<sup>45</sup> Contudo, além do desafio de entender tais interações, há também a possibilidade de um candidato a fármaco interagir com mais de um alvo, e conseqüentemente, gerar efeitos colaterais. Assim, mesmo que alguns compostos tenham grande afinidade com o alvo buscado, muitas vezes acabam por não se tornarem fármacos, devido a fatores como permeabilidade, absorção, metabolismo, excreção e toxicidade. Tais fatores podem ser incluídos um grupo de propriedades conhecidas como ADME ou ADMET (Absorção, Distribuição, Metabolismo, Excreção e Toxicidade). Sendo identificados previamente candidatos problemáticos, ocorrerão, também nesse caso, consideráveis reduções de custos e tempo do processo global. Existem programas capazes de calcular várias dessas propriedades em um composto e compará-las com cerca de 95% dos fármacos já existentes, disponíveis em um banco de dados virtual. E grande ênfase tem sido dada a essas propriedades, também conhecidas como *drug-like*, como faz a Regra dos Cinco de Lipinski.<sup>46</sup>

## 1.6 Docagem molecular

Quando a estrutura tridimensional de uma macromolécula é conhecida, a pesquisa sobre as interações de seus sítios receptores com determinadas moléculas pode ser realizada com o auxílio de uma ferramenta computacional conhecida como docagem (do inglês, *docking*) molecular. Essa técnica objetiva prever a afinidade entre um ligante e uma macromolécula, tornando possível caracterizar o comportamento de complexos.

Busca-se analisar a associação entre duas moléculas a partir das coordenadas atômicas das mesmas. Nessa associação pode-se prever a orientação e a conformação do ligante no respectivo sítio de interação na macromolécula. Com isso, podem ser identificados resíduos de aminoácidos cuja participação é fundamental nessa interação. Dessa forma, busca-se então a identificação do modo de ligação de uma pequena molécula (ligante) em uma macromolécula (proteína, por exemplo).

Uma consequência disso é que complexos experimentais podem ser reconstruídos a partir desse método.<sup>47</sup>

Pequenas moléculas que atuam no mesmo sítio de interação de uma macromolécula tendem a apresentar efeitos biológicos semelhantes, desde que apresentem semelhante capacidade de interação com esse alvo. Por isso, a informação obtida em análises de docagem molecular pode ser utilizada para desenvolver moléculas com tal semelhança, a fim de que os efeitos biológicos almejados possam ser potencializados. Quando se conhece a estrutura do complexo proteína-ligante, pode-se modificar a estrutura química do ligante a fim de que a afinidade desse com a proteína seja melhorada. Sendo assim, o processo de desenvolvimento de novos compostos pode ser otimizado, gerando uma maior diversidade de compostos analisados. O método de docagem molecular é um método que auxilia o planejamento de fármacos baseando-se na estrutura do receptor. É um método que tem logrado êxito no auxílio à descoberta de novos fármacos, como bloqueadores de canais iônicos, inibidores enzimáticos e agonistas/antagonistas de receptores.<sup>41,48</sup>

Basicamente, um programa de docagem molecular se baseia em um algoritmo de busca e uma função de pontuação. O algoritmo de busca explora a região determinada, onde variações de orientação e conformação do ligante são testadas, as quais são representadas numericamente pela função de pontuação. Ao final da etapa de cálculo, pode ser obtido um ranking com os melhores resultados, onde além da análise numérica, é feita também a análise visual da interação. Os métodos estocásticos ou randômicos são algoritmos de busca que aplicam alterações conformacionais aleatoriamente para um ligante, sendo que as melhores são selecionadas. Os dois métodos estocásticos mais conhecidos na docagem molecular são o Algoritmo Genético (AG) e o método de Monte Carlo. O método de Monte Carlo utiliza a distribuição de probabilidade de Boltzmann para calcular a probabilidade de aceite de uma mudança aleatória na conformação do ligante. E o método de Algoritmo Genético se baseia no processo evolucionário de sistemas biológicos para buscar interações favoráveis a tal conformação, utilizando técnicas como seleção, cruzamento e mutação.<sup>49</sup>

A flexibilidade dos resíduos de aminoácidos é uma das primeiras preocupações quando se pretende realizar um cálculo de docagem molecular o mais próximo possível da realidade. E essa é justamente uma das maiores dificuldades encontradas

nesse tipo de cálculo, já que existe um elevado número de graus de liberdade envolvidos, fazendo com que muitos programas limitem o número máximo de resíduos flexíveis. E com essa limitação, surge a necessidade de se escolher quais resíduos apresentarão flexibilidade durante os cálculos. Por isso, é importante que essa escolha seja feita após um conhecimento mais aprofundado sobre o complexo.

Para que o ligante se estabeleça com alta afinidade no sítio do receptor, se faz necessário que as formas tridimensionais do ligante e do receptor sejam complementares. Associado a isso, a capacidade do ligante formar ligações de hidrogênio em pontos considerados chave no sítio é um fator fundamental. E o entendimento de como essa complementaridade pode influenciar na formação do complexo ligante-receptor se torna de grande relevância. A caracterização de uma macromolécula pode ser realizada por cristalografia de raios-X ou por espectroscopia de ressonância magnética nuclear (RMN). Suas coordenadas atômicas ficam à disposição em bancos de dados de macromoléculas, como o *Protein Data Bank* (PDB).<sup>42,50,51</sup>

As interações intermoleculares são as principais responsáveis pela afinidade da interação ligante-receptor. São representadas, principalmente, pelas ligações de hidrogênio, forças de van der Waals e interações eletrostáticas. Os papéis mais significativos, tanto no complexo ligante-receptor quanto na totalidade da macromolécula, são desempenhados pelas ligações de hidrogênio, pois são elas que mantêm a conformação bioativa das proteínas.<sup>52</sup> A identificação de uma ligação de hidrogênio se dá, basicamente, a partir da verificação dos ângulos formados entre os átomos envolvidos na interação e das distâncias entre os mesmos. Os valores dos ângulos considerados para essa ligação estão entre 110° e 180° entre um doador X, o átomo de hidrogênio e um aceptor Y. E as distâncias características dessa ligação estão entre 2,5 e 3,3 Å.<sup>53</sup>

Devido a presença de cadeias apolares no complexo, ocorrem também as interações hidrofóbicas. Essas interações ocorrem quando duas ou mais superfícies hidrofóbicas se aproximam e rompem a solvatação feita pelas moléculas de água. A interação ligante-receptor é favorecida, nesse caso, pela compensação entrópica do sistema.<sup>54</sup> Dessa forma, se uma simulação for realizada com relativa precisão, essas interações poderão ser preditas e explicadas para grande parte dos ligantes.

### 1.6.1 O programa GOLD

O programa GOLD (*Genetic Optimization for Ligand Docking*) utiliza o método de Algoritmo Genético (AG) para explorar a flexibilidade conformacional do ligante e da proteína, ainda que com um número limitado de resíduos flexíveis. O cálculo desse software consiste, basicamente, numa população inicial, criada aleatoriamente, sendo que essa população representa um conjunto de possíveis soluções – orientações e conformações do ligante – onde cada membro é codificado com um cromossomo. Os cromossomos contêm informações relacionadas à conformação do ligante, sendo que os ângulos de torção e distâncias podem ser modificados, além dos potenciais átomos aceptores e doadores para ligações de hidrogênio, como também pontos hidrofóbicos.<sup>55-57</sup>

A partir das coordenadas tridimensionais e do espaço de abrangência, uma população inicial de poses (conformação e orientação) do ligante é gerada aleatoriamente, sendo atribuído a cada indivíduo (ou cromossomo) da população uma pontuação de ajuste com base na previsão de afinidade desse como o espaço de busca delimitado pelo usuário (sítio de ligação). No processo de posicionamento do ligante no seu respectivo sítio, o AG utiliza um processo de ajuste de mínimos quadrados, buscando formar o maior número possível de ligações de hidrogênio, baseado nas características dos cromossomos, para que seja definida a orientação do ligante. As conformações resultantes da rotação torcional são analisadas para verificar a possibilidade e a abrangência das ligações de hidrogênio.<sup>55-57</sup>

A energia de interação total do ligante com a proteína representa a aptidão dessa interação, e essa aptidão é avaliada a partir de uma função de energia. Essa função também é usada para ranquear as melhores interações obtidas nos cálculos, onde se tem uma pontuação associada a cada um dos cromossomos, baseado na afinidade do ligante com a proteína. As principais funções de pontuação executadas no programa GOLD são a GoldScore, a ChemScore, e a ChemPLP.

A função GoldScore combina a energia de van der Waals proteína-ligante, a energia de ligação de hidrogênio do complexo e a energia interna da conformação do ligante. Essa combinação pode ser representada por seis estágios de avaliação:

- Conformação do ligante no respectivo sítio;
- Posicionamento do ligante no interior do sítio;

- Obtenção da energia de ligação de hidrogênio do complexo;
- Obtenção da energia de van der Waals para a interação proteína-ligante;
- Obtenção da energia interna do ligante (mecânica molecular);
- Soma dos termos de energia com seu respectivo peso – pontuação final;

Buscando reproduzir estruturas já conhecidas de complexos formados entre ligantes e macromoléculas, alguns ajustes baseados em conhecimentos empíricos são adicionados aos pesos dos componentes de energia dessa função. A função GoldScore, por exemplo, pode ser descrita pela seguinte expressão:

$$GoldScore = S_{hb\_ext} + 1,375 S_{vdW\_ext} + S_{hb\_int} + S_{vdW\_int} \quad (1)$$

onde:

$S_{hb\_ext}$ : Energia das ligações de hidrogênio proteína-ligante.

$S_{vdW\_ext}$ : Energias de van der Waals do complexo proteína-ligante.

$S_{hb\_int}$ : Energia das ligações de hidrogênio intramoleculares do ligante.

$S_{vdW\_int}$ : Energia das tensões intramoleculares do ligante.

Multiplica-se o termo  $S_{vdW\_ext}$  por 1,375 buscando uma correção empírica para ajustar interações hidrofóbicas entre a proteína-ligante.<sup>58</sup>

Ao final, os maiores valores de pontuação representarão a maior aptidão, e conseqüentemente, tendem a ser as melhores opções para descrever interação ligante-receptor.

Para as contribuições de van der Waals na função de pontuação, o programa utiliza o potencial de Lennard-Jones. Para a contribuição interna, aplica-se um potencial de tipo 6-12, e para a contribuição externa, um potencial 4-8.

No caso de um par de átomos  $i$  e  $j$  do ligante, cujas coordenadas são  $r_i$  e  $r_j$ , a energia potencial de Van der Waals será representada pela equação 2:

$$E_{vdW}(r) = \sum_{j=1}^N \sum_{i=1}^N 4\varepsilon \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2)$$

onde

$$r_{ij} \equiv |r_i - r_j| \quad (3)$$

e  $\mathcal{E}$  caracteriza a intensidade da interação e  $\sigma$  define a distância onde o potencial será nulo. Os expoentes 12 e 6 indicam repulsão em pequena distância e atração em maiores distâncias, respectivamente. Com o aumento da distância entre os dois átomos, a energia se aproxima do zero.<sup>58</sup>

Outros parâmetros como raio de átomos, polarizabilidade, energia e direcionalidade da ligação de hidrogênio, potenciais de torção, entre outros, são parâmetros empíricos incorporados no software e são utilizados para quaisquer funções que venham a ser utilizadas. No programa GOLD, o usuário deve fornecer como entrada para o AG a localização, o ligante e o tamanho do sítio de interação. E o programa fornecerá como saída as conformações do ligante e proteína associadas ao cromossomo mais apto na população final. E, para análise visual, são também apresentados os resultados dos cálculos em tela gráfica (poses – conformação e orientação), a fim de que o melhor complexo receptor-ligante seja selecionado.<sup>55-57</sup>

### 1.7 Algoritmo genético

O método de Algoritmo Genético (AG), dentre diversas aplicações, tem sido utilizado em programas de docagem molecular com o objetivo de determinar o posicionamento dos ligantes nos seus respectivos receptores e também para predizer o melhor modelo de ligação, baseado nas funções de pontuação utilizadas. Em processos que buscam a otimização combinatorial, onde a demanda de cálculos é relativamente elevada, esse algoritmo é muito bem aproveitado. Como exemplo desses processos combinatoriais, temos o *design* de compostos, relações estrutura-atividade e o reconhecimento molecular.

Esse algoritmo mimetiza o processo de seleção natural, se baseando em características importantes para os organismos vivos, como genética e evolução biológica. E, mesmo existindo um grande número de variações na sua aplicação, a ideia principal do algoritmo é comum a todas.<sup>59</sup>

Uma possível solução ao problema, que é representada por uma estrutura de dados, no caso do AG é um cromossomo. E no caso da docagem molecular, são a

orientação e a conformação do ligante no seu respectivo sítio de interação no receptor. O AG, tendo por exemplo, algumas dezenas de cromossomos, é capaz de manipular essa população, ou seja, um grupo de soluções. Cada um dos cromossomos é representado por uma sequência de bits, onde cada posição de bit (0 ou 1) representa um gene desse cromossomo. E a sequência dos bits, analogamente a sistema biológicos, é chamada de genótipo.

De maneira bastante resumida, o procedimento do AG passa pela codificação das variáveis onde, em seguida, forma-se a população inicial. Dessa primeira população é obtida uma resposta sujeita a avaliação. Para dar prosseguimento ao processo, são realizados cruzamentos e mutações, os quais também são avaliados em cada etapa, a fim de que ocorram processos de seleção e também de herança.

Gera-se aleatoriamente a população inicial de indivíduos. E todos os indivíduos são testados conforme os parâmetros iniciais estabelecidos, onde se atribui uma avaliação numérica através de uma função de pontuação estabelecida. Essa função retorna uma pontuação, indicando os indivíduos mais aptos com os valores numéricos mais altos. Assim, tem-se a associação de cada cromossomo com um valor de aptidão. Os valores de aptidão são também úteis na seleção de indivíduos que serão utilizados como cromossomos parentais nas gerações posteriores. Depois da avaliação, a seleção é feita pela eliminação dos indivíduos da população com baixos valores para a função de ajuste, fazendo com que, somente os indivíduos mais aptos transmitam seus genes. Após a identificação de um cromossomo apto, manipula-se o mesmo com operadores genéticos, visando a formação de novos cromossomos em gerações subsequentes. Ao final dessa etapa, é feita, novamente, uma avaliação.<sup>60</sup>

Os operadores genéticos podem ser de cruzamento e de mutação. E são aplicados probabilisticamente aos indivíduos selecionados. O operador de cruzamento combina partes dos cromossomos mais aptos visando a geração de novos cromossomos. Necessita, portanto, de dois cromossomos parentais para produzir dois cromossomos filhos. Já o operador de mutação introduz, aleatoriamente, novas informações em um cromossomo. Assim, necessita de um cromossomo parental e produz um cromossomo filho.

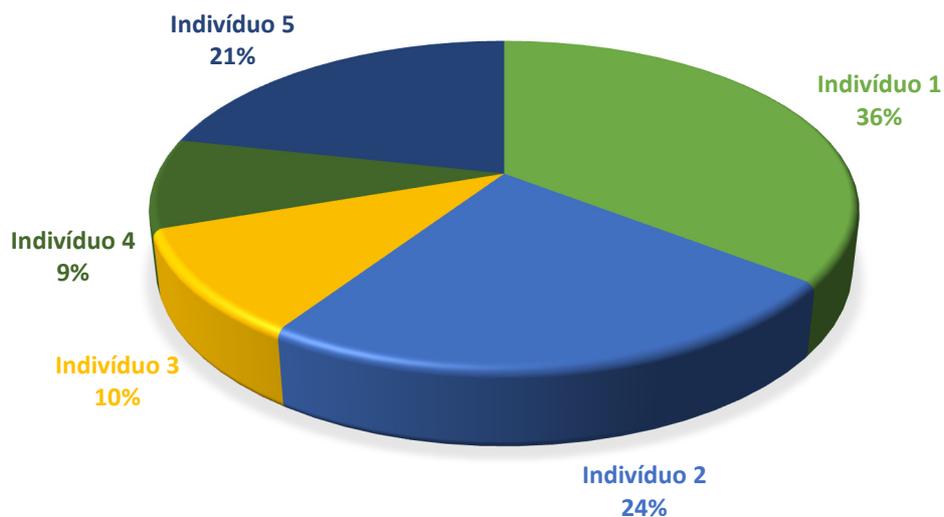
Além da manipulação direta de uma população de cromossomos, utiliza-se também o modelo de ilhas, que divide essa população em várias subpopulações e promove a migração de cromossomos individuais entre as mesmas. Nesse caso, a migração também é um operador genético. Porém ao invés desse operador

efetivamente migrar um indivíduo para outra ilha, ele apenas o copia. Geralmente, o modelo de ilhas tende a melhorar o desempenho do AG. Os operadores genéticos são aplicados repetidas vezes, sempre substituindo os cromossomos menos aptos da população, até que a condição estabelecida como satisfatória seja alcançada.

A seleção dos cromossomos é realizada com o auxílio de um método conhecido como Método da Roleta, conforme apresentado na Figura 5. Nesse método, a probabilidade de um cromossomo ser escolhido para ser o cromossomo pai é diretamente proporcional à sua aptidão. Isso também pode ser interpretado como uma pressão evolutiva no algoritmo. Pois, em analogia a uma roleta, quanto maior a aptidão do membro, maior será a sua fatia na roleta.<sup>55</sup>

Dessa forma, garantindo sempre a sobrevivência do mais apto, ao longo do tempo, a população se direciona a uma solução ótima, que no caso da docagem molecular, implica que o cálculo se direcione ao melhor modo de ligação. No Método da Roleta também é feita a escolha da aplicação do operador genético. Nesse caso, a probabilidade de um operador ser selecionado dependerá do peso atribuído ao mesmo, o qual pode ser definido inicialmente na definição dos parâmetros. A atribuição do peso dependerá da relevância que o usuário dará para o operador naquele caso.

Figura 5 – Método da roleta. A probabilidade de escolha de cada um dos indivíduos está diretamente relacionada à sua aptidão.



Fonte: Elaborada pelo autor.

Outro fator de grande relevância no AG é a formação de nichos. Essa técnica é bastante utilizada para evitar um excesso de soluções demasiadamente

semelhantes sejam obtidas, preservando assim a diversidade da população. Assim, quando um novo indivíduo é incluído na população, é feita uma comparação com todos os membros dessa população visando identificar se não há nenhum outro membro ocupando o mesmo nicho. Ou seja, verifica-se a existência de soluções semelhantes. No caso da docagem molecular, a comparação pode ser feita através do RMSD (*Root-Mean-Square deviation*) das coordenadas tridimensionais dos átomos de cada um dos ligantes, onde pode ser estabelecido um parâmetro de corte.<sup>55</sup>

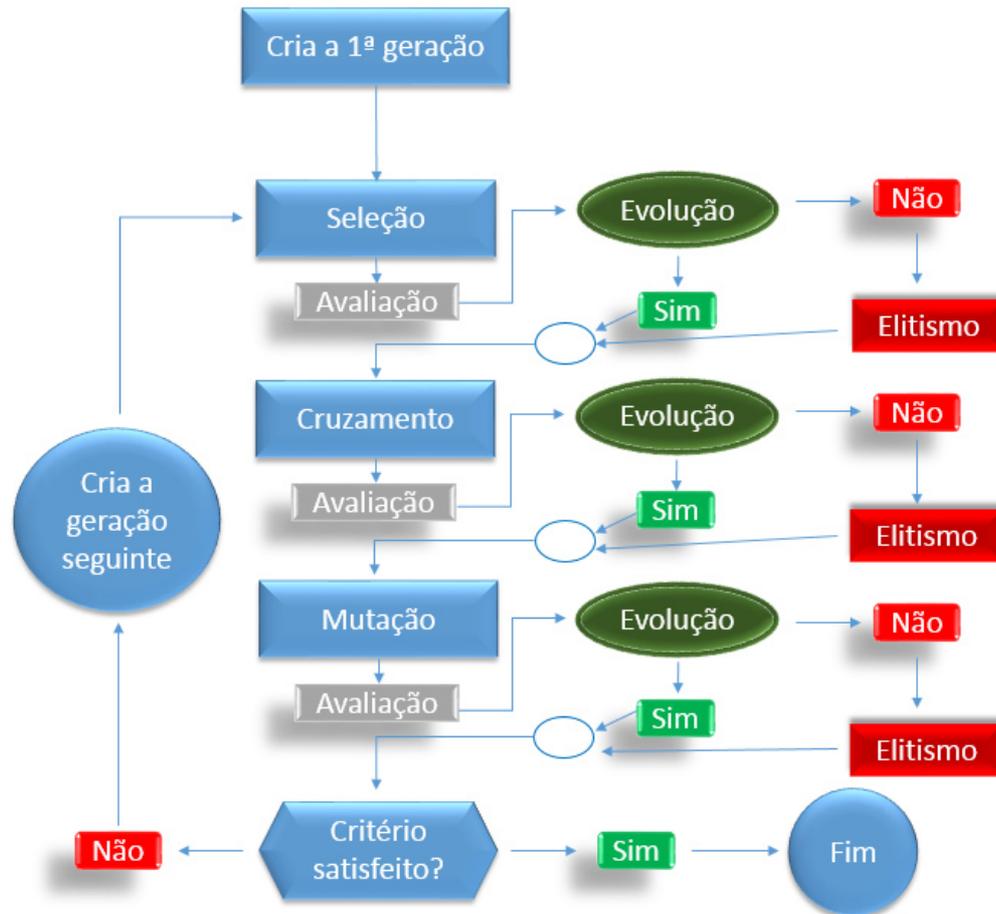
Resumidamente, pode-se descrever o procedimento do Algoritmo Genético nas seguintes etapas:<sup>55-56</sup>

- Escolhe-se um grupo de operadores genéticos (cruzamento, mutação, migração) e atribuiu-se um peso a cada um deles
- Forma-se uma população inicial aleatoriamente e a função de pontuação determina a aptidão dos seus membros;
- Escolhe-se um operador baseando-se em cálculos de probabilidade com os respectivos pesos de cada operador;
- O operador escolhe os pais através de seleção baseada na aptidão dos indivíduos;
- Cromossomos filhos são produzidos pelo operador e, com a função de pontuação, avalia-se a aptidão;
- Os filhos substituem os indivíduos menos aptos;
- O processo para se a solução for aceitável. Se não, volta-se para a etapa de escolha dos pais;

Esse algoritmo pode ser, de forma bastante simplificada, representado pela Figura 6.

Como na natureza, a informação disponível sobre os indivíduos encontra-se nos cromossomos. E buscando desenvolver cromossomos mais evoluídos, ou seja, com o melhoramento do valor da função de pontuação, são aplicados os operadores genéticos sucessivamente ao longo das gerações. Dessa forma, as soluções ou indivíduos mais adaptados ao seu meio se perpetuam através da reprodução, enquanto as menos adaptadas se encaminham para a extinção.

Figura 6 – Representação simplificada do algoritmo genético.



Fonte: Elaborada pelo autor.

### 1.8 Relação Quantitativa Estrutura-Atividade (QSAR)

O estudo de relações quantitativas entre estrutura e atividade de um conjunto de compostos possibilita correlacionar as propriedades estruturais e físico-químicas dos mesmos com suas respectivas respostas biológicas, para gerar modelos capazes de prever a atividade biológica de outros compostos não analisados. Essas relações são amplamente utilizadas para compreender como as características estruturais de moléculas biologicamente ativas contribuem para sua atividade.<sup>61</sup> Assim, é possível analisar esses compostos em termos dessas propriedades estruturais, ao invés de considerar apenas certos grupos farmacofóricos.

Também chamadas de modelos, as QSARs são apresentadas como equações, quantificando a atividade biológica como função das propriedades das moléculas. Esses modelos podem ser utilizados, por exemplo, para priorizar, em pesquisas

posteriores, moléculas que tenham apresentado um melhor desempenho no que diz respeito às predições das suas respectivas atividades biológicas.<sup>62</sup>

Para que um modelo de QSAR tenha êxito em seu desenvolvimento, vários aspectos devem ser considerados. Deve haver um critério bastante rigoroso na seleção dos descritores e nas análises estatísticas, para que parâmetros realmente relevantes sejam selecionados. Tão fundamental quanto essas análises, é a realização de testes de validação dos modelos, a fim de que a confiabilidade e funcionalidade do modelo sejam asseguradas. Para que um conjunto de moléculas possa ser analisado, essas devem possuir um esqueleto em comum, onde variações estruturais são admitidas. Isso se deve à intenção de avaliar o mecanismo de ação desse conjunto de compostos e, a partir dessa avaliação, estender esse entendimento a outros compostos semelhantes ao do conjunto.

Grande quantidade de dados é obtida de cada uma das moléculas otimizadas, e muitos desses são gerados a partir de descritores moleculares, que são tratados como variáveis. Assim, podem ser obtidas importantes informações relacionadas à atividade biológica dos compostos. Esses descritores podem ser classificados, por exemplo, como geométricos, físico-químicos, topológicos, eletrônicos, entre outros. Após a obtenção dos cálculos das geometrias das moléculas, é possível obter diversos outros descritores moleculares. E com todos os dados reunidos, pode-se utilizar métodos para separar as variáveis que são relevantes das não relevantes para explicar a variação da atividade biológica. Para uma seleção inicial das variáveis, pode ser usada a análise de correlação com os valores de atividade biológica, selecionando as variáveis que mantenham alta correlação com as respectivas atividades. O método de Algoritmo Genético pode ser utilizado na etapa seguinte, para seleção dos descritores.<sup>63</sup> Para a construção de um modelo matemático, podem ser utilizadas técnicas de Mínimos Quadrados Parciais (PLS - do inglês, *partial least squares*), bem como de Redes Neurais Artificiais (ANN - do inglês, *artificial neural network*). São técnicas de grande importância, já que são elas as responsáveis pelas combinações das variáveis.

### 1.8.1 Mínimos Quadrados Parciais (PLS)

O método de PLS, depois de fazer as devidas correlações entre as variáveis, fornece como resultado final um modelo matemático capaz de prever a atividade

biológica de um determinado conjunto de compostos. Nesse método, duas matrizes X e Y são relacionadas, onde X representa os descritores e Y a atividade biológica:

$$X = TP' + E = \sum t_n p'_n + E \quad (4)$$

$$Y = UQ' + F = \sum u_n q'_n + F \quad (5)$$

onde T é a matriz de escore da matriz X, e P é o seu peso, assim como U vem a ser a matriz de escore da matriz Y e Q o seu peso. Os resíduos são, respectivamente, E e F, e os somatórios são compostos por  $n$  variáveis.

Diversos parâmetros indicam a consistência e qualidade de um modelo de PLS, como por exemplo o coeficiente de correlação,  $r$ , o qual determina a razão entre o valor mensurado da atividade biológica e o valor predito pelo modelo matemático gerado. Outro parâmetro de grande relevância nessas análises é o  $r^2$ , que corresponde à fração da variabilidade total que é explicada pelo modelo, portanto quanto mais próximo o valor de  $r^2$  for de 1, maior a capacidade de o modelo explicar as variações em torno da média.<sup>64,65,66</sup>

$$r^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - y_{mean})^2} \quad (6)$$

onde:

$y_{iexp}$ : valor experimental da atividade da  $i$ -ésima amostra;

$\hat{y}_{ical}$ : previsão da atividade da  $i$ -ésima amostra utilizando o modelo com todas as amostras;

$y_{mean}$ : valor médio experimental da atividade;

A razão entre a variabilidade explicada pelo modelo e a variabilidade não explicada pelo modelo é denominada equação de Teste F:

$$Teste F = \frac{\sum (y_{iexp} - y_{mean})^2 - \sum (y_{iexp} - \hat{y}_{ical})^2}{\sum (y_{iexp} - \hat{y}_{ical})^2} \frac{(I - 1)}{VL} - 1 \quad (7)$$

nesse caso, o número de amostras é representado por  $I$  e número de variáveis latentes por  $VL$ . Assim, quanto maior for o valor de Teste F, maior será a proporção de características explicadas pelo modelo, ou seja, os descritores utilizados na geração do modelo foram bem escolhidos.

A consistência do modelo de regressão também pode ser analisada a partir dos valores de  $q^2$ , que mensura a capacidade de predição de novos compostos.

$$q^2 = 1 - \frac{\sum (y_{iexp} - \hat{y}_{ival})^2}{\sum (y_{iexp} - y_{mean})^2} \quad (8)$$

onde  $\hat{y}_{ival}$  representa a predição da atividade sem a  $i$ -ésima amostra, utilizando o modelo sem a mesma. O  $q^2$  é obtido a partir da validação cruzada. O mecanismo que faz essa validação é o método “*leave-one-out*”, que cria um modelo de regressão sem uma molécula do conjunto, e utilizando esse novo modelo para prever a atividade biológica do composto retirado do conjunto. O desvio padrão entre o valor predito e observado é calculado. Até que todas as moléculas do conjunto sejam retiradas e validadas, esse processo se repetirá.

Dentre os parâmetros ligados aos erros de predição, os três principais são: *PRESS*, *SEV* e *SEC*, os quais são apresentados a seguir.

- Soma dos quadrados dos erros de predição (*PRESS* - do inglês, *Prediction Residues Error Squares Sum*):

$$PRESS = \sum (y_{iexp} - \hat{y}_{ival})^2 \quad (9)$$

- Erro padrão de validação (*SEV* - do inglês, *Standard Error of Validation*):

$$SEV = \left[ \frac{PRESS}{I} \right]^{1/2} \quad (10)$$

- Erro padrão de calibração (*SEC* - do inglês, *Standard Error of Calibration*):

$$SEC = \left[ \frac{\sum (y_{iexp} - \hat{y}_{ical})^2}{I - VL - 1} \right]^{1/2} \quad (11)$$

Dessa maneira, vemos que quanto mais próximo de 1 os valores de  $r^2$  e  $q^2$  estiverem, melhor será o modelo. Assim como os menores valores possíveis de *PRESS*, *SEV* e *SEC*, e os maiores valores possíveis de Teste-F.<sup>64,65,66</sup>

### 1.8.2 Redes Neurais Artificiais

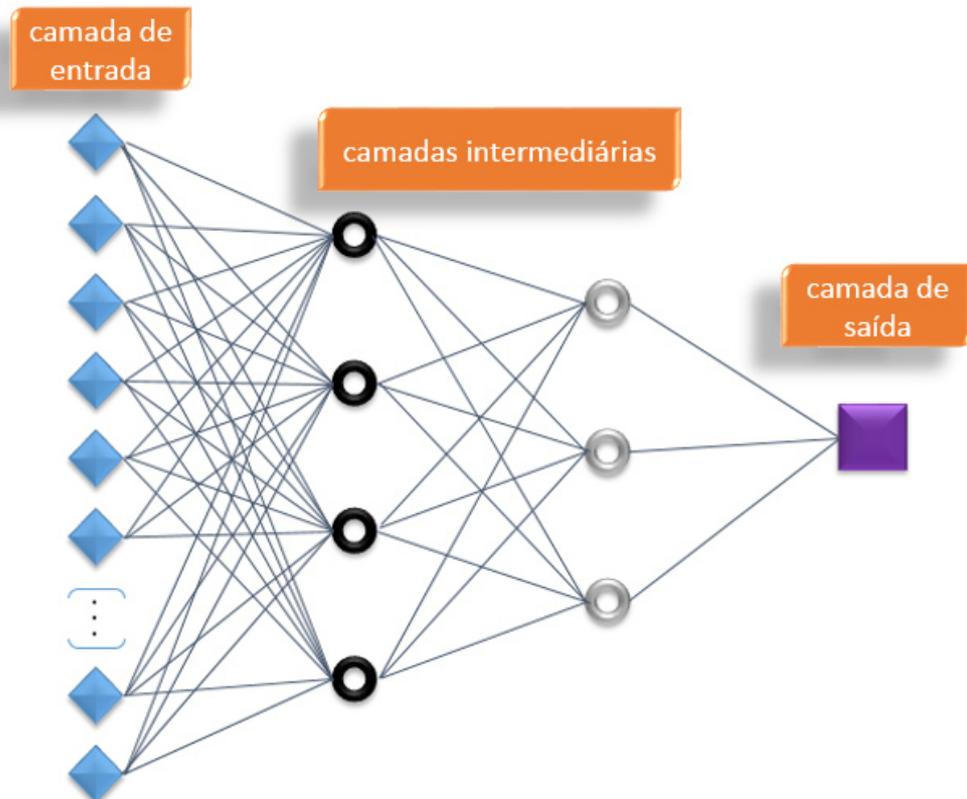
O método de Redes Neurais Artificiais (ANN) é um método computacional de reconhecimento de padrões que apresenta um modelo matemático como produto final. Ele é inspirado na estrutura neural de organismos inteligentes que adquirem conhecimento a partir da experiência.<sup>67</sup> Uma das características mais importantes desse método é a habilidade de aprender com seu ambiente, para com isso melhorar seu desempenho. E esse processo é feito de modo iterativo com ajustes aplicados a seus pesos, que é chamado de treinamento das redes. Quando a rede neural atinge uma solução generalizada para uma classe de problemas, pode ser dito que ocorreu o aprendizado. Existem diversos tipos de algoritmos de aprendizado, específicos para determinados modelos de redes neurais, que diferem entre si principalmente pela modificação dos pesos.

Basicamente, o funcionamento das redes neurais artificiais é realizado por etapas que podem ser divididas em camadas, que são classificadas em três grupos: camadas de entrada, intermediárias e de saída. Na camada de entrada, os padrões são apresentados à rede. Nas camadas intermediárias, a maior parte do processamento é realizada a partir das conexões ponderadas. Por fim, na camada de saída, o resultado final é concluído e apresentado. Um modelo que possibilita a utilização de camadas intermediárias é o do tipo Perceptron de Multicamadas, MLP (do inglês, *Multilayer Perceptron*), que está representado na Figura 7.<sup>68</sup>

Após serem multiplicados pelos sinais de entrada, os pesos são somados, sobre os quais é aplicada uma função de transferência, a fim de que seja fornecida uma saída. As funções de transferência, portanto, atuam sobre as entradas apresentadas ao neurônio. Como exemplo, podemos citar funções como a *tansig* (tangente hiperbólica sigmoidal), a *logsig* (logarítmica sigmoidal) e a *purelin* (linear).<sup>68,69</sup> Os valores numéricos são ajustados, a partir do algoritmo de aprendizagem, até se obter outra resposta, de maneira análoga à introduzida inicialmente. A função desse algoritmo é calcular o erro dos valores obtidos em relação

aos fornecidos à rede, após serem fornecidas as variáveis de entrada e a geração da resposta. O erro é então utilizado para estimar os erros das camadas intermediárias, a fim de que ele seja retropropagado até as conexões da camada de entrada – os pesos, os quais são ajustados para ser obtido um menor erro na camada de saída.<sup>70</sup>

Figura 7 – Representação do funcionamento das Redes Neurais Artificiais do tipo MLP.



Fonte: Elaborada pelo autor.

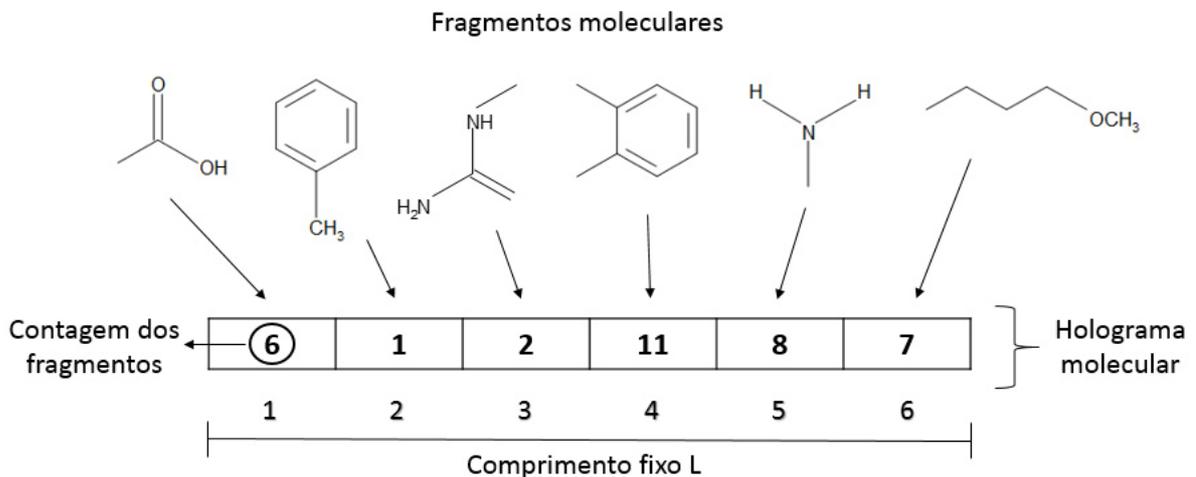
## 1.9 HQSAR

Baseado em fragmentos moleculares especializados, o holograma QSAR (HQSAR, do inglês, *hologram* QSAR) é um método no qual as moléculas são convertidas em sequências numéricas que codificam suas respectivas estruturas bidimensionais. E, a partir dessas informações codificadas, é possível estabelecer relações quantitativas entre tais estruturas e a atividade biológica. Nesse caso, essas relações são estabelecidas utilizando o método de PLS.<sup>71</sup>

O processo consiste em fragmentar, estruturalmente, cada uma das moléculas, onde os fragmentos resultantes são organizados em uma sequência numérica – o holograma molecular. Cada uma das estruturas do conjunto analisado tem o seu respectivo holograma. Dessa maneira, os hologramas moleculares atuam como impressões digitais moleculares, codificando o maior número de fragmentos possível, tais como lineares, ramificados e cíclicos. A ocorrência desses fragmentos é contabilizada, e fragmentos iguais são alocados na mesma posição do holograma.<sup>71,72</sup>

Sendo assim, o holograma pode ser representado como um arranjo de caixas, onde cada uma dessas caixas corresponde a um tipo específico de fragmento e, em cada uma dessas caixas está contido o número de fragmentos contidos nessa estrutura. Esse arranjo de caixas possui um comprimento L, conforme a Figura 8. A qualidade de um modelo HQSAR depende diretamente da maneira como os fragmentos estão distribuídos nos hologramas, já que reúnem diversas informações sobre as características dessas estruturas.<sup>73</sup>

Figura 8 – Processo de geração de um holograma molecular.



