



TAMIRIS MARIA DE ASSIS

**ESTUDOS DE QSAR DEPENDENTE DO
RECEPTOR EM ATIVADORES DA PROTEÍNA
GLUCOQUINASE**

LAVRAS – MG

2015

TAMIRIS MARIA DE ASSIS

**ESTUDOS DE QSAR DEPENDENTE DO RECEPTOR EM
ATIVADORES DA PROTEÍNA GLUCOQUINASE**

Dissertação 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 Mestre.

Orientadora

Dra. Elaine Fontes Ferreira da Cunha

LAVRAS - MG

2015

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Assis, Tamiris Maria de.

Estudos de QSAR dependente do receptor em ativadores da
proteína Glucoquinase / Tamiris Maria de Assis. – Lavras :
UFLA, 2015.

89 p. : il.

Dissertação (mestrado acadêmico)–Universidade Federal de
Lavras, 2015.

Orientador(a): Elaine Fontes Ferreira da Cunha.

Bibliografia.

1. Diabetes Mellitus tipo 2. 2. Enzima glucoquinase. 3.
QSAR-4D DR. I. Universidade Federal de Lavras. II. Título.

TAMIRIS MARIA DE ASSIS

**ESTUDOS DE QSAR DEPENDENTE DO RECEPTOR EM
ATIVADORES DA PROTEÍNA GLUCOQUINASE**

Dissertação 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 Mestre.

APROVADA em 26 de fevereiro de 2015.

Dra. Ana Paula Guimarães	UFV
Dra. Denise Alvarenga Rocha	UFLA
Dr. Teodorico de Castro Ramalho	UFLA

Dra. Elaine Fontes Ferreira da Cunha
Orientadora

LAVRAS - MG

2015

AGRADECIMENTOS

Primeiramente, 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 noivo 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, Denise e Teodorico, pela disposição e pelas contribuições que complementarão o trabalho.

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

RESUMO

O Diabetes Mellitus (DM) é um distúrbio metabólico resultante da falta absoluta ou resistência à insulina. Existem vários tipos de diabetes, sendo que a forma predominante em 90-95% dos casos é o diabetes tipo 2 (DM2). Este ocorre, principalmente, na fase adulta e tem o sedentarismo como principal fator agravante. Os medicamentos existentes para o controle do DM2 possuem vários efeitos colaterais e uma ineficiência quando utilizado por longos períodos. Objetivando-se propor novos candidatos a fármacos antidiabéticos, estudaram-se 54 ativadores da proteína glucoquinase (GQ), uma enzima que catalisa a fosforilação da glicose e, portanto, é uma peça importante no controle da homeostase da glicose plasmática, por meio da metodologia de relação quantitativa estrutura-atividade dependente do receptor (QSAR-4D DR). As conformações obtidas por simulação de dinâmica molecular foram sobrepostas, de acordo com os doze alinhamentos testados, em uma caixa virtual tridimensional composta por células de 2 Å. Os modelos foram gerados pela técnica que combina algoritmos genéticos (GA, do inglês *Genetic Algorithm*) e mínimos quadrados parciais (PLS, do inglês *Partial Least Square*). O melhor alinhamento gerou modelos com coeficiente de determinação (R^2) entre 0,674 e 0,743 e de validação cruzada (Q^2) entre 0,509 e 0,610, indicando uma boa capacidade preditiva. O modelo 1 do melhor alinhamento foi escolhido por apresentar o menor valor de LOF (do inglês lack-of-fit). Esse modelo forneceu informações importantes para o desenho racional de novos ativadores glucoquinase, como a presença de um grupo aromático como a piridina e de um grupo eletronegativo como o trifluórmethyl na molécula. Foi possível propor compostos como **P1**, **P2** e **P3**, que, de acordo com a equação gerada, possuem uma atividade maior que a apresentada para os compostos estudados, portanto são promissores protótipos antidiabéticos.

Palavras-chave: Diabetes Mellitus tipo 2. Enzima glucoquinase. QSAR-4D DR.

ABSTRACT

Diabetes Mellitus (DM) is a metabolic disorder resulting from complete failure or insulin resistance. There are many types of diabetes, with the predominant form in 90-95% of cases being Type 2 Diabetes (T2DM). This type occurs mainly in adult life and has sedentariness as its main aggravating factor. The medications for controlling T2DM have several side effects and they are inefficient when used over long periods of time. Aiming to propose new candidates for anti-diabetic drugs, we studied 54 glucokinase protein (GK) activators, an enzyme that catalyzes glucose phosphorylation and is, therefore, an important piece in controlling plasma glucose homeostasis by means of the quantitative structure-activity relationship receptor dependent (4D-QSAR RD) methodology. The conformations obtained by molecular dynamics simulation were superimposed according to the twelve alignments tested in a virtual three-dimensional box comprised by 2Å cells. The models were generated by the technique that combines genetic algorithms (GA) and partial least squares (PLS). The best alignment generated models with determination coefficient (R^2) between 0.674 and 0.743 and cross-validation (Q^2) between 0.509 and 0.610, indicating good predictive capacity. Model 1 of the best alignment was chosen for presenting the lowest LOF (lack-of-fit) value. This model provided important information for the rational design of new glucokinase activators such as the presence of an aromatic group such as pyridine and an electronegative group such as trifluoromethyl in the molecule. It was possible to propose compounds such as **P1**, **P2** and **P3**, which, according to equation generated, present higher activity than that presented for the studied compounds, thus being promising anti-diabetics prototypes.

Keywords: Type 2 Diabetes Mellitus. Glucokinase enzyme. 4D-QSAR RD.

LISTA DE FIGURAS

Figura 1	Número estimado de adultos com diabetes por grupo de idade, ano e categorias de países desenvolvidos e em desenvolvimento.....	18
Figura 2	Comparação por regiões do número de casos de diabetes em 2013 e a proporção dos casos que não são diagnosticados	20
Figura 3	Estrutura tridimensional da enzima Glucoquinase Humana	23
Figura 4	Reação de fosforilação da glicose	24
Figura 5	Comparação da atividade relativa das enzimas Hexoquinase I e IV pela concentração de glicose	24
Figura 6	Gráfico de distribuição dos valores de atividade biológica (pEC ₅₀) dos 54 compostos estudados	48
Figura 7	Imagem ilustrativa do corte realizado juntamente com o composto 1 em estudo	57
Figura 8	Valores de energia total para SDM de 50 ps para os complexos ligante-proteína (1).....	58
Figura 9	Valores de pEC ₅₀ experimentais versus preditos do conjunto de treinamento, usando o melhor modelo de QSAR-4D-DR do alinhamento 9.....	65
Figura 10	Modelo de QSAR-4D DR representativo do alinhamento 9, mostrando o composto 1 juntamente com os resíduos de aminoácidos selecionados da GQ.....	69
Figura 11	Representação da molécula 55 e do GCOD-1 (esfera verde) obtido do Modelo1 do alinhamento 9.....	70
Figura 12	Representação da molécula 27 e do GCOD-2 (esfera verde) obtido do Modelo1 do alinhamento 9.....	71

Figura 13	Representação da molécula 43 e do GCOD-3 (esfera verde) obtido do Modelo1 do alinhamento 9.....	73
Figura 14	Representação da molécula 17 e do GCOD-4 (esfera verde) obtido do Modelo1 do alinhamento 9.....	74
Figura 15	Representação da molécula 35 e do GCOD-5 (esfera verde) obtido do Modelo1 do alinhamento 9.....	75

LISTA DE TABELAS

Tabela 1	Estrutura dos 54 ativadores glucoquinase com os respectivos valores de pEC ₅₀	38
Tabela 2	Alinhamentos testados e os átomos de C _α da proteína selecionados para o alinhamento dos complexos	52
Tabela 3	Definição dos elementos de interação farmacofórica, IPEs	53
Tabela 4	Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os doze alinhamentos.....	60
Tabela 5	Resumo dos resultados dos modelos de QSAR-4D obtidos para o Alinhamento 9. LSE (Least Squares Error) é o erro dos mínimos quadrados e LOF (lack-of-fit) é o valor de desajuste de Friedman.....	62
Tabela 6	Matriz de correlação cruzada dos GCODs do melhor modelo do alinhamento 9	63
Tabela 7	Valores de pEC ₅₀ experimentais e preditos e valores residuais (pEC ₅₀ predito - pEC ₅₀ experimental) para os compostos do conjunto de treinamento	65
Tabela 8	Valores de pEC ₅₀ experimentais e preditos e valores residuais (pEC ₅₀ predito - pEC ₅₀ experimental) para os compostos do conjunto de teste.....	67
Tabela 9	Valores de ocupação percentual do GCOD-2 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos 5, 9, 27 e 36	72
Tabela 10	Valores de ocupação percentual do GCOD-3 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos 16, 17, 27 e 43	73

Tabela 11	Valores de ocupação percentual do GCOD-5 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos 4, 5, 22 e 35	76
Tabela 12	Valores de ocupação percentual do GCOD-6 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos 1, 2, 22 e 34	77
Tabela 13	Valores de atividade preditos ($pEC_{50predito}$) para os inibidores P1, P2, P3, P4, P5, P6 e 50	81

LISTA DE ABREVIATURAS

Ala	Alanina
Arg	Arginina
His	Histidina
Ile	Isoleucina
Met	Metionina
Thr	Treonina
Tyr	Tirosina
Val	Valina

LISTA DE SIGLAS

ATP	Trifosfato de adenosina
DM	Diabetes Mellitus
IDF	International Diabetes Federation
OMS	Organização Mundial de Saúde
TOTG	Teste Oral de Tolerância à Glicose
DM2	Diabetes mellitus tipo 2
GQ	Glucoquinase
PDB	<i>Protein Data Bank</i> (Banco de Dados de Proteínas)
PRGQ	<i>Proteína Reguladora da Glucoquinase</i>
AGQ	Ativadores Glucoquinase
IUPAC	<i>International Union of Pure and Applied Chemistry</i>
QSAR	Relação quantitativa estrutura-atividade
QSAR-4D DR	Quarta dimensão da relação quantitativa estrutura-atividade dependente do receptor
QSAR-4D IR	Quarta dimensão da relação quantitativa estrutura-atividade independente do receptor
SDM	Simulação da dinâmica molecular
GCODs	<i>Grid Cell Occupancy Descriptors</i>
IPE	<i>Interaction pharmacophore Elements</i>
MM	Mecânica Molecular
CEP	<i>Conformational Ensemble profile</i>
PLS	<i>Partial Least Squares</i>
GA	<i>Genetic Algorithm</i>
GFA	<i>Genetic Function Approximation</i>
LOF	<i>Lack-of-fit</i>

LISTA DE SÍMBOLOS

Å	Angstrom
EC ₅₀	concentração capaz de induzir a metade do efeito máximo
K	Kelvin
µM	Micromolar
M	Molar
kcal mol ⁻¹	Quilocaloria por mol
pEC ₅₀	Log de EC ₅₀

SUMÁRIO

1	INTRODUÇÃO	15
2	REFERENCIAL TEÓRICO.....	17
2.1	Diabetes Mellitus.....	17
2.1.1	Diabetes Mellitus Tipo 2.....	21
2.2	Proteína Glucoquinase.....	22
2.3	Ativadores Glucoquinase.....	25
2.4	A química medicinal e o planejamento de fármacos.....	26
2.5	Relação quantitativa estrutura- atividade	28
2.5.1	QSAR 4D.....	30
2.6	Métodos e índices estatísticos	31
2.6.1	Índice de Falta de Ajuste (LOF, <i>lack-of-fit</i>).....	31
2.6.2	Coefficiente de Correlação Linear Quadrático (r^2 ou R^2)	32
2.6.3	Validação Interna dos modelos.....	32
2.6.4	Validação Externa dos modelos.....	33
2.6.5	Validação Y-randomization ($R^2_{Y\text{-rand}}$ e R^2_P)	34
3	OBJETIVOS.....	36
4	MATERIAL E MÉTODOS.....	37
4.1	Dados Biológicos.....	37
4.2	Construção e Minimização dos compostos	48
4.3	Simulação de dinâmica molecular (SDM).....	50
4.4	Definição do alinhamento	50
4.5	Definição do tamanho das células.....	52
4.6	Definição dos elementos de interação farmacofórica.....	53
4.7	Definição dos tipos de ocupação	53
4.8	Obtenção dos descritores de ocupação das células (GCOD).....	54
4.9	Pré-tratamento dos descritores (GCOD).....	54
4.9.1	Obtenção dos modelos de QSAR-4D	55
4.9.2	Validações	56
4.9.3	Identificação dos compostos <i>outliers</i>	56
4.9.4	Seleção da "Conformação Preferencial"	56
5	RESULTADOS E DISCUSSÃO	57
5.1	Análise da simulação por Dinâmica Molecular.....	57
5.2	Seleção do melhor alinhamento de QSAR-4D	59
5.3	Análise do Alinhamento 9.....	62
5.3.1	<i>Outliers</i> do Conjunto de Treinamento e Teste.....	64
5.3.2	Análise dos descritores.....	67
5.4	Planejamento de potenciais ativadores Glucoquinase	79
6	CONCLUSÕES	83
	REFERÊNCIAS.....	84

1 INTRODUÇÃO

O Diabetes Mellitus (DM) é uma desordem crônica resultante de altas concentrações de glicose no sangue, em razão da falta absoluta ou resistência à insulina. Essa patologia é dividida em quatro classes clínicas: diabetes tipo 1, diabetes tipo 2, diabetes gestacional e outros tipos específicos. Dentre essas classes, o diabetes tipo 2 é o caso mais prevalente (CHEN et al., 2014).

Os principais sintomas do diabetes são: sede excessiva, aumento de apetite e poliúria (urina em excesso), porém, em muitos casos, os portadores da doença permanecem assintomáticos durante anos, o que pode acarretar complicações como, insuficiência renal, anomalias cardíacas e retinopatia (ALBERTI; ZIMMET, 1998).

O Diabetes tipo 2, normalmente se desenvolve na fase adulta e está associado à obesidade, portanto a base do tratamento consiste na mudança de estilo de vida, praticando atividades físicas e realizando uma boa alimentação. Contudo, com a progressão da doença é necessário o uso de medicamentos hipoglicemiantes. Existem vários tipos de medicamentos com essa finalidade no mercado, porém nenhum deles é capaz de alcançar o controle glicêmico por longo tempo, além de alguns mostrarem efeitos colaterais como: risco de hipoglicemia, riscos cardiovasculares e infecções genitais e urinárias (PARK et al., 2014).

A enzima glucoquinase catalisa o passo inicial para a glicólise e tem papel essencial na regulação dos níveis de glicose plasmática. Portanto, ela foi escolhida como alvo para o desenvolvimento de novos medicamentos antidiabéticos. Ela possui um sítio alostérico onde sabe-se que, quando certas moléculas estão ligadas nele, a sua afinidade pela glicose é aumentada. Essas moléculas são denominadas Ativadores Glucoquinase.

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. Nesse processo, a química computacional destaca-se pelo baixo custo e menor tempo, quando comparado a 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 (Lima, 2007).

Neste trabalho, procura-se determinar quantitativamente a influência de descritores estruturais na atividade de derivados heteroaril como ativadores da proteína Glucoquinase, a partir da aplicação de estudos de QSAR-4D. Além disso, pretende-se propor novos protótipos como futuros candidatos à fármacos hipoglicemiantes orais.

2 REFERENCIAL TEÓRICO

2.1 Diabetes Mellitus

A glicose é a principal fonte de energia do nosso organismo, pois ela pode ser facilmente metabolizada para gerar adenosina trifosfato (ATP), um nucleotídeo carreador de energia. Vários tecidos dependem de glicose para a produção de ATP e, assim, desempenhar suas funções celulares, entretanto, é necessário que a glicose seja distribuída pela corrente sanguínea (Wasserman, 2009). Para que a glicose encontrada no sangue entre nas células, é necessária à atuação de um hormônio produzido pelo pâncreas denominado insulina.

A falta absoluta ou resistência à insulina resulta em altas concentrações de glicose no sangue, ou seja, a hiperglicemia que provoca o desenvolvimento do Diabetes Mellitus (DM) (CHEN et al., 2014). O DM é uma desordem crônica que afeta o metabolismo de carboidratos, proteínas e lipídeos, e pode ser considerada uma doença epidêmica de caráter mundial (PAL, 2009). A Federação Internacional de Diabetes (IDF) estima que o número de portadores da doença fosse 382 milhões em 2013, e que, em 2035, esse número chegue a alcançar 592 milhões de pessoas (INTERNATIONAL DIABETES FEDERATION, 2013).

