

UNIVERSIDADE FEDERAL DE ALFENAS

THIAGO CASTILHO ELIAS

MB-ISOSTER: UM SOFTWARE PARA SIMULAÇÃO DE BIOISOSTERISMO

Alfenas/MG

2018

THIAGO CASTILHO ELIAS

MB-ISOSTER: UM SOFTWARE PARA SIMULAÇÃO DE BIOISOSTERISMO

Tese apresentada como parte dos requisitos para obtenção do título de Doutor em Química pela Universidade Federal de Alfenas. Área de concentração: Físico-Química.
Orientador: Nelson José Freitas da Silveira.

Alfenas/MG

2018

Dados Internacionais de Catalogação-na-Publicação (CIP)
Sistema de Bibliotecas da Universidade Federal de Alfenas
Biblioteca Central – Campus Sede

E42m Elias, Thiago Castilho.
MB-ISOSTER: um software para simulação de bioisosterismo / Thiago Castilho Elias. – Alfenas-MG, 2018.
172 f.: il. --

Orientador: Nelson José Freitas da Silveira.
Tese (Doutorado em Química) -Universidade Federal de Alfenas, 2018.
Bibliografia.

1. Compostos bioativos. 2. Química - Computador. 3. Química Farmacêutica. I. Silveira, Nelson José Freitas da. II. Título.

CDD-541

Thiago Castilho Elias

“MB-isoster: um software para simulação de bioisosterismo”

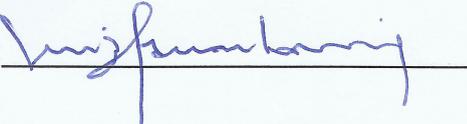
A Banca examinadora, abaixo assinada, aprova a Tese apresentada como parte dos requisitos para a obtenção do título de Doutor em Química pela Universidade Federal de Alfenas. Área de concentração: Físico- Química.

Aprovado em: 30 / 11 / 2018

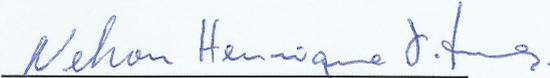
Prof. Dr. Nelson José Freitas da Silveira
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura: 

Prof. Dr. Luiz Eduardo da Silva
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura: 

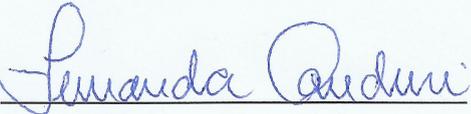
Prof. Dr. Nelson Henrique Teixeira Lemes
Instituição: Universidade Federal de Alfenas –
UNIFAL-MG

Assinatura: 

Prof. Dr. Walter Figueira de Azevedo Júnior
Instituição: Pontifícia Universidade Católica do
Rio Grande do Sul

Assinatura: 

Profa. Dra. Fernanda Canduri
Instituição: Universidade de São Paulo - USP

Assinatura: 

AGRADECIMENTOS

Agradeço à Universidade Federal de Alfenas – UNIFAL-MG e ao programa de Pós Graduação em Química – PPGQ, pela oportunidade que me foi oferecida de cursar doutoramento em química.

Agradeço à CAPES pelo apoio financeiro. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Agradeço ao meu orientador, Prof. Dr. Nelson José Freitas da Silveira, por ter aceitado me orientar e pelo tempo e dedicação disponibilizados durante minha trajetória acadêmica.

Agradeço à banca avaliadora que muito contribuiu para o aprimoramento do trabalho, agradeço à banca da Qualificação, composta pelos professores Profa. Dra. Márcia Paranho Veloso, Prof. Dr. Humberto César Brandão de Oliveira e Prof. Dr. Luciano Sindra Virtuoso; e à banca da Defesa, composta pelos professores Prof. Dr. Luiz Eduardo da Silva, Prof. Dr. Nelson Henrique Teixeira Lemes, Profa. Dra. Fernanda Canduri e Prof. Dr. Walter Filgueira de Azevedo Júnior.

Agradeço aos colegas do grupo de pesquisa MolMod-CS pelo companheirismo e dicas dadas para o aprimoramento da interface gráfica do software.

RESUMO

Na área de planejamento racional de fármacos, o bioisosterismo se constitui em uma ferramenta disponível para o químico medicinal que pretende melhorar o perfil de atuação de seus compostos líderes, ou seja, o bioisosterismo compreende a substituição de um fragmento molecular por outro, cujas propriedades físico-químicas lhe sejam similares. Dessa forma, é possível modular propriedades como absorção, aumento de tempo de meia-vida e redução da toxicidade. O programa MB-Isoster foi desenvolvido no intuito de auxiliar no desenho racional de fármacos baseando-se na estratégia do bioisosterismo. O usuário entra com a estrutura da molécula de interesse e seleciona o fragmento molecular que deseja modificar, o programa então consulta uma base de dados contendo relações bioisostéricas baseadas na literatura e constrói novas moléculas. Os bioisósteros gerados podem ser filtrados de acordo com propriedades físico-químicas calculadas internamente, como logP e logS, a fim de se selecionar moléculas com propriedades farmacocinéticas adequadas. Outra função disponível, a qual não necessita da seleção de fragmentos, consiste na substituição de um átomo de hidrogênio na molécula original por um substituinte dentre um conjunto de 35 radicais contendo funções orgânicas representativas (metil, hidroxil, fenil, etc). Uma terceira função permite que o usuário leia um complexo receptor-ligante de um arquivo pdb; interações não-ligante (van der Waals e eletrostáticas) são computadas entre os átomos do receptor e do ligante, de forma a indicar átomos do ligante com valores de energia mais elevados, como potenciais pontos para modificações bioisostéricas. Por fim, foi instalado um *plugin* que permite rodar *virtual screening* entre os bioisósteros e um receptor de interesse usando o programa AutoDock Vina como *engine* de *molecular docking*. O programa MB-Isoster está disponível gratuitamente no endereço eletrônico <http://molmod-cs.unifal-mg.edu.br/tools>. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

Palavras-chave: Bioisosterismo. Química Computacional. Química Medicinal.

ABSTRACT

In rational drug development field, bioisosterism is a tool to medicinal chemist who wants to improve their lead compounds performance, that is, bioisosterism refers to molecular fragment substitution by another that has similar physical chemistry properties. Thus, it is possible to modulate drug properties such as absorption, half-life increase and toxicity reduction. MB-Isoster software has been developed in order to help in rational drug design based on bioisosterism strategy. User starts with molecule of interest and selects molecular fragment to change, then, the software consults a database with bioisosteric relationship based on literature, and build new molecules. Generated bioisosteres can be filtered based on internally computed physical chemistry properties, such as logP and logS, in order to select molecules with appropriate pharmacokinetic properties. Another available function, to which it not needs to select a fragment, is substituting hydrogen atom in original molecule by a substituent among a set of 35 common organic radicals (methyl, hydroxyl, phenyl, and so on). A third function allows user to read receptor-ligand complex from a pdb file; non-bonded interactions (van der Waals and electrostatic) are computed between atoms from receptor and ligand, in order to indicate atoms in ligand with the highest energies as potential modification points. Lastly, it was installed a plugin that allows running virtual screening between bioisosteres and a receptor of interest using AutoDock Vina program as molecular docking engine. MB-Isoster is freely available at <http://molmod-cs.unifal-mg.edu.br/tools>. This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Key words: Bioisosterism. Computational Chemistry. Medicinal Chemistry.

LISTA DE FIGURAS

Figura 1 - Estrutura 2D de fármacos obtidos de produtos naturais ou derivados destes.....	18
Figura 2 - Estrutura do imatinib.....	19
Figura 3 - Exemplo de moléculas sintetizadas para estudo SAR a partir de um produto natural.....	20
Figura 4 - Bioisósteros do grupamento ácido carboxílico.....	21
Figura 5 - Regra do hidreto formulada por Grimm.....	22
Figura 6 - Fármaco zidovudina, um análogo do nucleosídeo timina.....	23
Figura 7 - Mecanismo de inibição da enzima timidilato sintase pelo fármaco 5-fluorouracil.....	24
Figura 8 - Relação bioisostérica entre os fármacos procaína e procainamida.....	25
Figura 9 - Desenvolvimento do fármaco tenoxicam por bioisosterismo de anéis.....	25
Figura 10 - Bioisosterismo não-clássico envolvendo a troca de hidroxila de grupo fenol pelo grupamento $-NHSO_2CH_3$	26
Figura 11 - Exemplo de bioisosterismo não-clássico por bióforos.....	27
Figura 12 - Relação bioisostérica de anelação entre os fármacos lidocaina e mepivacaina.....	27
Figura 13 - Bioisosterismo usado como forma de proteger a molécula da ação metabólica do organismo.....	28
Figura 14 - Bioisosterismo de anel com substituição de anel azola por purina.....	29
Figura 15 - Bioisosterismo por substituição de bifenil por 2-arilfurano.....	30
Figura 16 - Bioisosterismo por substituição de amida por 1,2,4-oxodiazola.....	31
Figura 17 - Bioisosterismo clássico envolvendo substituição de hidroxila por amina para o produto natural arctigenina.....	32
Figura 18 - Bioisosterismo não-clássico de substituição de anel benzênico por grupamento oximo.....	33
Figura 19 - Bioisosterismo de anéis com substituição de anel piridínico por anel benzênico com substituintes em posição meta.....	34
Figura 20 - Relação bioisostérica de tiosemicarbazona e tiazolidinonas com atividade antiparasitária.....	35
Figura 21 - Bioisosterismo clássico divalente envolvendo substituição de calcogênios.....	36
Figura 22 - Desenvolvimento de novos antiangiogênicos a partir do produto natural wondonin por meio de modificações bioisostéricas.....	37

Figura 23 - Bioisosterismo clássico monovalente cloro → metil para o composto atpenin A5, inibidor do complexo II da cadeia respiratória mitocondrial.....	38
Figura 24 - Diagrama de classes para o pacote mol.....	41
Figura 25 - Diagrama de classes mostrando a implementação do campo de força.....	42
Figura 26 - Diagrama de classes mostrando interações de leitura e escrita de arquivos moleculares.....	43
Figura 27 - Diagrama de classes para as funções de construção de novos bioisósteros.....	44
Figura 28 - Diagrama de classes para o cálculo das propriedades físico-químicas.....	45
Figura 29 - Diagrama de classes para a execução do <i>virtual screening</i>	46
Figura 30 - Arquivo de configuração usado para se definir as ordens de ligações covalentes..	50
Figura 31 - Exemplos de anéis e sua representação textual.....	53
Figura 32 - Sistema de anéis simples.....	54
Figura 33 - Sistema de anéis fundidos.....	55
Figura 34 - Sistema de anéis espiranos.....	55
Figura 35 - Sistema de anéis em ponte.....	56
Figura 36 - Sistema de anéis complexos.....	56
Figura 37 - Molécula quinina e sua representação como grafo.....	57
Figura 38 - Matriz de adjacência para o grafo da molécula benzeno.....	58
Figura 39 - Grafo da molécula quinina, mostrando a remoção de vértices não pertencentes a sistemas de anéis.....	60
Figura 40 - Simplificação da representação molecular por contração de vértices.....	61
Figura 41 - Estrutura de dados em lista encadeada para armazenar cadeias de átomos contidos em vértices contraídos.....	62
Figura 42 - Contração de vértices para anéis isolados.....	63
Figura 43 - Contração de vértices para anéis espiranos.....	64
Figura 44 - Contração de vértices para anéis fundidos, caso I.....	65
Figura 45 - Contração de vértices para anéis fundidos, caso II.....	65
Figura 46 - Contração de vértices para sistemas de ponte.....	66
Figura 47 - Exemplo de sistemas de anéis complexos.....	67
Figura 48 - Representação esquemática do vetor formado entre dois átomos.....	68
Figura 49 - Representação esquemática para o cálculo do ângulo entre três átomos.....	69
Figura 50 - Definição de ângulo diedro.....	70
Figura 51 - Representação do ângulo de Wilson.....	71
Figura 52 - Representação esquemática da distância covalente.....	72

Figura 53 - Arquivo MMFFBOND.PAR do campo de força MMFF94.....	93
Figura 54 - Alguns <i>templates</i> de anéis usados na minimização da geometria molecular.....	102
Figura 55 - Minimização de ligações covalentes.....	104
Figura 56 - Minimização dos ângulos diedro.....	108
Figura 57 - Algoritmo de gradientes conjugados não lineares.....	111
Figura 58 - Fragmentação molecular.....	112
Figura 59 - Fragmentos bioisósteros do grupo amida.....	114
Figura 60 - Construção dos bioisósteros.....	116
Figura 61 - Arquivo de configuração para construção da biblioteca de fragmentos.....	118
Figura 62 - Definição do sítio ativo para cálculo da interação entre átomos da molécula e do receptor.....	121
Figura 63 - Exemplos de moléculas para as quais se aplicam fatores de correção no cálculo do logP pelo algoritmo XLOP3-AA.....	125
Figura 64 - <i>Grid box</i> empregado no <i>docking molecular</i>	130
Figura 65 - Arquivo de configuração contendo os parâmetros do <i>grid box</i> para rodar <i>docking molecular</i> com o programa AutoDock Vina.....	130
Figura 66 - Árvore de diretórios organizada para se rodar <i>virtual screening</i>	132
Figura 67 - Sistema modelo usado para ilustrar o uso do programa MB-Isoster.....	133
Figura 68 - Script bash para rodar o programa MB-Isoster.....	136
Figura 69 - Ícone do programa MB-Isoster.....	136
Figura 70 - Tela inicial do programa MB-Isoster.....	137
Figura 71 - Bioisósteros exibidos no <i>Bioisosteres Panel</i>	138
Figura 72 - Janela de diálogo “Read Ligand from PDB-Complex”.....	142
Figura 73 - Janela de diálogo “Molecule Filter”.....	143
Figura 74 - Janela de diálogo “Virtual Screening”.....	145
Figura 75 - Tabela exibida após a conclusão do virtual screening.....	146
Figura 76 - Molécula chiquimato usada como protótipo para substituições bioisostérica.....	149
Figura 77 - Análogos por adição de substituintes do composto chiquimato.....	158
Figura 78 - Gráfico 2D da interação entre o chiquimato e análogos com a enzima chiquimato quinase.....	159

LISTA DE TABELAS

Tabela 1	- Inibição de hCA I, II, IV e IX dos compostos 10, 22 e SLC-0111 por ensaio de parada de fluxo de CO ₂	36
Tabela 2	- Atributos da classe Atom.....	47
Tabela 3	- Atributos da classe Bond.....	49
Tabela 4	- Representação da ordem de ligações covalentes para os formatos mol/sdf, mol2 e a representação interna ao programa.....	49
Tabela 5	- Atributos da classe Ring.....	51
Tabela 6	- Atributos da classe Molecule.....	73
Tabela 7	- Tipo de átomo conforme as especificações SYBYL.....	80
Tabela 8	- Tipos de átomos de acordo com a especificação AutoDock.....	83
Tabela 9	- Alguns tipos de átomos definidos pelo campo de força MMFF94.....	93
Tabela 10	- Atributos da classe Radical.....	113
Tabela 11	- Fragmentos bioisostéricos e sua representação linear.....	115
Tabela 12	- Exemplos de tipos de átomos definidos no algoritmo XLOP3-AA.....	125
Tabela 13	- Exemplos de tipos de átomos definidos no algoritmo de cálculo de logS.....	126
Tabela 14	- Exemplos de tipos de átomos definidos no algoritmo de cálculo de TPSA.....	127
Tabela 15	- Termos da função de <i>score</i> empregada pelo programa AutoDock Vina.....	129
Tabela 16	- Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,0001 kcal/mol.....	147
Tabela 17	- Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,001 kcal/mol.....	148
Tabela 18	- Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,01 kcal/mol.....	148
Tabela 19	- Cálculo das interações não ligante entre os átomos do chiquimato e o sítio ativo da enzima chiquimato quinase, obtidos com a estrutura pdb 2iyx.....	150
Tabela 20	- <i>Virtual screening</i> entre os bioisósteros do chiquimato e a estrutura 2iyx.....	151
Tabela 21	- <i>Virtual screening</i> entre os análogos por substituição de hidrogênio do chiquimato e a estrutura 2iyx.....	155
Tabela 22	- Rmsd entre o composto chiquimato presente na estrutura pdb 2iyx e a conformação de <i>docking</i> de seus análogos.....	157

LISTA DE ABREVIATURAS E SIGLAS

ATP	-	Adenosine triphosphate (Trifosfato de adenosina)
BFS	-	Breadth-First Search (Busca em Amplitude)
CADD	-	Computer-aided drug design (Desenho de fármacos assistido por computador)
CADDD	-	Computer-aided drug discovery and development (Descoberta e desenvolvimento de fármacos assistido por computador)
CCMS	-	Computational Center for Macromolecular Structure (Centro Computacional para Estruturas de Macromoléculas)
CDK	-	Chemistry Development Kit (Ferramenta de Desenvolvimento em Química)
DFT	-	Density Functional Theory (Teoria do Funcional da Densidade)
HTS	-	High Throughput Screening (Triagem de Alta Produtividade)
IUPAC	-	International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada)
LBDD	-	Ligand-Based Drug Design (Desenho de Fármaco Baseado no Ligante)
MMP	-	Matched Molecular Pairs (Pares de Emparelhamento Molecular)
MORT	-	Molecular Objects and Relevant Templates (Objetos Moleculares e Modelos Aplicáveis)
PDB	-	Protein Data Bank (Banco de Dados de Proteínas)
PPAR	-	Peroxisome Proliferator-Activated Receptors (Receptores Ativados por Proliferadores de Peroxissomos)
PPAR	-	Peroxisome Proliferator-Activated Receptors (Receptores Ativados por Proliferação de Peroxisomo)
PSA	-	Polar Surface Area (Área de Superfície Polar)
QSAR	-	Quantitative Structure-Activity Relationship (Relação Estrutura-Atividade Quantitativa)
SAR	-	Structure-Activity Relationship (Relação Estrutura-Atividade)
SBDD	-	Structure-Based Drug Design (Desenho de Fármacos Baseado na Estrutura)
TPSA	-	Topological Polar Surface Area (Área de Superfície Polar Topológica)
VEGF	-	Vascular Endothelial Growth Factor (Fator de Crescimento Endotelial Vascular)

SUMÁRIO

1	INTRODUÇÃO	13
2	REVISÃO DA LITERATURA	16
2.1	QUÍMICA MEDICINAL.....	16
2.2	BIOISOSTERISMO.....	21
2.3	TRABALHOS ENVOLVENDO O EMPREGO DE BIOISOSTERISMO.....	28
3	MATERIAIS E MÉTODOS	39
3.1	REPRESENTAÇÃO MOLECULAR.....	46
3.1.1	Classe Atom	47
3.1.2	Classe Bond	49
3.1.3	Classe Ring	51
3.1.3.1	Sistemas de anéis.....	54
3.1.3.2	Algoritmo de busca de anéis.....	57
3.1.4	Descritores Geométricos	68
3.1.5	Classe Molecule	73
3.2	LEITURA E ESCRITA DE ARQUIVOS DE ESTRUTURAS MOLECULARES.....	76
3.2.1	Formato mol	76
3.2.2	Formato sdf	78
3.2.3	Formato mol2	79
3.2.4	Formato pdb	81
3.2.5	Formato pdbqt	82
3.2.6	Algoritmos para leitura e edição de moléculas	84
3.3	OTIMIZAÇÃO DA GEOMETRIA MOLECULAR.....	88
3.3.1	Campo de força MMFF94s	89
3.3.2	Validação do campo de força MMFF94s	94
3.3.3	Implementação da álgebra matricial	96
3.3.4	Algoritmo de minimização de energia	100
3.3.4.1	Minimização de anéis.....	101
3.3.4.2	Minimização de ligações covalentes.....	103
3.3.4.3	Minimização dos ângulos.....	105
3.3.4.4	Minimização dos diedros.....	107
3.3.4.5	Refinamento.....	108

3. 4	CONSTRUÇÃO DE MOLÉCULAS ANÁLOGAS POR BIOISOSTERISMO....	111
3. 4. 1	Construção da biblioteca de fragmentos	117
3. 4. 2	Construção de novas moléculas por substituição de átomos de hidrogênio	119
3. 5	ANÁLISE DE INTERAÇÕES RECEPTOR-LIGANTE.....	120
3. 6	CÁLCULO DE PROPRIEDADES <i>DRUG-LIKE</i>	122
3. 7	<i>VIRTUAL SCREENING</i>	127
3. 8	ESTUDO DE CASO.....	132
4	RESULTADOS E DISCUSSÃO	135
4. 1	INTERFACE GRÁFICA.....	136
4. 1. 1	Barra de menus	138
4. 1. 2	Leitura de moléculas a partir de complexos receptor-ligante	140
4. 1. 3	Seleção de bioisósteros com propriedades <i>drug-like</i>	143
4. 1. 4	Execução de <i>virtual screening</i>	144
4. 2	VALIDAÇÃO DO CAMPO DE FORÇA.....	147
4. 3	ESTUDO DE CASO.....	149
4. 4	COMPARAÇÃO COM SOFTWARES SIMILARES.....	160
5	CONCLUSÃO	162
	REFERÊNCIAS	163

1 INTRODUÇÃO

Desenho de fármacos assistido por computador (do inglês, *Computer-aided drug design*, CADD) é um termo utilizado para referenciar o uso de ferramentas computacionais em projetos de pesquisa e desenvolvimento de fármacos, possuindo duas vertentes básicas: Desenho de fármacos baseado na estrutura (do inglês, *Structure-based drug design*, SBDD), e Desenho de fármaco baseado no ligante (do inglês, *ligand-based drug design*, LBDD). No primeiro caso, SBDD, a estrutura tridimensional do alvo molecular – podendo ser uma enzima, receptor de membrana, RNA ou DNA – é conhecida, dessa forma, pode-se modelar uma molécula que melhor se encaixe no sítio receptor do alvo. No segundo caso, LBDD, conhece-se um conjunto de ligantes estruturalmente relacionados que apresentam alguma atividade biológica de interesse, como atividade antimicrobiana ou antifúngica, e a partir da relação entre as propriedades físico-químicas e atividades biológicas dos mesmos, pode-se realizar modificações estruturais a fim de melhorar suas atividades (YU E MACKERELL JR., 2016). Outro termo utilizado é descoberta e desenvolvimento de fármacos assistido por computador (do inglês, *Computer-aided drug discovery and development*, CADDD), que envolve tanto as técnicas computacionais quanto as experimentais de forma complementar, tendo como características: 1. O uso de computadores com o intuito de agilizar as etapas de descoberta e desenvolvimento de fármacos; 2. O aumento das informações químicas e biológicas em relação a ligantes e alvos moleculares para identificação e otimização de novos fármacos; 3. Desenvolvimento de filtros para eliminar moléculas com características indesejadas (KAPETANOVIC, 2006). A química computacional pode auxiliar na busca por moléculas farmacologicamente interessantes ao reduzir o número de compostos e experimentos que deverão ser realizados até se chegar nos testes clínicos (TSUI, ORTWINE E BLANEY, 2017). Indústrias farmacêuticas tem investido no desenvolvimento de softwares para busca de novos fármacos contratando pesquisadores especializados, geralmente com título de doutor e com amplos conhecimentos em química computacional, biologia computacional, física e áreas correlacionadas, além de habilidades no desenvolvimento de softwares (LUTY E ROSE, 2016).

Atualmente estão disponíveis diversos programas destinados ao CADD, incluindo programas de *molecular docking* como AutoDock4, desenvolvido por Morris *et al.* (2009), e Auto Dock Vina, criado por Trott e Olson (2010), dinâmica molecular como Amber, descrito em Case *et al.* (2005), e GROMACS, de Hess *et al.* (2008), e identificação de farmacóforos

como MOE, segundo Labute *et al.* (2001) e PHASE (DIXON *et al.*, 2006). Além disso, existem bibliotecas de softwares prontas para serem utilizadas na construção de novos programas em química computacional, como é o caso do CDK (*Chemistry Development Kit*), descrito em Steinbeck *et al.* (2003) escrito em java, útil para desenvolvimento de softwares em quimio- e bioinformática, com métodos para renderização de estruturas químicas, leitura e escrita de arquivos, checagem de isomorfismo, entre outros. Outra biblioteca disponível é o MORT (*Molecular Objects and Relevant Templates*), de Zhang *et al.* (2014), escrita em C++ e usada em biologia computacional e CADD, possui um modelo relacional para representar moléculas e seus átomos e apresenta funções como cálculo da energia molecular.

A pesquisa e desenvolvimento de novos fármacos é um processo demorado e custoso, sendo necessário pelo menos 14 anos desde a concepção do projeto até a entrada do medicamento no mercado (MYERS E BAKER, 2001). Basicamente é constituído das seguintes etapas (HORIEN E YUAN, 2017):

- a) Identificação e validação de um alvo molecular relacionado com a doença;
- b) Busca por compostos capazes de interagir com o alvo molecular e intervir na doença;
- c) Testes *in vitro* e *in vivo* em modelos celulares e animais, a fim de investigar as propriedades farmacocinéticas (administração, distribuição, metabolismo e excreção) e toxicidade;
- d) Testes clínicos de fase I: administração do composto em pequeno número de voluntários saudáveis, a fim de testar a segurança e dosagem ideal para seres humanos;
- e) Testes clínicos de fase II: testes de eficácia e efeitos colaterais em pequeno número de pacientes com a doença que se deseja controlar;
- f) Testes clínicos de fase III e IV: número maior de pacientes, caracterização da eficácia e reações adversas.

Para que uma molécula venha a se tornar um novo fármaco, esta deve possuir algumas características: ser segura, eficaz, estável química e metabolicamente, capaz de atingir o alvo molecular, disponível a partir de fontes naturais ou síntese, e patenteável (LESK, 2008).

Na vanguarda do desenho racional de candidatos a fármaco está a Química Medicinal, uma ciência multidisciplinar cujo objeto de estudo é a “descoberta, desenvolvimento, identificação e interpretação do modo de ação de compostos biologicamente ativos a nível molecular” (KHAN, DEIMLING E PHILIP, 2011). Dentre as estratégias adotadas pela Química Medicinal para a melhoria de compostos líderes está o bioisosterismo, um conceito

derivado do isosterismo, cujas primeiras pesquisas datam do ano 1919 com o pesquisador Langmuir, que investigava as similaridades em propriedades físico-químicas de diferentes moléculas. Langmuir percebeu que moléculas como N_2 e CO eram similares em muitas de suas propriedades físico-químicas, concluindo que isto se devia ao número e arranjo de elétrons ser igual entre essas moléculas, cunhando assim o termo isosterismo (LANGMUIR, 1919). Mais tarde, em 1951, Friedman propõe o termo bioisosterismo, o qual engloba os conceitos do isosterismo, sendo estendido a moléculas que possuem atividade biológica similar, em alguns casos antagonista (FRIEDMAN, 1951). Burger ampliou este conceito dizendo que bioisósteros são compostos com formato e volume molecular próximos, com distribuição de elétrons similar, apresentando propriedades físicas em comum (BURGER, 1991). O que caracteriza o bioisosterismo é o fato de compostos bioisósteros interagirem com o mesmo alvo molecular, e em consequência, possuírem atividade biológica similar (PATANI E LAVOIE, 1996). O bioisosterismo é empregado na modificação de compostos líderes a fim de melhorar suas propriedades farmacocinéticas e farmacodinâmicas (MUGNAINI, PASQUINI, E CORELLI, 2012).