Fonte: Elaborada pelo autor.

Os dados de entrada requeridos para a aplicação do método de HQSAR são as estruturas moleculares em suas representações 2D e os respectivos valores de atividade biológica. Dessa maneira, se mostra como um método relativamente simples, se comparado a outros, onde se faz necessário, por exemplo, o alinhamento molecular tridimensional. O método de HQSAR emprega diversas combinações de

parâmetros que irão definir os melhores modelos para o conjunto de moléculas analisado. Entre tais parâmetros, podemos citar o comprimento do holograma, o tamanho dos fragmentos e a distinção dos fragmentos. O comprimento do holograma define quantas posições serão necessárias para alocar os diferentes fragmentos. O tamanho dos fragmentos está diretamente relacionado ao número de átomos admitido a cada um dos fragmentos, podendo ser definido um número máximo e mínimo de átomos. Já a distinção dos fragmentos define quais características moleculares serão responsáveis por diferenciar os fragmentos entre si, entre as quais podemos citar os tipos de átomo, ligação, conectividade, quiralidade, presença de átomos de hidrogênio e a presença de átomos doadores ou aceptores de ligações de hidrogênio, que são descritos com mais detalhes na Tabela 1.<sup>73,74</sup>

Para impedir que fragmentos distintos e importantes sejam agrupados numa mesma posição, várias combinações de distribuição de fragmentos são testadas, já que os parâmetros utilizados na distinção são ajustáveis, e são eles que irão definir o tipo de informação que será incluída na formação dos hologramas.

Tabela 1 – Parâmetros de distinção dos fragmentos utilizados no método HQSAR.

Distinção dos fragmentos	Definição
<b>Átomos (A)</b>	Distinção dos fragmentos feita a partir dos tipos de átomos.
<b>Ligação (B)</b>	Distinção a partir dos tipos de ligações (simples, dupla, tripla, aromática).
<b>Conectividade (C)</b>	Distinção baseada no estado de hibridização dos átomos daquele fragmento.
<b>Hidrogênio (H)</b>	Distinção a partir do número de átomos de hidrogênio.
<b>Quiralidade (Ch)</b>	Distinção baseada na presença de átomos com centros quirais.
<b>Doadores e Aceptores (DA)</b>	Distinção baseada na presença de átomos doadores ou aceptores de ligações de hidrogênio.

### 1.10 QSAR 3D

A análise comparativa dos campos moleculares (CoMFA – do inglês, *comparative molecular field analysis*) é um método de QSAR 3D onde se utiliza um conjunto de moléculas estruturalmente similares ou que têm apenas o farmacóforo em comum. Nesse método, a molécula deve estar alinhada com as demais, a fim de se obter para todas as moléculas do conjunto uma grade virtual tridimensional. Essa grade representa campos moleculares de interação (MIFs, do inglês, *molecular*

*interaction fields*), que são a combinação de um grande número de descritores das interações estéricas e eletrostáticas. Calcula-se, em todos os pontos dessa grade, a energia de interação de um átomo de prova (ou sondas) com cada molécula. Dessa maneira, pode ser estabelecida uma relação entre a energia de interação e a atividade biológica de um conjunto de compostos.<sup>75</sup>

Utiliza-se as energias de interação de Van der Waals e de Coulomb, calculadas a partir dos potenciais de Lennard-Jones e de Coulomb, para descrever o comportamento dos descritores estéricos e eletrostáticos.<sup>76</sup>

A utilização do QSAR 3D admite a geração de modelos a partir de um conjunto de compostos estruturalmente diversificado, desde que mantida uma única característica comum entre eles – o farmacóforo, que geralmente servirá como referência para o correto alinhamento do conjunto de moléculas. Esse alinhamento pode se dar a partir de um corpo rígido: uma estrutura comum que se faça presente e todas as moléculas do conjunto analisado, ou então também pode ser baseado em requerimentos farmacofóricos previamente identificados. Os métodos de QSAR 3D não necessitam, diretamente, de informações a respeito da estrutura do receptor.<sup>77</sup>

As energias de interação calculadas em cada ponto da grade, compõem os descritores (variáveis independentes) que serão correlacionados com a atividade biológica (variável dependente). Essa correlação geralmente é calculada com o método de PLS, tendo como resultado a geração de um modelo. A robustez interna desse modelo pode ser verificada a partir de técnicas de validação cruzada. Além do modelo gerado, tem-se também como resultado a representação gráfica dos mapas de contorno estéricos e eletrostáticos, sendo possível analisá-los ao redor das moléculas, evidenciando regiões favoráveis ou desfavoráveis para a atividade biológica desse grupo de compostos.<sup>76,78</sup>

O método CoMSIA (do inglês, *comparative molecular similarity indices analysis*) apresenta características bastante semelhantes ao CoMFA. No entanto, além de interações estéricas e eletrostáticas, é capaz descrever as moléculas com campos hidrofóbicos e também, com campos de doadores e aceptores envolvidos nas ligações de hidrogênio. Apesar de semelhantes, o método CoMSIA utiliza uma função gaussiana dependente da distância para descrever as energias de interação, resultando em inclinações mais suaves se comparada aos potenciais de Lennard-Jones e de Coulomb.<sup>76,79</sup>

## 2. OBJETIVOS

Os objetivos gerais deste trabalho são: prever a atividade biológica de um grupo de chalconas a partir de diversos modelos matemáticos, e propor novas estruturas, baseando-se em características pertinentes aos complexos formados por esse grupo de compostos e seu respectivo sítio de ligação.

Objetivos Específicos:

- ✓ Desenvolver modelos de QSAR, baseados em análises de PLS e ANN.
- ✓ Gerar um modelo de HQSAR utilizando as combinações de fragmentos moleculares de maior relevância.
- ✓ Propor modelos de QSAR 3D, a partir de análises de CoMFA e CoMSIA.
- ✓ Compreender o comportamento dos complexos formados por esse grupo de compostos e seu sítio de ligação, utilizando as análises de docking molecular.
- ✓ Propor novas estruturas, a partir das análises do comportamento dos ligantes nos complexos.
- ✓ Analisar e testar as novas estruturas nos modelos gerados, buscando a predição da atividade biológica das mesmas.

### 3. METODOLOGIA

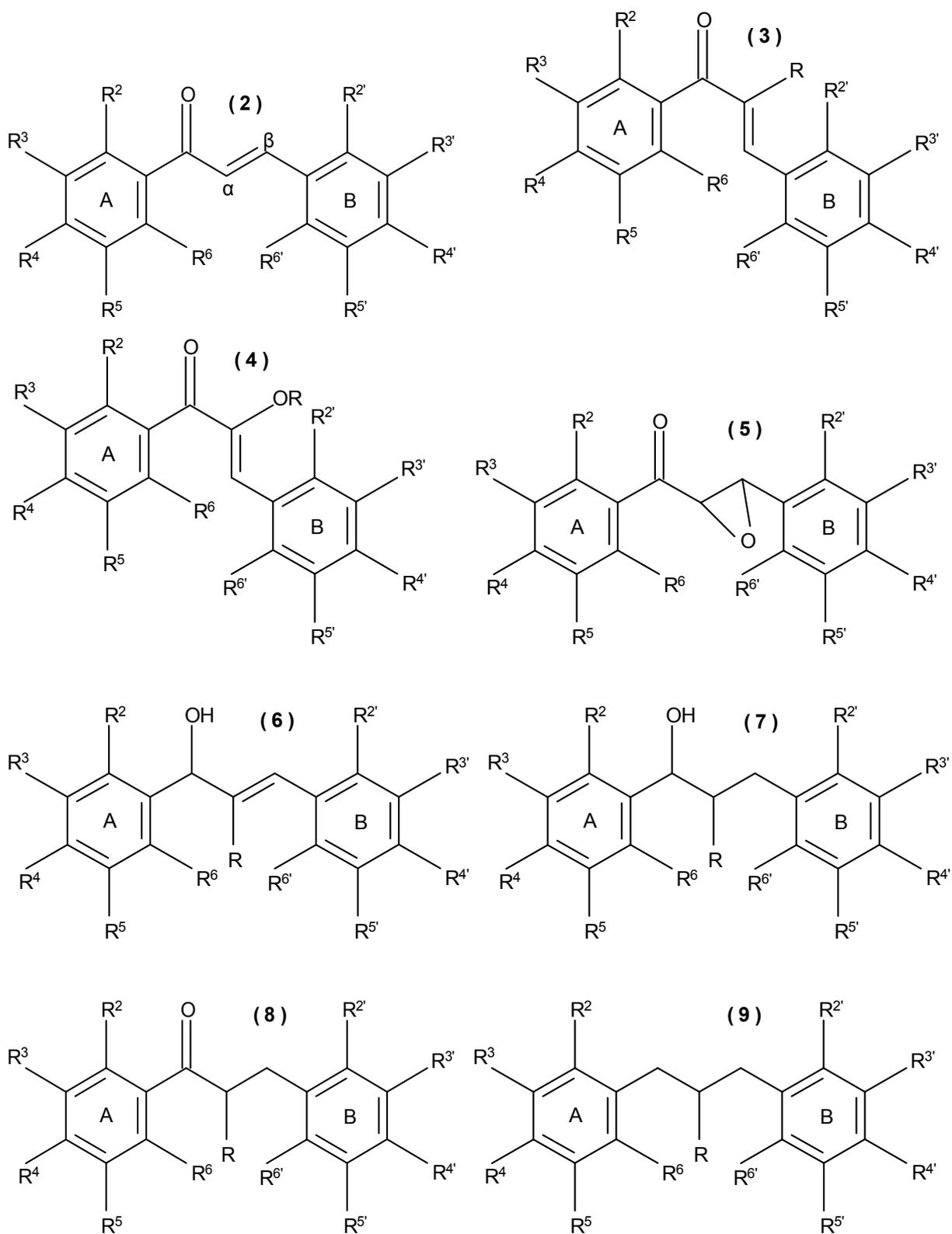
As moléculas utilizadas neste trabalho foram propostas no estudo de Ducki *et al*<sup>25</sup>, onde foram sintetizadas e testadas biologicamente. Desse estudo foram selecionadas as 87 chalconas que serão usadas neste trabalho. A descrição das moléculas selecionadas com suas respectivas representações estruturais e valores de atividade biológica ( $pIC_{50}$ ) estão apresentados na Tabela 20, do apêndice A. A atividade biológica medida em  $IC_{50}$  corresponde à mínima concentração em que uma substância seja capaz de inibir 50% de determinada cultura. Os esqueletos principais das oito estruturas base dessas chalconas estão apresentados na Figura 9.<sup>25</sup>

#### 3.1 Modelos de QSAR

Os cálculos relacionados a estrutura eletrônica desse conjunto de moléculas foram realizados com computadores (Core i7) do Grupo de Química Quântica do Instituto de Química de São Carlos. O software utilizado foi o Gaussian 09<sup>80</sup>, com o método de cálculo DFT, funcional híbrido B3LYP<sup>81</sup> e função de base 6-31g++(d,p),<sup>82</sup> a fim de otimizar a geometria dos compostos e calcular diversas características dessas moléculas, que foram utilizadas posteriormente para obtenção de diversos descritores eletrônicos.

Após a otimização, foram obtidos os descritores moleculares com a utilização da plataforma E-Dragon.<sup>83</sup> A partir de uma análise prévia do coeficiente de correlação de cada um dos descritores com a atividade biológica, foi possível selecionar os descritores com alta relevância, onde os descritores cuja correlação tenha sido menor que 0,3, foram excluídos. O método de Algoritmo Genético também foi utilizado como seletor de descritores, onde foi utilizado o software BuildQSAR.<sup>84</sup> Foram considerados diversos parâmetros como o número de gerações, descritores por modelo e alta correlação de variáveis. Os valores de R e de  $q^2$  ajudaram a indicar os melhores modelos e, conseqüentemente, os melhores descritores.

Figura 9 – Estruturas base dos compostos analisados. A maioria das moléculas é oriunda das estruturas base 2, 3 e 4.



Fonte: Elaborada pelo autor.

### 3.1.1 Metodologia de PLS

Após serem definidos os descritores de maior relevância, cujos valores foram autoescalados, foram realizadas análises de PLS, onde foi utilizado o software Pirouette 3.10.<sup>85</sup> A escolha do melhor modelo foi baseada em parâmetros estatísticos como  $q^2$ ,  $r^2$ , PRESS, SEC e SEV. Para validar o modelo escolhido, foram realizadas análises como o gráfico de resíduos de Student *versus* alavancagem (para identificar a presença de *outliers*), testes de leave-N-out (para verificar a robustez) e randomização da atividade biológica (para mostrar que o modelo não foi gerado ao acaso). Todos os testes de validação interna do modelo foram realizados com o uso do programa QSAR Modeling.<sup>86</sup>

### 3.1.2 Metodologia de ANN

O critério de seleção dos conjuntos de treinamento e teste foi o mesmo para os modelos de PLS e ANN. Foram selecionados 10% do total (9 compostos, denotados por um asterisco (\*) na Tabela 20, do apêndice A) para a validação interna do método de ANN (a fim de que não haja superajuste). Esses compostos foram retirados das análises de PLS, para que pudesse ser feita uma comparação direta entre ambos os modelos. Tanto nos modelos de PLS quanto de ANN, foram utilizados 70% dos compostos (61 compostos) no conjunto de treinamento e 20% (17 compostos, sublinhados na Tabela 20) no conjunto de teste. Ou seja, a fim de se fazer a comparação mais direta possível entre os dois métodos, foram retirados do modelo de PLS os compostos presentes na validação interna do modelo de ANN. Assim, os compostos presentes nos conjuntos de treinamento e teste de PLS e ANN passaram a ser exatamente os mesmos. Vale ressaltar que os descritores selecionados também são os mesmos.

A separação dos conjuntos de treinamento, validação e teste, tanto para as análises de PLS quanto para as análises de ANN foi feita por meio do método de ANN, seguindo, basicamente, o mesmo padrão de treinamento das redes utilizadas no decorrer do trabalho, e ainda assim, respeitando a quantidade mínima de 20% dos compostos no conjunto de teste para cada variação de uma unidade na atividade biológica.

Nas análises de ANN, o treinamento da rede se deu com o uso do programa Matlab.<sup>87</sup> O algoritmo utilizado foi o *backpropagation*. Baseados na normalização de valores máximos e mínimos, os dados foram escalados em uma faixa de 0 a 1. As redes de alimentação utilizadas foram do tipo MLP-ANN (Perceptron de multicamadas).

No processo de escolha do número de neurônios e das funções de transferência ideais para o modelo, passaram a ser considerados de 1 a 20 neurônios e 3 tipos de funções de transferência (*logsig*, *tansig* e *purelin*). Sendo utilizadas duas camadas intermediárias, havia a possibilidade de duas funções diferentes serem implementadas no mesmo cálculo. Assim, somam-se 9 variações de funções de transferência: *logsig-logsig*, *logsig-tansig*, *logsig-purelin*, *tansig-tansig*, *tansig-logsig*, *tansig-purelin*, *purelin-purelin*, *purelin-tansig*, *purelin-logsig*. E utilizando 20 neurônios para cada uma dessas variações, chegamos a um total de 180 combinações.

A escolha da melhor combinação se baseou em parâmetros como  $r^2$  e EQM (erro quadrático médio). O parâmetro  $r^2$  foi utilizado para os três conjuntos de compostos, portanto  $r^2_{\text{treinamento}}$ ,  $r^2_{\text{validação}}$  e  $r^2_{\text{teste}}$ . Além de buscar os melhores valores para esses 3 parâmetros, optava-se pelo melhor equilíbrio entre eles. Para evitar que houvesse superajuste nos modelos gerados, buscou-se manter somente os modelos cujos valores de EQM dos conjuntos de treinamento se mantivessem menores que os de EQM dos conjuntos de validação e teste.

$$EQM = \frac{1}{N} \sum_{i=1}^N (y_{ical} - y_{iexp})^2 \quad (12)$$

Ao final, dentre as 180 combinações possíveis, foram selecionadas 10% delas, com base no EQM, para compor o modelo *Consensus-ANN* (ANN-CM). Essa abordagem leva em consideração as diferentes assimilações de cada um dos modelos gerados durante a aprendizagem das redes. E cada um dos modelos é ponderado por um fator de normalização (FN), que se baseia nos valores dos seus respectivos EQMs.

$$FN = \frac{EQM^{-1}}{\sum EQM^{-1}} \quad (13)$$

### 3.2 Modelos de HQSAR

Para a construção dos modelos de HQSAR foi utilizado o software SYBYL-X 1.2, e várias combinações de distinção dos fragmentos foram consideradas, já que tais modelos são influenciados pelos parâmetros relacionados à geração dos hologramas. Os parâmetros analisados foram: átomo (A), ligação (B), conectividade (C), presença de átomos de hidrogênio (H), quiralidade (Ch) e átomos doadores e aceptores de ligação de hidrogênio (DA). Foram utilizadas diversas combinações dos parâmetros de distinção, a fim de encontrar a combinação que pudesse gerar o modelo com maior capacidade preditiva. Os comprimentos do holograma foram testados de 53 a 401 posições, utilizando as 12 séries padrão disponíveis no programa. Para tais análises, foi mantido um mínimo de 20% da quantidade total de compostos para compor o conjunto de teste, correspondente à validação externa dos modelos gerados.<sup>88,89</sup>

Inicialmente, o tamanho do fragmento foi adotado como 4-7, a partir do qual foram testadas as combinações de parâmetros e comprimentos do holograma, para posteriormente serem testadas outras possibilidades de tamanho do fragmento. Após ser estabelecida a melhor combinação dos parâmetros relacionados à geração dos hologramas, foi testado o melhor tamanho do fragmento, cujas variações foram de 2-5, 3-6, 4-7, 5-8, 6-9, 7-10, 8-11, 9-12 e 10-13. Além dos parâmetros relacionados à geração dos hologramas, também foram analisados os parâmetros inerentes às análises de PLS, tais como  $r^2$  (coeficiente de correlação),  $q^2$  (coeficiente de validação cruzada), StdErr (erros padrão), comprimento de holograma e número de componentes.<sup>90</sup>

Após definido o melhor modelo HQSAR para o conjunto de compostos, foi possível gerar mapas de contribuições individuais 2D, os quais podem fornecer indicações sobre os fragmentos moleculares diretamente relacionados à atividade biológica, contribuindo inclusive, para a proposição de novas estruturas.

### 3.3 Modelos de QSAR 3D

As estruturas das moléculas aqui estudadas são as mesmas estruturas otimizadas nas análises iniciais de QSAR, cujos cálculos estavam num nível satisfatório de refinamento. Para os estudos de QSAR 3D, o alinhamento estrutural é

um parâmetro fundamental e tem influência direta nos resultados dessas análises. A sobreposição de corpo rígido com um núcleo comum a todos os confôrmeros está entre as principais estratégias de alinhamento, e foi utilizada para o alinhamento do conjunto de compostos aqui estudado.<sup>91</sup>

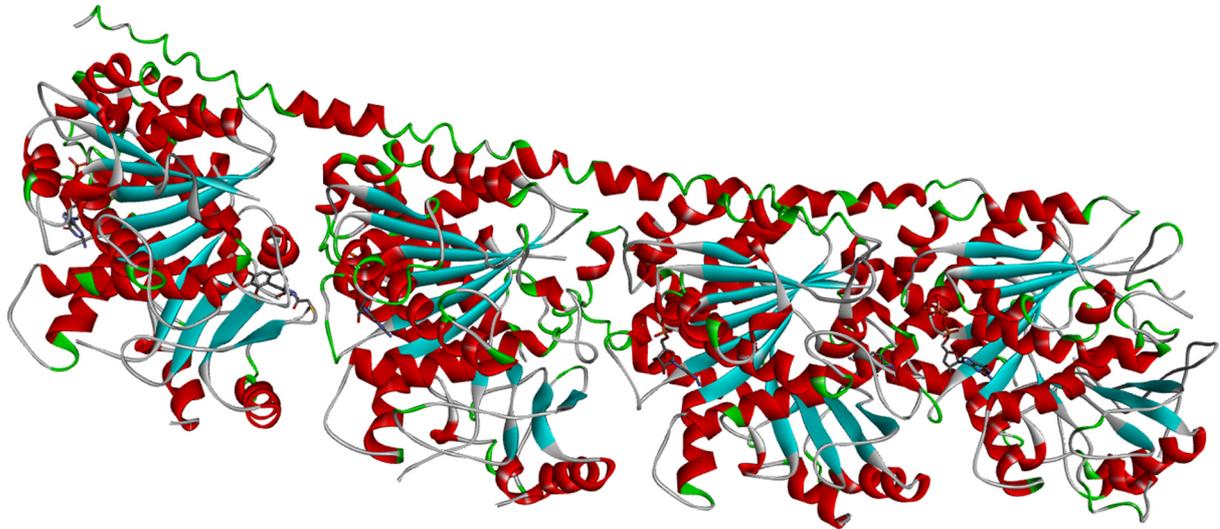
Os modelos foram gerados a partir de análises de PLS na própria plataforma do programa SYBYL-X 1.2, onde parâmetros como  $q^2$ ,  $r^2_{\text{cross}}$ ,  $r^2$  foram analisados, a fim de identificar os melhores modelos para o conjunto. Para a validação externa dos modelos gerados, também foi mantido um mínimo de 20% dos compostos no conjunto de teste. Nessas análises, um complemento à capacidade de predição da atividade biológica é a visualização de mapas de contorno 3D, que são regiões estéricas e eletrostáticas no espaço, capazes de indicar influências positivas e negativas na atividade biológica. Esses mapas de contorno podem ser bastante úteis na compreensão das interações dos compostos com a tubulina, e conseqüentemente, colaborar para a proposição de novas estruturas.<sup>92</sup>

### 3.4 Docagem molecular

As estruturas tridimensionais dos ligantes utilizadas nas análises de docagem molecular são as mesmas utilizadas nas análises anteriores, onde se mantiveram com a mesma otimização. Esses cálculos de otimização foram realizados durante os estudos de QSAR, e foram aqui oportunamente aproveitados devido ao seu refinamento. Dos ligantes apresentados na Tabela 20, foram selecionados para as simulações de docagem os ligantes que apresentavam maior representatividade, tanto nas suas estruturas como também nos valores de atividade biológica.

Em seguida, foi realizada a busca, através do banco de dados *Protein Data Bank* (PDB – [www.rcsb.org/pdb](http://www.rcsb.org/pdb)), das coordenadas tridimensionais da tubulina – código PDB 1SA0, na qual foi identificado o sítio de ligação de interesse (colchicina). Essa estrutura foi construída por difração de raios X, tendo resolução de 3,58 Å. Sua representação pode ser vista na Figura 10.<sup>93</sup>

Figura 10 – Representação estrutural da tubulina. Código PDB 1SA0.



Fonte: Elaborada pelo autor.

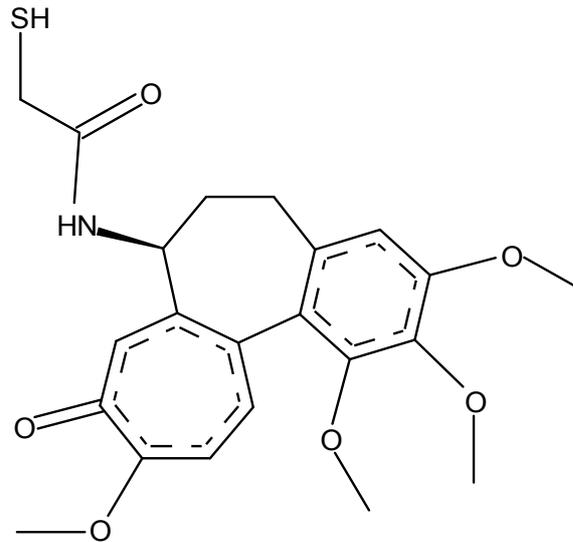
Como o sítio de ligação da colchicina, que é a região de interesse deste estudo, representa uma pequena parte de toda essa estrutura, foram excluídas do cálculo algumas regiões que não influenciavam nos resultados do mesmo.

Estando as estruturas dos ligantes e do receptor devidamente preparadas para o estudo, foram realizados os cálculos para simular a formação dos complexos. O programa utilizado para realizar tal simulação foi o GOLD 5.0<sup>94</sup>. E as representações 2D dos resultados foram geradas com o programa Discovery Studio 16.1.<sup>95</sup>

Foram testadas previamente as funções de pontuação GoldScore, ChemScore, ChemPLP e ASP, buscando verificar o desempenho de cada uma delas nas simulações de *redocking*, para então selecionar a melhor para os cálculos posteriores. O programa foi configurado para que fossem geradas 10 saídas após cada cálculo.

Foram considerados os resíduos dentro de um raio de 10 Å ao redor do ligante cristalográfico CN2, apresentado na Figura 11. Também foram adicionados átomos de hidrogênio à proteína. As moléculas de água foram mantidas nos cálculos, já que podem mediar ligações de hidrogênio entre a macromolécula e o ligante. A flexibilidade dos resíduos da proteína foi testada até o número máximo de 10, que é o máximo permitido por tal programa.<sup>96</sup>

Figura 11 – Ligante cristalográfico CN2 do sítio de ligação da colchicina, fornecido no PDB.<sup>93</sup>



Fonte: Elaborada pelo autor.

O programa GOLD é capaz de otimizar automaticamente o processo de docagem. Ele varia o número de operadores genéticos e também o número de ilhas conforme características do ligante e do sítio receptor. Quanto menor o ligante e menos flexibilidade o complexo apresentar, menos operadores serão aplicados na execução do algoritmo genético. Assim, conforme a Tabela 2, os parâmetros de evolução aplicados foram:

Tabela 2 – Parâmetros de evolução aplicados no cálculo de docagem molecular.

<b>Número de operadores</b>	Entre 10.000 e 125.000
<b>Tamanho da população</b>	100
<b>Pressão de seleção</b>	1,1
<b>Número de ilhas</b>	5
<b>Tamanho do nicho</b>	2
<b>Pesos operador de cruzamento</b>	95
<b>Peso operador de mutação</b>	95
<b>Peso operador de migração</b>	10

Os cálculos foram realizados num computador com processador core i7. Após os cálculos, foi feita a avaliação dos parâmetros obtidos, bem como a análise visual a partir da representação gráfica do complexo formado. Entre esses parâmetros, podemos citar as orientações dos ligantes, as interações desses com os resíduos de

aminoácidos da proteína, as pontuações associadas a cada orientação e também, se para um conjunto de compostos estudado, existe um modo de ligação característico. Após a análise desses parâmetros, é possível sugerir modos de ligação e as principais interações no complexo, além de prever quais complexos são os mais favoráveis.

## 4. RESULTADOS E DISCUSSÃO

### 4.1 Modelos de QSAR

Depois de reunidos os descritores eletrônicos e moleculares, somaram-se 769 descritores. Após as análises de seleção, a partir da correlação dos mesmos com a atividade biológica e do método de Algoritmo Genético, chegou-se ao número ideal de 7 descritores para compor os modelos de QSAR. No método de Algoritmo Genético, esse número foi obtido após 9.500 gerações, utilizando um critério de manter os que apresentassem correlação menor que 0,9 entre si.

Os descritores selecionados foram a energia do orbital HOMO-3 ( $E_{\text{HOMO-3}}$ ), polarizabilidade isotrópica (PI), unipolaridade (UNIP), *shape profile* nº 02 (SP02), função de distribuição radial (RDF045e, RDF155v e RDF035m) e índice total R (RTv), sendo  $E_{\text{HOMO-3}}$  e PI descritores eletrônicos obtidos a partir do software Gaussian. Os demais são moleculares, oriundos da plataforma E-Dragon. Uma tabela com os respectivos valores dos descritores para cada um dos compostos pode ser encontrada no apêndice B.

#### 4.1.1 Análises de PLS

Após estabelecidos os compostos pertencentes a cada um dos conjuntos (treinamento, validação e teste – indicados na Tabela 20), foi possível realizar os testes de validação interna para o conjunto de treinamento.

A fim de identificar compostos que pudessem acarretar em algum tipo de anomalia no modelo, foi realizado o teste de detecção de *outliers*, conforme apresentado na Figura 12. Nota-se que, mesmo que os compostos 2vv e 2dd apresentem um elevado grau de alavancagem no modelo, seus respectivos resíduos de Student encontram-se dentro do admitido. De maneira análoga, os compostos 2m e 3c apresentam resíduo de Student elevado, contudo, não exercem elevada alavancagem no modelo. Dessa forma, não houveram *outliers* no modelo gerado.

Figura 12 – Detecção de outliers.

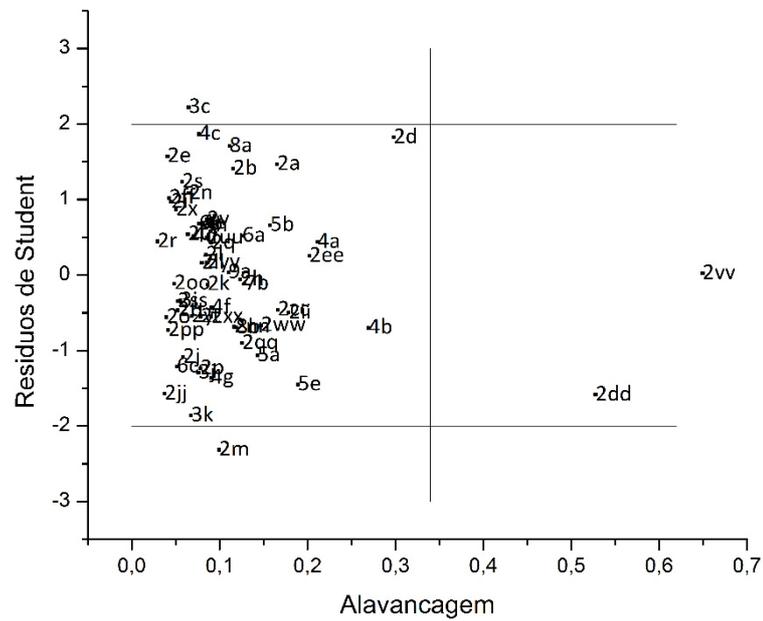
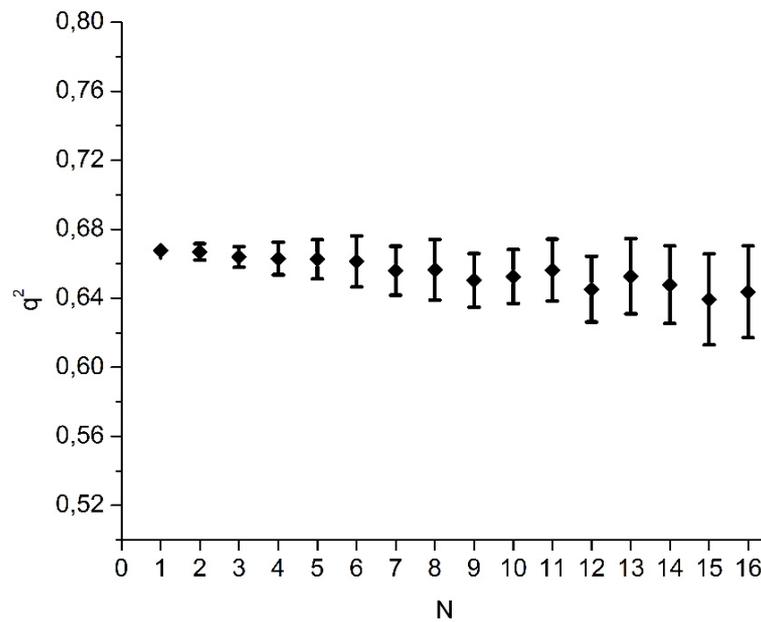
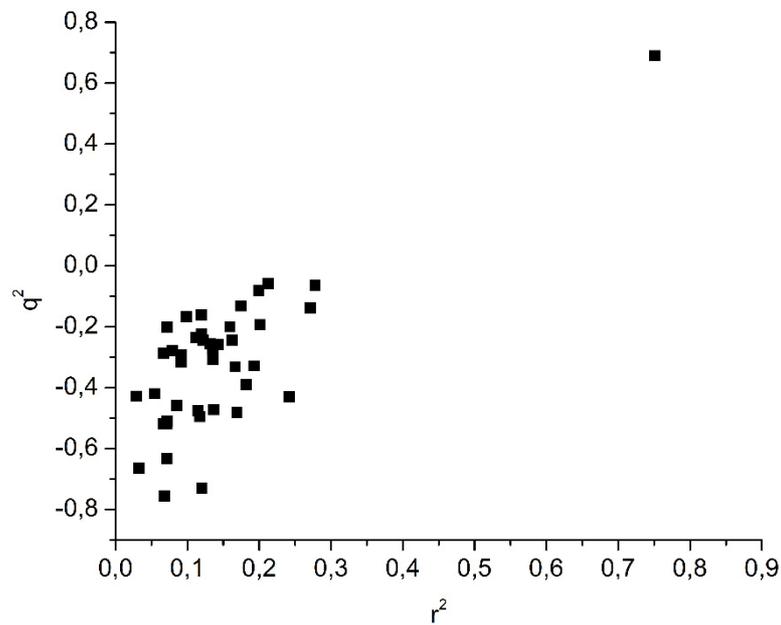


Figura 13 – Teste de *leave-N-out*.

Fonte: Elaborada pelo autor.

Conforme a Figura 14, ao ser feito o teste de randomização da atividade biológica, mostra-se que o modelo não foi gerado ao acaso. Nesse teste, foram aplicadas 40 aleatorizações e, conforme recomendações da literatura, os valores de  $q^2$  e  $r^2$  permaneceram abaixo de 0,4.<sup>97</sup>

Figura 14 – Teste de randomização da atividade biológica.



Fonte: Elaborada pelo autor.

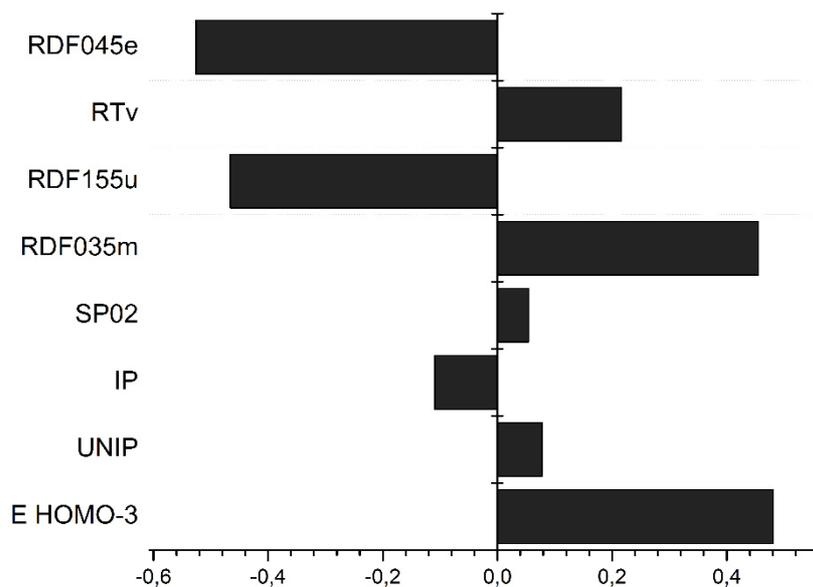
Podemos também observar, na Tabela 3, alguns parâmetros obtidos nas análises de PLS, onde o modelo com 7 variáveis latentes foi escolhido para prever a atividade biológica do conjunto de teste, onde o  $r^2_{\text{teste}}$  apresentou valor de 0,812. Como vemos, os valores de  $q^2$  e  $r^2$  do conjunto de treinamento encontram-se acima do mínimo admitido.<sup>97</sup>

Tabela 3 – Parâmetros estatísticos obtidos para o modelo de PLS.

Variáveis latentes	1	2	3	4	5	6	7	8
Variância	239,883	31,775	24,840	19,127	81,588	44,305	<b>21,947</b>	16,534
Percentual	49,976	6,620	5,175	3,985	16,998	9,230	<b>4,572</b>	3,445
Acumulativo (%)	49,976	56,595	61,771	65,755	82,753	91,983	<b>97,555</b>	100,000
SEV	0,895	0,895	0,745	0,692	0,628	0,611	<b>0,609</b>	0,609
PRESS	48,895	48,890	33,881	29,226	24,076	22,802	<b>22,645</b>	22,631
$q^2$	0,292	0,332	0,530	0,578	0,648	0,668	<b>0,670</b>	0,670
SEC	0,844	0,670	0,600	0,562	0,558	0,561	<b>0,566</b>	0,571
$r^2$	0,384	0,618	0,698	0,740	0,749	0,750	<b>0,751</b>	0,751

As contribuições de cada um dos descritores para geração do modelo de PLS estão representadas na Figura 15.

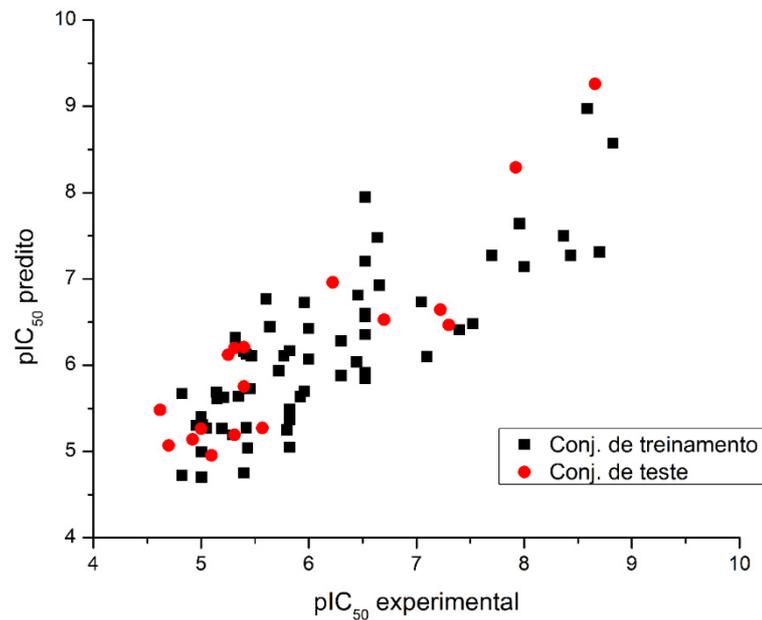
Figura 15. Contribuição de cada descritor para o vetor de regressão do modelo PLS.