O crescimento populacional, o envelhecimento, a urbanização, a crescente prevalência da obesidade e a falta de exercícios físicos são fatores que influenciam para o aumento do número de casos de diabetes. Segundo estimativas, esse aumento será notavelmente maior nos países em desenvolvimento (Figura 1) (WILD et al., 2004).

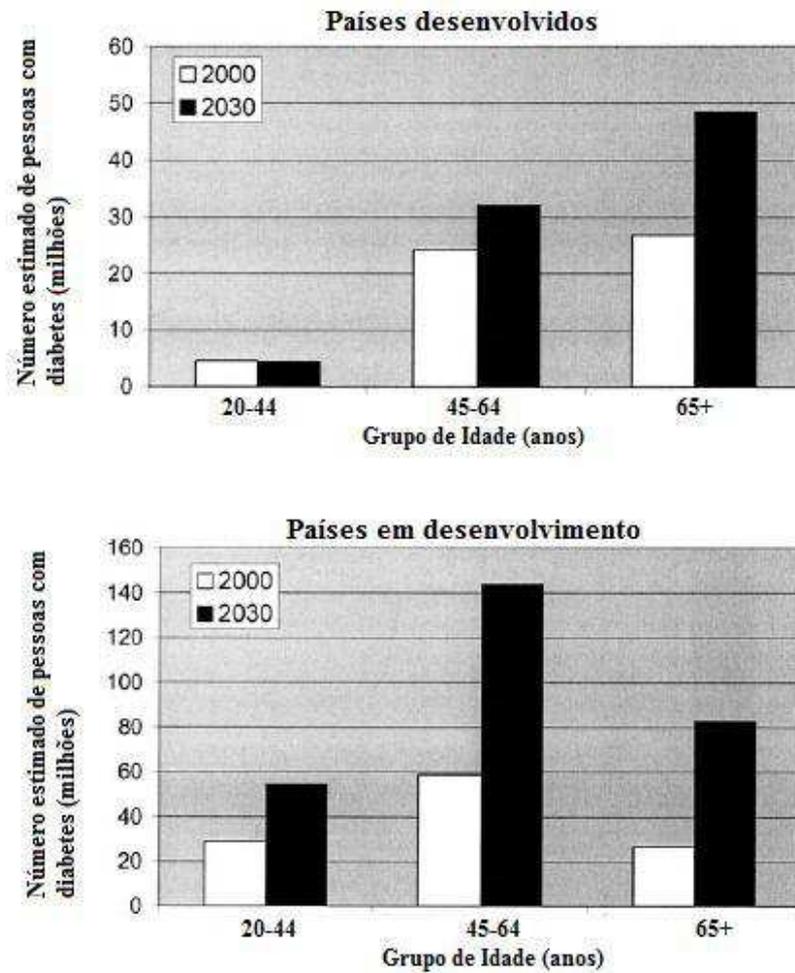


Figura 1 Número estimado de adultos com diabetes por grupo de idade, ano e categorias de países desenvolvidos e em desenvolvimento

Fonte: Adaptado Wild et al. (2004).

Atualmente, essa patologia é classificada conforme a etiologia (PASQUALOTTO; ALBERTON; FRIGERI, 2012). A classificação da Organização Mundial de Saúde (OMS) a divide em quatro classes clínicas: diabetes mellitus tipo 1 (DM1), diabetes mellitus tipo 2 (DM2), diabetes mellitus

gestacional e outros tipos específicos. O diabetes mellitus tipo 1 ou diabetes insulino-dependente resulta da destruição crônica das células β pancreáticas por mecanismos autoimunes mediados por células, como linfócitos T e macrófagos. Esse tipo de diabetes representa cerca de 5% de todos os casos de diabetes no mundo (VOLTARELLI et al., 2009). O diabetes mellitus tipo 2 ou não insulino-dependente envolve a desregulação do metabolismo de glicose e resistência à insulina. Esta prevalece em 90% a 95% dos casos de DM no mundo (PAL, 2009). O diabetes gestacional é definido como tolerância diminuída aos carboidratos, de grau e intensidade variados em gestantes. Geralmente, após o parto, ela se normaliza, contudo as mulheres que apresentaram diabetes gestacional são mais propícias a desenvolverem o diabetes tipo 2 tardiamente. Os outros tipos específicos de diabetes incluem: defeitos genéticos da célula beta e da ação da insulina, processos de doenças que danificam o pâncreas, diabetes relacionado a outras endocrinopatias e os casos decorrentes do uso de medicamentos (GROSS et al., 2002).

Os principais sintomas do diabetes são: poliúria, polidipsia, polifagia e perda de peso involuntária (ALBERTI; ZIMMET, 1998), porém, estima-se que, aproximadamente, 46% dos portadores da doença não são diagnosticados. Na Figura 2, mostra-se a comparação entre o número de casos totais em 2013 e a proporção de casos que não são diagnosticados (INTERNATIONAL DIABETES FEDERATION, 2013). Por esse motivo, a suspeita clínica ocorre a partir de fatores de risco para os diabéticos.

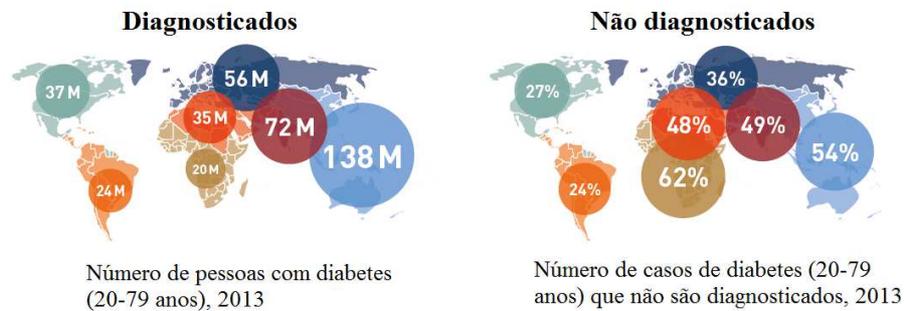


Figura 2 Comparação por regiões do número de casos de diabetes em 2013 e a proporção dos casos que não são diagnosticados

Fonte: Adaptado INTERNATIONAL DIABETES FEDERATION (2013)

Apesar de o diabetes ser classificado como uma simples doença, várias complicações podem ocorrer como, disfunção e insuficiência renal, anomalias cardíacas, retinopatia diabética, neuropatia e aterosclerose (CHEN et al., 2014).

Para o diagnóstico, além da presença dos sintomas, são feitos testes laboratoriais que medem o nível de glicose no sangue em diferentes situações, como glicemia em jejum, teste oral de tolerância à glicose (TOTG-75) e glicemia casual. A glicemia em jejum consiste em medir o nível de glicose sanguínea, após um jejum de 8 horas, se a glicemia for maior ou igual a 126mg.dL^{-1} , a pessoa é considerada diabética. No TOTG-75 a glicemia é medida em jejum e após 120 minutos da ingestão de 75g de glicose. Será considerada diabética a pessoa que apresentar glicemia maior ou igual a 200mg.dL^{-1} após a ingestão de glicose. Já, a glicemia casual pode ser realizada a qualquer hora do dia, sendo que esse teste será positivo para um nível de glicose maior ou igual a 200mg.dL^{-1} com a presença de sintomas clássicos (GROSS et al., 2002).

2.1.1 Diabetes Mellitus Tipo 2

O DM2 corresponde a 90% dos casos de diabetes e é causada pela combinação da diminuição da secreção e ação da insulina e o aumento da produção de glicose pelo fígado, frequentemente descritos como resistência à insulina (BRÄUER et al., 2005). Normalmente, essa doença desenvolve-se na idade adulta, e está associada à obesidade, pois esses fatores contribuem para a diminuição da sensibilidade à insulina nos tecidos periféricos (PASQUALOTTO; ALBERTON; FRIGERI, 2012). Assim, o sedentarismo da população é o principal fator agravante da doença, pois 80 a 90 % dos indivíduos que a apresentam são obesos (AZEVEDO; GROSS, 1990).

Os sintomas relacionados ao DM2 são semelhantes à maioria dos tipos de diabetes, a maior diferença é que, diferentemente do DM1, no qual os indivíduos perdem peso antes de ser submetido a um tratamento, no DM2, na maioria dos casos, isso não acontece (SANTOS; SILVEIRA; CAFFARO 2006), porém, indivíduos com DM2 podem permanecer assintomáticos durante anos ou décadas até que a deficiência de insulina progrida e os sintomas apareçam.

A base do tratamento consiste na mudança de estilo de vida, com prática de atividades físicas regulares e boa alimentação, porém, com a progressão da doença, torna-se necessário o uso de medicamentos hipoglicemiantes orais, para o adequado controle da glicemia (VOLTARELLI et al., 2009). Esses medicamentos têm a finalidade de baixar a glicemia e mantê-la normal, e são classificados de acordo com seu mecanismo de ação: os que aumentam a produção de insulina, aqueles que diminuem a produção e liberação de glicose pelo fígado, agentes que reduzem a resistência à insulina, aqueles que retardam a absorção de carboidratos, aqueles que aumentam a resposta das células beta em relação ao nível de glicose em circulação e promove a sensação de saciedade e aqueles que aumentam os efeitos da incretina, um hormônio que causa um

aumento na quantidade de insulina liberada a partir das células beta do pâncreas (INTERNATIONAL DIABETES FEDERATION, 2013).

Os medicamentos disponíveis para o tratamento do DM2 são: sulfonilureias, glinidas, metformina, biguanidas, glitazonas, inibidores da alfa-glicosidase, inibidores DPP-4, incretinomiméticos (análogos GLP-1), tiazolidinedionas (TZDs) e o mais novo aprovado, inibidores SGLT2. Contudo, apesar da disponibilidade e da eficácia em reduzir a glicose plasmática ou hemoglobina glicada, nenhum é capaz de alcançar o controle glicêmico por longo tempo, sendo necessária a combinação de dois ou mais medicamentos na terapia. Além disso, alguns medicamentos mostram efeitos colaterais que incluem um aumento no risco de hipoglicemia, riscos cardiovasculares, riscos pancreáticos e infecções genitais e urinárias (PARK et al., 2014). Por isso, a importância da descoberta de novos compostos antidiabéticos, com maior eficácia e menos efeitos adversos que os já existentes.

2.2 Proteína Glucoquinase

Para a identificação de um novo composto-protótipo ou uma nova entidade química que possa representar um novo candidato a fármaco, elege-se o melhor alvo-farmacológico para aquela aplicação terapêutica e, em função do nível de conhecimento estrutural disponível sobre esse receptor, identifica-se a melhor estratégia para a construção do novo ligante desejado (DUFFY et al., 2012).

A seleção de novos alvos para o desenvolvimento de medicamentos hipoglicemiantes orais para o DM2 é guiada pelo estabelecimento da homeostase da glicemia (MATSCHINSKY; PORTE JÚNIOR, 2010). Nesse sentido, a enzima glucoquinase (GQ) foi escolhida como alvo molecular para o

desenvolvimento de novos medicamentos antidiabéticos por sua função regulatória da glicose plasmática.

A GQ ou hexoquinase do tipo IV (Figura 3) é uma enzima citoplasmática pertencente à família das hexoquinases (enzimas que fosforilam a glicose e outros açúcares de seis carbonos) que catalisa o passo inicial para a glicólise, ou seja, a fosforilação da glicose para glicose-6-fosfato (Figura 4) (IRWIN. TAN, 2014). Ela é uma isoenzima composta de 465 aminoácidos (peso molecular ~50 kDa) (JAIN et al., 2013).

A enzima é expressa em tecidos com papéis essenciais na regulação dos níveis de glicose no sangue como, fígado, células β do pâncreas e em células selecionadas do intestino e cérebro. No fígado, a função da GQ é regular o metabolismo da glicose, enquanto que no pâncreas, células intestinais e neurônios funcionam como sensor de glicose. Para essa última função, ela regula a liberação de insulina em resposta aos níveis de glicose no sangue e no fígado pela síntese de glicogênio (IYNEDJIAN, 2009).

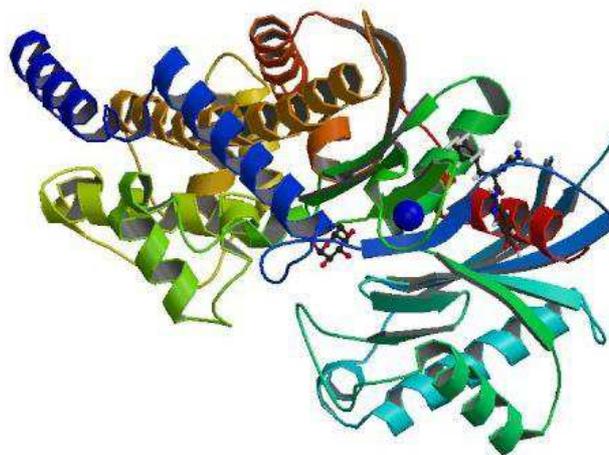


Figura 3 Estrutura tridimensional da enzima Glucoquinase Humana

Fonte: PDB

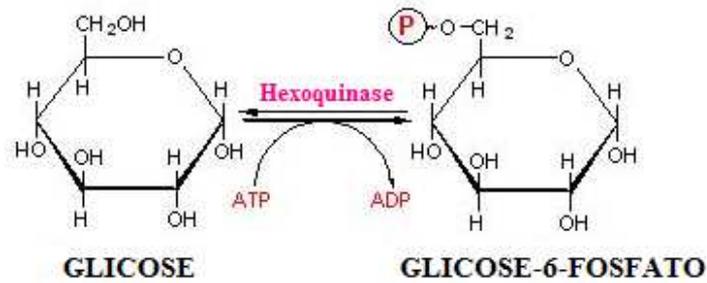


Figura 4 Reação de fosforilação da glicose
 Fonte: Adaptado Lehninger: Principles of Biochemistry.

Uma característica da glucoquinase, que permite que ela atue como sensor de glicose, é que sua atividade exibe uma dependência sigmoideal com a glicose (Figura 5), permitindo uma resposta rápida e pronunciada da variação dos níveis de glicose plasmática (KAMATA et al., 2004).

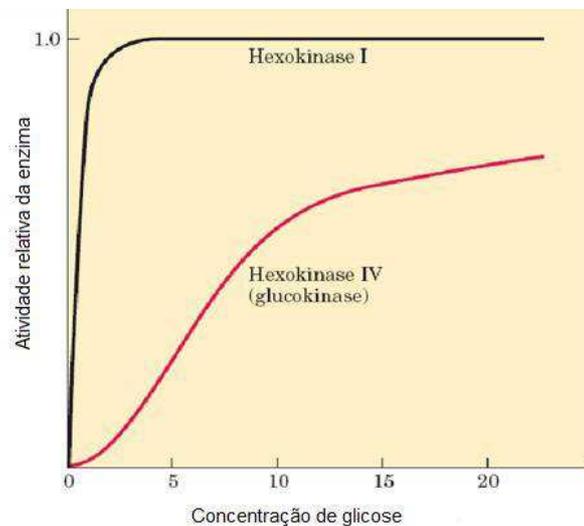


Figura 5 Comparação da atividade relativa das enzimas Hexoquinase I e IV pela concentração de glicose

Fonte: Adaptado Lehninger: Principles of Biochemistry

Além de possuir um maior peso molecular (~100kDa), as isoformas de hexoquinase I, II, III apresentam uma maior afinidade pela glicose do que a hexoquinase IV. A diferença entre os pesos moleculares é decorrente do fato da IV ser um monômero e as isoformas I, II e III serem dímeros (MATSCHINSKY et al., 2011). Além dessas características, a GQ apresenta um cooperativismo positivo (coeficiente Hill 1,7) e não é inibida pela concentração de glicose-6-fosfato (LIU et al., 2012). Porém, é regulada por uma proteína de 68 kDa denominada proteína reguladora da glucoquinase (PRGQ), que se encontra ligada à GQ dentro do núcleo, tornando-a inativa. Quando as concentrações de glicose plasmática se elevam, a PRGQ se desliga da GQ que passa à forma livre e ativa no citoplasma, após desempenhar sua função ela volta à forma inativa ligada a PRGQ (ARDEN et al., 2004). Além da concentração de glicose, as concentrações de frutose-1-fosfato e frutose-6-fosfato, no meio intracelular, também estão envolvidas no mecanismo de ativação/inativação da GQ. Concentrações aumentadas da primeira favorecem a ativação da GQ, enquanto que concentrações elevadas da segunda favorecem sua inativação, uma vez que a frutose-6-fosfato promove a associação das proteínas PRGQ e GQ e a frutose-1-fosfato promove a dissociação (VEIGA-DA-CUNHA; VAN SCHAFTINGEN, 2002).

