



TAMIRIS MARIA DE ASSIS

**MODELOS DE QSAR- 4D E ANCORAMENTO MOLECULAR
APLICADOS AO ESTUDO DE INIBIDORES DA PROTEÍNA
TGF- β 1 PARA O TRATAMENTO DO CÂNCER**

**LAVRAS - MG
2018**

TAMIRIS MARIA DE ASSIS

**MODELOS DE QSAR- 4D E ANCORAMENTO MOLECULAR APLICADOS AO
ESTUDO DE INIBIDORES DA PROTEÍNA TGF- β 1 PARA O TRATAMENTO DO
CÂNCER**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração em Química/Bioquímica, para a obtenção do título de Doutora.

Orientadora
Dra. Elaine Fontes Ferreira da Cunha

**LAVRAS - MG
2018**

de Assis, Tamiris Maria.

Modelos de qsar-4d e ancoramento molecular aplicados ao estudos de inibidores da proteína $\text{tgf-}\beta 1$ para o tratamento do câncer / Tamiris Maria de Assis. - 2018.

99 p. : il.

Orientador(a): Elaine Fontes Ferreira da Cunha.

Tese (doutorado) - Universidade Federal de Lavras, 2018.
Bibliografia.

1. Câncer. 2. Proteína $\text{TGF-}\beta 1$. 3. QSAR-4D. I. da Cunha, Elaine Fontes Ferreira. . II. Título.

TAMIRIS MARIA DE ASSIS

**MODELOS DE QSAR- 4D E ANCORAMENTO MOLECULAR APLICADOS AO
ESTUDO DE INIBIDORES DA PROTEÍNA TGF- β 1 PARA O TRATAMENTO DO
CÂNCER**

**4D-QSAR MODELS AND MOLECULAR DOCKING APPLIED TO THE STUDY OF
TGF- β 1 PROTEIN INHIBITORS THOUGHT CANCER TREATMENT**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Agroquímica, área de concentração em Química/Bioquímica, para a obtenção do título de Doutora.

APROVADA em 23 de fevereiro de 2018

Dra. Ana Paula Guimarães	UFV
Dra. Daiana Teixeira Mancini	UFLA
Dr. Teodorico de Castro Ramalho	UFLA
Dr. Alison Geraldo Pacheco	UFLA

Orientadora
Dra. Elaine Fontes Ferreira da Cunha

**LAVRAS - MG
2018**

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer a Deus, por tudo o que eu já vivi. Pelos momentos bons, ruins, felizes e, principalmente, por ter me dado força para superar os difíceis, pois foram eles que me fizeram ser quem eu sou.

À minha família que me ensinou a cultivar o amor, o carinho, a compreensão, o respeito e a responsabilidade. Valores esses que me proporcionaram muitas conquistas ao longo da minha trajetória, principalmente a minha mãe e meu pai, Vanderlete e Antonio, pelo amor, confiança e dedicação.

Aos meus irmãos, Tatiane e Weverton, pelo carinho, compreensão e amizade.

Ao meu esposo André, pela compreensão, amizade, confiança e paciência durante todos esses anos.

Ao meu filho João Pedro, que me mostrou o verdadeiro sentido da palavra amor.

Aos amigos que conquistei durante esta jornada, e a todos aqueles que estiveram juntos de mim nos momentos bons e ruins, obrigada pela amizade.

À minha orientadora, Prof^a Elaine, pela paciência, confiança, amizade, incentivo e apoio, principalmente pelo carinho.

Aos amigos e companheiros do Grupo de Química Computacional, a "era do gelo" pela grande parceria e amizade durante esses anos de convivência.

Aos membros da banca, Ana Paula, Daiana, Teodorico e Alison, pela disposição e pelas contribuições que complementarão o trabalho.

À Capes, Fapemig, UFLA e DQI, pelo suporte e oportunidade de realizar a pesquisa.

RESUMO

O câncer é uma doença que afeta milhões de pessoas em todo o mundo, e é caracterizado pelo crescimento anormal de células, que se manifestam de forma agressiva e incontrolável. Existem vários tipos de tratamento para o câncer, no entanto eles são mais eficazes nas fases iniciais, desse modo é essencial a descoberta de novos tratamentos para o câncer em estágios avançados. Nesta perspectiva, os membros da subfamília TGF- β são considerados alvos moleculares interessantes, uma vez que desempenham funções importantes no crescimento e desenvolvimento celular. Sendo assim, o objetivo deste trabalho foi determinar quantitativamente a influência de descritores estruturais na atividade de derivados aril pirimidina como inibidores da proteína TGF- β 1, a partir da aplicação de estudos de QSAR-4D e analisar o modo de interação entre esses inibidores e a proteína TGF- β 1 a partir do estudo de ancoramento molecular. Este estudo foi realizado a fim de entender a afinidade dos inibidores com a proteína e identificar quais fatores são relevantes para a inibição da via de sinalização. Os modelos de QSAR-4D apresentaram bons parâmetros estatísticos ($R^2= 0,792$, $Q^2= 0,584$, $R^2_{pred}= 0,648$, $r^2_m=0,547$, $R^2_p= 0,681$ e $R^2_{rand}= 0,207$) além de grupos farmacofóricos importantes para a atividade desses compostos. A partir do ancoramento foram analisadas as interações existentes entre proteína e ligantes, no qual foi averiguado que, resíduos de aminoácidos como His283, Asp351 e Lys232 fazem ligação de hidrogênio com a maioria dos compostos. Em contrapartida, os resíduos Leu340, Ile211 e Val219 realizam interações estéricas importantes com grupos que foram considerados favoráveis no QSAR. A partir desses resultados foram propostas dez novas estruturas, que tiveram sua atividade predita a partir do modelo obtido pelo QSAR-4D. Todas apresentaram boas energias de interação calculadas pelo ancoramento molecular e quatro apresentaram atividades preditas superiores àquelas dos compostos estudados. Portanto, os resultados obtidos sugerem que esses quatro compostos podem ser considerados potentes protótipos de inibidores da proteína TGF- β 1 para o tratamento do câncer.

Palavras-chave: Câncer. Proteína TGF- β 1. QSAR-4D. Ancoramento Molecular.

ABSTRACT

Cancer is a disease that affects millions of people around the world, it is characterized by abnormal cell growth, which manifests itself aggressively and uncontrollably. There are several types of treatment for cancer, however they are most effective in the early stages, so it is essential to discover new treatments for cancer in advanced stages. In this perspective, the members of the TGF- β subfamily are considered interesting molecular targets, since they play important function in growth and development cell. Therefore, the objective of this work was to quantitatively determine the influence of structural descriptors on the activity of aryl pyrimidine derivatives as inhibitors of the TGF- β 1 protein, from the application of QSAR-4D studies and to analyze the mode of interaction between these inhibitors and TGF- β 1 protein through the molecular docking. This study was carried out in order to understand the affinity of the inhibitors with the protein and to identify which factors are relevant for the inhibition of the signaling pathway. The 4D-QSAR models presented good statistical parameters ($R^2 = 0.792$, $Q^2 = 0.584$, $R^2_{\text{pred}} = 0.648$, $r^2_{\text{m}} = 0.547$, $R^2_{\text{p}} = 0.681$ and $R^2_{\text{rand}} = 0.207$) besides pharmacophores groups important for the activity of these compounds. From the docking, the interactions between protein and ligand were analyzed, in which it was verified that amino acid residues like His283, Asp351 and Lys232 perform hydrogen bonding with most of the compounds. In contrast, Leu340, Ile211 and Val219 residues perform important steric interactions with groups that were considered favorable in QSAR. From these results, ten new structures were proposed, which had their activity predicted from the model obtained by QSAR-4D. All presented good interaction energies calculated by molecular docking and four presented predicted activities superior to those of the compounds studied. Therefore, the results suggest that these four compounds can be considered as potent prototypes of inhibitors of the TGF- β 1 protein for the treatment of cancer.

Keywords: Cancer. TGF- β 1 protein. 4D-QSAR. Molecular Docking.

LISTA DE FIGURAS

Figura 1 – Distribuição dos casos de câncer por região.....	16
Figura 2 – Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para 2016 por sexo.....	17
Figura 3 – Estrutura TGF β 1.....	20
Figura 4 – Via de sinalização TGF β / Smad.....	22
Figura 5 – Estrutura química do Galunisertib e Vactosertib.....	23
Figura 6 – Estrutura de um núcleo pirimidínico.....	24
Figura 7 – Gráfico de distribuição dos valores de inibição de fosforilação dos 63 compostos estudados.....	54
Figura 8 – Composto 1, as letras a, b, c, d, e representam os átomos fixados durante a dinâmica molecular e/ou os átomos que foram utilizados nos alinhamentos.....	56
Figura 9 – Aminoácidos considerados flexíveis para o estudo de ancoramento molecular juntamente com o inibidor complexado da proteína TGF- β 1 (PDB:3FAA).....	61
Figura 10 – Modelos de QSAR-4D representativo do alinhamento 4, mostrando o composto 47 juntamente com descritores.....	68
Figura 11 – Representação da molécula 53 e do GCOD-1.....	68
Figura 12 – Representação da molécula 29 e do GCOD-2.....	69
Figura 13 – Representação da molécula 51 e do GCOD-3.....	70
Figura 14 – Representação da molécula 47 e do GCOD-4.....	70
Figura 15 – Representação da molécula 41 e dos GCOD-5 e 6.....	71
Figura 16 – Representação da molécula 15 e do GCOD-7.....	72
Figura 17 – Representação da molécula 57 e do GCOD-8.....	72
Figura 18 – Representação da molécula 46 e do GCOD-9.....	73
Figura 19 – Sobreposição entre a estrutura de raios-x cristalizada do receptor (amarelo) e a conformação obtida no estudo de docking (vermelho).....	75
Figura 20 – Ligação de hidrogênio do ligante cristalizado (2-aminoimidazol) na proteína de código PDB 3FAA. (A) interações apresentadas no PDB; (B) interações obtidas no redocking.....	76

Figura 21 – Pose do ancoramento molecular para o composto 50 e suas ligações de hidrogênio.....	82
Figura 22 – Mapa de energia estérica para a pose do composto 56	83
Figura 23 – Mapa de energia potencial eletrostático do composto 56	84
Figura 24 – Mapa de potencial eletrostáticas do composto 36	85
Figura 25 – Ilustração das moléculas propostas.....	86
Figura 26 – Composto proposto P6 dentro do sítio de ligação da proteína TGF- β 1.....	89

LISTA DE TABELAS

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (inibição da via de sinalização).....	39
Tabela 2 – Grupos teste estudados.....	55
Tabela 3 – Alinhamentos testados e os átomos selecionados.....	58
Tabela 4 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 1.....	62
Tabela 5 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 2.....	63
Tabela 6 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 3.....	63
Tabela 7 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 4.....	64
Tabela 8 – Resumo dos resultados dos modelos de QSAR-4D provenientes do alinhamento 4.....	65
Tabela 8 – Matriz de correlação cruzada dos GCODs do melhor modelo do alinhamento 4.....	66
Tabela 10 – Valores experimentais e preditos de inibição juntamente com o valor residual.....	74
Tabela 11 – Energias obtidas a partir do <i>redocking</i> (kcal.mol ⁻¹) e valor do desvio médio quadrático (RMSD).....	76
Tabela 10 – Energias obtidas a partir do ancoramento molecular para os derivados aril pirimidinas (kcal.mol ⁻¹).....	77
Tabela 11 – Valores de atividade predita e energia de docking (kcal.mol ⁻¹) para os inibidores propostos.....	87

LISTA DE ABREVIATURAS

His	Histidina
Asp	Aspartato
Leu	Leucina
Ile	Isoleucina
Lys	Lisina
Val	Valina
Ala	Alanina
Tyr	Tirosina
Phe	Fenilalanina
Ser	Serina

LISTA DE SIGLAS

OMS	Organização Mundial de Saúde
TGF- β	<i>Transforming Growth Factor β</i>
QSAR	Relação Quantitativa Estrutura Atividade
INCA	Instituto Nacional do Câncer
BMP	Bone Morphogenetic Proteins
GDF	<i>Growth and Differentiation Factors</i>
Vg 1	Vitelogenina 1
SLC	<i>Small Latente Complex</i>
SP	<i>Signal Peptide</i>
LAP	<i>Latency Associted Peptide</i>
LTBP	<i>Latency TGF-β Binding Protein</i>
LLC	<i>Large Latente Complex</i>
SMAD	<i>Mother Against Decapentaplegic Homolog</i>
IUPAC	União Internacional de Química Pura e Aplicada
RLS	Regressão Linear Simples
RLM	Regressão Linear Múltipla
RNL	Regressão não-linear
QSAR-IR	Relação Quantitativa Estrutura-Atividade Independente do Receptor
QSAR-DR	Relação Quantitativa Estrutura-Atividade Dependente do Receptor
SDM	Simulação de Dinâmica Molecular
CEP	Conformational Ensemble Profile
GCOD	Grid Cell Occupancy Descriptors
IPE	Interaction Pharmacophore Elements
PLS	Partial Least Squares
VL	Variáveis Latentes
GFA	Genetic Function Approximation
GA	Genetic Algorithm
LOF	Lack-of-fit

PDB	Protein Data Bank
MVD	Molegro Virtual Docker
RMSD	Root-Mean-Square Deviation

SUMÁRIO

1.	INTRODUÇÃO.....	14
2.	REFERENCIAL TEÓRICO.....	15
2.1	Câncer	13
2.2.	Superfamília TGF- β	18
2.2.1.	A subfamília TGF- β	19
2.3.	Sinalização de TGF- β	21
2.4.	Inibidores de TGF- β	22
2.5.	Compostos pirimidínicos	24
2.6.	A química medicinal e o planejamento de fármacos	25
2.7.	Relação quantitativa estrutura- atividade.....	26
2.7.1.	QSAR 4D.....	28
2.7.1.1	Simulação de Dinâmica Molecular	27
2.7.1.2.	Definição do tamanho das células	28
2.7.1.3.	Definição dos elementos de interação farmacofórica	29
2.7.1.4.	Definição dos tipos de ocupação	29
2.7.1.5.	Definição do alinhamento	30
2.8.	Ancoragem molecular (Docking)	30
2.9.	Métodos e índices estatísticos.....	32
2.9.1.	Regressão pelo método dos mínimos quadrados parciais (PLS).....	32
2.9.2.	Aproximação da Função Genética	32
2.9.3.	Índice de Falta de Ajuste (LOF, lack-of-fit).....	33
2.9.4.	Coeficiente de Correlação Linear Quadrático (r^2 ou R^2).....	34
2.9.5.	Validação Interna dos modelos	35
2.9.6.	Validação Externa dos modelos.....	35
2.9.7.	Validação Y-randomization (R^2 Y-rand) e R^2 P.....	35
3.	OBJETIVOS	36
4.	METODOLOGIA	37
4.1.	Dados Biológicos.....	38
4.2.	Construção e Minimização dos compostos.....	54
4.3.	Metodologia de QSAR-4D: Divisão do conjunto de treinamento e de teste	54
4.4.	Simulação de dinâmica molecular (SDM)	56
4.5.	Definição do tamanho das células	57
4.6.	Definição dos elementos de interação farmacofórica	57
4.7.	Definição dos tipos de ocupação	57
4.8.	Definição do alinhamento	57
4.9.	Construção dos modelos de QSAR-4D.....	58
4.9.1.	Pré-tratamento dos descritores (GCOD)	58
4.9.2.	Obtenção dos modelos de QSAR-4D	58
4.9.3.	Validações	59
4.9.4.	Valores residuais e Outliers	59
4.9.5.	Seleção da "Conformação Preferencial"	59
4.10.	Ancoramento Molecular.....	60
5.	RESULTADOS E DISCUSSÃO	62
5.1.	Seleção do melhor alinhamento de QSAR-4D	62
5.2.	Análise do Alinhamento 4.....	65
5.2.1.	Análise dos descritores.....	67
5.2.2.	Análise dos valores residuais e Outliers	73

5.3.	Análise do ancoramento molecular	75
5.4.	Planejamento de potenciais inibidores da enzima TGF- β 1.....	84
6.	CONSIDERAÇÕES FINAIS	90
	REFERÊNCIAS.....	91

1. INTRODUÇÃO

Câncer é o nome dado a um conjunto de doenças que são caracterizadas pelo crescimento anormal de células, que se dividem de forma incontrolável, formando tumores. E a sua nomenclatura é dada a partir dos diferentes tipos de células e tecidos existentes (INCA, 2017).

Ela é uma doença que afeta milhões de pessoas em todo o mundo e que aumenta cada vez mais devido ao crescimento e ao estilo de vida da população. Dados da organização mundial de saúde (OMS, 2017) indicam que mais de 14 milhões de pessoas desenvolvem câncer todos os anos, e que este número deve subir para 21 milhões em 2030.

Existem muitas formas de tratamentos para esta doença, no entanto, elas são mais eficazes nas fases iniciais, na qual o câncer ainda se encontra somente no tecido de origem. Por isso são necessários novos alvos para tratamento de alguns tipos de câncer em estágios avançados. Neste sentido, a superfamília de proteínas TGF- β é um alvo molecular interessante, pois desempenha papéis críticos no crescimento e desenvolvimento celular. Dentre os constituintes dessa superfamília, a subfamília TGF- β é a mais estudada e os inibidores da sua via de sinalização são os mais avançados no desenvolvimento clínico (NEUZILLET *et al.*, 2015).

O processo de descoberta de novos fármacos é complexo, longo e de alto custo. É de caráter multidisciplinar e engloba áreas como a farmacologia, bioquímica, química orgânica e computacional (GUIDO, ANDRICOPULO e OLIVA, 2010). Nesse processo, o último destaca-se pelo baixo custo e menor tempo, quando comparado aos métodos experimentais. Dentre os métodos computacionais mais utilizados no planejamento racional, encontra-se o QSAR (relação quantitativa estrutura-atividade), um método que correlaciona informações estruturais com a atividade biológica dos compostos, por meio de descritores moleculares, e o Ancoramento molecular, que simula o encaixe de um ligante em uma macromolécula (SANT, 2009).

Neste trabalho, procurou-se determinar quantitativamente a influência de descritores estruturais na atividade de derivados aril pirimidina como inibidores da proteína TGF- β 1, a partir da aplicação de estudos de QSAR-4D, a fim de identificar características importantes para a atividade biológica. Objetivou-se adicionalmente analisar o modo de ligação entre esses compostos e a proteína TGF- β 1 por ancoramento molecular e consequente proposição de novas moléculas, protótipos como candidatos à fármacos para o tratamento do câncer.

2. REFERENCIAL TEÓRICO

2.1. Câncer

Segundo o Instituto Nacional do Câncer (INCA), o câncer é o nome dado a um conjunto de mais de cem doenças que têm em comum o crescimento desordenado de células que invadem os tecidos e órgãos. Essas células exibem uma tendência agressiva e incontrolável, dividindo-se rapidamente e formando tumores (acúmulo de células cancerosas) ou neoplasias malignas. No entanto, existem tumores benignos, que são massas localizadas de células que se multiplicam vagarosamente e se assemelham ao seu tecido original, o que raramente constitui um risco de vida (INCA, 2017).

A proliferação anormal das células resulta em características invasivas, que podem chegar aos tecidos vizinhos e ganhar acesso aos vasos linfáticos e sanguíneos. Tais células podem ser transportadas para outras localidades do corpo, o que é conhecido como metástase. Além disso, essa doença concede algumas características especiais às células, como a morte celular programada (apoptose) e a produção de novos vasos sanguíneos (angiogênese) (DA SILVA *et al.*, 2013; SILVA; RAMALHO e DA CUNHA, 2014).

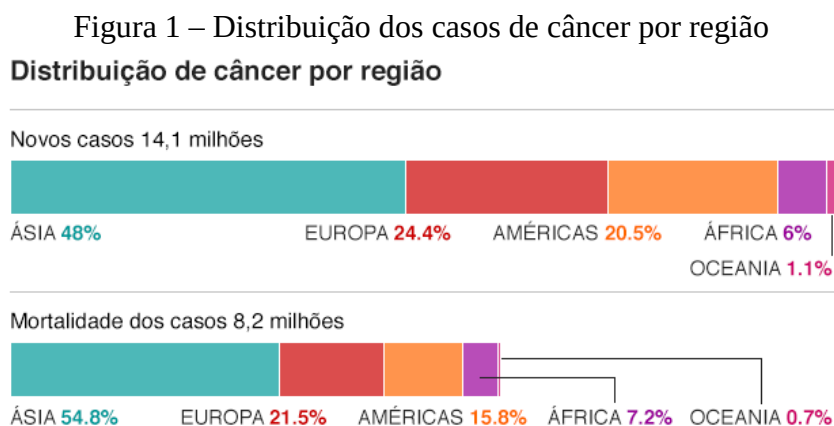
As causas de câncer ainda não são bem esclarecidas, mas ele surge devido a mutações genéticas espontâneas (interna) ou induzidas (externas) por agentes patogênicos como metais, radiações, radicais livres, inflamações crônicas, xenobióticos (tabaco, álcool, pesticidas), entre outros, que promovem desordem no ciclo celular, ocasionando excesso na taxa de proliferação e deficiência nas taxas de morte celular (DA SILVA *et al.*, 2017; FERNANDES, 2016; FERRARI ;TORRES, 2002).

As causas internas são na maioria das vezes, geneticamente pré-determinadas, pois relacionam a capacidade de defesa do organismo às agressões externas. Algumas pessoas apresentam uma predisposição genética para o desenvolvimento de algum tipo de câncer independente da ação do ambiente, entretanto, pessoas que herdaram um gene mutado podem não desenvolver a doença (NUCCI, 2003).

As diferentes categorias de câncer correspondem aos vários tipos de células no corpo, e a nomenclatura é dada conforme o tipo de tecido que lhes deu origem. As três principais categorias de cânceres conhecidos são carcinoma, sarcoma e as leucemias e linfomas. Os carcinomas tem

origem em tecidos epiteliais como pele e mucosa. Os sarcomas possuem origem em tecidos conjuntivos como osso, músculo ou cartilagem. Já as leucemias e linfomas tem origem no sistema linfático, a primeira se desenvolve na medula óssea e a segunda nos gânglios linfáticos. No entanto, há exceções nas quais a nomenclatura pode ser dada utilizando o nome dos cientistas que os descreveram pela primeira vez, sem citar que são tumores tal como a doença de Hodgkin (INCA, 2017).