O presente trabalho apresenta o software MB-Isoster, desenvolvido com o intuito de auxiliar o químico medicinal na busca por novas moléculas a partir de modificações bioisostéricas em um composto líder. Por meio de uma interface gráfica, o usuário pode abrir uma estrutura molecular e selecionar os átomos que deverão ser substituídos; o programa procura por fragmentos que possam ser adicionados, gerando uma coleção de novas moléculas. Estas poderão ser avaliadas quanto ao seu perfil farmacocinético por meio do cálculo de propriedades *drug-like*; e quanto ao seu perfil farmacodinâmico por meio de *molecular docking*. As estruturas mais interessantes poderão ser selecionadas para estudos em bancada (síntese e ensaios *in vitro* e *in vivo*). O trabalho encontra-se estruturado em cinco capítulos, sendo o primeiro esta Introdução. No capítulo 2 será feita uma revisão da literatura, abordando conceitos de química medicinal, bioisosterismo e trabalhos experimentais exemplificando o uso do bioisosterismo na busca por moléculas bioativas. No capítulo 3 se apresenta a construção lógica do programa, com a definição das classes e pacotes, as classes usadas para representar entidades que constituem a molécula (átomos, ligações covalentes, anéis, moléculas), a leitura e escrita de arquivos moleculares, a implementação do campo de força, os métodos para construção dos bioisósteros, cálculo de propriedades *drug-like*, cálculo da energia de interação em complexos receptor-ligante, *virtual screening* e um estudo de caso. O capítulo 4 apresenta a interface gráfica do programa e sua utilização. Por fim, o capítulo 5 faz uma conclusão do trabalho e aponta aprimoramentos futuros para o software.

2 REVISÃO DA LITERATURA

A revisão da literatura abordará os seguintes tópicos: primeiramente se falará a respeito da Química Medicinal, uma disciplina que trata do planejamento racional de novas moléculas que possam vir a se transformar em fármacos aprovados; serão elucidados os principais conceitos do bioisosterismo, uma das técnicas de desenho de novas moléculas utilizadas pela Química Medicinal; por fim, serão exemplificados trabalhos experimentais em que se empregou o bioisosterismo.

2.1 QUÍMICA MEDICINAL

A química medicinal possui papel de destaque na pesquisa e desenvolvimento de fármacos. Químicos medicinais são responsáveis por selecionar, preparar e avaliar novos compostos, os quais, se apresentarem atividade biológica interessante, serão eleitos como compostos líderes (LOMBARDINO E LOWE, 2004). A seguinte definição é feita pela IUPAC:

A Química Medicinal se ocupa da descoberta, do desenvolvimento, da identificação e da interpretação do mecanismo de ação de compostos com atividade biológica em nível molecular. Ênfase é colocada nos fármacos, mas o interesse da química medicinal não se restringe a estes mas inclui compostos bioativos em geral. A Química Medicinal também se preocupa com o estudo, identificação, e síntese dos produtos metabólicos desses fármacos e compostos relacionados (WERMUTH *et al.*, 1998).

A química medicinal é uma ciência multidisciplinar, abrangendo áreas como química orgânica, bioquímica, farmacologia, biologia molecular, genética, imunologia, farmacocinética, toxicologia, físico-química, cristalografia de difração de raios X, espectroscopia, e técnicas de simulação e análise computacional (IMMING, 2015). Três etapas principais estão envolvidas na prática da química medicinal (WERMUTH, 2006):

- a) Etapa da descoberta: escolha do alvo molecular, identificação e produção de substâncias capazes de interagir com o alvo selecionado. Os compostos produzidos nessa etapa são denominados compostos líderes, raramente serão moléculas prontas para entrarem no mercado;

- b) Etapa da otimização: melhorias adicionais à estrutura do composto líder. São realizadas pequenas modificações estruturais, gerando uma família de moléculas análogas, estas serão ensaiadas em um estudo SAR (do inglês, *Structure-Activity Relationship*, Relação Estrutura-Atividade), fornecendo o entendimento do mecanismo de ação a nível molecular. Seus principais objetivos são aumentar a potência e a seletividade, e diminuir a toxicidade. Moléculas com características farmacocinéticas insatisfatórias são eliminadas nessa fase;
- c) Etapa de desenvolvimento: as propriedades farmacocinéticas do composto selecionado são aprimoradas. Elabora-se uma formulação química pronta para entrar em testes clínicos.

Os produtos naturais são a principal fonte de inspiração para a obtenção e desenvolvimento de novos fármacos, a diversidade e quantidade de moléculas encontradas na natureza não pode ser alcançada pelas bibliotecas de compostos químicos sintéticos das grandes empresas farmacêuticas (NEWMAN E CRAGG, 2012). Exemplos de fármacos obtidos diretamente de fontes naturais incluem a artemisina, um composto isolado a partir de extratos da planta *Artemisia annua*, com atividade contra parasitas do gênero *Plasmodium*, causadores da malária (FIGURA 1). Esta molécula foi descoberta a partir do estudo de extratos de mais de 2000 plantas utilizadas na medicina tradicional chinesa, indicadas para o tratamento da febre (TU, 2011). A partir da artemisina, foram sintetizados fármacos análogos mais potentes que esta, como artemeter e artesunato (SHEN, 2015). Outro exemplo é o taxol, um diterpeno obtido de árvores da espécie *Taxus brevifolia*, apresenta atividade anticâncer por inibição da depolimerização de microtúbulos, fazendo com que células tumorais fiquem paralisadas na fase G2/M, impedindo a mitose e induzindo a apoptose (FIGURA 1) (HEINSTEIN E CHANG, 1994).

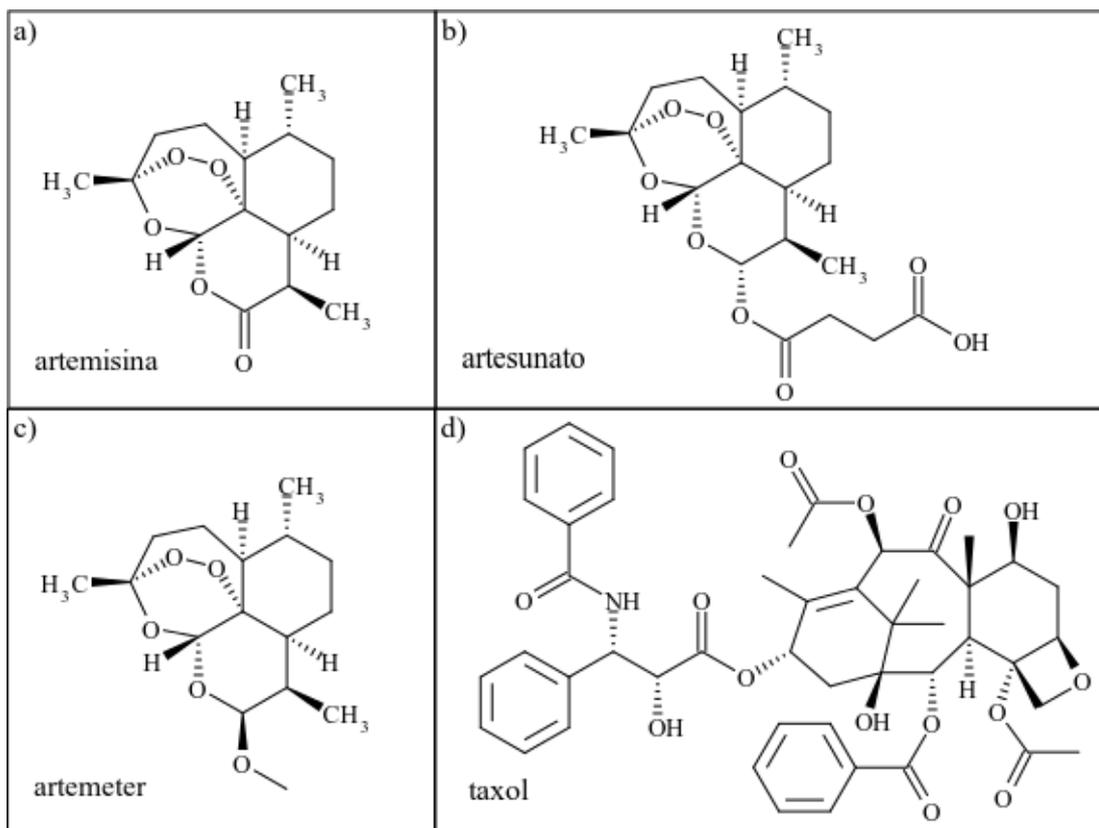


Figura 1 – Estrutura 2D de fármacos obtidos de produtos naturais ou derivados destes.

Fonte: Do autor.

Legenda: a) artemisina;
 b) artesunato;
 c) artemeter;
 d) taxol.

A química combinatória é outra técnica empregada para a descoberta de compostos líderes, é baseada na construção de uma biblioteca de diferentes moléculas formadas a partir da ligação covalente entre diversos blocos construtores, realizada de forma aleatória em fase sólida ou líquida (LIU, LI E LAM, 2017). Uma vez construída, a biblioteca química pode ser rastreada contra um ou mais alvos moleculares por meio de HTS (*High Throughput Screening*, Triagem de Alta Produtividade) (PANDEYA E THAKKAR, 2005). Um exemplo de fármaco obtido por química combinatória é o imatinib, cuja atividade anticâncer se dá pela inibição de tirosina quinase (EC: 2.7.10.2). Sua descoberta ocorreu a partir de um composto líder identificado em uma biblioteca química com substâncias mimetizadoras de ATP (FIGURA 2) (CAPDEVILLE *et al.*, 2002).

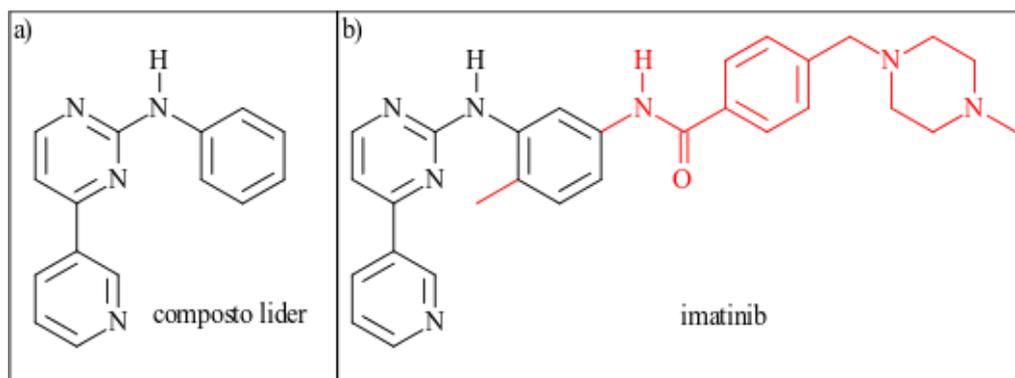


Figura 2 – Estrutura do imatinib.

Fonte: Do autor.

Legenda: a) composto líder identificado com atividade anticâncer;

b) imatinib. As diferenças estruturais entre o imatinib e seu composto líder são apresentadas em vermelho.

Após a síntese de uma coleção de análogos ao composto líder é realizado um ensaio de SAR (*Structure-Activity Relationship*, Relação Estrutura-Atividade). Diferentes grupos químicos são inseridos em posições determinadas da estrutura do composto líder e avalia-se como essas modificações interferem na atividade biológica do novo composto, seja melhorando ou reduzindo sua eficácia (MCKINNEY *et al.*, 2000). Dessa forma é possível inferir qual é o farmacóforo, a região da molécula indispensável para a interação com o receptor. Quando relações quantitativas são inseridas, tem-se o QSAR (*Quantitative Structure-Activity Relationship*, Relação Estrutura-Atividade Quantitativa). Pode-se relacionar a concentração mínima necessária para que o efeito biológico seja observado com propriedades físico-químicas dos compostos, como o logP, propriedades eletrônicas e efeitos estéricos (SHAHLAEI, 2013).

Em estudo recente, Kotoku e colaboradores (2017) realizaram um rastreamento de moléculas com atividade antiangiogênica a partir de substâncias isoladas da esponja *Corticium simplex*, identificando-se a cortistatina A, a qual foi potente e seletiva contra o modelo biológico de células endoteliais de veias umbilicais humanas (FIGURA 3). Devido à complexidade dessa estrutura, foi sintetizado um análogo simplificado com atividade comparável à molécula original. A partir deste, foram construídas cinco novas estruturas para estudos de SAR (compostos 23a, 23b, 28, 29 e 30). A introdução de uma acetamida ao anel A provocou aumento na atividade antiproliferativa seletiva contra o modelo celular utilizado.

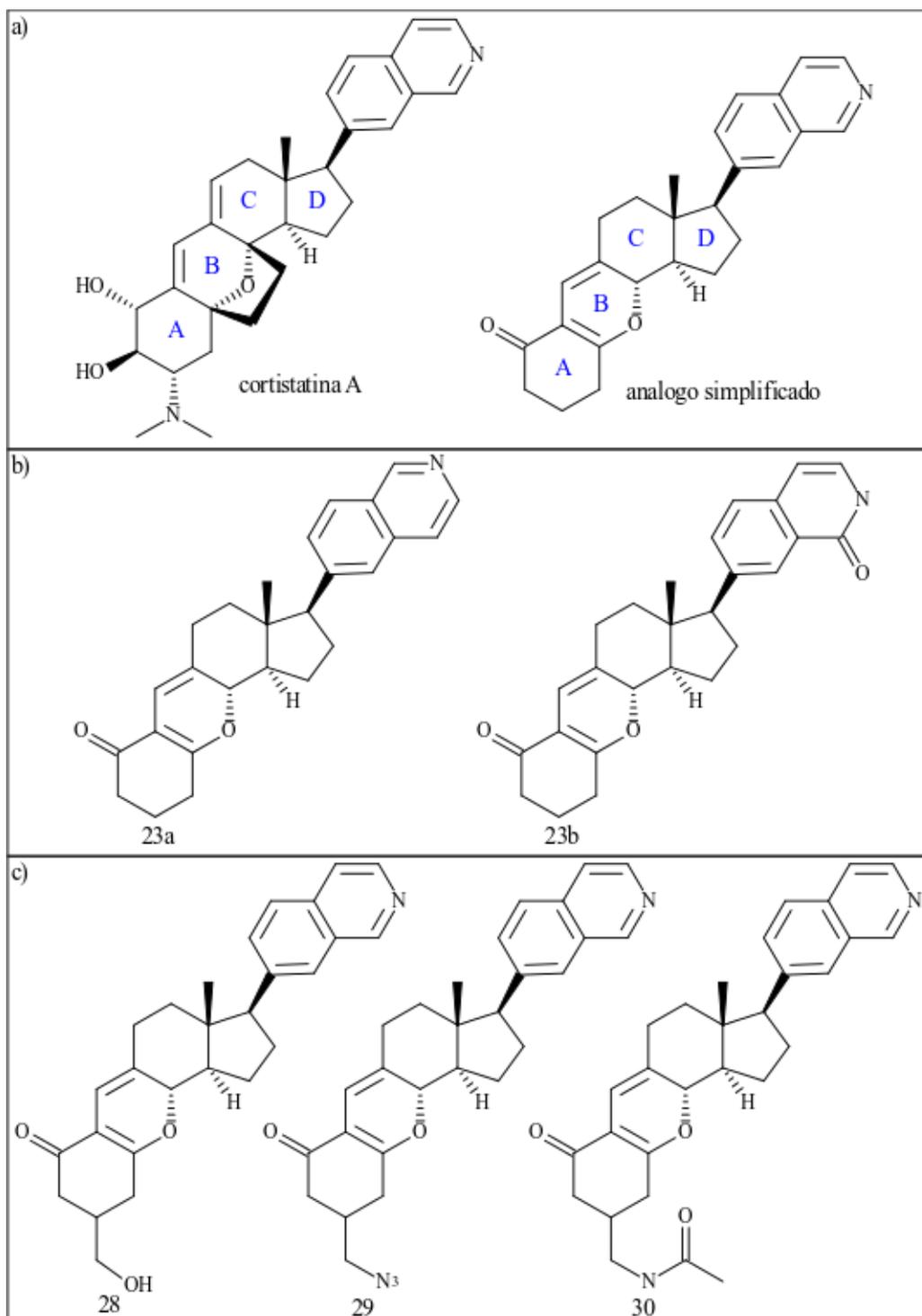


Figura 3 – Exemplo de moléculas sintetizadas para estudo SAR a partir de um produto natural.

Fonte: Adaptado de Kotoku *et al.* (2017, p. 1343-1344).

Legenda: a) estrutura da cortistatina A e seu análogo simplificado;

b) análogos com modificações no anel isoquinolinil;

c) análogos com adição de substituintes no anel A.

2. 2 BIOISOSTERISMO

O bioisosterismo constitui uma das estratégias empregadas por químicos medicinais para a modificação de compostos a fim de potencializar sua atividade. O termo bioisosterismo está relacionado a subunidades moleculares que apresentam propriedades físico-químicas em comum, tais como volume molecular, distribuição eletrônica, pKa, logP, e devido a essa similaridade, podem apresentar também a mesma atividade biológica (BARREIRO E FRAGA, 2008). Tendo em vista esse conceito, uma molécula líder pode ter um de seus fragmentos substituído por outro, o que pode modular suas propriedades farmacocinéticas e/ou farmacodinâmicas. Um exemplo é o grupamento ácido carboxílico que possui como bioisósteros um conjunto de subunidades apresentadas na figura 4.

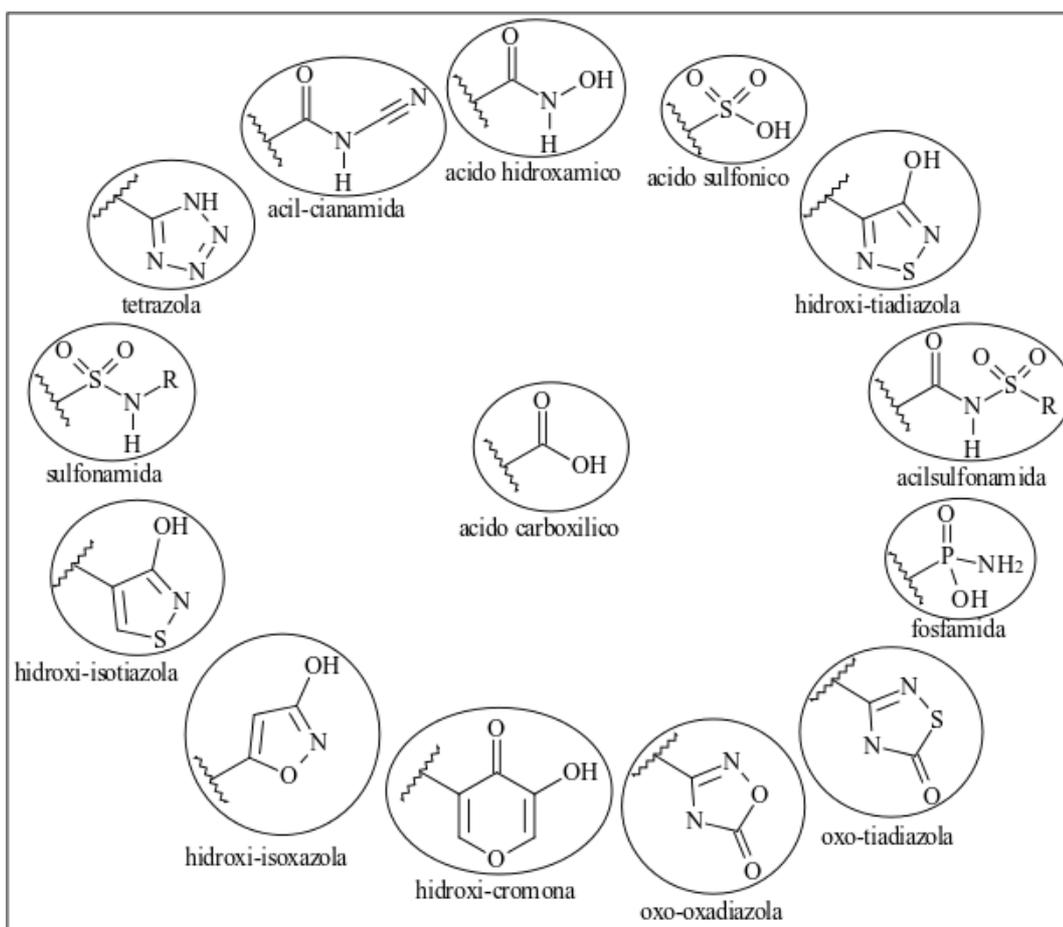


Figura 4 – Bioisósteros do grupamento ácido carboxílico.

Fonte: Adaptado de Barreiro e Fraga (2008, p. 278).

Os fundamentos do bioisosterismo se iniciam em 1919, quando o pesquisador Langmuir estudava o comportamento químico de grupos de moléculas que apresentavam

mesmo número de elétrons na camada de valência. Como resultado, observou-se que estas substâncias apresentavam propriedades físico-químicas em comum. Esse comportamento foi definido como isosterismo, e pares de entidades químicas apresentando essa característica são isósteras entre si (LANGMUIR, 1919). Em 1925 Grimm estabeleceu uma regra empírica, chamada regra do hidreto, segundo a qual, ao se adicionar um hidreto a um átomo, este conjunto (átomo + hidreto) se comporta de maneira similar ao átomo seguinte presente na tabela periódica (GRIMM, 1925). A figura 5 apresenta de forma esquemática a regra do hidreto formulada por Grimm.

Nº elétrons	6	7	8	9	10	11
	C	N	O	F	Ne	Na ⁺
	H ⁻ ↘	CH	NH	OH	FH	
		H ⁻ ↘	CH ₂	NH ₂	OH ₂	FH ₂ ⁺
			H ⁻ ↘	CH ₃	NH ₃	OH ₃ ⁺
				H ⁻ ↘	CH ₄	NH ₄ ⁺

Figura 5 – Regra do hidreto formulada por Grimm.

Fonte: Adaptado de Lima e Barreiro (2005, p. 24).

Em 1951, Friedman amplia o conceito de isosterismo, trazendo o termo bioisosterismo, o qual é aplicado para grupos de moléculas, que além de serem isósteras, também possuem atividade biológica em comum, uma vez que interagem com o mesmo sítio receptor. Quanto à atividade, porém, esta pode ser equivalente ou antagônica (FRIEDMAN, 1951). Em 1979, Thornber define o termo bioisósteros como subunidades, grupos ou moléculas que possuem propriedades físico-químicas que lhes garantem efeito biológico similar (THORNBUR, 1979).

Pode-se classificar o bioisosterismo como clássico e não clássico (BURGER, 1991). O bioisosterismo clássico baseia-se na valência em comum dos grupamentos bioisostéricos, nessa categoria inclui-se também os anéis equivalentes. Assim tem-se:

- Átomos e grupos monovalentes;
- Átomos e grupos divalentes;
- Átomos e grupos trivalentes;
- Átomos e grupos tetravalentes;
- Anéis equivalentes.

A zidovudina, ou AZT, é um fármaco utilizado no tratamento da AIDS, atuando sobre a enzima transcriptase reversa (EC: 2.7.7.49) do vírus HIV. Constitui um bioisómero do tipo monovalente em relação ao nucleosídeo timina, no qual um grupamento -OH está substituído por uma azida (FIGURA 6) (SWAPNA *et al.*, 1989).

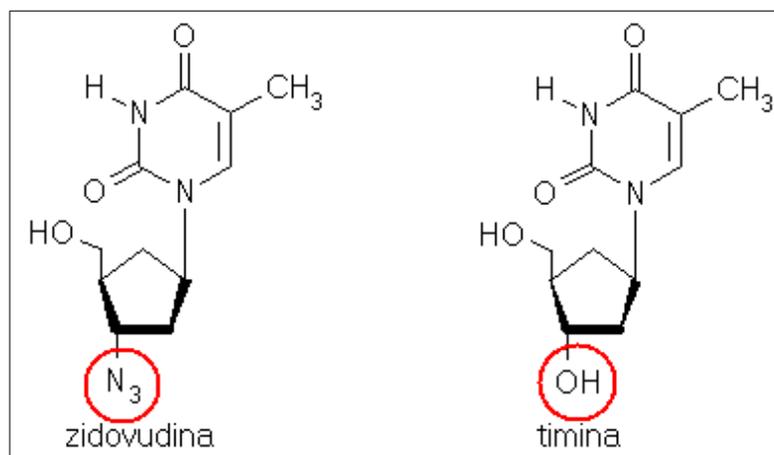


Figura 6 – Fármaco zidovudina, um análogo do nucleosídeo timina.
Fonte: Do autor.

A substituição de um átomo de hidrogênio por um flúor é amplamente empregada no desenvolvimento de fármacos antivirais e anti-tumorais. Tais fármacos são análogos de nucleosídeos naturais, de forma a inibir a síntese de DNA (MENG E QING, 2006). O flúor pode mimetizar o hidrogênio em virtude de seu raio de van der Waals ser de 1,35 Å, enquanto que para o hidrogênio tem-se um raio de van der Waals de 1,10 Å, tendo um mínimo aumento em efeitos estéricos (DUSCHINSKY, PLEVEN E HEIDELBERGER, 1957). Entretanto, a maior eletronegatividade do flúor, e a impossibilidade da remoção de um íon F⁺ da molécula, em contraste com a fácil remoção de íons H⁺ – comumente encontrada em reações enzimáticas – fazem com que nucleosídeos fluorados sejam resistentes à ação enzimática, bloqueando a rota biossintética (KIRK, 2006). Esse mecanismo é encontrado no fármaco anticâncer 5-fluorouracil, análogo do uracil, precursor do timidilato (HEIDELBERGER *et al.*, 1957). No organismo, o 5-fluorouracil é convertido em ácido 5-fluoro-2'-deoxiuridílico, que por sua vez inibe a enzima timidilato sintase por ligação covalente (FIGURA 7).