2.3 Ativadores Glucoquinase

No intuito de desenvolver novos medicamentos antidiabéticos, nos últimos anos foram produzidas pequenas moléculas com a hipótese de ativar o sítio alostérico da enzima GQ e produzir os efeitos hipoglicemiantes. Essas moléculas são denominadas ativadores glucoquinase (AGQs) e estão associadas a um duplo mecanismo de aumento da captação de glicose no fígado e da secreção de insulina nas células β do pâncreas (MAO et al., 2012).

Foi verificado que esses compostos ligam-se a um sítio alostérico da enzima, cerca de 20Å do sítio de ligação da glicose (Kamata *et al.*, 2004), e que todos AGQs aumentam a afinidade de ligação com a glicose, no entanto eles variam quanto à velocidade máxima ($V_{m\acute{a}x}$) (CHERUVALLATH *et al.*, 2013). Desde então, há muitos trabalhos visando a esse mecanismo de ativação (JAIN *et al.*, 2013; LANG *et al.*, 2011; MAO *et al.*, 2012; QIAN *et al.*, 2013).

Vários candidatos AGQs avançaram em estudos clínicos e demonstraram reduzir, tanto os níveis de glicose em jejum quanto os níveis pós-prandial em indivíduos saudáveis e em pacientes com DM2, no entanto a hipoglicemia é um dos efeitos adversos mais preocupantes (PARK *et al.*, 2014).

Grimsby *et al.* (2003) foram os primeiros a publicarem moléculas AGQs e, desde então, várias classes de compostos como, fenilacetamida, benzamida e imidazolacetamida têm sido identificadas como potentes ativadores. Um modo de ligação de ativadores à enzima GQ foi revelado por pesquisadores em 2009 (MITSUYA *et al.*, 2009) e em decorrência das informações geradas a partir de pesquisas sobre esses complexos, é possível projetar racionalmente compostos com maior atividade (PARK *et al.*, 2014).

O planejamento racional faz parte do estudo da química medicinal computacional, 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 *et al.*, 2002).

2.4 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), os 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, tornaram possível a descoberta de inovações terapêuticas, proporcionando melhorias significativas na qualidade de vida das diversas populações no mundo. 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 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 análise quimioinformáticas (uso de técnicas computacionais aplicadas a problemas químicos) (SCHEIBER, 2012). A combinação de tecnologias biológicas e do 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, passa por posterior 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).

A eleição do alvo-terapêutico é a etapa crucial no sucesso do processo de descoberta de novos candidatos a fármacos e depende estreitamente dos conhecimentos bioquímicos sobre a fisiopatologia da doença em estudo e seus mecanismos farmacológicos (BARREIRO; FRAGA, 2005). Os alvos biológicos podem ter a sua estrutura tridimensional (3D) conhecida ou não, fato que determina a priorização de estratégias de planejamento (ANDRICOPULO; SALUM; ABRAHAM, 2009).

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; OLIVA, 2010). Dentre os estudos em química medicinal, destacam-se os métodos de QSAR (relação quantitativa estrutura atividade) que têm como objetivos relacionar a informação em conjunto de dados (estruturas químicas e atividade biológica) e criar modelos estatísticos com capacidade preditiva (LIMA, 2007).

2.5 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). 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 (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-química, 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, regressão linear múltipla ou regressão não-linear (TAVARES, 2004).

Depois da metodologia de QSAR de Hansch e Fujita (1964), várias metodologias têm sido propostas, cada uma caracterizando-se por diferentes abordagens de cálculos, descritores moleculares e algoritmos estatísticos para a construção dos modelos resultantes (SANTOS-FILHO et al., 2009). 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). Na análise de QSAR-DR, os modelos são derivados a partir da estrutura 3D e das múltiplas conformações ligante-receptor, nela estão incluídos os QSARs 4D, 5D, 6D e 7D (Rodrigues *et al.*, 2012). Essa abordagem fornece uma explícita simulação do processo de encaixe, utilizando a

estrutura do complexo receptor-ligante, em que tanto ligante como o receptor podem ser completamente flexíveis, pela utilização da simulação de dinâmica molecular (ANDRADE et al., 2010).

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 (LIMA, 2007).

2.5.1 QSAR 4D

Em 1997, Hopfinger e colaboradores propuseram uma nova metodologia de QSAR chamada de QSAR-4D. A análise em QSAR-4D incorpora liberdade conformacional ao desenvolvimento de modelos de QSAR-3D, fazendo com que a mudança de estado molecular constitua a quarta dimensão.

Nessa metodologia, uma caixa gradeada é utilizada ao redor das estruturas dos ligantes (ou complexos), para determinar as regiões no espaço tridimensional que são responsáveis por possíveis interações, incorporando flexibilidade molecular e alinhamentos múltiplos, permitindo, assim, a identificação da conformação que potencializa a atividade biológica prevista (RODRIGUES et al., 2012). Nessa abordagem, os descritores representam as frequências de ocupação dos diferentes tipos de átomos nas caixas cúbicas, durante a simulação de dinâmica molecular (SDM), de acordo com o alinhamento de cada ensaio, o que corresponde a um conjunto médio de amostragem conformacional (ANDRADE et al., 2010).

Os descritores de ocupação das células da caixa, GCODs (*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 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).

2.6 Métodos e índices estatísticos

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.

2.6.1 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) e definido pela Equação 1.

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

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 escolhido (ROGERS; HOPFINGER, 1994).

2.6.2 Coeficiente de Correlação Linear Quadrático (r^2 ou R^2)

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 (r^2) é uma medida de ajustamento de um modelo estatístico linear generalizado em relação aos valores observados. Um valor de r^2 maior que 0,7, indica que o modelo está correlacionado e pode ser utilizado para calcular a atividade do conjunto de teste. Na Equação 2, mostra-se como o coeficiente de determinação é calculado (CALDAS; RAMALHO; CUNHA, 2014).

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

2.6.3 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_{cv}^2 , (Equação 3) (SENESE et al., 2004).

$$q_{cv}^2 = 1 - \frac{\sum_{i=1}^n (y_i^{obs} - y_i^{predcv})^2}{\sum_{i=1}^n (y_i^{obs} - y_{i_{med}}^{obs})^2} \quad (3)$$

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 $y_{\text{med}}^{\text{obs}}$ 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).

Os valores de q^2 devem ser transformados em q^2_{ajustado} , demonstrado pela Equação 4, que permite a comparação entre modelos com número diferente de variáveis e a punição quando o modelo inclui variáveis irrelevantes. Modelos com valores de q^2_{ajustado} , maiores que 0,5, são considerados bons modelos preditivos (CALDAS; RAMALHO; CUNHA, 2014).

$$q^2_{\text{ajust}} = 1 - \frac{(n-1)(1-q^2)}{(n-k-1)} \quad (4)$$

Nessa equação, q^2 corresponde ao quadrado do coeficiente de correlação após a validação cruzada, n corresponde ao número de compostos do conjunto de treinamento e k corresponde à quantidade de variáveis independentes do modelo.

2.6.4 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 5. Esse processo, verificará a capacidade do modelo de prever a atividade biológica de moléculas congêneres (KIRALJ; FERREIRA, 2009).

$$R_{pred}^2 = 1 - \frac{\sum_{j=1}^k (y_j^{obs} - y_j^{pred})^2}{\sum_{j=1}^k (y_j^{obs} - y_{med}^{obs})^2} \quad (5)$$

onde y_j^{obs} e y_j^{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_{pred}^2 deve ser maior que 0,5 (PRATIMROY et al., 2009).

O valor de R_{pred}^2 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_{pred}^2 , que dá a importância para a diferença entre r^2 e r_0^2 . Essa Equação está apresentada abaixo (Equação 6):

$$r_m^2 = r^2 (1 - |\sqrt{r^2} - r_0^2|) \quad (6)$$

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.

2.6.5 Validação Y-randomization (R_{Y-rand}^2) e R_p^2

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 7):

$$R_p^2 = r^2 \sqrt{r^2 - R_{Y\text{-rand}}^2} \quad (7)$$

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 determinar quantitativamente a influência de descritores estruturais na atividade de derivados heteroaril como ativadores da enzima GQ, a partir da aplicação de estudo de QSAR-4D-DR e propor novos protótipos como futuros candidatos à fármacos hipoglicemiantes orais.

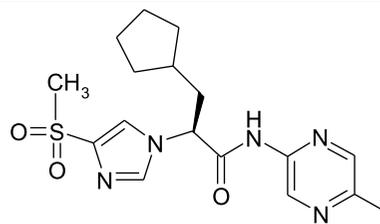
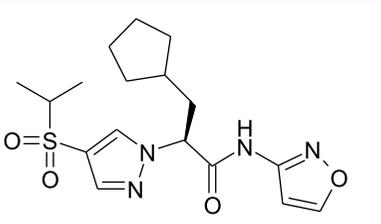
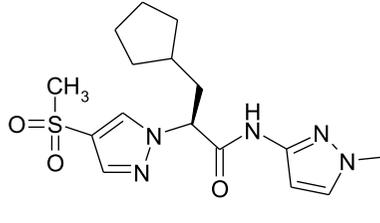
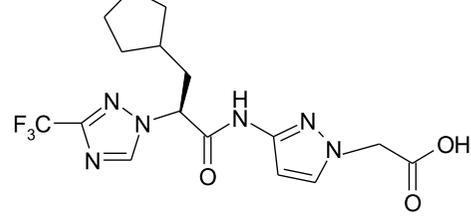
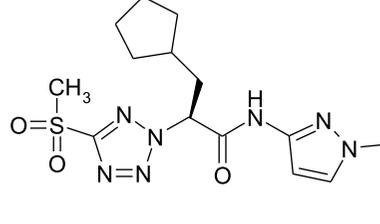
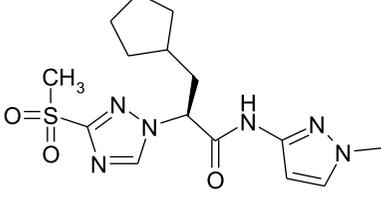
4 MATERIAL E MÉTODOS

Este estudo realizou-se 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.

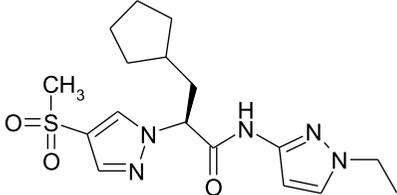
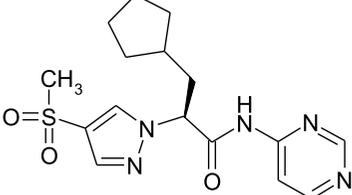
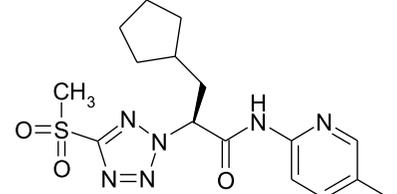
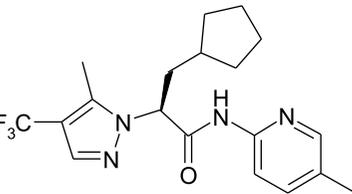
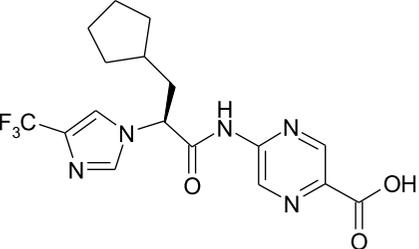
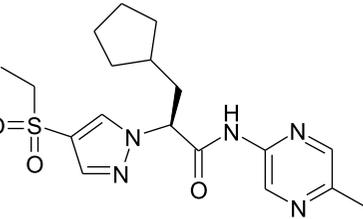
4.1 Dados Biológicos

Na Tabela 1, está apresentada uma série de análogos heteroaril, ativadores GQ, que foram estudados. Esses compostos foram sintetizados e avaliados farmacologicamente por Benbow e colaboradores (BENBOW et al., 2013). 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 EC_{50} expressos em μM foram transformados em pEC_{50} ($-\log EC_{50}$). EC_{50} representa a concentração do composto que induz metade do efeito máximo Na Tabela 1, relacionam-se os AGQs com seus respectivos valores de pEC_{50} . Os compostos apresentados sublinhados, na tabela, são aqueles que fazem parte do grupo teste.

Tabela 1 Estrutura dos 54 ativadores glucoquinase com os respectivos valores de pEC₅₀

No.	Ativador	pEC ₅₀	No.	Ativador	pEC ₅₀
1		4,21	2		4,22
3		4,26	4		4,35
5		4,40	6		4,48

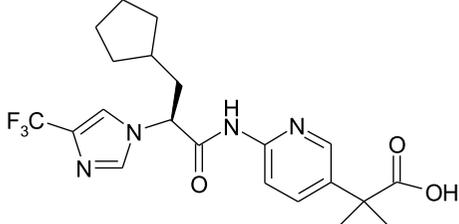
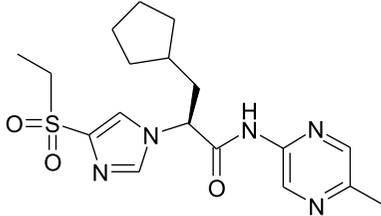
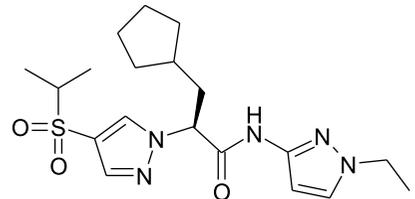
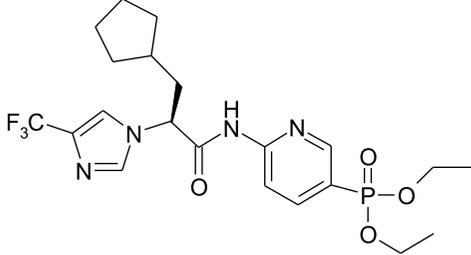
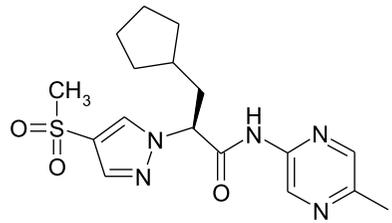
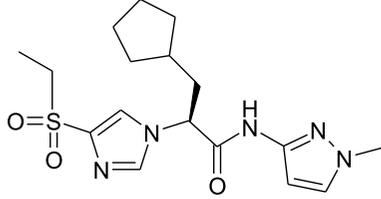
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
7		4,49	8		4,49
9		4,50	10		4,55
11		4,55	12		4,60

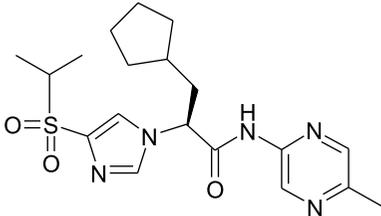
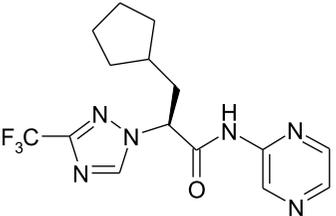
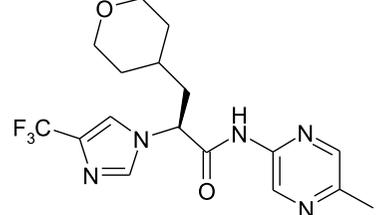
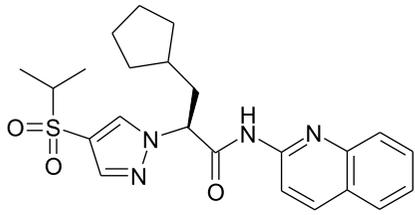
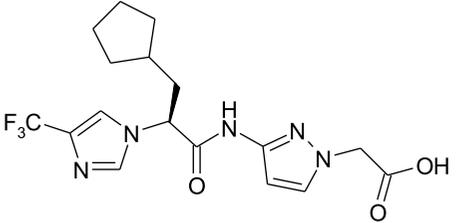
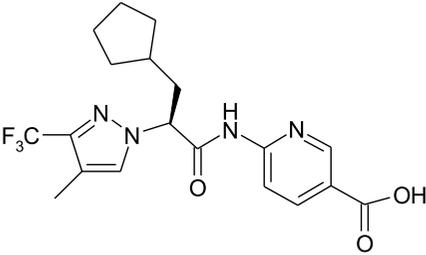
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
<u>13</u>		4,62	<u>14</u>		4,64
<u>15</u>		4,66	<u>16</u>		4,74
<u>17</u>		4,74	<u>18</u>		4,77