O câncer é um problema de saúde pública que afeta pessoas de todo o mundo e de todas as idades, e que está em constante progresso. Dados a Organização Mundial de Saúde (OMS) em 2017, relatam que ele é responsável por uma em cada seis mortes no mundo. Mais de 14 milhões de pessoas desenvolvem câncer todos os anos, e esse número deve subir para mais de 21 milhões de pessoas em 2030. A Figura 1 apresenta a distribuição de câncer por região, na qual se pode perceber que os países da Ásia são os principais afetados tanto pelo número de novos casos quanto pela mortalidade.



Fonte: Adaptado GLOBOCAN.

A cada ano 8,8 milhões de pessoas no mundo morrem de câncer, informações baseadas em dados de 2017. Em 2012 essa estimativa era de 8,2 milhões (OMS, 2017). Este problema, afeta principalmente os países em desenvolvimento, onde é esperado que nas próximas décadas, o impacto do câncer na população corresponda a 80% dos mais de 20 milhões de novos casos estimados para 2015 (INCA, 2017).

Dados do INCA relatam que, os tipos de câncer mais incidentes no mundo são pulmão, mama, intestino e próstata. Sendo que nos homens, os mais frequentes são pulmão e próstata, já nas mulheres, o câncer de mama é sem dúvidas o mais comum. A Figura 2 mostra estimativas para o ano de 2018 dos 10 tipos de câncer mais incidentes por sexo.

Figura 2 – Distribuição proporcional dos dez tipos de câncer mais incidentes estimados para 2018 por sexo

Localização Primária	Casos	%	Homens	Mulheres	Localização Primária	Casos	%
Próstata	68.220	31,7%			Mama Feminina	59.700	29,5%
Traqueia, Brônquio e Pulmão	18.740	8,7%			Cólon e Reto	18.980	9,4%
Cólon e Reto	17.380	8,1%			Colo do Útero	16.370	8,1%
Estômago	13.540	6,3%			Traqueia, Brônquio e Pulmão	12.530	6,2%
Cavidade Oral	11.200	5,2%			Glândula Tireoide	8.040	4,0%
Esôfago	8.240	3,8%			Estômago	7.750	3,8%
Bexiga	6.690	3,1%			Corpo do Útero	6.600	3,3%
Laringe	6.390	3,0%			Ovário	6.150	3,0%
Leucemias	5.940	2,8%			Sistema Nervoso Central	5.510	2,7%
Sistema Nervoso Central	5.810	2,7%			Leucemias	4.860	2,4%

Fonte: Estimativa 2018: incidência de câncer no Brasil / Instituto Nacional de Câncer José Alencar Gomes da Silva.

No Brasil, o câncer representa a segunda causa de óbito na população adulta. No biênio 2018-2019, estimativas apontam a ocorrência de cerca de 600 mil casos novos de câncer. O perfil epidemiológico mostra que o câncer de próstata em homens e o de mama em mulheres serão os mais frequentes (INCA, 2018).

A prevenção e o diagnóstico precoce são fatores importantes, que merecem a atenção dos órgãos governamentais, principalmente da área de saúde, mesmo que sem a eficiência e eficácia esperadas (NUCCI, 2003). Como prevenção, nos últimos anos foi desenvolvida a vacina contra o HPV (vírus papiloma humano), que pode prevenir câncer de colo de útero, garganta e ânus (BRASIL, 2013)

O diagnóstico e tratamento dos diferentes tipos de câncer, em todas as idades, sofreram consideráveis avanços nos últimos anos. Métodos de biologia molecular, análises bioquímicas e modernos métodos de imagem têm permitido o diagnóstico apurado, acompanhamento adequado e avaliação do prognóstico dos pacientes (BOING; VARGAS e BOING, 2007; BORSATTO; VIDAL e ROCHA, 2011).

O tratamento do câncer pode envolver algumas etapas como: cirurgia para diagnóstico (biópsia); radioterapia com fins curativos ou atenuantes; quimioterapia para controle temporário da doença, atenuante ou curativa; e hormonoterapia com função medicamentosa (FRIGO; DE OLIVEIRA ZAMBARDA, 2015).

Um número cada vez maior de tipos de câncer já é tratado com sucesso. O diagnóstico nas fases iniciais de desenvolvimento de tumor aumenta a chance de cura total da doença, porém novos tratamentos para alguns tipos de câncer em estágios avançados estão sendo desenvolvidos (BATISTA; DE MATTOS e DA SILVA, 2015).

Na célula normal existe um equilíbrio na regulação positiva e negativa de proliferação determinando um ciclo nuclear normal. No câncer, essa homeostase fica alterada pelo aumento do sinal estimulatório ou pela diminuição dos sinais inibitórios (KIMURA; MATSUO e RICARTE-FILHO, 2007).

Um dos fatores menos compreendidos associados à progressão do tumor é a invasão e metástase do tecido. Este processo muito complexo, geralmente é regulado por uma “conversa” cruzada entre várias vias de sinalização, das quais a mais investigada está a via do Fator de crescimento transformante β (TGF- β , do inglês *transforming growth factor β*), que controla múltiplos aspectos do comportamento celular (GUO; WANG, 2009; ROSEN; ASHURST e CHAP, 2010).

A expressão dessa via tem sido muito estudada na maioria dos cânceres epiteliais: câncer de próstata, pele, mama, pulmão, colorretal e pancreático (EFTIMIE; PEREZ e BUONO, 2017).

2.2. Superfamília TGF- β

Os membros da superfamília TGF- β são reconhecidos como citocinas multifuncionais que desempenham papéis essenciais na regulação do desenvolvimento e diferenciação de mamíferos. Elas regulam a homeostase do tecido, proliferação, diferenciação, apoptose, adesão, mobilidade, migração e invasão celular (EFTIMIE; PEREZ e BUONO, 2017; GUTGOLD *et al.*, 2017; LI *et al.*, 2017).

Essa superfamília consiste em mais de 30 membros diferentes, incluindo as subfamílias de proteínas TGF- β (E.C.2.7.11.30), as ativinas, inibinas, hormônios anti-müllerianos, proteínas

morfogenéticas dos ossos (BMP, do inglês *bone morphogenetic proteins*), fatores de crescimento e diferenciação (GDF, do inglês *growth and differentiation factors*) e família vitelogenina 1 (Vg-1) (HOODLESS; WRANA, 1998; ZHOU *et al.*, 2017). Esses membros são tipicamente homodímeros de duas subunidades de 12-15 kDa ricas em cisteínas, ligadas por uma única ligação dissulfeto (HOODLESS; WRANA, 1998).

A sinalização desregulada por muitos desses membros tem papéis cruciais, tanto no desenvolvimento de tumores, quanto na metástase. Entre os membros da superfamília, a subfamília TGF- β é a mais estudada e seus inibidores são os mais avançados no desenvolvimento clínico (NEUZILLET *et al.*, 2015).

2.2.1. A subfamília TGF- β

A subfamília TGF- β desempenha um papel importante na regulação da formação e progressão do câncer. Ela tem dupla ação, anti e pró-tumoral, dependendo do estágio do tumor e alteração genética (GUTGOLD *et al.*, 2017; NEUZILLET *et al.*, 2015).

Em células normais e carcinomas precoces ela atua como supressor de tumor, promovendo a prisão do ciclo celular e a apoptose. No entanto, troca suas funções para promover a progressão do tumor em casos de câncer mais avançados ou invasivos, possibilitando a mobilidade celular do câncer, invasão e até metástase (LEBRUN, 2012; NEUZILLET *et al.*, 2015). Contudo, ainda não está claro o momento no qual a proteína muda a sua função (DUMONT; ARTEAGA, 2003; EFTIMIE; PEREZ e BUONO, 2017).

Nos mamíferos a TGF- β é dividida em três isoformas distintas, TGF- β 1, TGF- β 2 e TGF- β 3, que são codificadas por um gene diferente e compartilham 70% de homologia (KIMURA; MATSUO e RICARTE-FILHO, 2007; LEBRUN, 2012), contudo estão localizadas em regiões distintas nos cromossomos humanos 19q13, 1q41 e 14q24, respectivamente. Dentre elas, a isoforma mais estudada é a TGF- β 1 (LEBRUN, 2012), pelo fato de sua sequência de aminoácidos ser altamente conservada, já que é idêntica entre humanos e bovinos, e difere em somente um aminoácido nos ratos (BARNARD; LYONS e MOSES, 1990).

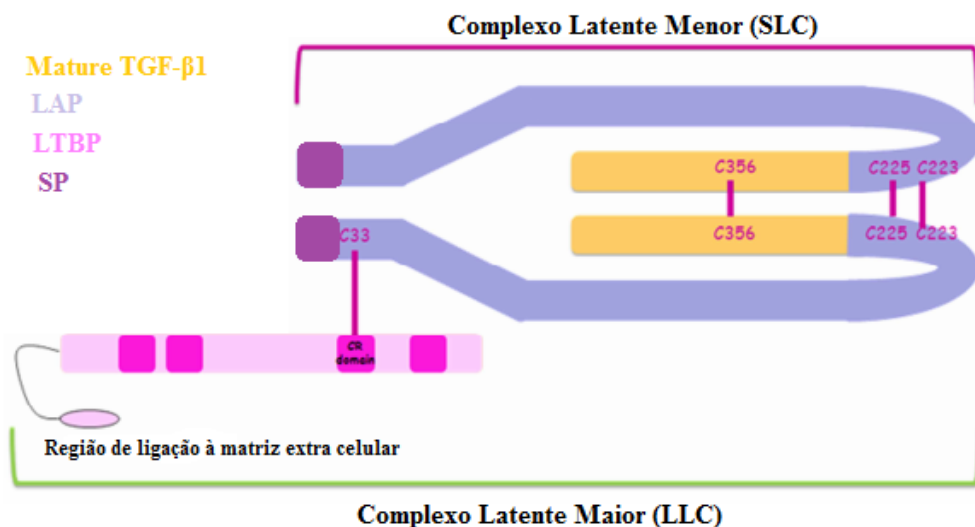
TGF- β 1 é sintetizada como uma proteína precursora de 390 aminoácidos, que pode ser chamada também de pré-pró-TGF- β 1 ou complexo latente menor (SLC, do inglês *small latente*

complex). Este complexo possui três regiões distintas, um fragmento de sinalização N-terminal (SP, do inglês *signal peptide*) contendo 29 resíduos de aminoácidos, o peptídeo associado à latência (LAP, do inglês *latency associated peptide*) que contém 249 resíduos e uma sequência C-terminal (também conhecida como *Mature TGF- β 1*) que contém 112 aminoácidos (Figura 3) (FUENTES-CALVO; MARTÍNEZ-SALGADO, 2013; LEBRUN, 2012).

Os SLCs são secretados por células e depositados na matriz extracelular como complexos covalentes com suas proteínas de ligação (LTBP, do inglês *latent TGF- β binding protein*) formando o complexo latente maior. As LTBP contribuem para estabilidade do TGF- β 1 (FUENTES-CALVO; MARTÍNEZ-SALGADO, 2013).

Em sua via secretora, ocorre a primeira clivagem da porção N-terminal através da glicosilação no peptídeo glicina-leucina transformando-se em pró-TGF- β 1. Na segunda fase ocorre uma clivagem proteolítica separando as cadeias entre os aminoácidos 278 e 279, bem como a dimerização de dois monômeros com três ligações dissulfeto através dos resíduos de cisteína 223, 225 (LAP-LAP) e 365 (*Mature TGF- β 1-Mature TGF- β 1*) (Figura 3). Além disso, LAP sofre mudanças conformacionais mantendo o complexo na sua forma inativa, evitando sua interação com o receptor (KIMURA; MATSUO e RICARTE-FILHO, 2007; ROBERTS, 1991).

Figura 3 – Estrutura TGF- β 1

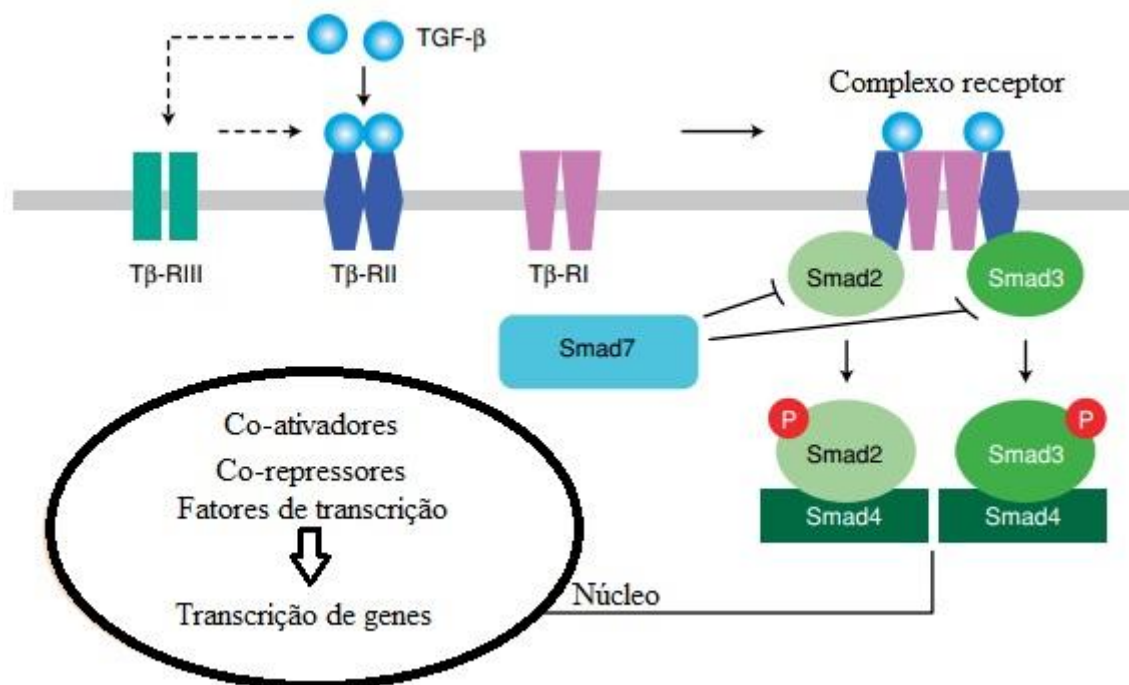


Fonte: Adaptado Fuentes-Calvo e Martínez-Salgado, 2013.

2.3. Sinalização de TGF- β

A sinalização de TGF- β é principalmente mediada por ligantes de TGF- β , dois tipos de receptores serina-treonina quinase e proteínas SMAD (*Mother against decapentaplegic homolog*) (LI *et al.*, 2017).

A transdução de sinais inicia-se com a formação de um complexo hexamérico que compreende dois receptores tipo II (T β RII), dois receptores tipo I (T β RI) e um ligante dimérico, na membrana plasmática (HIRSCHHORN *et al.*, 2017; HUI ;FRIEDMAN, 2003). Dentro desse complexo, os receptores tipo II fosforilam e ativam os receptores tipo I, que propagam a sinalização através da ativação e fosforilação de proteínas citoplasmáticas da família de SMADs, conhecidas como *receptor-regulated* SMADs (R-SMADs), SMAD2 e SMAD3. A fosforilação de R-SMADs permite a sua interação com *common-mediator* SMAD (co-SMAD), SMAD4, formando um complexo capaz de atingir o núcleo, onde irão modular a transcrição de genes-alvo (HIRSCHHORN *et al.*, 2017; KIMURA; MATSUO e RICARTE-FILHO, 2007). A sinalização é mostrada na Figura 4.

Figura 4 – Via de sinalização TGF β / Smad

Legenda: Nesta via, os dímeros de TGF- β (Mature TGF- β) se ligam a um receptor do tipo II (T β -RII). Esta ligação pode ser reforçada pela presença de receptores do tipo III. Depois de ligar ao TGF- β , o T β -RII recruta e fosforila receptores do tipo I (T β -RI). O receptor do tipo I, em seguida recruta e fosforila as proteínas Smad2 e 3, (processo inibido por Smad7) que então se liga ao Smad4 formando um complexo heterodimérico. Este complexo entra no núcleo da célula onde atua como um fator de transcrição de vários genes.

Fonte: Adaptado Hui e Friedman, 2003.

2.4. Inibidores de TGF- β

Visto que, em casos de cânceres avançados ou invasivos a TGF- β atua na progressão do tumor, a inibição da sua via de sinalização pode ser uma estratégia apropriada para o tratamento desses tipos de cânceres, uma vez que existem poucas alternativas terapêuticas (LEBRUN, 2012).

A via de sinalização de TGF/SMAD é controlada por múltiplos mecanismos autorreguladores que existem para manter reações induzidas por TGF- β fortemente reguladas.

Muitos inibidores da via TGF- β foram investigados no aspecto pré-clínico e alguns já estão em desenvolvimento clínico (NEUZILLET *et al.*, 2015).

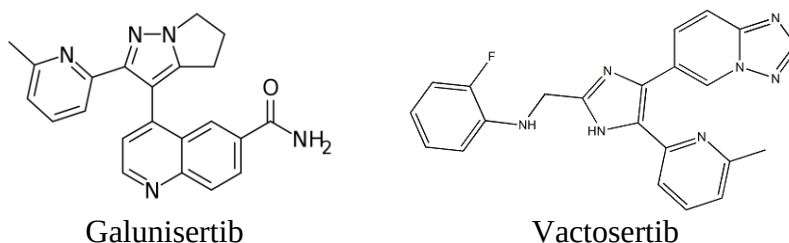
A inibição dessa via pode ser realizada em três níveis:

(I) Nível do ligante: nucleotídeos que previnem a síntese de TGF- β (como exemplo trabedersen e Lucanix);

(II) Nível ligante-receptor: armadilhas de ligante que previnem a interação entre ligante-receptor (como exemplo fresolmumab e disitertide) e;

(III) Nível intracelular: inibidores do receptor quinase de TGF- β que previnem o sinal de transdução (como exemplo galunisertib e vactosertib) (NEUZILLET *et al.*, 2015).

Figura 5 – Estrutura química do Galunisertib e Vactosertib



Nota: Galunisertib é um inibidor TGF- β 1 que está em desenvolvimento clínico mais avançado.

Fonte: Do autor (2018)

No nível intracelular, a primeira via inibidora de TGF- β /SMAD que foi descrita envolve um membro da família SMAD, o SMAD 7, que é classificado como SMAD inibitório (I-SMAD). Ele funciona através de um mecanismo de resposta negativo que encerra a sinalização, pois bloqueia a fosforilação dos R-SMADs, interferindo o acesso de SMAD 2/3 para o domínio da quinase do receptor tipo I (DERYNCK; AKHURST e BALMAIN, 2001; KIMURA; MATSUO e RICARTE-FILHO, 2007; LEBRUN, 2012).

Além disso, o SMAD 7 também recruta proteínas fosfatases e ubiquitina ligases ao receptor TGF- β ativado, favorecendo ainda mais para o término da sinalização (LEBRUN, 2012).

Os inibidores exercem sua atividade antitumoral principalmente afetando células responsivas ao TGF- β no microambiente do tumor. A sinalização influencia o microambiente tumoral através da promoção da fibrose, angiogênese e metástase. A inibição de TGF- β

normaliza a homeostase do microambiente do tumor, com impacto direto nas células cancerosas e limitados efeitos colaterais (NEUZILLET *et al.*, 2015).

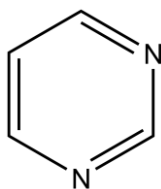
Como os inibidores de TGF- β objetivam principalmente o microambiente do tumor, eles devem ser usados em combinação com agentes citotóxicos para matar as últimas células. Essa combinação pode aumentar a sensibilidade do tumor à quimioterapia e radioterapia (NEUZILLET *et al.*, 2015). O estudo desses inibidores ainda está em progresso e são necessários muitos esforços para melhoria dos mesmos.

2.5. Compostos pirimidínicos

Vários medicamentos comercializados e produtos em fase de testes clínicos possuem em sua estrutura núcleos heterocíclicos. Esses compostos possuem uma grande importância pela sua vasta aplicação em diversos campos da química e devido à grande variedade e complexidade estrutural, o que possibilita criar uma variada série de estruturas novas, com propriedades físicas e químicas diversas (LUNA *et al.*, 2011).

Dentre os núcleos heterocíclicos, destacam-se os pirimidínicos (Figura 6) e seus derivados devido a sua importância medicinal e biológica. Eles possuem várias aplicações farmacológicas, sendo utilizados como agentes antiinflamatório, antitumorais (KAWAUCHI *et al.*, 1994), antivirais (MAGNUS; CONFALONE e STORACE, 2000), anti-hipertensivos e ansiolíticos (LUNA *et al.*, 2011).

Figura 6 – Estrutura de um núcleo pirimidínico



Fonte: Do autor (2018)

Portanto, o planejamento racional de fármacos pode ajudar no desenvolvimento de novos agentes antitumorais baseados em derivados pirimidínicos, uma vez que visa a acelerar o

processo de descoberta de novas substâncias que possuam propriedades úteis, capazes de representar novos protótipos de fármacos de uso seguro (BARREIRO, ELIEZER J *et al.*, 2002).

2.6. A química medicinal e o planejamento de fármacos

O processo de descoberta e desenvolvimento de novos fármacos é complexo, longo e de alto custo. Segundo Guido, Andricopulo e Oliva (2010), a descoberta de inovações terapêuticas proporcionaram melhorias significativas na qualidade de vida das diversas populações no mundo. Essas descobertas foram possíveis a partir de avanços nas áreas de química, biologia e a melhor compreensão de vias bioquímicas, alvos moleculares e de mecanismos que levam ao aparecimento e desenvolvimento de doenças. Nesse aspecto, a química medicinal destaca-se nos processos de produção e desenvolvimento de fármacos.

A química medicinal possui caráter multidisciplinar, e engloba áreas como a farmacologia, bioquímica, química orgânica sintética, química computacional, entre outras (VERLI; BARREIRO, 2005). De acordo com a União Internacional de Química Pura e Aplicada (IUPAC), a química medicinal é uma disciplina com base em química, envolvendo a invenção, a descoberta, o planejamento, a identificação, a preparação e a interpretação do mecanismo de ação molecular de compostos biologicamente ativos. Sendo assim, ela estuda as razões moleculares da ação dos fármacos, de maneira a descrever a relação entre a estrutura química e atividade farmacológica (BARREIRO, E. J. *et al.*, 2002).