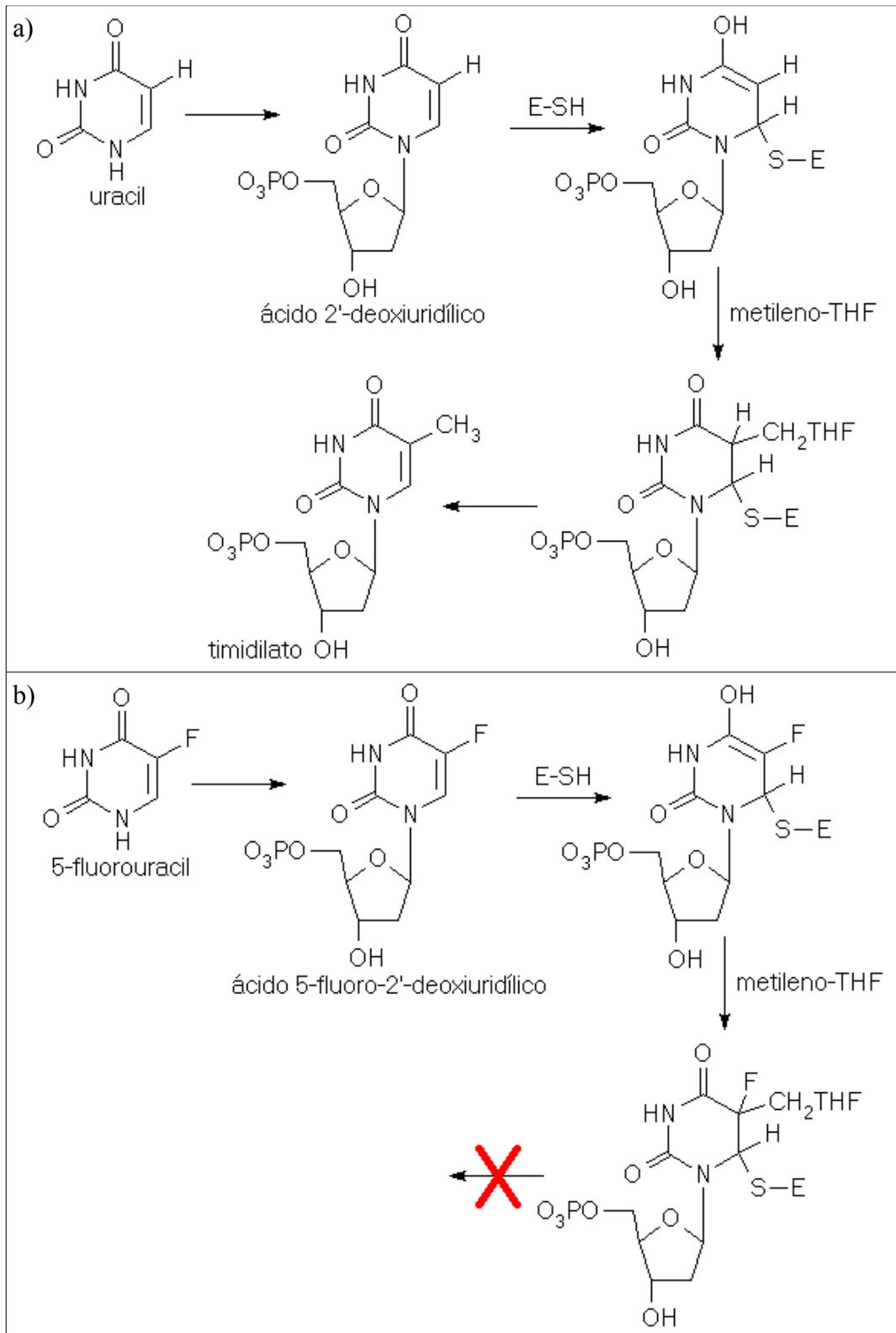


Figura 7 – Mecanismo de inibição da enzima timidilato sintase pelo fármaco 5-fluorouracil.

Fonte: Adaptado de Patani e LaVoie (1996, p. 3149).

Legenda: a) rota biossintética normal;

b) inibição com o 5-fluorouracil. E-SH: enzima timidilato sintase; metileno-THF: ácido 5,10-metilenotetrahidrofólico.

A relação bioisostérica de grupos divalentes é observada entre os fármacos procaína e procainamida (FIGURA 8). A procaína é utilizada como anestésico local, enquanto a procainamida é utilizada no tratamento da arritmia cardíaca. A diferença estrutural entre essas duas moléculas consiste na substituição de um átomo de oxigênio na procaína por um grupo NH na procainamida, resultando em uma transformação éster \rightarrow amida, fazendo a procainamida ser quatro vezes mais estável *in vivo* que a procaína (LOFTSSON, 2014).

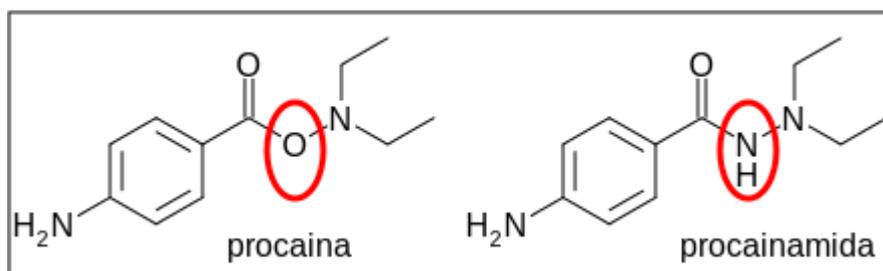


Figura 8 – Relação bioisostérica entre os fármacos procaína e procainamida.
Fonte: Do autor.

O bioisosterismo de anéis equivalentes é exemplificado pela relação bioisostérica entre o anel fenila e o tiofeno. O fármaco tenoxicam foi desenvolvido baseado nessa estratégia, conforme figura 9, seu precursor é o fármaco piroxicam. Ambas as moléculas agem sobre a enzima ciclooxigenase (EC: 1.14.99.1), tendo aplicação para o tratamento da artrite e osteoartrite, entretanto, o fármaco tenoxicam, com o grupo tiofeno, possui maior tempo de meia vida no plasma sanguíneo, permitindo a administração de doses menores de 20 mg apenas uma vez ao dia (BINDER *et al.*, 1987).

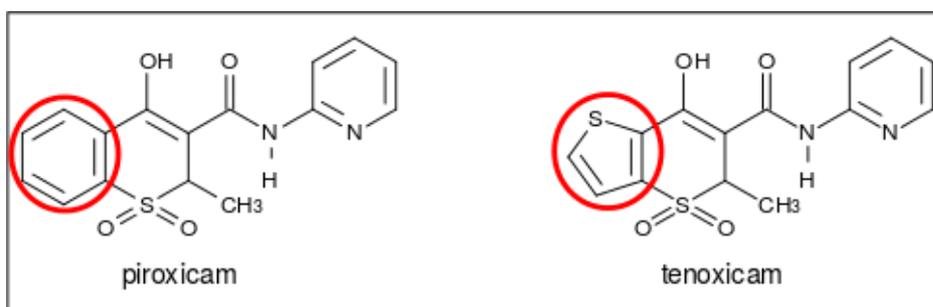


Figura 9 – Desenvolvimento do fármaco tenoxicam por bioisosterismo de anéis.
Fonte: Adaptado de Lima e Barreiro (2005, p. 32).

O bioisosterismo não-clássico inclui grupos funcionais com propriedades equivalentes, retro-isosterismo, bióforos isostéricos, anelação e retro-anelação (BARREIRO E FRAGA, 2008). Um exemplo de bioisosterismo não-clássico para grupos funcionais equivalentes envolve a substituição da hidroxila de uma função fenol pelo grupamento $-NHSO_2CH_3$

(FIGURA 10). Larsen e Lish sintetizaram uma série de fenetanolaminas com o grupo alquilsulfonamida acoplado ao anel benzênico, alguns dos quais apresentaram atividade sobre receptores adrenérgicos (FIGURA 10a) (LARSEN E LISH, 1964). Em outro estudo, feito por Lin e colaboradores, o mesmo tipo de substituição foi realizado agora em moléculas com atividade antagonista do hormônio gonadotropínico, sendo que o derivado sulfonamídico apresentou melhor desempenho (FIGURA 10b) (LIN *et al.*, 2001).

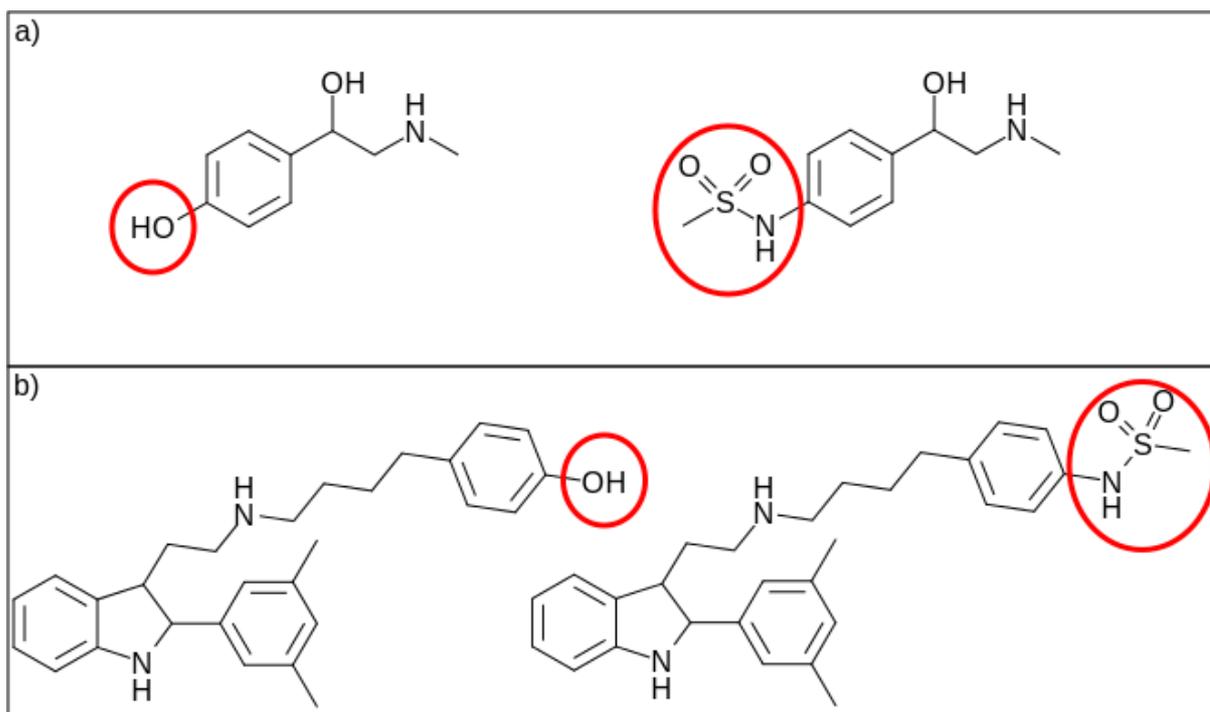


Figura 10 – Bioisosterismo não-clássico envolvendo a troca de hidroxila de grupo fenol pelo grupamento $-NHSO_2CH_3$.

Fonte: Do autor.

Legenda: a) compostos com atividade sobre receptores adrenérgicos;

b) compostos com atividade antagonista do hormônio gonadotropínico.

Fragmentos bióforos são aqueles capazes de interagir com o sítio receptor de uma biomacromolécula por meio de um mecanismo similar, como é o caso dos aceptores de ligação-H, exemplificados pelos fragmentos tetrazola, carboxilato e acilsulfonamida (FIGURA 11) (BARREIRO E FRAGA, 2008).

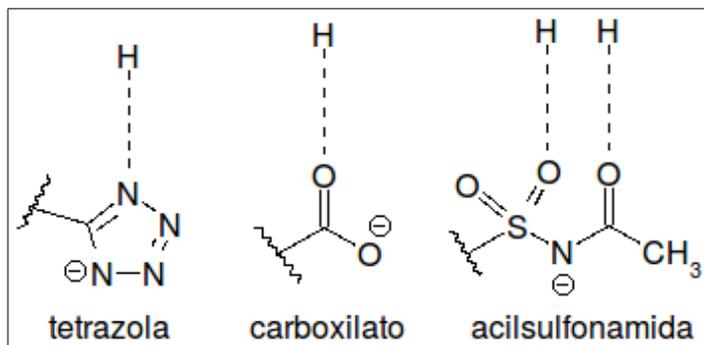


Figura 11 – Exemplo de bioisosterismo não-clássico por bióforos.
Fonte: Adaptado de Barreiro e Fraga (2008, p. 279).

A anelação ocorre quando uma estrutura de cadeia aberta é substituída por outra possuindo um anel. A restrição conformacional oriunda da anelação terá efeitos positivos caso a conformação ativa do bioisómero de cadeia aberta seja similar à apresentada pelo bioisómero anelado. Esta relação é observada entre os fármacos lidocaina e mepivacaina, o primeiro usado clinicamente a partir de 1948 e o segundo em 1957, ambos utilizados como anestésicos locais (FIGURA 12) (RUETSCH, BÖNI E BORGEAT, 2001). Embora possuam propriedades equivalentes, mepivacaina possui ação um pouco mais longa.

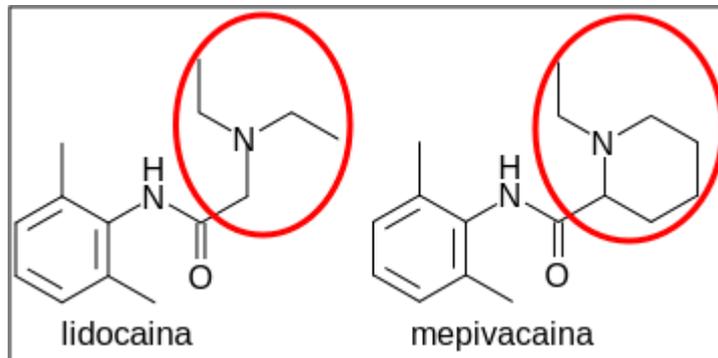


Figura 12 – Relação bioisostérica de anelação entre os fármacos lidocaina e mepivacaina.

Fonte: Do autor.

O bioisosterismo pode ser empregado no intuito de proteger grupos sensíveis aos processos metabólicos do organismo, como é o caso de moléculas que apresentam a função éster, facilmente clivada por esterases plasmáticas (EC: 3.1.1.1), porém, a substituição da função éster por um grupo oxazólico, previne esta metabolização (FIGURA 13a) (KIM *et al.*, 2002). Outro caso são moléculas com grupamento amida, alvo da ação de amidases (EC: 3.5.1.4); a troca de grupamento amida por uma pirrolo produz uma molécula resistente à degradação por amidases (FIGURA 13b) (VAN-WIJNGAARDEN *et al.*, 1987).

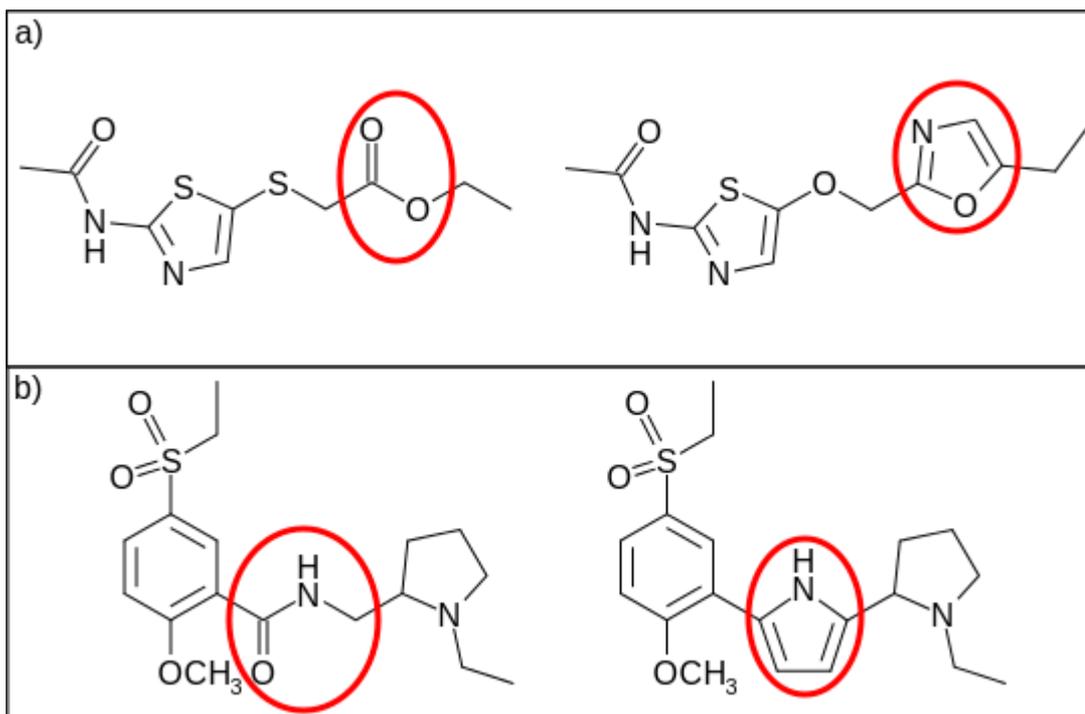


Figura 13 – Bioisosterismo usado como forma de proteger a molécula da ação metabólica do organismo.

Fonte: Do autor.

Legenda: a) substituição de grupamento éster por um grupo oxazólico;
b) substituição de uma amida por uma pirrola.

2. 3 TRABALHOS ENVOLVENDO O EMPREGO DE BIOISOSTERISMO

Lu e colaboradores (2016), em busca de substâncias não-nucleosídicas capazes de inibir a enzima transcriptase reversa (EC: 2.7.7.49) do vírus HIV, sintetizaram uma série de arilazolil(azinil)tioacetanilidas tendo como compostos líderes as estruturas VRX-480733 e RDEA-806 (FIGURA 14), as quais estavam em testes clínicos de fase II em 2008. A principal modificação dos novos compostos foi a substituição do anel azola de cinco membros por uma purina. 21 moléculas foram sintetizadas, variando-se substituintes em quatro posições específicas. O composto LAD-8 apresentou melhor atividade contra o vírus HIV tipo selvagem ($EC_{50} = 0,78 \mu M$), enquanto que LBD-6 mostrou atividade moderada contra os mutantes L100I ($EC_{50} = 5,64 \mu M$) e RES056 ($EC_{50} = 22,24 \mu M$).

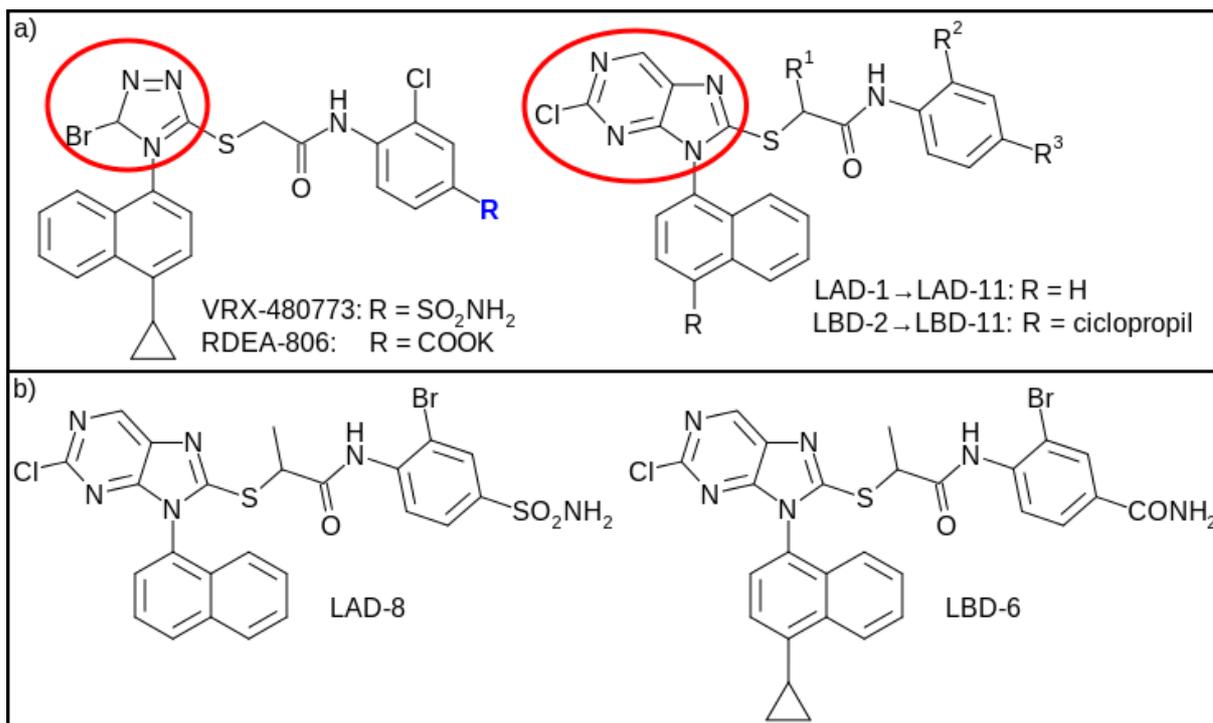


Figura 14 – Bioisosterismo de anel com substituição de anel azola por purina.

Fonte: Adaptado de Lu *et al.* (2016, p. 4425-4426).

Legenda: a) compostos VRX-480773 e RDEA-806 (protótipos retrovirais) e estrutura geral dos novos compostos sintetizados por Lu *et al.*

b) estrutura das moléculas LAD-8 e LBD-6, as quais apresentaram melhor atividade contra vírus HIV tipo selvagem e mutante, respectivamente.

Glicoproteína-P (EC: 7.6.2.2) é uma proteína de membrana responsável pelo transporte de moléculas nas células, tem a função de absorção e excreção de substâncias. A super-expressão desta proteína no cérebro faz com que fármacos antineoplásicos não consigam atingir o tumor. A fim de modular a atividade da glicoproteína-P, Capparelli e colaboradores (2014) estudaram derivados de tetrahydroisoquinolinas como potenciais inibidores. Partindo da estrutura do tariquidar, um fármaco em testes clínicos, foram sintetizadas moléculas menores com estrutura simplificada. Dentre estas, o composto MC113 apresentou boa atividade inibitória *in vitro*, porém baixa atividade *in vivo*. Por modificações bioisostéricas, foram sintetizados análogos de MC113, no qual o fragmento bifenil foi substituído por um grupamento 2-irilfurano. O derivado 24a foi o que apresentou melhor atividade (EC₅₀ = 0,91 μM) (FIGURA 15).

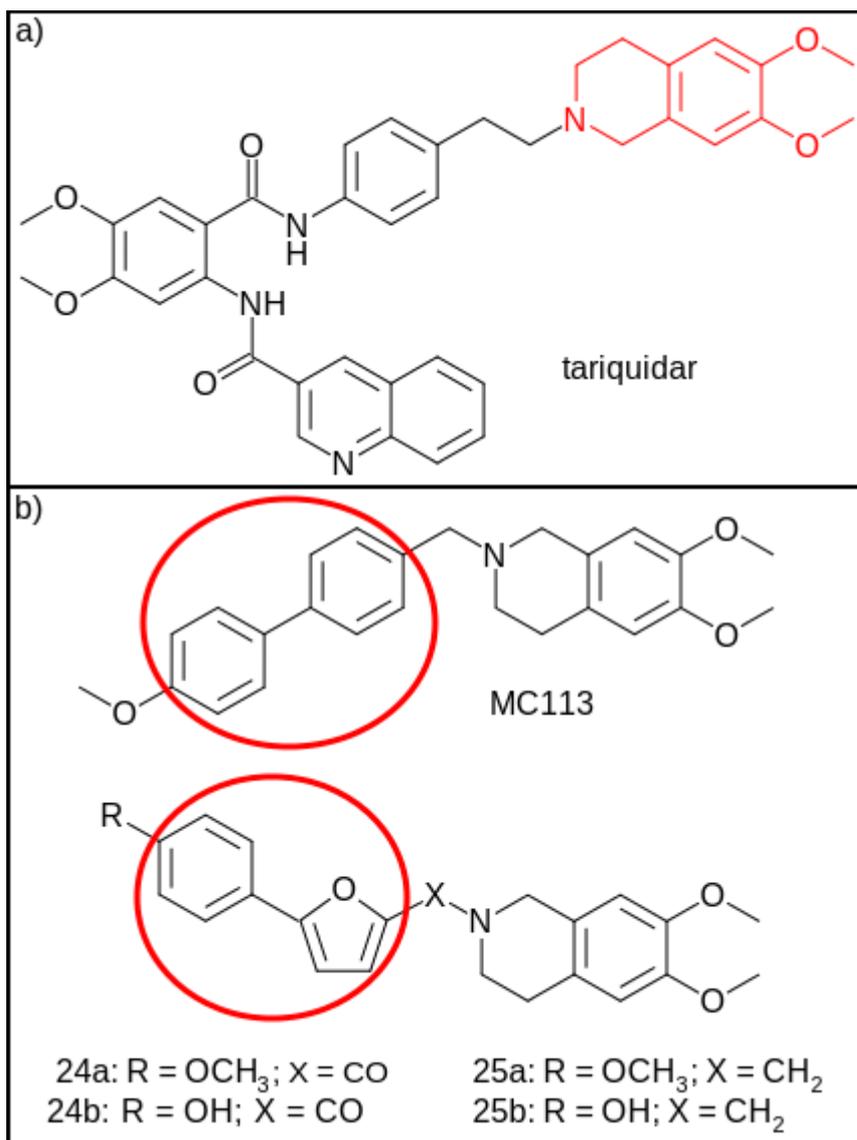


Figura 15 – Bioisosterismo por substituição de bifenil por 2-arilfurano.

Fonte: Adaptado de Capparelli *et al.* (2014, p. 9984).

Legenda: a) estrutura do tariquidar, em vermelho o farmacóforo presente nas estruturas simplificadas;

b) estrutura do composto MC113 e seus análogos.