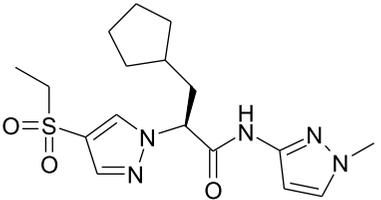
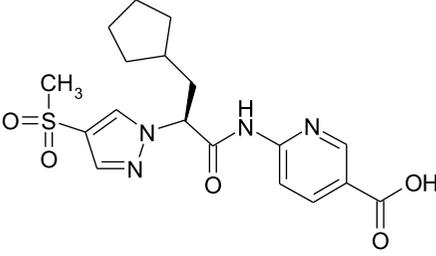
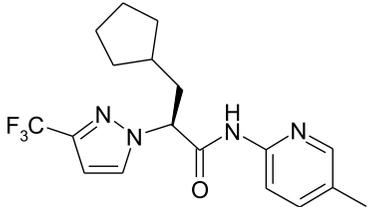
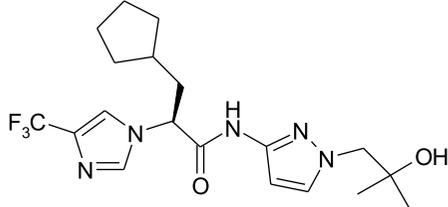
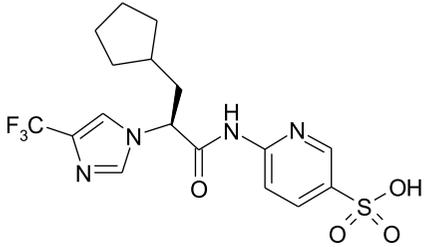
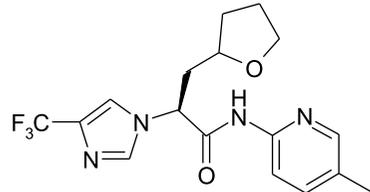
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
19		4,77	20		4,80
21		4,80	22		4,85
23		4,96	24		4,96

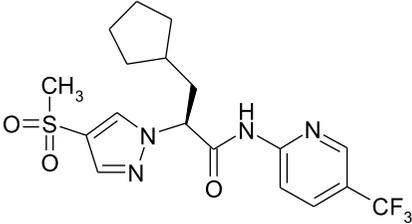
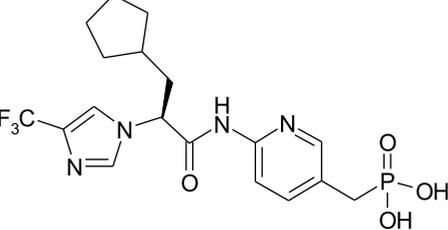
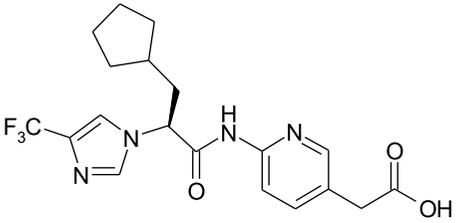
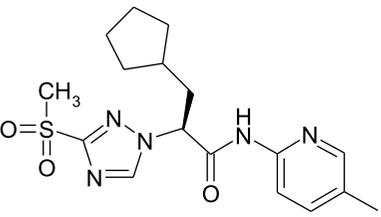
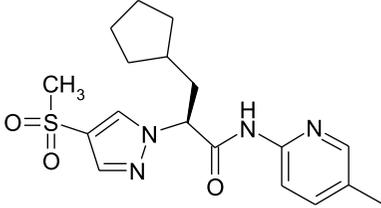
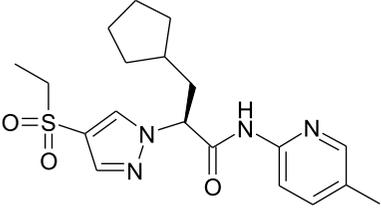
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
25		5,00	26		5,00
27		5,06	28		5,09
29		5,09	30		5,10

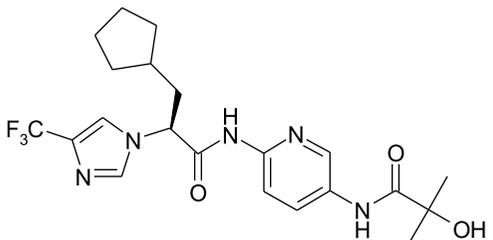
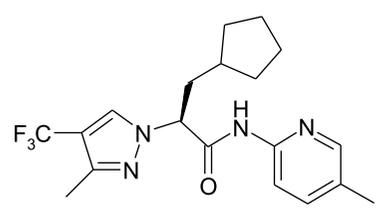
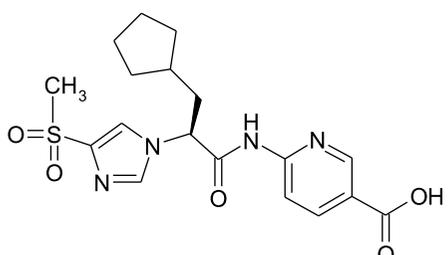
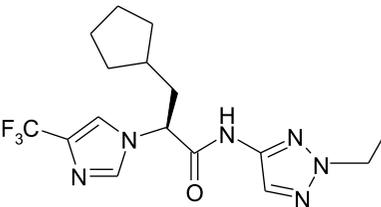
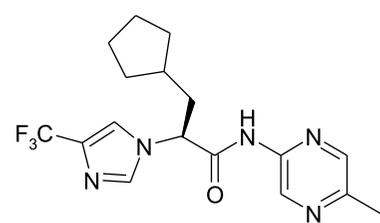
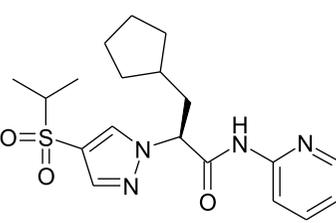
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
31		5,14	32		5,15
33		5,26	34		5,28
35		5,29	36		5,30

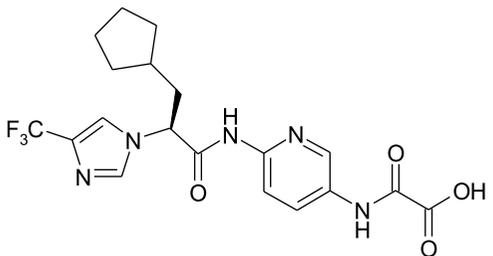
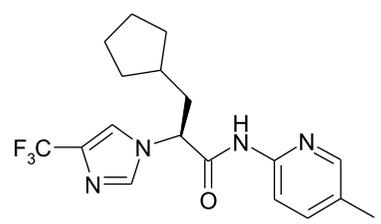
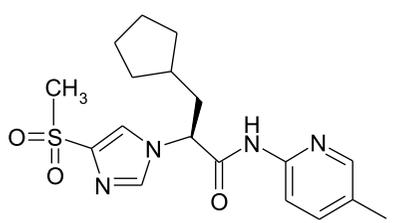
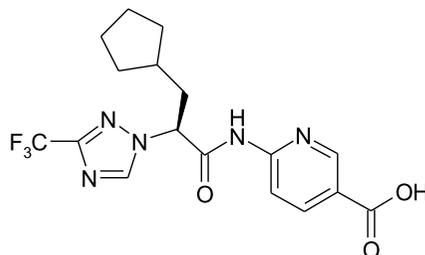
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
37		5,31	38		5,34
39		5,38	<u>40</u>		5,41
41		5,42	42		5,54

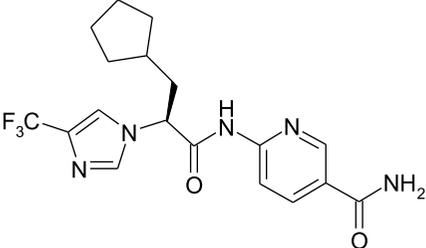
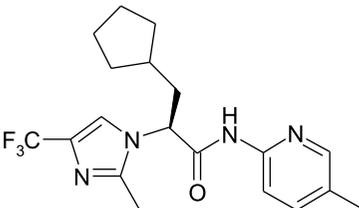
“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
43		5,55	44		5,55
45		5,64	46		5,66
47		5,70	48		5,72

“Tabela 1, continuação”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
49		5,80	50		5,96
51		6,10	52		6,15

“Tabela 1, conclusão”

No.	Ativador	pEC_{50}	No.	Ativador	pEC_{50}
53		6,40	54		6,40

Dados obtidos a partir dos estudos de Benbow et al. (2013).

A faixa dos valores de pEC_{50} dos 54 compostos abrange três ordens de magnitude (4,21 a 6,40) 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 pEC_{50} dos 54 compostos estudados.

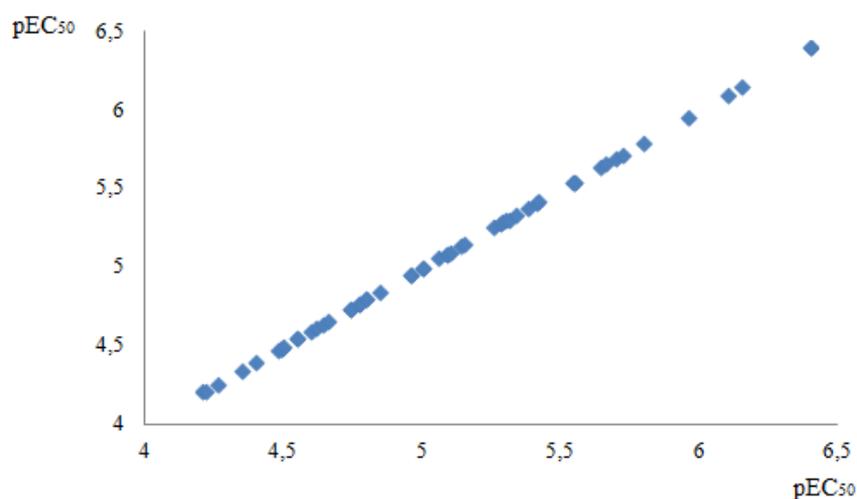


Figura 6 Gráfico de distribuição dos valores de atividade biológica (pEC_{50}) dos 54 compostos estudados

4.2 Construção e Minimização dos compostos

A estrutura tridimensional (3D), obtida por difração de raio-X da enzima GQ, cristalizada com a glicose, e um ativador (ácido-6-((2S)-3-ciclopentil-2-[4-(trifluormetil)-1H-imidazol-1-il-]propanoil}amino)-3-piridina), disponível no Protein Data Bank (PDB) (Berman *et al.*, 2000) sob o código 3VF6 com resolução de 1,86 Å (LIU *et al.*, 2012), foi utilizada como modelo para construir as estruturas 3D de todos os compostos, utilizando o programa *Accelrys*

Discovery Studio 2.1 (ACCELRYS SOFTWARE INCORPORATED, 2007). Essa estrutura foi escolhida, tendo em vista a similaridade do ativador cristalizado e os compostos que foram estudados. A enzima GQ consiste em 465 átomos, assim, a demanda computacional exigida pela simulação de dinâmica molecular (processo posteriormente explicado) para todos os 54 complexos proteína-ligante seria muito grande. Então, a técnica baseada no corte da enzima em um raio de cerca de 5 Å ao redor do ligante foi empregada. À estrutura obtida foram adicionados os átomos de hidrogênio correspondentes, a fim de manter a integridade geométrica do receptor. Em seguida, a energia de cada um dos complexos enzima-ligante foi minimizada (maior estabilidade) no mesmo programa citado anteriormente.

Em química medicinal, considera-se que dentre as conformações mais estáveis (ou de menores energias) da molécula isolada deve estar à bioativa (aquela que se liga ao sítio de ação), assim como as conformações da mesma, quando em solução e em sua forma cristalina.

Os modelos de QSAR-4D foram construídos, utilizando um conjunto de 42 compostos denominados de conjunto de treinamento, e validados externamente, usando um conjunto de 12 compostos, denominados de conjunto teste (sublinhado na Tabela 1).

A divisão entre compostos do conjunto de treinamento e do conjunto teste foi feita separando-se os 54 compostos, estruturalmente relacionados, em três subgrupos, de acordo com faixas de valores de atividade biológica e selecionando, aleatoriamente, 25% 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 pEC_{50} de 4,21 a 4,96; o grupo II compreende valores de pEC_{50} de 5,00 a 5,96 e o grupo III compreende os compostos com pEC_{50} de 6,10 a 6,40.

4.3 Simulação de dinâmica molecular (SDM)

As estruturas dos compostos otimizados no sítio ativo da GQ foram submetidas ao processo de SDM com o objetivo de gerar um perfil de amostragem conformacional (CEP, do inglês *Conformational Ensemble Profile*) para cada composto, permitindo, assim, investigar a flexibilidade conformacional do ligante (HOPFINGER et al., 1997).

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 (1984), incorporado ao programa 4D- QSAR. 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 5000 conformações para cada composto. Os cálculos de SDM foram executados, aplicando uma constante dielétrica dependente da distância de $3r$, no intuito de tentar modelar o efeito do solvente. Adicionalmente, todos os resíduos de aminoácidos selecionados foram fixados para prevenir uma grande mudança conformacional dos complexos, o que poderia resultar em conformações estericamente proibidas.

4.4 Definição do alinhamento

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). Como os

compostos estudados são congêneres, podemos assumir que eles irão apresentar o mesmo mecanismo de ativação na GQ.

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 bom alinhamento.

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).

Para os doze alinhamentos realizados, foram escolhidos átomos do carbono-alfa (C_{α}) da proteína (Tabela 2), já que estes são comuns a todos os complexos e abrangem a estrutura dos compostos estudados.

Tabela 2 Alinhamentos testados e os átomos de C_α da proteína selecionados para o alinhamento dos complexos

Alinhamento	C _α
1	Tyr61/Tyr214/Val455
2	Tyr214/Val455/ Tyr61
3	Val455/ Tyr61/ Tyr214
4	Thr65/Ile211/Met235
5	Ile211/Met235/Thr65
6	Met235/Thr65/ Ile211
7	Arg63/Met235/Val455
8	Met235/Val455/ Arg63
9	Val455/ Arg63/ Met235
10	Arg63/Tyr214/Val455
11	Tyr214/Val455/ Arg63
12	Val455/ Arg63/ Tyr214

4.5 Definição do tamanho das células

Um perfil de amostragem conformacional (CEP, do inglês *Conformational Ensemble Profile*), para cada composto, foi obtido, por meio da SDM e, em seguida, esse conjunto de distintas conformações foi inserido dentro de uma caixa tridimensional virtual. A caixa é composta por células cúbicas que podem ter dimensões variadas. 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). Neste trabalho, foram utilizadas células de 2,0 Å de aresta.

4.6 Definição dos elementos de interação farmacofórica

Na metodologia de QSAR-4D, pode-se definir sete tipos de elementos de interação farmacofórica (IPE, do inglês *Interaction Pharmacophore Elements*), os quais correspondem aos tipos de átomos que podem realizar interações específicas (HOPFINGER et al., 1997). Na Tabela 3, relacionam-se os seis tipos de IPEs utilizados neste trabalho.

Tabela 3 Definição dos elementos de interação farmacofórica, IPEs

IPEs	Definição	Código
1	interação não polar	np
2	interação polar positiva	p+
3	interação polar negativa	p-
4	acceptor de ligação hidrogênio	alh
5	doador de ligação hidrogênio	dlh
6	sistema aromático	ar

4.7 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. Neste estudo, será utilizada a ocupação absoluta, pois não havia um composto de referência.

4.8 Obtenção dos descritores de ocupação das células (GCOD)

Para cada alinhamento testado, um complexo por vez é posicionado automaticamente no espaço de referência da caixa, de acordo com a sequência do alinhamento anteriormente definida e são computados os valores de ocupação da caixa (*Grid Cell Occupancy Descriptors*, GCOD) referentes a cada elemento de interação farmacofórica (IPE). Esses valores são normalizados de acordo com o número de conformações selecionadas, isto é, transformados em valores percentuais (HOPFINGER et al., 1997). Os GCODs correspondem às variáveis independentes na metodologia QSAR-4D.