Cada medicamento comercializado carrega uma riqueza de dados clínicos que podem ser aproveitados em conjunto com informações sobre a estrutura química, e, portanto, dentro de todos os tipos de análises quimioinformáticas (uso de técnicas computacionais aplicadas a problemas químicos) (SCHEIBER, 2012). A combinação de tecnologias biológicas com o meio ambiente rico em informações, criado a partir de métodos computacionais modernos, fornece uma oportunidade de acelerar o processo de identificação de candidatos a novos fármacos (DUFFY *et al.*, 2012). Esse processo inicia-se pela escolha correta do alvo-terapêutico relacionado à patologia que se pretende tratar, que posteriormente passa por validação terapêutica e pela identificação ou descoberta de novos padrões moleculares de substâncias que representem

entidades químicas capazes de serem reconhecidas com níveis adequados de seletividade pelos alvos terapêuticos eleitos, promovendo resposta biológica (BARREIRO; FRAGA, 2005).

As moléculas bioativas podem ser identificadas a partir de triagens reais ou virtuais de produtos naturais, compostos sintéticos ou coleções combinatórias ou, ainda, mediante planejamento racional. Com o auxílio de métodos de química medicinal, é possível explorar o imenso espaço químico, delineando o trabalho de identificação, seleção e otimização de moléculas bioativas (GUIDO; ANDRICOPULO e OLIVA, 2010).

A atividade biológica de um composto é o resultado das interações desta molécula com o sistema biológico. O estudo dessas interações pode ser realizado por duas aproximações: métodos independentes ou dependentes do receptor e/ou enzimas. No primeiro, as interações com a macromolécula são indiretas utilizando modelos de correlação entre a estrutura das substâncias e suas atividades, sendo conhecida como relações quantitativas estrutura-atividade (QSAR), no segundo caso, as interações dos compostos com a macromolécula são consideradas diretamente, como exemplo os métodos de ancoramento molecular (SANT, 2009).

2.7. Relação quantitativa estrutura- atividade

Pequenas modificações estruturais alteram a atividade biológica de moléculas, esses efeitos podem ser explorados com base no conhecimento de química, experiência e intuição, bem como com o auxílio de ferramentas computacionais (WASSERMANN *et al.*, 2012). Esse fato, apesar de já ser conhecido anteriormente, só teve relevância há cerca de cinco décadas, quando Hansch e Fujita (1964) propuseram um método voltado para o estudo de relações quantitativas entre estrutura química e atividade biológica (QSAR) (HANSCH; FUJITA, 1964).

A teoria subjacente do QSAR é que a atividade biológica das estruturas está relacionada com características moleculares específicas (descritores), e que técnicas de regressão linear podem ser utilizadas para avaliar a importância relativa das características que contribuem para o efeito biológico. Os modelos de QSAR são construídos utilizando-se a atividade biológica como variável resposta (y) e os descritores moleculares como as variáveis independentes (x) (SANTOS-FILHO *et al.*, 2009).

A análise de Hansch e Fujita (1964) se aplica a estudos de uma série de compostos que apresentam estruturas químicas similares, sendo que, em linhas gerais, tais estruturas diferem entre si por variações de grupos substituintes em posições definidas. Neste modelo há a correlação entre valores de atividade biológica de série de compostos com suas respectivas propriedades físico-químicas, na tentativa de identificar e caracterizar a influência de cada fator físico-químico envolvido no desencadeamento da atividade biológica. Cita-se que esta correlação pode se dar por regressão linear simples (RLS), regressão linear múltipla (RLM) ou regressão não-linear (RNL). As regressões lineares simples e múltipla consideram que a relação entre as variáveis é descrita por uma função linear, e se diferem pelo fato que, na RLS existe apenas uma variável independente enquanto que na RLM podem existir várias. Já para regressão não linear, a relação entre as variáveis não pode ser descrita por uma função linear (TAVARES, 2004).

Depois da metodologia de QSAR de Hansch e Fujita (1964), ocorreram vários avanços interdisciplinares, e o crescimento de dados químicos e bancos de dados estimulou o crescimento de pesquisas, o que resultou na criação de várias metodologias com diferentes abordagens de cálculos, descritores moleculares e algoritmos estatísticos para a construção dos modelos resultantes (CHERKASOV *et al.*, 2014).

Os estudos de QSAR podem ser agrupados em dois grandes grupos: independentes do receptor (QSAR-IR) e dependente do receptor (QSAR-DR). O primeiro grupo inclui o QSAR unidimensional (1D), bidimensional (2D), tridimensional (3D) e 4D que inclui a quarta dimensão. Os descritores calculados são características moleculares reconhecíveis como peso molecular e soma das propriedades atômicas (1D), descritores topológicos (2D), coordenadas atômicas, geometria ou energia dos descritores na caixa (3D), e combinação de coordenadas atômicas e amostragem de conformações (4D) (SANTOS-FILHO *et al.*, 2009).

Na análise de QSAR-DR, estão incluídos os modelos de QSARs 4D, 5D, 6D e 7D. (RODRIGUES *et al.*, 2012). O QSAR-4D (HOPFINGER *et al.*, 1997) é derivado da estrutura 3D e das múltiplas conformações ligante-receptor, QSAR-5D (VEDANI; DOBLER, 2002) que considera o encaixe induzido, QSAR-6D (VEDANI; DOBLER e LILL, 2005) que avalia diferentes modelos de solvatação simultaneamente e finalmente o QSAR-7D (POLANSKI, 2009)

que inclui dados dos receptores reais ou modelos de receptores virtuais baseados em alvo como, por exemplo, os obtidos por modelagem de proteínas por homologia.

Se utilizados corretamente, os métodos computacionais para análise da correlação estrutura-atividade podem fornecer um menor custo e um menor tempo no desenvolvimento de um novo fármaco quando comparados a procedimentos empíricos, que consistem em testar diversas substâncias químicas obtidas por extração de matérias-primas naturais ou por síntese química na tentativa de procurar uma determinada atividade farmacológica (LIMA, 2007).

2.7.1. QSAR 4D

Em 1997, Hopfinger e colaboradores propuseram uma nova metodologia de QSAR chamada de QSAR-4D. Este método envolve o cálculo de simulações de dinâmica molecular (SDM) para cada molécula, gerando um número de confôrmeros que constituem a quarta dimensão. Além da flexibilidade conformacional, a possibilidade de se definir mais de um alinhamento é uma evolução quando comparado ao QSAR-3D.

Esta metodologia possui alguns passos que são: simulação de dinâmica molecular, definição do tamanho das células, definição dos elementos de interação farmacofórica, definição dos tipos de ocupação e definição dos alinhamentos, que estão detalhadas a seguir.

2.7.1.1. Simulação de dinâmica molecular (SDM)

Na metodologia de QSAR-4D é importante prever o estado conformacional de mais baixa energia do conjunto de amostragem de cada composto (DA ROCHA PITA, 2006). Em função disso, a Simulação por Dinâmica Molecular (SDM) é aplicada para fornecer informações sobre o comportamento dinâmico do sistema molecular e gerar um perfil de amostragem conformacional (CEP, do inglês *Conformational Ensemble Profile*) (HOPFINGER et al., 1997).

2.7.1.2. Definição do tamanho das células

Cada conformação do CEP de cada molécula é inserida em uma caixa tridimensional virtual. A caixa é composta por células cúbicas que registram a distribuição de ocupação espacial de cada átomo de cada ligante. Quanto menor o tamanho da célula, maior será o

número de variáveis geradas e, conseqüentemente, o tempo de análise aumenta, em compensação, os modelos obtidos serão mais refinados em termos dos tipos de átomos selecionados (HOPFINGER et al., 1997).

2.7.1.3. Definição dos elementos de interação farmacofórica

Os descritores representam as frequências de ocupação dos diferentes tipos de átomos nas caixas cúbicas, durante a SDM (ANDRADE *et al.*, 2010). Esses descritores, chamados de GCODs (do inglês, *grid cell occupancy descriptors*) podem ser gerados a partir de diferentes tipos de interações farmacofóricas, que em QSAR-4D são chamados de IPEs (*interaction pharmacophore elements*). Esses IPEs são definidos como A (qualquer tipo), NP (não polar), P+ (carga polar positiva), P- (carga polar negativa), HA (aceptor de ligação de hidrogênio), HB (doador de ligação de hidrogênio) e Ar (aromático), e correspondem as interações que podem ocorrer no sítio ativo. Assim, os IPEs estão relacionados com a natureza dos descritores na análise de QSAR-4D, enquanto os GCODs são relacionados com as coordenadas de IPE mapeado numa rede comum. Na metodologia de QSAR-4D, um conjunto contendo o perfil conformacional de cada composto é utilizado para gerar as variáveis independentes (GCODs), em vez de, apenas, uma conformação de partida (ANDRADE *et al.*, 2010).

2.7.1.4. Definição dos tipos de ocupação

A frequência de ocupação das células da caixa virtual pelos átomos classificados de acordo com o tipo de IPE forma o conjunto de descritores que será utilizado na construção dos modelos de QSAR-4D. A frequência de ocupação das células da caixa virtual pode ser calculada, utilizando-se três tipos de medida: a ocupação absoluta que corresponde à frequência de ocupação das células pelos átomos das conformações dos complexos, independente de um complexo de referência; a ocupação parcial, onde se subtrai o valor da ocupação de um complexo de referência; e ocupação reunida, na qual soma-se o valor da ocupação de um complexo de referência (HOPFINGER et al., 1997). A frequência de ocupação das células da caixa virtual é computada para a faixa de conformações selecionadas da trajetória da SDM para todos os complexos em estudo.

2.7.1.5. Definição do alinhamento

A geração de modelos de QSAR consistentes depende da composição dos conjuntos de treinamento e teste. O conjunto de treinamento é aquele que será utilizado para gerar os modelos de QSAR, e o conjunto teste é utilizado para validação desses modelos (SILVA; RAMALHO e DA CUNHA, 2014).

O alinhamento é a maneira pela qual os compostos do conjunto de treinamento são comparados. A qualidade do modelo gerado depende do alinhamento testado, sendo assim, esta é considerada uma das etapas mais importantes na metodologia de QSAR-4D (HOPFINGER et al., 1997).

Em geral, os alinhamentos são escolhidos de modo a abranger uma estrutura em comum dos compostos dos grupos treinamento e teste. Alinhamentos usando átomos da direita, esquerda e do meio, ou que usam átomos que abrangem toda a região similar dos ligantes devem ser usados para garantir um bons resultados (DA CUNHA *et al.*, 2005).

A ordem dos átomos escolhidos para cada alinhamento é importante, já que, durante o procedimento, o primeiro átomo selecionado é posicionado na origem dos três eixos cartesianos e a molécula é girada de modo que o segundo átomo se posicione no eixo x, e o terceiro se posicione no eixo y (HOPFINGER et al., 1997).

2.8. Ancoragem molecular (Docking)

A aplicação de métodos computacionais para estudar a formação de complexos intermoleculares é alvo de muitas pesquisas. Dentre estes métodos, o ancoramento proteína-ligante se destaca por ser uma técnica importante para o planejamento de novos agentes quimioterápicos (GONÇALVES, 2008).

A atividade biológica de um composto é consequência das interações deste com o sistema biológico, as quais são de diferentes intensidades e naturezas químicas. Intensidades que dependem da complementaridade estérica e eletrostática entre ligante e receptor (GONÇALVES, 2008).

O processo de ancoramento molecular, conhecido também por *Docking*, envolve a predição da conformação do ligante e sua orientação dentro do sítio de interação da proteína,

utilizando a energia total envolvida no processo para identificar a conformação do ligante energeticamente mais favorável quando ligado ao alvo (KITCHEN *et al.*, 2004).

Durante uma simulação de *docking* são geradas várias conformações para o ligante dentro do sítio de interação da proteína. Portanto, torna-se necessário um modelo, que permita avaliar todas as possíveis posições adquiridas pelo ligante e sua estabilidade ao longo da simulação, e então escolher a estrutura mais estável do complexo. Este modelo de seleção pode ser expresso pela combinação de dois componentes: um algoritmo de busca e uma função de pontuação (RAMALHO *et al.*, 2016; THOMSEN; CHRISTENSEN, 2006).

O algoritmo de busca realiza a predição da conformação e orientação do ligante em relação ao sítio de ligação da macromolécula receptora e a função de pontuação é capaz de classificar a afinidade ou energia livre de ligação intermolecular, a partir da hipótese de que o *score* de energia mais baixo representa a melhor orientação após o *docking* (RAMALHO *et al.*, 2016). Dessa forma, essa metodologia determina a conformação mais provável do ligante na enzima, levando em conta o modo de interação de menor energia (THOMSEN; CHRISTENSEN, 2006).

Os valores da função de desempenho do ancoramento (Docking Scoring Function), E_{score} , são definidos pela Eq. 1

$$E_{score} = E_{inter} + E_{intra} \quad (1)$$

Em que, o termo E_{inter} , corresponde a energia de interação ligante-proteína, definido de acordo com a Eq. 2.

$$E_{inter} = \sum_{i \in \text{ligante}} \sum_{j \in \text{proteína}} \left[E_{PLP}(r_{ij}) + 332.0 \frac{q_i q_j}{4r_{ij}^2} \right] \quad (2)$$

O primeiro termo E_{PLP} representado na Eq. 2 é um termo de energia potencial inibidor por partes que usa dois conjuntos diferentes de parâmetros: um para a aproximação do termo estérico (van der Waals) entre átomos, e outro potencial para a ligação hidrogênio. O segundo é um

potencial de Coulomb que descreve as interações eletrostáticas entre átomos carregados e possui uma constante dielétrica dependente da distância ($D(r) = 4r$). O valor numérico de 332,0 fixa as unidades de energia eletrostática para kcal.mol^{-1} (RAMALHO *et al.*, 2009; THOMSEN; CHRISTENSEN, 2006).

O termo E_{intra} que corresponde a energia interna do ligante é definido de acordo com a Eq. 3.

$$E_{\text{intra}} = \sum_{i \in \text{ligante}} \sum_{j \in \text{ligante}} E_{\text{PLP}}(r_{ij}) + \sum_{\text{ligações flexíveis}} A[1 - \cos(m \cdot \theta - \theta_0)] + E_{\text{penalizada}} \quad (3)$$

O primeiro termo refere-se à soma de todos os pares de átomos do ligante excluindo os pares de átomos conectados por duas ligações. O segundo termo refere-se à energia de torção, em que θ é o ângulo de torção da ligação. A média da contribuição de torção da ligação da energia é usada se diversas torções são determinadas. O último termo, $E_{\text{penalizada}}$ (termo de correção), atribui uma penalidade de 1.000 se a distância entre dois átomos pesados (mais de duas ligações distantes) for menor que 2,0 Å, punindo conformações inviáveis do ligante. Em resumo, essas funções são usadas para ancorar automaticamente uma molécula flexível em uma molécula molde (proteína).

2.9. Métodos e índices estatísticos

2.9.1. Regressão pelo método dos mínimos quadrados parciais (PLS)

As técnicas de seleção de variáveis têm como objetivo encontrar um subconjunto de preditores capaz de melhorar os resultados ou pelo menos mantê-los constantes em termos de erro.

Os métodos lineares de regressão multivariada relacionam duas ou mais variáveis independentes (x) à variável dependente (y) como demonstrado na Equação 4.

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n + e \quad (4)$$

em que b_i são os coeficientes de regressão, n o número de variáveis independentes e “ e ” o valor residual (FERREIRA, 2002; GELADI; KOWALSKI, 1986).

Abordagens como regressão por mínimos quadrados parciais (PLS, do inglês *partial least squares*) utilizam da transformação da matriz de descritores X em uma nova matrix X' que pode ser apresentada em um espaço de menor dimensão sem perder muita informação. As novas variáveis formadas são denominadas variáveis latentes (VL), que são combinações lineares das variáveis originais, e são utilizadas para maximizar a covariância entre as variáveis independentes e dependentes (FERREIRA, 2002; GELADI; KOWALSKI, 1986; PIRHADI; SHIRI e GHASEMI, 2015).

Desse modo, o PLS permite relacionar um grande número de variáveis com uma ou mais variáveis resposta, sendo, portanto bastante empregado em estudos de QSAR (FERREIRA; MONTANARI e GAUDIO, 2002; PIRHADI; SHIRI e GHASEMI, 2015).

2.9.2. Aproximação da função genética (GFA)

A aproximação da função genética (GFA, do inglês *Genetic Function Approximation*), é um algoritmo desenvolvido por Rogers e Hopfinger (1994) baseado em algoritmo genético (GA, do inglês *Genetic Algorithm*), sendo, portanto um algoritmo de otimização inspirado em princípios evolucionários.

No GFA, os modelos iniciais são formados por acaso, selecionando algumas variáveis independentes, descritores, para criação de uma população de modelos. Nesta etapa, são realizados processos de cruzamento e mutação nos modelos, que, combinando os melhores termos, gera modelos cada vez melhores. Esses modelos são classificados segundo os valores do fator de desajuste de Friedman (LOF) (ROGERS; HOPFINGER, 1994).

2.9.3. Índice de Falta de Ajuste (LOF, lack-of-fit)

A metodologia GFA/PLS inclui um índice denominado de “índice de falta de ajuste” (em inglês, *lack-of-fit*, LOF) implementado por Friedman (KARKI; KULKARNI, 2001) definido pela Equação 5.

$$LOF = \frac{LSE}{\left\{1 - \left(c + \frac{dp}{M}\right)\right\}^2} \quad (5)$$

em que C é o número de termos no modelo, d é *smoothing factor*, p é o número total de variáveis contido em cada termo e M é o número de compostos, LSE erro dos mínimos quadrados, calculado pela diferença entre os valores experimentais e preditos da atividade. Assim, quando dois métodos têm o mesmo erro de mínimos quadrados, aquele que possuir o menor número de variáveis independentes terá o menor valor de LOF e, portanto, será o escolhido (ROGERS; HOPFINGER, 1994).

2.9.4. Coeficiente de Correlação Linear Quadrático (r² ou R²)

A validação é um passo crucial para a construção dos modelos de QSAR, pois é ela que vai nos mostrar se os modelos obtidos são confiáveis e capazes de prever a atividade de moléculas congêneres.

A correlação é um instrumento adequado para descobrir e medir a relação entre duas variáveis. O coeficiente de correlação linear quadrático, conhecido também como coeficiente de determinação (r²) é uma medida de ajustamento de um modelo estatístico linear generalizado em relação aos valores observados. Um valor de r² maior que 0,7, indica que o modelo está correlacionado e pode ser utilizado para calcular a atividade do conjunto de teste. (CALDAS; RAMALHO e DA CUNHA, 2014). Na Equação 6, mostra-se como o coeficiente de determinação é calculado

$$r^2 = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{pred})^2}{\sum_{i=1}^n (y_i^{obs} - y_{med}^{obs})^2} \quad (6)$$

Em que y_i^{obs} e y_i^{pred} são os valores de atividade experimental e atividade predita para o grupo de treinamento e y_i^{obs}_{med} representa o valor médio das atividades experimentais para o grupo de treinamento.

2.9.5. Validação Interna dos modelos

A técnica de validação interna é denominada de “*leave-one-out cross-validation*” (LOO-cv, q^2), na qual cada composto do conjunto de treinamento é eliminado por vez e os coeficientes da equação são recalculados até que todos os componentes tenham sido eliminados uma vez. O índice estatístico gerado nesse processo é o coeficiente de correlação quadrático da validação cruzada, q^2 ou r^2_{cv} , (Equação 7) (SENESE *et al.*, 2004).

$$q^2_{cv} = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{predcv})^2}{\sum_{i=1}^n (y_i^{obs} - y_{med}^{obs})^2} \quad (7)$$

Nessa equação, y_i^{obs} são os valores de atividade experimental para o grupo de treinamento, y_i^{predcv} são os valores de atividade calculados na validação cruzada e y_{med}^{obs} é o valor médio das atividades experimentais para o grupo de treinamento. Um modelo é considerado aceitável quando o valor de q^2 excede 0,5 (PRATIM ROY *et al.*, 2009).

2.9.6. Validação Externa dos modelos

O propósito da validação externa é testar o poder preditivo do modelo. Para isso, são utilizadas moléculas que não fizeram parte da construção do modelo e que são denominadas de grupo teste. As equações são geradas com base nos compostos do conjunto de treinamento e, por meio do modelo obtido calcula-se a atividade predita dessas moléculas. A capacidade preditiva dos modelos é julgada com base na previsão R^2_{pred} , cujo valor é calculado de acordo com a Equação 8. Esse processo, verificará a capacidade do modelo de prever a atividade biológica de moléculas congêneres (KIRALJ; FERREIRA, 2009).

$$R^2_{pred} = 1 - \frac{\sum_{j=1}^k (y_i^{obs} - y_j^{pred})^2}{\sum_{j=1}^k (y_i^{obs} - y_{med}^{obs})^2} \quad (8)$$

onde y_i^{obs} e y_i^{pred} , correspondem aos valores experimentais e preditos dos compostos do conjunto teste, respectivamente e y_{med}^{obs} corresponde à média das atividades experimentais dos

compostos do conjunto de treinamento. Para um modelo preditivo de QSAR, o valor de R^2_{pred} deve ser maior que 0,5 (PRATIM ROY *et al.*, 2009).

O valor de R^2_{pred} é altamente dependente da média do grupo de treinamento então, somente o seu uso não garante uma verdadeira reflexão da predição externa. Portanto, Pratim Roy e colaboradores (2009) definiram uma equação modificada para R^2_{pred} , que dá a importância para a diferença entre r^2 e r^2_0 . Essa Equação está apresentada abaixo (Equação 9):

$$r_m^2 = r^2(1 - |\sqrt{r^2} - r_0^2|) \quad (9)$$

em que r^2 é o quadrado do coeficiente de correlação entre os valores observados e preditos e r_0^2 é o quadrado do coeficiente de correlação entre os valores observados e preditos com interseção no eixo zero. Valores maiores que 0,5 podem ser um bom indicativo de uma boa previsibilidade externa.

2.9.7. Validação Y-randomization (R^2 Y-rand) e R^2P

A finalidade do teste Y-randomization é detectar e quantificar a chance de correlação entre as variáveis independentes e os descritores. A significância estatística dessa relação pode ser verificada pelo teste de aleatorização, que é quando os valores das variáveis independentes são trocados de forma aleatória. Para um modelo de QSAR aceitável, o valor de $R^2_{Y\text{-rand}}$ deve ser menor que o coeficiente de correlação do modelo não randomizado (KIRALJ; FERREIRA, 2009).

Para essa validação, usa-se um parâmetro que penaliza o modelo pela diferença entre $R^2_{Y\text{-rand}}$ e R^2 . Esse parâmetro pode ser calculado pela seguinte equação (Equação 10):

$$R_p^2 = r^2 \sqrt{r^2 - R_{Y\text{-rand}}^2} \quad (10)$$

Na equação, r^2 é o coeficiente de determinação e $R^2_{Y\text{-rand}}$ é o valor encontrado após a randomização. O valor encontrado para esse parâmetro deve ser maior que 0,5 para um modelo aceitável (PRATIM ROY *et al.*, 2009).