Fonte: Elaborada pelo autor.

Na Figura 16, estão apresentados os valores de predição dos conjuntos de treinamento e teste obtidos pelo método de PLS.

Figura 16 – Predição da atividade biológica do modelo de PLS.



Fonte: Elaborada pelo autor.

#### 4.1.2 Análises de ANN

Adotando uma nova metodologia no emprego das Redes Neurais Artificiais, 180 variações foram testadas. Para cada uma dessas variações, diversos modelos eram gerados, dos quais, selecionava-se o melhor representante.

Ao final, haviam 180 modelos, um de cada variação. Desse total, 10% foi selecionado para compor o modelo *Consensus* (ANN-CM), os quais encontram-se na Tabela 4. Nota-se que a maioria das funções de transferência que compõem esse modelo são do tipo tansig-tansig, e que quanto melhores são os valores de  $r^2_{\text{teste}}$  e EQM (teste) de um modelo, maior é sua influência (FN) no modelo *Consensus*.

Tabela 4 – Parâmetros estatísticos dos 18 melhores modelos de MLP-ANN, e valores de  $r^2_{\text{teste}}$  e EQM do ANN-CM, gerados a partir do fator de normalização (FN).

Função de transferência	Neurônios	$r^2_{\text{teste}}$	EQM (conj. teste)	FN para ANN-CM
Tansig-Tansig	1	0,805	0,323	0,054
Tansig-Tansig	2	0,812	0,322	0,054
Tansig-Tansig	3	0,821	0,301	0,058
Tansig-Tansig	4	0,820	0,309	0,056
Tansig-Tansig	6	0,822	0,310	0,056
Tansig-Tansig	9	0,818	0,302	0,058
Tansig-Tansig	10	0,843	0,276	0,063
Tansig-Tansig	13	0,815	0,312	0,056
Tansig-Tansig	14	0,823	0,298	0,059
Tansig-Purelin	1	0,804	0,316	0,055
Purelin-Tansig	10	0,809	0,327	0,053
Purelin-Tansig	12	0,809	0,327	0,053
Purelin-Tansig	20	0,783	0,324	0,054
Purelin-Purelin	1	0,812	0,325	0,054
Purelin-Purelin	2	0,812	0,325	0,054
Purelin-Purelin	3	0,812	0,325	0,054
Purelin-Purelin	4	0,812	0,325	0,054
Logsig-Tansig	1	0,808	0,318	0,055
<b>ANN-CM</b>		<b>0,829</b>	<b>0,286</b>	

Podemos então, comparar os valores  $r^2_{\text{teste}}$  e EQM dos métodos de PLS e ANN-CM. Na Tabela 5, nota-se que os valores do ANN-CM são ligeiramente superiores aos do PLS.

Tabela 5 – Valores estatísticos de predição dos modelos PLS e ANN-CM.

Modelos	$r^2_{\text{teste}}$	EQM (conj. teste)
PLS	0,812	0,327
ANN-CM	0,829	0,286

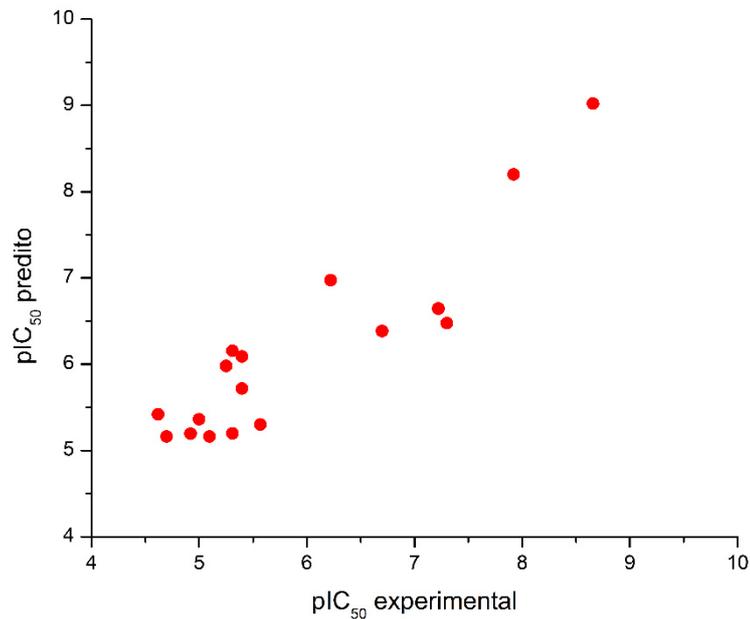
Na Tabela 6, temos a comparação direta da capacidade preditiva dos dois modelos. O valor médio dos resíduos também foi ligeiramente menor para o ANN-CM (0,473) em relação ao PLS (0,500).

Tabela 6 – Predição da atividade biológica dos modelos PLS e ANN-CM.

Comp.	PLS			ANN-CM	
	plC <sub>50</sub> Experimental	plC <sub>50</sub> Predito	Resíduo	plC <sub>50</sub> Predito	Resíduo
<b>2aa</b>	<b>5,569</b>	5,273	0,296	5,301	0,268
<b>2α</b>	<b>4,699</b>	5,070	-0,371	5,164	-0,465
<b>2β</b>	<b>4,620</b>	5,482	-0,863	5,419	-0,799
<b>2f</b>	<b>6,699</b>	6,527	0,172	6,384	0,315
<b>2gg</b>	<b>5,398</b>	6,209	-0,811	6,090	-0,692
<b>2kk</b>	<b>5,310</b>	5,192	0,118	5,199	0,111
<b>2mm</b>	<b>5,252</b>	6,122	-0,870	5,980	-0,728
<b>2rr</b>	<b>5,097</b>	4,955	0,142	5,164	-0,067
<b>3b</b>	<b>8,658</b>	9,261	-0,603	9,021	-0,363
<b>3e</b>	<b>7,921</b>	8,292	-0,372	8,201	-0,281
<b>3g</b>	<b>7,301</b>	6,466	0,835	6,476	0,825
<b>3h</b>	<b>7,222</b>	6,645	0,577	6,645	0,576
<b>5c</b>	<b>5,310</b>	6,198	-0,888	6,156	-0,846
<b>5d</b>	<b>4,921</b>	5,142	-0,221	5,195	-0,274
<b>7a</b>	<b>6,222</b>	6,960	-0,738	6,974	-0,752
<b>7c</b>	<b>5,398</b>	5,755	-0,357	5,718	-0,320
<b>8c</b>	<b>5,000</b>	5,264	-0,264	5,362	-0,362

Na Figura 17, temos a representação visual da predição do conjunto de teste feita com o modelo *Consensus* ANN.

Figura 17 – Predição da atividade biológica do modelo *Consensus* ANN – conjunto de teste.



Fonte: Elaborada pelo autor.

Também vale comentar que os resultados obtidos a partir dessa metodologia de ANN são consideravelmente superiores aos obtidos na pesquisa de mestrado. Comparando diretamente a metodologia de ANN desenvolvida na pesquisa de mestrado com a metodologia aqui apresentada, podemos notar, na Tabela 7, tal diferença:

Tabela 7 – Diferença entre os valores de  $r^2_{\text{teste}}$  e resíduo médio das metodologias aplicadas na pesquisa anterior e na ANN-CM.

	$r^2_{\text{teste}}$	Resíduo médio
<b>ANN (anterior)</b>	0,633	0,687
<b>ANN-CM</b>	0,829	0,473

Assim, a elaboração de um modelo *Consensus* ANN nos mostra que uma maior exploração dos fatores envolvidos nesse tipo de cálculo (funções de transferência e neurônios) fornece maior segurança nas etapas de seleção, já que se tem um melhor panorama sobre as potencialidades de tal método.

## 4.2 Análises de HQSAR

Com as combinações de distinção dos fragmentos, foram gerados 15 modelos a partir do tamanho de fragmento (4-7), os quais estão apresentados na Tabela 8, juntamente com os parâmetros estatísticos do PLS e comprimento do holograma. Como pode ser visto, o modelo que apresentou os melhores resultados para esse tamanho de fragmento foi o **Modelo 10** (A.B.H.DA), possuindo os maiores valores de  $q^2$  e  $r^2$ , e o menor valor de Erro Padrão Estimado (SEE).

Tabela 8 – Resultado das combinações de distinção de fragmentos quando utilizado o tamanho padrão de fragmento (4-7). N é o número de componentes associadas ao melhor modelo, e HL é o comprimento do holograma capaz de gerar o melhor modelo HQSAR.

Modelo	Distinção de fragmento	Parâmetros estatísticos				
		$q^2$	$r^2$	SEE	N	HL
1	A.B	0,637	0,800	0,538	5	353
2	A.B.C	0,605	0,781	0,562	5	199
3	A.B.C.H	0,629	0,886	0,409	6	401
4	A.B.C.H.Ch	0,633	0,892	0,398	6	401
5	A.B.H	0,631	0,878	0,424	6	353
6	A.B.C.Ch	0,613	0,851	0,468	6	199
7	A.B.C.Ch.DA	0,598	0,855	0,462	6	199
8	A.B.DA	0,703	0,876	0,427	6	257
9	A.B.C.DA	0,594	0,853	0,464	6	199
<b>10</b>	<b>A.B.H.DA</b>	<b>0,756</b>	<b>0,901</b>	<b>0,382</b>	<b>6</b>	<b>97</b>
11	A.B.C.H.DA	0,655	0,885	0,411	6	257
12	A.B.Ch.DA	0,703	0,869	0,438	6	257
13	A.B.Ch	0,664	0,850	0,469	6	307
14	A.B.H.Ch	0,604	0,872	0,433	6	353
15	A.B.C.H.Ch.DA	0,663	0,894	0,394	6	401

Após ser estabelecida a melhor combinação dos parâmetros relacionados à geração dos hologramas, foi testado o melhor tamanho do fragmento para o **Modelo 10**, cujas variações foram de 2-5, 3-6, 4-7, 5-8, 6-9, 7-10, 8-11, 9-12, 10-13. Conforme pode ser observado da Tabela 9, o modelo que apresentou os melhores resultados

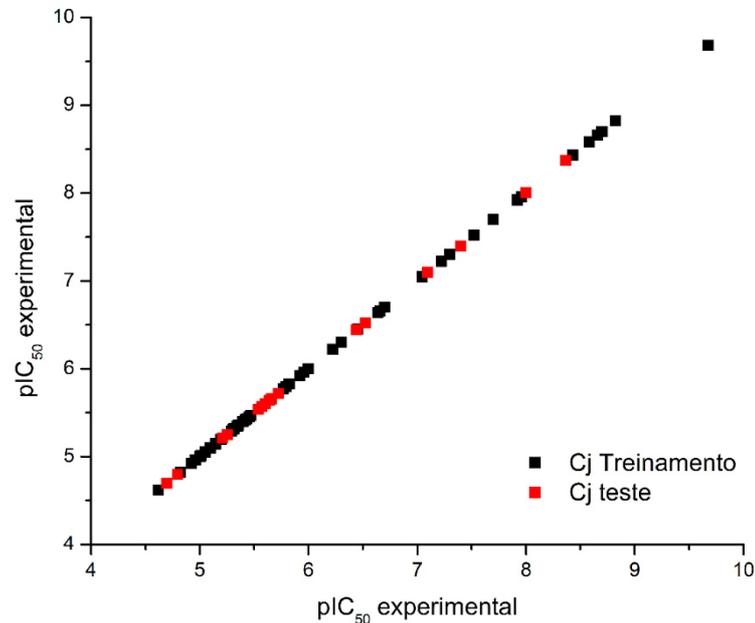
para os parâmetros analisados foi o modelo com o tamanho de fragmento (8-11), com  $q^2 = 0,772$ ,  $r^2 = 0,920$  e  $SEE = 0,343$ .

Tabela 9 – Resultados das análises da influência do tamanho dos fragmentos a partir da melhor combinação de fragmentos – A.B.H.DA.

Tamanho dos fragmentos	Parâmetros estatísticos				
	$q^2$	$r^2$	SEE	N	HL
2-5	0,583	0,846	0,475	6	257
3-6	0,729	0,889	0,404	6	97
4-7	0,756	0,901	0,382	6	97
5-8	0,704	0,902	0,380	6	401
6-9	0,702	0,908	0,368	6	151
7-10	0,696	0,916	0,351	6	307
<b>8-11</b>	<b>0,772</b>	<b>0,920</b>	<b>0,343</b>	<b>6</b>	<b>257</b>
9-12	0,737	0,903	0,378	6	257
10-13	0,647	0,895	0,392	6	307

O melhor modelo foi gerado com 70 moléculas no conjunto de treinamento (80%) e 17 moléculas do conjunto de teste (20%). Esses compostos foram separados seguindo o critério de separar o mesmo percentual de compostos em cada faixa de atividade biológica. Nesse caso, para os valores de  $pIC_{50}$  de 4,0 a 9,0, a cada variação de uma unidade de  $pIC_{50}$ , eram selecionados aproximadamente 20% dos compostos para o conjunto de teste, garantindo assim que não fossem selecionados compostos concentrados em apenas uma faixa de atividade biológica, o que pode ser verificado na Figura 18.

Figura 18 – Verificação da distribuição dos valores de atividade biológica na separação dos conjuntos de treinamento e de teste para o modelo de HQSAR.



Fonte: Elaborada pelo autor.

A partir do modelo escolhido, pôde ser testada a capacidade preditiva do mesmo via validação externa, ou seja, prever a atividade biológica de compostos que não fizeram parte do conjunto de treinamento. Tais compostos encontram-se destacados por um asterisco (\*), juntamente com seus valores, na Tabela 10. Da mesma forma, tal modelo é capaz de prever a atividade biológica de moléculas análogas que venham a ser propostas, o que poderá ser verificado posteriormente.

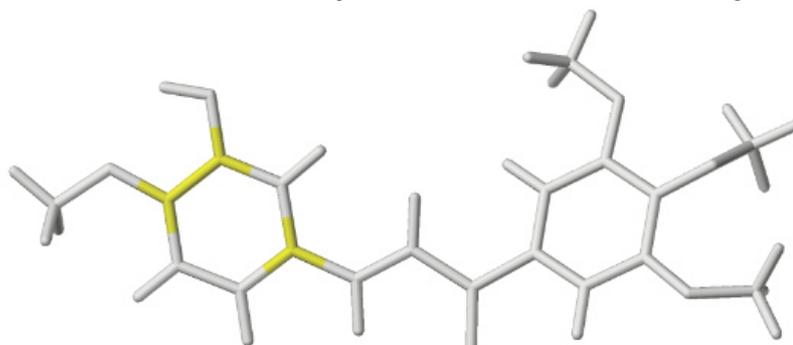
Nas Figuras 19, 20 e 21 são apresentados os mapas de contribuições individuais 2D para os compostos 2a, 3a e 4a, a partir do modelo HQSAR escolhido. Tais mapas podem fornecer indicações sobre os fragmentos moleculares diretamente relacionados à atividade biológica se realizada uma interpretação dos fragmentos estruturais incorporados no modelo. As cores vermelho/laranja indicam contribuições negativas, enquanto as cores verde/amarelo indicam contribuições favoráveis à atividade biológica. As cores em branco representam contribuições intermediárias.

Tabela 10 – Predição dos valores da atividade biológica (pIC<sub>50</sub>) para os conjuntos de treinamento e de teste a partir das análises de HQSAR.

	<b>Exper.</b>	<b>HQSAR</b>						
			<b>2nn</b>	5,051	5,250	<b>3h</b>	7,222	7,824
<b>2a*</b>	8,367	6,703	<b>2o</b>	6,523	5,971	<b>3i*</b>	5,721	7,046
<b>2aa*</b>	6,523	5,321	<b>2oo</b>	5,824	5,301	<b>3j*</b>	5,638	6,734
<b>2α*</b>	5,658	5,616	<b>2p</b>	5,398	5,558	<b>3k*</b>	5,602	8,272
<b>2b*</b>	5,569	6,657	<b>2pp</b>	5,013	5,210	<b>4a</b>	8,824	9,364
<b>2bb*</b>	5,252	5,588	<b>2q</b>	6,523	5,871	<b>4b</b>	8,585	8,902
<b>2β*</b>	4,796	5,708	<b>2qq</b>	5,824	5,961	<b>4c</b>	8,432	8,278
<b>2c*</b>	4,699	7,075	<b>2r</b>	5,357	6,113	<b>4d</b>	7,959	7,673
<b>2cc*</b>	6,523	6,072	<b>2rr</b>	5,004	5,226	<b>4e</b>	7,699	7,406
<b>2d*</b>	8,000	6,174	<b>2s</b>	6,523	6,371	<b>4f</b>	6,658	6,739
<b>2dd*</b>	5,208	5,731	<b>2ss</b>	5,824	5,874	<b>4g</b>	6,638	7,004
<b>2e*</b>	5,538	5,680	<b>2t</b>	5,347	5,940	<b>4h*</b>	6,444	7,590
<b>2ee</b>	6,456	5,918	<b>2tt</b>	5,004	5,111	<b>5a</b>	5,469	5,606
<b>2f</b>	4,620	4,450	<b>2u</b>	6,523	5,780	<b>5b</b>	5,432	5,338
<b>2ff</b>	5,194	5,711	<b>2uu</b>	5,796	5,654	<b>5c</b>	5,310	5,301
<b>2g</b>	8,000	5,715	<b>2v</b>	5,319	5,458	<b>5d</b>	4,921	4,965
<b>2gg</b>	5,959	5,845	<b>2vv</b>	5,000	5,501	<b>5e</b>	4,824	4,851
<b>2h</b>	5,456	5,343	<b>2w</b>	6,523	5,996	<b>6a</b>	7,046	6,865
<b>2hh</b>	5,149	5,417	<b>2ww</b>	5,796	5,552	<b>6b</b>	6,301	6,078
<b>2i*</b>	7,398	5,782	<b>2x</b>	5,310	5,958	<b>6c</b>	5,398	5,316
<b>2ii</b>	5,959	5,809	<b>2xx</b>	4,959	5,276	<b>7a</b>	6,222	6,076
<b>2j</b>	5,420	5,439	<b>2y</b>	6,523	6,469	<b>7b</b>	6,000	6,041
<b>2jj</b>	5,143	5,439	<b>2yy</b>	5,770	5,654	<b>7c</b>	5,398	5,278
<b>2k*</b>	7,097	4,233	<b>2z</b>	5,292	5,446	<b>8a</b>	7,523	7,308
<b>2kk</b>	5,921	5,717	<b>2zz</b>	4,824	5,022	<b>8b</b>	6,000	5,855
<b>2l</b>	5,420	5,269	<b>3a</b>	9,678	9,042	<b>8c</b>	5,000	5,190
<b>2ll</b>	5,097	5,361	<b>3b</b>	8,658	8,952	<b>8d</b>	4,921	4,880
<b>2m</b>	6,699	6,310	<b>3c</b>	8,699	8,052	<b>9a</b>	6,301	6,271
<b>2mm</b>	5,824	5,821	<b>3e</b>	7,921	8,119			
<b>2n</b>	5,398	5,456	<b>3g</b>	7,301	7,496			

Para a estrutura 2a, a região indicada com contribuição positiva à atividade biológica corresponde a parte do anel esquerdo da molécula, onde pode ocorrer interações hidrofóbicas com o receptor.

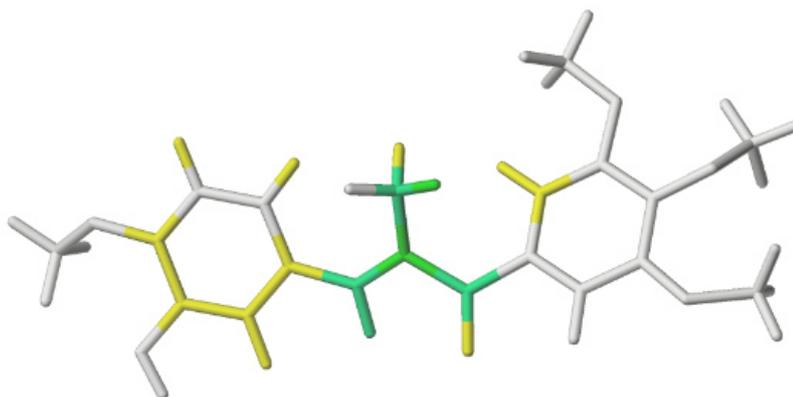
Figura 19 – Mapa de contribuições individuais da molécula 2a para o modelo HQSAR. A cor amarela indica contribuição favorável à atividade biológica.



Fonte: Elaborada pelo autor.

Para a estrutura 3a, as indicações de contribuição positiva à atividade biológica estão principalmente na região central, abrangendo também parte do anel esquerdo. Na região central, destaca-se um grupo metila, além de um átomo de oxigênio.

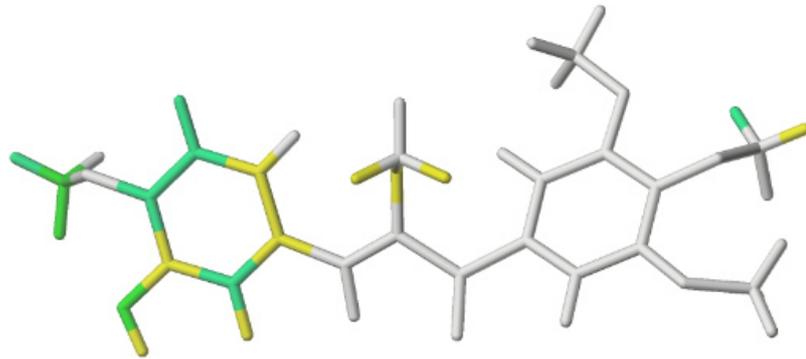
Figura 20 – Mapa de contribuições individuais da molécula 3a para o modelo HQSAR. As cores verde e amarelo indicam contribuições favoráveis à atividade biológica.



Fonte: Elaborada pelo autor.

Na estrutura 4a, o lado esquerdo da molécula é indicado quase que integralmente com contribuições positivas à atividade biológica, onde se localizam grupos hidroxila e metoxila, além de um grupo metoxila na região central da molécula.

Figura 21 – Mapa de contribuições individuais da molécula 4a para o modelo HQSAR. As cores verde e amarelo indicam contribuições favoráveis à atividade biológica.



Fonte: Elaborada pelo autor.

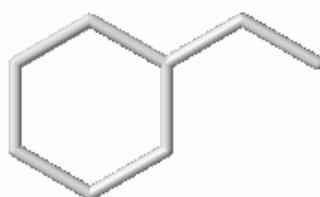
Pode-se perceber, a partir das estruturas selecionadas para fins de comparação (2a, 3a e 4a), que o modelo HQSAR gerado contempla muito pouco (ou quase nada) a extremidade direita das estruturas. Porém, a partir de estudos preliminares de interação do complexo, essa extremidade (de compostos semelhantes) participa ativamente de diversos tipos de interações que ocorrem no complexo formado no sítio de ligação da colchicina. Deve-se lembrar que, devido ao fato de as análises de HQSAR não levarem em consideração a estrutura do receptor, as indicações de regiões favoráveis podem vir a ser insuficientes para uma completa compreensão do complexo e para a proposição de novas estruturas. Mas, ainda assim, indicam regiões verdadeiramente favoráveis à atividade biológica, tornando bastante útil o modelo gerado por esse método, para fins de predição da atividade biológica de compostos análogos.

### 4.3 Análises de QSAR-3D

Para esse estudo, buscou-se, inicialmente, utilizar duas abordagens de alinhamento: corpo rígido e docagem molecular. Contudo, durante as análises de docagem, foi percebido que as mesmas não apresentavam uma uniformidade do padrão de comportamento: seus posicionamentos eventualmente se apresentavam invertidos, conforme variava-se o perfil das estruturas analisadas (fato que poderá ser compreendido depois de apresentadas as análises de docagem molecular). Devido a

esse comportamento, optou-se por não basear esse alinhamento de docagem nos estudos de QSAR-3D. Portanto, o alinhamento das moléculas foi baseado na sobreposição de corpo rígido da subestrutura comum entre todos os confôrmeros de menor energia. Na Figura 22 está apresentado o núcleo comum a todas as estruturas, definido para o alinhamento. A molécula de referência escolhida para esse alinhamento foi a 3a, sendo essa a de maior valor de atividade biológica dentre todas as analisadas.

Figura 22 – Subestrutura comum a todos os compostos, selecionada para o alinhamento dos mesmos nas análises de QSAR 3D.

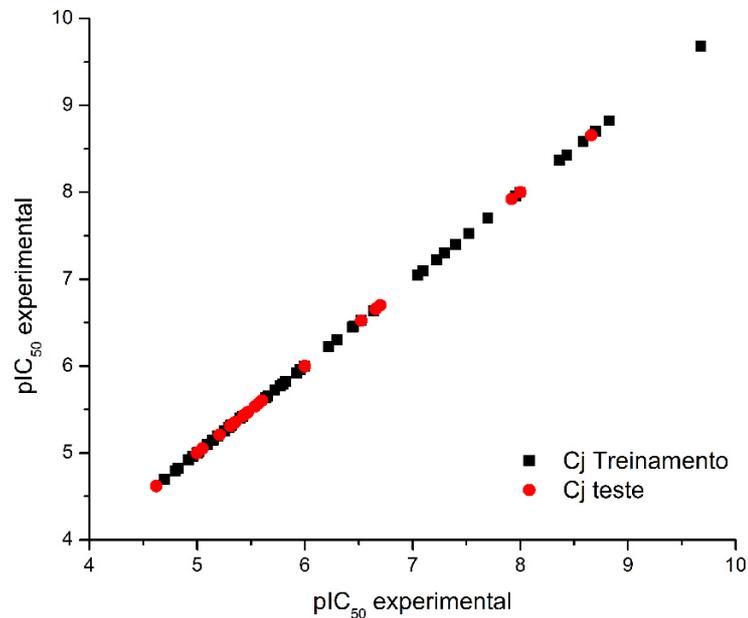


Fonte: Elaborada pelo autor.

O alinhamento das análises de QSAR 3D é útil tanto para as análises de CoMFA quanto para as de CoMSIA. E a separação dos conjuntos de treinamento e teste foi a mesma para ambas as análises. Dos 87 compostos, foram separados 67 para o conjunto de treinamento (77%) e 20 para o conjunto de teste (23%). Os compostos do conjunto de teste encontram-se destacados com um asterisco (\*) na Tabela 12. Esses compostos também foram separados seguindo o critério de retirar o mesmo percentual de compostos em cada faixa de atividade biológica. Assim, a partir de  $pIC_{50} = 4,0$ , a cada variação de uma unidade de  $pIC_{50}$ , eram selecionados aproximadamente 23% dos compostos para o conjunto de teste, conforme pode ser observado na Figura 23.

Nas análises de CoMFA, o modelo foi gerado a partir dos campos estérico e eletrostático. Nas análises de CoMSIA, o melhor modelo foi gerado quando considerados os campos estérico, eletrostático, hidrofóbico e doador/aceptor de ligação de Hidrogênio.

Figura 23 – Verificação da distribuição dos valores de atividade biológica na separação dos conjuntos de treinamento e de teste para os modelos de QSAR 3D.



Fonte: Elaborada pelo autor.

Os parâmetros obtidos nas análises de PLS dos modelos CoMFA e CoMSIA estão apresentados na Tabela 11. Como pode-se perceber, o modelo CoMFA apresentou valores de parâmetros ligeiramente melhores que o modelo CoMSIA.

Tabela 11 – Resultados dos parâmetros estatísticos dos melhores modelos de QSAR 3D (CoMFA e CoMSIA).

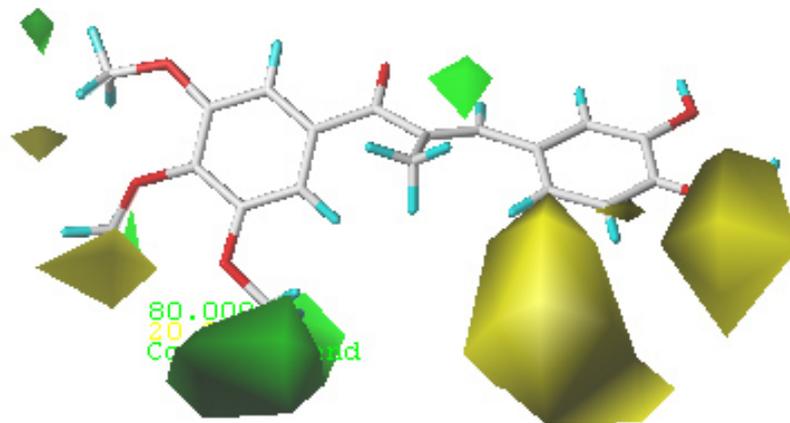
Parâmetros	CoMFA	CoMSIA
$r^2$	0,902	0,869
$q^2$	0,610	0,588
$r^2_{cross}$	0,617	0,569
<b>F</b>	91,805	66,570

Os mapas de contorno estérico e eletrostático para o modelo CoMFA encontram-se nas Figuras 24 e 25, respectivamente. Nas Figuras 26 a 29 estão apresentados os mapas de contorno gerados pelo modelo CoMSIA.

#### 4.3.1 Modelos de CoMFA

**Estéricos:** Para a molécula 3a, os campos verdes do mapa de contorno correspondem a grupos metila e metoxila na região central e no lado esquerdo da molécula. Para o modelo gerado, os grupos volumosos próximos a esses campos verdes estão associados a um aumento da atividade biológica. Já as regiões amarelas correspondem, em geral, a átomos de hidrogênio que, quando próximos aos campos amarelos, representam também um aumento da atividade biológica, por serem pouco volumosos.

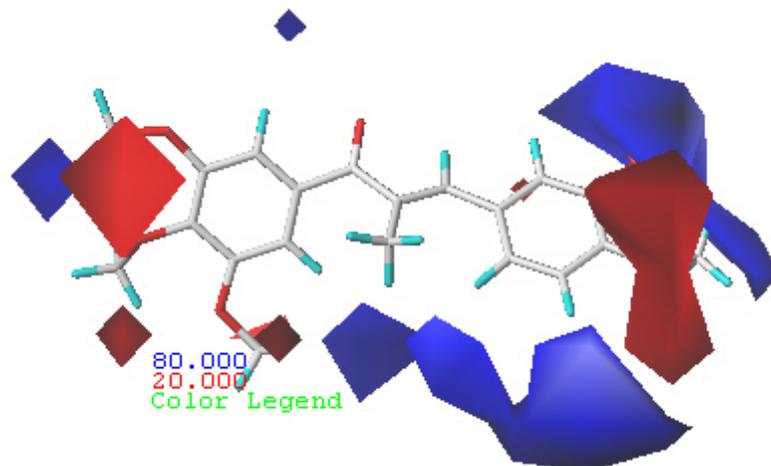
Figura 24 – Mapas de contorno estéricos para a molécula 3a, gerados a partir da plataforma CoMFA para o modelo de QSAR 3D.



Fonte: Elaborada pelo autor.

**Eletrostáticos:** Os valores de atividade biológica são maiores quando há mais carga positiva próxima aos campos azuis e mais carga negativa próxima aos campos vermelhos. Para a estrutura 3a, os mapas destacam a importância da carga negativa em regiões onde se fazem presentes grupos hidroxila e metoxila, da mesma forma que os radicais alquila são indicados como importantes representantes para a composição de uma região com predominância de carga positiva.

Figura 25 – Mapas de contorno eletrostáticos para a molécula 3a, gerados a partir da plataforma CoMFA para o modelo de QSAR 3D.

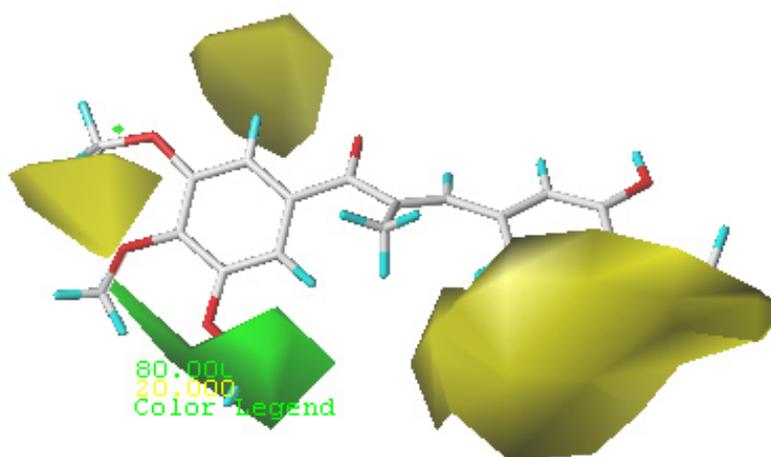


Fonte: Elaborada pelo autor.

#### 4.3.2 Modelos de CoMSIA

**Estéricos:** Ainda tendo como exemplo, a molécula 3a, as regiões em amarelo indicam as regiões onde não é interessante a presença de substituintes volumosos. Análogo aos mapas de contorno do modelo CoMFA, essas regiões são, em sua maioria, radicais alquila.

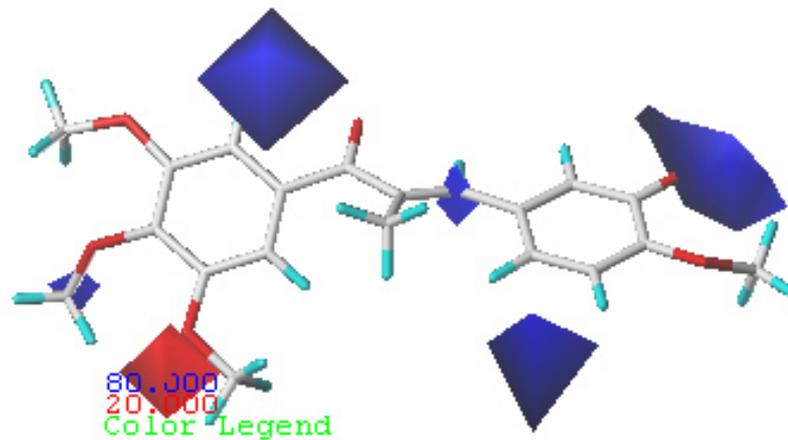
Figura 26 – Mapas de contorno estéricos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D.



Fonte: Elaborada pelo autor.

**Eletrostáticos:** A predominância das regiões em azul indica de carga positiva próxima a essas regiões favorece o aumento da atividade biológica, onde radicais alquila se fazem presentes.

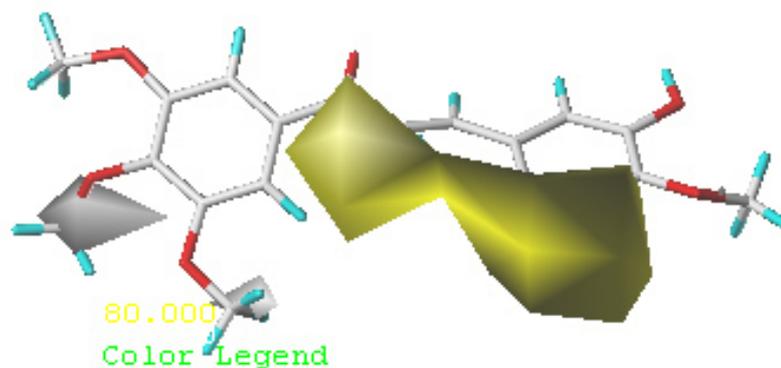
Figura 27 – Mapas de contorno eletrostáticos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D.



Fonte: Elaborada pelo autor.

**Hidrofóbicos:** Regiões em amarelo favoráveis para substituintes hidrofóbicos, e regiões em branco desfavoráveis para substituintes hidrofóbicos. Na maior parte das regiões em amarelo, encontram-se os grupos alquila, ao passo que nas regiões em branco encontra-se o grupo metoxila.

Figura 28 – Mapas de contorno hidrofóbicos para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D.

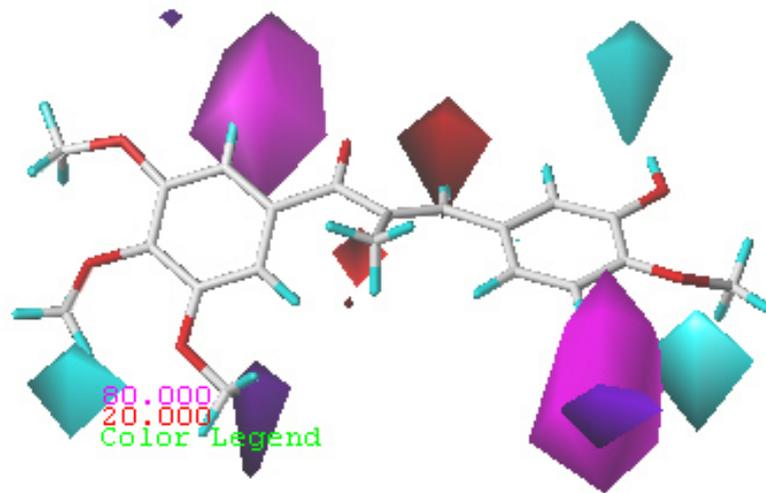


Fonte: Elaborada pelo autor.

**Doadores e aceptores de ligação de hidrogênio:** As regiões favoráveis para doadores de ligação de hidrogênio encontram-se em azul ciano, e as desfavoráveis para doadores, em roxo. Nas regiões em azul ciano, encontram-se os grupos metoxila e hidroxila. Nas regiões em roxo, encontram-se principalmente grupos alquila.