4.9 Pré-tratamento dos descritores (GCOD)

A análise de QSAR-4D gera um enorme 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 realizados dois tipos de pré-tratamento do banco de dados, antes da etapa de geração dos modelos (equações). No primeiro, empregamos um critério para a exclusão de variáveis, no qual os GCODs que apresentaram valores de ocupação diferentes de zero em dez compostos ou menos, foram eliminados. E no segundo, os GCODs que

apresentaram valores de ocupação com variância menor que 0,001 foram excluídos, descartando, assim, GCODs pouco representativos em função do número de compostos.

4.9.1 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 uma metodologia que combina algoritmo de aproximação da função genética (GFA, do inglês *Genetic Function Approximation*) e mínimos quadrados parciais (PLS, do inglês *Partial Least Squares*), (ROGERS; HOPFINGER, 1994). A grande vantagem desse método é a construção e otimização de múltiplos modelos, ao invés da otimização de apenas um único modelo (ROGERS; HOPFINGER, 1994).

Os cálculos de GFA 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, oito termos. Equações com mais de oito 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, 1993).

Os melhores modelos resultantes de cada alinhamento testado no estudo de QSAR-4D foram ordenados pelo valor do fator de desajuste (LOF, do inglês *lack-of-fit*) de Friedman.

4.9.2 Validações

Os parâmetros de validação calculados foram: LOF, r^2 , q^2 , q^2_{ajust} , R^2_{pred} , r^2_m , $R^2_{Y\text{-rand}}$ 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 (sublinhado na Tabela 1).

4.9.3 Identificação dos compostos *outliers*

Os compostos designados como *outliers*, são compostos que não foram bem preditos pela equação ou modelo. Eles são identificados pela observação dos valores residuais calculados para os compostos do conjunto de treinamento e do conjunto teste. Neste trabalho, eles foram definidos como aqueles cujos valores modulares da diferença entre pEC_{50} experimental e o pEC_{50} predito são superiores ao dobro do desvio padrão desses resíduos (HOPFINGER et al., 1997).

4.9.4 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). A conformação foi selecionada dentro de uma faixa de 5 kcal.mol^{-1} da conformação de menor energia, que maximiza a atividade biológica (HOPFINGER et al., 1997).

5 RESULTADOS E DISCUSSÃO

5.1 Análise da simulação por Dinâmica Molecular

No estudo de QSAR-4D, a SDM foi utilizada para avaliar o espaço conformacional de cada um dos complexos, ligante-proteína, resultantes dos processos de otimização de geometria; para isso, todos os átomos da proteína foram mantidos fixos, visando à manutenção da integridade da estrutura molecular.

Na Figura 7, representa-se o corte realizado na enzima Glucoquinase (código 3VF6) a 5Å de distância do ligante para todos os 54 compostos.

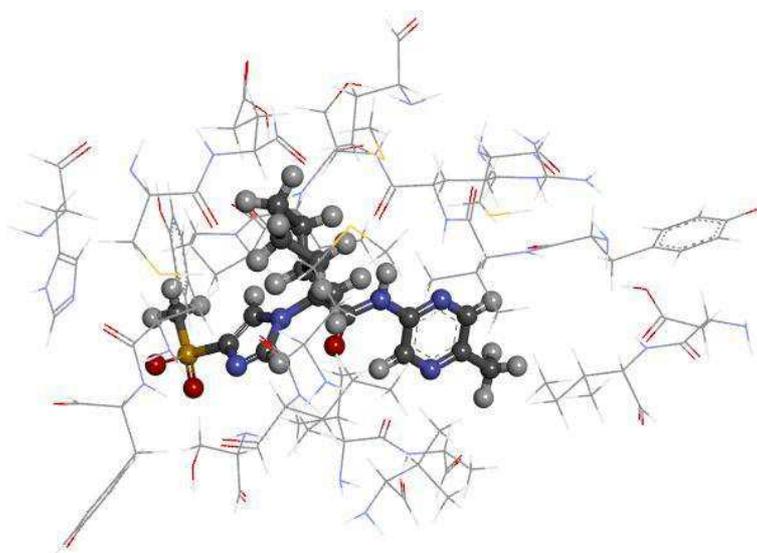


Figura 7 Imagem ilustrativa do corte realizado juntamente com o composto 1 em estudo

Na Figura 8, representa-se o gráfico de variação de energia total, em função do número de passos utilizados na dinâmica molecular referente ao

composto 1. Nessa figura, podemos observar que a energia total se mantém estável durante quase toda simulação, e que, como já era esperado, a maior energia encontrada foi no início da simulação, decorrente da força aplicada.

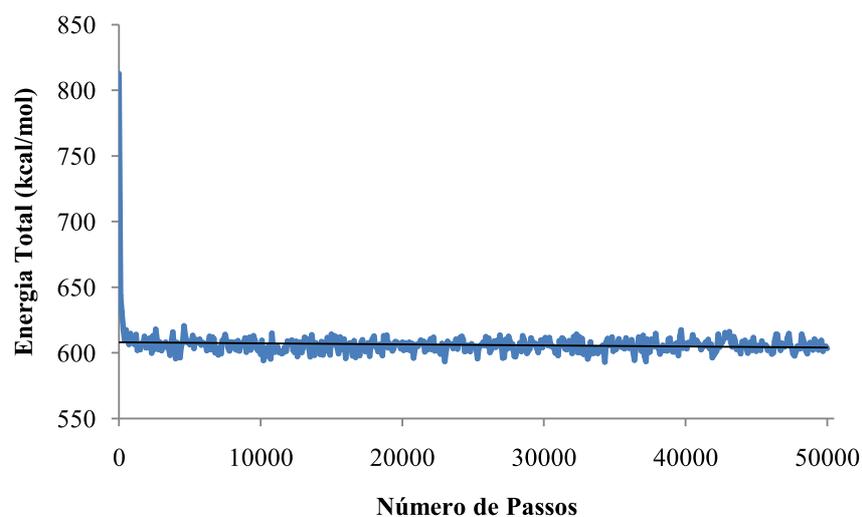


Figura 8 Valores de energia total para SDM de 50 ps para os complexos ligante-proteína (1)

Esta etapa teve por finalidade gerar um perfil de amostragem conformacional, produzindo 5000 conformações para cada composto. Seis tipos de átomos foram utilizados para definir os elementos de interação farmacofórica (IPEs): np (1), p+ (2), p- (3), alh (4), dlh (5) e ar (6). As conformações de cada um dos complexos foram alinhadas de doze formas diferentes como apresentado na Tabela 2.

5.2 Seleção do melhor alinhamento de QSAR-4D

A possibilidade da análise de múltiplos alinhamentos a partir de uma mesma amostragem conformacional obtida por SDM é encontrada na metodologia de QSAR-4D. Assim, analisaram-se doze alinhamentos testados em função do número de termos (variáveis independentes) dos modelos (equações) obtidos, usando como parâmetros os valores de Q^2_{ajustado} e R^2_{pred} que refletem a capacidade preditiva do modelo. O valor de Q^2_{ajustado} , (coeficiente de correlação linear quadrático da validação cruzada normalizado em função do número de variáveis) é considerado o índice mais adequado para comparar modelos de QSAR que contêm números de termos diferentes, desde que o número de observações (N) seja igual (neste estudo N= 42, que corresponde aos 42 complexos ligante-enzima utilizados no conjunto de treinamento). Com um conjunto de treinamento de 42 compostos, o número máximo de termos a ser obtido em um modelo ou equação é oito. Então, selecionamos modelos contendo sete e oito termos, com os maiores valores de Q^2 , para cada alinhamento.

O valor de R^2_{pred} é correspondente à validação externa e verificará a capacidade do modelo de prever a atividade biológica de moléculas semelhantes.

Os modelos de QSAR-4D foram obtidos utilizando uma metodologia que combina algoritmos genéticos (GA) e mínimos quadrados parciais (PLS), implementados no programa 4D-QSAR.

Na Tabela 4, são apresentados os valores de R^2 , $RMSE_C$, Q^2 , $RMSE_{CV}$, Q^2_{ajustado} , número de GCOD no modelo, Q^2_{externo} e Rm^2_{teste} para os doze alinhamentos testados.

Tabela 4 Resumo dos resultados dos melhores modelos de QSAR-4D obtidos para os doze alinhamentos

Alinhamento	R²	RMSE_C	q²_{cv}	RMSE_{Cv}	q²_{ajustado}	GCOD	R²_{pred}	r_m²_{teste}	R²_{Y-rand}	R_p²
<i>1</i>	0,761	0,248	0,622	0,330	0,516	8	0,353	0,293	0,102	0,618
<i>2</i>	0,735	0,278	0,537	0,361	0,446	8	-0,014	0,012	0,298	0,486
<i>3</i>	0,747	0,359	0,633	0,326	0,525	8	0,430	0,347	0,181	0,562
<i>4</i>	0,794	0,245	0,694	0,298	0,576	8	0,218	0,204	0,276	0,571
<i>5</i>	0,752	0,269	0,665	0,312	0,552	8	-0,198	-0,067	0,096	0,609
<i>6</i>	0,771	0,258	0,482	0,380	0,412	7	-0,085	-0,048	0,422	0,455
<i>7</i>	0,739	0,276	0,620	0,331	0,515	8	0,567	0,513	0,219	0,533
<i>8</i>	0,733	0,279	0,596	0,341	0,495	8	0,054	0,049	0,146	0,561
<i>9</i>	0,738	0,276	0,610	0,336	0,521	7	0,58	0,500	0,292	0,500
<i>10</i>	0,768	0,259	0,636	0,325	0,543	7	-0,169	-0,064	0,053	0,649
<i>11</i>	0,779	0,254	0,636	0,325	0,528	8	0,119	0,117	0,102	0,640
<i>12</i>	0,759	0,268	0,606	0,338	0,503	8	0,62	0,496	0,122	0,606

Observando a tabela acima, percebemos que todos os alinhamentos apresentam um coeficiente de determinação (R^2) maior que 0,7 e um valor de Q^2 (parâmetro calculado na validação leave-one-out) maior que 0,5, que são valores de parâmetros aceitáveis para um bom modelo de QSAR. Porém, quando analisamos os resultados de Q^2_{ajustado} , percebemos que os alinhamentos 2, 6 e 8 apresentam um valor para esse parâmetro menor que 0,5 que é valor mínimo aceitável para um bom modelo de QSAR, portanto não devem ser utilizados. Outro parâmetro utilizado na validação dos modelos foi o R^2_{pred} e, de acordo com os resultados obtidos nesta validação, somente os alinhamentos 7, 9 e 12 apresentam valores acima do mínimo aceitável (0,5) e, portanto, os outros alinhamentos não podem ser utilizados.

Como o valor de R^2_{pred} é altamente dependente da média do grupo de treinamento, fizemos os cálculos dos valores de $r_m^2_{\text{teste}}$, que é uma validação adicional para o grupo teste. Nessa validação, somente os alinhamentos 7 e 9 apresentaram resultados satisfatórios e, portanto, somente esses dois alinhamentos foram considerados.

Os resultados de $R^2_{Y\text{-rand}}$, foram satisfatórios para todos os alinhamentos pois, todos apresentaram valores menores que os respectivos R^2 . Já, para o parâmetro R^2_p , somente os valores encontrados para os alinhamentos 2 e 6, não foram satisfatórios.

Como os alinhamentos 7 e 9 apresentaram resultados satisfatórios em todas as validações realizadas, o método utilizado para verificar qual modelo seria melhor foi a análise dos valores de Q^2_{ajustado} . Para o alinhamento 7, o valor encontrado foi 0,515 e o modelo apresenta 8 descritores, já, para o alinhamento 9 o valor encontrado foi 0,521 e o modelo apresenta 7 descritores. Portanto, escolhemos o alinhamento 9 para uma análise mais detalhada.

5.3 Análise do Alinhamento 9

Os dez modelos provenientes do alinhamento 9 apresentam boa capacidade preditiva ($Q^2 = 0,509 - 0,610$) e o coeficientes de determinação R^2 entre 0,674 e 0,743 (Tabela 5).

Tabela 5 Resumo dos resultados dos modelos de QSAR-4D obtidos para o Alinhamento 9. LSE (Least Squares Error) é o erro dos mínimos quadrados e LOF (lack-of-fit) é o valor de desajuste de Friedman.

Modelos	1	2	3	4	5	6	7	8	9	10
R^2	0,738	0,743	0,686	0,707	0,705	0,675	0,674	0,701	0,700	0,700
q^2_{CV}	0,610	0,589	0,570	0,571	0,548	0,560	0,557	0,509	0,526	0,543
GCOD	7	8	6	7	7	6	6	7	7	7
LSE	0,076	0,075	0,091	0,085	0,086	0,095	0,095	0,087	0,087	0,087
LOF	0,136	0,147	0,148	0,152	0,153	0,154	0,154	0,155	0,155	0,155

Na Tabela acima, observa-se que, dentre todos os modelos gerados neste alinhamento, somente os modelos 3, 6 e 7 não apresentaram resultados satisfatórios, ou seja, eles possuem um valor de R^2 menor que 0,7 e, portanto, não são considerados bons modelos. Dentre os modelos que possuem bons parâmetros de validação, o modelo 1 foi escolhido como o melhor modelo, pois, apesar do modelo 2 apresentar o maior valor de R^2 , ele não apresenta o maior valor de q^2_{CV} e nem o menor valor de LOF.

A fim de verificar se o Modelo 1 apresenta o mesmo tipo de informação estrutura-atividade, foi calculada a matriz de correlação cruzada entre os valores de ocupação das células (GCODs) (Tabela 6).

Tabela 6 Matriz de correlação cruzada dos GCODs do melhor modelo do alinhamento 9

GCOD	3,3,1(p-)	-1,-1,4(ar)	0,-2,3(np)	1,3,3(np)	-1,-2,3(alh)	1,2,2(np)	-2,-1,2(np)
3,3,1 (p-)	1,00						
-1,-1,4(ar)	0,06	1,00					
0,-2,3(np)	0,12	-0,07	1,00				
1,3,3(np)	0,09	-0,19	0,09	1,00			
-1,-2,3(alh)	0,28	0,22	-0,10	-0,04	1,00		
1,2,2(np)	-0,02	-0,13	0,12	-0,22	0,02	1,00	
-2,-1,2(np)	0,11	-0,14	0,12	-0,12	0,19	-0,01	1,00

A partir da tabela acima, pode-se verificar que nenhum par de descritor do modelo selecionado está correlacionado, pois apresentam $r < 0,5$. 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 DR (KUBINYI, 1993).

5.3.1 *Outliers* do Conjunto de Treinamento e Teste

Na Tabela 7, estão relacionados os valores de atividade biológica experimentais ($pEC_{50\text{experimental}}$), os valores de atividade biológica calculados ($pEC_{50\text{predito}}$), utilizando a equação do modelo escolhido no alinhamento 9, e os respectivos valores residuais.

Na Figura 9, apresentam-se os valores de pEC_{50} experimentais versus pEC_{50} preditos para os compostos do conjunto de treinamento.

Neste trabalho, foram considerados como *outliers* os compostos com valor residual duas vezes superior ao valor de desvio padrão dos resíduos ($pEC_{50\text{Exp}} - pEC_{50\text{Pred}}$). O desvio padrão (DP), para o conjunto de treinamento, foi de 0,279, então, os compostos com valores residuais maiores que 0,558 são considerados *outliers*. Portanto, o modelo tem apenas um composto do conjunto de treinamento considerado um *outlier*, o composto **46** (sublinhado na tabela).