3. OBJETIVOS

Conduziu-se este trabalho, com o objetivo de propor novos compostos que possam ser futuros protótipos para o tratamento de diversos tipos de câncer, tendo como objetivos específicos:

- Determinar quantitativamente a influência de descritores estruturais na atividade de derivados aril pirimidina como inibidores da proteína TGF- β 1, a partir do estudo de QSAR-4D.
- Analisar o modo de interação desses inibidores junto à proteína a fim de verificar aminoácidos importantes para a estabilidade do complexo.
- A partir dos resultados observados, propor novas estruturas que sejam mais ativas do que as estudadas.
- Testar os novos compostos a fim de verificar as interações realizadas e energia de ancoramento molecular.

4. METODOLOGIA

Este estudo foi realizado no Laboratório de Modelagem Molecular do departamento de Química da Universidade Federal de Lavras- DQI/UFLA. Foram utilizados dois computadores: um Intel Core i7 e uma Silicon Graphics O2.

4.1. Dados Biológicos

Na Tabela 1, estão relacionados os sessenta e três compostos derivados aril pirimidina que foram utilizados neste trabalho e seus respectivos valores de inibição de sinalização. Esses compostos foram sintetizados e avaliados farmacologicamente por Gelman e colaboradores (GELMAN *et al.*, 2016). É importante ressaltar que todos os dados foram obtidos do mesmo teste farmacológico, eliminando potenciais erros que poderiam ser introduzidos no conjunto de dados, caso estes fossem provenientes de diferentes fontes. Os valores de inibição expressos em μM foram convertidos em unidades molar e posteriormente em unidades negativa de logaritmo (-log).

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

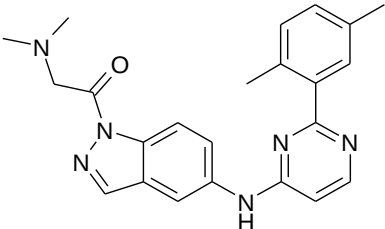
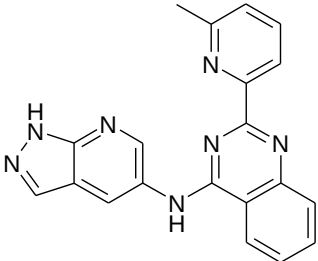
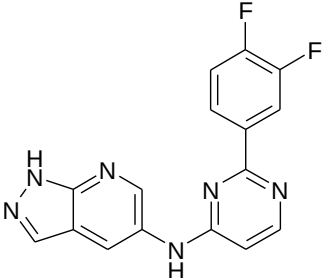
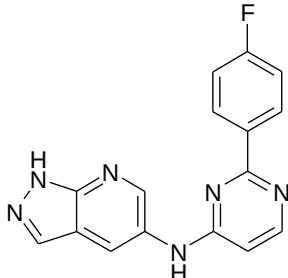
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
1		6,25	2		6,49
3		5,81	4		6,08

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

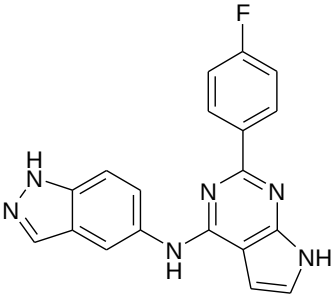
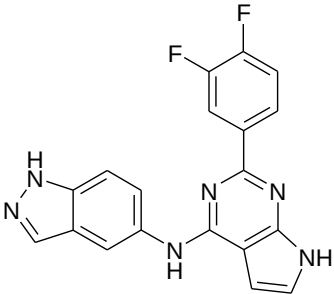
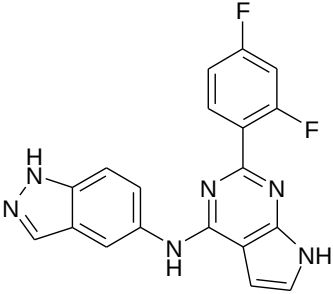
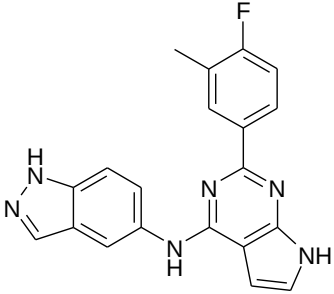
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
5		6,12	6		5,77
7		6,06	8		6,13

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

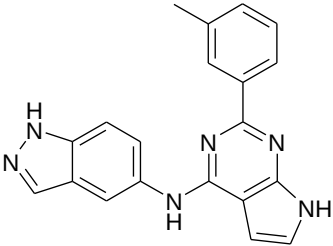
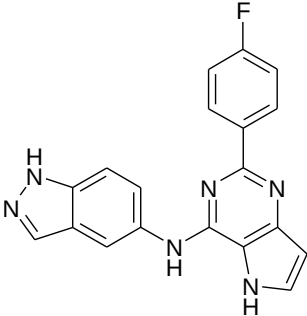
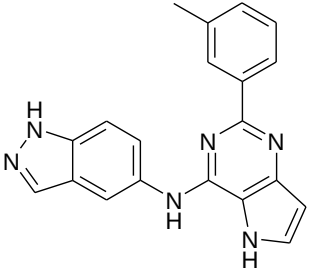
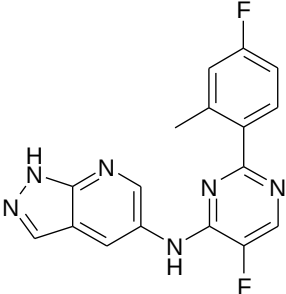
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
9		5,84	10		5,29
11		5,27	12		5,79

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

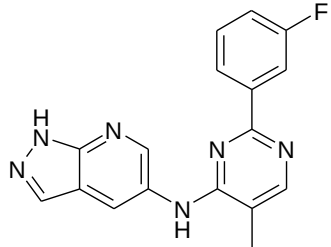
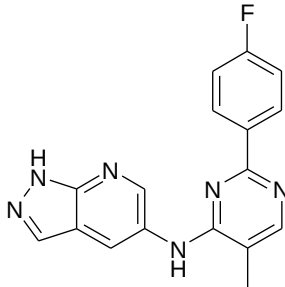
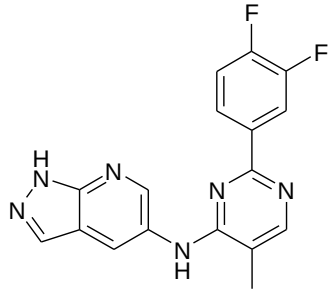
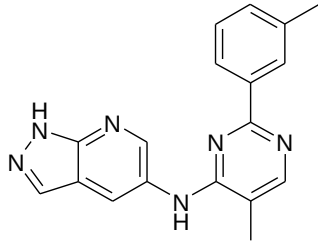
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
13		5,56	14		6,14
15		6,06	16		6,39

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

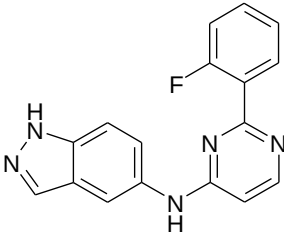
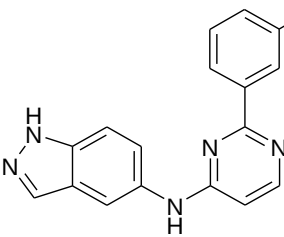
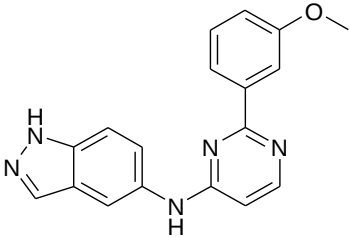
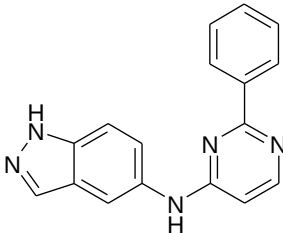
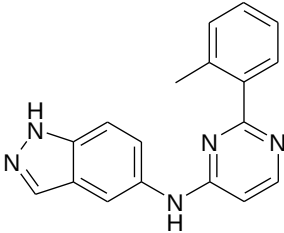
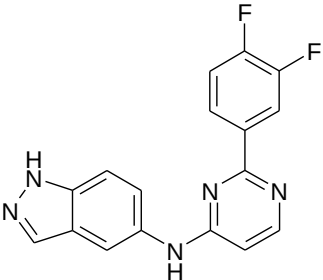
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
17		5,66	18		5,87
19		5,39	20		5,76
21		5,91	22		6,95

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

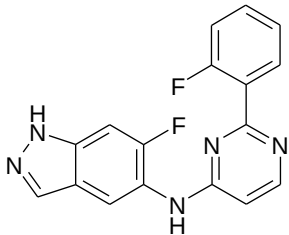
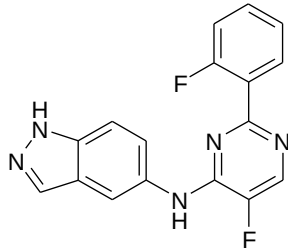
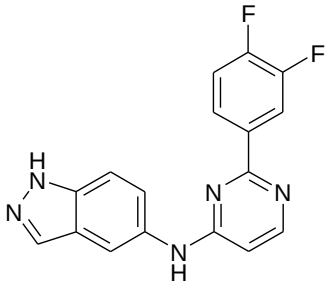
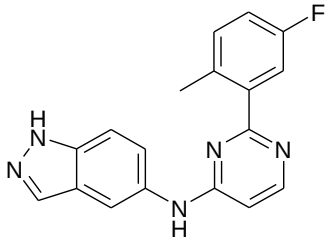
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
23		5,88	24		5,39
25		6,49	26		6,00

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

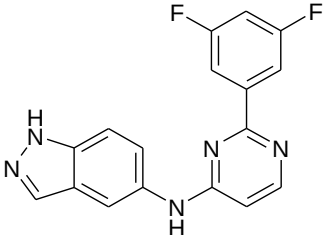
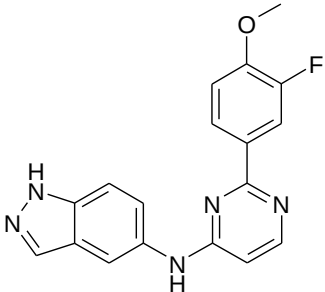
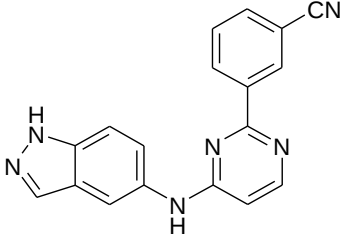
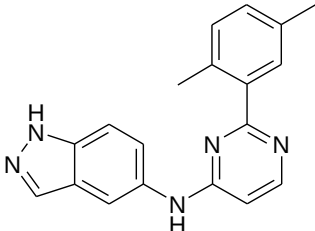
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
27		5,66	28		5,67
29		7,07	30		6,07

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

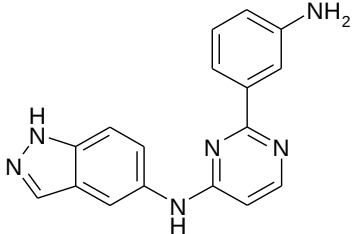
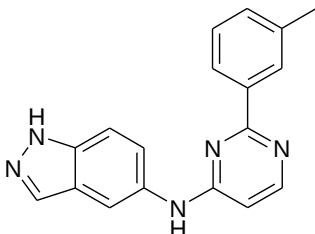
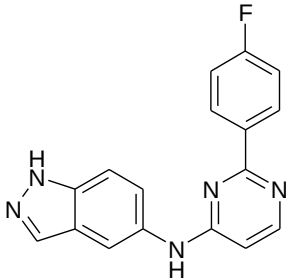
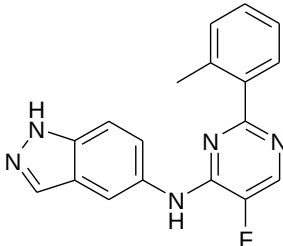
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
31		5,57	32		6,26
33		6,69	34		5,63

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

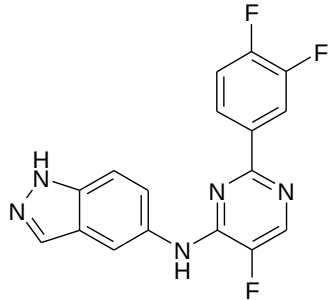
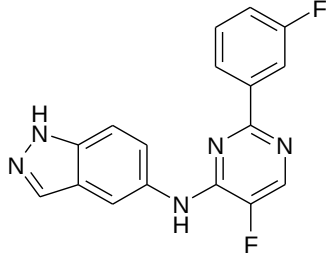
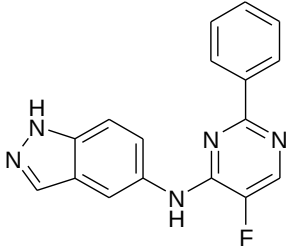
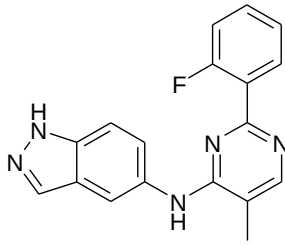
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
35		5,69	36		5,51
37		5,49	38		6,47

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

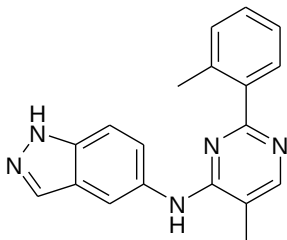
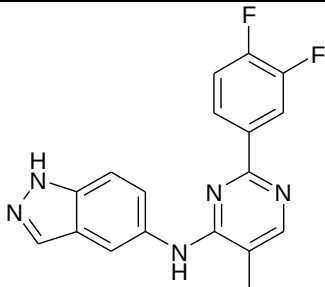
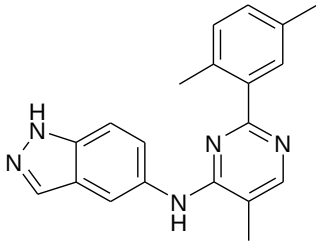
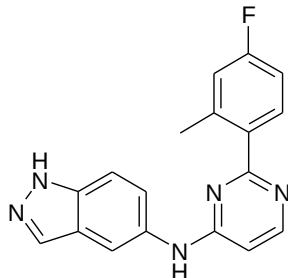
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
39		6,36	40		6,76
41		6,77	42		7,27

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

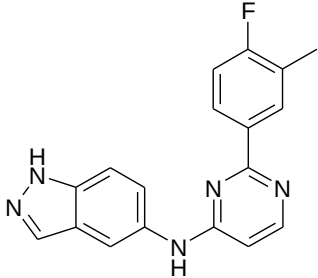
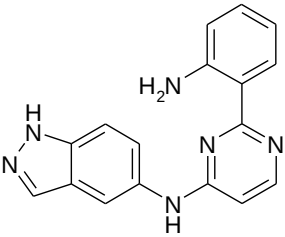
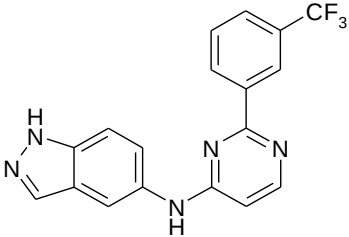
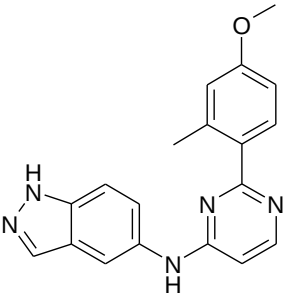
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
43		6,73	44		5,73
45		5,35	46		7,18

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

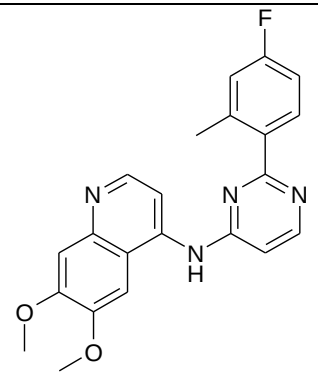
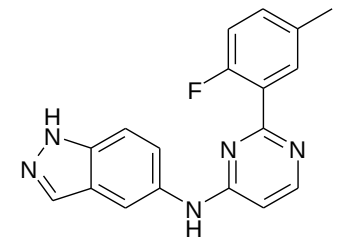
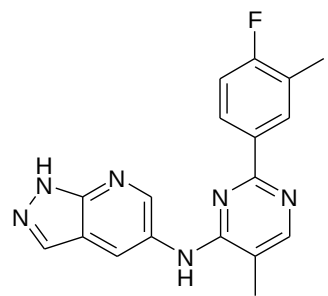
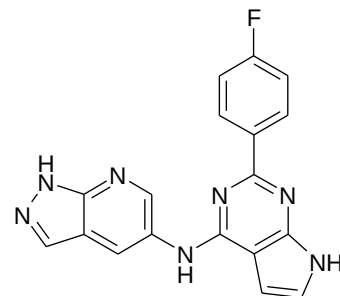
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
47		7,35	48		6,23
49		6,46	50		5,46

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

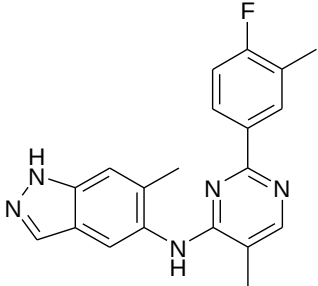
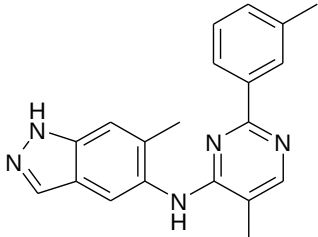
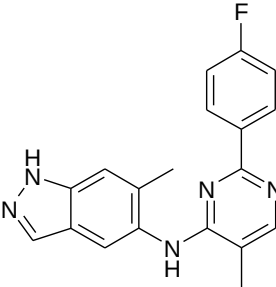
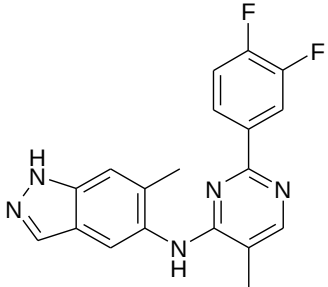
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
51		6,08	52		6,49
53		6,27	54		6,18

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (continua).

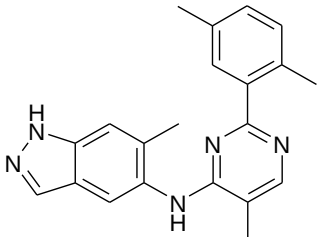
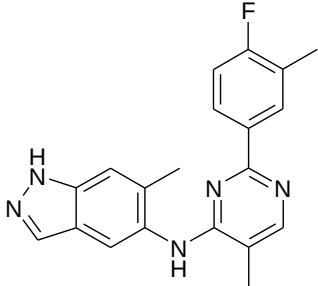
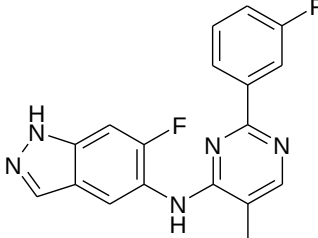
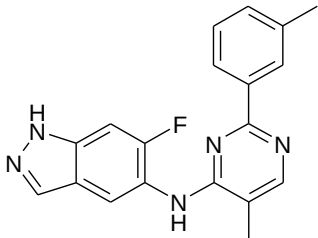
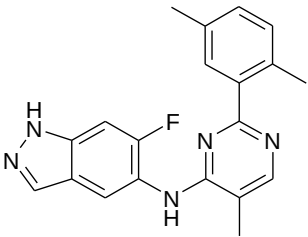
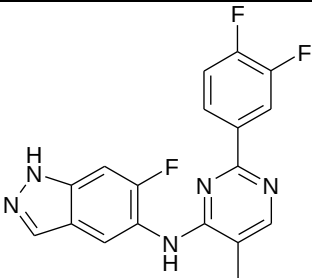
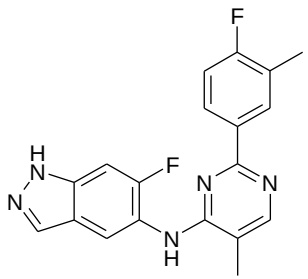
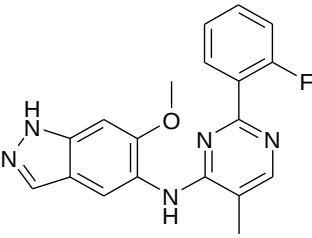
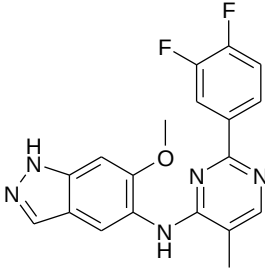
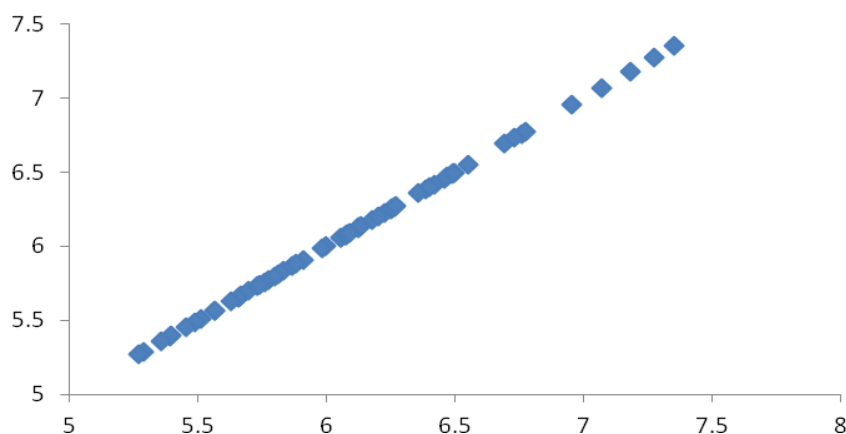
No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
55		6,20	56		6,55
57		5,74	58		6,09

Tabela 1 – Estrutura dos 63 inibidores TGF- β com os respectivos valores de atividade (-log da inibição da via de sinalização) (conclusão).