As regiões favoráveis para aceptores de ligação de hidrogênio encontram-se em coloração magenta, e as desfavoráveis para aceptores, em vermelho. Nas regiões em magenta, temos também os radicais alquila; e nas regiões em vermelho, temos os grupos alquila localizados na região central da molécula.

Figura 29 – Mapas de contorno de doadores e aceptores de ligação de hidrogênio para a molécula 3a, gerados a partir da plataforma CoMSIA para o modelo de QSAR 3D.



Fonte: Elaborada pelo autor.

Analisando os mapas de contorno dos dois modelos gerados, pode-se perceber que o modelo CoMFA, além de fornecer os melhores valores de parâmetros estatísticos com relação ao modelo CoMSIA, também gerou mapas de contorno mais condizentes com as características dos compostos. Por exemplo, para os campos eletrostáticos, o modelo CoMFA foi capaz de destacar melhor as regiões favoráveis a substituintes eletronegativos.

Os valores de predição da atividade biológica ( $pIC_{50}$ ), tanto para o modelo CoMFA quanto para o CoMSIA, encontram-se na Tabela 12.

Tabela 12 – Predição dos valores da atividade biológica (pIC<sub>50</sub>) para os conjuntos de treinamento e de teste a partir das análises de QSAR 3D (CoMFA e CoMSIA).

	<b>EXPERIM.</b>	<b>CoMFA</b>	<b>CoMSIA</b>				
				<b>2q</b>	6,523	6,074	5,972
<b>2a</b>	8,367	7,887	7,831	<b>2qq</b>	5,824	6,167	5,845
<b>2aa</b>	6,523	6,137	6,206	<b>2r*</b>	5,357	7,523	7,271
<b>2α</b>	5,658	5,763	5,741	<b>2rr</b>	5,004	5,262	5,469
<b>2b*</b>	5,569	7,902	7,341	<b>2s</b>	6,523	5,738	5,955
<b>2bb</b>	5,252	5,469	5,413	<b>2ss</b>	5,824	5,483	5,504
<b>2β</b>	4,796	5,397	5,152	<b>2t*</b>	5,347	7,800	7,303
<b>2c</b>	4,699	5,400	5,419	<b>2tt</b>	5,004	5,117	5,277
<b>2cc</b>	6,523	6,611	6,320	<b>2u</b>	6,523	7,728	7,564
<b>2d*</b>	8,000	6,174	6,194	<b>2uu</b>	5,796	5,792	5,752
<b>2dd*</b>	5,208	8,462	7,991	<b>2v</b>	5,319	5,445	5,463
<b>2e*</b>	5,538	7,466	7,173	<b>2vv</b>	5,000	4,962	5,279
<b>2ee</b>	6,456	5,648	5,763	<b>2w</b>	6,523	6,589	6,516
<b>2f*</b>	4,620	8,841	7,991	<b>2ww</b>	5,796	6,327	6,504
<b>2ff</b>	5,194	5,069	5,416	<b>2x</b>	5,310	5,653	5,305
<b>2g</b>	8,000	7,327	7,318	<b>2xx</b>	4,959	5,404	5,551
<b>2gg</b>	5,959	5,714	5,457	<b>2y*</b>	6,523	5,374	4,828
<b>2h*</b>	5,456	7,580	7,257	<b>2yy</b>	5,770	5,792	5,752
<b>2hh</b>	5,149	5,180	4,955	<b>2z</b>	5,292	5,416	6,008
<b>2i</b>	7,398	7,712	7,488	<b>2zz</b>	4,824	5,088	4,549
<b>2ii</b>	5,959	5,613	5,379	<b>3a</b>	9,678	9,741	9,904
<b>2j</b>	5,420	5,216	5,409	<b>3b*</b>	8,658	7,123	6,901
<b>2jj</b>	5,143	5,216	5,409	<b>3c</b>	8,699	8,146	7,981
<b>2k</b>	7,097	7,015	6,671	<b>3e*</b>	7,921	6,526	6,335
<b>2kk</b>	5,921	5,725	5,586	<b>3g</b>	7,301	7,614	7,753
<b>2l</b>	5,420	5,587	5,372	<b>3h</b>	7,222	7,461	7,644
<b>2ll</b>	5,097	5,337	5,254	<b>3i</b>	5,721	6,173	6,712
<b>2m*</b>	6,699	5,888	5,838	<b>3j</b>	5,638	5,662	5,587
<b>2mm</b>	5,824	5,388	5,376	<b>3k*</b>	5,602	7,215	7,579
<b>2n</b>	5,398	5,600	5,626	<b>4a</b>	8,824	8,684	8,444
<b>2nn*</b>	5,051	7,228	7,314	<b>4b</b>	8,585	8,551	8,767
<b>2o</b>	6,523	5,719	5,577	<b>4c</b>	8,432	7,929	7,634
<b>2oo</b>	5,824	5,661	5,561	<b>4d</b>	7,959	7,808	7,659
<b>2p</b>	5,398	5,575	5,629	<b>4e</b>	7,699	7,699	7,672
<b>2pp</b>	5,013	4,866	5,432	<b>4f*</b>	6,658	7,551	7,635

<b>4g</b>	6,638	7,075	7,446	<b>6c</b>	5,398	5,691	5,406
<b>4h</b>	6,444	6,338	6,960	<b>7a</b>	6,222	5,850	6,369
<b>5a*</b>	5,469	7,145	7,044	<b>7b</b>	6,000	5,890	5,933
<b>5b*</b>	5,432	6,474	6,409	<b>7c</b>	5,398	5,150	5,647
<b>5c*</b>	5,310	7,596	7,134	<b>8a</b>	7,523	7,752	7,380
<b>5d</b>	4,921	4,888	4,783	<b>8b*</b>	6,000	7,463	6,795
<b>5e</b>	4,824	4,683	4,758	<b>8c*</b>	5,000	7,614	7,598
<b>6a</b>	7,046	6,879	7,122	<b>8d</b>	4,921	4,973	5,380
<b>6b</b>	6,301	6,510	6,053	<b>9a</b>	6,301	6,508	6,042

#### 4.4 Análises de docagem molecular

A partir da estrutura cristalográfica integral da tubulina, extraída do PDB (código 1SA0 - Figura 10), e visando diminuir o tempo de cálculo, foram mantidas apenas as partes dessa estrutura que têm uma relação direta com o ligante no cálculo de docagem molecular.

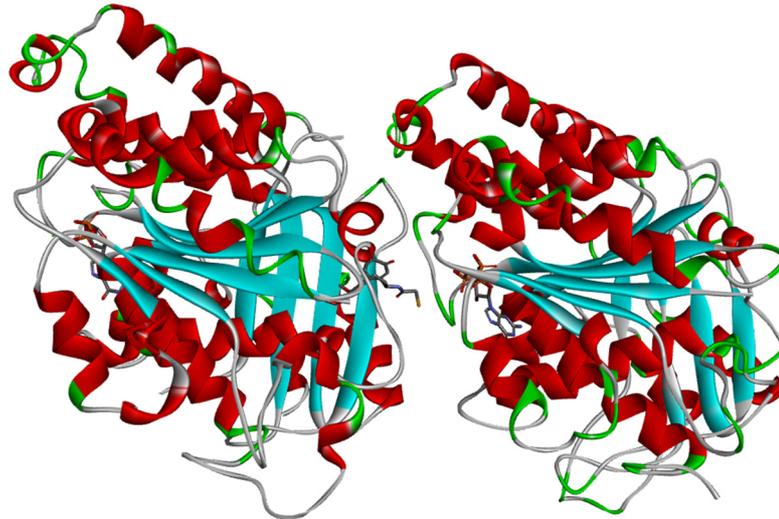
A fim de garantir que a melhor descrição possível do sistema seja feita, foram realizados cálculos de *redocking* com a  $\beta$ -tubulina, estrutura na qual se encontra o sítio de ligação da colchicina, e também com a estrutura dimérica  $\alpha\beta$ -tubulina, já que a  $\alpha$ -tubulina se encontra bastante próxima do sítio de ligação da colchicina. Na Tabela 13, podem ser observados os valores de pontuação e RMSD para os cálculos envolvendo apenas a  $\beta$ -tubulina e os cálculos com a  $\alpha\beta$ -tubulina.

Tabela 13 – Comparação entre as funções de pontuação e RMSD testadas para a  $\beta$ -tubulina e para a  $\alpha\beta$ -tubulina, baseada nos valores de pontuação e RMSD para as funções ChemPLP, ChemScore e GoldScore.

$\beta$ -tubulina	$\alpha\beta$ -tubulina	Pontuação	RMSD (Å)
<b>ChemPLP</b>		62,394	1,068
	<b>ChemPLP</b>	64,756	1,224
<b>ChemScore</b>		25,149	0,800
	<b>ChemScore</b>	28,572	0,813
<b>GoldScore</b>		59,034	1,078
	<b>GoldScore</b>	69,078	0,978

Portanto, como houve aumento significativo na energia do complexo com as duas estruturas juntas em todas as funções testadas, a  $\alpha$ -tubulina foi mantida com a  $\beta$ -tubulina nos cálculos posteriores, cuja estrutura dimérica,  $\alpha\beta$ -tubulina, está apresentada na Figura 30:

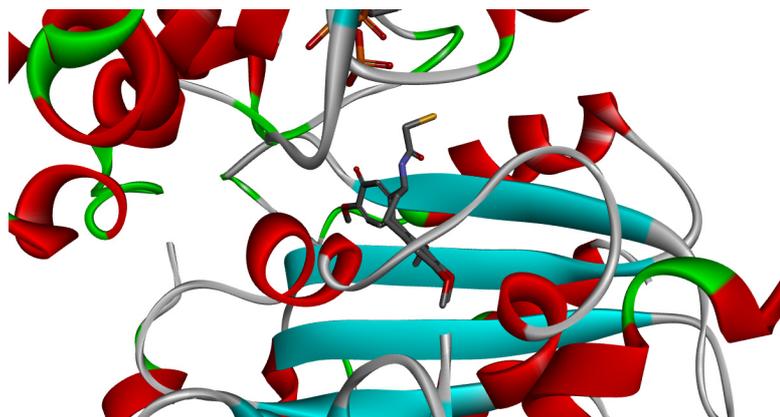
Figura 30 – Estrutura remanescente utilizada para os cálculos de docagem molecular, sendo mantidas a  $\alpha$ -tubulina (direita) e a  $\beta$ -tubulina (esquerda).



Fonte: Elaborada pelo autor.

E, na mesma estrutura, aproximando a visualização, podemos observar com mais detalhes o sítio de ligação da colchicina, conforme a Figura 31:

Figura 31 – Sítio de ligação da colchicina, no qual encontra-se o ligante cristalográfico CN2.



Fonte: Elaborada pelo autor.

Após estabelecida a estrutura da  $\alpha\beta$ -tubulina para os cálculos, foi feita a comparação, com cálculos de *redocking*, entre as três funções de pontuação (GoldScore, ChemPLP e ChemScore), a fim de selecionar uma delas para realizar os

cálculos posteriores. Visando a maior aproximação possível da realidade, foi adicionada flexibilidade a 10 resíduos, pois esse é o número máximo admitido pelo programa.

A partir de informações obtidas no banco de dados PDB e em outros trabalhos publicados, as quais foram devidamente verificadas, foi dada flexibilidade, inicialmente, aos resíduos THR179 $\alpha$ , VAL238 $\beta$ , CYS241 $\beta$ , LEU248 $\beta$ , LEU255 $\beta$ , ASN258 $\beta$ , MET259 $\beta$ , ASN349 $\beta$ , LYS352 $\beta$ , ILE378 $\beta$ , a qual foi denominada **flexibilidade 1**. Os sufixos nos resíduos denotam o monômero a que pertencem:  $\alpha$  ou  $\beta$ -tubulina.<sup>93,98-103</sup> Os resultados estão apresentados na Tabela 14.

Tabela 14 – Valores de pontuação e RMSD para as funções GoldScore, ChemPLP e ChemScore.

Função	Pontuação	RMSD (Å)
GoldScore	66,909	0,988
ChemPLP	64,870	1,43
ChemScore	31,140	1,297

Após serem identificados os menores valores de RMSD e analisar a conformidade visual dos ligantes, a função de pontuação que apresentou melhor desempenho foi a Goldscore, principalmente quando sujeita a cálculos de flexibilidade de resíduos. Portanto, foi ela a escolhida para realizar todos os demais cálculos.

Estando definidas a estrutura do receptor e a função de pontuação a serem utilizadas nos cálculos, foram selecionados alguns compostos a serem utilizados como ligantes nesses cálculos, adotando a flexibilidade dos resíduos citados anteriormente. Da Tabela 20, no apêndice A, foram selecionados os compostos 2a, 2 $\beta$ , 2c, 2d, 2g, 2nn, 2rr, 2vv, 2xx, 2zz, 3a, 3b, 3c, 4a, 4b, 4c, 4d, 5d, 6a, 8a e 8d, buscando estabelecer alguma relação entre a energia do complexo e os valores de atividade biológica. As pontuações calculadas encontram-se na Tabela 15. Contudo, como pode ser percebido nessa tabela, não foi possível estabelecer nenhuma relação da pontuação com a atividade biológica dos compostos.

Os resultados dos complexos da  $\alpha\beta$ -tubulina com os compostos da tabela acima foram analisados tanto na interface gráfica do programa Gold, quanto no Discovery Studio, no qual foi possível analisar as interações a partir de diagramas 2D, dos quais alguns são apresentados a seguir nas Figuras 32 a 35, que se referem aos

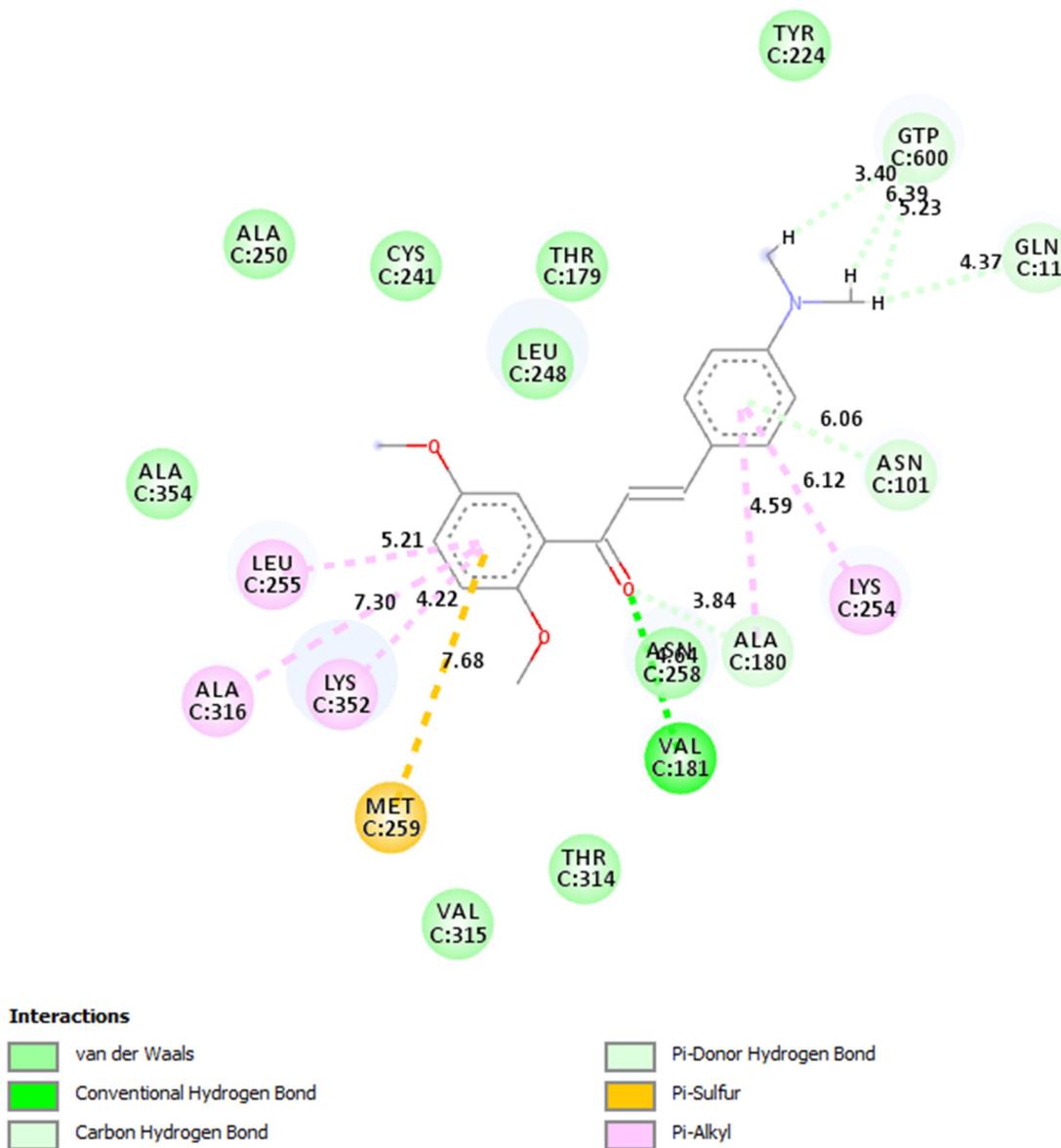
ligantes 2 $\beta$ , 2d, 3a e 4a, respectivamente. A distância da interação é medida em angström (Å).

Tabela 15 – Valores de pontuação para a melhor conformação de cada um dos compostos selecionados para as análises de docagem molecular (flexibilidade 1).

<b>Composto</b>	<b>Pontuação (Fitness)</b>		
<b>2a</b>	52,46	<b>3a</b>	57,29
<b>2<math>\beta</math></b>	49,67	<b>3b</b>	53,79
<b>2c</b>	52,39	<b>3c</b>	54,03
<b>2d</b>	53,11	<b>4a</b>	51,56
<b>2g</b>	55,32	<b>4b</b>	52,76
<b>2nn</b>	54,20	<b>4c</b>	49,46
<b>2rr</b>	52,07	<b>4d</b>	52,70
<b>2vv</b>	56,42	<b>5d</b>	49,22
<b>2xx</b>	51,85	<b>6a</b>	54,93
<b>2zz</b>	51,90	<b>8a</b>	57,31
		<b>8d</b>	53,51

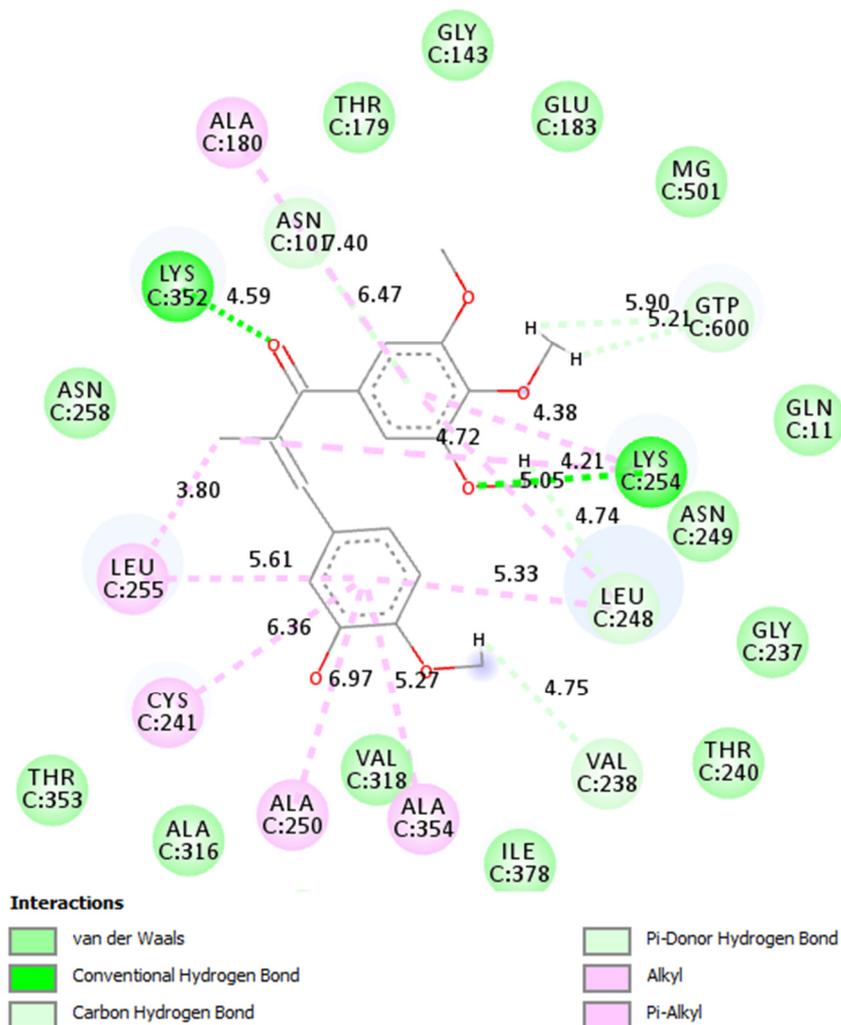


Figura 33 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 2d no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1.



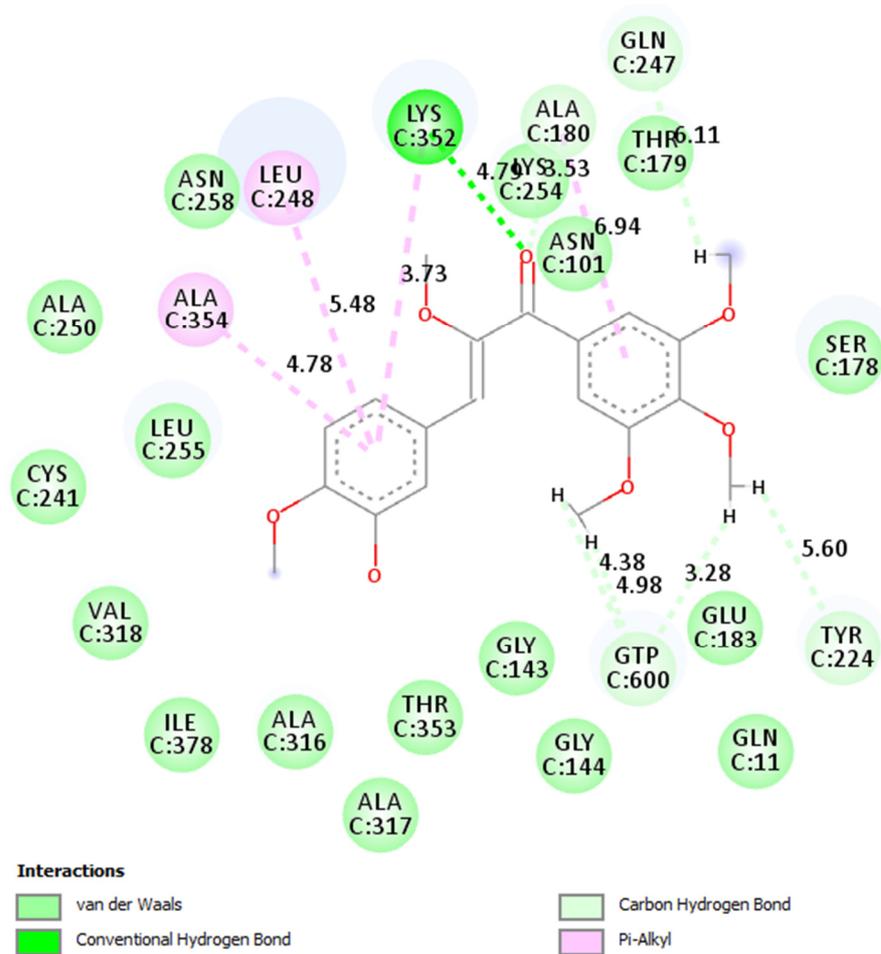
Fonte: Elaborada pelo autor.

Figura 34 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1.



Fonte: Elaborada pelo autor.

Figura 35 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 4a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 1.



Fonte: Elaborada pelo autor.

Entretanto, foi percebido também que alguns resíduos pertencentes a  $\alpha$ -tubulina se fizeram presentes nessas interações frequentemente, como também a molécula de GTP. Isso pode sugerir que os ligantes analisados, mesmo sendo considerados do sítio da  $\beta$ -tubulina, também interajam com resíduos da  $\alpha$ -tubulina, já que esse sítio de ligação se encontra numa região interface entre os dois monômeros. Sendo assim, foram realizados testes com os mesmos ligantes, mas fornecendo flexibilidade a alguns resíduos da  $\alpha$ -tubulina, buscando que os dois monômeros se equilibrassem em termos de quantidade de resíduos flexíveis. Sendo assim, temos agora resíduos cujas flexibilidades foram baseadas tanto em buscas na literatura, como também em nossas análises prévias, que nos forneceram características das interações desse conjunto de ligantes no referido sítio.

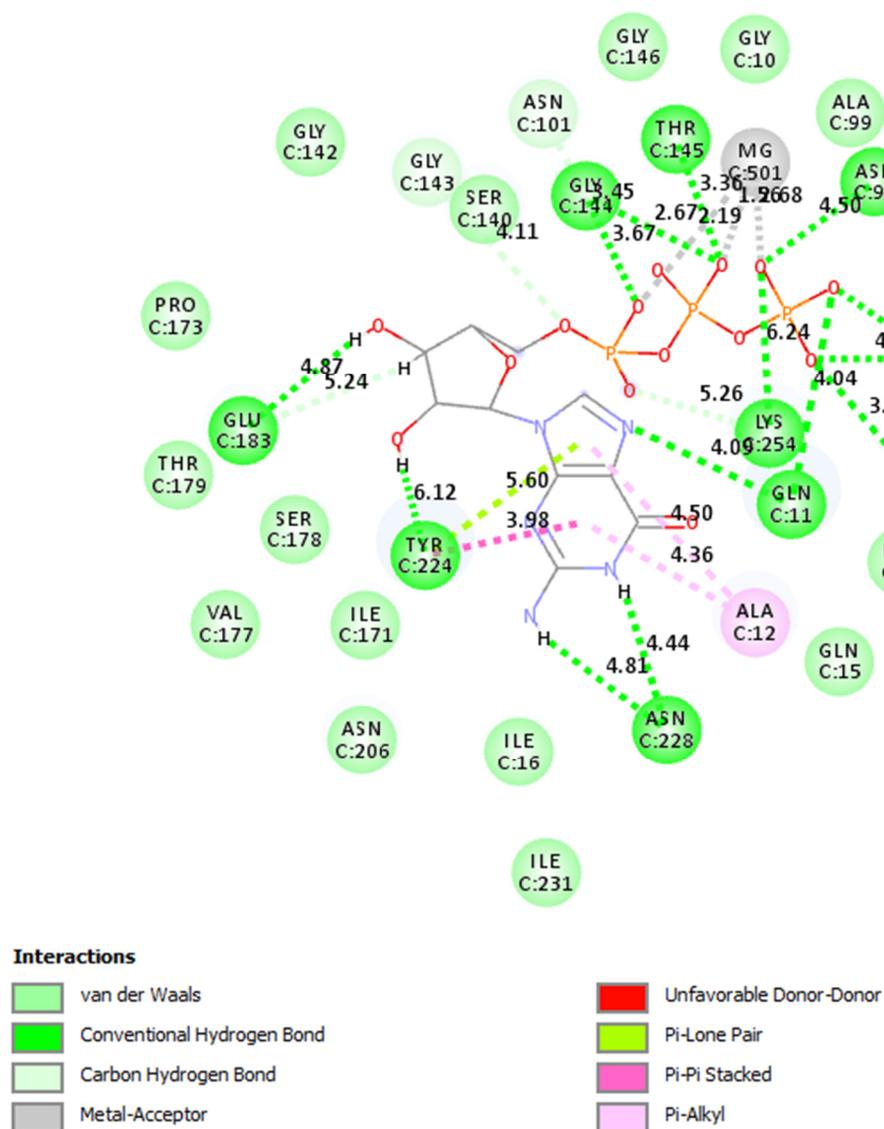
Os resíduos escolhidos para ser fornecida flexibilidade foram: LYS254 $\beta$ , LYS352 $\beta$ , ANS258 $\beta$ , LEU248 $\beta$ , TYR224 $\alpha$ , SER178 $\alpha$ , THR179 $\alpha$ , GLU183 $\alpha$ , ASN101 $\alpha$ , GLN11 $\alpha$ , a qual foi denominada **flexibilidade 2**. A justificativa para a escolha desses resíduos da  $\alpha$ -tubulina é o fato de aparecerem frequentemente nas análises gráficas, e com os diagramas 2D, ficam mais perceptíveis suas interações.

Vale ressaltar também que muitos desses resíduos escolhidos da  $\alpha$ -tubulina interagem com a molécula de GTP. Isso pode indicar que, para alguns resíduos de aminoácidos, os compostos aqui estudados e a molécula de GTP são concorrentes, mesmo não formando complexos no mesmo sítio. E a molécula de GTP, como visto na introdução deste trabalho, está diretamente relacionada com a capacidade de polimerização dos microtúbulos.

Para fins de comparação dos resíduos de aminoácidos que interagem com os compostos aqui analisados e com o GTP, na Figura 36 está a representação do diagrama 2D para o complexo da molécula de GTP no seu respectivo sítio, onde pode ser percebido que alguns resíduos que interagem com essa molécula interagem também com os ligantes analisados. Por exemplo, LYS254 $\beta$ , TYR224 $\alpha$ , GLU183 $\alpha$ , ASN101 $\alpha$ , GLN11 $\alpha$  e GLY144 $\alpha$  são resíduos que se fizeram presentes nesse cálculo de interação com o GTP e também se fazem presentes nos cálculos com os ligantes estudados, além de outros resíduos de menor ocorrência.

Após estarem definidos quais seriam os resíduos flexíveis durante os cálculos de docagem, e vendo a necessidade da manutenção da molécula de GTP nesses mesmos cálculos, foi possível obter novos resultados nas simulações e, conseqüentemente, novos valores de pontuação para cada uma das simulações. Os valores de pontuação para as moléculas analisadas podem ser vistos na Tabela 16. Para facilitar a comparação, nessa mesma tabela, estão apresentados também os valores de pontuação para os cálculos realizados antes de ser inteiramente definida a flexibilidade dos resíduos (Tabela 15). Nota-se o quanto a mudança nos valores de pontuação foi expressiva.

Figura 36 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e a molécula de GTP no seu respectivo sítio de ligação. Os resíduos de aminoácidos foram mantidos rígidos para fins de comparação.



Fonte: Elaborada pelo autor.

Tabela 16 – Comparação dos valores de pontuação entre o primeiro e segundo testes de flexibilidade fornecida a alguns dos resíduos de aminoácidos da  $\alpha$  e  $\beta$ -tubulina. Pode ser percebido o contraste entre as escolhas de flexibilidade.

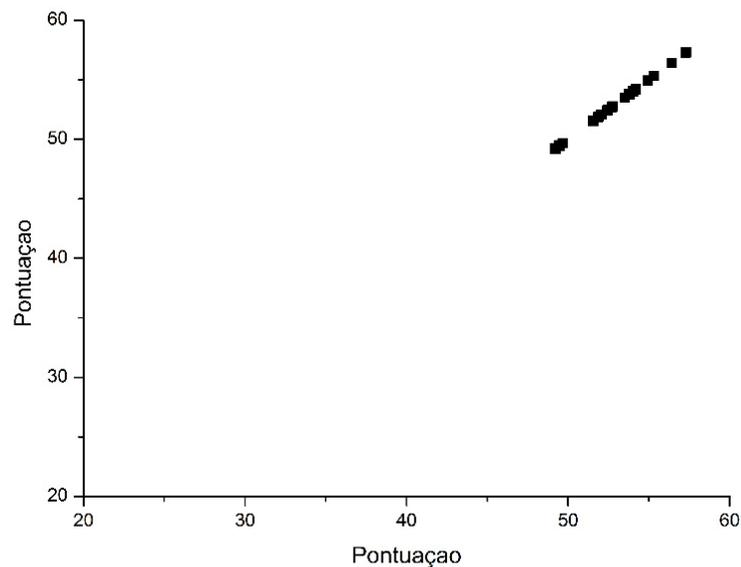
<b>Composto</b>	<b>Pontuação (Fitness) Flexibilidade 1</b>	<b>Pontuação (Fitness) Flexibilidade 2</b>
<b>2a</b>	52,46	41,00
<b>2<math>\beta</math></b>	49,67	26,23
<b>2c</b>	52,39	36,60
<b>2d</b>	53,11	46,67
<b>2g</b>	55,32	45,39
<b>2nn</b>	54,20	32,80
<b>2rr</b>	52,07	33,61
<b>2vv</b>	56,42	36,03
<b>2xx</b>	51,85	32,01
<b>2zz</b>	51,90	30,69
<b>3a</b>	57,29	42,49
<b>3b</b>	53,79	44,96
<b>3c</b>	54,03	42,10
<b>4a</b>	51,56	47,14
<b>4b</b>	52,76	47,12
<b>4c</b>	49,46	49,73
<b>4d</b>	52,70	43,11
<b>5d</b>	49,22	30,00
<b>6a</b>	54,93	46,02
<b>8a</b>	57,31	49,38
<b>8d</b>	53,51	32,36

Como pode ser notado, ao ser considerada a flexibilidade de resíduos da  $\alpha$ -tubulina, passou a existir um contraste entre os valores de pontuação. E isso tornou possível estabelecer um indicativo de moléculas de alta atividade biológica a partir dos valores de pontuação dos complexos. Os valores para o complexo em que não se levava em consideração resíduos flexíveis da  $\alpha$ -tubulina são praticamente uniformes: praticamente todos os valores estão entre 50 e 60. Já nas análises nas quais se levou em consideração a flexibilidade de alguns resíduos da  $\alpha$ -tubulina, nota-se que há um contraste bastante considerável entre os ligantes: a maioria dos ligantes de baixa atividade biológica não apresentaram valores de pontuação maiores do que

40 (26 a 36). Já os de alta atividade biológica apresentaram valores de pontuação maiores de 40 (41 a 49) em sua maioria, o que significa maior interação do complexo.

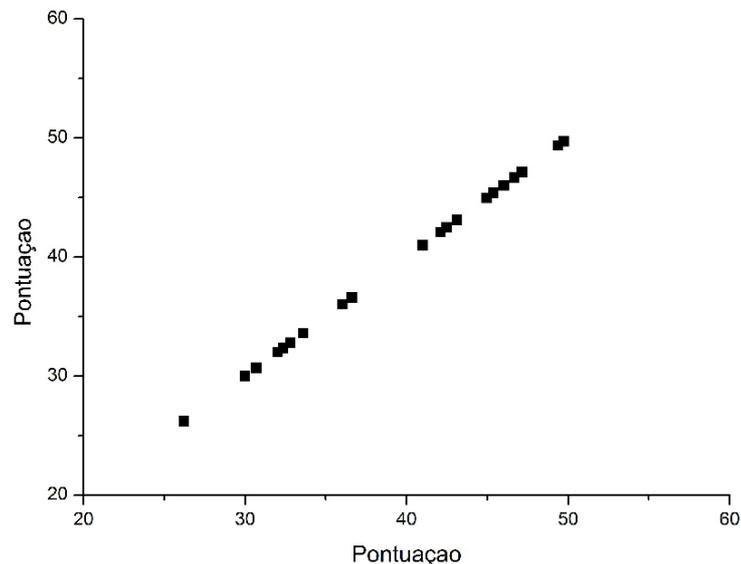
Nas Figuras 37 e 38 estão apresentados os gráficos correspondentes a Tabela 16. Na Figura 37, correspondente aos cálculos que não contemplaram a flexibilidade dos resíduos da  $\alpha$ -tubulina, temos o gráfico de pontuação *versus* pontuação. Já na Figura 38, temos outro gráfico de pontuação *versus* pontuação, com os valores de pontuação dos cálculos que levaram em consideração a flexibilidade de alguns resíduos da  $\alpha$ -tubulina. Nesse segundo gráfico, é facilmente notável uma formação de dois grupos principais: um acima dos valores de 40, e o outro abaixo desse valor. O grupo de compostos que se localiza acima dos valores de 40 corresponde às moléculas de alta atividade biológica. E o grupo localizado abaixo desse valor corresponde às moléculas de baixa atividade biológica.

Figura 37 – Plotagem dos valores de pontuação *versus* pontuação dos complexos formados entre os ligantes selecionados e a  $\alpha\beta$ -tubulina quando os resíduos flexíveis foram os mesmos da flexibilidade 1. Nota-se a baixa dispersão dos valores no gráfico.



Fonte: Elaborada pelo autor.

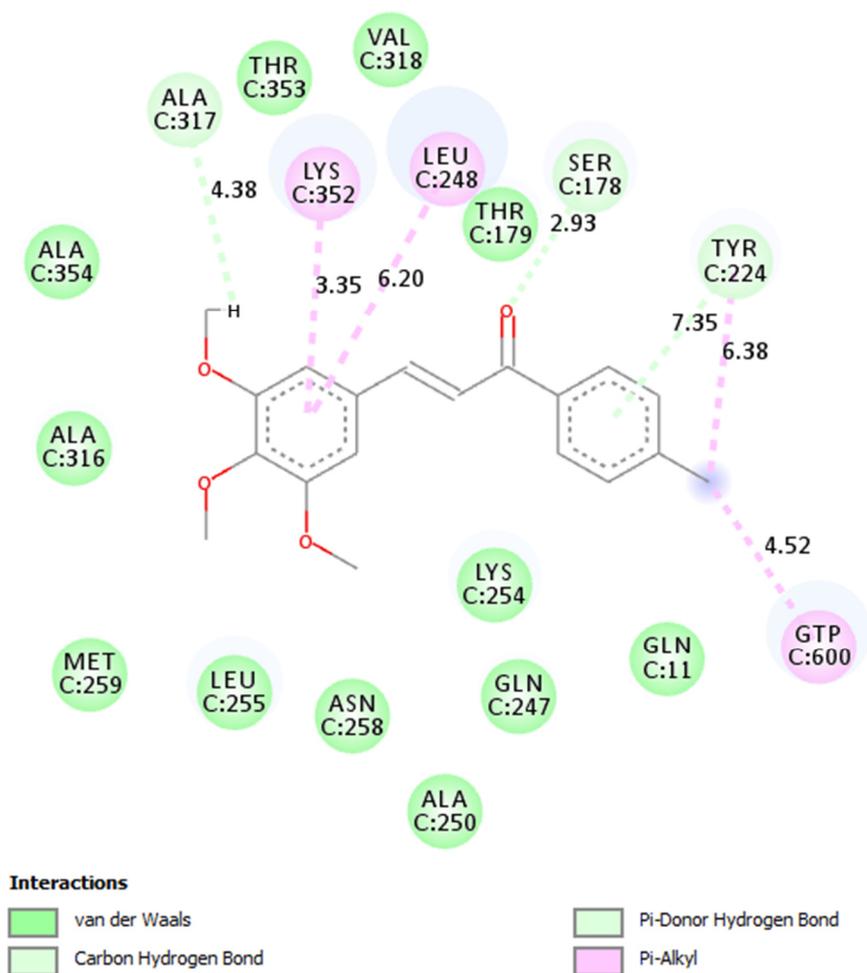
Figura 38 – Plotagem dos valores de pontuação *versus* pontuação dos complexos formados entre os ligantes selecionados e a  $\alpha\beta$ -tubulina quando os resíduos flexíveis foram os mesmos da flexibilidade 2 nos cálculos de docagem. Nota-se uma maior dispersão dos pontos se comparados aos da Figura 36, e também, dois agrupamentos principais: um para valores de pontuação abaixo de 40, e outro acima de 40.