Diamidas antranílicas constituem uma família de moléculas usadas como pesticidas no combate a insetos em lavouras. Estas atuam sobre um sítio alostérico dos receptores de rianodina de insetos, mas não de mamíferos, ativando a liberação de cálcio para contração muscular. Diversos estudos tem sido realizados em busca de análogos mais potentes dessa classe de moléculas (LIU *et al.*, 2017). A partir da estrutura do pesticida já em uso no mercado clorantraniliprole, foram desenvolvidos 29 compostos, no qual o grupamento amida foi substituído por um anel 1,2,4-oxadiazola, que por sua vez recebeu diferentes substituintes R; também realizou-se a substituição do cloro ligado ao anel benzênico por um átomo de bromo ou hidrogênio para alguns dos análogos. O composto 3III foi o mais promissor, com 90% de

atividade larvicida a concentração de 0,5 mg L⁻¹ e LC₅₀ de 0,20 mg L⁻¹ (FIGURA 16).

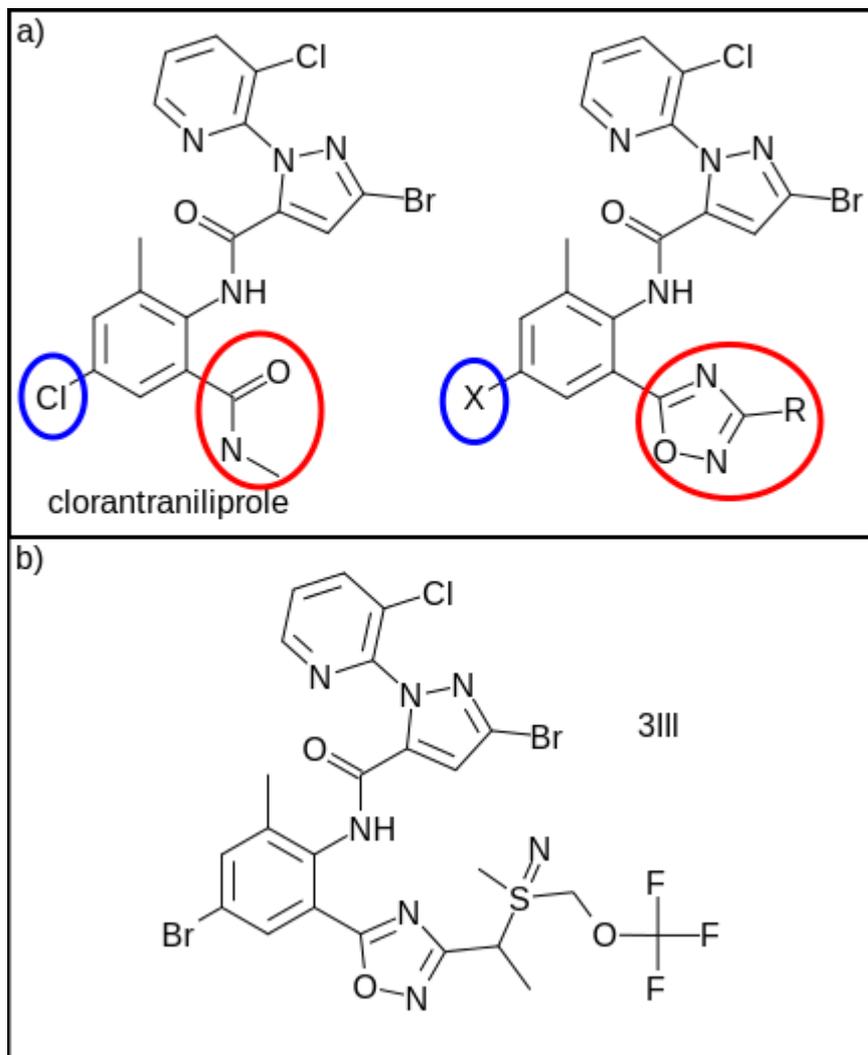


Figura 16 – Bioisosterismo por substituição de amida por 1,2,4-oxodiazola.
Fonte: Adaptado de Liu *et al.* (2017, p. 918-919).

Legenda: a) estrutura do inseticida clorantraniliprole usado como composto protótipo para a síntese de inseticidas mais potentes e estrutura geral dos novos compostos sintetizados por Liu *et al.*;

b) composto 3III, o qual apresentou melhor atividade.

Em tumores sólidos, a carência de glicose aumenta o potencial invasivo e metastático de células tumorais. Arctigenina é um produto natural obtido de sementes de *Arctium lappa*, possuindo atividade contra células cancerígenas em estado de carência de glicose. O grupamento 4-hidroxil da arctigenina é importante para seu efeito citotóxico. A fim de avaliar mais a fundo a contribuição do grupo 4-hidroxil para a atividade da arctigenina, Lei e colaboradores (2015) sintetizaram análogos no qual esse grupamento foi trocado por substituintes aminados. O análogo mais simples, no qual o substituinte era uma amina, apresentou atividade melhor que o composto original, com IC₅₀ de 2,85 μM, contra IC₅₀ de

6,57 μM , respectivamente. Análogos com substituintes maiores, em que um dos hidrogênios da amina foi modificado para uma cadeia alquílica ou carbonílica, não apresentaram atividade. Também sintetizou-se um análogo em que o grupo 4-hidroxil apenas foi removido, sem adição de outro substituinte, de forma similar não houve atividade (FIGURA 17).

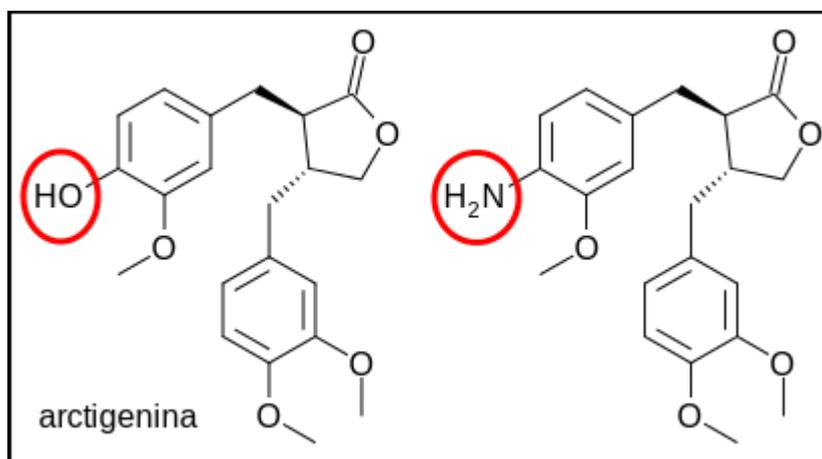


Figura 17 – Bioisosterismo clássico envolvendo substituição de hidroxila por amina para o produto natural arctigenina.

Fonte: Adaptado de Lei *et al.* (2015, p. 435-436).

A síndrome metabólica é uma condição que envolve fatores de risco como dislipidemia aterogênica, pressão sanguínea elevada, hiperglicemia, estado pró-trombótico e estado pró-inflamatório, podendo conduzir a doenças cardiovasculares e diabetes. A família de receptores PPAR (*Peroxisome Proliferator-Activated Receptors*) foi identificada com alvos moleculares promissores para o desenvolvimento de fármacos para o tratamento de síndrome metabólica. Moléculas com atividade agonista dual para os receptores PPAR α e PPAR γ atuam simultaneamente na correção da resistência à insulina e no balanço de lipídios, entretanto, os melhores resultados são apontados para moléculas com atividade agonista total para receptores PPAR α e atividade agonista parcial para receptores PPAR γ , em vista da redução de efeitos colaterais. Em estudos anteriores, o composto LT175 foi identificado como *hit*, possuindo atividade dual frente aos receptores PPAR α e PPAR γ , com agonismo parcial para o subtipo γ (PIEMONTESE *et al.*, 2015). Isso motivou a investigação de um novo composto com variação bioisostérica, com adição de um grupo oximo no lugar de um anel aromático (FIGURA 18). A partir deste, foram sintetizados 13 novos compostos com acréscimo de substituintes em dois pontos específicos do grupamento oximo, sendo um o hidrogênio ligado ao carbono e outro o hidrogênio ligado ao oxigênio. O composto (S)-3 apresentou resultado mais interessante, com EC_{50} de 3,8 μM frente a ambos receptores PPAR α e PPAR γ .

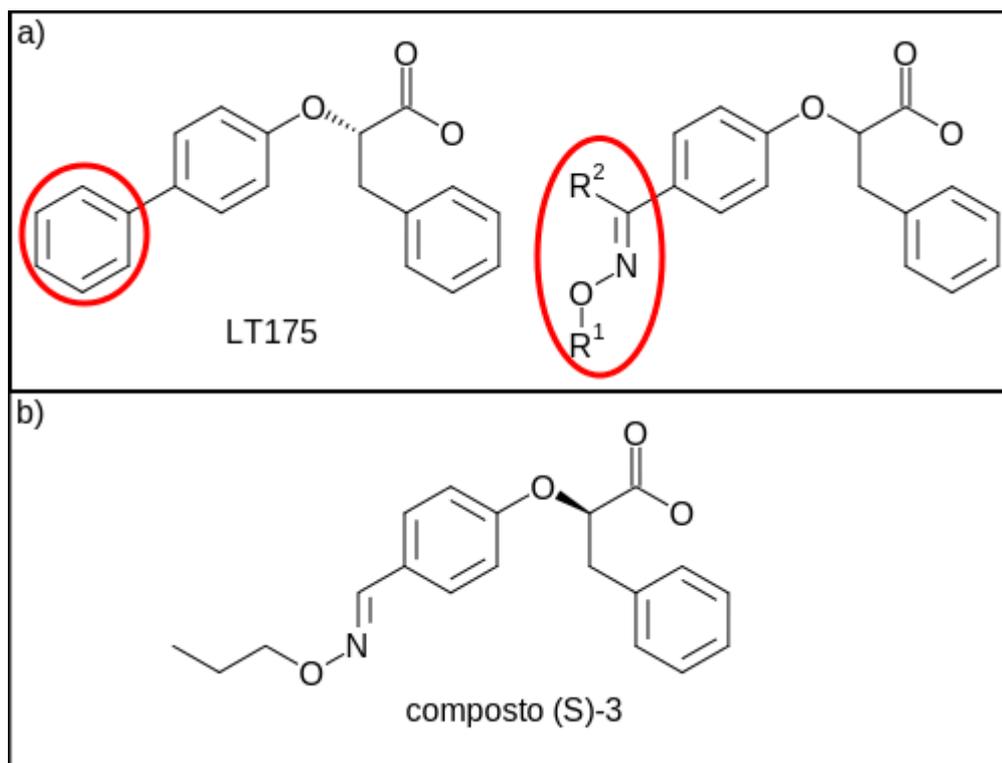


Figura 18 – Bioisosterismo não-clássico de substituição de anel benzênico por grupamento oximo.

Fonte: Adaptado de Piemontese *et al.* (2015, p. 584).

Legenda: a) composto LT175 e estrutura geral de análogos sintetizados por Piemontese *et al.*;
b) composto (S)-3, o qual apresentou melhor atividade.

A ação agonista sobre receptores nicotínicos de acetilcolina tem sido estudada para aplicação na terapia de distúrbios do sistema nervoso, como o vício do tabaco e álcool, e depressão. O receptor do subtipo $\alpha 4\beta 2$ é um dos mais pesquisados como alvo terapêutico. Em vista de encontrar um agonista potente para o receptor $\alpha 4\beta 2$ e ao mesmo tempo seletivo em relação ao subtipo $\alpha 3\beta 4$, Bolchi e colaboradores (2016) desenvolveram uma série de ligantes tendo como partida o composto A-84543, descoberto como *hit* em pesquisas anteriores. A estratégia adotada foi alterar o anel piridínico por um anel benzênico, além de adicionar substituintes nas posições meta desse anel. Os compostos possuindo um substituinte 6-hidroxil-1-hexinil numa das posições meta e um substituinte hidroxil ou nitróxido na outra posição meta apresentaram os melhores resultados, com razão de interação $\alpha 3\beta 4$ Ki / $\alpha 2\beta 2$ Ki de 131 e 85, respectivamente. Por outro lado, o composto possuindo apenas o substituinte 6-hidroxil-1-hexinil não apresentou atividade e seletividade favorável (FIGURA 19).

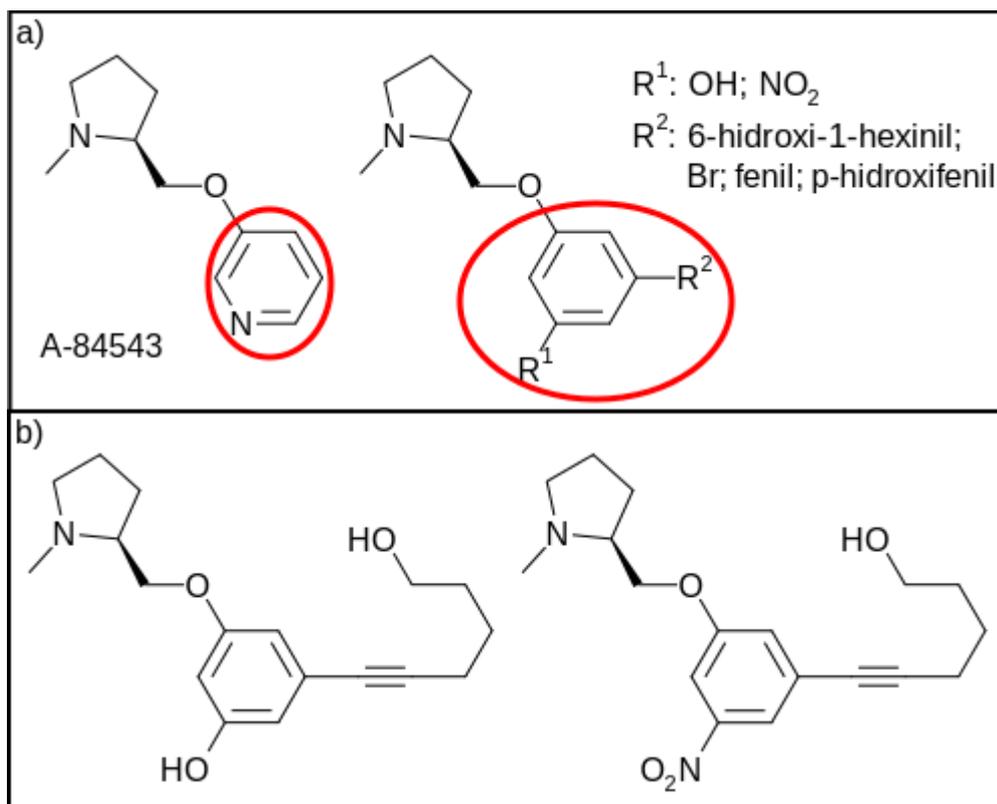


Figura 19 – bioisosterismo de anéis com substituição de anel piridínico por anel benzênico com substituintes em posição meta.

Fonte: Adaptado de Bolchi *et al.* (2016, p. 5614, 5616).

Legenda: a) composto *hit* A-84543 em paralelo com a estrutura geral dos análogos bioisostéricos;

b) análogos com substituinte 6-hidroxi-1-hexenil em R^2 , apresentaram melhor seletividade frente aos receptores $\alpha 3\beta 4$ / $\alpha 2\beta 2$.

A doença de chagas é causada pelo protozoário *Trypanosoma cruzi*. Sem vacina disponível, é tratada com o fármaco benznidazola, apresentando baixa eficácia na fase crônica sintomática, além de severos efeitos colaterais. Baseado em estudos prévios demonstrando o potencial antiparasitário de tiosemicarbazonas frente ao *T. cruzi*, Oliveira Filho e colaboradores (2015) sintetizaram uma série de compostos com potencial atividade tripanomicida. Selecionou-se o composto *hit* 3,4-diclorofenil-tiosemicarbazona, identificado como inibidor de cruzaina (EC: 3.4.22.51) (FIGURA 20a). Partindo dos resultados de estudos de SAR indicando que o substituinte 3,4-diclorofenil é fundamental para a ação tripanomicida e devido à baixa drogabilidade de tiosemicarbazonas, substituiu-se esta por anéis tiazolidinonas. O composto **20**, com substituinte fenil em N3 e isopropil em C5, apresentou melhor atividade antiparasitária com IC₅₀ de 1,7 ± 0,17 μM, mais eficiente que o controle benznidazola (IC₅₀ de 10,6 ± 0,87 μM). Além disso, não apresentou citotoxicidade em concentração superior a 100 μM frente a linhagem de macrófagos J774 (FIGURA 20b).

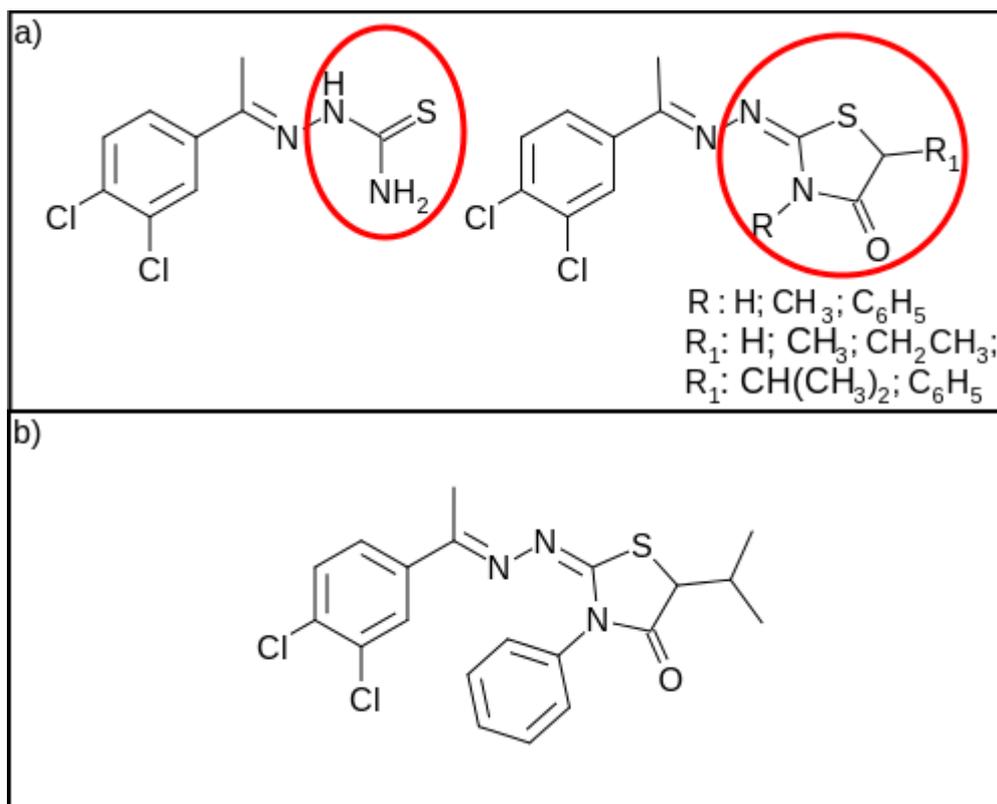


Figura 20 – Relação bioisostérica de tiosemicarbazona e tiazolidinonas com atividade antiparasitária.

Fonte: Adaptado de Oliveira Filho *et al.* (2015, p. 7479).

Legenda: a) composto 3,4-diclorofenil-tiosemicarbazona com atividade tripanomicida e estrutura geral de seus análogos bioisostéricos;

b) composto com melhor atividade tripanomicida frente ao T. Cruzi.

Anidrase carbônica (EC: 4.2.1.1) é uma metaloenzima envolvida na conversão de dióxido de carbono e água em ácido carbônico. Está associada a algumas patologias como cancerogênese e progressão. O composto SLC-0111, possuindo um grupamento ureído em sua estrutura, entrou em fase clínica I para o tratamento de tumores onde se observa a superexpressão de anidrase carbônica isoforma IX. Tendo em vista as propriedades antioxidantes do selênio e utilizando princípios do bioisosterismo clássico de grupos divalentes, Angeli e colaboradores (2017) desenvolveram 15 análogos do SLC-0111 para os quais o oxigênio do grupamento ureído foi substituído por um átomo de selênio; para uma análise SAR mais detalhada, também sintetizaram e avaliaram o bioisóstero com a substituição do oxigênio por enxofre (FIGURA 21). Os compostos foram testados contra as anidrases carbônicas de isoforma I, II, IV e IX, sendo comparados com o fármaco de referência acetazolamida. Ambos bioisósteros contendo selênio ou enxofre apresentaram melhor interação com as anidrases carbônicas isoforma I e II quando comparados ao protótipo SLC-0111; porém a interação foi reduzida para a anidrase carbônica isoforma IV e somente o bioisóstero sulfurado demonstrou melhoria de interação com a isosforma IX (TABELA 1).

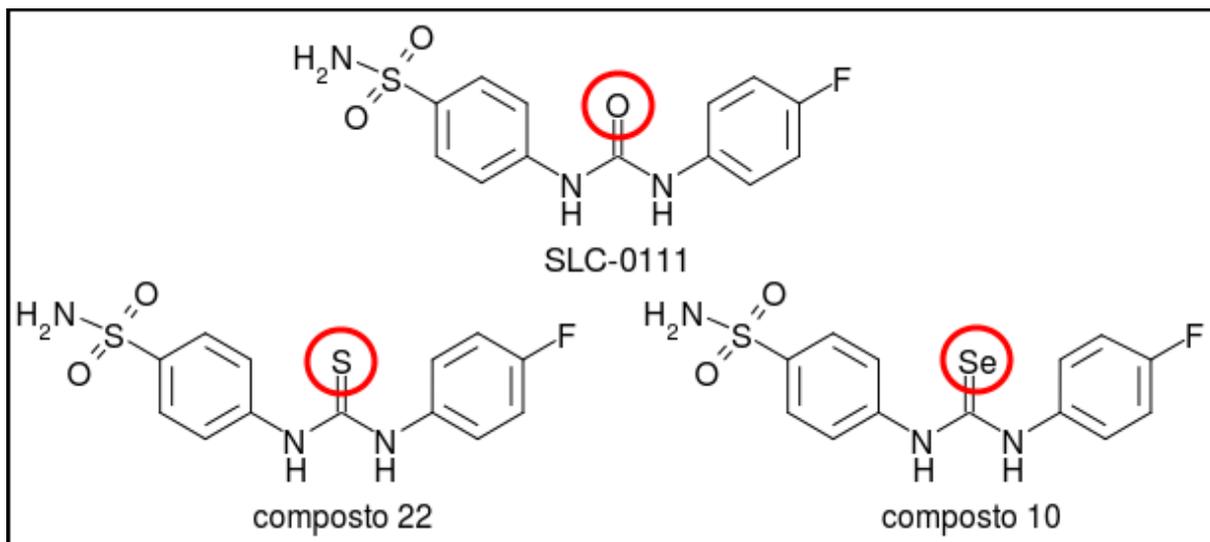


Figura 21 – Bioisosterismo clássico divalente envolvendo substituição de calcogênios.

Fonte: Adaptado de Angeli *et al.* (2017, p. 963).

Tabela 1 – Inibição de hCA I, II, IV e IX dos compostos 10, 22 e SLC-0111 por ensaio de parada de fluxo de CO₂.

composto	K_1 (nM)			
	hCA I	hCA II	hCA IV	hCA IX
10	152,3	66,3	7557	63,0
22	35,0	14,2	4797	32,1
SLC-0111	5080	960,0	286,0	45,0

Fonte: Adaptado de Angeli *et al.*, 2017.

Wondonin é um produto natural obtido de esponjas das espécies *Poecillatra wondoensis* e *Jaspis sp.* Apresenta atividade antiangiogênica por inibição do fator de crescimento endotelial vascular (VEGF - *Vascular Endothelial Growth Factor*, Fator de Crescimento Endotelial Vascular), sem citotoxicidade aparente, tornando-se um agente terapêutico em potencial para doenças como artrite e câncer. Entretanto, apresenta grupos químicos instáveis e de difícil síntese. Para contornar esses problemas, Yu e colaboradores (2017) propuseram uma série de modificações bioisostéricas partindo da molécula original (FIGURA 22). A primeira transformação foi a substituição do anel imidazol pelo anel 1,2,3-triazol e a troca do grupo instável vinil sulfato por etil (**I**). Em seguida, foram sintetizados análogos em que o anel acetal foi substituído por heterociclos aromáticos, tais como benzofurano, benzotiazol e benzoxazol. O análogo com benzotiazol apresentou melhor potência e foi selecionado como novo protótipo (**II**). As hidroxilas do grupamento catecol foram substituídas por metoxilas a fim de proteger grupos sensíveis ao metabolismo (**III**). Por fim, foram testados diversos substituintes no grupamento N-dimetil, resultando no composto

final com substituinte piperidina (IV), o qual apresentou o maior índice de seletividade de até 75, tendo $IC_{50} = 0,631 \mu M$ em ensaios de formação de tubos induzido por VEGF.

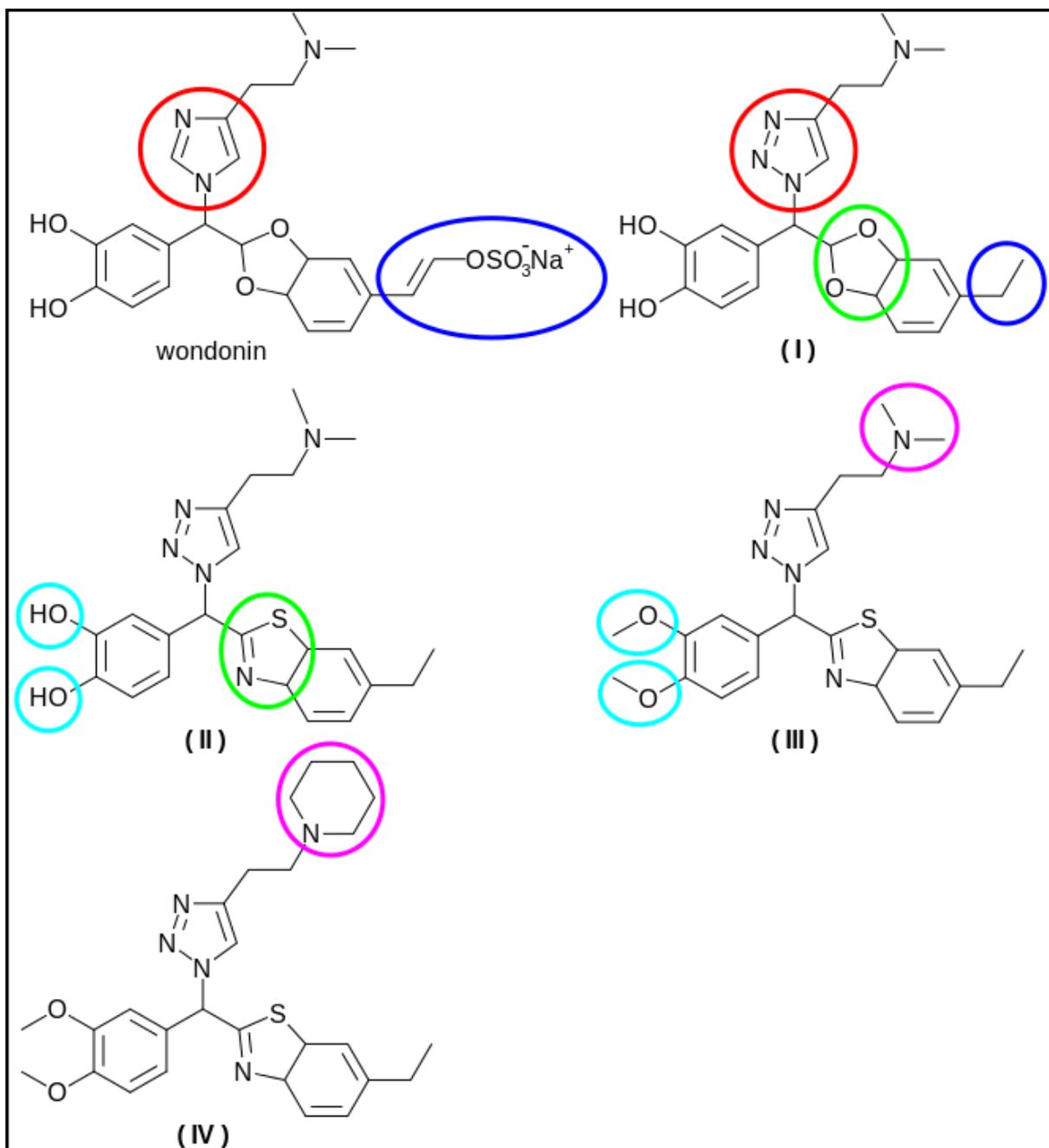


Figura 22 – Desenvolvimento de novos antiangiogênicos a partir do produto natural wondonin por meio de modificações bioisostéricas.