No.	Inibidor	Ativ.	No.	Inibidor	Ativ.
59		6,25	60		5,98
61		6,27	62		6,42
63		6,40			

A faixa dos valores de inibição dos 63 compostos abrange três ordens de magnitude (5,27 a 7,35) em unidade logarítmica, e os valores de atividade biológica apresentam uma distribuição regular dentro dessa faixa. O gráfico da Figura 6, corresponde à distribuição dos valores de inibição de fosforilação dos 63 compostos estudados.

Figura 7 – Gráfico de distribuição dos valores de inibição de fosforilação dos 63 compostos estudados



Fonte: Do Autor (2018)

4.2. Construção e Minimização dos compostos

As estruturas tridimensionais de cada um dos sessenta e três compostos apresentados na Tabela 1 foram construídas utilizando o programa PC Spartan Pro (PRO, 2000). Em seguida, as estruturas foram minimizadas utilizando o campo de força MMFF, e suas cargas atômicas parciais foram calculadas utilizando o método semi-empírico PM3 (STEWART, 1989).

4.3. Metodologia de QSAR-4D: Divisão do conjunto de treinamento e de teste

Os modelos de QSAR-4D foram construídos, utilizando um conjunto de 49 compostos formando o conjunto de treinamento, e validados externamente, usando um conjunto de 14 compostos, denominados de conjunto teste.

Foram avaliados 4 grupos testes, sendo que a escolha dos compostos do conjunto de treinamento e do conjunto teste foi feita de duas maneiras diferentes a fim de abranger a maior

heterogeneidade dos compostos. Para a seleção de 2 grupos teste foi utilizado o algoritmo de Kennard-Stone. O processo se inicia selecionando as duas amostras com maior distância Euclidiana entre si. Para cada uma das amostras restantes, calcula-se a distância mínima em relação às amostras selecionadas. Depois, são escolhidas as amostras que possuem a maior distância mínima até se alcançar o número desejado para calibração (KENNARD; STONE, 1969; SOUZA; MADARI e GUIMARÃES, 2012).

Os outros 2 grupos testes foram selecionados separando-se os 63 compostos, estruturalmente relacionados, em três subgrupos, de acordo com faixas de valores de atividade biológica e selecionando, aleatoriamente, 20% dos compostos de cada subgrupo para compor o conjunto teste e os remanescentes correspondem ao conjunto de treinamento. O grupo I compreende os compostos que apresentam valores de inibição de 5,27 a 5,98; o grupo II compreende valores de 6,00 a 6,95 e o grupo III compreende os compostos com inibição de 7,07 a 7,35. Usando essa abordagem, temos compostos no conjunto de teste que abrangem todos os valores de inibição (DA CUNHA *et al.*, 2005). Na Tabela 2 são apresentados os quatro grupos teste estudados.

Tabela 2 – Grupos teste estudados.

Grupo Teste	Compostos
1	7, 8, 11, 21, 22, 26, 31, 37, 39, 53, 55, 56, 57, 60
2	1, 12, 13, 14, 17, 21, 27, 28, 36, 39, 43, 44, 50, 53
3	3, 4, 5, 13, 17, 21, 22, 26, 39, 40, 42, 48, 52, 58
4	13, 14, 17, 18, 20, 21, 27, 28, 30, 37, 38, 39, 45, 53

Os grupos teste 1 e 2 foram selecionados pelo algoritmo de Kennard-Stone e os grupos 3 e 4 foram selecionados manualmente.

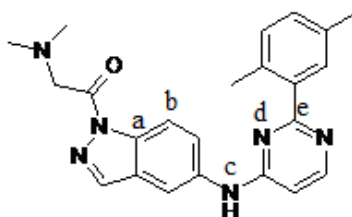
Fonte: Do Autor (2018)

4.4. Simulação de dinâmica molecular (SDM)

O processo de SDM foi executado usando o programa Molsim 3.0 (MOLSIM, 1997) com o campo de força MM2 (WEINER *et al.*, 1986), que inclui parametrizações descritas por Hopfinger e Pearlstein (HOPFINGER; PEARLSTEIN, 1984), incorporado ao programa 4D-QSAR. Como a intenção não é avaliar interações intermoleculares e/ou intramoleculares e sim, a diferença na orientação espacial durante a sobreposição das conformações dentro da caixa gradeada, acredita-se que não haverá perdas de informações importantes para gerar as equações matemáticas. De fato, haverá um erro energético constante para todos os compostos usando-se um campo de força com parâmetros antigos.

A temperatura para cada SDM foi de 300K, próxima à temperatura usada nos ensaios farmacológicos dos compostos em estudo, com um tempo total de simulação de 50ps, integradas em intervalos de 0,001ps. Uma conformação foi registrada a cada 0,01ps no arquivo de trajetória, totalizando 500 conformações para cada composto. Os cálculos de SDM foram executados utilizando solvente implícito, aplicando uma constante dielétrica, dependente da distância de 3r, no intuito de tentar modelar o efeito do solvente. Adicionalmente, foram fixados três átomos, bce (Figura 7), comuns aos compostos, com o objetivo de evitar uma grande mudança conformacional dos ligantes na ausência da estrutura da proteína.

Figura 8 – Compostos 1, as letras a, b, c, d, e representam os átomos fixados durante a dinâmica molecular e/ou os átomos que foram utilizados nos alinhamentos.



Fonte: Do autor (2018).

4.5. Definição do tamanho das células

As células da caixa virtual cúbica podem ser formadas por diferentes tamanhos de arestas. Quanto menor o tamanho da célula, mais refinados serão os modelos e mais descritores serão gerados, no entanto o tempo de simulação aumenta. Neste trabalho, utilizou-se o menor tamanho de célula, que é de 1,0 Å de aresta.

4.6. Definição dos elementos de interação farmacofórica

Os elementos de interação farmacofórica utilizados neste estudo foram: não polar (np), polar com densidade de carga positiva (p+), polar com densidade de carga negativa (p-), aceptor de ligação de hidrogênio (hba), doador de ligação de hidrogênio (hbd) e aromático (ar). (SILVA; RAMALHO e DA CUNHA, 2014).

4.7. Definição dos tipos de ocupação

Neste trabalho, foi utilizada a ocupação absoluta, pois não havia um composto de referência.

4.8. Definição do alinhamento

Foram testados nove alinhamentos. Para definição de cada um foram escolhidos três átomos comuns a todos os compostos. A Figura 7 apresentada anteriormente, mostra o composto 1 e as letras **a**, **b**, **c**, **d** e **e** representam os átomos escolhidos para os alinhamentos. A Tabela 3 apresenta todos os alinhamentos realizados.

Tabela 3 – Alinhamentos testados e os átomos selecionados

Alinhamento	1º átomo	2º átomo	3º átomo
1	e	c	a
2	c	a	e
3	a	e	c
4	c	b	e
5	e	c	b
6	b	e	c
7	c	b	d
8	d	c	b
9	b	d	c

Fonte: Do Autor (2018)

4.9. Construção dos modelos de QSAR-4D

4.9.1. Pré-tratamento dos descritores (GCOD)

A análise de QSAR-4D gera um grande número de variáveis (os descritores), em razão do grande número de IPEs e células da caixa utilizados. Além disso, existe a possibilidade de haver descritores altamente correlacionados, o que dificulta a interpretação e validação dos modelos.

Para reduzir o número de variáveis independentes (GCODs) geradas e melhorar a qualidade dos GCODs, foram utilizados dois critérios antes da etapa de geração dos modelos (equações). No primeiro, utilizou-se o método de PLS para selecionar os descritores com maior contribuição para explicar a variação da atividade biológica. E no segundo, os GCODs com variância igual a zero para todos os compostos foram eliminados.

4.9.2. Obtenção dos modelos de QSAR-4D

Os GCODs com um alto peso em cada banco de dados provenientes do pré-tratamento foram processados em conjunto com os valores de atividade biológica (variável dependente), utilizando a metodologia GFA, (ROGERS; HOPFINGER, 1994).

Os cálculos foram realizados no programa Wolf 6.2 (ROGERS, 1994) e iniciados com 100 modelos (equações) gerados, randomicamente, com uma taxa de probabilidade de mutação de 50%. O *crossover* foi variado entre 50000 e 90000 e o *smoothing factor*, que controla o número de descritores (GCODs) nos modelos de QSAR, foi variado de 0,1 a 0,7 no intuito de se obter equações com, no máximo, nove termos. Equações com mais de nove termos, possivelmente, provocarão um super ajuste (*overfitting*) nos dados, pois o número máximo de termos pode ser definido pelo número de compostos utilizados no conjunto de treinamento, dividido por 5 (KUBINYI, 2008).

Os modelos gerados foram ordenados segundo os valores do fator de desajuste (LOF) de Friedman e o melhor modelo resultante de cada alinhamento foi submetido aos métodos de validação.

4.9.3. Validações

Os parâmetros de validação calculados foram: LOF, r^2 , q^2 , R^2_{pred} , r^2_m , R^2_{Y-rand} e R^2_p . Os modelos foram obtidos a partir de um grupo de moléculas denominado grupo de treinamento e validados externamente por um grupo denominado teste.

4.9.4. Valores residuais e Outliers

Os valores residuais ou resíduos são calculados pela diferença entre a atividade experimental e atividade predita ($pAtiv_{EXP.} - pAtiv_{PRED.}$). A partir do desvio padrão desses valores, podem ser identificados compostos que não foram bem preditos pelo modelo e, portanto são considerados *Outliers*. Neste trabalho, eles foram definidos como aqueles cujos valores modulares dos resíduos são superiores ao dobro do desvio padrão desses resíduos. (HOPFINGER *et al.*, 1997).

4.9.5. Seleção da "Conformação Preferencial"

O objetivo, nesta etapa, é prever uma conformação de cada composto, utilizando as conformações de menor energia do perfil da amostragem conformacional (CEP). A metodologia de QSAR-4D assume que esta seria a conformação que maximiza a atividade do composto no

modelo selecionado. Um critério é utilizar todas as conformações de cada composto, dentro de um intervalo de energia (ΔE). Neste trabalho a conformação de cada composto foi selecionada dentro de uma faixa de 5 kcal.mol^{-1} . Posteriormente, dentre essas conformações, aquela com maior atividade conforme previsto pelo modelo foi determinada como a conformação representativa (PASQUALOTO *et al.*, 2004).

4.10. Ancoramento Molecular

A fim de compreender melhor a interação dos derivados aril pirimidina com a proteína TGF- β 1 realizou-se o estudo de ancoramento molecular.

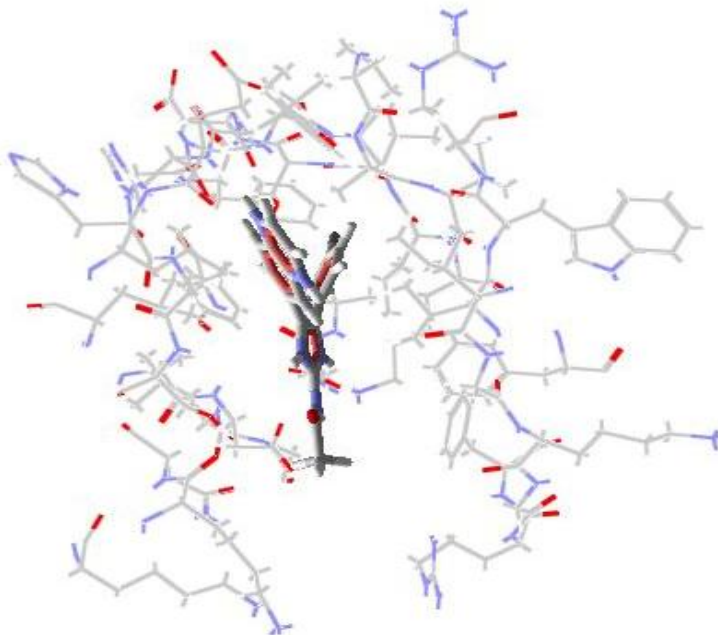
As estruturas 3D dos compostos em estudo foram preparadas conforme descrito anteriormente para o estudo de QSAR-4D. Como receptor, foi utilizada a estrutura tridimensional da proteína TGF- β 1, obtida por difração de raios-X, complexada com um inibidor 2-aminoimidazol, disponível no *Protein Data Bank* (PDB) (SUSSMAN *et al.*, 1998) sob o código 3FAA (BONAFOUX *et al.*, 2009) com resolução de $3,35\text{\AA}$, que foi escolhida pelo fato de apresentar um inibidor complexado com estrutura semelhante às aquelas em estudo. A proteína foi anteriormente preparada no programa *Accelrys Discovery Studio 2.1* (ACCELRYS SOFTWARE INC., 2007), onde os átomos de hidrogênio foram adicionados, a estrutura minimizada utilizando o campo de força DREIDING (MAYO; OLAFSON e GODDARD, 1990) e as cargas atômicas parciais calculadas através da mecânica molecular.

Esta enzima encontra-se no PDB na forma de pentâmero, entretanto, como o monômero é independente, o estudo foi desenvolvido apenas com a cadeia A dessa enzima (BONAFOUX *et al.*, 2009).

Os estudos de ancoragem molecular foram realizados utilizando o programa *Molegro Virtual Docker* (MVD). Este programa foi escolhido por produzir mais exatidão de ancoragem molecular superior a outros programas como *Glide*, *Surflex*, *FlexX*, *Gold* e *Autodock* (ARAÚJO *et al.*, 2011; THOMSEN; CHRISTENSEN, 2006). Os cálculos foram realizados considerando flexíveis os resíduos de aminoácidos próximos a $6\text{\AA}$ do inibidor complexado, o que simula melhor o processo biológico real, no qual, tanto o ligante quanto a proteína podem mudar a sua conformação. Foram considerados flexíveis os resíduos Gly353, Leu352, Asp351, Ala350, Leu

340, Asn338, Lys337, Lys335, Ser287, Gly286, His285, Glu284, His283, Tyr282, Asp281, Ser280, Val279, Leu278, Ala264, Ile263, Phe262, Gly261, Leu260, Tyr249, Ala246, Glu245, Phe234, Lys232, Val231, Ala230, Val229, Arg221, Trp220, Val219, Phe216, Arg215, Gly214, Lys213, Gly212, Ile211 e são apresentados na Figura 9 juntamente com o inibidor complexado.

Figura 9 – Aminoácidos considerados flexíveis para o estudo de ancoramento molecular juntamente com o inibidor complexado da proteína TGF- β 1 (PDB:3FAA).



Fonte: Do Autor (2018)

O primeiro passo é a detecção das cavidades, locais onde os ligantes podem ser ancorados. A cavidade utilizada neste trabalho possui um volume de $325,1\text{\AA}^3$, e foi escolhida por apresentar o inibidor complexado no local. A função de pontuação de energia empregada foi o MolDock Score [GRID] com resolução de $0,3\text{\AA}$.

Para cada complexo foi executada 1 corrida, retornando 50 poses, utilizando o algoritmo de busca MolDock optimizer. A região de ancoramento foi restrita a uma esfera de $9\text{\AA}$ de raio, centrada nas coordenadas $x = 58,21$; $y = 47,95$ e $z = 12,27$, que é a região de ligação do inibidor complexado. O tamanho da esfera foi escolhido a fim de incluir todos os resíduos importantes do sítio ativo. As melhores soluções dos complexos obtidos foram analisadas de acordo com a sua

conformação (comparada com o inibidor complexado) e com as interações intermoleculares (proteína-ligante), como ligação de hidrogênio, interações eletrostáticas e hidrofóbicas.

Para validação da metodologia utilizada, foi realizado o Re-docking, que consiste em retirar o ligante do complexo enzimático e realizar o docking, a fim de avaliar se a conformação foi reproduzida (LEACH, 2001). Esta validação é analisada em termos do desvio médio quadrático das posições atômicas (RMSD), entre a estrutura ligante obtida experimentalmente e a encontrada pelo programa. Valores de RMSD menores de 2,0 Å indicam que as poses encontradas foram corretamente preditas (LEACH, 2001).

5. RESULTADOS E DISCUSSÃO

5.1. Seleção do melhor alinhamento de QSAR-4D

Na metodologia de QSAR-4D foram analisados nove alinhamentos. Para cada alinhamento, foram gerados vários modelos, no entanto, foi analisado apenas o melhor modelo proveniente de cada alinhamento. Como forma de assegurar a qualidade dos modelos, foram analisados quatro conjuntos de teste.

Na Tabela 4, 5, 6 e 7, estão apresentados os parâmetros estatísticos dos modelos de QSAR-4D dos nove alinhamentos, para os conjuntos de teste 1, 2, 3 e 4 respectivamente.

Tabela 4 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 1

Alinhamento	R^2	q^2_{cv}	GCOD	R^2_{pred}	r_m^2 teste	R^2_{Y-rand}	R_p^2
1	0,707	0,510	9	0,495	0,297	0,206	0,596
2	0,738	0,069	9	0,614	0,191	0,205	0,627
3	0,522	0,224	9	0,486	0,216	0,195	0,413
4	0,778	0,564	9	0,666	0,593	0,207	0,667
5	0,752	0,454	9	0,672	0,386	0,154	0,670
6	0,679	0,081	9	0,539	0	0,200	0,571
7	0,642	0,028	9	0,639	0,523	0,179	0,545
8	0,638	0,001	9	0,694	0,433	0,131	0,569
9	0,687	0,508	9	0,302	0,174	0,199	0,579

Fonte: Do Autor (2018).

Tabela 5 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 2

Alinhamento	R^2	q^2_{cv}	GCOD	R^2_{pred}	r_m^2 teste	R^2_{Y-rand}	R_p^2
1	0,671	0,538	9	0,474	0,352	0,127	0,603
2	0,716	0,172	9	0,734	0,693	0,207	0,604
3	0,479	0,158	9	0,758	0,709	0,215	0,355
4	0,785	0,569	9	0,741	0,698	0,211	0,671
5	0,619	0,472	9	0,359	0	0,157	0,535
6	0,653	0,272	9	0,699	0,648	0,210	0,538
7	0,639	0,016	9	0,672	0,476	0,195	0,533
8	0,628	0,131	9	0,653	0,463	0,211	0,511
9	0,624	0,436	9	0,646	0,545	0,156	0,541

Fonte: Do Autor (2018).

Tabela 6 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 3

Alinhamento	R^2	q^2_{cv}	GCOD	R^2_{pred}	r_m^2 teste	R^2_{Y-rand}	R_p^2
1	0,729	0,544	9	0,401	0	0,234	0,601
2	0,837	0,668	9	0,322	0	0,198	0,731
3	0,778	0,661	9	0,001	0	0,226	0,655
4	0,822	0,589	9	0,409	0	0,192	0,719
5	0,716	0,285	9	0,035	0	0,096	0,464
6	0,752	0,595	9	0,329	0	0,177	0,658
7	0,846	0,417	9	0,047	0	0,197	0,741
8	0,807	0,009	9	0,148	0	0,192	0,704
9	0,718	0,570	9	0,148	0	0,139	0,645

Fonte: Do Autor (2018).

Tabela 7 – Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os nove alinhamentos com o conjunto de teste 4

Alinhamento	R^2	q^2_{CV}	GCOD	R^2_{pred}	r_m^2 teste	R^2_{Y-rand}	R_p^2
1	0,732	0,546	9	0,280	0	0,192	0,629
2	0,737	0,131	9	0,473	0,243	0,194	0,632
3	0,529	0,218	9	0,258	0	0,206	0,413
4	0,792	0,584	9	0,648	0,547	0,207	0,681
5	0,648	0,145	9	0,453	0	0,231	0,519
6	0,649	0,359	9	0,616	0,204	0,202	0,538
7	0,627	0,049	9	0,542	0,261	0,186	0,526
8	0,660	0,057	9	0,390	0	0,166	0,571
9	0,615	0,422	9	0,559	0,464	0,209	0,499

Fonte: Do Autor (2018).

Analisando as Tabelas 4, 5, 6 e 7, percebe-se que os resultados variaram muito, dependendo do alinhamento e do conjunto de teste analisados. Nota-se que os grupos testes 1 e 2 apresentaram em média, valores satisfatórios de R^2_{pred} (maiores que 0,5) para mais de 60% dos alinhamentos testados, o que mostra que esses conjuntos representam bem o conjunto de treinamento. No entanto os outros parâmetros não foram adequados.

Já para o grupo teste 3, acontece ao contrário. Ele apresenta valores aceitáveis nas validações internas (R^2 , q^2_{CV}), porém nas validações externas (R^2_{pred} , r_m^2 teste) isso não é observado. O conjunto de teste 4 divergiu muito em função de cada alinhamento.

Nas Tabelas acima nota-se também, que o alinhamento 4 apresentou valores satisfatórios para todos os parâmetros estatísticos, quando foi testado com os conjuntos de teste 1, 2 e 4. Para o conjunto de teste 3, somente a validação externa ($R^2_{pred}= 0,409$) não foi satisfatória.

Para escolha do melhor alinhamento foram analisados todos os parâmetros estatísticos. A partir desta análise verificou-se que o alinhamento 4 foi o que obteve melhores resultados com todos os grupos testes estudados, portanto ele foi escolhido para representar o conjunto de dados. A fim de escolher o melhor grupo teste foram avaliados os valores de validação cruzada (q^2), pois ele mostra a significância estatística de cada componente do modelo. Desse modo, como o valor

de q^2 apresentado para o grupo teste 4 foi maior ($q^2 = 0,584$), o modelo gerado com o conjunto de teste 4 para o alinhamento 4 foi selecionado para representar o conjunto de dados e prosseguir com a análise dos descritores.

5.2. Análise do Alinhamento 4

Os dez modelos provenientes do alinhamento 4 apresentaram boa capacidade preditiva e coeficientes de determinação R^2 adequados (Tabela 8). Portanto, a escolha do melhor modelo foi feita analisando os valores de LOF, que possuem relação com o erro dos mínimos quadrados. Assim quanto menor o valor de LOF, menor o erro. Dessa forma, foi escolhido o modelo 1 do alinhamento 4 para ser analisado. O modelo selecionado está representado na equação 11:

$$\text{Ativ} = 4,93 + 0,19 (0,0,2,np) + 0,84 (0,5,0,np) - 2,05 (1,5,0,np) + 1,71 (2,-2,2,np) \\ + 1,13 (-1,1,-1,np) + 0,72 (0,0,-1, np) + 0,48 (1,4,2,p-) - 0,55 (0,4,0,p-) + 1,53 (1,5,1,np) \quad (11)$$

Tabela 8 – Resumo dos resultados dos modelos de QSAR-4D provenientes do alinhamento 4.