Fonte: Elaborada pelo autor.

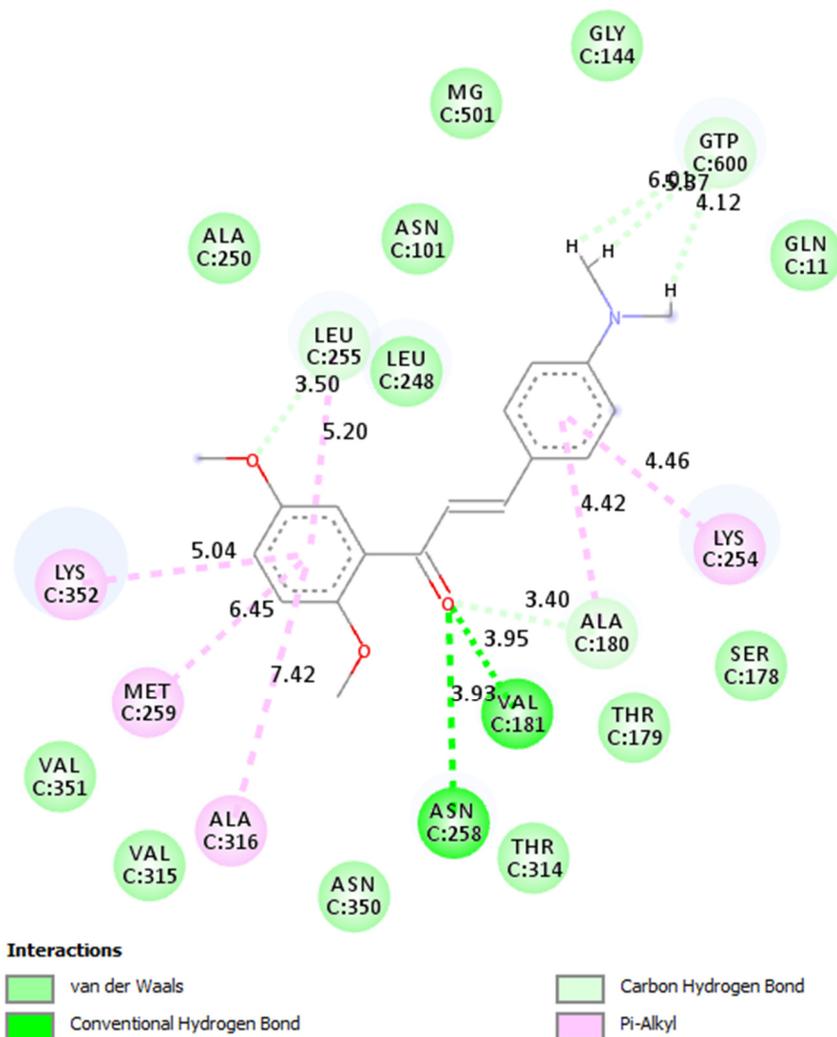
Isso demonstra que, além de apenas um contraste, há a indicação que moléculas de alta atividade biológica desse conjunto de compostos apresentam alta afinidade de interação com resíduos da  $\alpha$ -tubulina, e isso pode ser usado como um indicativo de moléculas de alta atividade biológica, tanto qualitativamente (análise de interações) quanto quantitativamente (valores de pontuação). Essas observações nos levam a crer que a interação com resíduos da  $\alpha$ -tubulina é um fator que deve ser levado em consideração para um melhor entendimento do comportamento do complexo, e por isso, tais flexibilidades dos resíduos foram mantidas nos cálculos posteriores. Com os complexos bem definidos, podem ser analisadas mais detalhadamente suas interações. Nas Figuras 39 a 42, estão apresentados os diagramas 2D para os ligantes 2 $\beta$ , 2d, 3a e 4a, respectivamente, após ser fornecida a flexibilidade a um diferente grupo de resíduos (flexibilidade 2). Os diagramas dos demais ligantes selecionados para as análises de docagem encontram-se no apêndice C, nas Figuras 56 a 64.

Figura 39 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante  $2\beta$  no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



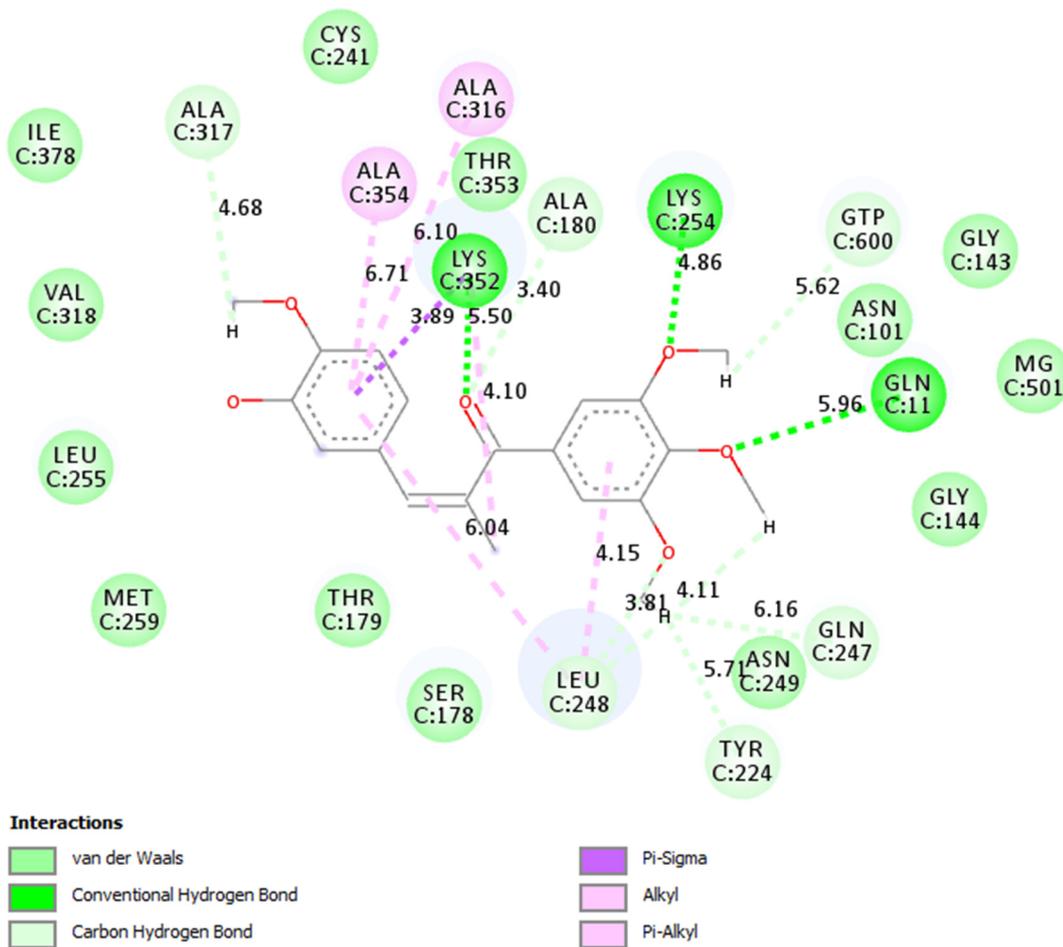
Fonte: Elaborada pelo autor.

Figura 40 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 2d no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



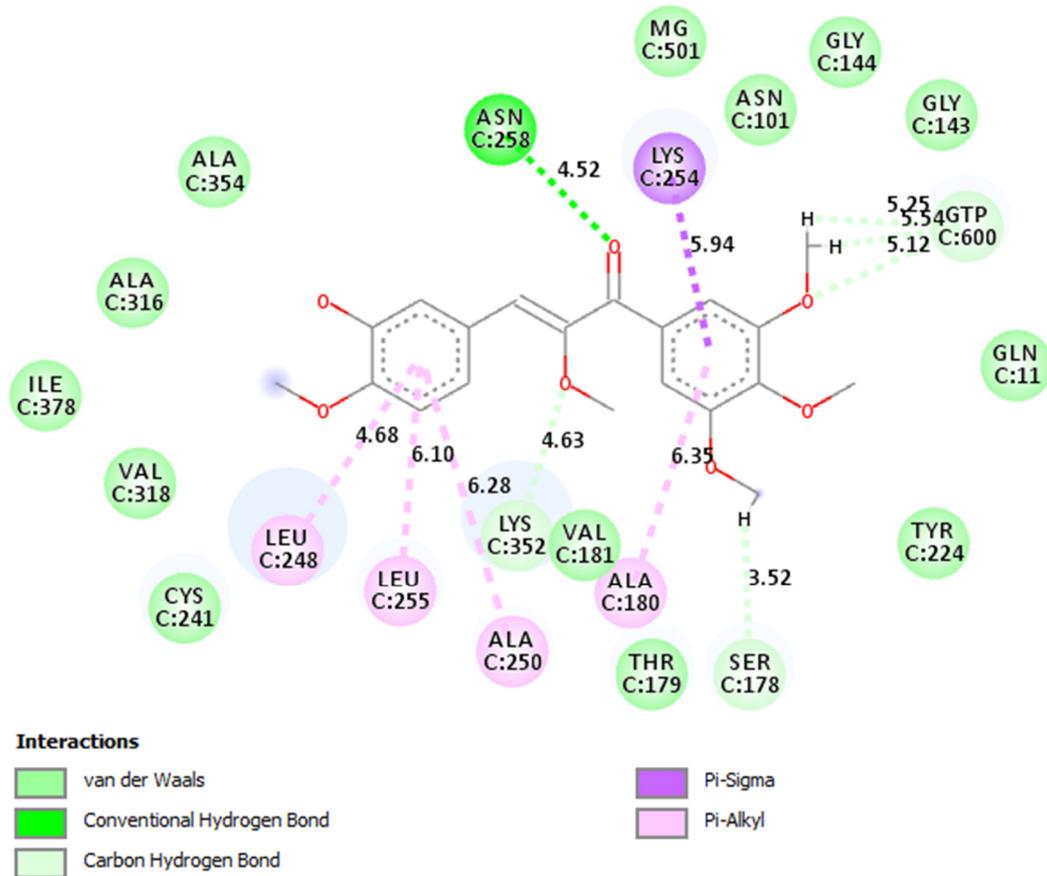
Fonte: Elaborada pelo autor.

Figura 41 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

Figura 42 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 4a no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

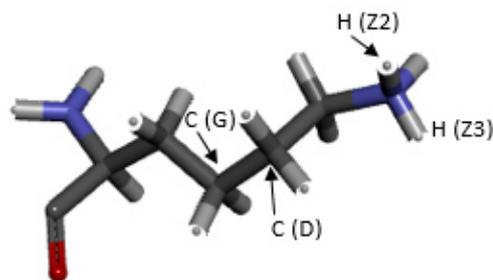
Analisando os diagramas 2D apresentados, podemos observar a capacidade de interação desses ligantes, bem como os tipos de interação relacionados. Nota-se que as interações mais significativas que ocorreram na formação dos complexos foram as ligações de hidrogênio e também as interações hidrofóbicas, em grande parte representadas por interações do tipo  $\pi$ -alquil. As ligações de hidrogênio geralmente ocorriam com os resíduos LYS254, LYS352, THR179, SER178, GLN11 e a molécula de GTP. Já as interações hidrofóbicas ocorriam com os resíduos LYS254, LYS352 e LEU248. Além disso, ocorre também uma grande quantidade de interações de van der Waals de diversos resíduos além dos acima citados, com todos os ligantes analisados. Dessa forma, podemos também, analisar com mais detalhes quais são os grupos e átomos envolvidos nas interações desses complexos.

#### 4.4.1 LYS254

As ligações de hidrogênio provenientes da interação dos ligantes com a LYS254 ocorrem a partir da extremidade  $\text{NH}_3$  desse resíduo. Mais especificamente, se dão a partir do átomo de hidrogênio localizado na posição Z3 desse grupo  $\text{NH}_3$  e, em alguns casos a partir do átomo de hidrogênio localizado na posição Z2 desse mesmo grupo, conforme indicado na Figura 43. Nota-se que o hidrogênio da posição Z3 da LYS254 tem um papel fundamental nas ligações de hidrogênio com os mais diversos ligantes. Lembrando que H (Z2) e H (Z3) compõem o mesmo grupo  $\text{NH}_3$  da LYS254. Nos ligantes, essas ligações de hidrogênio ocorrem, em geral, nos grupos ligados ao anel aromático A, desde a posição  $\text{R}^3$  até a posição  $\text{R}^6$ .

As ligações  $\pi$ -alquil que ocorrem entre os ligantes e a LYS254 ocorrem, em sua maioria, a partir dos átomos de carbono localizados nas posições G e D nesse resíduo, também indicadas na Figura 43. Em geral, interagem com o anel aromático A. Esses átomos de carbono (C (G) e C (D)) estão localizados numa região central do resíduo, de perfil hidrofóbico. Essa região se mostrou como fundamental para as interações hidrofóbicas da LYS254 com os ligantes, mostrando também a capacidade desse resíduo interagir com os ligantes tanto a partir de ligações de hidrogênio quanto de interações hidrofóbicas.

Figura 43 – Representação da LYS254, com os átomos de hidrogênio indicados nas posições Z2 e Z3, bem como os átomos de carbono das posições G e D.



Fonte: Elaborada pelo autor.

#### 4.4.2 LYS352

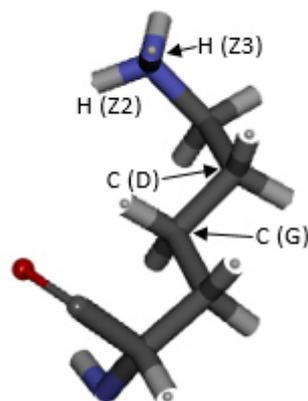
As ligações de hidrogênio entre os ligantes e a LYS352 se dá, como na LYS254, através de uma extremidade  $\text{NH}_3$ , a partir do átomo de hidrogênio localizado na

posição Z2 e, em alguns casos, a partir do átomo de hidrogênio localizado na posição Z3, indicados na Figura 44. Dessa forma, nota-se a capacidade de interação desse grupo e a sua importância nas ligações de hidrogênio com os ligantes. A maior diferença com relação à LYS254 é que essa interação ocorre na região central dos ligantes, mais precisamente com o átomo de oxigênio da carbonila, denotando assim a região de interação desse resíduo com os ligantes analisados.

Já nas ligações do tipo  $\pi$ -alquil, esse resíduo interage principalmente com o anel aromático B, a partir dos átomos de carbono localizados nas posições G e D desse resíduo (Figura 44). Vale ressaltar que essa região hidrofóbica da LYS352 corresponde ao mesmo tipo de região descrito nas interações da LYS254, com a diferença de que essa interação hidrofóbica da LYS352 ocorre com o anel aromático B. Dessa forma, nota-se que a semelhança desses dois resíduos é bastante considerável, ainda que interajam com os ligantes em regiões diferentes.

Portanto, temos dois resíduos com características bastante semelhantes, que se mostraram fundamentais para a formação dos complexos analisados, onde ambos são capazes de interagir de duas maneiras distintas (ligações de hidrogênio e  $\pi$ -alquil), e contribuindo assim para a estabilidade do ligante nessa região.

Figura 44 – Representação da LYS352, com os átomos de hidrogênio indicados nas posições Z2 e Z3, assim como os átomos de carbono das posições G e D.

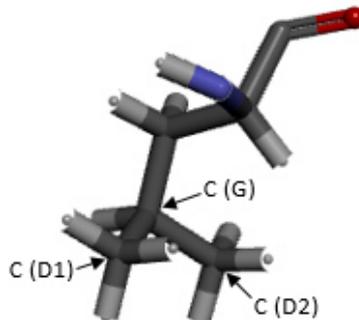


Fonte: Elaborada pelo autor.

#### 4.4.3 LEU248

As ligações do tipo  $\pi$ -alquil caracterizaram as interações da LEU248 com os ligantes. Esse resíduo foi capaz de interagir com os dois anéis dos ligantes (anéis A e B) por ficar posicionado mais próxima à região central dos ligantes. Dessa forma, a interação poderia ocorrer tanto no anel aromático A quanto no B, conforme o posicionamento e as características estruturais dos ligantes. Essas interações sempre ocorreram a partir dos átomos de carbono do resíduo localizados nas posições G e D, caracterizando a região hidrofóbica do resíduo, apresentados na Figura 45. Eventualmente, ocorreram também interações hidrofóbicas desse resíduo com o grupo metoxila localizado na região central de alguns ligantes, sendo esse grupo um substituinte na posição R, em alguns casos. É interessante notar também que a estrutura desse tipo de ligante, com 2 anéis aromáticos, distribuídos nas extremidades, mesmo com variações na estrutura e no posicionamento, favoreceu muito esse tipo de interação.

Figura 45 – Representação da LEU248, com indicação dos átomos de carbono das posições G e D (D1 e D2).



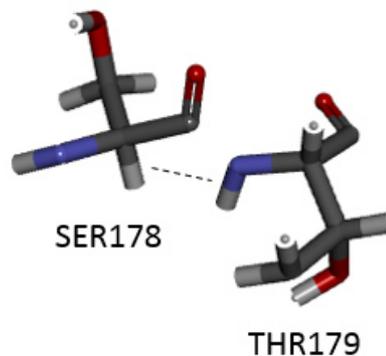
Fonte: Elaborada pelo autor.

#### 4.4.4 THR179 e SER178

Esses dois resíduos interagem de maneira bastante semelhante a partir de ligações de hidrogênio (convencionais e do tipo carbono hidrogênio) com a região central dos ligantes. Interagem tanto com oxigênio da carbonila quanto com o grupo metoxila na posição R, quando presente no ligante. Eventualmente, podem interagir

com grupos metoxila das posições R<sup>3</sup> e R<sup>5</sup>. A interação do THR179 com os ligantes se dá de duas maneiras: a partir de átomos de O do ligante, e a partir de átomos de O do resíduo. Já a interação da SER178 ocorre principalmente a partir de seu grupo hidroxila. Em ambos os casos, há uma tendência de interação com a região central da molécula. As representações desses resíduos estão apresentadas na Figura 46.

Figura 46 – Representação dos resíduos SER178 e THR179, os quais foram separados para uma melhor visualização. Conforme a linha tracejada, a ligação entre esses resíduos ocorre a partir do átomo de carbono de SER178 e o átomo de nitrogênio da THR179.



Fonte: Elaborada pelo autor.

#### 4.4.5 GLN11

As ligações de hidrogênio que a GLN11 realiza com os ligantes se dá através da extremidade NH<sub>2</sub> da GLN11, mais especificamente a partir de um dos átomos de hidrogênio dessa extremidade, da posição E21, como pode ser observado na Figura 47. Interagem, em geral, com grupos das posições R<sup>4</sup> e R<sup>5</sup> dos ligantes onde, na maioria dos casos, encontra-se um grupo metoxila. Dessa forma, nota-se que há uma tendência de interação com esse grupo quando posicionado próximo ao anel aromático A dos ligantes, e nas posições que confirmam um maior alongamento ao ligante, levando em consideração que a GLN11 é um resíduo pertencente à  $\alpha$ -tubulina.

Figura 47 – Representação da GLN11, com indicação do átomo de hidrogênio da posição E21, responsável pelas ligações de hidrogênio no complexo formado.

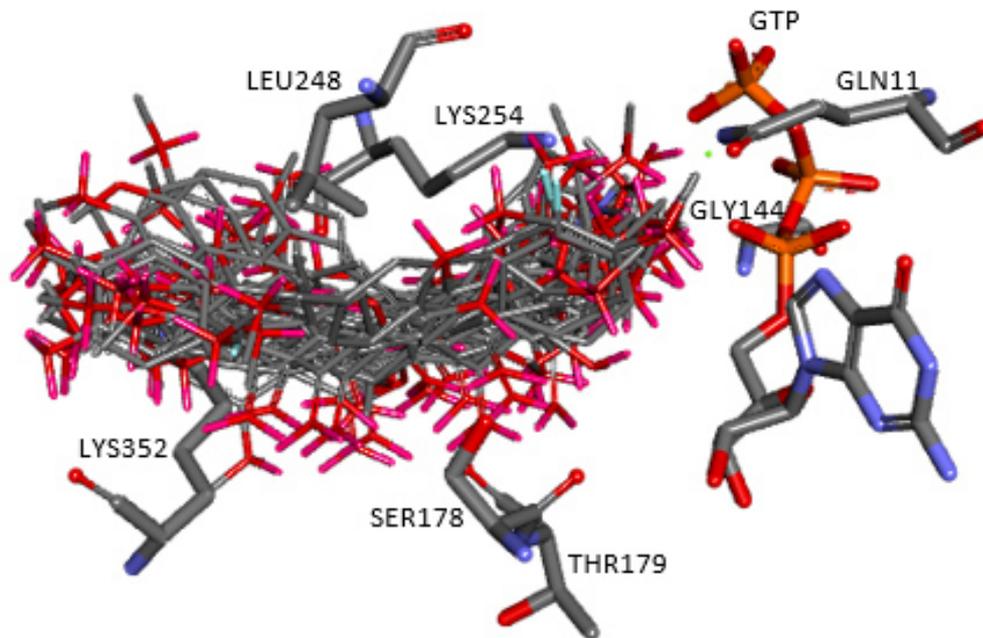


Fonte: Elaborada pelo autor.

#### 4.4.6 GTP

A molécula de GTP interage com os ligantes principalmente a partir de ligações de hidrogênio do tipo carbono hidrogênio, onde um átomo de oxigênio do grupo fosfato do GTP interage, geralmente, com grupos metoxila. No ligante, as interações ocorrem quando há a presença desses grupos metoxila nas posições R<sup>3</sup> e R<sup>4</sup>, denotando as regiões mais alongadas do ligante. Isso é bastante significativo, levando em consideração que o sítio de interação do GTP está localizado na  $\alpha$ -tubulina, e o do ligante está localizado na  $\beta$ -tubulina. Assim, se nota a capacidade de interação desses ligantes na região de interface da estrutura dimérica  $\alpha\beta$ -tubulina, como pode ser observado no modo de ligação obtido para esse conjunto de compostos, apresentado da Figura 48.

Figura 48 – Modo de ligação obtido para as chalconas no sítio de ligação da colchicina (2a, 2 $\beta$ , 2c, 2d, 2g, 2nn, 2rr, 2vv, 2xx, 2zz, 3a, 3b, 3c, 4a, 4b, 4c, 4d, 5d, 6a, 8a e 8d), sendo omitidos os átomos de hidrogênio. Estão apresentados os principais resíduos envolvidos na formação desses complexos, assim como a molécula de GTP.



Fonte: Elaborada pelo autor.

#### 4.5 Proposição de estruturas

Após uma minuciosa análise do complexo formado entre os ligantes estudados e os resíduos do sítio de ligação da colchicina, além da elucidação de um modo de ligação característico para esse grupo de ligantes, foi possível propor algumas alterações na estrutura dos mesmos. Ao todo, foram propostas 36 novas estruturas, as quais foram devidamente analisadas a fim de verificar seus comportamentos no referido sítio, bem como prever suas respectivas atividades biológicas.

Durante as análises de docagem molecular, bem como nas demais análises de QSAR, foi percebido que as estruturas base 2, 3 e 4 apresentaram o maior potencial de proposição de novos compostos, devido, principalmente, à conformação que essas estruturas apresentavam no sítio de interação. As estruturas base 5, 7, 8 e 9 possuem na região central da estrutura, uma diferença bastante significativa com relação às estruturas 2, 3 e 4: elas não possuem uma dupla ligação entre os carbonos  $\alpha$  e  $\beta$ . A estrutura 6 possui essa dupla ligação, porém não apresenta o átomo de oxigênio da carbonila. Ao invés disso, apresenta um grupo hidroxila ligado nessa posição. Essa diferença de estrutura na região central das moléculas comprometeu sensivelmente suas capacidades de interação na formação do complexo, e por isso, se mantiveram nessa etapa, como estruturas base, as estruturas 2, 3 e 4.

Nas análises de Ducki *et al*, as estruturas 2, 3 e 4 são as que apresentaram os maiores valores de atividade biológica, bem como uma maior quantidade de compostos com altos valores de atividade.<sup>25</sup> Isso pode ser tomado como um indicativo de maior potencial destas, frente às demais estruturas. Além de que, nas análises iniciais de docagem molecular, tais estruturas apresentaram maior uniformidade em seus modos de ligação, o que acaba por conferir maior estabilidade às análises e à proposição de novas estruturas. Em particular, as estruturas 3 e 4 têm um comportamento bastante interessante e parecido. A estrutura 2, apesar de ter vários compostos de baixa atividade biológica, também apresentou diversos compostos com altos valores de atividade biológica. A partir dessas estruturas base, foram propostas algumas alterações e, conforme respondiam positivamente ou negativamente às alterações, novas estruturas iam sendo propostas. Também, é importante enfatizar que, por muitas vezes, a simples adição de um grupo, ou então a troca de posições dos grupos, acaba por gerar um comportamento não tão previsível na conformação dessa molécula no seu respectivo sítio. Em vários casos, ocorrem mudanças bastante

significativas, como por exemplo, a inversão horizontal da molécula com a simples troca/adição de um pequeno grupo. Dessa forma, é interessante que seja testado um número razoável de possibilidades, a fim de que uma parcela desses compostos apresente o comportamento esperado.

Ainda nas análises de docagem molecular anteriores, foi percebido que alguns grupos exerciam um papel bastante significativo nas interações dos complexos. Por exemplo, o grupo hidroxila favoreceu muito em diversas interações, principalmente em ligações de hidrogênio. Já o grupo metoxila contribuiu muito para aumentar o volume dos ligantes, sem que fosse perdida a capacidade de interação. Em geral, grupos que eram constituídos de algum halogênio não apresentaram resultados satisfatórios na capacidade de interação dos ligantes. A seguir, será descrito com mais detalhes o papel de cada um desses grupos nas diferentes posições dos ligantes analisados.

Dada a posição dos ligantes no complexo, o anel aromático A corresponde à região de maior proximidade da  $\alpha$ -tubulina, e juntamente com seus grupos substituintes das regiões de R<sup>2</sup> a R<sup>6</sup>, a região mais suscetível a interações, por se encontrar na região mais afastada da  $\beta$ -tubulina, ou seja, na interface entre a  $\alpha$  e a  $\beta$ -tubulina. Nesse anel, podemos notar uma tendência de predominância de grupos substituintes que está relacionada com a alta atividade biológica, que foi também tomada por base para propor algumas estruturas.

Na posição R<sup>2</sup> desse anel, notamos que, além do átomo de hidrogênio, não houveram outros grupos significativos. Isso pode favorecer a presença de grupos mais volumosos em posições próximas, como também a possibilitar interações hidrofóbicas.

Já nas posições R<sup>3</sup>, R<sup>4</sup> e R<sup>5</sup>, há uma predominância muito forte do grupo hidroxila, o que favorece, entre outras interações, à ligação de hidrogênio. Essas posições, principalmente a R<sup>3</sup> e R<sup>4</sup>, também aceitam, em certo grau, a inclusão de grupos mais volumosos, como o grupo metoxila, já que se trata de uma região com maior disponibilidade de espaço.

Na posição R<sup>6</sup> desse anel aromático, assim como na posição R<sup>2</sup>, não há grupos substituintes além do átomo de hidrogênio. Nota-se que essas posições (R<sup>2</sup> e R<sup>6</sup>) podem se tratar de posições estratégicas para os grupos localizados na região central dos ligantes.

Nota-se também que, mesmo que o grupo hidroxila tenha se mostrado como essencial nos grupos substituintes R<sup>3</sup>, R<sup>4</sup> e R<sup>5</sup>, ainda existe a necessidade de algum outro grupo mais volumoso que ele em alguma dessas posições, a depender do perfil do ligante analisado.

O anel aromático B corresponde, na maioria dos ligantes, à região da molécula que se encontra mais inserida na  $\beta$ -tubulina. Nota-se que os grupos substituintes ligados a esse anel aromático apresentam um perfil um pouco menos volumoso se comparado aos grupos ligados ao anel aromático A.

Na posição R<sup>2'</sup>, assim como na posição R<sup>2</sup>, há uma forte tendência da presença apenas do átomo de hidrogênio como substituinte. Já nas posições R<sup>3'</sup>, R<sup>4'</sup> e R<sup>5'</sup>, o grupo hidroxila se fez novamente presente na maioria dos casos. Contudo, a aceitação de grupos mais volumosos, como o metoxila, diminui. Na posição R<sup>6'</sup>, assim como na R<sup>2'</sup>, a tendência é de que haja apenas ligações com átomos de hidrogênio.

Além de haver uma tendência de grupos substituintes de menor volume, também é percebido que há uma tendência de um menor número de grupos substituintes ligados a esse anel aromático.

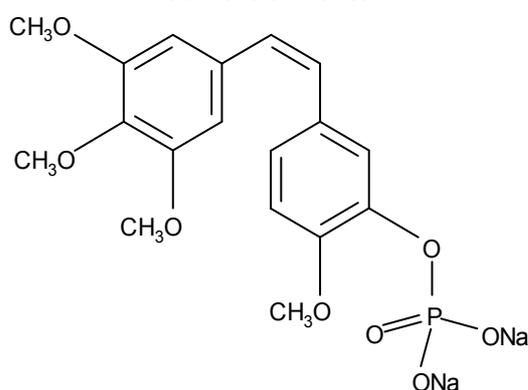
Na região central da molécula, algumas estruturas base, como a 2 e a 5, não apresentam disponibilidade de ligação com grupos substituintes. Mas, nas demais estruturas base aqui estudadas, existe a possibilidade de inclusão de grupos substituintes, que se localizam na posição R. Nessa posição, há a predominância de grupos metila e metoxila, que varia conforme o perfil do ligante. Isso indica a tendência e a importância da presença de um grupo mais volumoso nessa região, já que, em alguns casos, essa região central também se localiza na interface  $\alpha\beta$ -tubulina, o que significa uma região de maior espaço, e um grupo mais volumoso tende a ter mais possibilidades de interação.

Posteriormente, foi também analisado o grupo dihidroxi fosfato (H<sub>2</sub>PO<sub>4</sub>), o qual se mostrou ser um grupo bastante interessante. Por ser formado por um grupo fosfato, apresenta algumas características de interação semelhantes à molécula de GTP, ou seja, pode interagir com resíduos que, a princípio, interagiriam com o GTP, e isso é algo bastante interessante quando se pretende inibir a polimerização dos microtúbulos.

Já é conhecido um fármaco que apresenta uma evolução com relação à combretastatina A4, que é o CA4DP, ou também conhecido como fosbretabulina dissódica, que apresenta um grupo Na<sub>2</sub>PO<sub>4</sub> ligado à combretastatina A4, como pode

ser observado na Figura 49. Esse fármaco já está sendo comercializado, e tem se mostrado de uma eficiência bastante satisfatória, principalmente pela sua capacidade de ser associado com outros tratamentos tradicionais de quimioterapia, radioterapia e cirurgia, sem maiores intercorrências.<sup>104,105</sup> Dessa forma, é interessante analisar o comportamento de alguns ligantes quando adicionado um grupo fosfato (nesse caso, dihidroxi fosfato).

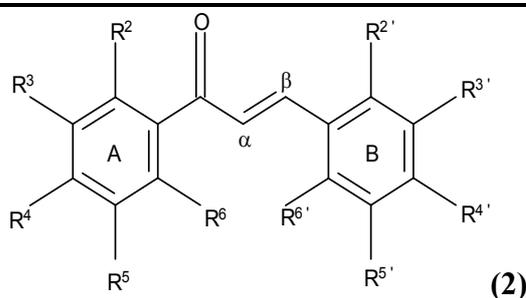
Figura 49 – Representação da fosbretabulina dissódica. Fármaco que já está disponível comercialmente



Fonte: Elaborada pelo autor.

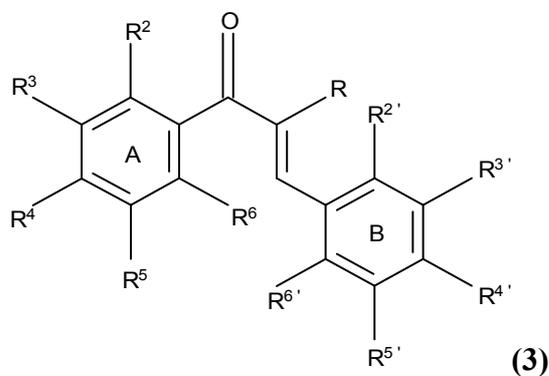
A partir de todas as análises realizadas, é possível que sejam propostas novas estruturas, a fim de que seja melhorada a atividade biológica desse grupo de compostos. No total, foram propostas 36 novas estruturas, que estão apresentadas na Tabela 17.

Tabela 17 – Estruturas propostas para o sítio de ligação da colchicina.