Fonte: Adaptado de Yu *et al.* (2017, p. 1066-1067).

Com o objetivo de investigar a relação bioisostérica entre cloro e metil como substituintes em cadeias alifáticas saturadas, Krautwald e colaboradores (2016) utilizaram o composto atpenin A5 como modelo experimental. Atpenin é um agente antifúngico capaz de inibir o complexo II da cadeia respiratória mitocondrial (EC: 1.3.5.1), possui uma cadeia

alifática com substituintes cloro e metil (FIGURA 23). Foram sintetizados dois análogos, um em que foi acrescido um terceiro átomo de cloro e outro em que todos os átomos de cloro foram trocados por grupamentos metil. A ação inibitória destes foi testada contra enzimas respiratórias do nematoide *Ascaris suum* e contra enzimas respiratórias de coração bovino. Para a enzima NADH-fumarato redutase (EC: 1.3.1.6) do nematoide, o IC_{50} foi de 4,1 nM para o análogo com cloro e 4,0 nM para o análogo com metil, enquanto para a enzima bovina succinato-citocromo c redutase (EC: 1.9.3.1) o IC_{50} foi de 10 nM e 15 nM, respectivamente, demonstrando a relação bioisostérica cloro \rightarrow metil para o composto atpenin A5.

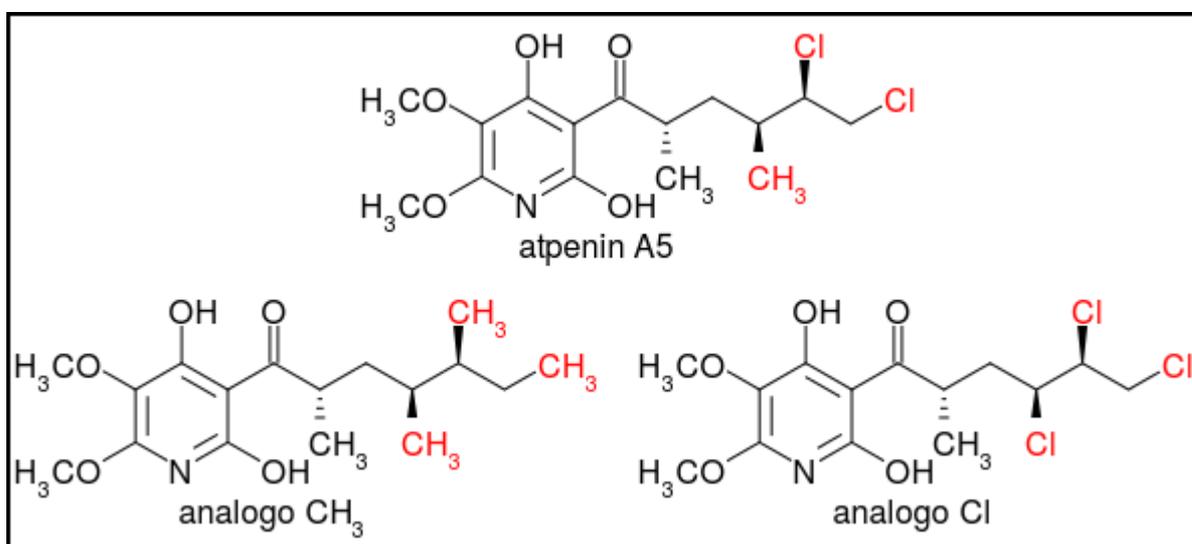


Figura 23 – Bioisosterismo clássico monovalente cloro \rightarrow metil para o composto atpenin A5, inibidor do complexo II da cadeia respiratória mitocondrial.

Fonte: Adaptado de Krautwald *et al.* (2016, p. 4049).

3 MATERIAIS E MÉTODOS

O programa MB-Isoster foi modelado utilizando-se o paradigma de orientação a objetos. As classes foram agrupadas em oito pacotes conforme listado a seguir:

- a) Pacote mol: possui classes responsáveis por construir a estrutura de uma molécula;
- b) Pacote io: suas classes são responsáveis pela leitura e escrita de arquivos moleculares;
- c) Pacote forcefield: implementa o campo de força MMFF94s (HALGREN, 1999) utilizado na minimização de energia molecular e cálculo de energia potencial;
- d) Pacote bioisostermaker: possui uma classe com a função de construir os bioisósteros;
- e) Pacote math: implementa as entidades matemáticas matriz e vetor;
- f) Pacote properties: possui classes responsáveis por calcular propriedades físico-químicas das moléculas, sendo estas: peso molecular, número de doadores e de aceptores de interação de hidrogênio, logP, logS e TPSA;
- g) Pacote virtualscreening: responsável por implementar *virtual screening* para os bioisósteros e um receptor selecionado pelo usuário, usando o programa AutoDock Vina como *engine de molecular docking*;
- h) Pacote gui: constrói a interface gráfica de usuário. A interface gráfica será apresentada na seção “4. RESULTADOS E DISCUSSÃO”.

O pacote mol define as seguintes classes:

- a) Atom: representa um átomo da molécula;
- b) Bond: representa uma ligação covalente entre dois átomos da molécula;
- c) Ring: representa um anel na molécula;
- d) RingSearch: utilitário cuja função é localizar os anéis presentes na molécula;
- e) Molecule: representa uma molécula completa;
- f) Residue: derivado de Molecule, é um resíduo lido a partir de arquivo pdb;
- g) Chain: consiste em uma cadeia macromolecular, podendo ser proteína, DNA, ou RNA; formada por uma sequência específica de resíduos lida a partir de arquivo pdb, podendo conter também ligantes cocrystalizados;
- h) Receptor: representa um receptor lido de arquivo pdb, com uma ou mais cadeias;
- i) Fragment: derivado de Molecule, é uma molécula que perdeu parte de sua estrutura por meio do rompimento de uma de suas ligações covalentes, a qual

passa a ser um radical podendo ser unido a outro fragmento;

j) Radical: representa um radical de um fragmento molecular.

O pacote `io` define as classes `IOMol`, utilizada para leitura e escrita de arquivos moleculares em formatos de arquivo comumente usados: `mol`, `sdf` (*Structure-Data File*), `mol2` e `pdb` (*Protein Data Bank*); `IOMolVS`, para escrita de arquivos moleculares em formato `pdbqt` usados em *molecular docking*; e `AssignPar`, a qual realiza algumas edições nas moléculas lidas dependendo do formato de arquivo utilizado. Dentro do pacote `forcefield` estão organizadas as seguintes classes, de acordo com a sequência de etapas empregadas na minimização de energia:

- a) `AtomTyping`: para cada átomo da molécula, associa um tipo atômico definido pelo campo de força `MMFF94`;
- b) `ParameterSetter`: identifica os parâmetros de campo de força para uma molécula em particular, armazenando-os para futura utilização;
- c) `Energy`: calcula a energia potencial da molécula. Calcula tanto a energia total quanto os termos individuais de energia (estiramento de ligação covalente, interações de van der Waals, etc);
- d) `Optimizer`: realiza a minimização da geometria molecular por meio de gradientes conjugados não lineares.

O pacote `bioisostermaker` define uma única classe com mesmo nome, `BioisosterMaker`, cuja tarefa é construir os bioisósteros, enquanto que o pacote `math` possui as classes `Matrix` para realizar operações matriciais e `CartesianVector` para operações de vetores cartesianos. No pacote `properties` há uma classe em específico para cálculo de cada uma das propriedades físico-químicas mencionadas anteriormente e uma classe geral `MolProperties`, que irá chamar as demais classes para o cálculo das propriedades. O pacote `virtualescreening` possui as seguintes classes:

- a) `ActiveSite`: define o sítio ativo do receptor;
- b) `SearchExecutable`: procura pelo programa `AutoDock Vina` no computador do usuário;
- c) `VirtualScreening`: realiza o *virtual screening*;
- d) `VSSorting`: ordena os bioisósteros em ordem crescente da função de score obtida no *molecular docking*;
- e) `VSTable`: exibe uma tabela com os resultados obtidos.

A relação entre as diferentes classes pode ser visualizada nos diagramas seguintes (FIGURAS 24 a 29).

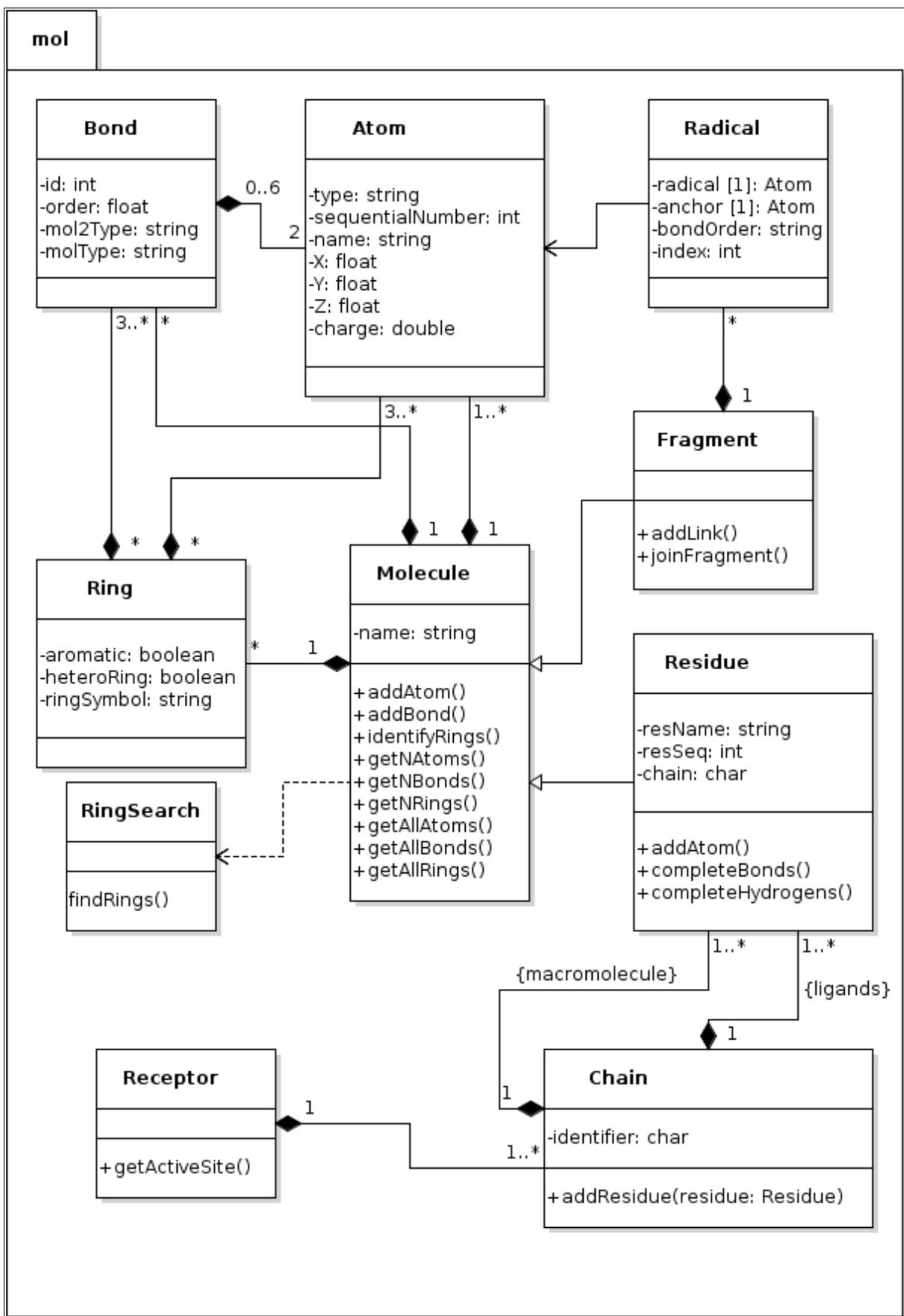


Figura 24 – Diagrama de classes para o pacote mol.

Fonte: Do autor.

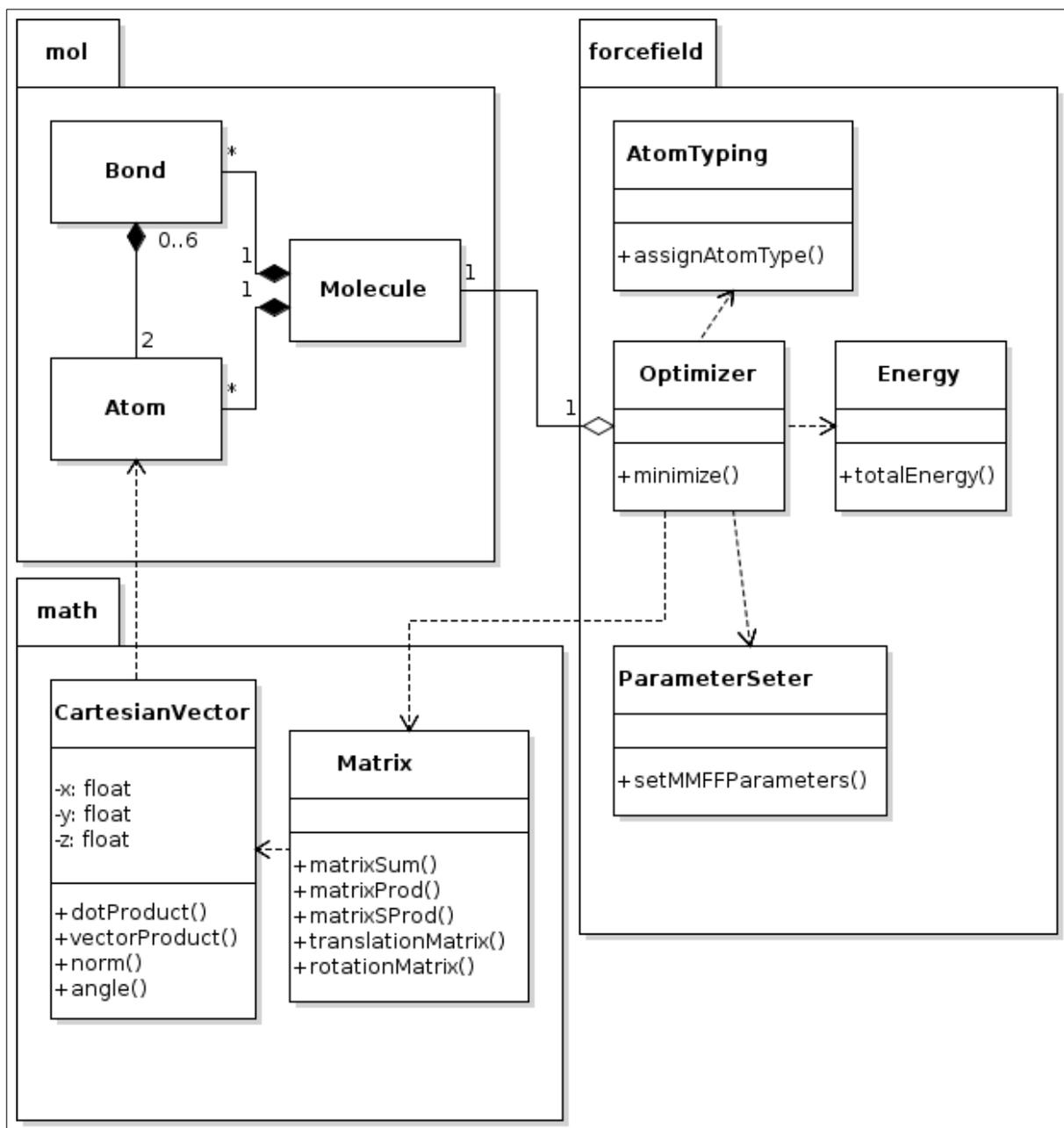


Figura 25 – Diagrama de classes mostrando a implementação do campo de força.

Fonte: Do autor.

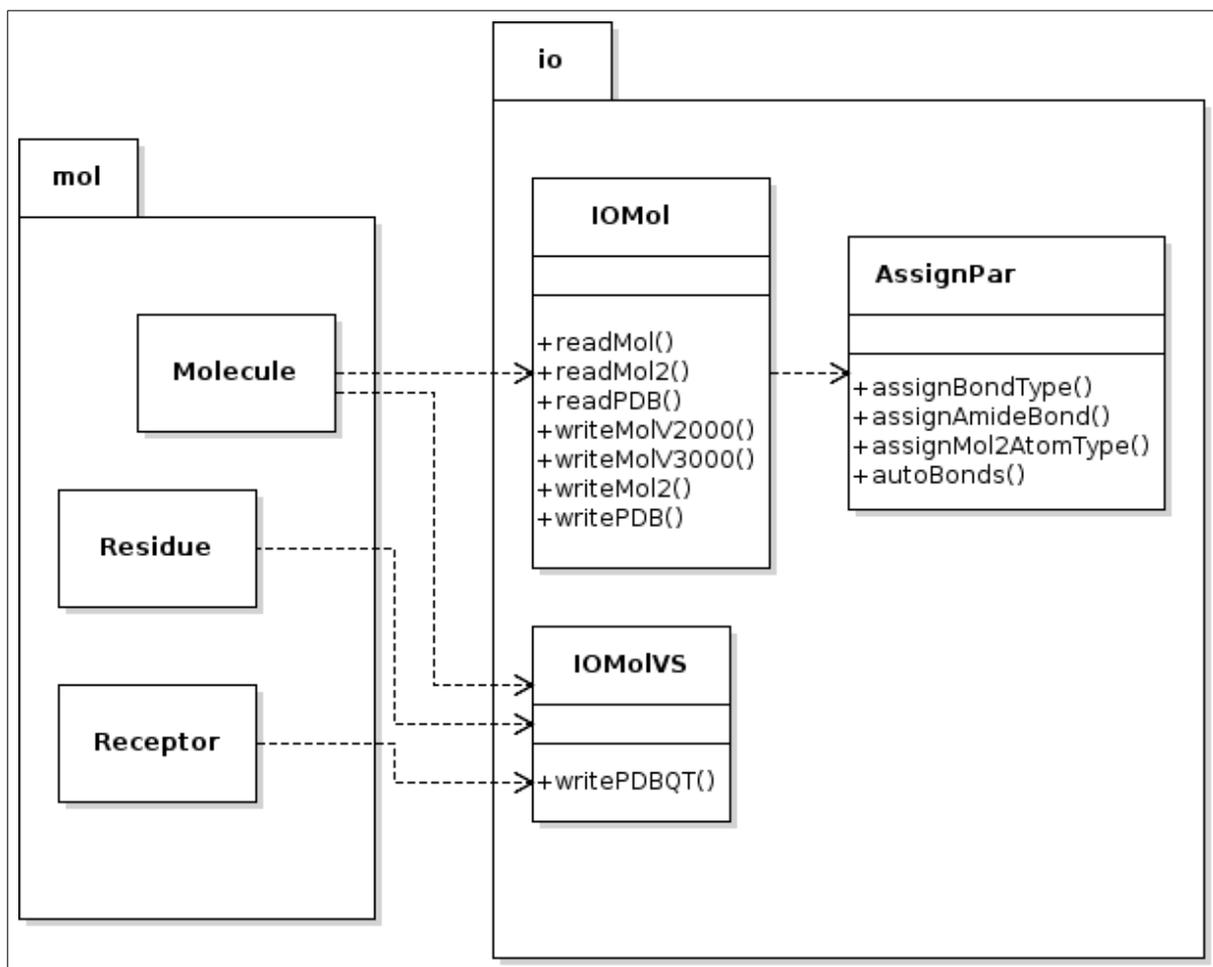


Figura 26 – Diagrama de classes mostrando interações de leitura e escrita de arquivos moleculares.
Fonte: Do autor.

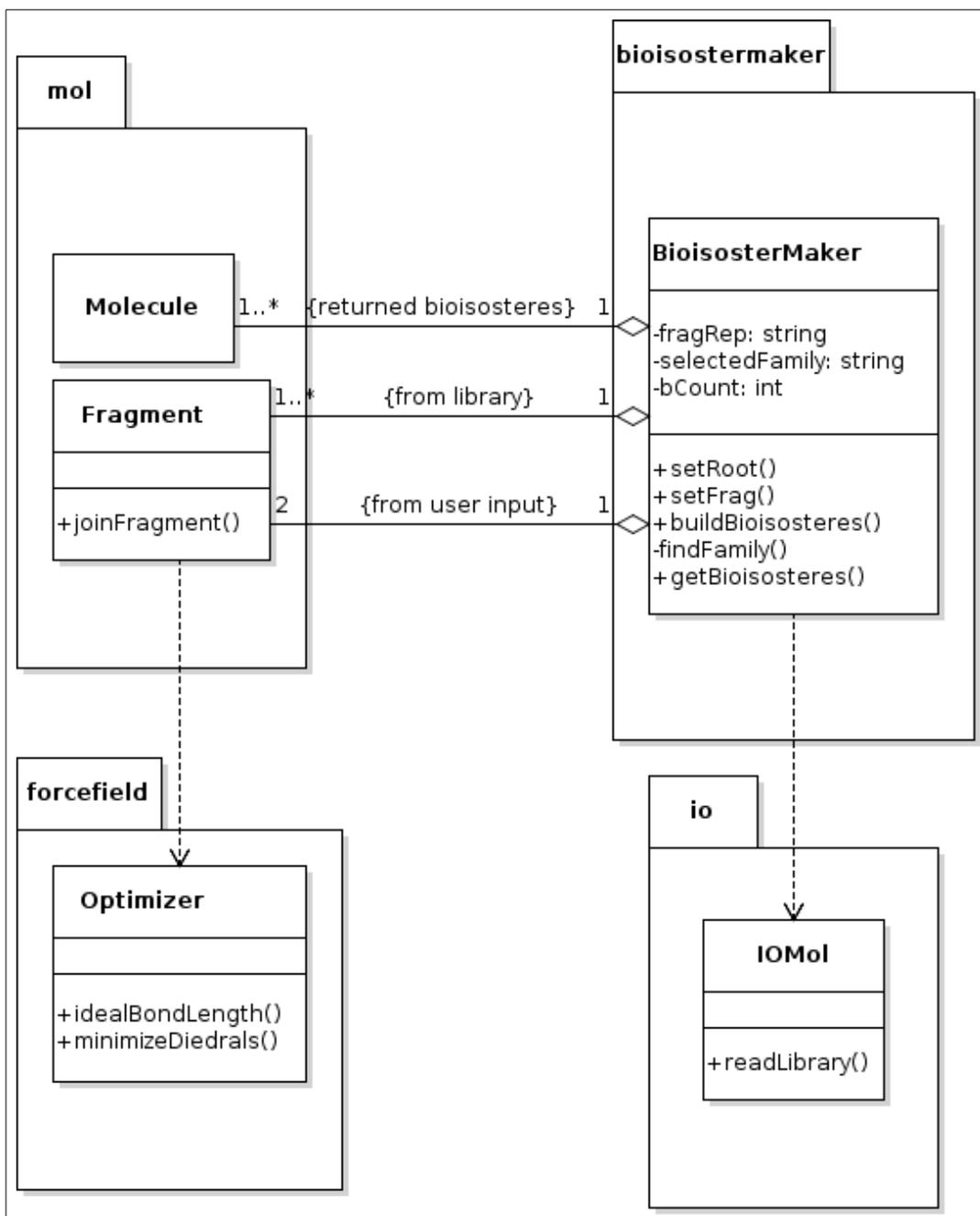


Figura 27 – Diagrama de classes para as funções de construção de novos bioisósteros.

Fonte: Do autor.