Modelo	LSE	LOF	R^2	q^2
1	0,043	0,070	0,792	0,584
2	0,051	0,089	0,782	0,574
3	0,047	0,089	0,799	0,527
4	0,048	0,091	0,793	0,574
5	0,052	0,091	0,776	0,512
6	0,048	0,092	0,792	0,581
7	0,048	0,092	0,792	0,590
8	0,052	0,092	0,774	0,526
9	0,048	0,092	0,791	0,523
10	0,053	0,093	0,771	0,589

Fonte: Do Autor (2018).

A fim de verificar se o Modelo 1 apresenta o mesmo tipo de informação estrutural-atividade, foi calculada a matriz de correlação cruzada entre os valores de ocupação das células (GCODs) (Tabela 8). São considerados correlacionados aqueles descritores com $r > 0,7$ (DA CUNHA *et al.*, 2005).

Tabela 9 – Matriz de correlação cruzada dos GCODs do melhor modelo do alinhamento 4.

GCOD	0,0,2(np)	0,5,0(np)	1,5,0(np)	2,-2,2(np)	-1,1,-1(np)	0,0,-1(np)	1,4,2(p-)	0,4,0(p-)	1,5,1(np)
0,0,2(np)	1,00								
0,5,0(np)	0,02	1,00							
1,5,0(np)	0,05	-0,48	1,00						
2,-2,2(np)	0,07	0,08	0,01	1,00					
-1,1,-1(np)	0,38	-0,01	-0,23	0,05	1,00				
0,0,-1(np)	0,11	0,05	0,23	0,51	-0,10	1,00			
1,4,2(p-)	-0,11	0,26	-0,04	-0,12	-0,05	-0,12	1,00		
0,4,0(p-)	0,05	0,21	-0,47	-0,05	0,39	-0,07	0,01	1,00	
1,5,1(np)	-0,05	-0,12	0,30	-0,02	-0,12	0,01	-0,05	-0,28	1,00

Fonte: Do Autor (2018).

A partir da Tabela 9, pode-se verificar que nenhum par de descritor do modelo selecionado está correlacionado, pois apresentam $r < 0,7$. O que indica pouca tendência de um descritor estar influenciando o outro, ou seja, cada descritor contribui de maneira única na construção do modelo de QSAR-4D.

5.2.1. Análise dos descritores

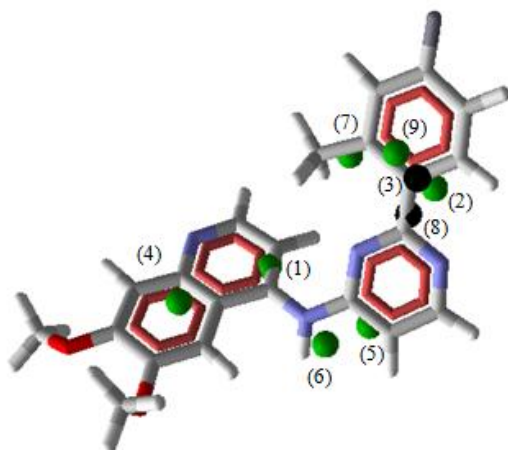
O melhor modelo do alinhamento 4 apresentado na equação 11, gerou 9 descritores (GCODs), sendo que sete apresentaram coeficiente positivo, (0,0,2,np), (0,5,0,np), (2,-2,2,np), (-1,1,-1,np), (0,0,-1,np), (1,4,2,p-), (1,5,1,np) e correspondem a interações favoráveis entre os substituintes dos compostos e os resíduos de aminoácidos da enzima TGF- β . Assim, substituintes com as características selecionadas (IPEs) nessas posições aumentam a potência dos compostos. Os GCODs (1,5,0,np) e (0,4,3,p-) apresentaram coeficientes negativos, correspondendo às interações desfavoráveis entre os substituintes dos compostos e os resíduos de aminoácidos da enzima.

Neste trabalho, os descritores são representados por esferas. As esferas verdes representam os GCODs com coeficientes positivos, e correspondem às regiões na molécula onde a presença de determinados substituintes favorece para aumentar sua potência, e as esferas pretas representam os GCODs com coeficientes negativos, correspondendo às regiões na molécula, onde a presença de certo substituinte irá diminuir a potência dos compostos.

Os GCODs referentes a cada molécula representam a frequência de ocupação das células, considerando todas as conformações de cada um dos compostos obtidos na SDM. Assim, a conformação representada graficamente para cada composto (conformação representativa) não possui, necessariamente, todos os GCODs descritos pelo modelo para a referida molécula.

Na Figura 8, ilustra-se o composto mais ativo (molécula 47) e os nove descritores obtidos no Modelo1 do alinhamento 4. Os descritores são numerados de acordo com a sequência da equação 11.

Figura 10 – Modelos de QSAR-4D representativo do alinhamento 4, mostrando o composto **47** juntamente com descritores

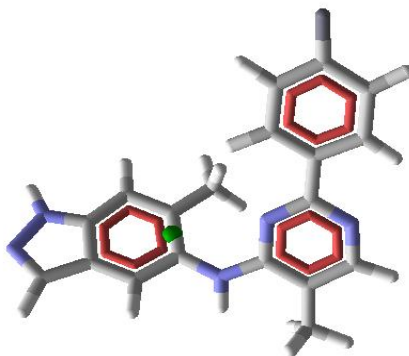


Nota: A ocupação dos GCODs representados por esferas verdes que contribui para o aumento da potência dos compostos e pelas esferas pretas, para a diminuição da potência.

Fonte: Do Autor (2018).

O GCOD-1 (Figura 10) possui coeficiente positivo e é do tipo não polar. Ele apresenta maior frequência de ocupação para os compostos que possuem um anel indazol ligado ao nitrogênio central, como o composto **53** (Figura 11), que também apresenta um grupo metil como substituinte, o que torna esta região mais apolar. Compostos que possuem um grupo piridina ligado ao nitrogênio não possuem frequência de ocupação nesta região, como o composto **2**. Assim, pode-se concluir que grupos apolares nessa posição são favoráveis para o aumento da potencia desses compostos.

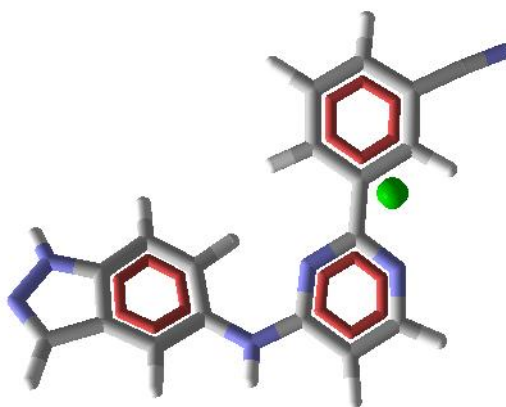
Figura 11 – Representação da molécula **53** e do GCOD-1



Fonte: Do Autor (2018)

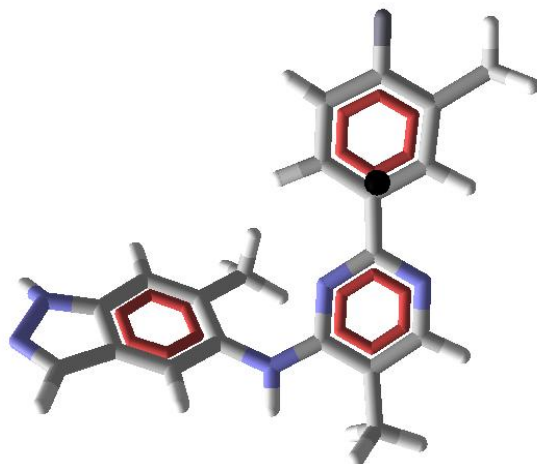
O GCOD-2 (Figura 10) também possui um coeficiente positivo, e, desse modo, também influencia no aumento da potência dos compostos. Esse descritor mostra que a presença de substituintes apolares nessa posição contribui para aumentar a potência dos compostos. Essa célula apresentou maior frequência de ocupação para o composto **29** (Figura 12) que, como outros compostos, apresenta um substituinte fenil nesta região, e nenhuma ocupação para os compostos **2**, **3**, **4**, **5**, entre outros. Ele se encontra em uma parte comum dos compostos, no entanto o composto **2** apresenta um anel piridina como substituinte. O que confirma que a presença de grupos apolares é favorável.

Figura 12 – Representação da molécula **29** e do GCOD-2



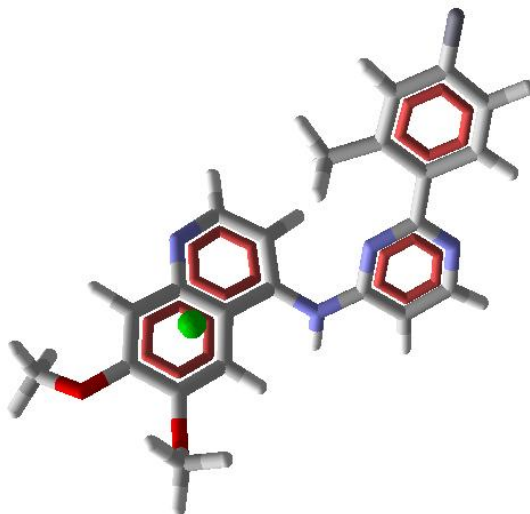
Fonte: Do Autor (2018)

O GCOD-3 (Figura 10) possui muita influência na atividade dos compostos, uma vez que ele apresenta o coeficiente mais negativo. Portanto, ele sugere que grupos nessas posições são desfavoráveis para a atividade dos compostos. Ele representa uma interação não polar, e as maiores frequências de ocupação encontradas para esse descritor são para os compostos **51** (Figura 13) e **61**, que apresentam grupos metilas ligados ao anel benzênico. Compostos como **3**, **4** e **5** os quais apresentam substituintes polares, não possuem frequência de ocupação nesta região. Isso indica que grupos apolares ligados ao anel benzênico nesta região são desfavoráveis para a potencia dos compostos.

Figura 13 – Representação da molécula **51** e do GCOD-3

Fonte: Do Autor (2018)

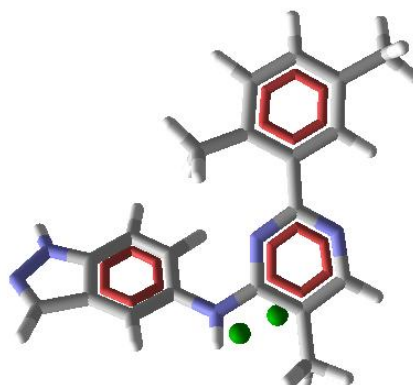
O maior coeficiente positivo é do GCOD-4 (Figura 10), que apresenta um IPE do tipo não polar. O composto que apresenta maior frequência de ocupação para esse descritor é o **47** (Figura 14), que apresenta um substituinte quinolina ligado ao nitrogênio central. Isso corrobora com os dados experimentais, uma vez que o composto **47** também é o de maior atividade. Portanto, substituintes apolares nesta região são favoráveis para o aumento da atividade.

Figura 14 – Representação da molécula **47** e do GCOD-4

Fonte: Do Autor (2018)

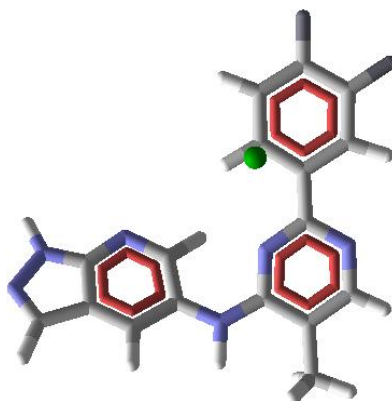
Os descritores 5 e 6 (Figura 10), nos trazem o mesmo tipo de informação, visto que os dois estão próximos, possuem coeficientes positivos e também representam o mesmo tipo de interação, não polar. Eles apresentam os maiores valores de ocupação para os compostos que possuem um grupo metila ligado ao anel pirimidina, como os compostos **39**, **41** (Figura 15), e **54**. Compostos que possuem grupos polares nesta região como **12** e **24** não apresentaram frequência de ocupação nesta região. Sendo assim, a presença de grupos apolares nesta região é favorável para a atividade dos compostos.

Figura 15 – Representação da molécula **41** e dos GCOD-5 e 6



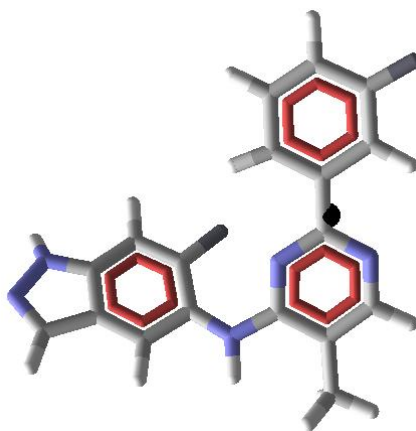
Fonte: Do Autor (2018)

O GCOD-7 (Figura 10), que apresenta um IPE do tipo polar negativo, possui um coeficiente positivo indicando o aumento na potência dos compostos que possuem elevada frequência de ocupação para esse descritor. Os compostos que possuem maior ocupação para esse descritor são os **15** (Figura 16) e **49**. Como pode ser observado na Figura 16, ele está localizado perto do anel benzênico, e possui maior ocupação para os compostos que possuem o átomo de Flúor como substituinte. No entanto, a frequência é zero para compostos que possuem o Flúor somente nas posições *orto* e *meta*, indicando que para o aumento da atividade, é importante a presença do Flúor como substituinte na posição *para*.

Figura 16 – Representação da molécula **15** e do GCOD-7

Fonte: Do Autor (2018)

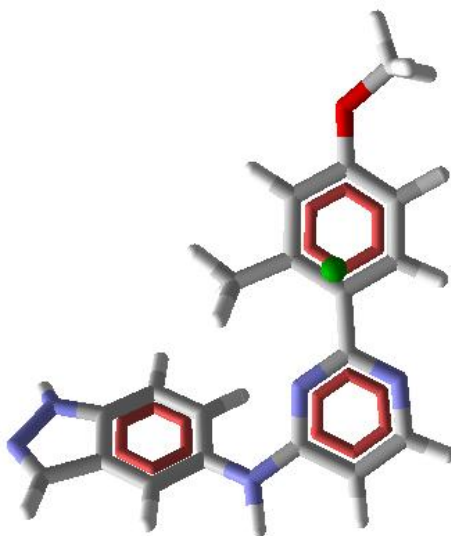
O outro descritor com coeficiente negativo é o GCOD-8 (Figura 10). Ele representa elementos de interação farmacofórica polar com carga negativa e possui maior frequência de ocupação para os compostos que possuem um átomo de flúor ligado ao anel indazol, como no composto **57** (Figura 17). Desse modo, átomos ou grupos polares com carga negativa nesta região da molécula são desfavoráveis para a atividade.

Figura 17 – Representação da molécula **57** e do GCOD-8

Fonte: Do Autor (2018)

O GCOD-9 (Figura 10) possui o segundo coeficiente mais positivo, que mostra que ele influencia bastante no aumento da atividade dos compostos. Ele é do tipo não polar e apresentou altas frequências de ocupação para os compostos que possuem grupos metila ligados ao anel benzênico, como no composto **46** (Figura 18). Entretanto a posição em que este grupo está ligado é importante, pois as frequências de ocupação se dão para os compostos que possuem este substituinte na posição *orto*. Logo, grupos não polares nessa região da molécula pode aumentar a atividade dos compostos.

Figura 18 – Representação da molécula **46** e do GCOD-9



Fonte: Do Autor (2018)

5.2.2. Análise dos valores residuais e Outliers

A Eq. 11 foi utilizada para calcular os valores de inibição preditos para todos os compostos estudados. Na Tabela 10 estão apresentados os valores de inibição experimental, predito e o valor residual.

Tabela 10 – Valores experimentais e preditos de inibição juntamente com o valor residual

Comp	Inib exp	Inib pred	Res	Comp	Inib exp	Inib pred	Res	Comp	Inib exp	Inib pred	Res	Comp	Inib exp	Inib pred	Res
1	6,25	6,30	-0,05	17	5,66	5,98	-0,32	33	6,69	6,33	0,36	49	6,46	6,69	-0,23
2	6,50	6,11	0,39	18	5,87	5,99	-0,12	34	5,63	5,85	-0,21	50	5,46	5,56	-0,10
3	5,81	5,78	0,03	19	5,39	6,90	-1,51	35	5,70	5,57	0,12	51	6,08	6,35	-0,27
4	6,08	5,98	0,10	20	5,76	5,99	-0,24	36	5,51	5,31	0,20	52	6,49	6,74	-0,25
5	6,13	5,71	0,41	21	5,91	6,04	-0,13	37	5,49	5,33	0,16	53	6,27	6,40	-0,13
6	5,77	5,86	-0,09	22	6,95	6,32	0,63	38	6,47	6,13	0,34	54	6,18	5,97	0,21
7	6,06	6,26	-0,20	23	5,88	5,94	-0,06	39	6,36	6,47	-0,11	55	6,20	6,36	-0,16
8	6,13	5,90	0,23	24	5,40	5,38	0,02	40	6,76	6,30	0,46	56	6,55	6,68	-0,13
9	5,84	5,98	-0,14	25	6,49	6,22	0,26	41	6,77	6,39	0,38	57	5,74	5,73	0,01
10	5,29	5,32	-0,03	26	6,00	5,58	0,42	42	7,27	6,31	0,96	58	6,09	6,03	0,06
11	5,27	5,32	-0,05	27	5,66	5,94	-0,28	43	6,73	6,55	0,19	59	6,25	6,25	0,00
12	5,79	5,78	0,015	28	5,67	6,16	-0,49	44	5,73	5,89	-0,16	60	5,98	5,82	0,16
13	5,57	5,90	-0,34	29	7,07	6,86	0,21	45	5,36	5,48	-0,12	61	6,27	6,28	-0,01
14	6,14	6,32	-0,19	30	6,07	5,92	0,15	46	7,18	7,19	-0,01	62	6,42	6,31	0,11
15	6,06	6,31	-0,25	31	5,57	5,99	-0,42	47	7,35	7,31	0,04	63	6,40	6,33	0,07
16	6,39	6,06	0,33	32	6,26	6,17	0,09	48	6,23	6,16	0,07				

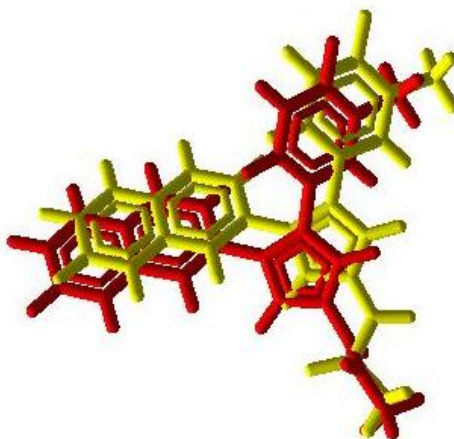
Neste trabalho, foram considerados como *outliers* os compostos com valor residual duas vezes superior ao valor de desvio padrão dos resíduos. O valor do desvio padrão dos resíduos foi 0,325, então, os compostos com valores residuais maiores que 0,651 são considerados *outliers*.

Ao analisar a Tabela 10, percebe-se que dois compostos (**19** e **42**) apresentam resíduos superiores a 0,651 e, portanto podem ser considerados *outliers*. Esse fato pode ser explicado através dos valores de atividade desses compostos. O composto **19** apresenta baixa atividade (5,39), além de possuir um substituinte incomum aos outros compostos, já o composto **42**, apesar de apresentar uma estrutura bastante comum com os demais, possui atividade alta de 7,27.

5.3. Análise do ancoramento molecular

Para validar o protocolo de ancoramento molecular foi realizada a simulação de *redocking*, na qual obteve um desvio médio quadrático (RMSD) de 1,22 Å (Figura 19). O que indica que o protocolo utilizado foi capaz de encontrar a posição do ligante, sendo, portanto utilizado como padrão para o estudo de *docking*. Segundo Leach e colaboradores (2001), valores de RMSD menores que 2 Å indicam que as poses foram corretamente preditas. Na Tabela 11 estão apresentados os valores de energia obtidos no *redocking*.

Figura 19 – Sobreposição entre a estrutura de raios-x cristalizada do receptor (amarelo) e a conformação obtida no estudo de docking (vermelho)



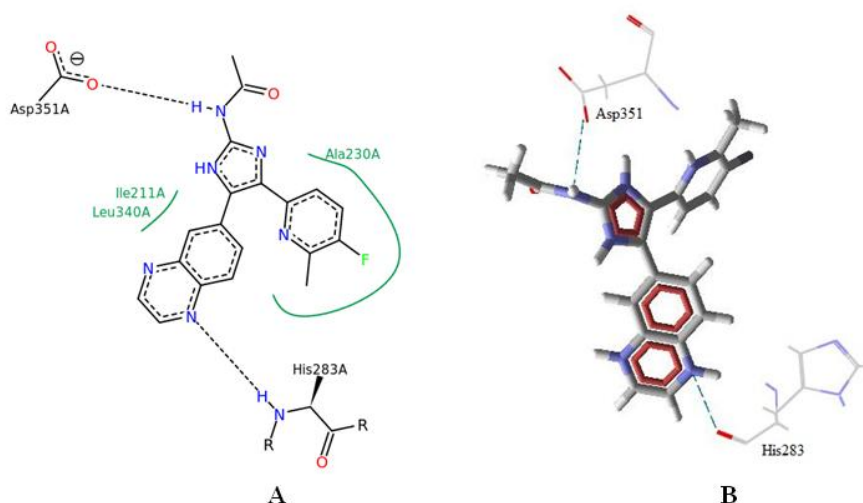
Fonte: Do Autor (2018)

Tabela 11 – Energias obtidas a partir do *redocking* (kcal.mol⁻¹) e valor do desvio médio quadrático (RMSD)

MolDock Score	Energia Estérica	Energia Intermolecular	Energia Eletrostática	Energia Total de Ligação de Hidrogênio	RMSD (Å)
-138,58	-141,51	-142,89	2,04	-0,79	1,22

Analisando os resultados obtidos no *redocking* pode-se dizer que o composto cristalizado apresenta boas energias de interação com a proteína. Também foi possível confirmar as interações de ligação de hidrogênio que são apresentadas no PDB, indicando mais uma vez que o protocolo de *docking* utilizado foi capaz de encontrar a posição do ligante (Figura 20).

Figura 20 – Ligação de hidrogênio do ligante cristalizado (2-aminoimidazol) na proteína de código PDB 3FAA. (A) interações apresentadas no PDB; (B) interações obtidas no *redocking*.



Fonte: Do Autor (2018)

Na Tabela 12 estão apresentados os resultados obtidos no estudo de *docking* para os 63 compostos estudados.