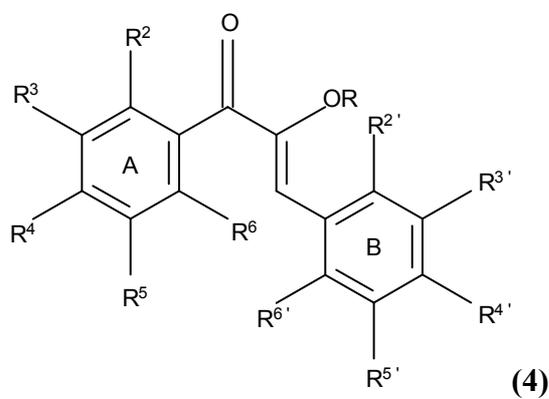


	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	R <sup>6</sup>	R <sup>2'</sup>	R <sup>3'</sup>	R <sup>4'</sup>	R <sup>5'</sup>	R <sup>6'</sup>
<b>2a1</b>	H	OMe	OMe	OMe	H	H	OH	H <sub>2</sub> PO <sub>4</sub>	H	H
<b>2a2</b>	H	OMe	OMe	OMe	H	H	OH	OMe	H <sub>2</sub> PO <sub>4</sub>	H
<b>2a3</b>	H	H <sub>2</sub> PO <sub>4</sub>	OMe	OMe	H	H	OH	OMe	H	H
<b>2a4</b>	H	OMe	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OMe	H	H
<b>2a5</b>	H	H <sub>2</sub> PO <sub>4</sub>	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OMe	H	H

<b>2a6</b>	H	H <sub>2</sub> PO <sub>4</sub>	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H	OH	OMe	H	H
<b>2a7</b>	H	H <sub>2</sub> PO <sub>4</sub>	OMe	OMe	H	H	OH	OEt	H	H
<b>2a8</b>	H	OMe	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OEt	H	H



	<b>R</b>	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>
<b>3a1</b>	Et	H	OMe	OMe	OMe	H	H	OH	OMe	H	H
<b>3a2</b>	Me	H	OMe	OMe	OMe	H	H	H	OMe	OH	H
<b>3a3</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	OMe	H
<b>3a4</b>	Me	H	OH	OH	OH	H	H	OH	OMe	H	H
<b>3a5</b>	Me	H	OH	OMe	OMe	H	H	OH	OMe	H	H
<b>3a6</b>	Me	H	H	OMe	OMe	OMe	H	OH	OMe	H	H
<b>3a7</b>	Me	H	OMe	OMe	OMe	H	H	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H
<b>3a8</b>	Me	H	OMe	OMe	OMe	H	H	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H
<b>3a9</b>	Me	H	OMe	OMe	OMe	H	H	OMe	H	H <sub>2</sub> PO <sub>4</sub>	H
<b>3a10</b>	H <sub>2</sub> PO <sub>4</sub>	H	OMe	OMe	OMe	H	H	OH	OMe	H	H
<b>3a11</b>	H <sub>2</sub> PO <sub>4</sub>	H	OH	OH	OH	H	H	OH	OMe	H	H
<b>3a12</b>	H <sub>2</sub> PO <sub>4</sub>	H	OMe	OH	OMe	H	H	OH	OMe	H	H
<b>3a13</b>	Me	H	H <sub>2</sub> PO <sub>4</sub>	OMe	OMe	H	H	OH	OMe	H	H
<b>3a14</b>	Me	H	OMe	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OMe	H	H
<b>3a15</b>	Me	H	OMe	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H	OH	OMe	H	H

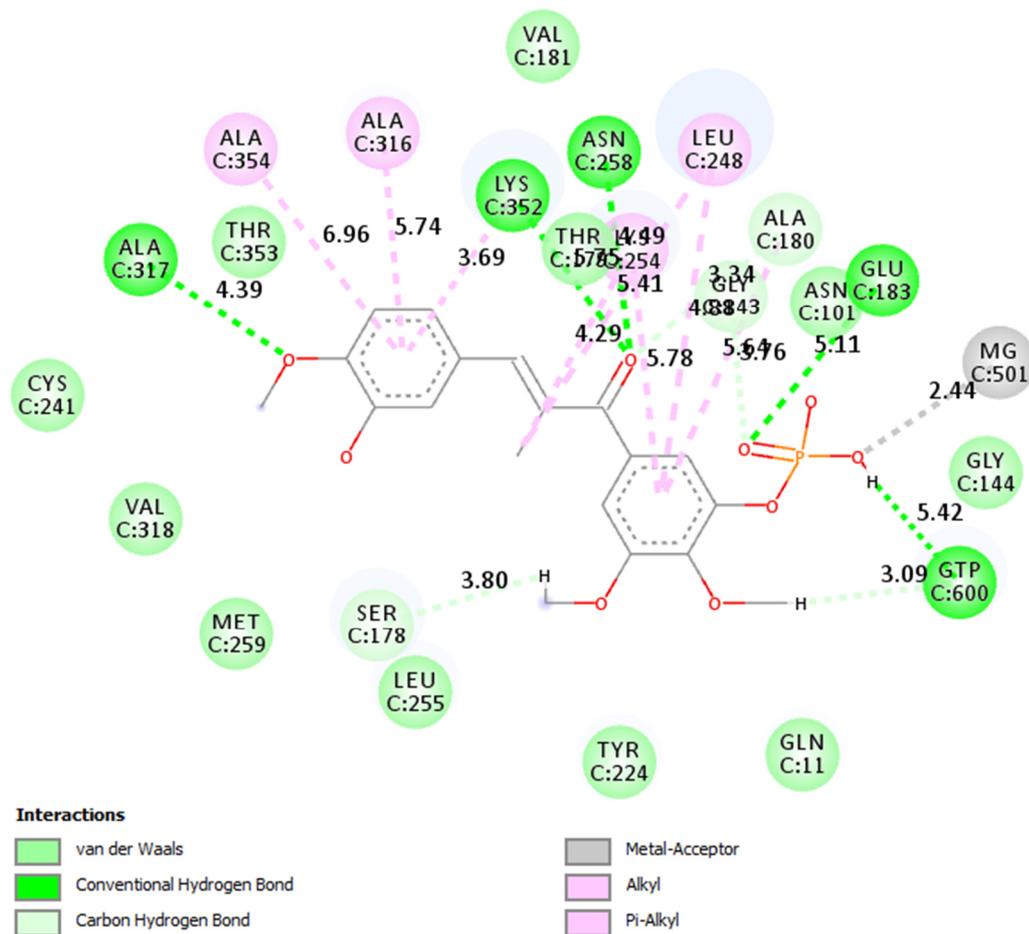


	<b>R</b>	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>
<b>4a1</b>	Me	H	OMe	OMe	OMe	H	H	H	OMe	OH	H
<b>4a2</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	OMe	H
<b>4a3</b>	Me	H	H	OMe	OMe	OMe	H	OH	OMe	H	H

<b>4a4</b>	Me	H	OMe	OMe	OMe	H	H	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H
<b>4a5</b>	Me	H	OMe	OMe	OMe	H	H	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H
<b>4a6</b>	Me	H	OMe	OMe	OMe	H	H	OMe	H	H <sub>2</sub> PO <sub>4</sub>	H
<b>4a7</b>	Me	H	H <sub>2</sub> PO <sub>4</sub>	OMe	OMe	H	H	OH	OMe	H	H
<b>4a8</b>	Me	H	OMe	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OMe	H	H
<b>4a9</b>	Me	H	OMe	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H	OH	OMe	H	H
<b>4a10</b>	Me	H	H <sub>2</sub> PO <sub>4</sub>	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OMe	H	H
<b>4a11</b>	Me	H	H <sub>2</sub> PO <sub>4</sub>	OMe	H <sub>2</sub> PO <sub>4</sub>	H	H	OH	OMe	H	H
<b>4a12</b>	Me	H	H <sub>2</sub> PO <sub>4</sub>	OMe	OMe	H	H	OH	OEt	H	H
<b>4a13</b>	Me	H	OMe	H <sub>2</sub> PO <sub>4</sub>	OMe	H	H	OH	OEt	H	H

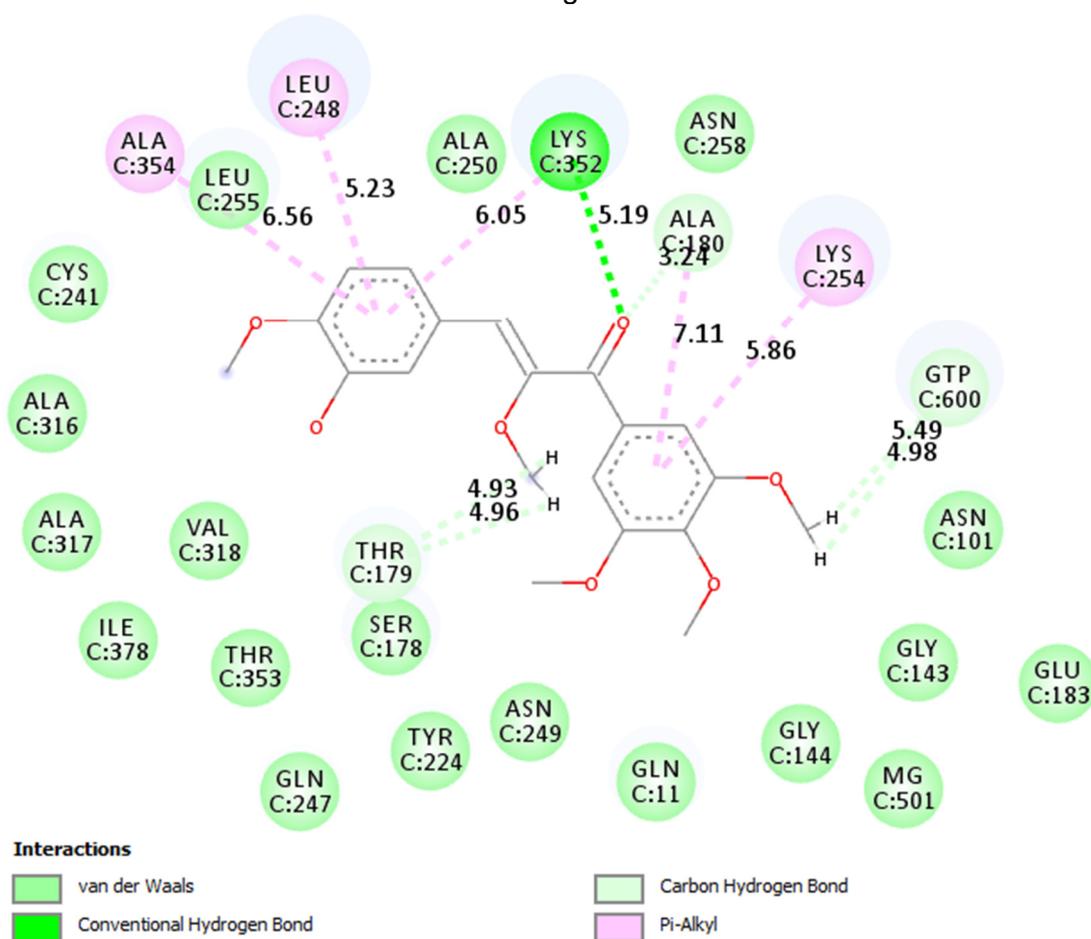
Essas estruturas foram devidamente analisadas nas simulações de docagem molecular, a fim de verificar se apresentavam comportamento semelhante às demais moléculas já analisadas anteriormente. Nas Figuras 50, 51 e 52, são apresentados os diagramas 2D das moléculas propostas 3a13, 4a1 e 4a9, respectivamente, onde podem ser analisadas as interações com os diversos resíduos de aminoácidos, tanto da  $\alpha$  quanto da  $\beta$ -tubulina, além da molécula de GTP, que foi mantida nos cálculos.

Figura 50 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a13 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



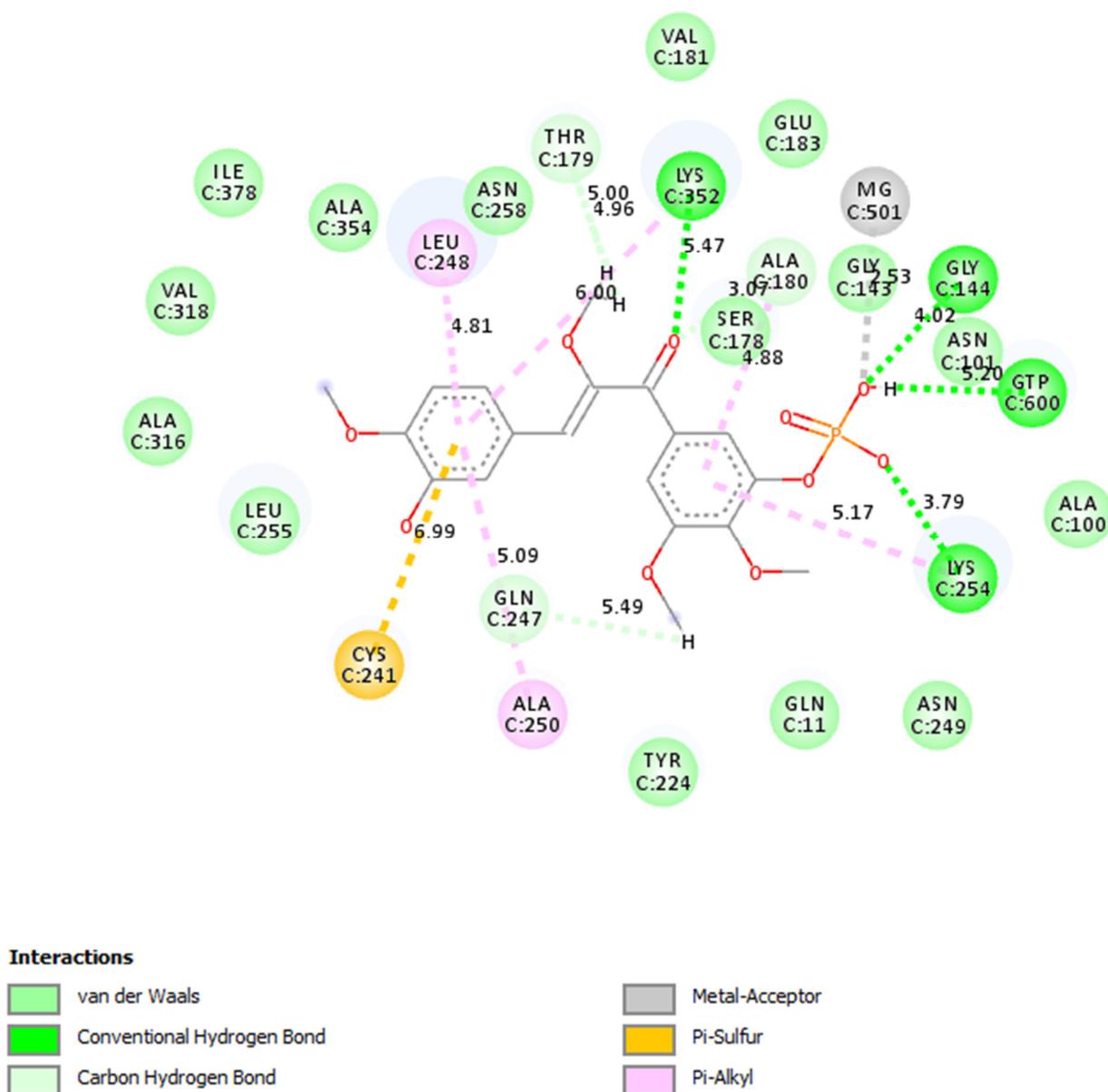
Fonte: Elaborada pelo autor.

Figura 51 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 4a1 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

Figura 52 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 4a9 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

Analisando os complexos formados e comparando com o modo de ligação descrito nas análises anteriores de docagem molecular, pode-se perceber que há grande similaridade entre ambos, ou seja, aquele mesmo modo de ligação pode ser usado para descrever as moléculas que foram propostas posteriormente. Isso indica que as moléculas propostas estão em concordância com as moléculas das quais derivaram, apresentando comportamento semelhante na formação do complexo. Isso

é bastante importante pelo fato de que não é desejável que ocorram mudanças radicais no comportamento do complexo, e sim, uma maior capacidade de interação.

Um grupo que foi adicionado durante a proposição e que não se fazia presente nos compostos inicialmente estudados, é o grupo dihidroxi fosfato. Esse grupo foi incorporado a algumas estruturas no momento em que foi percebido que havia certa disponibilidade para grupos mais volumosos, por se tratar de uma região de interface. Outro motivo para a inclusão desse grupo foi o fato de ele apresentar alta capacidade de interação com resíduos que compõem o sítio de interação do GTP, já que apresentam em comum os grupos fosfato. Como esperado, esse grupo, quando adicionado ao ligante, foi capaz de interagir com alguns resíduos que interagem com o GTP, inclusive, interagiu com o próprio GTP, principalmente quando presentes nas posições R<sup>3</sup> e R<sup>4</sup>. As interações com o GTP se deram a partir de ligações de hidrogênio, onde um átomo de oxigênio do GTP interagia com um dos átomos de hidrogênio do grupo dihidroxi fosfato do ligante. Também ocorreram ligações de hidrogênio do tipo carbono hidrogênio, com ocorrência na região central do ligante, bem como nas posições R<sup>3</sup> e R<sup>4</sup>.

Com a adição do grupo dihidroxi fosfato, se torna possível ter uma noção do quanto o GTP e seus respectivos resíduos interagem com esse grupo adicionado ao ligante. Lembrando que já existe o fármaco fosbretabulina, assim, estamos propondo alguns ligantes que se inspiram em um fármaco já consolidado, onde podemos estudar seus respectivos comportamentos em nossas análises, bem como predizer os valores de suas respectivas atividades biológicas.

As interações do íon magnésio com os ligantes ocorrem, geralmente, nas posições R<sup>3</sup> e R<sup>4</sup> do ligante, na maioria dos casos quando esse ligante apresenta um grupo dihidroxi fosfato, onde ocorre a interação do íon magnésio com um dos átomos de oxigênio do grupo dihidroxi fosfato. Apesar das interações dos ligantes com o íon magnésio serem de baixa ocorrência, vale comentar sobre sua existência, a fim de exemplificar como a presença de um grupo dihidroxi fosfato no ligante pode influenciar a dinâmica do processo de polimerização de microtúbulos, já que esse grupo pode interferir nas interações que o íon magnésio pode ter, por exemplo, com a molécula de GTP.

Após a proposição das novas estruturas e suas respectivas análises nas simulações de docagem molecular, foi realizado o teste de predição das mesmas em cada um dos modelos gerados no trabalho de doutorado: QSAR, HQSAR, QSAR-3D

(CoMFA), QSAR-3D (CoMSIA). Feito isso, os valores preditos de atividade biológica podem ser comparados, a fim que seja verificada a consistência na predição dos modelos e também a capacidade de inibição ( $pIC_{50}$ ) de cada uma das moléculas propostas. Os valores preditos de atividade biológica ( $pIC_{50}$ ) para todos os modelos estão dispostos na Tabela 18.

Tabela 18 – Valores preditos de atividade biológica ( $pIC_{50}$ ) pelos métodos de QSAR (PLS), HQSAR e QSAR 3D (CoMFA e CoMSIA) para as 36 moléculas propostas.

	QSAR	HQSAR	CoMFA	CoMSIA
<b>2a1</b>	4,842	5,287	7,439	8,963
<b>2a2</b>	5,553	5,605	6,985	8,420
<b>2a3</b>	5,373	6,045	8,044	8,257
<b>2a4</b>	6,032	5,474	7,678	8,214
<b>2a5</b>	5,124	4,349	8,393	8,420
<b>2a6</b>	5,123	5,820	8,176	7,754
<b>2a7</b>	5,500	5,925	8,359	8,177
<b>2a8</b>	6,194	5,354	7,278	8,100
<b>3a1</b>	<b>8,041</b>	<b>8,442</b>	<b>9,245</b>	<b>9,361</b>
<b>3a2</b>	<b>7,813</b>	<b>8,046</b>	<b>9,544</b>	<b>9,298</b>
<b>3a3</b>	6,734	6,814	6,660	6,525
<b>3a4</b>	6,547	7,760	8,180	8,977
<b>3a5</b>	6,981	8,340	10,004	9,336
<b>3a6</b>	<b>8,244</b>	<b>7,669</b>	<b>9,404</b>	<b>9,966</b>
<b>3a7</b>	6,884	8,340	9,033	9,406
<b>3a8</b>	6,713	7,669	9,303	9,481
<b>3a9</b>	6,197	8,704	8,967	9,696
<b>3a10</b>	<b>8,622</b>	<b>8,005</b>	<b>7,841</b>	<b>7,675</b>
<b>3a11</b>	6,628	8,892	9,617	8,767
<b>3a12</b>	6,459	8,860	8,854	9,423
<b>3a13</b>	<b>7,135</b>	<b>7,927</b>	<b>9,384</b>	<b>9,838</b>
<b>3a14</b>	<b>7,707</b>	<b>6,914</b>	<b>9,183</b>	<b>9,380</b>
<b>3a15</b>	6,715	7,572	8,039	8,990
<b>4a1</b>	<b>8,018</b>	<b>9,364</b>	<b>9,084</b>	<b>9,102</b>
<b>4a2</b>	<b>8,327</b>	<b>6,971</b>	<b>8,761</b>	<b>9,818</b>
<b>4a3</b>	<b>8,653</b>	<b>8,227</b>	<b>9,489</b>	<b>8,708</b>
<b>4a4</b>	6,568	8,503	9,794	9,575
<b>4a5</b>	6,458	8,007	8,755	9,456

<b>4a6</b>	6,063	9,138	8,740	8,769
<b>4a7</b>	6,177	8,758	8,309	9,136
<b>4a8</b>	6,840	8,332	9,082	8,467
<b>4a9</b>	<b>7,003</b>	<b>7,395</b>	<b>9,366</b>	<b>8,572</b>
<b>4a10</b>	6,379	8,117	7,773	8,542
<b>4a11</b>	5,814	8,635	9,798	9,368
<b>4a12</b>	6,379	8,139	9,290	9,448
<b>4a13</b>	6,875	8,635	9,227	8,782

Dentre as 36 moléculas propostas, podemos destacar as que apresentaram valores elevados de  $pIC_{50}$  em todos os modelos. Foi utilizado como critério de seleção moléculas que apresentaram valores de  $pIC_{50}$  maiores que 7,0 em todos os modelos. Dessa forma, foram selecionadas 10 moléculas com tal característica, que estão apresentadas na Tabela 19. Além disso, nas análises de docagem molecular, as 10 moléculas selecionadas apresentaram um ótimo comportamento, tanto na análise visual quanto em valores de pontuação, que permaneceram na faixa de pontuação acima de 40, conforme observado inicialmente nas análises de docagem, onde todas as moléculas de alta atividade biológica apresentaram valores de pontuação acima de 40. É interessante observar que, dentre as 10 melhores moléculas propostas, 4 apresentam um grupo dihidroxi fosfato em sua composição, mostrando assim a utilidade desse tipo de proposição.

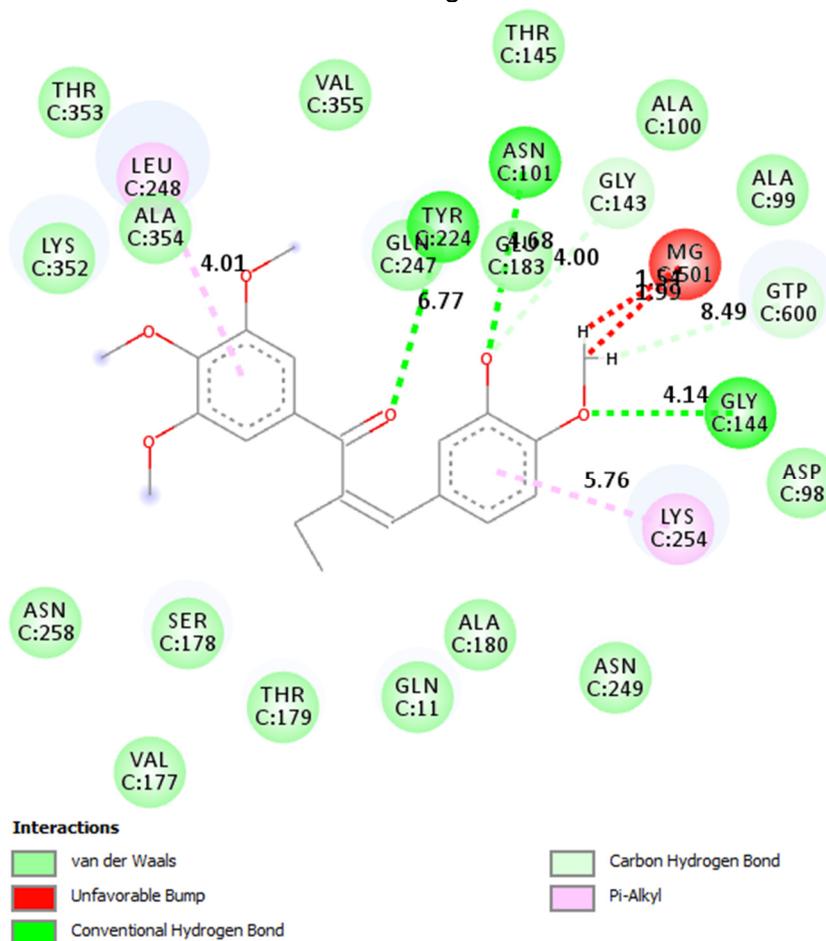
Tabela 19 – Moléculas propostas que apresentaram os melhores valores de atividade biológica ( $pIC_{50}$ ) nas predições de todos os modelos gerados.

	QSAR	HQSAR	CoMFA	CoMSIA
<b>3a1</b>	8,041	8,442	9,245	9,361
<b>3a2</b>	7,813	8,046	9,544	9,298
<b>3a6</b>	8,244	7,669	9,404	9,966
<b>3a10</b>	8,622	8,005	7,841	7,675
<b>3a13</b>	7,135	7,927	9,384	9,838
<b>3a14</b>	7,707	6,914	9,183	9,380
<b>4a1</b>	8,018	9,364	9,084	9,102
<b>4a2</b>	8,327	6,971	8,761	9,818
<b>4a3</b>	8,653	8,227	9,489	8,708
<b>4a9</b>	7,003	7,395	9,366	8,572

Dentre as moléculas propostas, algumas apresentaram comportamento diferenciado das demais, como foi o caso das estruturas 3a1 e 3a2, que durante as análises dos complexos formados, após os cálculos de docagem molecular, foi percebido que esses ligantes se encontravam com seu posicionamento horizontalmente invertido com relação às outras moléculas, cujo modo de ligação já foi descrito. Nos outros complexos, os ligantes apresentavam o seu anel aromático A voltado para a  $\alpha$ -tubulina, e o seu anel aromático B voltado para a  $\beta$ -tubulina. Já nesse caso das moléculas 3a1 e 3a2, o anel aromático que mais interage com a  $\alpha$ -tubulina é o B, e o que interage mais com o  $\beta$ -tubulina é o A, como pode ser verificado nas Figuras 53 e 54. Vale ressaltar que foi realizada uma minuciosa análise dos complexos formados, selecionando as poses mais estatisticamente representativas, para garantir que essa seria a formação mais adequada do complexo nesses casos. É interessante que, mesmo tendo um comportamento atípico na formação do complexo, essas moléculas foram capazes de ter uma boa interação nesse sítio, e também apresentaram bons valores de atividade biológica, quando submetidas aos modelos gerados neste trabalho, permanecendo entre a 10 moléculas propostas de maior valor de atividade biológica, conforme apresentadas na Tabela 19.

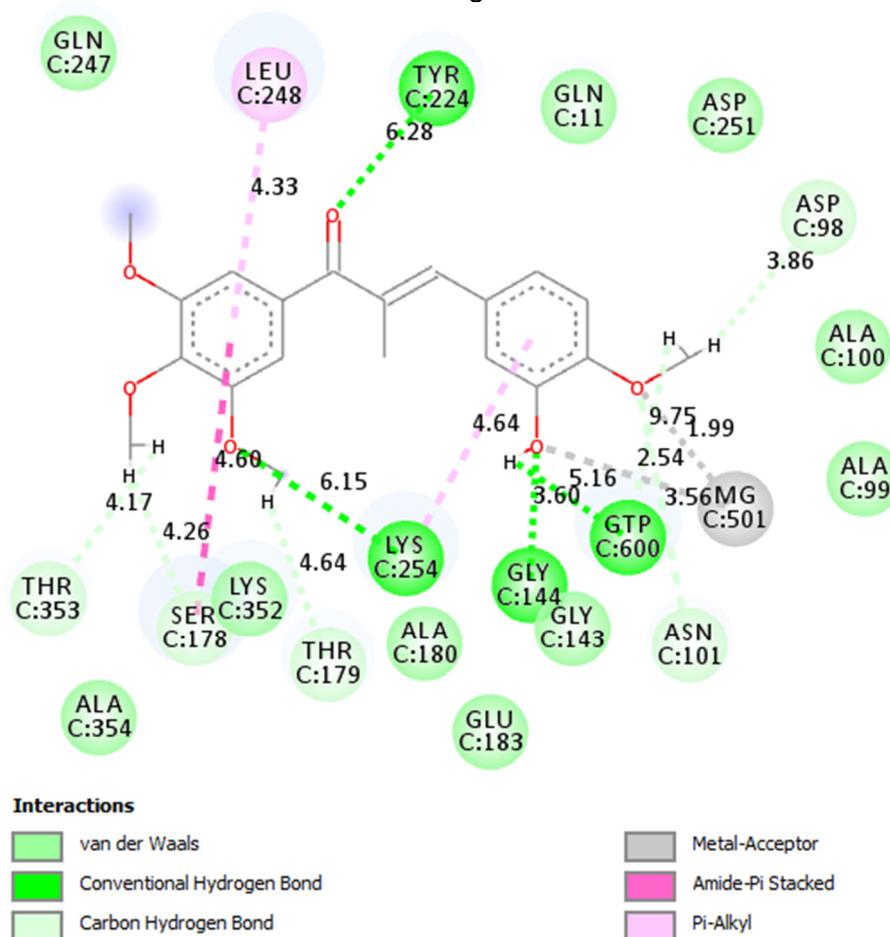
Chama a atenção a capacidade que esses ligantes tiveram de interagir com os resíduos através de ligações de hidrogênio, além das interações hidrofóbicas. Esse comportamento atípico (inversão), seguido de bons resultados nas análises, pode ser compreendido quando levado em consideração o perfil desse tipo de ligante. O equilíbrio entre os seus dois anéis aromáticos, bem como a semelhança dos grupos substituintes nas regiões de  $R^2$  a  $R^6$  e de  $R^{2'}$  a  $R^{6'}$ , acompanhados da boa disponibilidade de interação do átomo de oxigênio localizado na região central da molécula, possibilitaram que não apenas o ligante pudesse se inverter, mas também que apresentasse uma melhor capacidade de interação com os resíduos nessa diferenciada conformação. A consequência dessa inversão foi uma pequena alteração no modo de ligação dessas duas estruturas, onde resíduos como a GLY144, TYR224 e ASN101 passaram a ter uma maior interação com tais ligantes. As estruturas desses resíduos estão apresentadas na Figura 55.

Figura 53 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a1 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

Figura 54 – Diagrama bidimensional do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 3a2 no sítio de ligação da colchicina. Os resíduos de aminoácidos aos quais foi fornecida flexibilidade durante os cálculos de docagem são os mesmos da flexibilidade 2.



Fonte: Elaborada pelo autor.

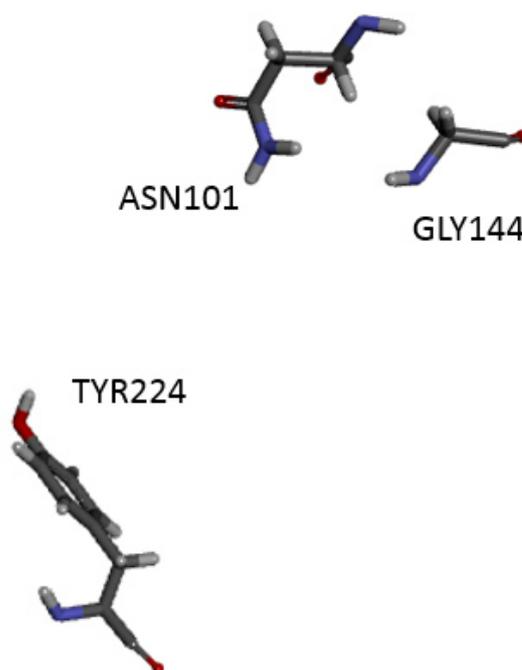
A TYR224 interage com os ligantes a partir de seu grupo hidroxila, que fazia ligações de hidrogênio com o oxigênio da carbonila, tanto para a 3a1 quanto para a 3a2. A LEU248, a partir da região dos átomos de carbono C (G) e C (D), interagiu com ligações do tipo  $\pi$ -alquil com o anel aromático A dos ligantes, em ambos os casos. A LYS254 interagiu de forma semelhante à LEU248, a partir da região dos seus átomos de carbono C (G) e C (D) fazendo ligações do tipo  $\pi$ -alquil, porém interagindo com o anel aromático B dos ligantes, além de também realizar uma ligação de hidrogênio, no caso da molécula 3a2. A GLY144 realizou ligações de hidrogênio a partir do seu grupo amina com os átomos de oxigênio dos grupos metoxila (na 3a1, posição R<sup>4</sup>) e hidroxila (na 3a2, posição R<sup>5</sup>). A ASN101, a partir da amina, fez ligações de hidrogênio com o grupo hidroxila localizado na posição R<sup>5</sup> do ligante 3a1. Já no ligante 3a2,

interagiu com ligações de hidrogênio do tipo carbono hidrogênio do tipo, do átomo de hidrogênio (A) com o grupo metoxila localizado na posição R<sup>4'</sup>.

Esse ligantes interagiram também com a molécula de GTP. No caso do ligante 3a1, essa interação ocorreu a partir de ligações de hidrogênio do tipo carbono hidrogênio, do átomo de oxigênio (3G) do GTP com o um dos átomos de hidrogênio do grupo metoxila, na posição R<sup>4'</sup>. No ligante 3a2, ocorreu a ligação de hidrogênio a partir do átomo de oxigênio (3A) do GTP com o átomo de hidrogênio do grupo hidroxila, na posição R<sup>5'</sup>.

Também houve interações com o íon magnésio. No caso do ligante 3a1, ocorreu uma ligação não favorável desse átomo com átomos de hidrogênio e carbono do grupo metoxila da posição R<sup>4'</sup>. No ligante 3a2 ocorreram duas interações com o íon magnésio. Uma com o átomo de oxigênio do grupo metoxila da posição R<sup>4'</sup>, e outra com o átomo de oxigênio do grupo hidroxila da posição R<sup>5'</sup>. Vale ressaltar que esses ligantes também foram capazes de interagir com resíduos da  $\alpha$ -tubulina.

Figura 55 – Representação dos resíduos GLY144, TYR224 e ASN101, todos pertencentes à  $\alpha$ -tubulina. Para fins de uma melhor compreensão do comportamento, seus respectivos posicionamentos em um dos complexos foram mantidos. Além, de estarem na formação do complexo dos ligantes 3a1 e 3a2, também se fizeram presentes em diversas análises anteriores, porém com uma menor frequência que os resíduos já citados anteriormente.



Fonte: Elaborada pelo autor.

## 5. CONCLUSÃO E PERSPECTIVAS

Como pôde ser verificado, tanto o modelo de PLS quanto o de ANN apresentaram boa capacidade preditiva, confirmando assim a importância desses métodos, já que tais modelos se mostraram com maior capacidade de predição se comparado aos modelos de HQSAR e QSAR 3D. Ainda que os modelos de HQSAR e QSAR 3D não tenham tido os mesmos resultados que os de QSAR, foram bastante úteis, pois auxiliaram na predição da atividade biológica das moléculas propostas, as quais foram submetidas a cada um dos modelos e, com suas particularidades, puderam avaliar diversas características dessas moléculas, indicando quais moléculas se mostrariam de alta atividade biológica em cada um dos modelos testados.

Além de se ter modelos com boa capacidade preditiva, a elucidação das principais interações que ocorrem no complexo formado pelos ligantes e o seu sítio de ligação foram de fundamental importância para compreender quais características poderiam ser melhoradas nesse ligantes, bem como os grupos substituintes mais propícios a interagirem naquele sítio de ligação. Para isso, as análises de docagem molecular tiveram grande importância, e a riqueza de detalhes que podem ser obtidos nessas análises são informações que garantem maior segurança nas proposições.

As novas estruturas propostas neste trabalho nos mostram que, mesmo com sutis alterações nos grupos substituintes, podem ocorrer mudanças bastante significativas na capacidade de interação do ligante no complexo, e conseqüentemente, no valor da atividade biológica. A partir das estruturas aqui propostas, abrem-se muitas possibilidades, tanto de aprofundamento de análises e de proposições, quanto de utilização de metodologia semelhante em outros trabalhos.

Alguns trabalhos têm sido publicados nesse sentido em nosso grupo de pesquisa, que apresentam pleno potencial de aprimoramento e de aplicabilidade no que diz respeito à proposição de novas estruturas, já que, nos últimos anos, as metodologias relacionadas às análises de QSAR foram aprimoradas, principalmente a de ANN.<sup>106,107,108,109,110</sup>

Para andamento deste trabalho, seriam bastante úteis análises como a de dinâmica molecular, principalmente pelas características do complexo, e da presença de outros ligantes numa região próxima, como é o caso da molécula de GTP.

## REFERÊNCIAS

- 1 LODISH, H.; BERK, A.; ZIPURSKI, S. L.; MATSUDAIRA, P.; BALTIMORE, D.; DARNELL, J. **Molecular cell biology**. 5 ed. New York: W. H. Freeman, 2000.
- 2 KERSSEMAKERS, J. W.; MUNTEANU, E. L.; LAAN, L.; NOETZEL, T. L.; JANSON, M. E.; DOGTEROM, M. Assembly dynamics of microtubules at molecular resolution. **Nature**, London, v. 442, n. 7103, p. 709-712, 2006.
- 3 ORGANIZAÇÃO PAN-AMERICANA DA SAÚDE. **Câncer**. Brasília: OPAS, 2018. Disponível em: [https://www.paho.org/bra/index.php?option=com\\_content&view=article&id=5588:folha-informativa-cancer&Itemid=1094](https://www.paho.org/bra/index.php?option=com_content&view=article&id=5588:folha-informativa-cancer&Itemid=1094). Acesso em: 15 ago. 2020.
- 4 INSTITUTO NACIONAL DO CÂNCER. **Estimativa 2020**: Brasil – estimativa de novos casos. Rio de Janeiro: INCA, 2020. Disponível em: <https://www.inca.gov.br/estimativa/estado-capital/brasil>. Acesso em 16 ago. 2020.
- 5 ALBERTS, B.; JOHNSON, A.; LEWIS, J.; RAFF, M.; ROBERTS, K.; WALTER, P. **Molecular biology of the cell**. 6 ed. New York: Garland Science, 2015.
- 6 LAZO, J. S.; TAKAMURA, K.; VOGT, A.; JUNG, J. K.; RODRIGUEZ, S.; BALACHANDRAN, R.; DAY, B. W.; WIPF, P. Antimitotic actions of a novel analog of the fungal metabolite palmarumycin CP1. **Journal of Pharmacology and Experimental Therapeutics**, Baltimore, v. 296, n. 2, p. 364-371, 2001.
- 7 JORDAN, A.; HADFIELD, J. A.; LAWRENCE, N. J.; MCGOWN, A. T. Tubulin as a target for anticancer drugs: agents which interact with the mitotic spindle. **Medicinal Research Reviews**, New York, v. 18, n. 4, p. 259-296, 1998.
- 8 WILSON, L.; JORDAN, M. A. Microtubules as a target for anticancer drugs. **Nature**, London, v. 4, p.252-265, 2004.
- 9 BERGSTRALH, D. T.; TING, J. P. Microtubule stabilizing agents: their molecular signaling consequences and the potential enhancement drug combination. **Cancer Treatment Reviews**, New York, v. 32, n. 3, p. 166-179, 2006.
- 10 CONDE, C.; CÁCERES, A. Microtubule assembly, organization and dynamics in axons and dendrites. **Nature Reviews Neuroscience**, London, v. 10, p. 319-332, 2009.
- 11 DESAI, A.; MITCHISON, T. J. Microtubule polymerization dynamics. **Annual Review of Cell and Developmental Biology**, Palo Alto, v.13, p. 83-117, 1997.
- 12 GUPTA, S.; BHATTACHARYYA, B. Antimicrotubular drugs binding to vinca domain of tubulin. **Molecular and Cellular Biochemistry**, The Hague, v. 253, n. 1-2, p. 41-47, 2003.

13 INSTITUTO NACIONAL DO CÂNCER. **Leucemia**. Rio de Janeiro: INCA, 2020. Disponível em: <https://www.inca.gov.br/tipos-de-cancer/leucemia>. Acesso em: 25 ago. 2020.

14 SAWYERS, C. L. Molecular genetics of acute leukaemia. **The Lancet**, New York, v. 349, p. 196-200, 1997.

15 HAMERSCHLAK, N.; SOUZA, C.; CORNACCHIONI, A. L.; PASQUINI, R.; TAKAK, D.; SPECTOR, N.; STEAGALL, M. Quality of life chronic myeloid leukemia patients in Brazil: ability to work as a key factor. **Support Care Cancer**, Berlin, v. 22, p. 2113-2118, 2014.

16 BAGG, A. Chronic myeloid leukemia. **Journal of Molecular Diagnostics**, Bethesda, v. 4, n. 1, p. 1-10, 2002.

17 ESFAHANI, M. K.; MORRIS, E. L.; DUTCHER, J. P.; WIERNIK, P. H. Blastic phase of chronic myelogenous leukemia. **Current Treatment Options in Oncology**, Philadelphia, v. 7, p. 189-199, 2006.