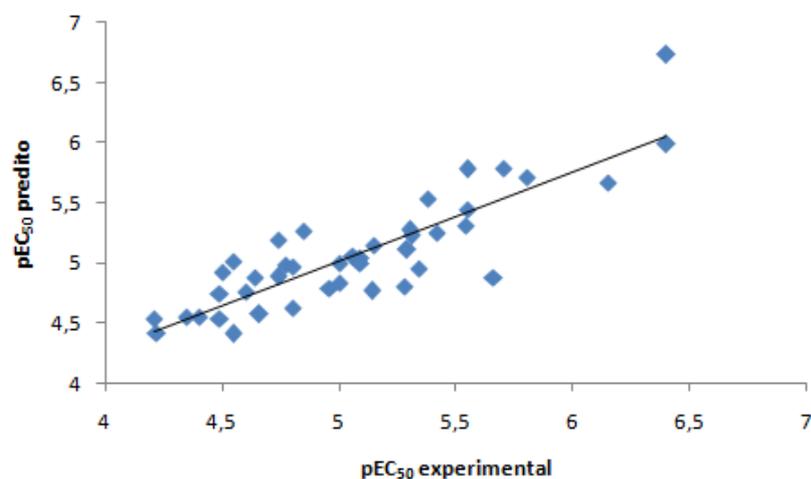


Figura 9 Valores de pEC_{50} experimentais versus preditos do conjunto de treinamento, usando o melhor modelo de QSAR-4D-DR do alinhamento 9

Tabela 7 Valores de pEC_{50} experimentais e preditos e valores residuais ($pEC_{50predito} - pEC_{50experimental}$) para os compostos do conjunto de treinamento

Composto	pIC_{50} Experimental	pIC_{50} Predito	Valores residuais
1	4,21	4,54	0,33
2	4,22	4,42	0,20
4	4,35	4,55	0,20
5	4,40	4,55	0,15
7	4,49	4,54	0,05
8	4,49	4,74	0,25
9	4,50	4,92	0,42
10	4,55	4,41	-0,14
11	4,55	5,02	0,47
12	4,60	4,76	0,16
14	4,64	4,88	0,24
15	4,66	4,57	-0,09
16	4,74	5,19	0,45
17	4,74	4,88	0,14
19	4,77	4,98	0,21
20	4,80	4,96	0,16
21	4,80	4,63	-0,17

“Tabela 7, conclusão”

Composto	pIC ₅₀ Experimental	pIC ₅₀ Predito	Valores residuais
22	4,85	5,27	0,42
24	4,96	4,79	-0,17
25	5,00	5,00	0,00
26	5,00	4,84	-0,16
27	5,06	5,06	0,00
28	5,09	5,05	-0,04
29	5,09	4,99	-0,10
31	5,14	4,77	-0,37
32	5,15	5,15	0,00
34	5,28	4,81	-0,47
35	5,29	5,11	-0,18
36	5,30	5,28	-0,02
37	5,31	5,23	-0,08
38	5,34	4,95	-0,39
39	5,38	5,53	0,15
41	5,42	5,25	-0,17
42	5,54	5,31	-0,23
43	5,55	5,78	0,23
44	5,55	5,45	-0,10
46	5,66	4,88	-0,78
47	5,70	5,79	0,09
49	5,80	5,71	-0,09
52	6,15	5,67	-0,48
53	6,40	5,99	-0,41
54	6,40	6,73	0,33

A molécula **46**, considerada como *outlier* neste modelo, é a única que possui um substituinte triazol, no grupo acetamida da molécula. Portanto, o modelo não foi capaz de prever sua atividade por não existir células da caixa “treinada” para descrever esse substituinte nessa posição.

O conjunto de teste, constituído por 12 compostos, não participa da construção dos modelos e, assim, o cálculo da atividade predita para esses compostos, utilizando o Modelo 1, corresponde a um processo de validação externa.

Os compostos *outliers* do grupo de teste foram calculados da mesma forma: os compostos com valor residual duas vezes superior ao valor do DP. O

valor do DP encontrado para o grupo de teste foi de 0,39, portanto compostos com valores residuais maiores que 0,78 são considerados *outliers*. Como nenhuma molécula apresentou um valor de resíduo superior a este, considera-se que o modelo foi capaz de prever bem os valores de atividade e não apresentou nenhum composto do conjunto de teste como *outlier*. Na Tabela 8, apresentam-se os valores de atividade biológica experimentais ($pEC_{50\text{experimental}}$), os valores de atividade biológica preditos ($pEC_{50\text{predito}}$) e os respectivos valores residuais para os compostos do conjunto de teste.

Tabela 8 Valores de pEC_{50} experimentais e preditos e valores residuais ($pEC_{50\text{predito}} - pEC_{50\text{experimental}}$) para os compostos do conjunto de teste

Composto	pIC_{50} Experimental	pIC_{50} Predito	Valores residuais
3	4,26	4,98	0,72
6	4,48	4,55	0,07
13	4,62	4,73	0,11
18	4,77	4,88	0,11
23	4,96	4,73	-0,23
30	5,10	5,59	0,49
33	5,26	5,54	0,28
40	5,41	5,20	-0,21
45	5,64	5,54	-0,10
48	5,72	5,13	-0,59
50	5,96	5,95	-0,01
51	6,10	5,49	-0,61

5.3.2 Análise dos descritores

O melhor modelo do alinhamento 9 (Modelo 1) gerou 7 descritores (GCODs), sendo que cinco apresentaram coeficiente positivo, (3,3,1,p-), (-1,-1,-4,ar), (1,3,3,np), (1,2,2,np), (-2,-1,2,np), e correspondem a interações favoráveis entre os substituintes dos compostos e os resíduos de aminoácidos do sítio alostérico da GQ. Assim, substituintes com as características selecionadas (IPes) nessas posições aumentam a potência dos compostos. Os GCODs (0,-2,3,np) e (-

1,-2,3,alh) apresentaram coeficientes negativos, correspondendo às interações desfavoráveis entre os substituintes dos compostos e os resíduos de aminoácidos da GQ.

Na Equação 8, apresenta-se o Modelo 1 do alinhamento 9. Os números entre parênteses indicam as coordenadas cartesianas x, y e z da célula da caixa e as letras, entre parênteses, indicam o tipo de IPE dos átomos que ocupam a célula.

$$\begin{aligned}
 pEC_{50} = & 2,96 + 1,53 (3,3,1, p -) + 1,16 (-1, -1,4, ar) - 0,39 (0, -2,3, np) \\
 & + 1,05 (1,3,3, np) - 0,98 (-1, -2,3, alh) + \\
 & 0,37 (1,2,2, np) + 0,57 (-2, -1,2, np) \quad (8)
 \end{aligned}$$

$$n = 42; r^2 = 0,738; q^2 = 0,610; q_{ajust}^2 = 0,521; LSE 0,56; LOF = 1,07$$

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 azuis 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 10, ilustra-se o composto menos ativo (molécula 1) e os sete descritores obtidos no Modelo1 do alinhamento 9.

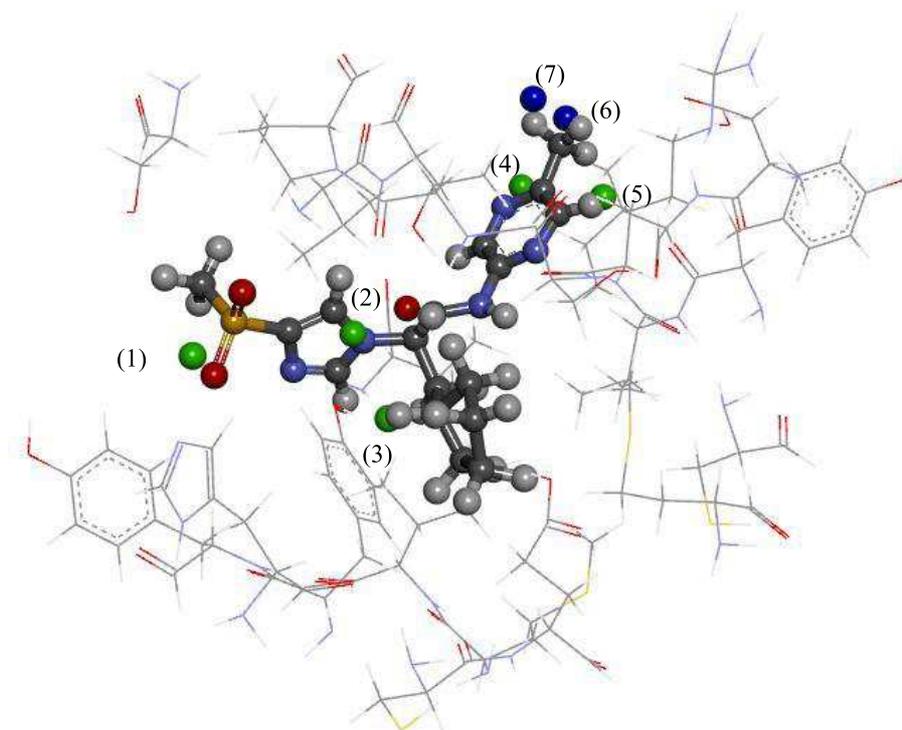


Figura 10 Modelo de QSAR-4D DR representativo do alinhamento 9, mostrando o composto **1** juntamente com os resíduos de aminoácidos selecionados da GQ

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

O GCOD-1 (3,3,1,p-) (Figura 10) é o descritor que mais contribui para o aumento na potência dos compostos, pois como pode ser observado na Equação 7, ele é o que possui um maior coeficiente positivo (1,53). Essa célula da caixa apresentou maior frequência de ocupação para o composto 54 e nenhuma frequência de ocupação para os compostos 1, 2, 5, 7, 8, 9, 12, 14, 15, 20, 21, 24, 25, 28, 31, 32, 37, 41 e 42. Isso é decorrente do fato de que esse descritor está relacionado a átomos polares negativos, e o composto **54** apresenta o substituinte

trifluórmetil, enquanto os demais compostos citados não apresentam. A presença desse descritor positivo, nessa região da molécula, pode ser explicada pela proximidade do resíduo His218 (Figura 11), que é um aminoácido polar básico, e, portanto, pode haver uma interação entre esse resíduo e os átomos de flúor desse grupo substituinte.

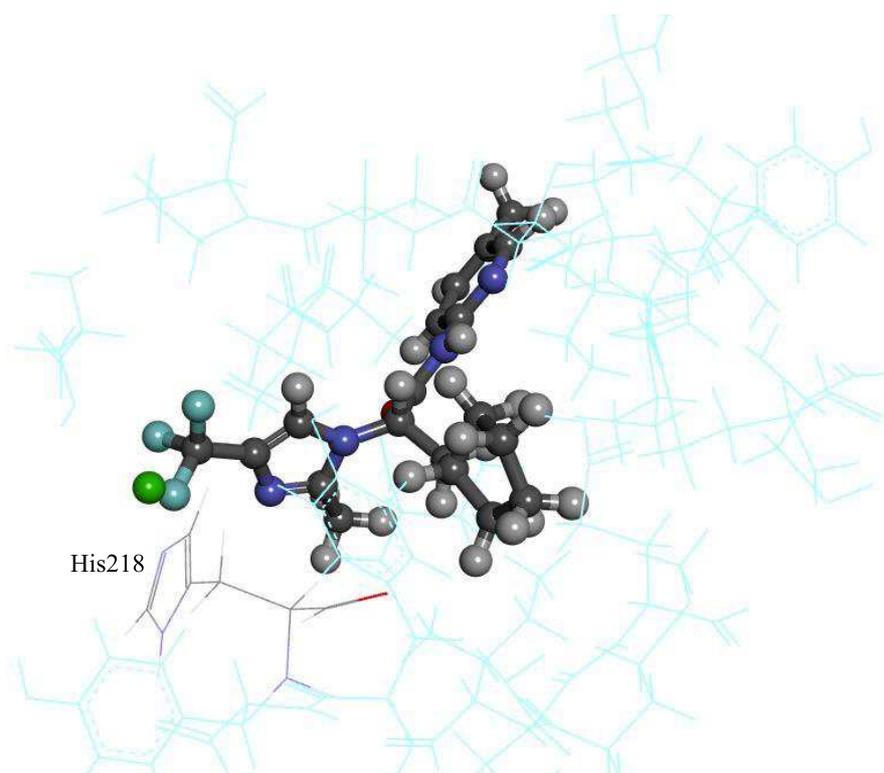


Figura 11 Representação da molécula **55** e do GCOD-1 (esfera verde) obtido do Modelo1 do alinhamento 9

O GCOD-2 (1,2,2,np) (Figura 10) também possui um coeficiente positivo, 0,37 (Eq. 8) 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 os compostos **27** e **36** (Tabela 9),

que possuem um anel imidazol nessa região. Já, para os compostos **5** e **9** a frequência de ocupação encontrada foi menor (Tabela 9), pela presença do substituinte tetrazol. Esse GCOD, localiza-se próximo ao resíduo de aminoácido Val455 (Figura 11) que é um aminoácido apolar, portanto a presença de substituintes apolares nessa posição pode favorecer interações de van der Waals entre esse aminoácido e a molécula.

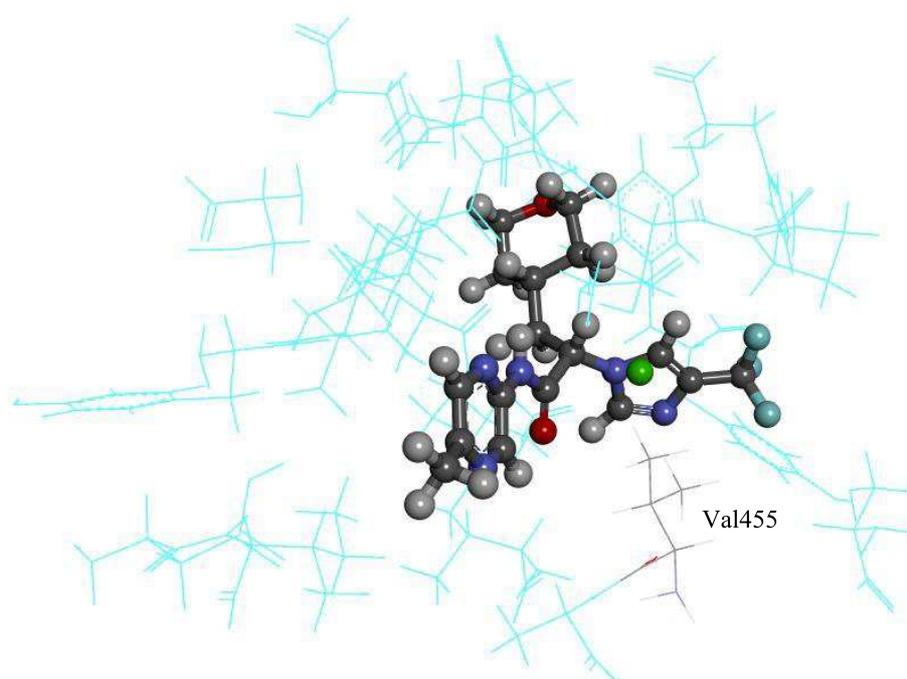


Figura 12 Representação da molécula **27** e do GCOD-2 (esfera verde) obtido do Modelo1 do alinhamento 9

Tabela 9 Valores de ocupação percentual do GCOD-2 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos **5, 9, 27 e 36**

Composto	Porcentagem de ocupação percentual da célula da caixa do GCOD-2 (+) (1,2,2,np)
5	0,034
9	0,02
27	1,358
36	1,276

O GCOD-3 (1,3,3,np) (Figura 10) também possui um coeficiente positivo, igual a 1,05, indicando o aumento na potência dos compostos que têm elevada frequência de ocupação para esse descritor. Este apresenta um IPE do tipo não polar, ou seja, assim como o GCOD-2, a presença de substituintes apolares nessa posição contribui para aumentar a potência dos compostos. A presença desse descritor pode ser explicada pela proximidade com resíduo de aminoácido Met235 (Figura13), que é um aminoácido apolar. Portanto, grupamentos apolares próximos a esse resíduo favorecem interações de van der Waals, entre eles. Para esse descritor, a maior diferença nas frequências de ocupação é em relação aos compostos 16, 17 e 27 (Tabela 10) que possuem um grupamento oxano, no lugar em que os demais compostos possuem um ciclopentano. Portanto, analisando nesta região, esses compostos são mais polares que os demais, sendo assim as frequências de ocupação foram menores. A maior frequência encontrada foi para a molécula 43, porém os valores para as demais não mudam muito.