Tabela 12 – Energias obtidas a partir do ancoramento molecular para os derivados aril pirimidinas (kcal mol⁻¹) (continua).

Composto	MolDock Score	Energia estérica	Energia Intermolecular	Energia Eletrostática	Energia Total de Ligação de Hidrogênio	Resíduos envolvidos nas Ligações de Hidrogênio
1	-137,45	-149,66	-148,40	0,16	-3,72	Lys232
2	-127,69	-149,45	-151,29	-0,31	-6,09	Lys232/ His283
3	-106,03	-128,78	-124,76	-0,68	-0,71	His283
4	-103,95	-124,04	-121,83	0,50	-2,84	His283
5	-127,52	-136,24	-139,32	1,74	-6,69	Lys232/ His283
6	-135,87	-150,20	-148,99	0,03	-4,34	Lys232/ Aps351/ His283
7	-141,98	-153,82	-154,82	0,55	-4,32	Aps351/His283
8	-137,12	-152,82	-149,52	0,21	-2,54	Lys232/ Aps351/ His283
9	-137,02	-150,24	-148,98	0,57	-4,18	Lys232/ Aps351/ His283
10	-125,81	-140,06	-137,09	-1,57	-3,28	His283/Ser287
11	-123,78	-139,04	-134,82	-1,38	-1,96	His283/Ser287
12	-133,33	-142,31	-143,10	0,80	-4,28	Lys232/ His283
13	-119,76	-130,90	-131,11	0,54	-4,08	Lys232/ His283
14	-109,559	-129,73	-126,58	-1,68	-1,36	His283
15	-119,27	-134,61	-133,88	-0,60	-2,57	Lys232/ His283
16	-116,23	-136,16	-134,59	-0,35	-2,51	Lys232/ His283
17	-104,33	-127,44	-124,07	-0,58	-1,41	His283
18	-111,62	-129,93	-126,96	0,13	-1,56	His283
19	-116,98	-134,23	-132,46	-1,03	-2,72	His283

Tabela 12 – Energias obtidas a partir do ancoramento molecular para os derivados aril pirimidinas (kcal mol⁻¹) (continua).

Composto	MolDock Score	Energia estérica	Energia Intermolecular	Energia Eletrostática	Energia Total de Ligação de Hidrogênio	Resíduos envolvidos nas Ligações de Hidrogênio
20	-106,59	-124,86	-124,49	-0,76	-3,52	Lys232/ His283
21	-112,65	-131,55	-129,75	-0,35	-2,17	Lys232/His283
22	-109,13	-130,25	-129,43	0,20	-3,31	Lys232/His283
23	-126,66	-137,86	-136,41	0,28	-2,61	Lys232/ His283
24	-121,09	-134,22	-133,72	0,41	-2,72	Lys232/His283
25	-128,71	-141,98	-142,49	-0,03	-4,02	Lys232/ Aps351/ His283
26	-124,63	-142,29	-140,51	-0,15	-2,49	Lys232/His283
27	-109,62	-127,95	-123,84	-0,03	-1,21	His283
28	-116,47	-133,48	-130,83	-1,05	-3,15	Tyr282/His283
29	-130,09	-145,32	-147,24	0,74	-5,91	Lys232/His283
30	-122,12	-131,78	-130,47	-0,34	-2,70	Lys232/ His283
31	-104,49	-112,83	-116,39	0,52	-8,70	Ile211/Ala230/ Lys232/ Leu278/ Ser280
32	-100,75	-113,08	-112,04	-0,63	-2,59	Ile211/Lys232
33	-110,51	-129,59	-126,53	-0,57	-1,21	His283
34	-110,07	-124,82	-123,69	-1,41	-2,72	Lys232/ His283
35	-117,85	-133,13	-131,95	0,77	-2,90	Lys232/ His283
36	-116,95	-137,26	-135,17	-0,42	-2,40	Lys232/ His283
37	-97,55	-120,76	-117,23	-0,28	-1,35	His283
38	-104,79	-123,13	-121,88	-0,86	-2,80	Lys232/His283
39	-112,46	-126,15	-124,50	0,47	-2,71	Lys232/His283

Tabela 12 – Energias obtidas a partir do ancoramento molecular para os derivados aril pirimidinas (kcal mol⁻¹) (continua).

Composto	MolDock Score	Energia estérica	Energia Intermolecular	Energia Eletrostática	Energia Total de Ligação de Hidrogênio	Resíduos envolvidos nas Ligações de Hidrogênio
40	-120,39	-134,54	-133,48	-0,29	-2,63	Lys232/His283
41	-124,14	-134,49	-133,06	-0,03	-2,72	Lys232/His283
42	-129,17	-138,16	-137,23	-0,37	-2,55	Lys232/His283
43	-115,81	-130,37	-129,26	0,53	-2,69	Lys232/His283
44	-113,80	-124,45	-125,07	0,69	-6,42	Lys232/Glu245/ Tyr249/His283
45	-139,59	-160,88	-159,90	0,39	-3,06	Lys232/His283
46	-118,59	-134,34	-130,14	0,09	-0,98	Lys232/His283
47	-148,55	-168,34	-166,69	0,51	-2,60	Ser280/His283
48	-111,94	-132,97	-130,18	0,00	-1,43	Lys232/His283
49	-112,24	-133,58	-129,91	-1,66	-1,31	His283
50	-115,34	-128,38	-127,84	-0,54	-3,59	Lys232/Aps351/ His283
51	-124,60	-136,13	-137,70	-0,25	-5,44	Lys232/His283
52	-119,28	-128,96	-130,39	-0,18	-5,76	Lys232/His283
53	-123,78	-137,47	-134,03	-1,17	-1,51	His283
54	-132,26	-143,87	-142,50	-0,16	-2,59	Lys232/His283
55	-141,97	-148,79	-146,96	-1,45	-2,72	Lys232/His283
56	-130,82	-142,83	-141,44	0,39	-2,66	Lys232/His283
57	-128,73	-137,37	-137,51	-0,20	-2,81	Lys232/His283

Tabela 12 – Energias obtidas a partir do ancoramento molecular para os derivados aril pirimidinas (kcal mol^{-1}) (conclusão).

Composto	MolDock Score	Energia estérica	Energia Intermolecular	Energia Eletrostática	Energia Total de Ligação de Hidrogênio	Resíduos envolvidos nas Ligações de Hidrogênio
58	-127,29	-137,46	-136,20	0,05	-2,78	Lys232/His283
59	-137,78	-144,93	-143,52	0,08	-2,68	Lys232/His283
60	-131,69	-143,95	-143,16	0,07	-2,82	Lys232/His283
61	-129,07	-141,89	-140,71	0,03	-2,83	Lys232/His283
62	-129,01	-140,27	-138,37	0,57	-2,52	Lys232/His283
63	-131,53	-140,91	-139,63	-1,03	-2,60	Lys232/His283

Para selecionar a pose de cada composto foi utilizada como padrão a molécula cristalizada na proteína TGF- β 1, avaliou-se no entanto, a sobreposição de um dos anéis, uma vez que os compostos não eram análogos não sendo possível calcular o RMSD.

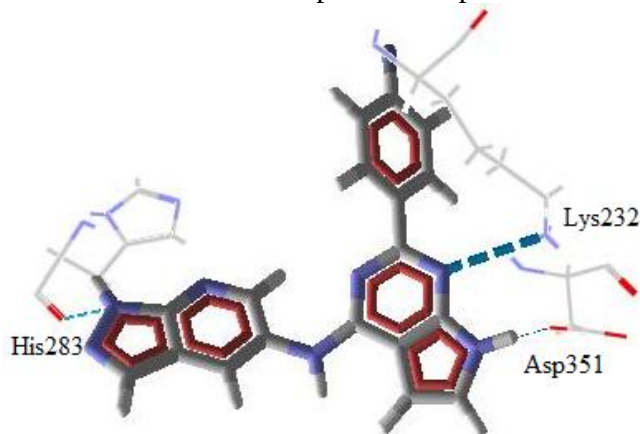
Ao analisar a tabela acima percebe-se que os compostos apresentaram boas energias de interação com a proteína TGF- β 1. Verificou-se também que os valores de energia total (MolDock Score), energia estérica e intermolecular foram melhores para o composto **47** (maior atividade biológica) corroborando com os resultados experimentais e com os resultados obtidos no QSAR-4D. No resultado de QSAR-4D o descritor mais influente no modelo obtido (2,-2,2,np), apresentou maior frequência de ocupação para esse composto. Percebe-se também que as ligações de hidrogênio são importantes, mas não determinantes para a atividade desses compostos, e que outras interações existentes devem ser consideradas como, interação estérica, intermolecular e eletrostática, uma vez que os compostos de maior atividade nem sempre apresentaram valores de ligação de hidrogênio mais estáveis.

A partir dos resultados de *docking* pode-se verificar que as interações existentes entre o ligante cristalizado e a proteína TGF- β 1 (Figura 20) também estiveram presentes nos compostos estudados. Essas interações são: ligação de hidrogênio com os resíduos His283 e Asp351, e

interações hidrofóbicas com os resíduos Leu340 e Ile211. Além desses, outros resíduos também se mostraram importantes, como Lys232, que apresentou ligação de hidrogênio com vários compostos e Val219 que apresentou interações hidrofóbicas com grupos apolares que foram considerados favoráveis para a atividade no estudo de QSAR.

As ligações de hidrogênio estão presentes em regiões comuns a quase todos os compostos. Na Figura 21, apresentamos o composto **50** e os resíduos de aminoácidos que fazem ligação de hidrogênio com ele. O resíduo His283 interage com nitrogênio do grupo indazol, o resíduo Asp351 interage com o nitrogênio do grupo pirrol e o resíduo Lys232 interage com o nitrogênio do grupo pirimidina.

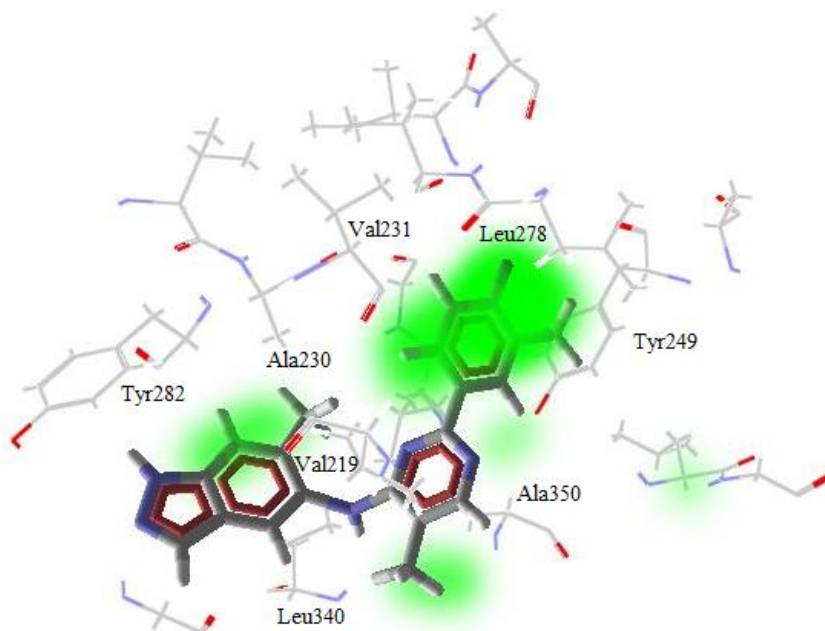
Figura 21 – Pose do ancoramento molecular para o composto **50** e suas ligações de hidrogênio



Fonte: Do Autor (2018)

Como foi visto nos resultados de QSAR, alguns grupos não polares foram considerados favoráveis para o aumento da atividade dos compostos. Isto pode ser explicado devido à presença de resíduos de aminoácidos apolares como Val219, Ala230, Leu278, Leu340, Ala350, dentre outros, perto dos grupos substituintes, que podem interagir por interação hidrofóbica. Na Figura 22 está apresentado o mapa de energia estérica para a pose do composto **56**. Nele pode-se perceber que as regiões hidrofóbicas favoráveis (representadas em verde) são exatamente aquelas relacionadas com os grupos que foram considerados positivos para a atividade dos compostos no QSAR, como os descritores **1** (0,0,2,np), **5** (-1,1,-1,np), **6** (0,0,-1,np) e **9** (1,5,1,np). Nesta figura são apresentados também, os resíduos de aminoácidos que fazem interações de van der Waals com os respectivos grupos metilas presentes no composto **56**.

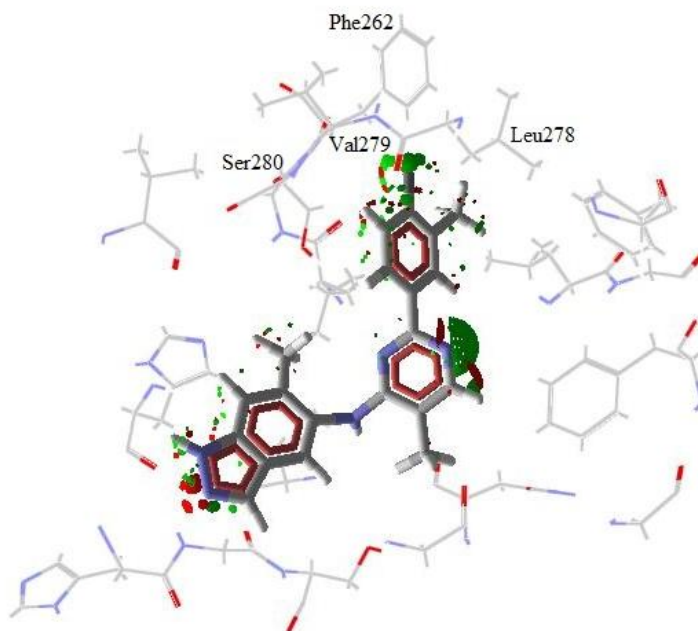
Figura 22 – Mapa de energia estérica para a pose do composto 56.



Fonte: Do Autor (2018)

Além dos grupos hidrofóbicos, substituintes polares negativos (como o Flúor) ligados na posição *para* do anel benzênico, também foram considerados favoráveis para a atividade como visto a partir do GCOD 7 (1,4,2,np). Apesar de ser um grupo aceptor de ligação de hidrogênio, não foi constatada nenhuma ligação de hidrogênio nesta região, portanto, pode-se atribuir esta estabilidade às interações eletrostáticas que estão presentes em toda a superfície do composto, como pode ser observado na Figura 23 para o composto 56.

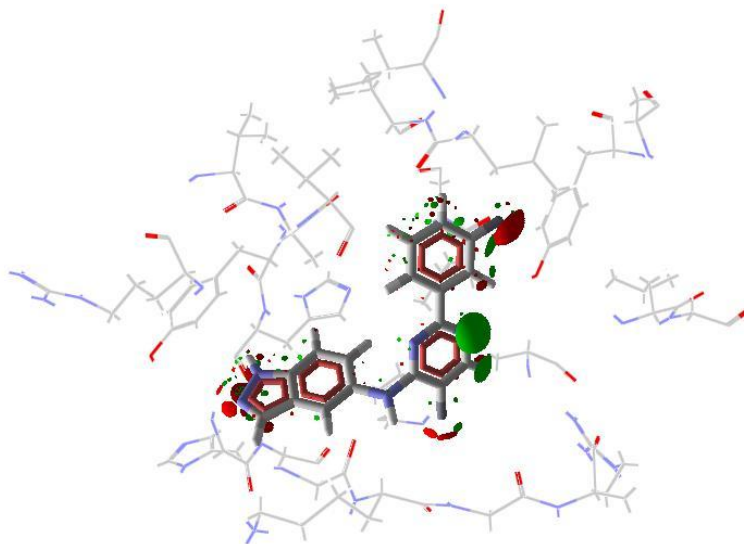
Figura 23 – Mapa de energia potencial eletrostático do composto **56**



Fonte: Do Autor (2018)

O átomo de flúor neste composto realiza interações eletrostáticas com os resíduos de aminoácidos Phe262, Leu278, Val279, e Ser280. E mais uma vez, os resultados do ancoramento corroboram com os do QSAR, pois o átomo de Flúor na posição *para* não apresenta nenhuma interação eletrostática negativa, enquanto que, quando o Flúor está na posição *meta* ele não apresenta interação eletrostática positiva, somente negativa, como pode ser observado na Figura 24, para o composto 36.

Figura 24 – Mapa de potencial eletrostáticas do composto **36**

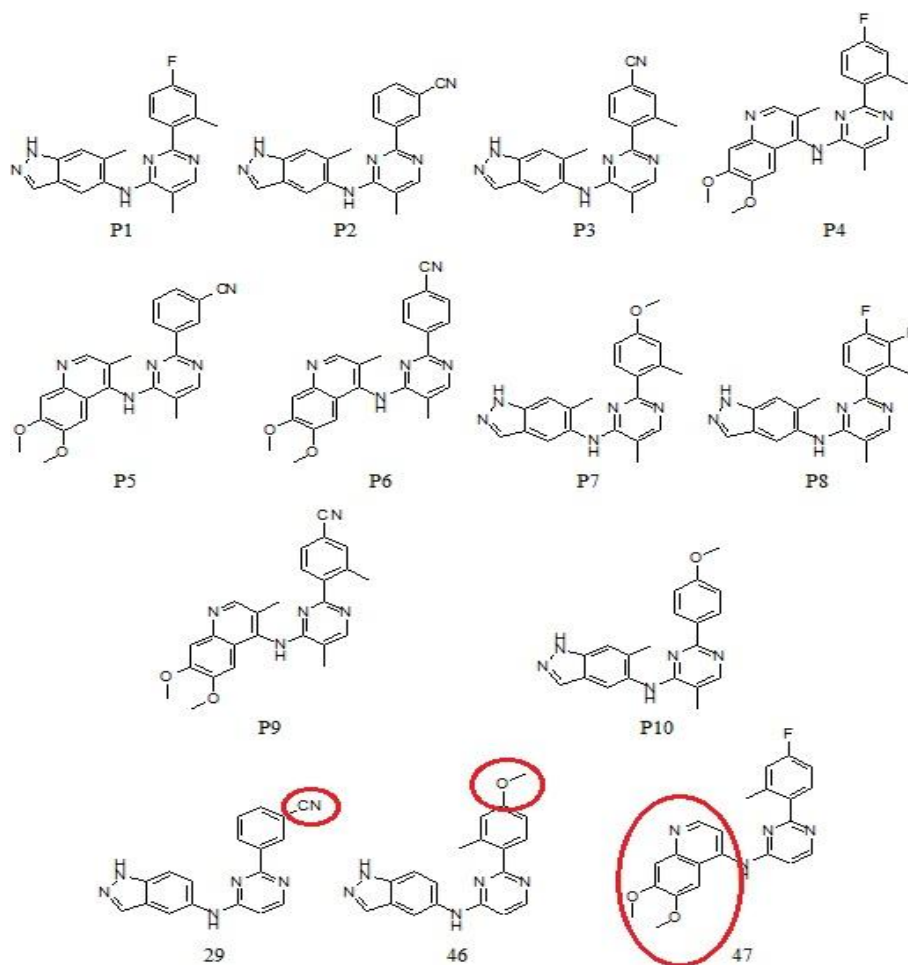


Nota: O potencial eletrostático positivo é mostrado em verde e o negativo em vermelho.
Fonte: Do Autor (2018)

5.4. Planejamento de potenciais inibidores da enzima TGF- β 1

A partir dos resultados de QSAR e *docking* foram propostas dez moléculas, as quais foram calculadas as atividades previstas a partir da equação 11 e também foram analisadas quanto às interações realizadas entre os compostos e a proteína TGF- β 1 pelo estudo de *docking*. As estruturas das 10 moléculas propostas são apresentadas na Figura 25.

Figura 25 – Estruturas das moléculas propostas e dos compostos 29, 46 e 47, que foram utilizados como referência para construção das propostas.



Fonte: Do Autor (2018)

Além dos grupos que foram considerados favoráveis para a atividade, obtidos a partir do QSAR-4D, para propor esses compostos foram considerados as estruturas das moléculas de maior atividade biológica (**29**, **46**, e **47**), pois foi verificado a partir do *docking* que grupos presentes

nessas moléculas (destacados em vermelho) podem realizar várias interações com a proteína, estabilizando ainda mais o complexo.

Os valores de atividade predita (inibição da sinalização) para as moléculas propostas, assim como os valores de energia de *docking* são apresentadas na Tabela 11.

Tabela 11 – Valores de atividade predita e energia de docking para os inibidores propostos

Moléculas Propostas	Atividade Predita	MolDock Score	Energia estérica	Energia inter	Energia Eletrost	Ligação de Hidrogênio
P1	6,54	-142,71	-147,78	-146,23	-0,17	-2,55
P2	6,32	-152,92	-158,95	-162,39	-2,42	-3,31
P3	<u>7,74</u>	-140,56	-147,44	-148,96	3,43	-2,71
P4	6,58	-153,97	-165,25	-162,46	1,49	-2,60
P5	7,08	-144,59	-155,46	-157,43	0,14	-3,27
P6	<u>8,93</u>	-122,77	-130,39	-134,44	-2,75	-4,86
P7	6,76	-141,63	-148,53	-147,15	-0,08	-2,5
P8	6,28	-138,77	-147,33	-145,80	0,02	-2,5
P9	<u>8,84</u>	-144,58	-147,72	-157,91	-3,05	-8,99
P10	<u>8,39</u>	-127,83	-140,57	-139,16	0,05	-2,54

Fonte: Do Autor (2018)

A partir da tabela acima, pode-se notar que dos dez compostos propostos, quatro apresentaram atividades preditas maiores (sublinhado na Tabela 11) do que à maior atividade experimental, que é a do composto **47** (7,35). Dentre esses quatro compostos, três (**P3**, **P6** e **P9**) apresentam um grupo nitrila em substituição ao átomo de flúor e em uma posição diferente àquela encontrada nos compostos estudados, o que parece melhorar as energias de interação entre proteína-ligante e consequentemente as atividades. Isto pode ser comprovado quando comparamos os valores de MolDock Score (energia total) para o composto **29** (-130,09 kcal.mol⁻¹) que possui o grupo nitrila na posição *meta* do anel benzênico e o composto **P3**

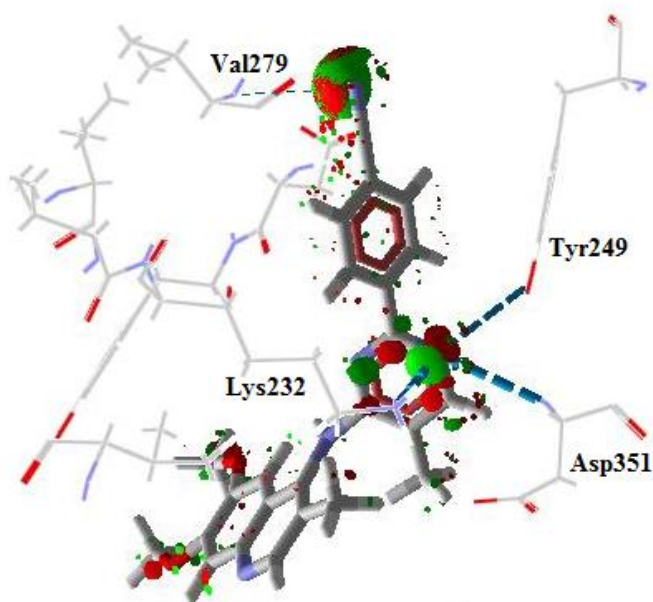
(-140,56 kcal.mol⁻¹) que possui esse grupo na posição *para*. Além da diferença na posição do grupo nitrila, o composto **P3** ainda possui três grupos metilas, o que fez com que a energia de interação estérica melhorasse e consequentemente também a energia total, passando de -145,32 kcal.mol⁻¹ do composto **29** para -147,44 kcal.mol⁻¹ do composto **P3**.