18 TEFFERI, A.; THIELE, J.; ORAZI, A.; KVASNICKA, H. M.; BARBIU, T.; HANSON, C. A.; BAROSI, G.; VERSTOVSEK, S.; BIRGEGARD, G.; MESA, R.; REILLY, J. T.; GISSLINGER, H.; VANNUCCHI, A. M.; CERVANTES, F.; FINAZZI, G.; HOFFMAN, R.; GILLILAND, D. G.; BLOOMFIELD, C. D.; VARDIMAN, J. W. Proposals and rationale for revision of the World Health Organization diagnostic criteria for polycythemia vera, essential thrombocythemia, and primary myelofibrosis: recommendations from an ad hoc international expert panel. **Blood**, New York, v. 110, n. 4, p. 1092-1097, 2007.

19 SILVER, R. T. The blast phase of chronic myeloid leukaemia. **Best Practice & Research Clinical Haematology**, London, v. 22, p. 387-394, 2009.

20 LOZZIO, C. B.; LOZZIO, B. B. Human chronic myelogenous leukemia cell-line with positive Philadelphia chromosome. **Blood**, New York, v. 45, n. 3, p. 321-334, 1975.

21 TSIFTSOGLU, A. S.; PAPPAS, I. S.; VIZIRIANAKIS, I. S. Mechanisms involved in the induced differentiation of leukemia cells. **Pharmacology & Therapeutics**, Oxford, v. 100, p. 257-290, 2003.

22 GIGANT, B.; WANG, C.; RAVELLI, R. B. G.; ROUSSI, F.; STEINMETZ, M. O.; CURMI, P. A.; SOBEL, A.; KNOSSOW, M. Structural basis for the regulation of tubulin by vinblastine. **Nature**, London, v. 435, n. 7041, p. 519-522, 2005.

23 BISWAS, B. B.; SEN, K.; GHOSH CHOUDHURY, G.; BHATTACHARYYA, B. Molecular biology of tubulin: its interaction with drugs and genomic organization. **Journal of Biosciences**, Bangalore, v. 6, n. 4, p. 341-457, 1984.

24 BAI, R.; COVELL, D. G.; PEI, X. F.; EWELL, J. B.; NGUYEN, N. Y.; BROSSI, A.; HAMEL, E. Mapping the binding site of colchicinoids on beta-tubulin. 2-chloroacetyl-2-demethylthiocolchicine covalently reacts predominantly with cysteine 239 and secondarily with cysteine 354. **Journal of Biological Chemistry**, Baltimore, v. 275, n. 51, p. 40443-40452, 2000.

25 DUCKI, S.; RENNISON, D.; WOO, M.; KENDALL, A.; CHABERT, J. F. D.; McGOWN, A. T.; LAWRENCE, N. Combretastatin-like chalcones as inhibitors of microtubule polymerization. Part 1: synthesis and biological evaluation of antivasular activity. **Bioorganic & Medicinal Chemistry**, Oxford, v. 17, p. 7698-7710, 2009.

26 OLIVEIRA, F. C. S.; BARROS, R. F. M.; MOITA NETO, J. M. Plantas medicinais utilizadas em comunidades rurais de Oeiras, semiárido piauiense. **Revista Brasileira de Plantas Mediciniais**, Botucatu, v. 12, p.282-301, 2010.

27 LOPES, L. S.; MARQUES, R. B.; PEREIRA, S. S.; AYRES, M. C.; CHAVES, M. H.; CAVALHEIRO, A. J.; VIEIRA JR, G. M.; ALMEIDA, F. R. C. Antinociceptive effect on mice of the hydroalcoholic fraction and epicatechin obtained from *Combretum leprosum* Mart & Eich. **Brazilian Journal of Medicinal and Biological Research**. Ribeirão Preto, v. 43, p. 1184-1192, 2010.

28 PETTIT, G. R.; SINGH, S. B.; HAMEL, E.; LIN, C. M.; ALBERTS, D. S.; KENDALL, D. G. Isolation and structure of the strong cell growth and tubulin inhibitor combretastatin A-4. **Experientia**, Basel, v. 45, n. 2, p. 209-211, 1989.

29 PETTIT, G. R.; SINGH, S. B.; BOYD, M. R.; HAMEL, E.; PETTIT, R. K.; SCHMIDT, J. M.; HOGAN, F. Antineoplastic agents. 291. Isolation and synthesis of combretastatins A-4, A-5 and A-6. **Journal of Medicinal Chemistry**, Washington, v. 38, p. 1666-1672, 1995.

30 PETTIT, G. R.; LIN, C. M.; SINGH, S. B.; CHU, P. S.; DEMPCY, R. O.; SCHMIDT, J. M.; HAMEL, E. Interactions of tubulin with potent natural and synthetic analogs of the antimetabolic agent combretastatin: a structure-activity study. **Molecular Pharmacology**, New York, v. 34, p. 200-208, 1988.

31 PETTIT, G. R.; SINGH, S. B.; SCHMIDT, J. M. Isolation, structure, synthesis and antimetabolic properties of combretastatins B-3 and B-4 from *Combretum Caffrum*. **Journal of Natural Products**, Cincinnati, v. 51, p. 517-527, 1988.

32 PETTIT, G. R.; CRAGG, G. M.; HERALD, D. L.; SCHMIDT, J.M.; LOHAVANUAYA, P. Isolation and structure of combretastatin. **Canadian Journal of Chemistry**, Ottawa, v. 60, N. 11, P.1374-1376, 1982.

33 PETTIT, G. R.; SINGH, S. B. Isolation, structure, and synthesis of combretastatin A-2, A-3 and B-2. **Canadian Journal of Chemistry**, Ottawa, v. 65, 2390, 1987.

34 NAM, N. H. Combretastatin A-4 analogues as antimetabolic antitumor agents. **Current Medicinal Chemistry**, Schiphol. v. 10, p. 1697-1722, 2003.

- 35 DORR, R. T.; DVORAKOVA, K.; SNEAD, K.; ALBERTS, D. S.; SALMON, S. E.; PETTIT, G. R. Antitumor activity of combretastatin-A4 phosphate, a natural product tubulin inhibitor. **Investigational New Drugs**, Boston, v. 14, p. 131-137, 1996.
- 36 OHSUMI, K.; NAKAGAWA, R.; FUKUDA, Y.; HATANAKA, T.; MORINAGA, Y.; NIHEI, Y.; OHISHI, K.; SUGA, Y.; AKIYAMA, Y.; TSUJI, T. Novel combretastatin analogues effective against murine solid tumors: design and structure – activity relationships. **Journal of Medicinal Chemistry**, Washington, v. 41, p. 3022-3032, 1998.
- 37 WANG, L.; WOODS, K. W.; LI, Q.; BARR, K. J.; McCROSKEY, R. W.; HANNICK, S. M.; GHERKE, L.; CREDO, R. B.; HUI, Y. H.; MARSH, K.; WARNER, R.; LEE, J. Y.; MOZNG, N. Z.; FROST, D.; ROSENBERG, S. H.; SHAM, H. L. Potent, orally active heterocycle-based combretastatin A-4 analogues: synthesis, structure – activity relationship, pharmacokinetics, and in vivo antitumor activity evaluation. **Journal of Medicinal Chemistry**, Washington, v. 45, p. 1697-1711, 2002.
- 38 McGOWN, A. T.; FOX, B. W. Differential cytotoxicity of combretastatins A1 and A4 in two daunorubicin-resistant P388 cell lines. **Cancer Chemotherapy and Pharmacology**, Berlin, v. 26, p.79-81, 1990.
- 39 RAHMAN, M. A. Chalcone: a valuable insight into the recent advances and potential pharmacological activities. **Chemical Sciences Journal**, Birmingham, CSJ-29, 2011.
- 40 ASSOCIAÇÃO BRASILEIRA DE LINFOMA E LEUCEMIA. **Tratamento LMC**. São Paulo: ABRALÉ, 2017. Disponível em: <http://www.abrale.org.br/lmc/tratamento>. Acesso em: 22 ago. 2020.
- 41 GRINTER, S. Z.; ZOU, X. Challenges, applications, and recent advances of protein-ligand docking in structure-based drug discovery. **Molecules**, Basel, v. 19, p. 10150-10176, 2014.
- 42 WANG, R.; GAO, Y.; LAI, L. LigBuilder: a multi-purpose program for structure-based drug design. **Journal of Molecular Modeling**, Berlin, v. 6, p. 498-516, 2000.
- 43 BENSON, D. A.; KARSCH-MIZRACHI, I.; LIPMAN, D. J.; OSTELL, J.; WHEELER, D. L. GenBank. **Nucleic Acids Research**, London, v. 36, p. 25-30, 2007.
- 44 DIMASI, J. A.; GRABOWSKI, H. G.; HANSEN, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. **Journal of Health Economics**, Amsterdam, v. 47, p. 20-33, 2016.
- 45 HUNG, C.; CHEN, C. Computational approaches for drug discovery. **Drug Development Research**, New York, v. 75, p. 412-418, 2014.
- 46 LIPINSKI, C. A.; LOMARDO, F.; DOMINY, B. W.; FEENEY, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, Amsterdam, v. 46, p. 3-26, 2001.

- 47 KONTOYIANNI, M.; McCLELLAN, L. M.; SOKOL, G. S. Evaluation of docking performance: comparative data on docking algorithms. **Journal of Medicinal Chemistry**, Washington, v. 47, p. 558-565, 2004.
- 48 JORGENSEN, W. L. Pulled from a protein's embrace. **Nature**, London, v. 466, p. 42-43, 2010.
- 49 YUAN, Y.; PEI, J.; LAI, L. LigBuilder 2: a practical de novo drug design approach. **Chemical Information and Modeling**, Washington, v. 51, p. 1083-1091, 2011.
- 50 RCSB Protein Data Bank. [S. l.: s. n.], 2017. Disponível em: <https://www.rcsb.org>. Acesso em: 10 mai. 2017.
- 51 McCARTHY, D. J. Computational approaches to structure-based ligand design. **Pharmacology & Therapeutics**, Oxford, v.84, p. 179-191, 1999.
- 52 BRANDL, M.; WEISS, M. S.; JABS, A.; SUHNEL, J.; HILGENFELD, R. C-H $\cdots$  $\pi$ -Interactions in proteins. **Journal of Molecular Biology**, London, v. 307, p. 357-377, 2001.
- 53 GOHLKE, H.; KLEBE, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. **Angewandte Chemie International Edition**, Weinheim, v. 41, p. 2644-2676, 2002.
- 54 CARACELLI, I.; ZUKERMAN, J. S. **Introdução à biofísica estrutural**. São Carlos: EdUFSCar, 2006. 69p.
- 55 JONES, G.; WILLETT, P.; GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. **Journal of Molecular Biology**, London, v. 245, p. 43-53, 1995.
- 56 JONES, G. J.; WILLETT, P.; GLEN, R. C.; LEACH, A. R.; TAYLOR, R. Development and validation of a genetic algorithm for flexible docking. **Journal of Molecular Biology**, London, v. 267, p. 727-748, 1997.
- 57 WILLETT, P. Genetic algorithms in molecular recognition and design. **Trends Biotechnology**, Amsterdam, v. 13, p. 516-521, 1995.
- 58 KITCHEN, D. B.; DECORNEZ, H.; FURR, J. R.; BAJORATH, J. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature**, London, v. 3, p. 935-949, 2004.
- 59 SOLMAJER, T.; ZUPAN, J. Optimization algorithms and natural computing in drug discovery. **Drug Discovery Today: technologies**, London, v. 1, n. 3, p. 247-252, 2004.
- 60 FORREST, S. Genetic algorithms: principles of natural selection applied to computation. **Science**, Washington, v. 261, p. 872-878, 1993.

- 61 KUBINNYI, H. **QSAR: Hansch analysis and related approaches**. New York: VHC, 1993. 240p.
- 62 CRAMER, C. J. **Essentials of computational chemistry**. 2 ed. Chichester: John Wiley & Sons, 2004.
- 63 OLIVEIRA, D. B.; GAUDIO, A. C. BuildQSAR: a new computer program for QSAR analysis. **Quantitative Structure-Activity Relationships**, New York, v. 19, p. 599-601, 2000.
- 64 BEEBE, K. R.; PEEL, R. J.; SEASHOLTZ, M.B. **Chemometrics: a practical guide**. New York: John Wiley, 1998. 348p.
- 65 KIRALJ, R.; FERREIRA, M. M. C. Basic validation procedures for regression models in QSAR and QSPR studies: theory and application. **Journal of the Brazilian Chemical Society**, São Paulo, v. 20, n. 4, p. 770–787, 2009.
- 66 GAUDIO, A. C.; ZADONA, E. Preposição, validação e análise dos modelos que correlacionam estrutura química e atividade biológica. **Química Nova**, São Paulo, n. 5, v. 24, 2001.
- 67 DOROFKI, M.; ELSHAFIE, A.; JAAFAR, O.; KARIM, O. A. Comparison of artificial neural network transfer functions abilities to simulate extreme runoff data. **IPCBE**, Singapura, v. 33, p. 39-44, 2012.
- 68 PASTI, R.; DE CASTRO, L. N. An immune and a gradient-based method to train multi-layer perceptron neural networks. *In*: INTERNATIONAL JOINT CONFERENCE ON NEURAL NETWORKS, 2006, Vancouver. **Proceedings** [...]. [S. l.]: IEEE, 2006.
- 69 CASTRO, L. N.; VON ZUBEN, F. J. **Optimized training techniques for feedforward neural networks**. Technical report, DCA-RT 03/98. Campinas, UNICAMP, 1998.
- 70 HASANI, M.; EMAMI, F. Evaluation of feed-forward back propagation and radial basis function neural networks in simultaneous kinetic spectrophotometric determination of nitroaniline isomers. **Talanta**, London, v. 75, p. 116-126, 2008.
- 71 NAUMANN, T.; LOWIS, D. HQSAR: a new, highly predictive QSAR technique. *In*: INTERNATIONAL ELECTRONIC CONFERENCE ON SYNTHETIC ORGANIC CHEMISTRY, 1., Basel, 1997. **Proceedings** [...]. Basel: MDPI, 1997. p. 1-8.
- 72 AVERY, M. A.; ALVIM-GASTOM, M.; RODRIGUES, C. R.; BARREIRO, E. J.; COHEN, F. E.; SABNIS, Y. A.; WOOLFREY, J. R. Structure-activity relationships of the antimalarial agent artemisinin. 6. The development of predictive in vitro potency models using CoMFA and HQSAR methodologies. **Journal of Medicinal Chemistry**, Washington, v. 45, p. 292-303, 2002.
- 73 HQSAR™ manual, SYBYL-X 1.2. St. Louis: Tripos Inc., 2010.

74 HONORIO, K. M.; GARRATT, R. C.; ANDRICOPULO, A. D. Hologram quantitative structure-activity relationships for a series of farnesoid X receptor activators. **Bioorganic & Medicinal Chemistry Letters**, Oxford, v. 15, p. 3119-3125, 2005.

75 CRAMER, R. D.; PATTERSON, D. E.; BUNCE, J. D. Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. **Journal of the American Chemical Society**, Washington, v.110, p. 5959-5967, 1988.

76 KUBINYI, H. QSAR and 3D QSAR in drug design. Part 1: methodology. **Drug Discovery Today**, Oxford, v. 2, p. 457-467, 1997.

77 ARAKAWA, M.; HASEGAWA, K.; FUNATSU, K. The recent trend in QSAR modeling – variable selection and 3D QSAR methods. **Current Computer-Aided Drug Design**, Amsterdam, v. 3, p. 254-262, 2007.

78 POTEKIN, A. V.; GRISHINA, M. A.; POTEKIN, V. A. Grid-based continual analysis of molecular interior for drug discovery, QSAR and QSPR. **Current Drug Discovery Technologies**, Amsterdam, v. 14, p. 181-205, 2017.

79 KLEBE, G.; ABRAHAM, U.; MIETZNER, T. Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. **Journal of Medicinal Chemistry**, Washington, v. 37, p. 4130-4146, 1994.

80 FRISCH, M. J.; TRUCKS, G. W.; SCHLEGEL, H. B.; SCUSERIA, G. E.; ROBB, M. A.; CHEESEMAN, J. R.; SCALMANI, G.; BARONE, V.; MENNUCCI, B.; PETERSSON, G. A.; NAKATSUJI, H.; CARICATO, M.; LI, X. HRATCHIAN, H.P.; IZMAYLOV, A. F.; BLOINO, J.; ZHENG, G.; SONNENBERG, J. L.; HADA, M.; EHARA, M.; TOYOTA, K.; FUKUDA, R.; HASEGAWA, J.; ISHIDA, M.; NAKAJIMA, T.; HONDA, Y.; KITAO, O.; NAKAI, H.; VREVEN, T.; MONTGOMERY, JR.; J. A.; PERALTA, J. E.; OGLIARO, F.; BEARPARK, M.; HEYD, J. J.; BROTHERS, E.; KUDIN, K N.; STAROVEROV, V. N.; KOBAYASHI, R.;NORMAND, J.; RAGHAVACHARI, K.; RENDELL, A.; BURANT, J. C.; IYENGAR, S.S.; TOMASI, J.; COSSI, M.; REGA, N.; MILLAM, N. J.; KLENE, M.; KNOX, J. E.; CROSS, J. B.; BAKKEN, V.; ADAMO, C.; JARAMILLO, J.; GOMPERS, R.; STRATMANN, R. E.; YAZYEV, O.; AUSTIN, A. J.; CAMMI, R.; POMELLI, C.; OCHTERSKI, J. W.; MARTIN, R. L.; MOROKUMA, K.; ZAKRZEWSKI, V. G.; VOTH, G. A.; SALVADOR, P.; DANNENBERG, J. J.; DAPPRICH,S.; DANIELS, A.D.; FARKAS,O.; FORESMAN, J. B.; ORTIZ, J. V.; CIOSLOWSKI, J.;FOX, D. J. **Gaussian**. Wallingford: Gaussian, 2009.

81 LEE, C.; YANG, W.; PARR, R. G. **Physical Review B**, College Park, n. 37, v. 785, 1988.

82 POOPLE, J. A.; SEEGER, R.; BINKLEY, J. S.; KRISHNAN, R. Self-consistent molecular orbital methods. XX. A basis set for correlated wavefunctions. **Journal of Chemical Physics**, New York, v. 72, 650, 1980.

- 83 VCCLAB. Virtual computational chemistry laboratory. Munich: [s. n.], 2005. Disponível em: <http://www.vcclab.org>. Acesso em 12 ago. 2018.
- 84 OLIVEIRA, D. B.; GAUDIO, A. C. BuildQSAR: a new computer program for QSAR analysis. **Quantitative Structure-Activity Relationships**, New York, v. 19, n. 6, p. 599–601, 2001.
- 85 INFOMETRIX INC. **Pirouette 3.11**. Woodinville, 2002.
- 86 MARTINS, J. P. A.; FERREIRA, M. M. C. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. **Química Nova**, São Paulo, v. 36, n. 4, p. 554–560, 2013.
- 87 MATHWORKS. **Matlab 7.12**. [S. l.], 2011.
- 88 SHAMSARA, J. SHAHIR-SADR, A. A predictive HQSAR model for a series of tricycle core containing MMP-12 inhibitors with dibenzofuran ring. **International Journal of Medicinal Chemistry**, Cairo, v. 2014, p. 1-10, 2014.
- 89 GARCIA, T. S.; HONÓRIO, K. M. Two-dimensional quantitative structure-activity relationship studies on bioactive ligands of peroxisome proliferator-activated receptor  $\delta$ . **Journal of the Brazilian Chemical Society**, São Paulo, v. 22, p.65-72, 2011.
- 90 HONÓRIO, K. M.; SALUM, L. B.; GARRAT, R. C.; POLIKARPOV, I.; ANDRICOPULO, A. D. Two- and three-dimensional quantitative structure-activity relationships studies on a series of liver X receptor ligands. **The Open Medicinal Chemistry Journal**, Hilversum, v. 2, p. 86-96, 2008.
- 91 CHAUDHAERY, S. S.; ROY, K. K.; SAXENA, A. K. Consensus superiority of the pharmacophore-based alignment, over maximum common substructure (MCS): 3D-QSAR studies on carbamates as acetylcholinesterase inhibitors. **Journal of Chemical Information and Modeling**, Washington, v. 49, p. 1590-1601, 2009.
- 92 XU, M.; ZHANG, A.; HAN, S.; WANG, L. Studies of 3D-quantitative structure-activity relationships on a set of nitroaromatic compounds: CoMFA, advanced CoMFA and CoMSIA. **Chemosphere**, Oxford, v. 48, p. 707-715, 2002.
- 93 RCSB Protein Data Bank. **Code 1SA0**. [S. l.: s. n.], 2017. Disponível em: <https://www.rcsb.org/pdb/explore/explore.do?structureId=1sa0>. Acesso em: 10 maio 2017.
- 94 JONES, G. J.; WILLETT, P.; GLEN, R. C.; LEACH, A. R.; TAYLOR, R. Development and validation of a genetic algorithm for flexible docking. **Journal of Molecular Biology**, London, v. 267, p. 727-748, 1997.
- 95 DASSAULT SYSTÈMES BIOVIA. **Discovery studio modeling environment**. San Diego: Dassault Systèmes, 2016.

96 VERDONK, M. L.; CHESSARI, G.; COLE, J. C.; HARTSHORN, M. J.; MURRAY, C. W.; NISSINK, J. W. M.; TAYLOR, R. D.; TAYLOR, R. Modeling water molecules in protein-ligand docking using GOLD. **Journal of Medicinal Chemistry**, Washington, v. 48, p. 6504-6515, 2005.

97 FERREIRA, M. M. C.; KIRALJ, R. Métodos Quimiométricos em Relações Quantitativas Estrutura-Atividade (QSAR). *In*: MONTANARI, C. A. **Química medicinal: métodos e fundamentos em planejamento de fármacos**. São Paulo: Edusp, 2011. 732p.

98 MADHAVI, S.; SREENIVASULU, R.; ANSARI, M. Y.; AHSAN, M. J.; RAJU, R. R. Synthesis, biological evaluation and molecular docking studies of pyridine incorporated chalcone derivatives as anticancer agents. **Letters in Organic Chemistry**, Sharjah, v. 13, p. 682-692, 2016.

99 HUCZYNSKI, A.; MAJCHER, U.; MAJ, E.; WIETRZYK, J.; JANCZAK, J.; MOSHARI, M.; TUSZYNSKI, J. A.; BARTL, F. Synthesis, antiproliferative activity and molecular docking of colchicine derivatives. **Bioorganic Chemistry**, New York, v. 64, p. 103-112, 2016.

100 KUMBHAR, B. V.; BOROGAON, A.; PANDA, D.; KUNWAR, A. Exploring the origin of differential binding affinities of human tubulin isotypes  $\alpha\beta$ II,  $\alpha\beta$ III and  $\alpha\beta$ IV for DAMA-colchicine using homology modelling, molecular docking and molecular dynamics simulations. **PLOS ONE**, San Francisco, p. 1-22, 2016.

101 AHMED, R. I.; OSMAN, E. E. A.; AWADALLAH, F. M.; EL-MAGHAZY, S. M. Design, synthesis and molecular docking of novel diarylcyclohexene and diaryllindazole derivatives as tubulin polymerization inhibitors. **Journal of Enzyme Inhibition and Medicinal Chemistry**, London, v. 32, n.1, p. 176-188, 2017.

102 PATEL, V. K.; RAJAK, H. Synthesis, biological evaluation and molecular docking studies of 2-amino-3,4,5-trimethoxyaroylindole derivatives as novel anticancer agents. **Bioorganic & Medicinal Chemistry Letters**, Oxford, v. 26, p. 2115-2118, 2016.

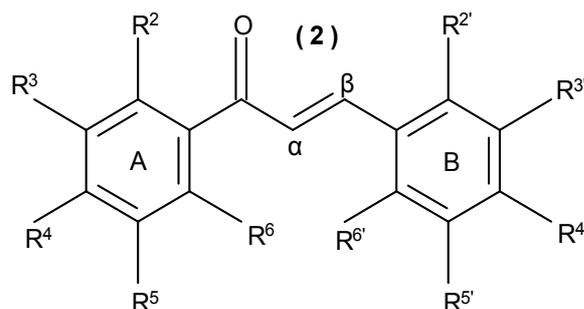
103 AMERI, A.; KHODARAHMI, G.; HASSANZADEH, F.; FOROOTANFAR, H.; HAKIMELAHI, G. Novel aldimine-type Schiff bases of 4-amino-5[(3,4,5-trimethoxyphenyl)methyl]-1,2,4,-triazole-3-thione/thiol: docking study, synthesis, biological evaluation, and anti-tubulin activity. **Archiv der Pharmazie Journal**, Weinheim, v. 349, p. 662-681, 2016.

104 WAKASKAR, R. R.; BATHENA, S. P. R.; TALLAPAKA, S. B.; AMBARDEKAR, V. V.; GAUTAM, N.; THAKARE, R.; SIMET, S. M.; CURRAN, S. M.; SINGH, R. K.; DONG, Y.; VETRO, J. A. Peripherally cross-linking the shell of core-shell polymer micelles decreases premature release of physically loaded combretastatin A4 in whole blood and increases its mean residence time and subsequent potency against primary murine breast tumors after IV administration. **Pharmaceutical Research**, New York, v. 32, p. 1028-1044, 2015.

- 105 NACIONAL CANCER INSTITUTE. **Fosbretabulin disodium (Code C2503)**. [S. l.]: NCI, 2020. Disponível em: [https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI\\_Thesaurus&ns=NCI\\_Thesaurus&code=C2503](https://ncit.nci.nih.gov/ncitbrowser/ConceptReport.jsp?dictionary=NCI_Thesaurus&ns=NCI_Thesaurus&code=C2503). Acesso em: 05 out. 2020.
- 106 LIPINSKI, C. F.; OLIVEIRA, A. A.; HONORIO, K. M.; OLIVEIRA, P. R.; DA SILVA, A. B. F. A molecular modeling study of combretastatin-like chalcones as anticancer agents using PLS, ANN and consensus models. **Structural Chemistry**, New York, v. 29, p. 957-965, 2018.
- 107 LIPINSKI, C. F.; MATAROLLO, V.; OLIVEIRA, P. R.; DA SILVA, A. B. F.; HONORIO, K. M. Advances and perspectives in applying Deep Learning for drug design and discovery. **Frontiers in Robotics and AI**, Lausanne, v. 6, n. 108, 2019.
- 108 OLIVEIRA, A. A.; LIPINSKI, C. F.; PEREIRA, E. B.; HONORIO, K. M.; OLIVEIRA, P. R.; WEBER, K. C.; ROMERO, R. A. F.; DE SOUZA, A. G.; DA SILVA, A. B. F. New consensus multivariate models based on PLS and ANN studies of sigma-1 receptor antagonists. **Journal of Molecular Modeling**, Berlin, v. 23, n. 302, 2017.
- 109 CHIARI, L. P. A.; DA SILVA, A. P.; OLIVERIA, A. A.; LIPINSKI, C. F.; HONORIO, K. M.; DA SILVA, A. B. F. Drug design of new sigma-1 antagonists against neuropathic pain: a QSAR study using partial least squares and artificial neural networks. **Journal of Molecular Structure**, Amsterdam, v. 1223, id. 129156, 2021.
- 110 SANTOS, G. R.; CHIARI, L. P. A.; DA SILVA, A. P.; LIPINSKI, C. F.; OLIVEIRA, A. A.; HONORIO, K. M.; SOUZA, A. G.; DA SILVA, A. B. F. A partial least squares and artificial neural network study for a series of arylpiperazines as antidepressant agents. **Journal of Molecular Modeling**, Berlin, v. 27, n. 10, p. 297, 2021.

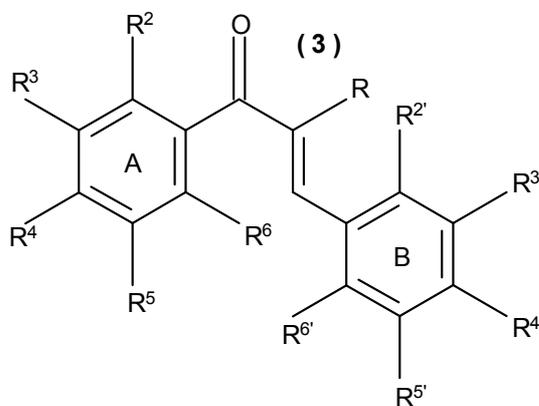
## APÊNDICES

## APÊNDICE A

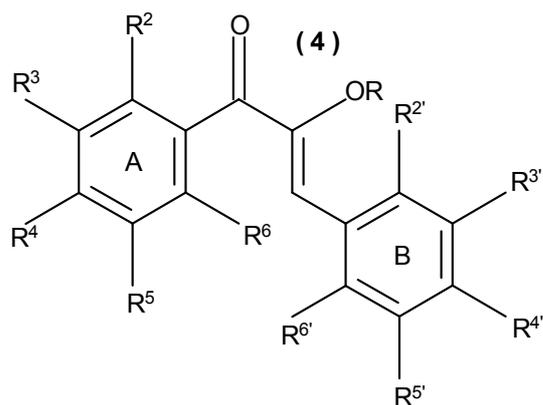
Tabela 20 – Compostos e respectivos valores de atividade biológica (pIC<sub>50</sub>) das 87 chalconas analisadas neste trabalho.

	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	R <sup>6</sup>	R <sup>2'</sup>	R <sup>3'</sup>	R <sup>4'</sup>	R <sup>5'</sup>	R <sup>6'</sup>	pIC <sub>50</sub> (μM)
<b>2a</b>	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	8,367
<b>2b</b>	H	OMe	OMe	OMe	H	H	NH <sub>2</sub>	OMe	H	H	5,569
<b>2c*</b>	OMe	OMe	OMe	H	H	H	OH	OMe	H	H	4,699
<b>2d</b>	OMe	H	H	OMe	H	H	H	NMe <sub>2</sub>	H	H	8,000
<b>2e</b>	H	OMe	OMe	OMe	H	H	O-(CH <sub>2</sub> ) <sub>2</sub> -O		H	H	5,538
<b>2f</b>	H	OMe	OMe	OMe	H	OMe	H	OMe	H	OMe	4,620
<b>2g*</b>	H	OMe	OMe	OMe	H	H	OMe	OMe	H	H	8,000
<b>2h</b>	H	OMe	OMe	OMe	H	H	O-CH <sub>2</sub> -O		H	H	5,456
<b>2i</b>	H	OMe	OMe	OMe	H	H	F	OMe	H	H	7,398
<b>2j</b>	H	OMe	OMe	OMe	H	H	H	OMe	H	H	5,420
<b>2k</b>	H	OMe	OMe	H	H	OMe	H	OMe	H	OMe	7,097
<b>2l</b>	OMe	H	OMe	H	OMe	H	OMe	OMe	OMe	H	5,420
<b>2m</b>	OMe	OMe	OMe	H	H	H	H	OMe	H	H	6,699
<b>2n</b>	OMe	H	H	OMe	H	H	O-(CH <sub>2</sub> ) <sub>2</sub> -O		H	H	5,398
<b>2o</b>	OMe	OMe	OMe	H	H	H	F	OMe	H	H	6,523
<b>2p</b>	H	O-CH <sub>2</sub> -O		H	H	OMe	H	OMe	OMe	OMe	5,398
<b>2q</b>	H	OMe	OMe	H	H	H	OMe	OMe	OMe	H	6,523
<b>2r</b>	H	OMe	OMe	OMe	H	H	OH	OH	H	H	5,357
<b>2s</b>	H	H	Me	H	H	H	OH	OMe	H	H	6,523
<b>2t</b>	H	OMe	OMe	OMe	H	H	OMe	OMe	OMe	H	5,347
<b>2u</b>	H	OMe	OMe	OMe	H	H	H	Cl	H	H	6,523
<b>2v</b>	H	OMe	OMe	OMe	H	H	H	F	H	H	5,319
<b>2w*</b>	H	H	H	H	H	H	OMe	OMe	OMe	H	6,523
<b>2x</b>	H	H	F	H	H	H	OMe	OMe	OMe	H	5,310
<b>2y</b>	OMe	OMe	OMe	H	H	H	OH	OH	H	H	6,523
<b>2z*</b>	H	OMe	OMe	H	H	H	OMe	OMe	H	H	5,292
<b>2aa</b>	OMe	H	OMe	H	H	H	H	OMe	H	H	6,523
<b>2bb*</b>	H	O-CH <sub>2</sub> -O		H	H	H	OMe	OMe	OMe	H	5,252
<b>2cc</b>	H	OMe	OH	H	H	H	OMe	OMe	OMe	H	6,523
<b>2dd</b>	H	OMe	OMe	OMe	H	Cl	H	Cl	H	H	5,208

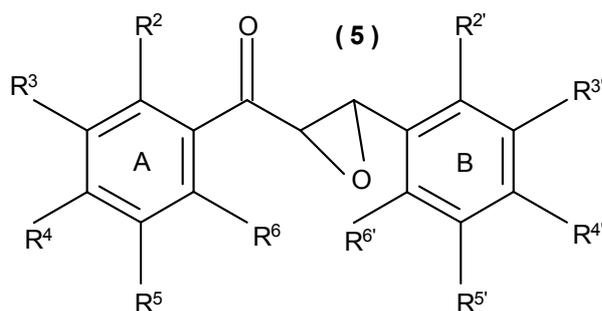
<b>2ee</b>	H	H	H	H	H	H	H	H	H	H	6,456
<b>2ff</b>	F	H	OMe	H	H	OMe	H	H	OMe	H	5,194
<b>2gg</b>	H	OMe	OMe	OMe	H	H	NO <sub>2</sub>	OMe	H	H	5,959
<b>2hh*</b>	OH	H	OMe	H	H	H	OMe	OMe	OMe	H	5,149
<b>2ii</b>	H	H	OMe	H	H	H	OMe	OMe	OMe	H	5,959
<b>2jj</b>	OMe	OMe	OMe	H	H	H	F	OMe	F	H	5,143
<b>2kk</b>	H	H	F	H	H	H	H	F	H	H	5,921
<b>2ll</b>	H	H	F	H	H	H	O-CH <sub>2</sub> -O		H	H	5,097
<b>2mm</b>	H	O-(CH <sub>2</sub> ) <sub>2</sub> -O		H	H	H	OMe	OMe	OMe	H	5,824
<b>2nn</b>	H	OMe	OMe	OMe	H	H	F	OMe	F	H	5,051
<b>2oo</b>	H	O-CH <sub>2</sub> -O		H	H	OMe	H	H	OMe	H	5,824
<b>2pp</b>	H	O-CH <sub>2</sub> -O		H	H	H	OMe	OMe	H	H	5,013
<b>2qq</b>	OMe	H	OMe	H	OMe	OMe	H	H	OMe	H	5,824
<b>2rr</b>	H	O-(CH <sub>2</sub> ) <sub>2</sub> -O		H	H	H	O-CH <sub>2</sub> -O		H	H	5,004
<b>2ss</b>	H	H	OH	H	H	H	OMe	OMe	OMe	H	5,824
<b>2tt</b>	OH	H	OMe	H	H	H	H	Cl	H	H	5,004
<b>2uu</b>	H	H	OMe	H	H	H	H	OMe	H	H	5,796
<b>2vv</b>	H	OMe	OMe	OMe	H	H	H	OBn	H	H	5,000
<b>2ww</b>	H	OMe	OMe	H	H	H	H	OMe	H	H	5,796
<b>2xx</b>	H	H	Me	H	H	H	H	Me	H	H	4,959
<b>2yy</b>	H	H	OMe	H	H	H	H	OMe	H	H	5,770
<b>2zz*</b>	OH	H	OMe	H	H	H	OMe	OMe	H	H	4,824
<b>2α</b>	H	H	Me	H	H	H	H	OMe	H	H	5,658
<b>2β</b>	H	H	Me	H	H	H	OMe	OMe	OMe	H	4,796



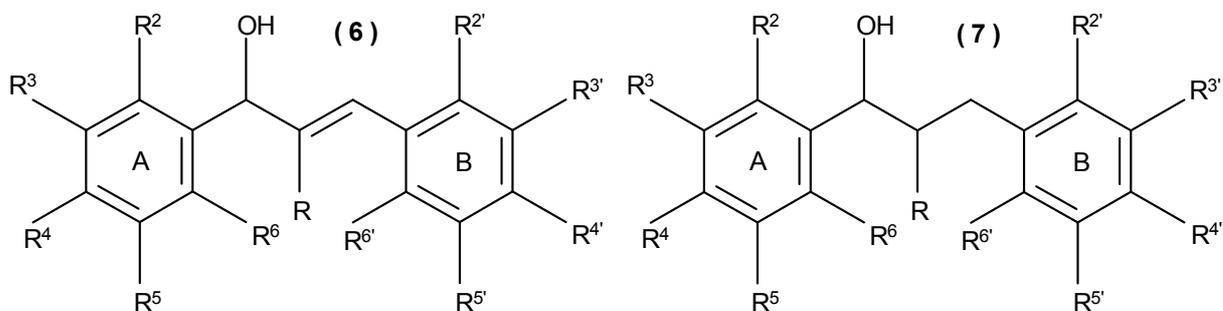
	<b>R</b>	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>	<b>pIC<sub>50</sub>(μM)</b>
<b>3a*</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	9,678
<b>3b</b>	Pr	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	8,658
<b>3c</b>	Me	H	OMe	OMe	OMe	H	H	F	OMe	H	H	8,699
<b>3e</b>	Me	OMe	H	H	OMe	H	H	H	NMe <sub>2</sub>	H	H	7,921
<b>3g</b>	Me	H	OMe	OMe	OMe	H	H	F	OMe	F	H	7,301
<b>3h</b>	Me	H	OMe	OMe	OMe	H	H	NO <sub>2</sub>	OMe	H	H	7,222
<b>3i</b>	Me	H	H	OMe	H	H	H	H	OMe	H	H	5,721
<b>3j</b>	Me	H	H	OMe	H	H	OMe	H	H	OMe	H	5,638
<b>3k</b>	Me	H	OMe	OMe	OMe	H	H	OMe	OMe	OMe	H	5,602



	<b>R</b>	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>	<b>pIC<sub>50</sub>(<math>\mu</math>M)</b>
<b>4a</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	8,824
<b>4b</b>	Et	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	8,585
<b>4c</b>	Me	H	OMe	OMe	OMe	H	H	F	OMe	H	H	8,432
<b>4d</b>	Et	H	OMe	OMe	OMe	H	H	F	OMe	H	H	7,959
<b>4e</b>	Pr	H	OMe	OMe	OMe	H	H	F	OMe	H	H	7,699
<b>4f</b>	Et	H	OMe	OMe	OMe	H	H	F	OMe	F	H	6,658
<b>4g</b>	Et	H	OMe	OMe	OMe	H	H	F	OMe	F	H	6,638
<b>4h</b>	Me	H	OMe	OMe	OMe	H	H	F	OMe	F	H	6,444

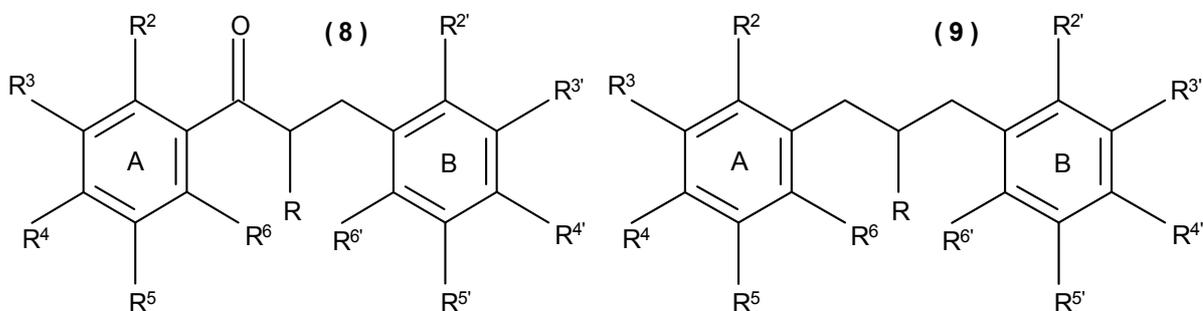


	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>	<b>pIC<sub>50</sub>(<math>\mu</math>M)</b>
<b>5a</b>	H	OMe	OMe	OMe	H	H	H	Me	H	H	5,469
<b>5b</b>	H	OMe	OMe	OMe	H	H	OMe	OMe	OMe	H	5,432
<b>5c</b>	H	OMe	OMe	OMe	H	H	H	OMe	H	H	5,310
<b>5d</b>	H	H	Me	H	H	H	H	Me	H	H	4,921
<b>5e</b>	H	H	Me	H	H	H	H	OMe	H	H	4,824



	<b>R</b>	<b>R<sup>2</sup></b>	<b>R<sup>3</sup></b>	<b>R<sup>4</sup></b>	<b>R<sup>5</sup></b>	<b>R<sup>6</sup></b>	<b>R<sup>2'</sup></b>	<b>R<sup>3'</sup></b>	<b>R<sup>4'</sup></b>	<b>R<sup>5'</sup></b>	<b>R<sup>6'</sup></b>	<b>pIC<sub>50</sub>(<math>\mu</math>M)</b>
<b>6a</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	7,046

<b>6b</b>	H	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	6,301
<b>6c</b>	H	H	OMe	OMe	OMe	H	H	H	OMe	H	H	5,398
<b>7a</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	6,222
<b>7b</b>	H	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	6,000
<b>7c</b>	H	H	OMe	OMe	OMe	H	H	H	OMe	H	H	5,398



	R	R <sup>2</sup>	R <sup>3</sup>	R <sup>4</sup>	R <sup>5</sup>	R <sup>6</sup>	R <sup>2'</sup>	R <sup>3'</sup>	R <sup>4'</sup>	R <sup>5'</sup>	R <sup>6'</sup>	pIC <sub>50</sub> ( $\mu$ M)
<b>8a</b>	Me	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	7,523
<b>8b</b>	H	H	OMe	OMe	OMe	H	H	OH	OMe	H	H	6,000
<b>8c</b>	H	H	OMe	OMe	OMe	H	H	H	OMe	H	H	5,000
<b>8d*</b>	H	H	OMe	OMe	OMe	H	H	H	OMe	H	H	4,921
<b>9a</b>	H	H	OMe	OMe	OMe	H	H	H	OMe	H	H	6,301

## APÊNDICE B

Tabela 21 – Valores dos descritores para cada um dos compostos.