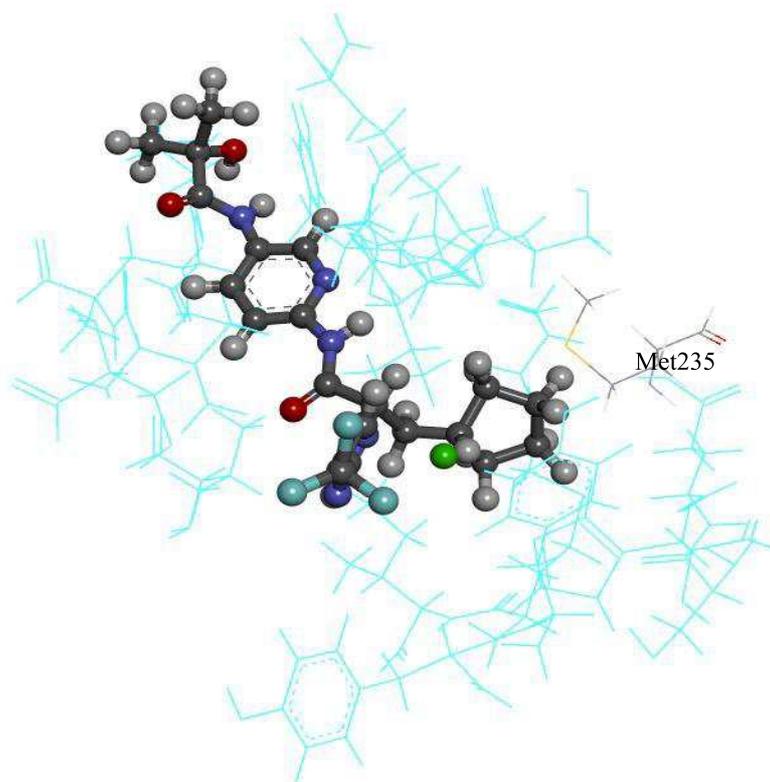


Figura 13 Representação da molécula **43** e do GCOD-3 (esfera verde) obtido do Modelo1 do alinhamento 9

Tabela 10 Valores de ocupação percentual do GCOD-3 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos **16**, **17**, **27** e **43**

Composto	Porcentagem de ocupação percentual da célula da caixa do GCOD-3 (+) (1,3,3,np)
16	0,704
17	0,434
27	0,676
43	1,528

O GCOD-4 (-2,-1,2,np) (Figura 10), que apresenta um IPE do tipo não polar, possui um coeficiente positivo de 0,57, indicando o aumento na potência dos compostos que possuem elevada frequência de ocupação para esse descritor. Ele se encontra próximo ao resíduo de aminoácido Ile159 (Figura 14), um aminoácido apolar que pode interagir com essa parte da molécula por interações de van der Waals. O composto **17** foi o que apresentou maior frequência de ocupação para esse GCOD, enquanto que o composto **2** não apresentou nenhuma. Essa diferença se deve à presença do grupo piridina no composto **17** e do grupo oxazol (que é mais polar) no composto **2**.

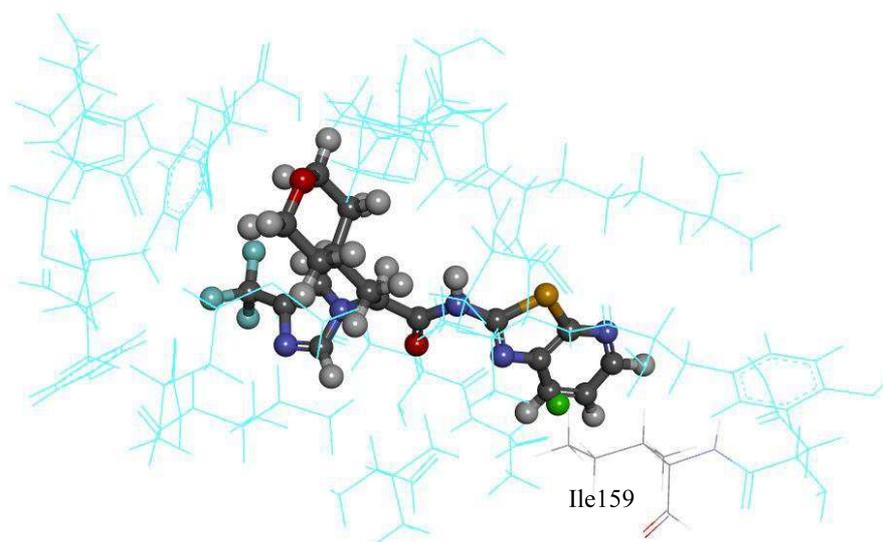


Figura 14 Representação da molécula **17** e do GCOD-4 (esfera verde) obtido do Modelo1 do alinhamento 9

O GCOD-5 (-1,-1,4,ar) (Figura 10) possui o segundo coeficiente mais positivo (1,16), que mostra que o IPE do tipo aromático influencia bastante no aumento da atividade dos compostos. Esse descritor apresentou altas frequências

de ocupação (Tabela 11) para as moléculas que possuem um anel piridina nessa posição como as moléculas **22** e **35**, e nenhuma frequência para as moléculas que possuem um anel pirazol, como as moléculas **4** e **5**. Portanto, podemos inferir que, apesar do grupo pirazol também ser um grupo aromático, a sua presença não contribui para o aumento da atividade. Na Figura 15, podemos observar a presença do resíduo Arg63 próximo ao anel piridina no composto **35**. Esse aminoácido possui cadeia lateral básica, o que permite que ele faça interações do tipo cátion- π com sistemas aromáticos, portanto, esse pode ser um tipo de interação muito importante para o aumento da atividade dos compostos estudados.

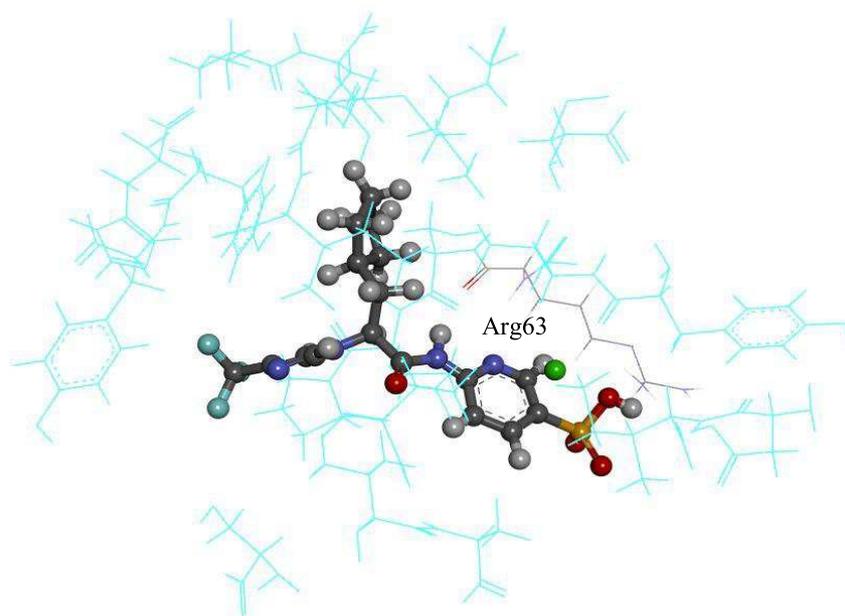


Figura 15 Representação da molécula **35** e do GCOD-5 (esfera verde) obtido do Modelo1 do alinhamento 9

Tabela 11 Valores de ocupação percentual do GCOD-5 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos **4**, **5**, **22** e **35**

Composto	Porcentagem de ocupação percentual da célula da caixa do GCOD-5 (+) (-1,-1,4,ar)
4	0
5	0
22	0,766
35	0,854

O GCOD-6 (0,-2,3,np) (Figura 10) possui um coeficiente negativo, (-0,39), e nos indica que a ocupação dessa célula da caixa por átomos apolares resulta em um decréscimo na potência dos compostos. Esse descritor está próximo ao resíduo de aminoácido Thr65 (Figura 16) que é um resíduo polar, então, moléculas que possuem alta frequência de ocupação para o GCOD-6 podem não interagir com o resíduo Thr65 e, conseqüentemente, a potência desses compostos deve ser menor. Os compostos que apresentaram alta frequência de ocupação para esse GCOD foram **19** e **22**, já compostos como **1** e **2**, não apresentaram nenhuma frequência de ocupação para esse descritor (Tabela 12).

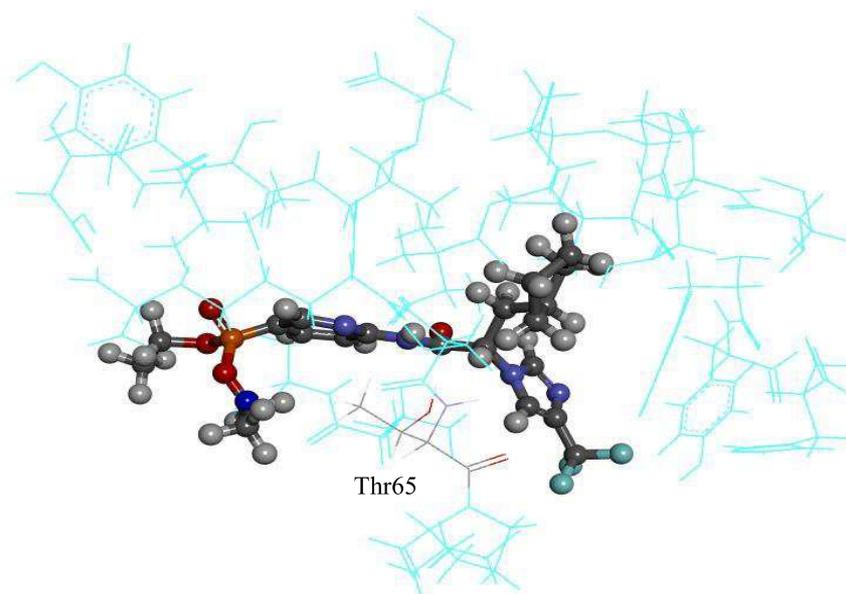


Figura 16 Representação da molécula **22** e do GCOD-6 (esfera azul) obtido do Modelo1 do alinhamento 9

Tabela 12 Valores de ocupação percentual do GCOD-6 do Modelo 1 do alinhamento 9 do QSAR-4D DR para os compostos **1**, **2**, **22** e **34**

Composto	Porcentagem de ocupação percentual da célula da caixa do GCOD-6 (-) (0,-2,3,np)
1	0
2	0
22	1,466
34	1,462

O GCOD-7 (-1,-2,3,alh) (Figura 10), possui uma grande influência na atividade, visto que ele possui o coeficiente mais negativo de todos, (-0,98). Esse descritor nos mostra que a ocupação dessa célula da caixa por átomos aceptores de ligação de hidrogênio resulta em uma diminuição na potência dos compostos.

Na Figura 17, podemos ver o resíduo de aminoácido Ala456 perto desse descritor. Como esse resíduo é apolar, seria necessária a presença de grupos apolares em sua proximidade, para que ocorresse interação do tipo van der Waals, aumentando, assim, a atividade. As maiores frequência de ocupação para esse GCOD foram encontradas para os compostos **35** e **39**, que possuem, nesse local, carbonilas como aceptores de ligação de hidrogênio. Já, compostos como **1** e **2**, que não apresentam grupos como este, nessa região da molécula, não apresentaram nenhuma frequência de ocupação para esse descritor.

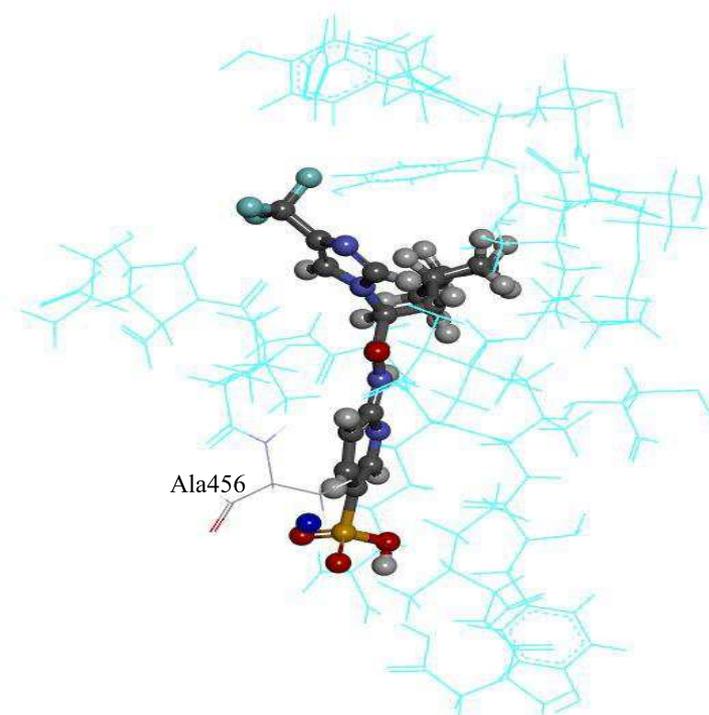


Figura 17 Representação da molécula **35** e do GCOD-7 (esfera azul) obtido do Modelo1 do alinhamento 9

5.4 Planejamento de potenciais ativadores Glucoquinase

O desenvolvimento de novos medicamentos é, sem dúvida, uma das tarefas mais difíceis da ciência (ALONSO; BLIZNYUK; GREADY, 2006). Essa tarefa envolve múltiplos fatores, responsáveis pela resposta terapêutica desejada e decorrente da complexidade dos sistemas biológicos. Diversas estratégias de desenho molecular de novos protótipos são conhecidas, porém quando se conhece o receptor, especialmente o sítio de interação, é possível o desenho de inibidores por processos de complementaridade molecular planejada, podendo, então, prever interações no sítio estudado.

Analisando os resultados obtidos a partir do estudo de QSAR-4D-DR, observamos que as principais características que aumentam a atividade é a presença do grupo trifluórmetil em uma das extremidades das moléculas pela possibilidade de ocorrer interação com o resíduo de aminoácido His218, e a presença de um grupo aromático, como a piridina, perto do resíduo Arg63, possibilitando uma interação cátion- π . O principal fator negativo encontrado foi a presença de átomos aceptores de ligação de hidrogênio perto do resíduo Ala456.

Tendo em vista esses resultados, foram propostas novas estruturas (Figura 18) para síntese e teste de atividade. Para essas moléculas ainda não sintetizadas, foi feito um ensaio virtual de atividade, aplicando a Equação 7, no qual foram obtidos os valores de atividade predita.

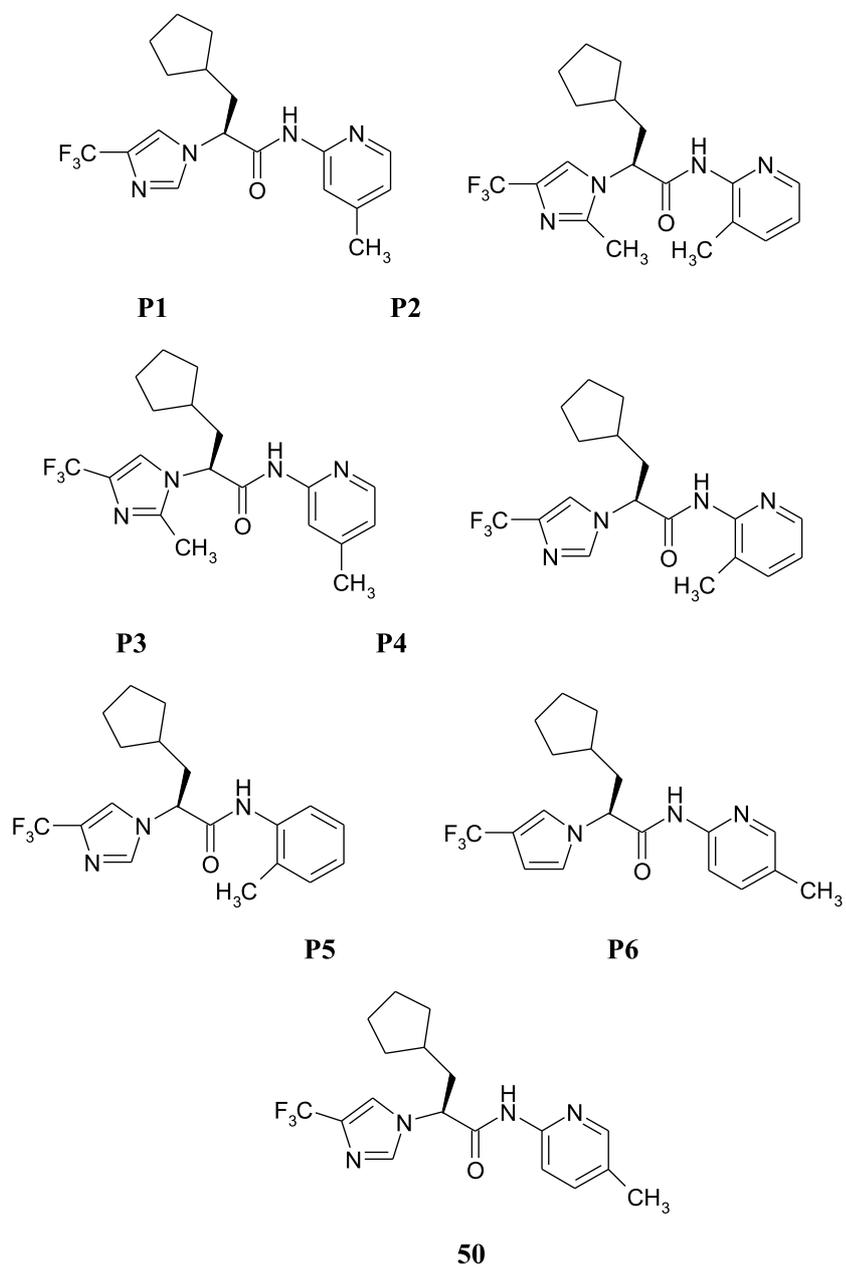


Figura 18 Ilustração das moléculas propostas: **P1**, **P2**, **P3**, **P4**, **P5**, **P6** e do composto **50**

As moléculas **P1** e **P4** foram propostas com base na estrutura do composto **53**, sendo excluído o grupo amida e adicionado um grupo metil, a fim de que possam ocorrer interações com resíduo de aminoácido Ala456. As moléculas **P2** e **P3** foram propostas baseadas na estrutura do composto **54** e mudando a posição do grupo metil. A molécula **P5** foi proposta, a fim de avaliar a importância do nitrogênio no anel aromático, fato que não foi verificado a partir dos descritores obtidos no modelo, comparando, então, com a atividade calculada para a molécula **P4**. Já, a molécula **P6** foi proposta com base no composto **50** (grupo de teste), e sua finalidade foi verificar a importância do N, substituindo o grupo imidazol pelo grupo pirrol. Os valores de atividade pEC_{50} preditos para as moléculas **P1**, **P2**, **P3**, **P4**, **P5** e **50** encontram-se na tabela abaixo (Tabela 13).