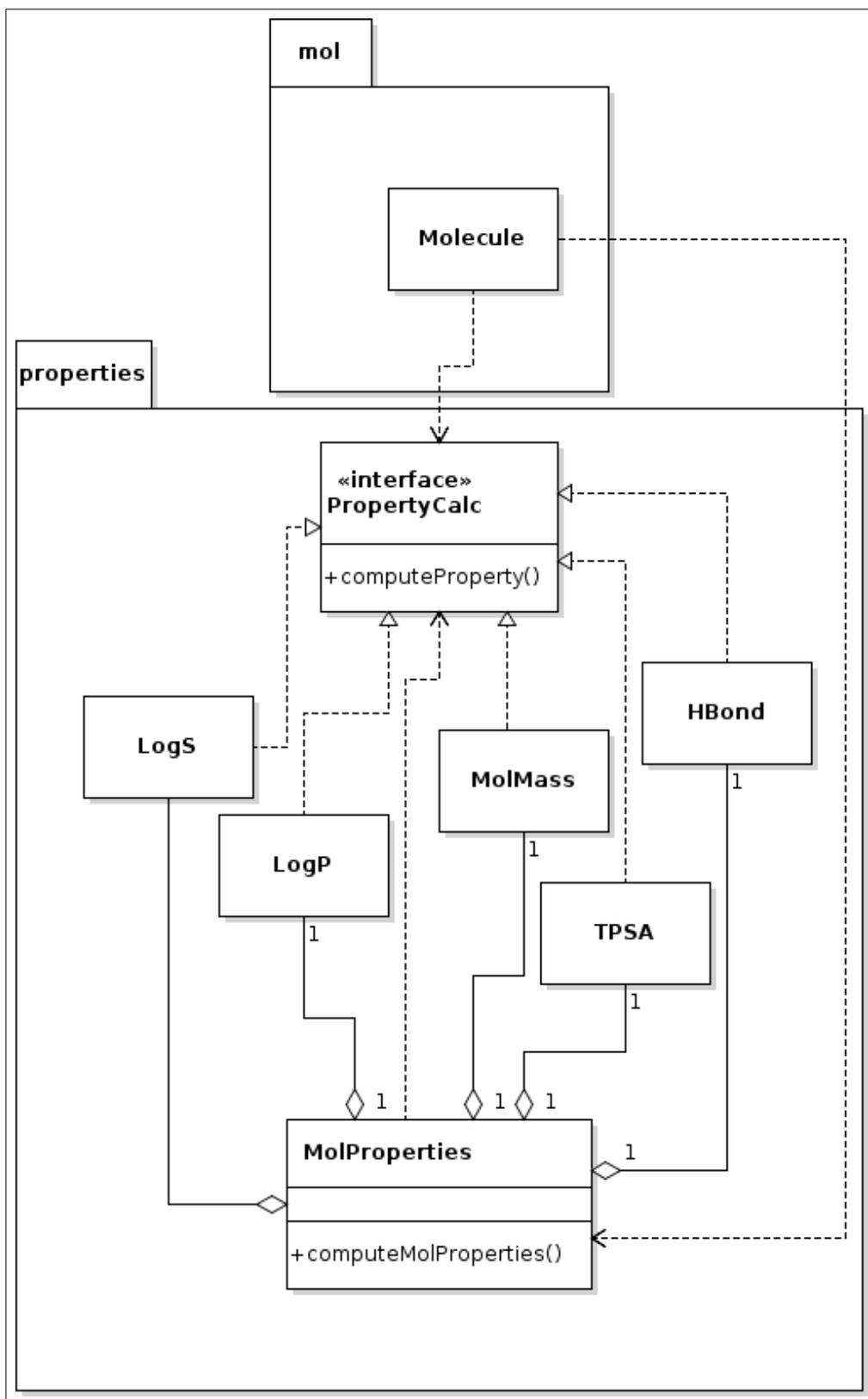


Figura 28 – Diagrama de classes para o cálculo das propriedades físico-químicas.
 Fonte: Do autor.

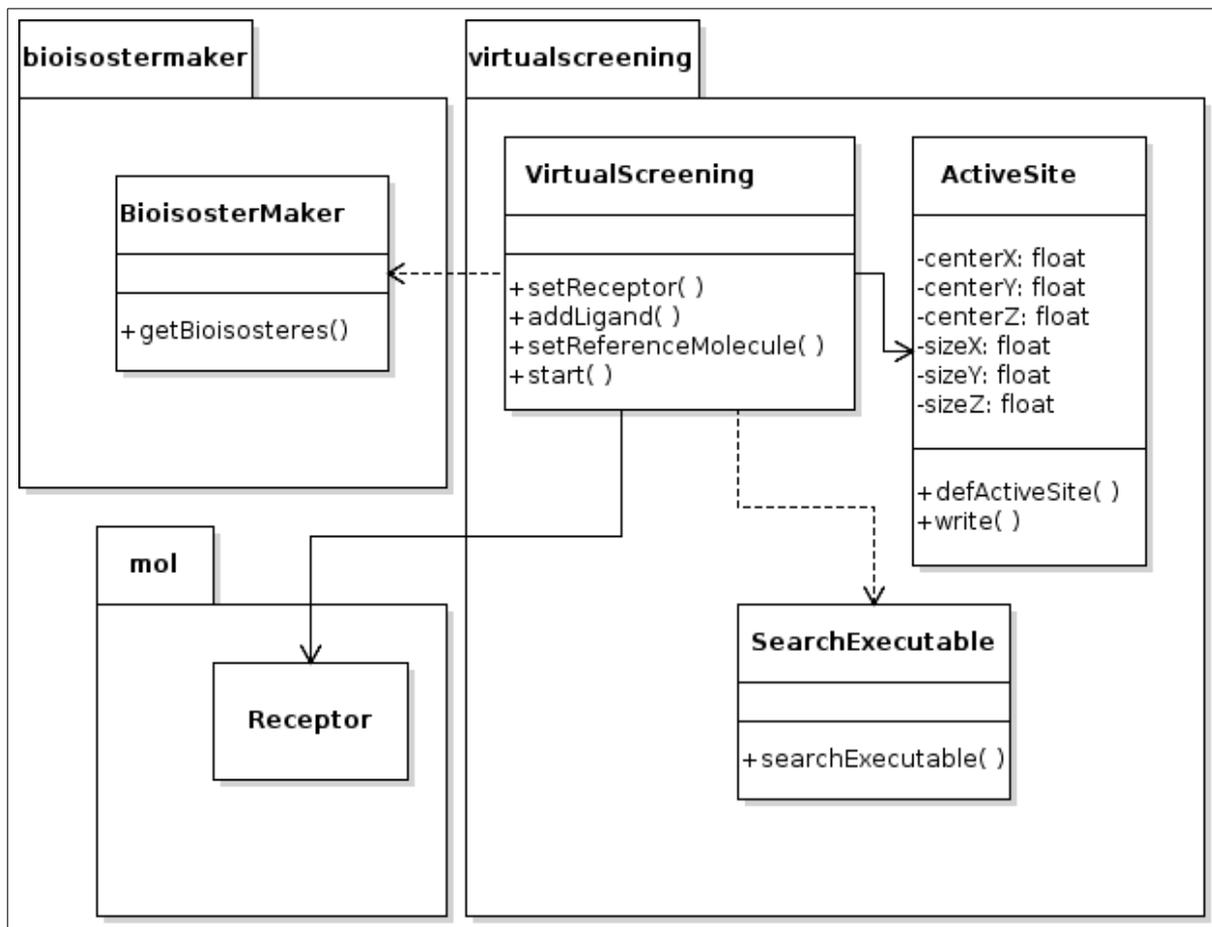


Figura 29 – Diagrama de classes para a execução do *virtual screening*.
Fonte: Do autor.

3. 1 REPRESENTAÇÃO MOLECULAR

De uma maneira simplificada, uma molécula pode ser definida como um conjunto de átomos conectados entre si por ligações covalentes. Dessa forma, uma maneira conveniente de representar computacionalmente uma molécula é definir uma classe Molecule que por sua vez será composta por uma lista de átomos (definidos pela classe Atom) e uma lista de ligações covalentes (definidas pela classe Bond). Alguns átomos podem estar conectados de tal forma a fechar um ciclo, formando sistemas de anéis, os quais podem ser identificados por meio de algoritmos baseados na teoria dos grafos. Uma vez que os átomos ocupam uma posição no espaço tridimensional, a geometria da molécula pode ser definida por alguns descritores, como distâncias interatômicas, ângulos entre três átomos consecutivos, ângulos diedro, ângulos de Wilson ou diedro impróprio e distância covalente.

3. 1. 1 Classe Atom

A classe Atom possui os atributos descritos na tabela 2:

Tabela 2 – Atributos da classe Atom

Atributo	Tipo	Descrição
element	String	Define o elemento do átomo. C para carbono, O para oxigênio, N para nitrogênio e assim por diante.
sequentialNumber	Inteiro	Número sequencial do átomo na molécula. O primeiro átomo da molécula recebe o número 1, o segundo o número 2, e assim por diante.
name	String	Define nomes diferentes para átomos do mesmo elemento na molécula; assim, os átomos de carbono podem ser nomeados como C1, C2, C3, etc, ou CA, CB, CC, etc.
nameIndicator	String	É o nome do átomo subtraído de seu tipo, assim, se o átomo é do tipo C e tem nome C1, o nameIndicator terá valor “1”.
X	Real	Coordenada cartesiana do átomo no eixo X.
Y	Real	Coordenada cartesiana do átomo no eixo Y.
Z	Real	Coordenada cartesiana do átomo no eixo Z.
mol2AtomType	String	Classificação do átomo de acordo com as convenções SYBYL para arquivos em mol2.
atomMMFFType	Inteiro	Classificação do átomo definido pelo campo de força MMFF.
symbol	String	Símbolo do átomo definido pelo campo de força MMFF.
charge	Real	Carga elétrica do átomo.
isInRing	Booleano	Define se o átomo é membro de algum anel.
DA	Caractere	Especifica se o átomo é um acceptor ou doador de interações de hidrogênio, ou nenhum.
valence	Real	Valência do átomo.
compVal	Real	Valência aparente, valência que o átomo aparenta ter devido a hidrogênios implícitos.
bondedAtoms	Lista de Atom	Lista dos átomos conectados covalentemente a este átomo.
properties	Lista de propriedades	Lista de propriedades do átomo definidas pelo campo de força MMFF.

Fonte: Do autor.

Aqui cabem algumas explicações quanto aos tipos de dados apresentados na segunda coluna da tabela 2:

- a) O tipo Inteiro é um valor numérico inteiro;
- b) O tipo Real é um valor numérico real;
- c) O tipo Caractere representa um único caractere, como a letra 'A' ou o caractere '1';
- d) O tipo String é uma cadeia de caracteres ou uma palavra, como “metano”;
- e) O tipo Booleano é um valor lógico que pode ser *verdadeiro* ou *falso*;
- f) A Lista é uma coleção de dados que pertencem ao mesmo tipo.

A lista de propriedades do átomo possui informações utilizadas para se atribuir parâmetros do campo de força MMFF e para se gerar parâmetros *default* quando estes não estão definidos. As propriedades são as seguintes:

- a) *aspec*: número atômico do átomo;
- b) *crd*: número de átomos vizinhos conectados ao átomo;
- c) *val*: valência do átomo; alguns átomos podem ser representados com mais de um tipo de valência, como é o caso de dois oxigênios conectados a um carbono em um carboxilato, um dos oxigênios é representado com valência 1 e o outro com valência 2;
- d) *pi*: define se o átomo possui um par isolado de elétrons pi que pode participar em interações de ressonância;
- e) *mltb*: define se o átomo obrigatoriamente deve participar de uma ligação covalente dupla ou tripla, ou em uma ligação intermediária como uma ligação amida;
- f) *arom*: define se o átomo faz parte de um anel aromático;
- g) *lin*: define se é esperado que o átomo forme um ângulo linear caso seja o átomo central entre três átomos consecutivos, como é o caso de carbonos sp, os quais realizam duas ligações duplas, ou uma tripla e uma simples;
- h) *sbmb*: definido para pares de átomos que podem formar entre si tanto ligações múltiplas como ligações delocalizadas simples, como é o caso dos dois carbonos centrais na molécula butadieno.

Os dados mínimos para se iniciar um átomo são o seu elemento, seu número sequencial e suas coordenadas cartesianas, os demais atributos podem ser calculados internamente por algoritmos específicos como será demonstrado nas seções seguintes. A lista de átomos conectados é construída de maneira indireta quando um objeto da classe Bond é criado.

3. 1. 2 Classe Bond

Os atributos da classe Bond são apresentados na tabela 3:

Tabela 3 – Atributos da classe Bond.

Atributo	Tipo	Descrição
atom1	Atom	Um dos átomos conectados pela ligação covalente.
atom2	Atom	O outro átomo conectado pela ligação covalente.
id	Inteiro	Número sequencial da ligação covalente. A primeira ligação covalente da molécula possui id = 1, a segunda id = 2 e assim por diante.
order	Real	Ordem de ligação definida da seguinte forma: Ligações simples: order = 1; ligações duplas: order = 2; ligações triplas: order = 3; ligações aromáticas: order = 1,5.
mol2Type	String	Ordem de ligação definida pelo formato de arquivo mol2.
molType	String	Ordem de ligação definida pelo formato de arquivo mol/sdf.
bt	Inteiro	Índice de tipo de ligação definido pelo campo de força MMFF.
isCyclic	Booleano	Define se a ligação covalente ocorre em um anel.

Fonte: Do autor.

Para se iniciar uma ligação covalente é necessário especificar os dois átomos que a compõe, o id, e opcionalmente a ordem de ligação. A correspondência entre as ordens de ligação definidas nos formatos mol e mol2, e aquela definida internamente (atributo *order*) é apresentada na tabela 4:

Tabela 4 – representação da ordem de ligações covalentes para os formatos mol/sdf, mol2 e a representação interna ao programa.

Ordem de ligação	Representação mol2	Representação mol/sdf	Representação interna
Ligação simples	1	1	1,0
Ligação dupla	2	2	2,0
Ligação tripla	3	3	3,0
Ligação aromática	ar	4	1,5
Ligação amida	am	1	1,0

Fonte: Do autor.

Uma vez que o objeto Bond é iniciado, cada um de seus átomos é notificado da formação da ligação covalente, ou seja, *atom2* é adicionado à lista de átomos vizinhos de *atom1* (atributo *bondedAtoms*) e *atom1* é adicionado à lista de vizinhos de *atom2*. Se a ordem de ligação não for especificada, é possível inferi-la baseando-se na distância interatômica e na natureza dos átomos participantes da ligação. O arquivo de configuração apresentado parcialmente na figura 30 é consultado para se definir as ordens de ligação:

C	C	1.4591	1.5675	1
C	C	1.3976	1.4591	ar
C	C	1.2962	1.3976	2
C	C	0.8900	1.2962	3
C	F	1.1500	1.6800	1
C	H	1.0800	1.1200	1
C	I	1.9900	2.2500	1
C	N	1.3513	1.4900	1
C	N	1.2625	1.3513	2
C	N	1.0100	1.2625	3
C	O	1.3400	1.4500	1
C	O	1.2839	1.3400	ar
C	O	1.2022	1.2839	2
C	P	1.2500	2.0200	1
C	S	1.7513	1.9000	1
C	S	1.7155	1.7513	ar
C	S	1.5600	1.7155	2
C	CL	1.1100	2.0400	1

Figura 30 – Arquivo de configuração usado para se definir as ordens de ligações covalentes.

Fonte: Do autor.

No arquivo apresentado na figura 30, as duas primeiras colunas correspondem aos elementos químicos dos átomos participantes da ligação, a terceira e a quarta colunas são a mínima e a máxima distância que pode existir entre o par de átomos para que seja atribuída determinada ordem de ligação, a qual é especificada na quinta coluna. Assim, se dois átomos de carbono estão a uma distância superior a 1,45 Å e inferior a 1,56 Å, a ordem de ligação entre estes será do tipo simples. Para se construir o arquivo de configuração apresentado acima, adotou-se o seguinte procedimento:

- Realizou-se o download de uma tabela do banco de dados PDB (*Protein Data Bank*) contendo os ligantes que aparecem em estruturas com resolução de 1,5 Å ou superior, num total de 3266 ligantes (“tabularResults.csv”, disponível em http://www.rcsb.org/pdb/results/grid_reports/gridReport.do?reportTitle=ligandSummary&format=csv&qrid=DBD4D06E&service=file);

- b) Fez-se o download do arquivo “components.cif”, um dicionário que define todos os ligantes que aparecem no banco de dados PDB, inclui informações sobre os tipos de ligação covalente em cada molécula (disponível em <ftp://ftp.wwpdb.org/pub/pdb/data/monomers/components.cif.gz>);
- c) Para cada ligante presente na tabela “tabularResults.csv”, realizou-se o download do mesmo;
- d) Para cada ligação covalente presente em um ligante, computou-se seu comprimento e se verificou seu tipo de acordo com o especificado no arquivo “components.cif”;
- e) Para cada par de átomos com os mesmos tipos atômicos e mesmo tipo de ligação foi reportado no arquivo de configuração o menor e maior comprimento de ligação encontrado;
- f) Por fim realizou-se uma edição manual para remover a sobreposição de intervalos em pares atômicos com os mesmos elementos mas diferentes ordens de ligação.

3. 1. 3 Classe Ring

A classe Ring, representando um anel na molécula, possui os atributos mostrados na tabela 5:

Tabela 5 – Atributos da classe Ring.

Atributo	Tipo	Descrição
members	Lista de Atom	Lista de átomos que pertencem ao anel.
bonds	Lista de Bond	Lista das ligações covalentes entre os átomos que pertencem ao anel.
size	Inteiro	Número de átomos do anel.
aromatic	Booleano	Define se o anel é aromático.
heteroRing	Booleano	Define se o anel possui heteroátomos, que são átomos diferentes de carbono.
ringSymbol	String	Representação textual do anel.

Fonte: Do autor.

A inicialização de um anel é feita informando-se seus átomos constituintes, os quais são identificados por meio de um algoritmo de busca de anéis apresentado na seção “3. 1. 3. 2

Algoritmo de busca de anéis”. O anel será classificado como aromático somente se possuir entre cinco e seis átomos e se todos os seus átomos constituintes possuírem a propriedade “arom” da lista de propriedades com valor verdadeiro; e será um hétero-anel se pelo menos um de seus átomos for diferente de carbono. Esses atributos são determinados conforme pseudo-código apresentado a seguir:

```

Início
Se size > 4 e size < 7 Então
    aromatic ← Verdadeiro;
Senão
    aromatic ← Falso;
Fim Se;

heteroRing ← Falso;

Para cada Atom de Ring Faça
    Se Atom.element != "C" Então
        heteroRing ← Verdadeiro;
    Fim Se;

    Se Atom.properties.arom = Falso Então
        aromatic ← Falso;
    Fim Se;

Fim Para;
Fim.

```

A representação textual do anel é útil quando se quer comparar anéis distintos e determinar se são equivalentes. Na figura 31 pode-se ver alguns exemplos de anéis e sua representação textual. Os critérios para se construir a representação textual são os seguintes:

- Inicia-se com uma chave de abertura e termina com chave de fechamento ('[' e ']');
- Cada átomo é representado por seu elemento, sendo que átomos em sequência estão ligados covalentemente, o último átomo da sequência está ligado ao primeiro átomo;
- Entre cada par de átomos consecutivos, há um caractere representando o tipo de ligação entre os mesmos da seguinte maneira: “-” para ligação simples; “=” para ligação dupla; e “:” para ligação aromática. Após o último átomo, segue o símbolo que representa a ligação covalente entre o último e o primeiro átomo da sequência;

Uma vez que é possível se fazer diferentes representações para um mesmo anel dependendo de qual átomo é selecionado como primeiro da sequência e de qual orientação se faça a leitura dos átomos (horária ou anti-horária), adotou-se as seguintes prioridades:

- a) Ligações de ordem maior devem aparecer primeiro, respeitando-se a sequência de ligações consecutivas, de acordo com a prioridade: “=” > “:” > “-”;
- b) Átomos mais comuns devem aparecer primeiro, respeitando-se a sequência de ligações consecutivas, de acordo com a prioridade: “C” > “N” > “O” > “S” > “P” > “Si” > “demais elementos”.

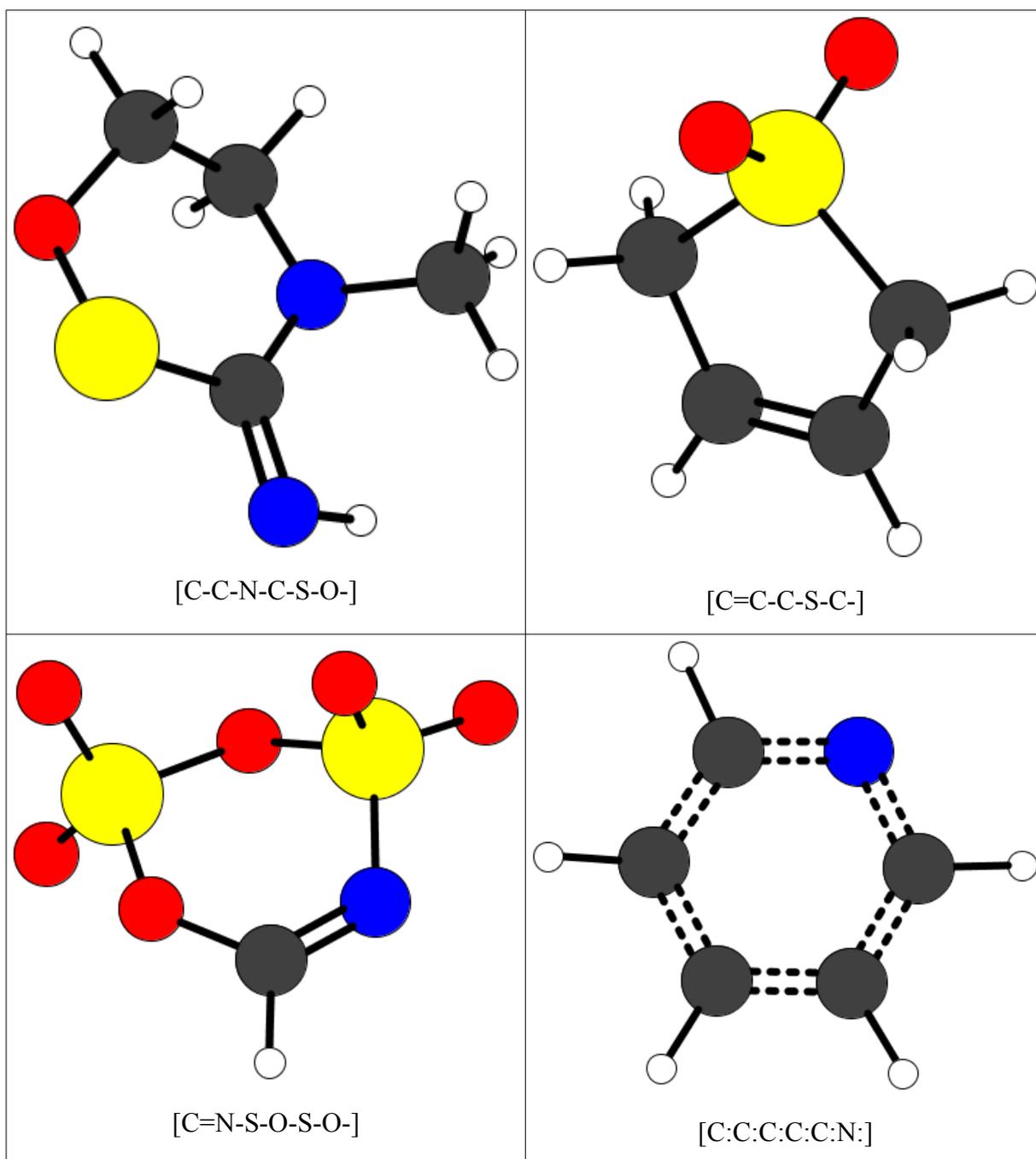


Figura 31 – Exemplos de anéis e sua representação textual. O seguinte esquema de cores é usado: cinza: carbono; azul: nitrogênio; vermelho: oxigênio; amarelo: enxofre; branco: hidrogênio.

Fonte: Do autor.

3. 1. 3. 1 Sistemas de anéis

Existe um conjunto de sistemas de anéis que podem ser encontrados em uma molécula, os quais podem ser anéis simples, anéis fundidos, anéis espiranos, sistemas em ponte e sistemas mais complexos em que os anéis se organizam como as faces de um poliedro. Em um anel simples, todos os átomos constituintes fazem parte de um único anel, como é apresentado na figura 32. Em anéis fundidos, dois átomos são comuns a dois anéis adjacentes (FIGURA 33), enquanto que em anéis espiranos um único átomo é comum a dois anéis adjacentes (FIGURA 34). Em sistemas de ponte, dois átomos de um mesmo anel não conectados entre si estão ambos conectados a um terceiro átomo (ou cadeia de átomos) não pertencente ao anel, de forma que três anéis podem ser identificados (FIGURA 35). Por fim, em um sistema em poliedro os átomos estão organizados de forma a ocupar os vértices ou as arestas de um sólido geométrico (FIGURA 36), os anéis deste sistema são mais difíceis de serem identificados.

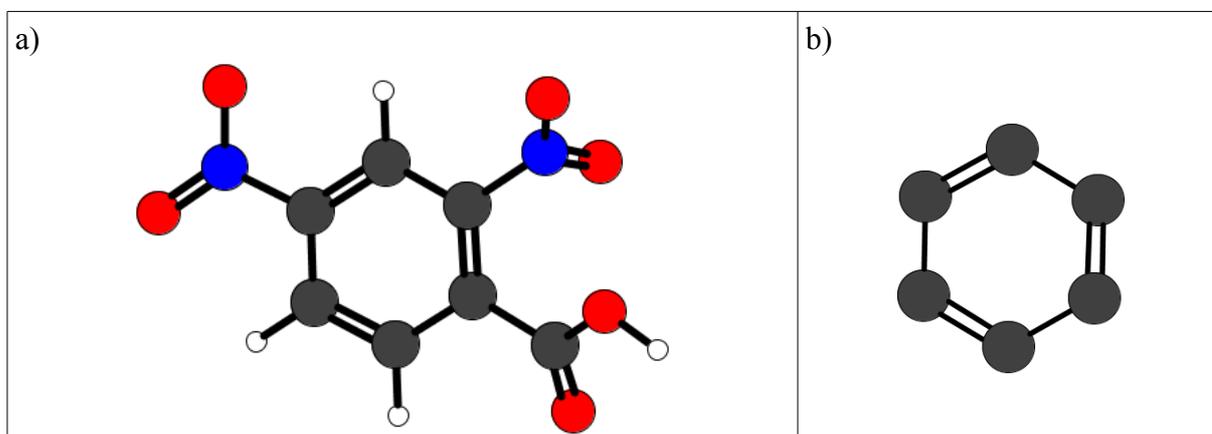


Figura 32 – Sistema de anéis simples.

Fonte: Do autor.

Legenda: a) exemplo de molécula contendo anel simples;

b) anel extraído da molécula exemplo.

Nota: Esquema de cores: cinza: carbono; azul: nitrogênio; vermelho: oxigênio; branco: hidrogênio.