Composto	E <sub>HOMO-3</sub>	UNIP	PI	SP02	RDF035m	RDF155u	RTv	RDF045e
2a	-6,207	96	263,480	6	4	6	6	18
2aa	-6,582	79	240,490	5	2	0	6	12
2b	-6,445	96	269,920	6	4	6	6	18
2bb	-6,561	97	260,010	6	4	0	6	19
2c	-6,332	94	261,430	6	4	3	6	22
2cc	-6,499	97	261,720	6	3	0	5	22
2d	-6,013	85	266,390	5	2	0	6	16
2dd	-7,018	87	294,650	5	4	0	7	15
2e	-6,494	103	271,170	6	4	1	7	17
2ee	-6,889	46	174,410	5	0	0	6	4
2f	-6,495	112	293,870	6	5	6	6	20
2ff	-6,646	83	235,530	5	2	0	6	14
2g	-6,477	103	276,470	6	4	6	6	19
2gg	-6,518	108	267,910	6	6	2	6	22
2h	-6,504	95	258,220	6	6	0	6	15
2hh	-6,389	97	258,240	6	3	0	6	20
2i	-6,565	96	259,940	6	4	0	6	17
2ii	-6,609	92	259,000	6	2	1	5	20
2j	-6,402	90	262,040	6	4	1	6	19
2jj	-6,518	100	259,100	6	5	2	6	18
2k	-6,169	101	278,830	6	2	3	6	17
2kk	-6,862	58	177,860	5	2	0	6	9
2l	-6,316	112	282,270	6	5	0	7	21
2ll	-6,701	69	202,930	5	4	0	5	10
2m	-6,311	88	257,830	6	4	3	6	20
2mm	-6,513	105	272,380	6	3	6	6	19
2n	-6,345	92	251,760	6	2	0	6	14
2nn	-6,714	102	261,090	6	5	3	6	14
2o	-6,392	94	257,840	6	4	1	6	20
2oo	-6,407	83	238,730	5	3	1	5	12
2p	-6,266	101	276,460	6	4	0	6	20
2pp	-6,570	87	245,270	5	4	2	6	14

<b>2q</b>	-6,480	105	276,340	6	3	0	6	23
<b>2qq</b>	-6,232	95	264,280	5	3	0	6	17
<b>2r</b>	-6,472	88	247,030	5	4	0	6	17
<b>2rr</b>	-6,557	86	240,610	5	3	0	6	12
<b>2s</b>	-6,569	70	222,390	5	1	0	5	12
<b>2ss</b>	-6,605	83	242,020	5	2	0	6	18
<b>2t</b>	-6,298	116	291,110	6	5	1	6	24
<b>2tt</b>	-6,673	67	217,660	5	1	0	6	9
<b>2u</b>	-6,620	81	249,950	5	3	0	6	13
<b>2uu</b>	-6,722	72	226,280	5	1	1	5	11
<b>2v</b>	-6,605	66	194,990	5	2	0	6	10
<b>2vv</b>	-6,314	149	325,700	6	5	9	7	21
<b>2w</b>	-6,994	75	264,890	5	2	0	6	17
<b>2ww</b>	-6,530	81	244,920	5	1	4	5	15
<b>2x</b>	-6,682	83	235,460	5	3	0	6	19
<b>2xx</b>	-6,616	58	208,440	5	0	0	5	7
<b>2y</b>	-6,309	86	245,070	5	3	0	6	20
<b>2yy</b>	-6,722	72	226,280	5	1	1	5	11
<b>2z</b>	-6,521	94	261,560	6	2	4	6	18
<b>2zz</b>	-6,381	87	243,540	5	2	0	6	15
<b>2α</b>	-6,679	65	218,700	5	1	0	5	9
<b>2β</b>	-6,534	83	248,990	5	2	0	6	17
<b>3a</b>	-6,459	98	268,150	6	5	6	7	27
<b>3b</b>	-6,384	105	279,620	6	7	4	8	37
<b>3c</b>	-6,539	98	264,980	6	5	2	7	27
<b>3e</b>	-5,838	87	273,740	5	2	3	7	23
<b>3g</b>	-6,655	104	263,790	6	6	1	7	22
<b>3h</b>	-6,711	112	274,190	6	7	3	7	27
<b>3i</b>	-6,651	73	230,590	5	2	1	6	17
<b>3j</b>	-6,449	80	234,200	5	2	0	7	21
<b>3k</b>	-6,476	117	292,920	6	6	1	7	29
<b>4a</b>	-6,473	101	276,840	6	7	7	7	25
<b>4b</b>	-6,489	105	287,230	6	7	7	7	32
<b>4c</b>	-6,567	101	272,960	6	7	1	7	22
<b>4d</b>	-6,584	105	283,010	6	7	2	7	28
<b>4e</b>	-6,587	109	294,100	6	7	1	7	27
<b>4f</b>	-6,700	114	294,860	6	8	2	7	25

<b>4g</b>	-6,691	110	282,030	6	9	2	7	26
<b>4h</b>	-6,669	107	274,200	6	7	0	6	22
<b>5a</b>	-6,835	84	217,420	5	5	0	8	21
<b>5b</b>	-6,691	117	266,290	6	6	0	7	22
<b>5c</b>	-6,666	92	231,180	6	4	1	7	19
<b>5d</b>	-6,873	59	188,060	5	0	0	7	9
<b>5e</b>	-6,644	75	197,860	5	4	0	7	10
<b>6a</b>	-6,514	98	250,930	5	7	0	7	31
<b>6b</b>	-6,660	96	236,080	5	7	0	7	24
<b>6c</b>	-6,642	90	238,000	5	6	0	7	21
<b>7a</b>	-6,457	98	237,560	5	7	0	7	36
<b>7b</b>	-6,438	96	235,460	5	5	0	7	27
<b>7c</b>	-6,507	90	222,390	5	4	0	7	24
<b>8a</b>	-6,585	98	242,700	5	8	0	7	30
<b>8b</b>	-6,563	96	230,180	6	5	0	7	22
<b>8c</b>	-6,621	90	225,950	5	4	0	7	19
<b>8d</b>	-6,749	92	225,700	6	3	0	7	19
<b>9a</b>	-6,404	88	221,050	5	5	0	7	22

## APÊNDICE C

Figura 56 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 2a e 2d, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

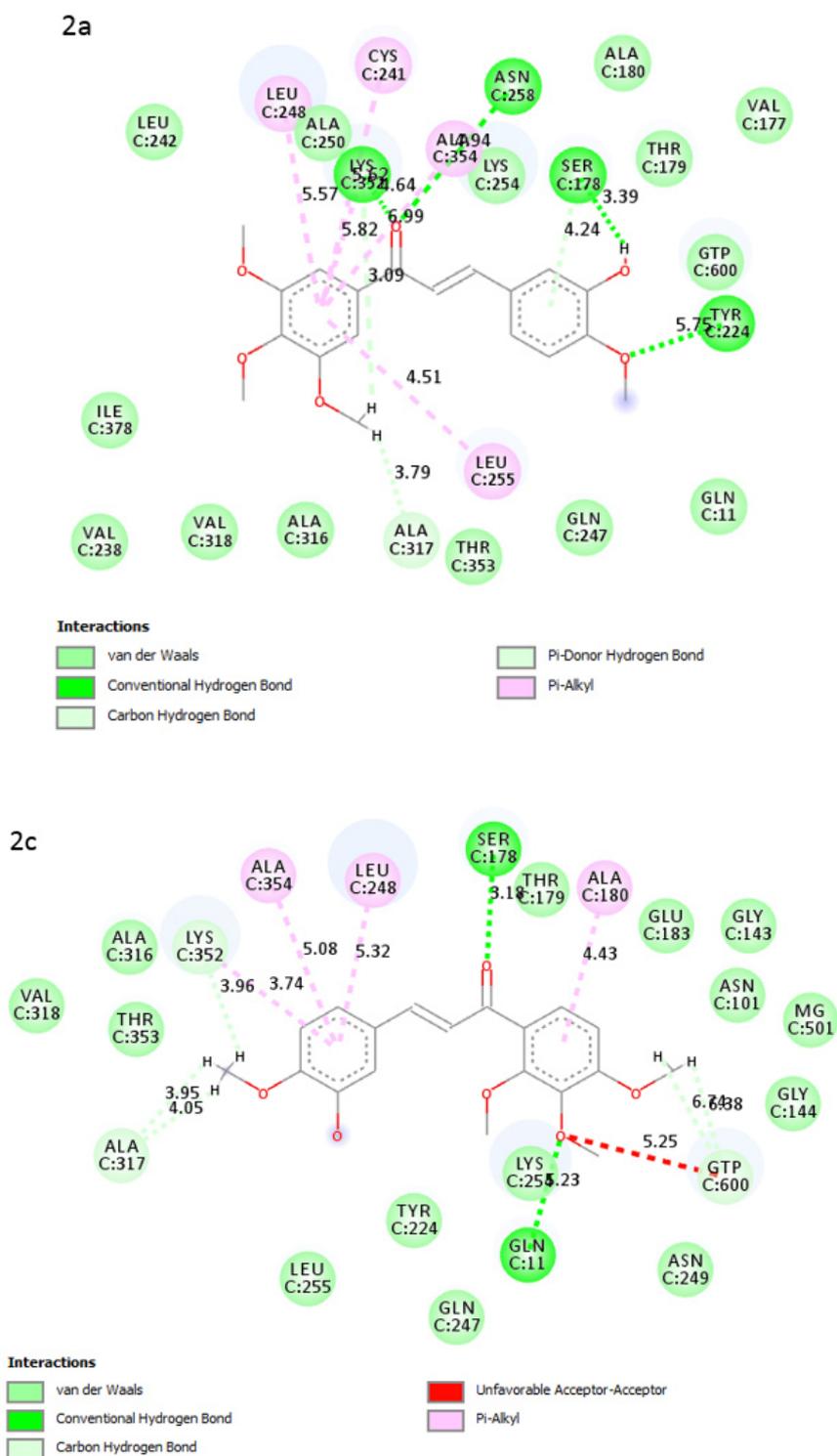


Figura 57 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 2g e 2nn, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

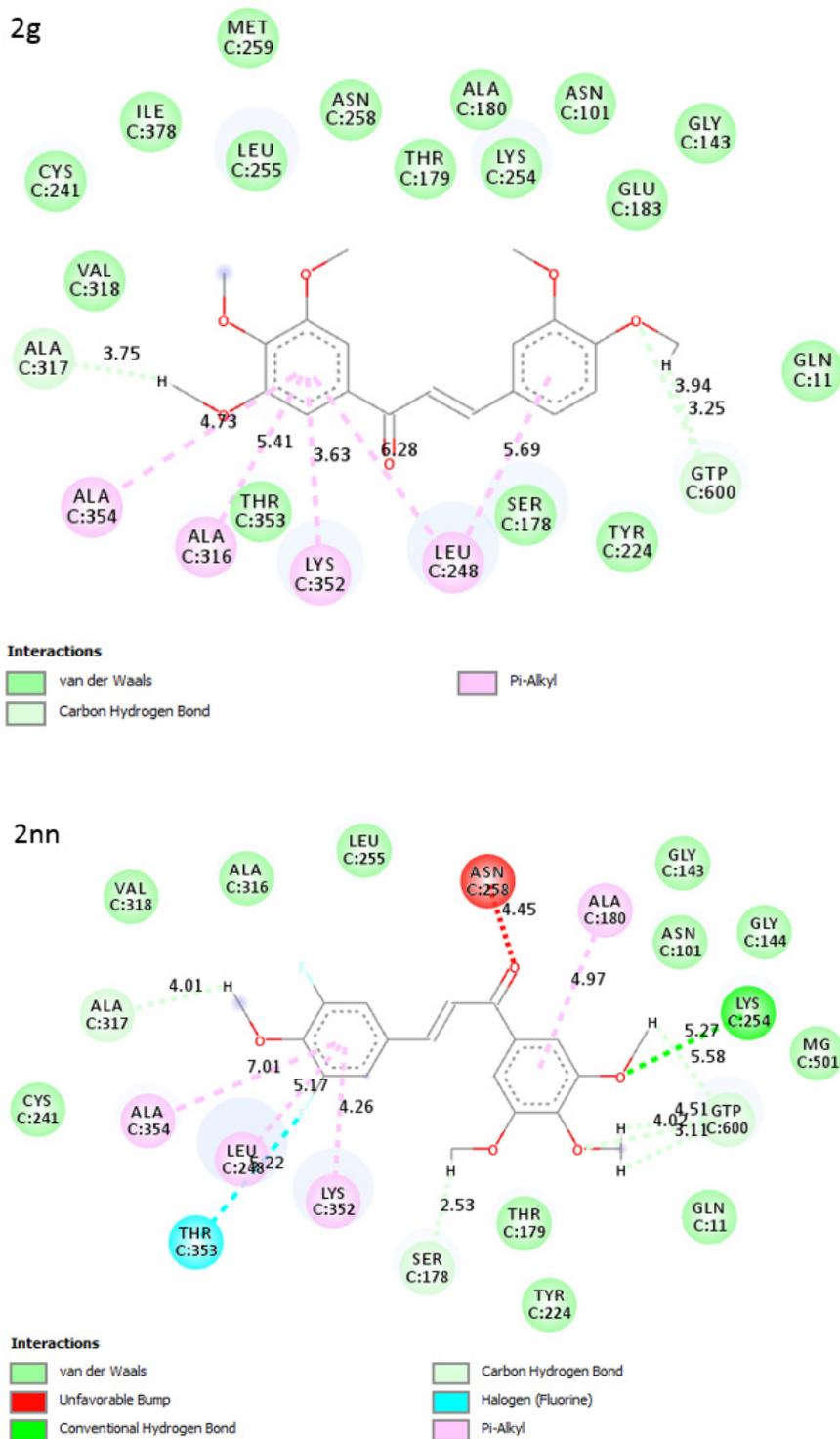




Figura 59 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 2xx e 2zz, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

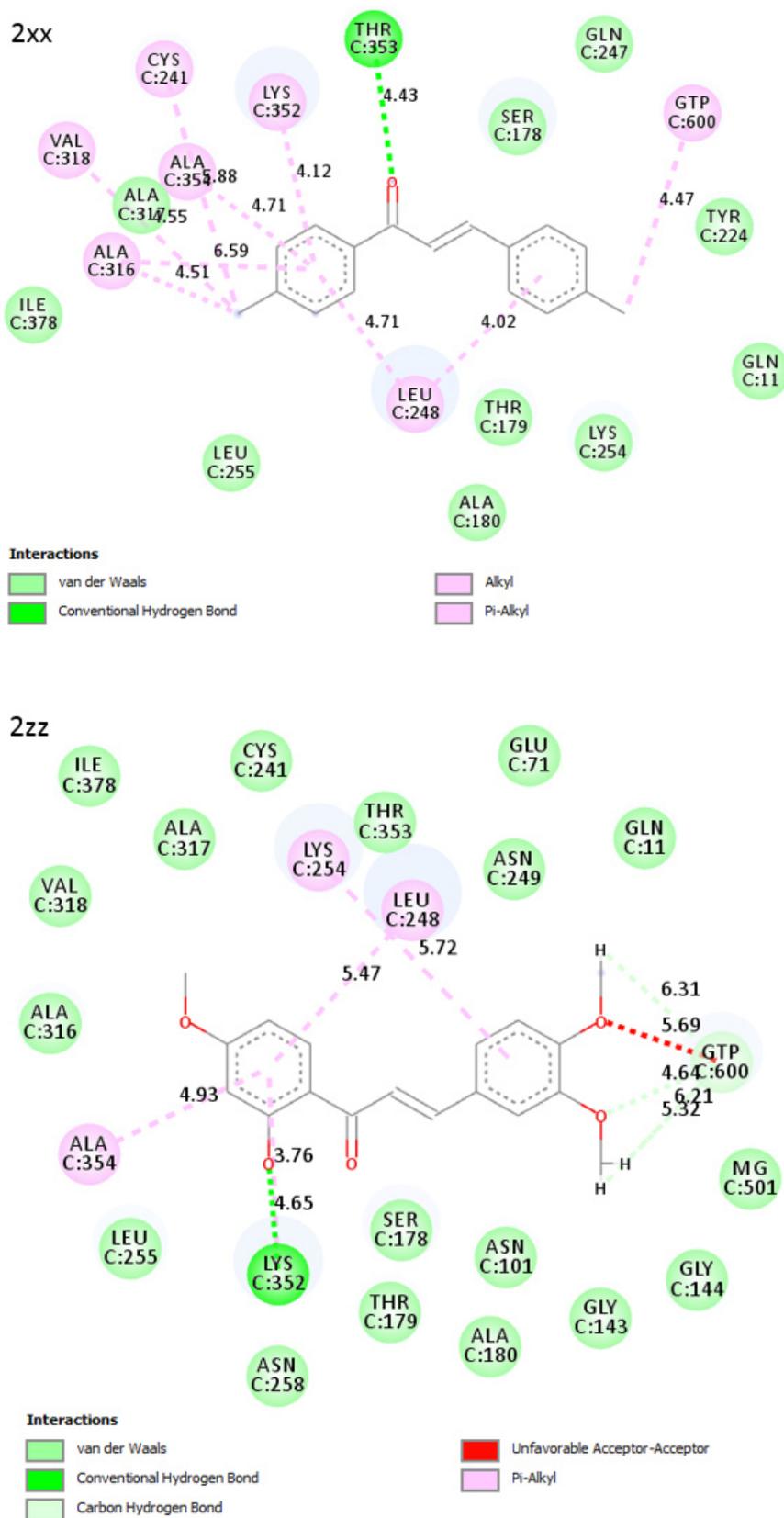


Figura 60 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 3b e 3c, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

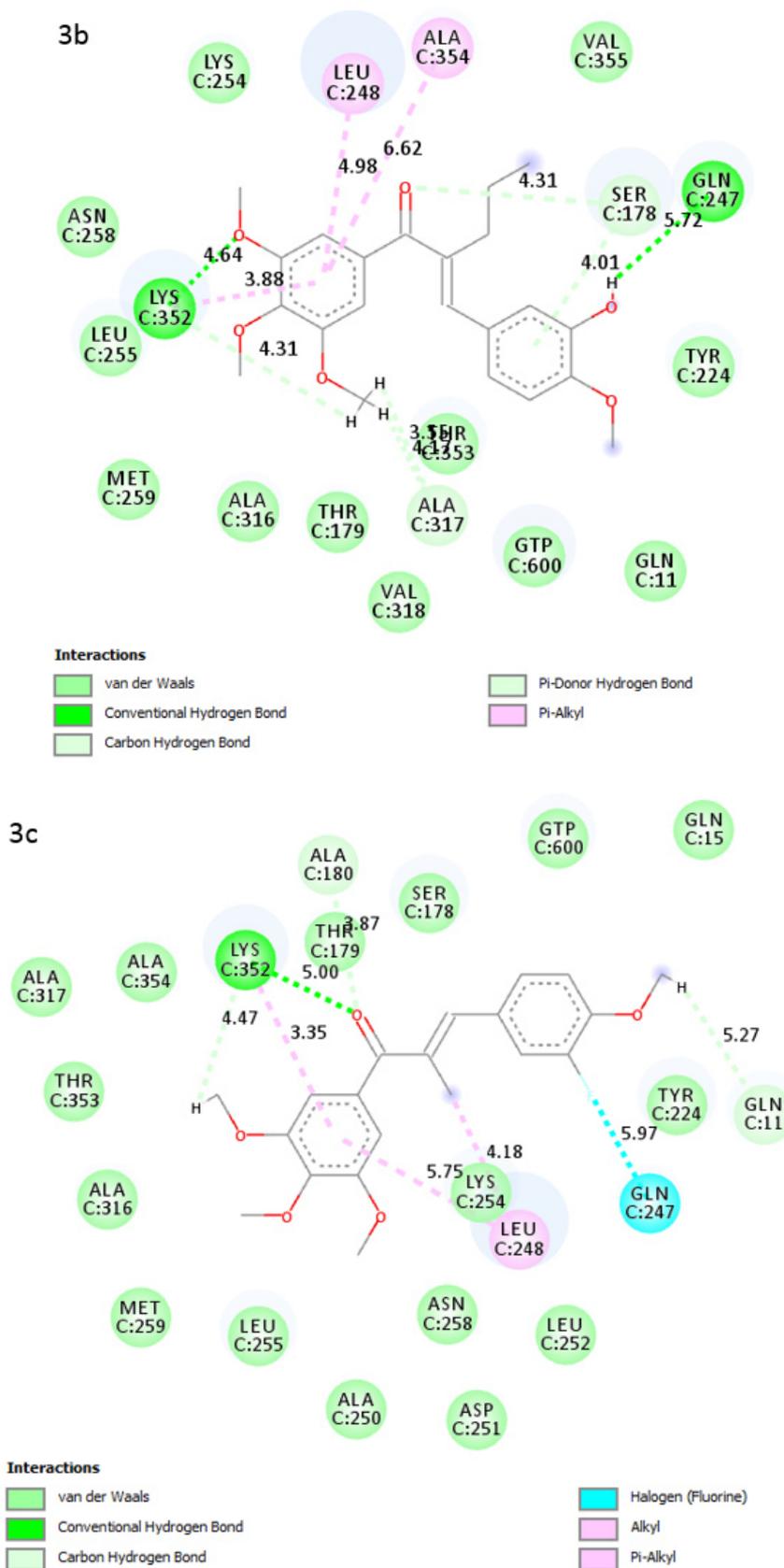


Figura 61 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 4b e 4c, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

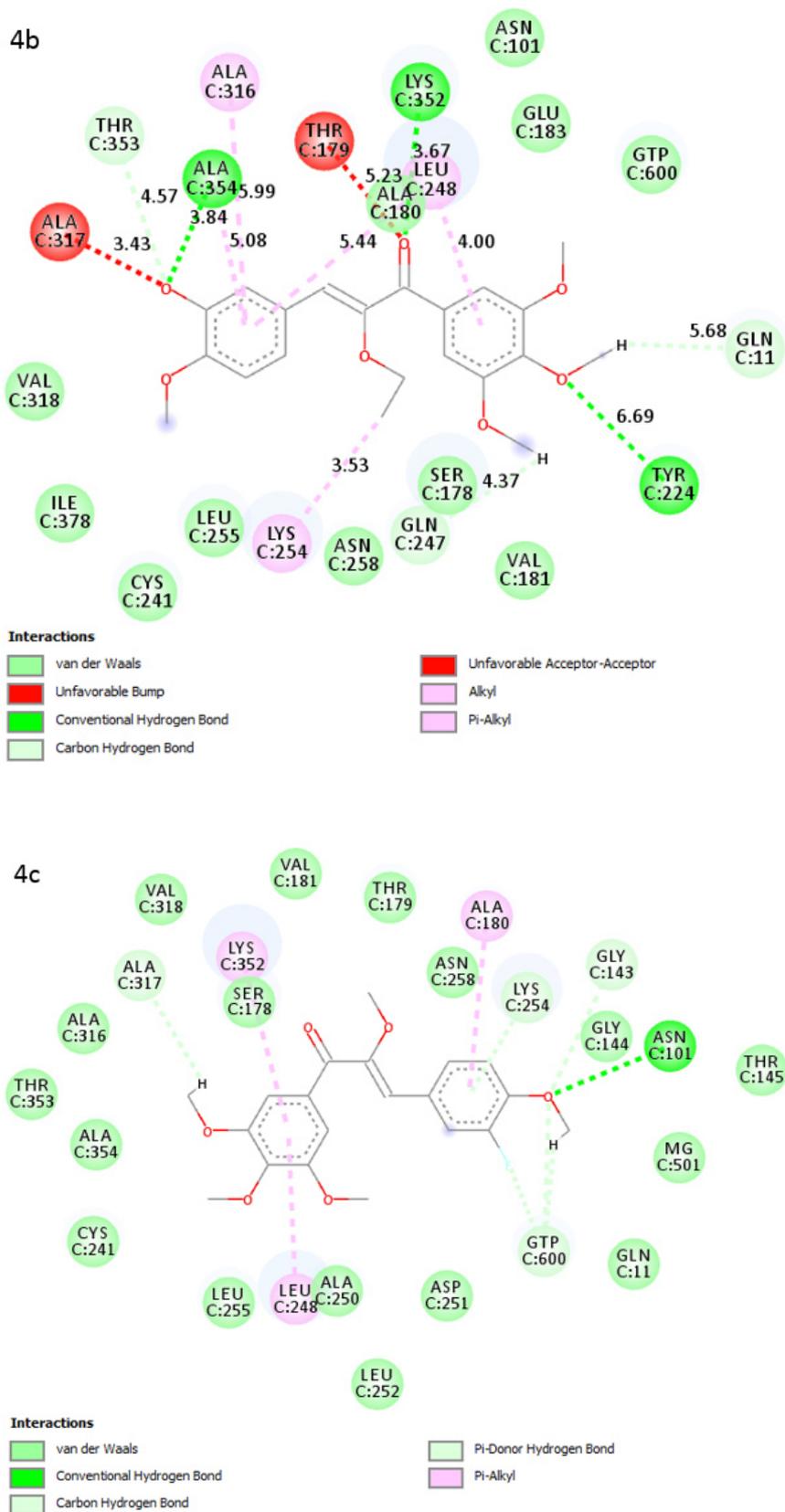


Figura 62 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 4d e 5d, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

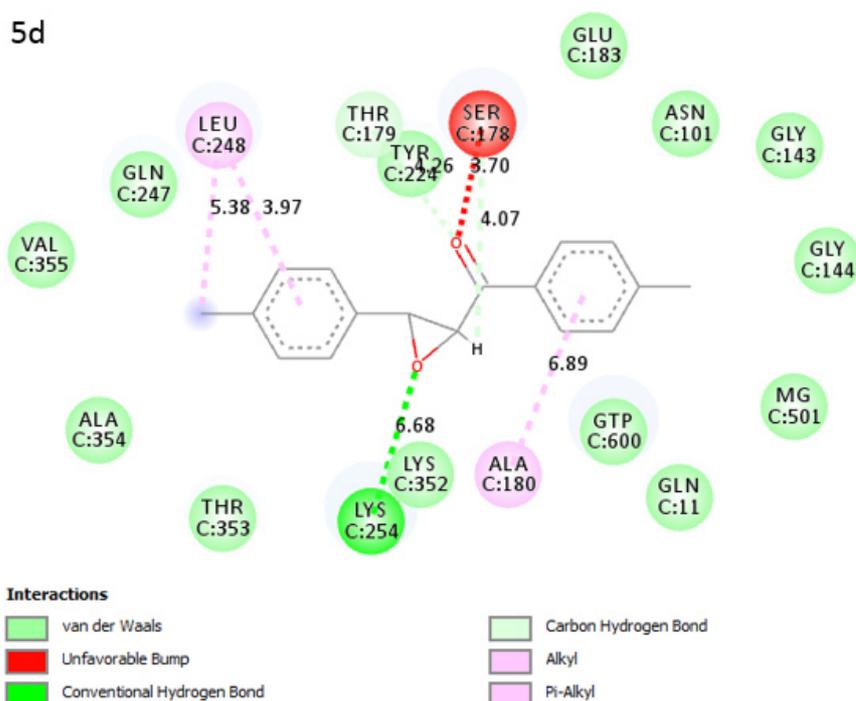
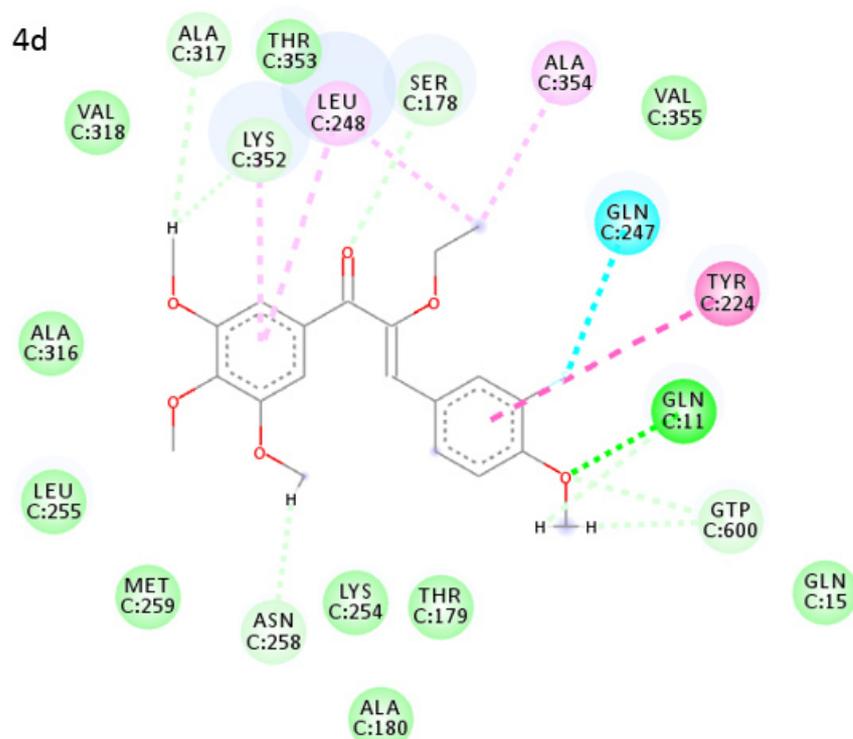


Figura 63 – Diagramas 2D dos complexos formados entre a  $\alpha\beta$ -tubulina e os ligantes 6a e 8a, selecionados para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

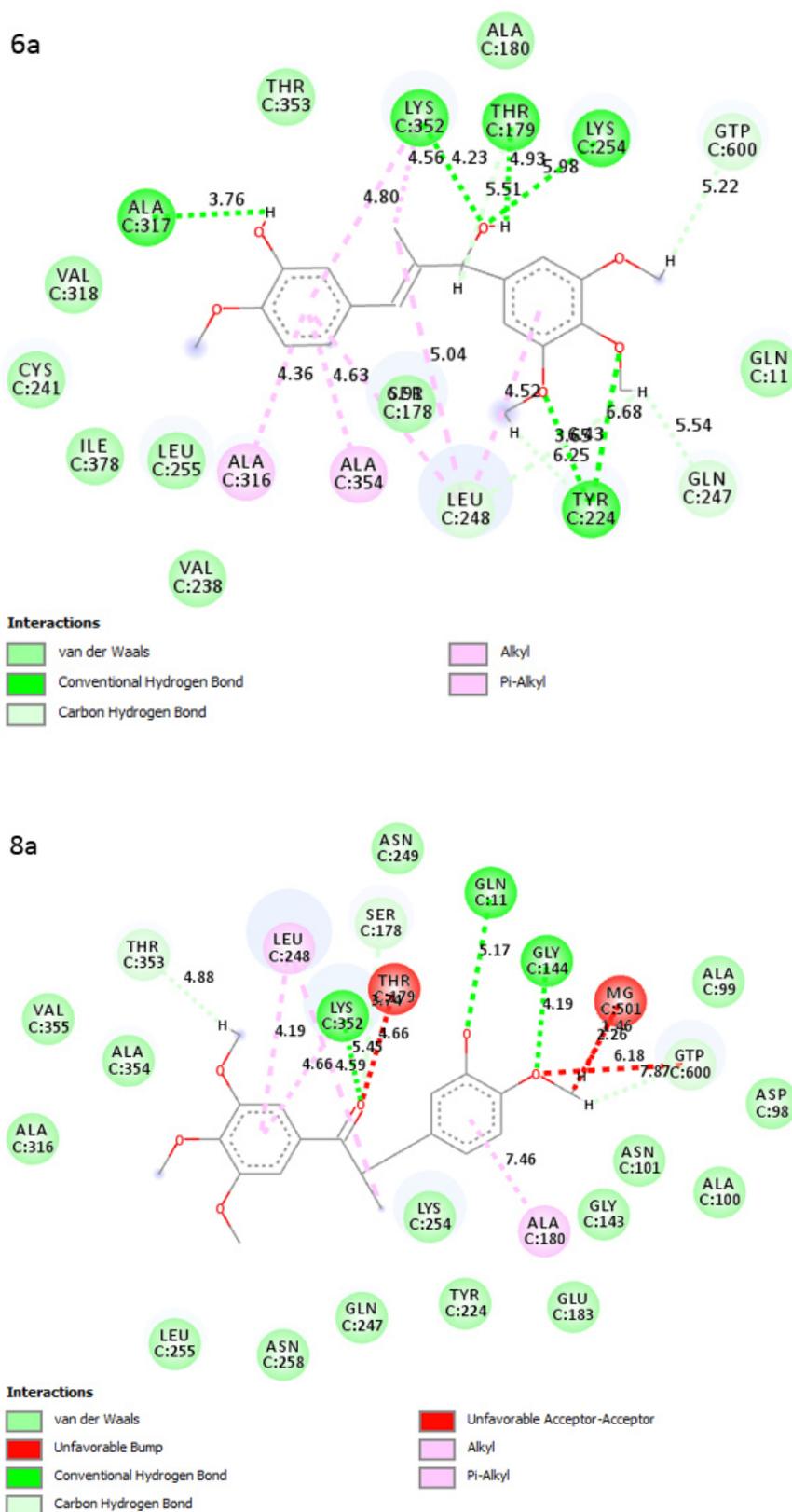


Figura 64 – Diagramas 2D do complexo formado entre a  $\alpha\beta$ -tubulina e o ligante 8d, selecionado para as análises de docagem molecular, após ser fornecida a **flexibilidade 2** a alguns dos resíduos.