Tabela 13 Valores de atividade preditos ($pEC_{50\text{predito}}$) para os inibidores **P1**, **P2**, **P3**, **P4**, **P5**, **P6** e **50**

	P1	P2	P3	P4	P5	P6	50
$pEC_{50\text{predito}}$	6,67	6,65	7,20	6,29	5,18	5,57	5,95

A partir dos resultados observados acima, vemos que os compostos propostos **P1**, **P2** e **P3** apresentaram uma atividade prevista maior que a atividade experimental dos compostos **53** e **54**, que são os compostos mais ativos da série com uma atividade de 6,64. A molécula **P3** foi a que apresentou maior atividade predita dentre as propostas e as estudadas, maior até que a do composto **54**, que apresentou atividade predita de 6,73. Para a molécula **P4**, apesar do alto valor de atividade, este não foi superior ao valor dos compostos mais ativos. Comparando a molécula **P4** e **P5** percebemos uma importância anteriormente não relatada, que é a presença do N no anel aromático. Quando ele é substituído por C, o valor de atividade diminui visivelmente, o que

confirma a importância do anel piridina para a atividade. Comparando a molécula **P6** com o composto **50**, vemos a importância do grupo imidazol, pois quando ele foi trocado pelo grupo pirrol na molécula **P6**, a atividade predita foi consideravelmente menor.

Na Figura 19, mostra-se a estrutura tridimensional da molécula **P3** dentro do sítio alostérico da enzima Glucoquinase. Na figura, podemos perceber a presença dos grupos metil perto dos resíduos de aminoácidos Ile211, Ala456 e Ile159. Esses resíduos são apolares, portanto podem interagir com os grupos metil existentes na molécula. Um grupo metil ligado ao anel piridina, nessa posição, confirma os GCODs 4 e 7, que indicaram que grupos apolares, perto dos resíduos Ile159 e Ala456 são bons para o aumento da atividade.

Sendo assim, a partir da metodologia de QSAR-4D DR foi possível propor três moléculas (**P1**, **P2** e **P3**) que apresentaram uma atividade superior àquela encontrada para o conjunto estudado, portanto podem ser consideradas como promissores ativadores Glucoquinase.

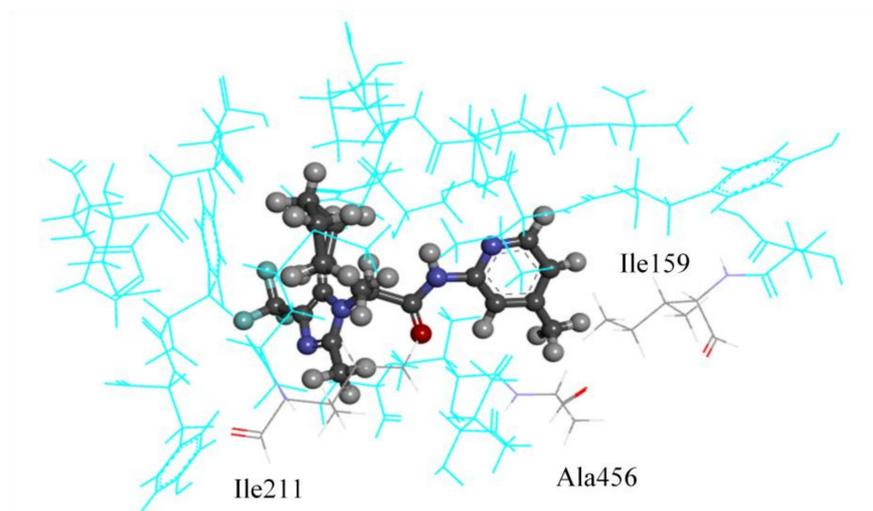


Figura 19 Representação da molécula proposta **P3** dentro da proteína Glucoquinase

6 CONCLUSÕES

Neste estudo, construímos e avaliamos modelos obtidos do método de QSAR-4D-DR, para ativadores da proteína qlucoquinase. Os modelos gerados fornecem informações importantes que podem auxiliar no planejamento de fármacos.

O melhor alinhamento (9) apresentou modelos com valores de coeficiente de determinação entre 0,674 e 0,743 e valores de validação cruzada entre 0,509 e 0,610, que mostram uma boa capacidade preditiva. O Modelo 1, desse alinhamento, foi escolhido por apresentar menor valor de LOF (*lack-of-fit*). Nesse modelo, somente a molécula **46** do grupo de treinamento foi considerada um *outlier*, já, o grupo de teste não apresentou nenhum.

A equação gerada apresenta cinco GCODs positivos e 2 negativos. A análise dos descritores mostrou que a grupos aromáticos, como a piridina perto do resíduo de aminoácido Arg63, e grupos polares negativos, como o triflúormetil perto da His218 são muito importantes para o aumento da atividade. Já, grupos aceptores de ligação de hidrogênio, perto do resíduo Ala456, não favorecem o seu aumento.

Portanto, propomos seis novos compostos para a síntese. Três deles apresentaram um valor de pEC₅₀ predito superior aos valores experimentais dos compostos estudados, portanto podem ser considerados promissores protótipos antidiabéticos.

REFERÊNCIAS

ACCELRYS SOFTWARE INCORPORATED. **Discovery studio modeling environment, release 2.1**. San Diego: [s.n.], 2007.

ALBERTI, K. G. M. M.; ZIMMET, P. F. Definition, diagnosis and classification of diabetes mellitus and its complications. Part 1: diagnosis and classification of diabetes mellitus. Provisional report of a WHO consultation. **Diabetic Medicine**, Chichester, v. 15, n. 7, p. 539-553, July 1998.

ALONSO, H.; BLIZNYUK, A. A.; GREASY, J. E. Combining docking and molecular dynamic simulations in drug design. **Medicinal Research Reviews**, New York, v. 26, n. 5, p. 531-568, Sept. 2006.

ANDRADE, C. H. et al. 4D-QSAR: perspectives in drug design. **Molecules**, Washington, v. 15, n. 5, p. 3281-3294, May 2010.

ANDRICOPULO, A. D.; SALUM, L. B.; ABRAHAM, D. J. Structure-based drug design strategies in medicinal chemistry. **Current Topics in Medicinal Chemistry**, Hilversum, v. 9, n. 9, p. 771-790, June 2009.

ARDEN, C. et al. Glucokinase is an integral component of the insulin granules in glucose-responsive insulin secretory cells and does not translocate during glucose stimulation. **Diabetes**, New York, v. 53, n. 9, p. 2346-2352, Sept. 2004.

AZEVEDO, M. J.; GROSS, J. L. Aspectos especiais da dieta no tratamento do diabete melito. **Revistada AMRIGS**, Porto Alegre, v. 34, n. 3, p. 181-6, jul./set. 1990.

BARREIRO, E. J. et al. A química medicinal de N-acilidrazonas: novos compostos-protótipos de fármacos analgésicos, antiinflamatórios e anti-trombóticos. **Química Nova**, São Paulo, v. 25, n. 1, p. 129-148, jan. 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, Nov. 2005.

BENBOW, J. W. et al. **New substituted heteroaryl compound useful for treating or delaying the progression or onset of Type II diabetes and diabetes-related disorders in mammals**. New York: Pfizer Incorporated, 2013.

BERMAN, H. M. et al. The protein data bank. **Nucleic acids Research**, Oxford, v. 28, n. 1, p. 235-242, Jan. 2000.

BRÄUER, S. et al. Evolutionary chemistry approach toward finding novel inhibitors of the type 2 diabetes target glucose-6-phosphate translocase. **Journal of Combinatorial Chemistry**, Washington, v. 7, n. 2, p. 218-226, Fev. 2005.

CALDAS, G. B.; RAMALHO, T. C.; CUNHA, E. F. da. Application of 4D-QSAR studies to a series of benzothiophene analogs. **Journal of Molecular Modeling**, Amsterdam, v. 20, n. 10, p. 1-10, Sept. 2014.

CHEN, H. et al. The diagnostics of diabetes mellitus based on ensemble modeling and hair/urine element level analysis. **Computers in Biology and Medicine**, Elmsford, v. 50, n. 7, p. 70-75, July 2014.

CHERUVALLATH, Z. S. et al. Design, synthesis and SAR of novel glucokinase activators. **Bioorganic & Medicinal Chemistry Letters**, New York, v. 23, n. 7, p. 2166-2171, Abr. 2013.

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, Sept. 2012.

GRIMSBY, J. et al. Allosteric activators of glucokinase: potential role in diabetes therapy. **Science**, London, v. 301, n. 5631, p. 370-373, July 2003.

GROSS, J. L. et al. Diabetes melito: diagnóstico, classificação e avaliação do controle glicêmico. **Arquivos Brasileiros de Endocrinologia e Metabologia**, São Paulo, v. 46, n. 1, p. 16-26, fev. 2002.

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, set. 2010.

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, Apr. 1964.

HOPFINGER, A. 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, Oct. 1997.

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

INTERNATIONAL DIABETES FEDERATION. **IDF diabetes atlas**. 6.ed. Oxford: IDF, 2013.

IRWIN, D. M.; TAN, H. Evolution of glucose utilization: glucokinase and glucokinase regulator protein. **Molecular Phylogenetics and Evolution**, Orlando, v. 70, p. 195-203, Jan. 2014.

IYNEDJIAN, P. Molecular physiology of mammalian glucokinase. **Cellular and Molecular Life Sciences**, Basel, v. 66, n. 1, p. 27-42, Jan. 2009.

JAIN, S. et al. Newer anti-hyperglycemic agents in type 2 diabetes mellitus—Expanding the horizon. **Apollo Medicine**, Amsterdam, v. 10, n. 2, p. 108-112, June 2013.

JIN, B.; HOPFINGER, A. J. A proposed common spatial pharmacophore and the corresponding active conformations of some TxA2 receptor antagonists. **Journal of Chemical Information and Computer Sciences**, Washington, v. 34, n. 4, p. 1014-1021, July 1994.

KAMATA, K. et al. Structural basis for allosteric regulation of the monomeric allosteric enzyme human glucokinase. **Structure**, Philadelphia, v. 12, n. 3, p. 429-438, Mar. 2004.

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, Dec. 2001.

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

KUBINYI, H. **QSAR: hanch analysis and related approaches. methods and principles in medicinal chemistry**. Germany: Wiley-VHC, 1993. 1993.

LANG, M. et al. Discovery and hit-to-lead optimization of novel allosteric glucokinase activators. **Bioorganic & Medicinal Chemistry Letters**, New York, v. 21, n. 18, p. 5417-5422, Sept. 2011.

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

LIU, S. et al. Insights into mechanism of glucokinase activation observation of multiple distinct protein conformations. **Journal of Biological Chemistry**, Bethesda, v. 287, n. 17, p. 13598-13610, Apr. 2012.

MAO, W. et al. Design, synthesis, and pharmacological evaluation of benzamide derivatives as glucokinase activators. **Bioorganic & Medicinal Chemistry**, Oxford, v. 20, n. 9, p. 2982-2991, May 2012.

MATSCHINSKY, F. M. et al. Glucokinase activators for diabetes therapy may 2010 status report. **Diabetes Care**, Alexandria, v. 34, n. Supplement 2, p. S236-S243, May 2011.

MATSCHINSKY, F. M.; PORTE JÚNIOR, D. Glucokinase activators (GKAs) promise a new pharmacotherapy for diabetics. **F1000 Medicine Reports**, Philadelphia, v. 2, n. 2, p. 1776-1783, June 2010.

MITSUYA, M. et al. Discovery of novel 3, 6-disubstituted 2-pyridinecarboxamide derivatives as GK activators. **Bioorganic & Medicinal Chemistry Letters**, New York, v. 19, n. 10, p. 2718-2721, May 2009.

MOLSIM, D. D. **Molecular mechanics and dynamics simulation software, users guide, version 3.2**. Lake Forest: The Chem21 Group, 1997.

PAL, M. Medicinal chemistry approaches for glucokinase activation to treat type 2 diabetes. **Current Medicinal Chemistry**, Schiphol, v. 16, n. 29, p. 3858-3874, Oct. 2009.

PARK, K. et al. Discovery of 3-(4-methanesulfonylphenoxy)-N-[1-(2-methoxyethoxymethyl)-1H-pyrazol-3-yl]-5-(3-methylpyridin-2-yl)-benzamide as a novel glucokinase activator (GKA) for the treatment of type 2 diabetes mellitus. **Bioorganic & Medicinal Chemistry Letters**, New York, v. 22, n. 7, p. 2280-2293, Apr. 2014.

PASQUALOTTO, K. R.; ALBERTON, D.; FRIGERI, H. R. Diabetes mellitus e Complicações. **Journal of Biotechnology and Biodiversity**, London, v. 3, n. 4, p. 134-145, Nov. 2012.

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

QIAN, Y. et al. Identification of RO4597014, a glucokinase activator studied in the clinic for the treatment of type 2 diabetes. **ACS Medicinal Chemistry Letters**, New York, v. 4, n. 4, p. 414-418, Mar. 2013.

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

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 Sciences**, Washington, v. 34, n. 4, p. 854-866, July 1994.

ROGERS, D. **WOLF 6.2 GFA program, molecular, simulation**. San Diego: [s.n.], 1994.

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, July 2009.

SANTOS, V. P. D.; SILVEIRA, D. R. D.; CAFFARO, R. A. Risk factors for primary major amputation in diabetic patients. **São Paulo Medical Journal**, São Paulo, v. 124, n. 2, p. 66-70, Mar. 2006.

SCHEIBER, J. Backtranslating clinical knowledge for use in cheminformatics: what is the potential? **Bioorganic & Medicinal Chemistry**, Oxford, v. 20, n. 18, p. 5461-5463, Sept. 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, Aug. 2004.

SOFTWARE FOR COMPUTER-ASSISTED MOLECULAR DESIGN. **4D-QSAR User's Manual Version 3.0**. Lake Forest: The Chem21 Group, 2001.

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

VEIGA-DA-CUNHA, M.; VAN SCHAFTINGEN, E. Identification of fructose 6-phosphate-and fructose 1-phosphate-binding residues in the regulatory protein of glucokinase. **Journal of Biological Chemistry**, Bethesda, v. 277, n. 10, p. 8466-8473, Mar. 2002.

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

VOLTARELLI, J. C. et al. Cell therapy in diabetes mellitus. **Revista Brasileira de Hematologia e Hemoterapia**, Santos, v. 31, supl. 1, p. 149-156, May 2009.

WASSERMAN, D. H. Four grams of glucose. **American Journal of Physiology-Endocrinology and Metabolism**, Bethesda, v. 296, n. 1, p. E11-E21, Jan. 2009.

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, Dec. 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, Apr. 1986.

WILD, S. et al. Global prevalence of diabetes estimates for the year 2000 and projections for 2030. **Diabetes Care**, Alexandria, v. 27, n. 5, p. 1047-1053, May 2004.