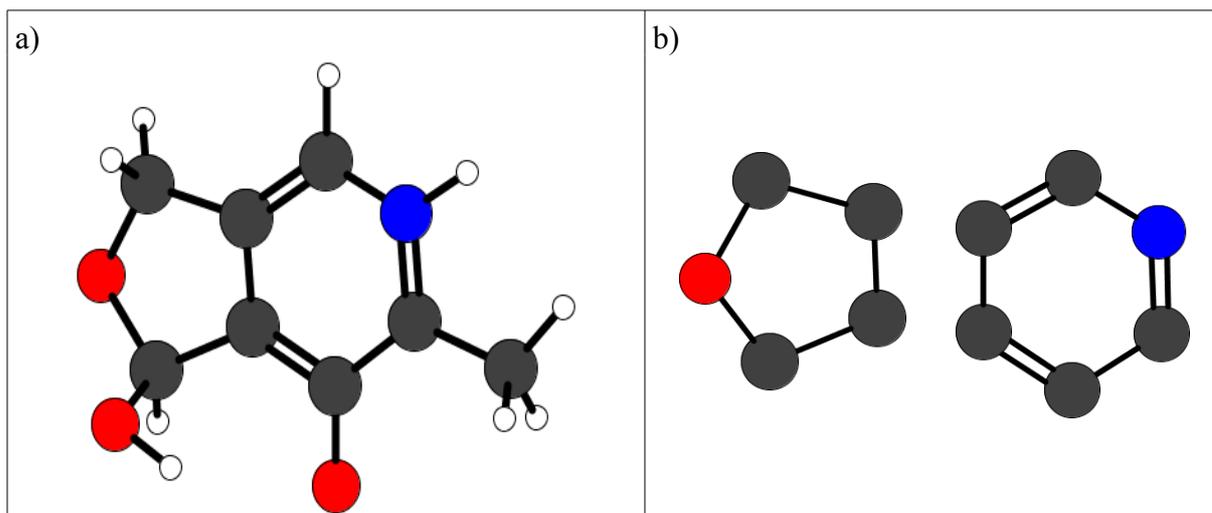


Figura 33 – Sistema de anéis fundidos.

Fonte: Do autor.

Legenda: a) exemplo de molécula contendo anéis fundidos;
b) anéis extraídos da molécula exemplo.

Nota: Esquema de cores: cinza: carbono; azul: nitrogênio; vermelho: oxigênio; branco: hidrogênio.

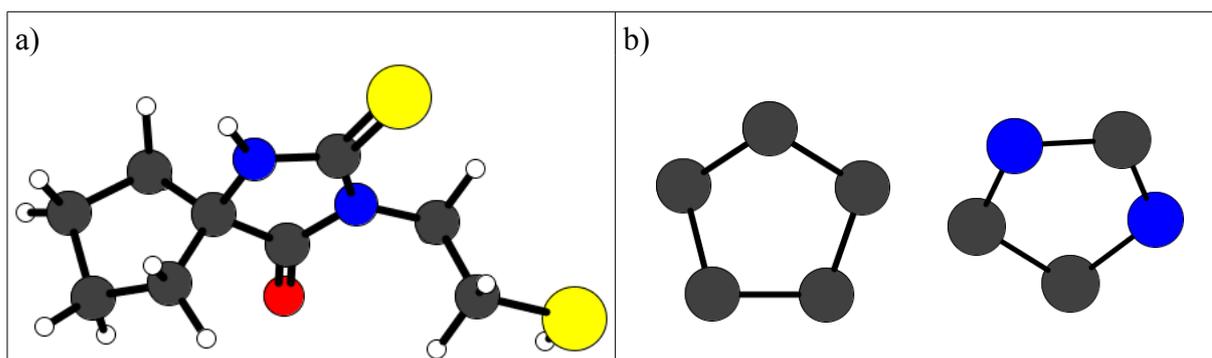


Figura 34 – Sistema de anéis espiranos.

Fonte: Do autor.

Legenda: a) exemplo de molécula contendo anéis espiranos;
b) anéis extraídos da molécula exemplo.

Nota: Esquema de cores: cinza: carbono; azul: nitrogênio; vermelho: oxigênio; amarelo: enxofre; branco: hidrogênio.

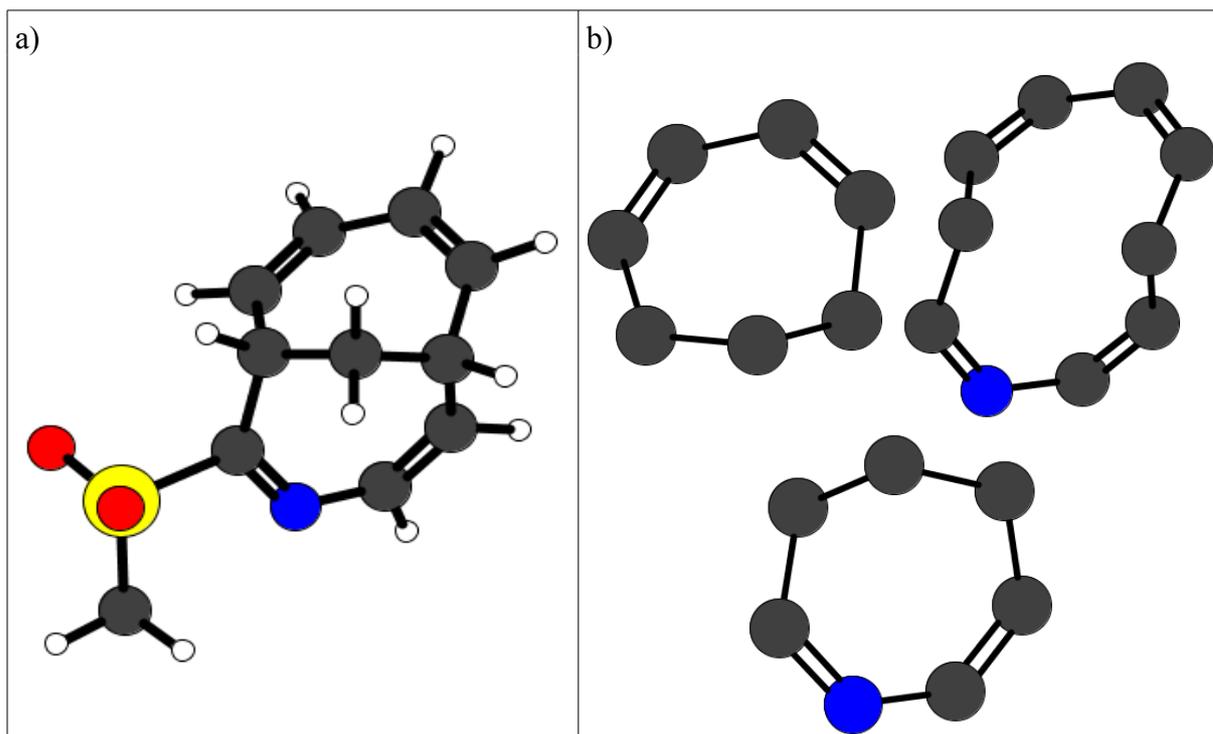


Figura 35 – Sistema de anéis em ponte.

Fonte: Do autor.

Legenda: a) exemplo de molécula contendo anéis em ponte;
b) anéis extraídos da molécula exemplo.

Nota: Esquema de cores: cinza: carbono; azul: nitrogênio; vermelho: oxigênio; amarelo: enxofre; branco: hidrogênio.

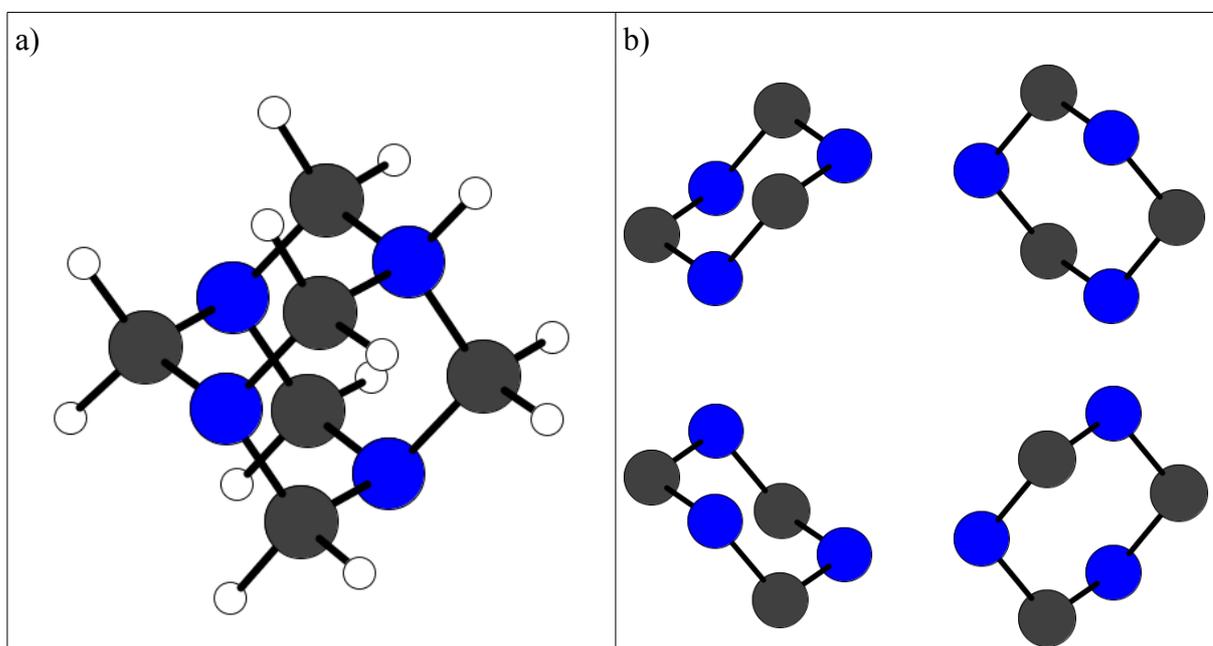


Figura 36 – Sistema de anéis complexos.

Fonte: Do autor.

Legenda: a) exemplo de molécula contendo anéis complexos;
b) anéis extraídos da molécula exemplo.

Nota: Esquema de cores: cinza: carbono; azul: nitrogênio; branco: hidrogênio.

3. 1. 3. 2 Algoritmo de busca de anéis

O algoritmo de busca de anéis é implementado pela classe RingSearch, utilizando procedimentos baseados na teoria dos grafos. Formalmente, um grafo G consiste em um conjunto de vértices V , conectados por um conjunto de arestas E , de forma que podemos escrever $G(V, E)$. Se um vértice é denotado por v e outro por w , a aresta que conecta a ambos é denotada por vw ou wv (NETTO E JURKIEWICZ, 2017). Se uma aresta conecta os vértices v e w , então ela é dita ser incidente a v e a w . Na figura 37 temos a molécula quinina e sua representação em forma de grafo. Cada vértice do grafo (o qual representa um átomo na molécula) é desenhado como um ponto indexado a partir de 1, cada aresta (a qual representa uma ligação covalente) é desenhada como uma linha conectando dois vértices.

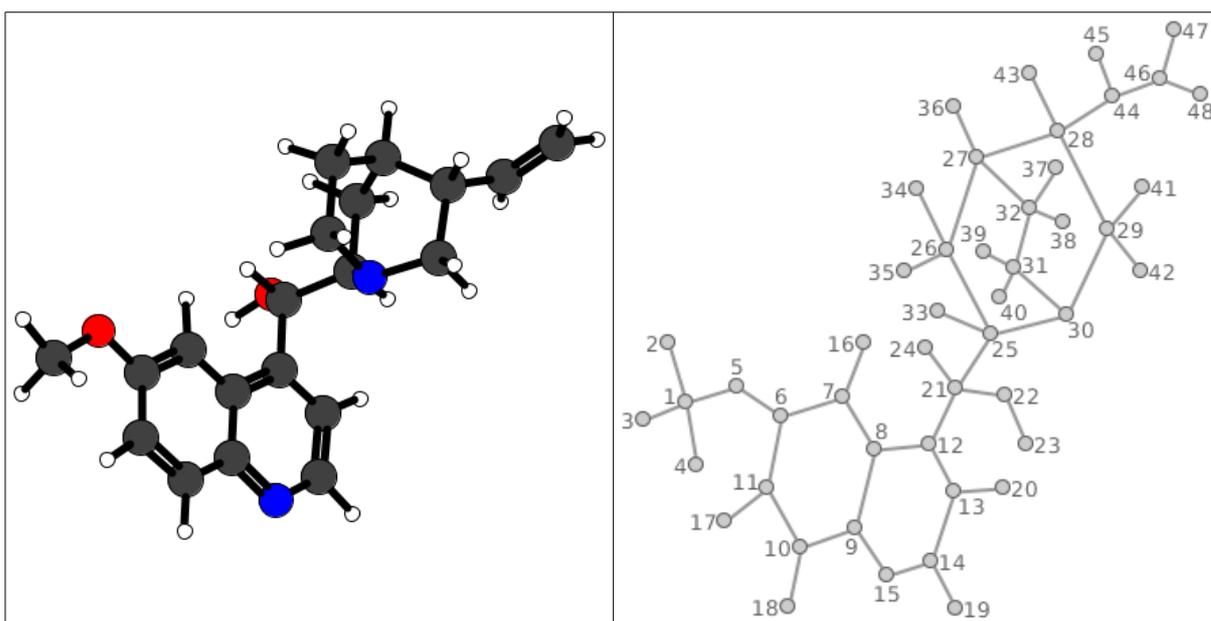


Figura 37 – Molécula quinina e sua representação como grafo.
Fonte: Do autor.

A representação de um grafo de maneira gráfica conforme apresentado na figura 37 não é compreensível por um programa de computador, sendo necessário representá-lo de outra forma. Optou-se pela representação por meio da matriz de adjacência, a qual é uma matriz $n \times n$, em que n é o número de vértices do grafo. Cada linha da matriz representa um vértice i e cada coluna representa um vértice j , se o vértice i estiver conectado ao vértice j , a entrada A_{ij} da matriz de adjacência terá valor 1, caso contrário, será zero. Assumindo a não existência de laços, ou seja, um vértice não pode estar conectado a ele mesmo, as entradas A_{ii}

sempre terão valor zero. Na figura 38 é apresentada matriz de adjacência para o grafo da molécula benzeno.

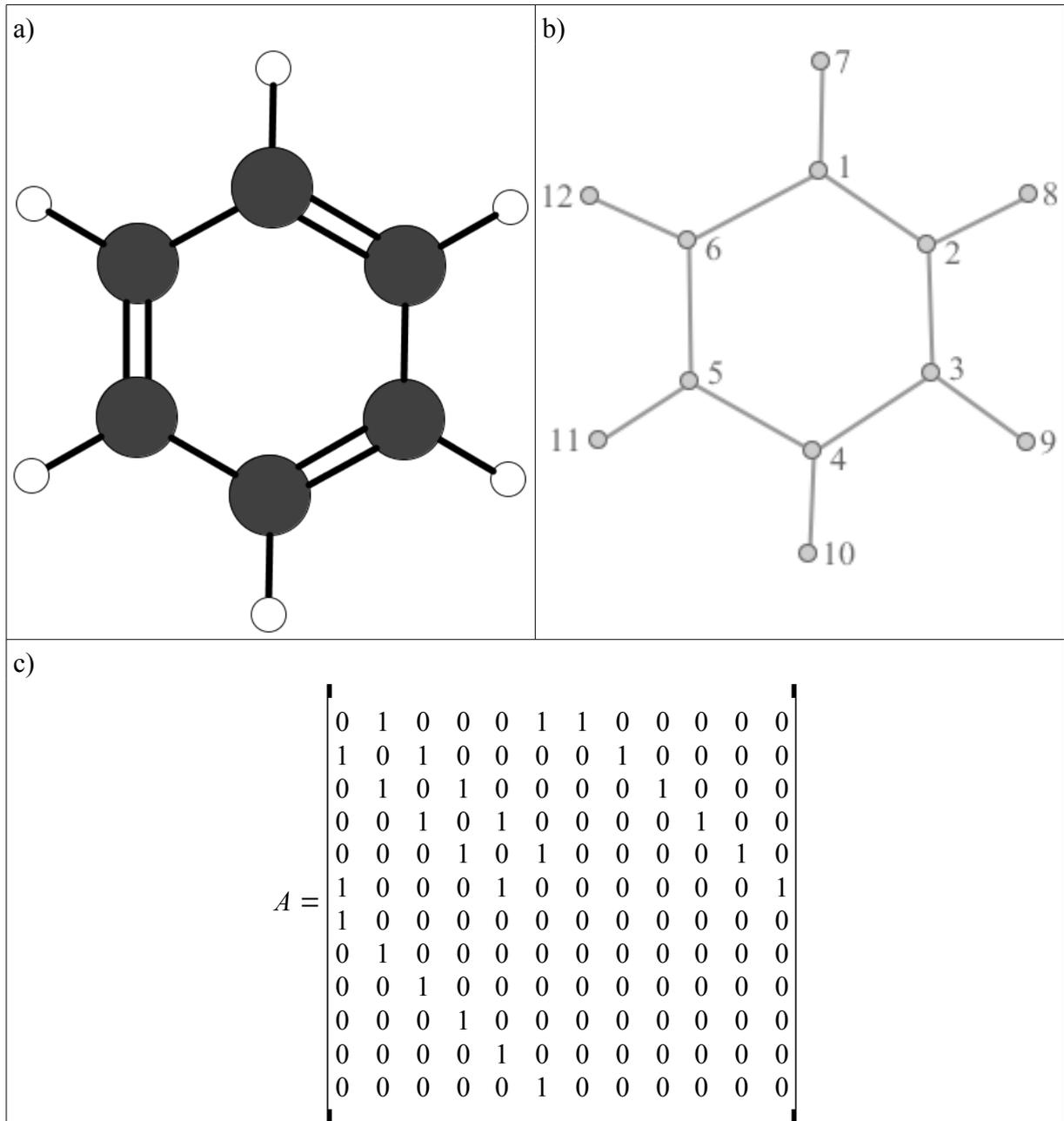


Figura 38 – Matriz de adjacência para o grafo da molécula benzeno.

Fonte: Do autor.

Legenda: a) Molécula de benzeno;

b) Grafo para a molécula benzeno;

c) Matriz de adjacência para o grafo apresentado em "b".

O grau de um vértice i é definido como o número de vértices aos quais i está conectado, dessa forma, na figura 38, o vértice de índice 4 possui grau 3, uma vez que está conectado a três vértices (vértices 3, 5 e 10). Junto com a matriz de adjacência A é definida

uma matriz coluna G contendo o grau de cada vértice; a matriz coluna G para a molécula de benzeno da figura 38 é apresentada a seguir:

$$G = \begin{pmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 3 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Antes de se iniciar a busca por anéis, é realizado um teste para se verificar se a molécula possui ou não anéis, avaliando se a seguinte equação é verdadeira:

$$\frac{1}{2} \sum_{i=1}^n g_i = n - 1$$

soma-se o grau de todos os vértices e divide-se o resultado por dois, se o valor encontrado for igual ao número de vértices menos um, a molécula não tem anéis e o algoritmo é encerrado.

Vértices com grau 1 não podem pertencer a um anel, então eles são eliminados do grafo antes de se iniciar a busca pelos anéis. Quando estes vértices são eliminados, podem surgir novos vértices que anteriormente possuíam grau maior que 1, mas agora possuem grau 1, o processo é então repetido até que não sobrem mais vértices de grau 1, conforme pseudocódigo abaixo. Uma variável global de controle, aqui nomeada *sumValency*, armazena a soma dos graus de todos os vértices, a cada remoção de vértices esta é decrementada.

Início

```

booleano : inOperation;
inOperation ← Verdadeiro;
inteiro : i, j, k;
k ← maior índice válido de G;
Enquanto inOperation = Verdadeiro Faça
    inOperation ← Falso;
    Para i = 0 até k Faça

```

```

Se G[i] = 1 Então
  inOperation ← Verdadeiro;
  Para j = 0 até k Faça
    A[i][j] = 0;
    A[j][i] = 0;
    G[i] = 0;
    G[j] = G[j] - 1;
    sumValency = sumValency - 2
  Fim Para
Fim Se
Fim Para
Fim Enquanto
Fim

```

Se a molécula possuir apenas um único sistema de anéis, como é o caso do benzeno (FIGURA 38), o algoritmo descrito é suficiente para remover todos os vértices que não participam de anéis, entretanto, moléculas mais complexas como a quinina (FIGURA 37), possuem sistemas de anéis separados em sua estrutura, de forma que ao se remover vértices de grau 1, ainda restarão vértices que não pertencem a anéis, mas conectam sistemas de anéis separados, como é o caso do vértice nº 13 para o grafo mostrado na figura 39.

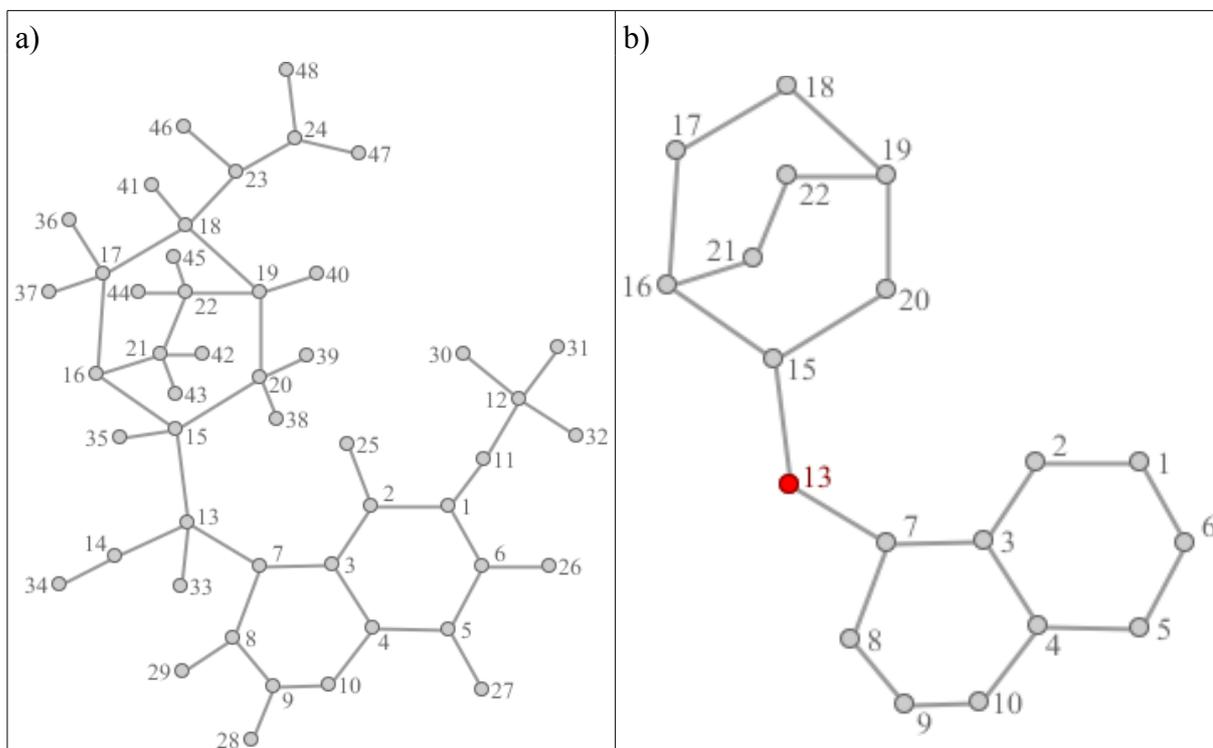


Figura 39 – Grafo da molécula quinina, mostrando a remoção de vértices não pertencentes a sistemas de anéis.

Fonte: Do autor.

Legenda: a) Grafo original;

b) Grafo após a remoção de vértices de grau 1. O vértice nº 13, embora não pertença a um sistema de anel, ainda se encontra presente.

Um vértice que é membro de um anel e que ao mesmo tempo está conectado a outro vértice não pertencente a um anel, possui grau 3 ou maior, logo, esses vértices são pontos candidatos para se localizar as cadeias abertas internas, como se pode ver para os vértices nº 7 e nº 15 da figura 39. A remoção destas cadeias segue o seguinte procedimento:

Início

Para cada vértice v com grau 3 ou maior Faça
 Para cada aresta e incidente a v Faça

$A[i][j] = 0$; {rompe a aresta provisoriamente}
 $A[j][i] = 0$; {rompe a aresta provisoriamente}

por meio de uma DFS (busca em profundidade),
 procurar outro caminho a partir de i que chegue a j

Se um caminho for encontrado Então
 os vértices i e j fazem parte do mesmo anel
 $A[i][j] = 1$ {a aresta é restabelecida}
 $A[j][i] = 1$ {a aresta é restabelecida}

Senão
 o vértice j pertence a uma cadeia aberta
 interna, a aresta se mantém rompida.

Fim Se

Fim Para

Fim Para

Fim

Ao final, os vértices das cadeias abertas que inicialmente estavam conectados a um vértice membro de anel, possuirão grau 1, o método anterior de remoção de cadeias laterais é chamado novamente, completando a retirada dos vértices internos. Feito isso, apenas os anéis estarão presentes. Para simplificar ainda mais o processo de busca, realiza-se uma contração de vértices, no qual vértices de grau 2 são “fundidos” em um único ponto, conforme figura 40.

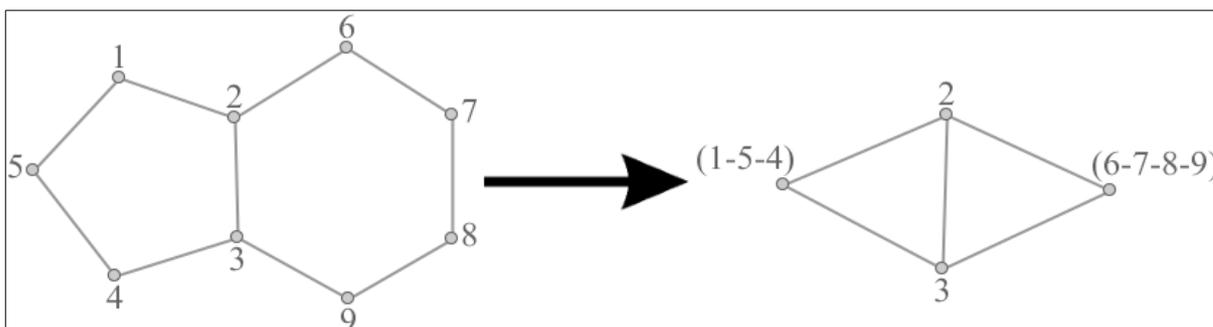


Figura 40 – Simplificação da representação molecular por contração de vértices.

Fonte: Do autor.

Para armazenar a informação dos átomos que fazem parte dos vértices contraídos, implementou-se uma matriz coluna contendo uma referência a cada átomo correspondendo a um vértice no grafo. Organizou-se uma estrutura de lista encadeada para organizar as cadeias de átomos para os vértices contraídos conforme apresentado na figura 41.