O composto **P8** foi proposto a fim de verificar a influência do átomo de flúor quando ligado ao anel benzênico. Comparando **P1** e **P8**, percebe-se que quando o átomo de flúor está presente na posição *meta* a atividade diminui, o que corrobora com os resultados de QSAR e *docking*, que mostrou que quando o Flúor se encontra nessa posição existem muitas interações eletrostáticas negativas.

Dentre os compostos propostos, o **P6** foi o que apresentou maior atividade predita. Ele possui em sua estrutura química além dos grupos encontrados no QSAR, parte do composto **47**. Comparando os valores de energias obtidas para estas duas moléculas, percebe-se que **P6** obteve energias de interação (MolDock Score, Estérica e Intermolecular) maiores do que o composto **47**. No entanto os valores de energia de interação eletrostática e de ligação de hidrogênio foram inferiores, o que pode ter aumentado a atividade.

Na Figura 26 está apresentado o composto proposto **P6** dentro do sítio de ligação da proteína. Nesta figura pode verificar que este composto realiza quatro ligações de hidrogênio, duas com resíduos já apresentados (Lys 232 e Asp351) e outras duas com resíduos que não foram relatados no ancoramento dos compostos em estudo (Tyr249 e Val279). Além disso, observam-se também as interações eletrostáticas, que já foram reconhecidas como importantes para atividade.

Figura 26 – Composto proposto **P6** dentro do sítio de ligação da proteína TGF- β 1



Nota: As ligações de Hidrogênio são apresentadas por linhas pontilhadas. As interações eletrostáticas são apresentadas por esferas verdes (positiva) e vermelhas (negativa).

Fonte: Do Autor (2018)

6. CONSIDERAÇÕES FINAIS

O câncer é sem dúvidas um problema de saúde pública mundial, uma vez que ele afeta pessoas de todas as idades e de todos os continentes. Neste estudo, construímos e avaliamos modelos que fornecem informações importantes que podem auxiliar no planejamento de fármacos para serem utilizados como inibidores da TGF- β 1. Dos nove alinhamentos testados, o alinhamento 4 foi o que apresentou melhores resultados com todos os grupos testes avaliados. O melhor modelo desse alinhamento apresentou coeficiente de determinação $R^2 = 0,79$, bons parâmetros de validação interna, $q^2 = 0,58$, e validação externa, $R^2_{\text{pred}} = 0,65$, além de grupos farmacofóricos importantes (como os grupos metilas e o átomo de flúor) para a atividade biológica. A partir do ancoramento molecular foi possível analisar quais aminoácidos fazem interações com os compostos e verificar a importâncias destas interações para a atividade.

Com base nestes resultados foram propostos dez compostos, dos quais quatro apresentaram atividades preditas maiores do que as atividades dos compostos estudados, e também apresentaram energias de interação favoráveis no ancoramento molecular. Portanto, esses compostos podem ser considerados como promissores protótipos para o tratamento do câncer.

REFERÊNCIAS

- ACCELRY'S SOFTWARE INCORPORATED. **Discovery Studio Modeling Environment, Release 2.1**. San Diego: [s.n], 2007.
- ANDRADE, C. H. et al. 4D-QSAR: Perspectives in Drug Design. **Molecules**, Washington, v. 15, n. 5, p. 3281-3294, 2010.
- ARAÚJO, J. Q. et al. Docking of the alkaloid geissospermine into acetylcholinesterase: a natural scaffold targeting the treatment of Alzheimer's disease. **Journal of molecular modeling**, Berlin, v. 17, n. 6, p. 1401-1412, 2011.
- BARNARD, J. A.; LYONS, R. M.; MOSES, H. L. The cell biology of transforming growth factor β . **Biochimica et Biophysica Acta (BBA)-Reviews on Cancer**, Boston, v. 1032, n. 1, p. 79-87, 1990.
- BARREIRO, E. J. et al. A química medicinal de N-acilidrazonas: novos compostos-protótipos de fármacos analgésicos, anti-inflamatórios e anti-trombóticos. **Química Nova**, São Paulo, v. 25, n. 1, p. 129-148, 2002.
- BARREIRO, E. J.; FRAGA, C. A. M. The question of innovative drugs in Brazil: the proposal of the Pronfar creation. **Química Nova**, São Paulo, v. 28, p. S56-S63, 2005.
- BATISTA, D. R. R.; DE MATTOS, M.; DA SILVA, S. F. Convivendo com o câncer: do diagnóstico ao tratamento. **Revista de Enfermagem da UFSM**, Santa Maria, v. 5, n. 3, p. 499-510, 2015.
- BOING, A. F.; VARGAS, S. A. L.; BOING, A. C. A carga das neoplasias no Brasil: mortalidade e morbidade hospitalar entre 2002-2004. **Revista da Associação Médica Brasileira**, São Paulo, v. 88040, p. 900, 2007.
- BONAFOUX, D. et al. 2-Aminoimidazoles inhibitors of TGF- β receptor 1. **Bioorganic & Medicinal Chemistry Letters**, London, v. 19, n. 3, p. 912-916, 2009.
- BORSATTO, A. Z.; VIDAL, M. L. B.; ROCHA, R. Vacina contra o HPV e a prevenção do câncer do colo do útero: subsídios para a prática. **Revista Brasileira de Cancerologia**, Rio de Janeiro, v. 57, n. 1, p. 67-74, 2011.
- BRASIL. **Guia Prático sobre o HPV: perguntas e respostas**. Brasília: Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância de Doenças Transmissíveis. Coordenação Geral do Programa Nacional de Imunizações, 2013.

CALDAS, G. B.; RAMALHO, T. C.; DA CUNHA, E. F. Application of 4D-QSAR studies to a series of benzothiophene analogs. **Journal of Molecular Modeling**, Amsterdam, v. 20, n. 10, p. 2420, 2014.

CHERKASOV, A. et al. QSAR modeling: where have you been? Where are you going to? **Journal of Medicinal Chemistry**, Washington, v. 57, n. 12, p. 4977-5010, 2014.

DA CUNHA, E. F. F. et al. 4D-QSAR Models of HOE/BAY-793 Analogues as HIV-1 Protease Inhibitors. **Molecular Informatics**, Weinheim, v. 24, n. 2, p. 240-253, 2005.

DA ROCHA PITA, S. S. Modelos de QSAR-4D dependente do receptor de inibidores peptídicos da tripanotona redutase. **Instituto de Química, Universidade Federal do Rio de Janeiro**, Rio de Janeiro, 2006.

DA SILVA, C. S. et al. PSICOLOGIA E CÂNCER: UM OLHAR MULTI DISCIPLINAR. **Gep News**, Maceió, v. 1, n. 2, p. 64-68, 2017.

DA SILVA, P. L. N. et al. O significado do câncer: percepção de pacientes. **Revista de enfermagem UFPE on line**, Recife, v. 7, n. 12, p. 6828-6833, 2013.

DERYNCK, R.; AKHURST, R. J.; BALMAIN, A. TGF- β signaling in tumor suppression and cancer progression. **Nature genetics**, New York, v. 29, n. 2, p. 117-129, 2001.

DUFFY, B. C. et al. Early phase drug discovery: cheminformatics and computational techniques in identifying lead series. **Bioorganic & Medicinal Chemistry**, Oxford, v. 20, n. 18, p. 5324-5342, 2012.

DUMONT, N.; ARTEAGA, C. L. Targeting the TGF β signaling network in human neoplasia. **Cancer Cell**, Cambridge, v. 3, n. 6, p. 531-536, 2003.

EFTIMIE, R.; PEREZ, M.; BUONO, P.-L. Pattern formation in a nonlocal mathematical model for the multiple roles of the TGF- β pathway in tumour dynamics. **Mathematical Biosciences**, Atlanta, v. 289, p. 96-115, 2017.

FERLAY, J. et al. **GLOBOCAN 2012 v1. 0, Cancer Incidence and Mortality Worldwide: IARC CancerBase. n. 11**, Lyon, 2012.

FERNANDES, A. G. Zinco e câncer: uma revisão. **Saúde. com**, Vitória da Conquista, v. 1, n. 2, 2016.

- FERRARI, C. K. B.; TORRES, E. Novos compostos dietéticos com propriedades anticarcinogênicas. **Revista Brasileira de Cancerologia**, Rio de Janeiro, v. 48, n. 3, p. 375-382, 2002.
- FERREIRA, M. Multivariate Qsar. **Journal of the Brazilian Chemical Society**, Campinas, v. 13, n. 6, p. 742-753, 2002.
- FERREIRA, M. M. C.; MONTANARI, C. A.; GAUDIO, A. C. Variable selection in QSAR. **Química Nova**, São Paulo, v. 25, n. 3, p. 439-448, 2002.
- FRIGO, L. F.; DE OLIVEIRA ZAMBARDA, S. Câncer do colo de útero: efeitos do tratamento. **Cinergis**, Santa Cruz do Sul, v. 16, n. 3, 2015.
- FUENTES-CALVO, I.; MARTÍNEZ-SALGADO, C. TGFB1 (transforming growth factor, beta 1). **Atlas of Genetics and Cytogenetics in Oncology and Haematology** Poitiers, v. 17, n. 8, 2013.
- GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, Louisville, v. 185, p. 1-17, 1986.
- GELMAN, M. et al. **TGF- β INHIBITORS**. WO 2016/160833 A1: 1-98 p., 1 April 2015, 6 October 2016.
- GONÇALVES, R. **Desenvolvimento e validação de Novos Métodos de distribuição da população Inicial em Algoritmos Genéticos para o problema de Docking Proteína-Ligante**. 2008. 100 p. Dissertação (Mestrado em Modelagem Computacional) - Laboratório Nacional de Computação Científica, Petrópolis, 2008.
- GUIDO, R. V.; ANDRICOPULO, A. D.; OLIVA, G. Planejamento de fármacos, biotecnologia e química medicinal: aplicações em doenças infecciosas. **Estudos Avançados**, São Paulo, v. 24, n. 70, p. 81-98, 2010.
- GUO, X.; WANG, X.-F. Signaling cross-talk between TGF- β /BMP and other pathways. **Cell Research**, Shanghai, v. 19, n. 1, p. 71-88, 2009.
- GUTGOLD, N. et al. TGF β splicing and canonical pathway activation in high-grade serous carcinoma. **Virchows Archiv**, Florence, v. 470, n. 6, p. 665-678, 2017.
- HANSCH, C.; FUJITA, T. ρ - σ - π Analysis. A method for the correlation of biological activity and chemical structure. **Journal of the American Chemical Society**, Easton, v. 86, n. 8, p. 1616-1626, 1964.

HIRSCHHORN, T. et al. Differential molecular regulation of processing and membrane expression of Type-I BMP receptors: implications for signaling. **Cellular and Molecular Life Sciences**, Basel, p. 1-18, 2017.

HOODLESS, P.; WRANA, J. Mechanism and function of signaling by the TGF β superfamily. In: Pawson A.J. (Ed.) Protein Modules in Signal Transduction. **Current Topics in Microbiology and Immunology**, Berlin, v. 228, p.235-272, Springer, 1998.

HOPFINGER, A. J.; PEARLSTEIN, R. A. Molecular mechanics force-field parameterization procedures. **Journal of Computational Chemistry**, New York, v. 5, n. 5, p. 486-499, 1984.

HOPFINGER, A. J. et al. Construction of 3D-QSAR Models Using the 4D-QSAR Analysis Formalism. **Journal of the American Chemical Society**, Easton, v. 119, n. 43, p. 10509-10524, 1997.

HUI, A. Y.; FRIEDMAN, S. L. Molecular basis of hepatic fibrosis. **Expert Reviews in Molecular Medicine**, Cambridge, v. 5, n. 5, p. 1-23, 2003.

INCA. Estimativa 2018: Incidência de câncer no Brasil. **Instituto Nacional de Câncer José Alencar Gomes da Silva**, Rio de Janeiro, 128 p., 2017.

INCA. ABC do Câncer: abordagens básicas para o controle do câncer. **Instituto Nacional de Câncer José Alencar Gomes da Silva**, Rio de Janeiro, 3. ed., 108 p., 2017.

KARKI, R. G.; KULKARNI, V. M. Three-dimensional quantitative structure–activity relationship (3D-QSAR) of 3-aryloxazolidin-2-one antibacterials. **Bioorganic & Medicinal Chemistry**, Oxford, v. 9, n. 12, p. 3153-3160, 2001.

KAWAUCHI, K. et al. **Japanese Kokai Tokkyo Koho**, JP06172377 [Chem. Abstr. 1995, 123, 33585f], 1994.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, Evanston, v. 11, n. 1, p. 137-148, 1969.

KIMURA, E. T.; MATSUO, S. E.; RICARTE-FILHO, J. C. TGF β , activin and SMAD signalling in thyroid cancer. **Arquivos Brasileiros de Endocrinologia & Metabologia**, São Paulo, v. 51, n. 5, p. 683-689, 2007.

KIRALJ, R.; FERREIRA, M. 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.

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

KUBINYI, H. **QSAR: Hansch Analysis and Related Approaches**. Methods and Principles in Medicinal Chemistry, Germany, John Wiley & Sons, 2008.

LEACH, A. R. **Molecular Modelling: Principles and Applications**. Prentice Hall, Harlow, 2. ed., 2001.

LEBRUN, J.-J. The dual role of TGF in human cancer: from tumor suppression to cancer metastasis. **ISRN Molecular Biology**, New York, v. 2012, 2012.

LI, J. et al. A low-frequency variant in SMAD7 modulates TGF- β signaling and confers risk for colorectal cancer in Chinese population. **Molecular Carcinogenesis**, Hoboken, v.56, n. 7, p. 1798-1807, 2017.

LIMA, L. M. Química Medicinal Moderna: desafios e contribuição brasileira. **Química Nova**, São Paulo, v. 30, n. 6, p. 1456, 2007.

LUNA, D. et al. Interfacial properties of a novel pyrimidine derivative and poly (ethylene glycol)-grafted phospholipid floating monolayers. **Colloids and Surfaces A: Physicochemical and Engineering Aspects**, Biejing, v. 373, n. 1-3, p. 22-28, 2011.

MAGNUS, N. A.; CONFALONE, P. N.; STORACE, L. A new asymmetric 1, 4-addition method: application to the synthesis of the HIV non-nucleoside reverse transcriptase inhibitor DPC 961. **Tetrahedron Letters**, Oxford, v. 41, n. 17, p. 3015-3019, 2000.

MAYO, S. L.; OLAFSON, B. D.; GODDARD, W. A. DREIDING: A generic force field for molecular simulations. **Journal of Physical Chemistry**, Washington, v. 94, n. 26, p. 8897-8909, 1990.

MOLEGRO, **MVD Molegro Virtual Docker**. DK-8000 Aarhus C, Denmark.

MOLSIM, D. D. **Molecular Mechanics and Dynamics Simulation Software, users guide, version 3.2**. Lake Forest: The Chem21 Group, 1997.

NEUZILLET, C. et al. Targeting the TGF β pathway for cancer therapy. **Pharmacology & therapeutics**, Kansas City, v. 147, p. 22-31, 2015.

NUCCI, N. A. G. **Qualidade de Vida e Câncer: um estudo compreensivo**. 225 p. Tese (Doutorado em Psicologia) Universidade de São Paulo, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, 2003.

OMS, Disponível em: < <https://nacoesunidas.org/oms-cancer-mata-88-milhoes-de-pessoas-anualmente-no-mundo/> >.

PASQUALOTO, K. F. et al. Rational design of new antituberculosis agents: receptor-independent four-dimensional quantitative structure– activity relationship analysis of a set of isoniazid derivatives. **Journal of Medicinal Chemistry**, Washington, v. 47, n. 15, p. 3755-3764, 2004.

PIRHADI, S.; SHIRI, F.; GHASEMI, J. B. Multivariate statistical analysis methods in QSAR. **RSC Advances**, Cambridge, v. 5, n. 127, p. 104635-104665, 2015.

POLANSKI, J. Receptor dependent multidimensional QSAR for modeling drug-receptor interactions. **Current Medicinal Chemistry**, Cambridge, v. 16, n. 25, p. 3243-3257, 2009.

PRATIM ROY, P. et al. On two novel parameters for validation of predictive QSAR models. **Molecules**, Washington, v. 14, n. 5, p. 1660-1701, 2009.

PRO, P. S. **1.0. 5; Wavefunction**. Inc.: Irvine, 2000.

RAMALHO, T. C. et al. Construction and assessment of reaction models of class I EPSP synthase: Molecular docking and density functional theoretical calculations. **Journal of Biomolecular Structure and Dynamics**, Albany, v. 27, n. 2, p. 195-207, 2009.

RAMALHO, T. C. et al. Computational enzymology and organophosphorus degrading enzymes: Promising approaches toward remediation technologies of warfare agents and pesticides. **Current Medicinal Chemistry**, Cambridge, v. 23, n. 10, p. 1041-1061, 2016.

ROBERTS, A. B.; SPORN, M. B. **The transforming growth factor- β s**. In: Peptide Growth Factors and Their Receptors I, Springer, New York, p. 419-472, 1991.

RODRIGUES, R. P. et al. Estratégias de Triagem Virtual no Planejamento de Fármacos. **Revista Virtual de Química**, Niterói, v. 4, n. 6, p. 739-776, 2012.

ROGERS, D. **WOLF 6.2 GFA Program, Molecular Simulation**, Inc.: San Diego, CA, 1994.

ROGERS, D.; HOPFINGER, A. J. Application of Genetic Function Approximation to Quantitative Structure-Activity Relationships and Quantitative Structure-Property Relationships.

Journal of Chemical Information and Computer Science, Washington, v. 34, n. 4, p. 854-866, 1994.

ROSEN, L. S.; ASHURST, H. L.; CHAP, L. Targeting signal transduction pathways in metastatic breast cancer: a comprehensive review. **The Oncologist**, Boston, v. 15, n. 3, p. 216-235, 2010.

SANT, C. M. R. Métodos de modelagem molecular para estudo e planejamento de compostos bioativos: Uma introdução. **Revista Virtual de Química**, Niterói, v. 1, n. 1, p. 49-57, 2009.

SANTOS-FILHO, O. A. et al. The Receptor-Dependent QSAR Paradigm: An Overview of the Current State of the Art. **Medicinal Chemistry**, New York, v. 5, n. 4, p. 359-366, 2009.

SCHEIBER, J. Backtranslating clinical knowledge for use in cheminformatics: What is the potential? **Bioorganic & Medicinal Chemistry**, Oxford, v. 20, n. 18, p. 5461-5463, 2012.

SENESE, C. L. et al. 4D-fingerprints, universal QSAR and QSPR descriptors. **Journal of Chemical Information and Computer Sciences**, Washington, v. 44, n. 5, p. 1526-1539, 2004.

SILVA, D. R.; RAMALHO, T. C.; DA CUNHA, E. F. F. 4D-QSAR model for compounds with binding affinity towards dopamine D2 receptors. **Letters in Drug Design & Discovery**, San Antonio, v. 11, n. 5, p. 649-664, 2014.

SOUZA, D. M.; MADARI, B. E.; GUIMARÃES, F. F. Aplicação de técnicas multivariadas e inteligência artificial na análise de espectros de infravermelho para determinação de matéria orgânica em amostras de solos. **Química Nova**, São Paulo, v. 35, n. 9, p. 1738-1745, 2012.

STEWART, J. J. Optimization of parameters for semiempirical methods II. Applications. **Journal of Computational Chemistry**, Ann Arbor, v. 10, n. 2, p. 221-264, 1989.

SUSSMAN, J. L. et al. Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. **Acta Crystallographica Section D: Biological Crystallography**, Liverpool, v. 54, n. 6, p. 1078-1084, 1998.

TAVARES, L. C. QSAR: A abordagem de Hansch. **Química Nova**, São Paulo, v. 27, n. 4, p. 631-639, 2004.

THOMSEN, R.; CHRISTENSEN, M. H. MolDock: a new technique for high-accuracy molecular docking. **Journal of Medicinal Chemistry**, Washington, v. 49, n. 11, p. 3315-3321, 2006.

VEDANI, A.; DOBLER, M. 5D-QSAR: the key for simulating induced fit? **Journal of Medicinal Chemistry**, Washington, v. 45, n. 11, p. 2139-2149, 2002.

VEDANI, A.; DOBLER, M.; LILL, M. A. Combining protein modeling and 6D-QSAR. Simulating the binding of structurally diverse ligands to the estrogen receptor. **Journal of Medicinal Chemistry**, Washington, v. 48, n. 11, p. 3700-3703, 2005.

VERLI, H.; BARREIRO, E. J. A medicinal chemistry paradigm: ligands and receptor flexibility. **Química Nova**, São Paulo, v. 28, n. 1, p. 95-102, 2005.

WASSERMANN, A. M. et al. Advances in Computational Medicinal Chemistry: Matched Molecular Pair Analysis. **Drug Development Research**, New York v. 73, n. 8, p. 518-527, 2012.

WEINER, S. J. et al. An all atom force field for simulations of proteins and nucleic acids. **Journal of Computational Chemistry**, New York, v. 7, n. 2, p. 230-252, 1986.

ZHOU, D. et al. Cytidine monophosphate kinase is inhibited by the TGF- β signalling pathway through the upregulation of miR-130b-3p in human epithelial ovarian cancer. **Cellular Signalling**, Farmington, v. 35, p. 197-207, 2017.