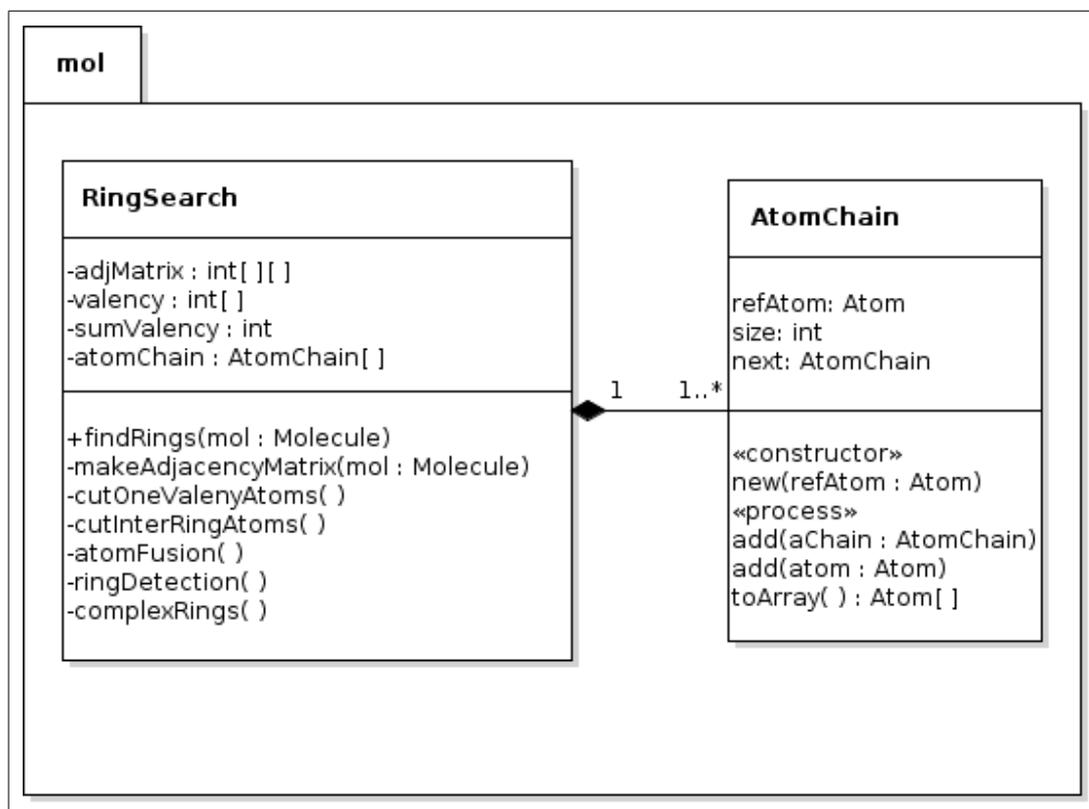


Figura 41 – Estrutura de dados em lista encadeada para armazenar cadeias de átomos contidos em vértices contraídos.

Fonte: Do autor.

Sendo C a matriz coluna contendo a lista encadeada mostrada acima, a operação de contração de vértices é feita da seguinte forma:

```

Início
  Para cada vértice  $i$  de grau 2 Faça
    Para cada vértice  $j$  de grau 2 conectado a  $i$  Faça:
      {remover a ligação entre  $i$  e  $j$ }
       $A[i][j] = 0;$ 
       $A[j][i] = 0;$ 
       $G[j] = 0;$ 
      Para cada vértice  $k$  conectado a  $j$  Faça:
        {desconectar  $k$  de  $j$ }
         $A[j][k] = 0;$ 
         $A[k][j] = 0;$ 
        {conectar  $k$  a  $i$ , assume-se a possibilidade de
           $k$  e  $i$  serem conectados por mais de uma
          aresta}
         $A[i][k] = A[i][k] + 1;$ 

```

```

A[k][i] = A[k][i] + 1;
{adicionar em C a informação de quais átomos
estão contraídos}
C[i] adiciona C[j];
Fim Para
Fim Para
Fim Para
Fim

```

Ao final da contração de vértices, a matriz de adjacência A apresentará muitas linhas cujas entradas serão todas zero, correspondendo aos vértices que foram eliminados, seja porque não fazem parte de anel, seja porque foram contraídos em um único vértice. Realiza-se uma redução da matriz de adjacência A , removendo-se as linhas e colunas cujas entradas sejam todas zero. As matrizes coluna G e C sofrem redução equivalente.

Na busca dos anéis, diferentes procedimentos são implementados dependendo do sistema de anel encontrado (isolado, espirano, fundido, pontes, e sistema complexo). Em anéis isolados, ambos os vértices que formam o anel possuem na matriz A a entrada $A[i][j] = 2$ (FIGURA 42), quando essa situação é encontrada, a união das listas encadeadas $C[i]$ e $C[j]$ é reportada como um anel encontrado; remove-se este anel para não ser contado novamente ($A[i][j] = 0$; $A[j][i] = 0$; $G[i] = 0$; $G[j] = 0$). Na figura 42 e seguintes as entradas da matriz C são representadas como o índice dos vértices correspondentes.

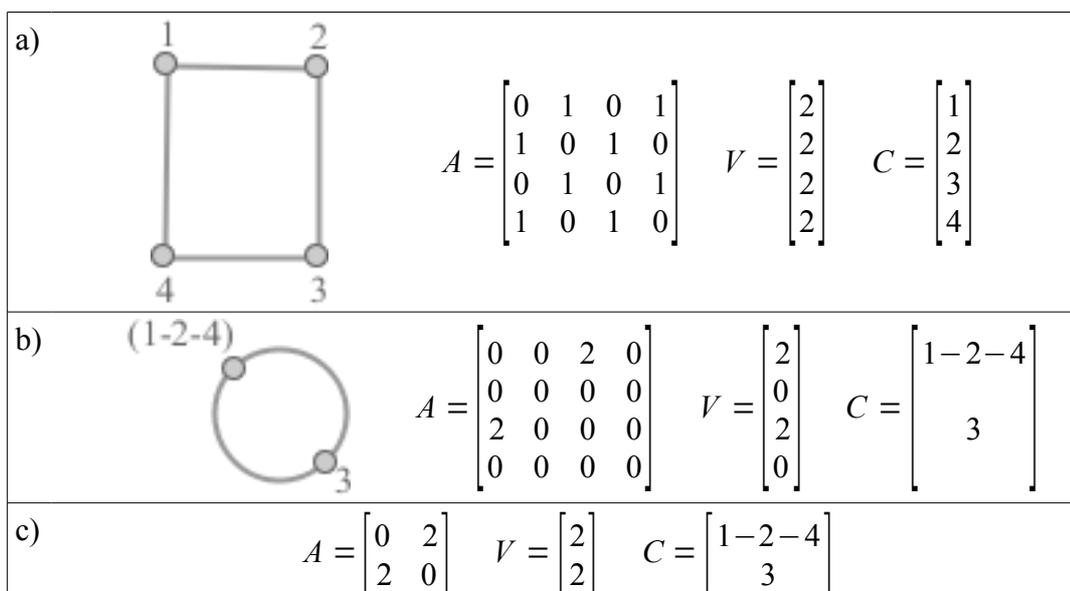


Figura 42 – Contração de vértices para anéis isolados.

Fonte: Do autor.

Legenda: a) representação inicial do anel isolado;
b) representação após contração de vértices;
c) matriz reduzida.

Em sistemas espiranos, as entradas $A[i][j]$ também são iguais a 2, mas um dos vértices possui grau 4, neste caso o anel é reportado como anteriormente (união das listas encadeadas $C[i]$ e $C[j]$), mas apenas o vértice de grau 2 é removido e o vértice de grau 4 tem o mesmo decrementado para 2 (FIGURA 43).

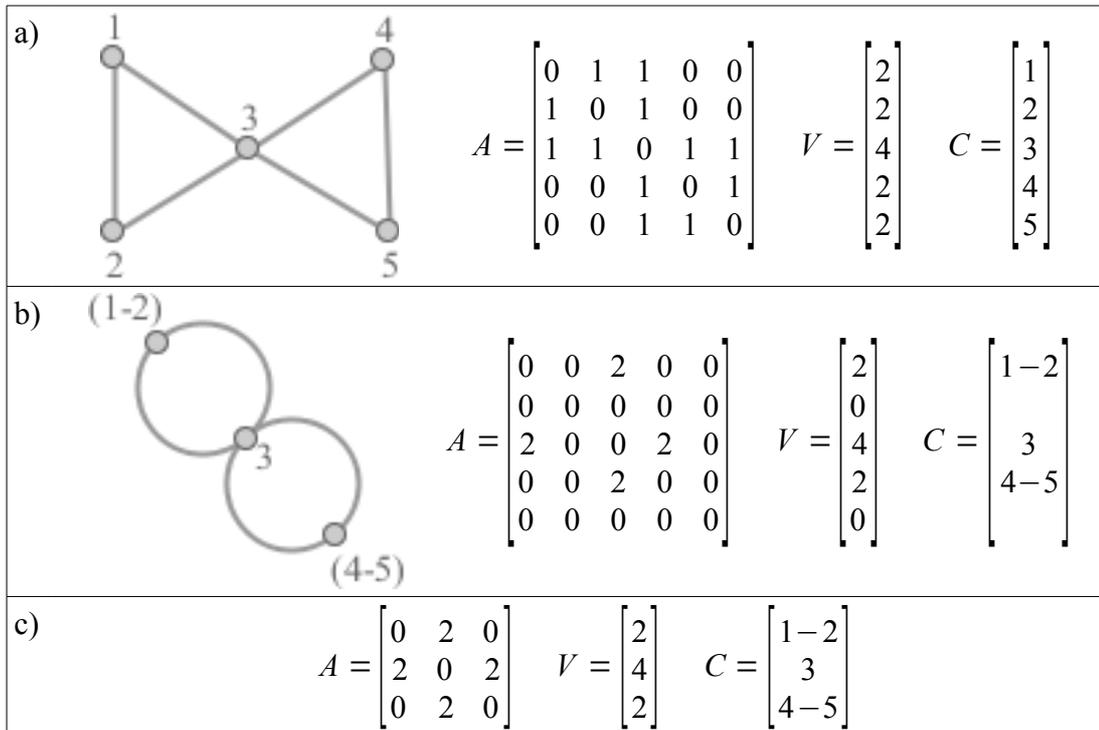


Figura 43 – Contração de vértices para anéis espiranos.

Fonte: Do autor.

- Legenda: a) representação inicial do anel isolado;
 b) representação após contração de vértices;
 c) matriz reduzida.

Em anéis fundidos, dois sub casos podem ocorrer: um em que o vértice i de grau 2 está conectado aos vértices $p1$ e $p2$, ambos com grau igual ou maior que 3, sendo que $p1$ e $p2$ estão diretamente conectados (FIGURA 44), e outro caso em que $p1$ e $p2$ não estão diretamente conectados, mas estão conectados a um vértice em comum c (FIGURA 45). No primeiro caso, o anel é reportado como união das listas encadeadas $C[i]$, $C[p1]$ e $C[p2]$ e no segundo caso como a união de $C[i]$, $C[p1]$, $C[c]$ e $C[p2]$, o vértice i é então removido e os graus $G[p1]$ e $G[p2]$ decrementados em uma unidade.

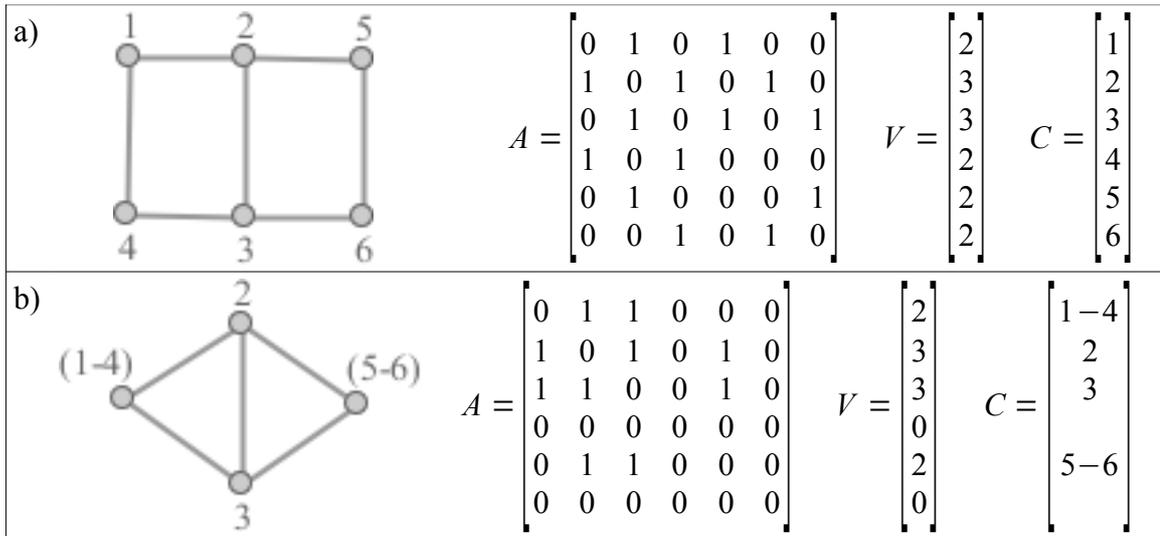


Figura 44 – Contração de vértices para anéis fundidos, caso I.

Fonte: Do autor.

Legenda: a) representação inicial do anel fundido;
 b) representação após contração de vértices.

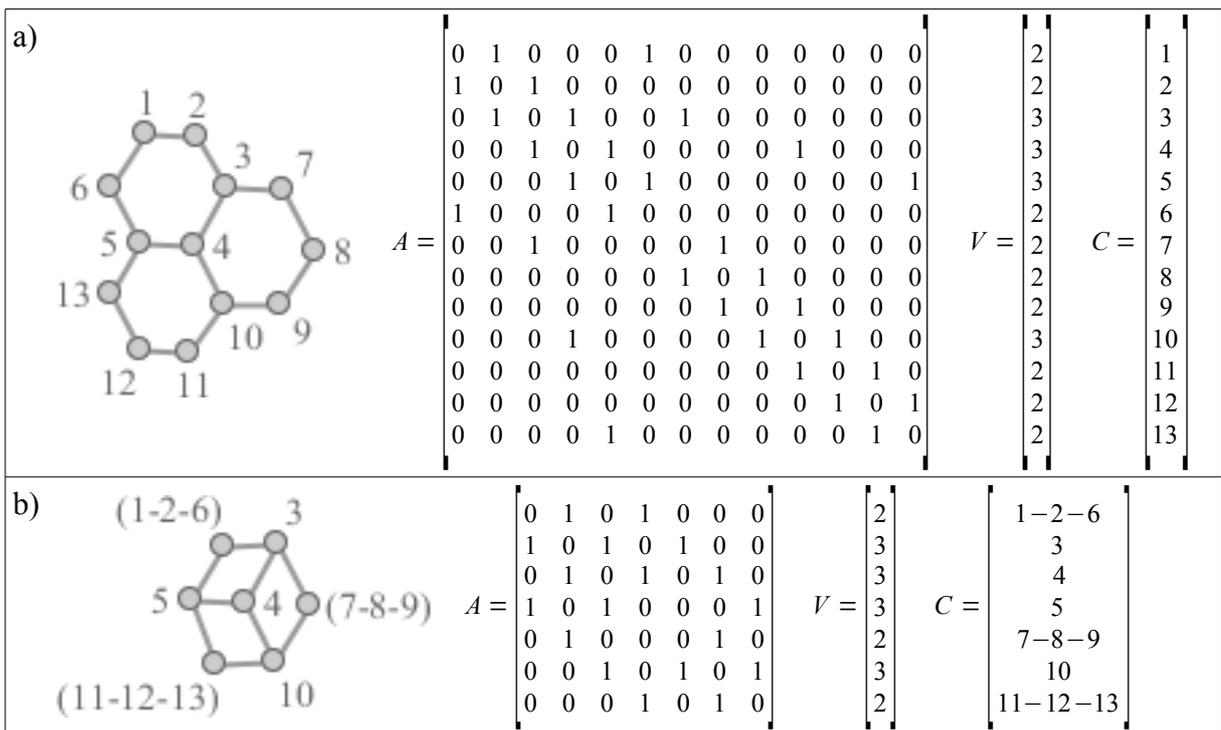


Figura 45 – Contração de vértices para anéis fundidos, caso II.

Fonte: Do autor.

Legenda: a) representação inicial do anel fundido;
 b) representação após contração de vértices e redução da matriz.

A identificação de sistemas de anéis em ponte é similar ao segundo sub caso para anéis fundidos, com a diferença de que não haverá um único vértice c comum a $p1$ e $p2$, mas pelo menos dois vértices $c1$ e $c2$, os anéis são reportados para cada combinação dois a dois entre os vértices $i, c1$ e $c2$, identificando-se assim três anéis: I) união das listas encadeadas $C[i], C[p1]$,

$C[p2]$ e $C[c1]$; II) união das listas encadeadas $C[i]$, $C[p1]$, $C[p2]$ e $C[c2]$; III) união das listas encadeadas $C[c1]$, $C[p1]$, $C[p2]$ e $C[c2]$ (FIGURA 46).

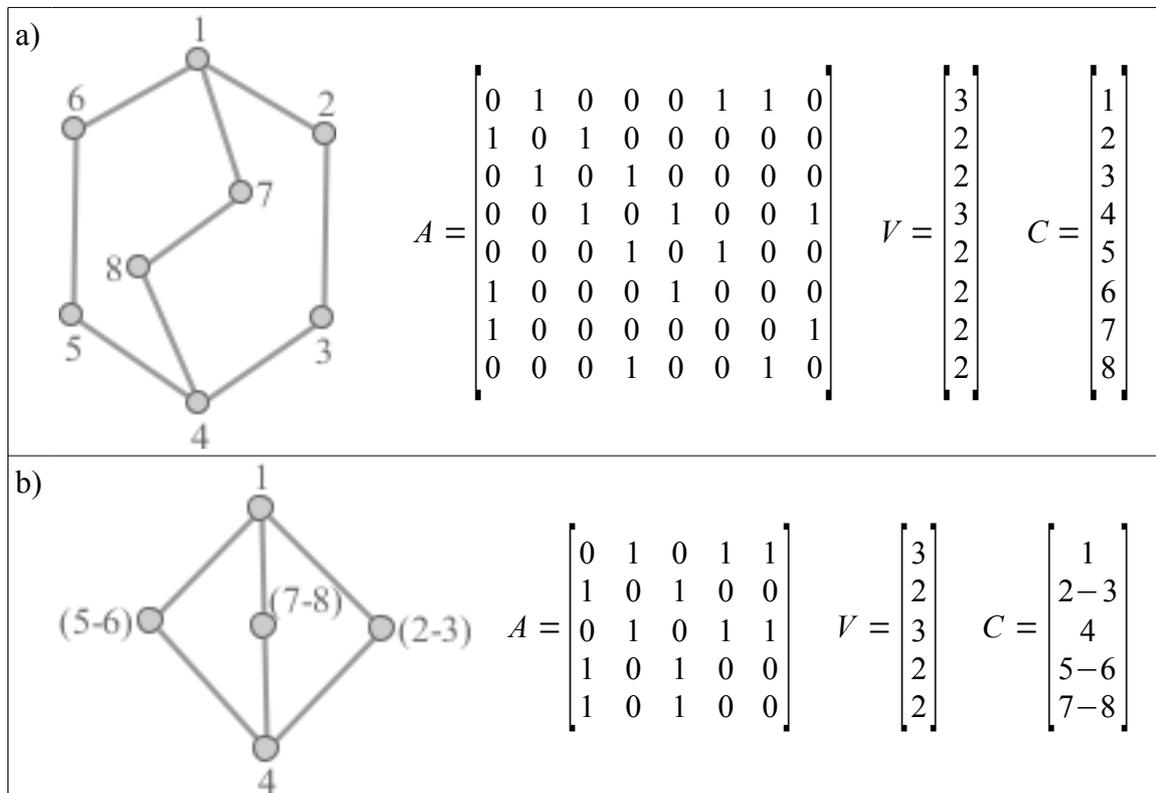


Figura 46 – Contração de vértices para sistemas de ponte.

Fonte: Do autor.

Legenda: a) representação inicial;

b) representação após contração de vértices e redução da matriz.

Após uma primeira iteração, a variável de controle *sumValency* é decrementada a medida em que anéis são identificados e os respectivos vértices são removidos do grafo. A remoção de vértices fará com que vértices adjacentes passem a ter grau 1, então, o algoritmo de remoção de vértices de grau 1 é chamado novamente e se passa para uma próxima iteração pela busca de novos anéis. Quando o valor de *sumValency* chega a zero, todos os anéis foram identificados e o algoritmo se encerra. Entretanto, sistemas de anéis complexos não podem ser identificados pelos algoritmos descritos anteriormente. A figura 47 mostra uma situação deste tipo.

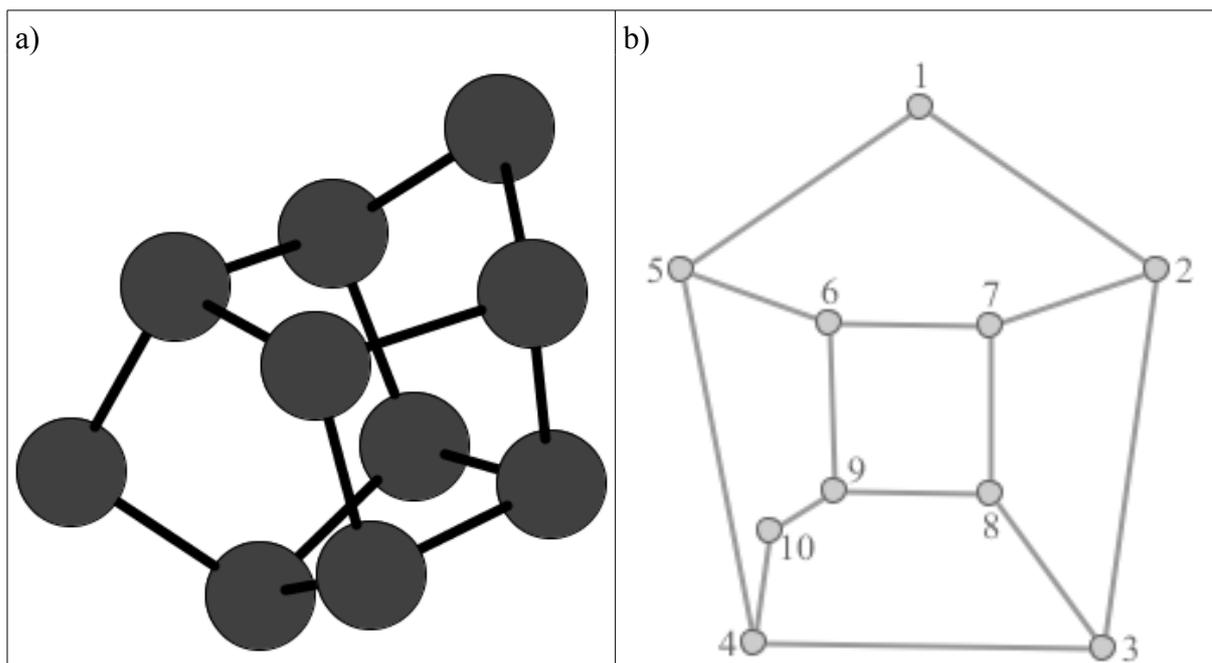


Figura 47 – Exemplo de sistemas de anéis complexos. Em anéis complexos, os átomos se organizam como vértices de um poliedro; os anéis individuais correspondem às faces desse poliedro.

Fonte: Do autor.

Legenda: a) Esquema mostrando uma estrutura de anéis formando um poliedro de seis faces;

b) Grafo representando o sistema de anéis mostrado em “a”.

Em situações como essa, os átomos se organizam geometricamente no espaço como vértices ou arestas de um poliedro. Procura-se identificar os anéis que formam as faces desse poliedro, tais como os anéis 1-2-7-6-5, 2-3-8-7, 3-4-10-9-8, 4-5-6-9-10, 6-7-8-9 e 1-2-3-4-5 vistos na figura 47. Primeiro calcula-se o número de faces desse poliedro da seguinte forma:

$$faces = arestas - vértices + 2$$

O número de vértices do poliedro é o somatório de todos os vértices do grafo cujo grau seja igual ou superior a três, enquanto que o número de arestas do poliedro é a metade do somatório dos graus dos vértices do grafo com grau igual ou superior a três. Por meio de uma análise combinatorial, os vértices são tomados três a três, quatro a quatro, e assim sucessivamente. Para cada subconjunto de vértices testa-se se estes formam um ciclo, da seguinte maneira:

- a) A partir da matriz de adjacência em análise (da qual foram eliminados os vértices não pertencentes a anéis ou que pertencem a sistemas de anéis simples, fundidos, espiranos ou pontes), extrai-se uma submatriz contendo apenas as linhas e colunas correspondentes ao subconjunto de vértices testados;

- b) Se o somatório de cada linha da submatriz for igual a dois, então o subconjunto de vértices faz parte de um anel, o qual é reportado;
- c) Caso contrário, o subconjunto é rejeitado;
- d) O teste é realizado até que o número de anéis encontrados seja igual ao número de faces calculadas.

3. 1. 4 Descritores Geométricos

Os descritores usados para caracterizar a geometria de uma molécula são a distância interatômica entre dois átomos, o ângulo formado entre três átomos consecutivos, o ângulo diedro formado entre quatro átomos consecutivos, o ângulo de Wilson ou diedro impróprio e a distância covalente. Tais descritores são necessários para o cálculo da energia potencial molecular e para os métodos de minimização de energia, conforme descrito em tópicos posteriores. O cálculo dos descritores geométricos é realizado por métodos da matemática vetorial. Dados dois átomos a e b , o vetor \vec{ab} é aquele que tem origem em a e destino em b (FIGURA 48).

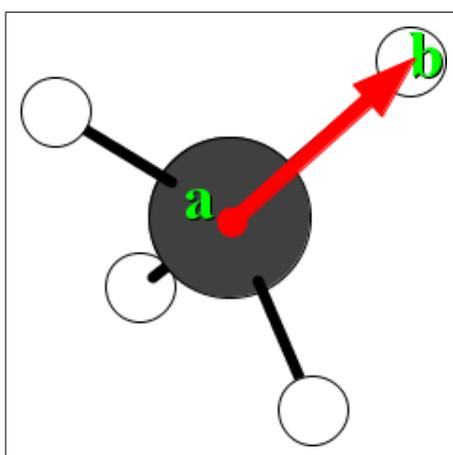


Figura 48 – Representação esquemática do vetor formado entre dois átomos.

Fonte: Do autor.

As coordenadas cartesianas x , y e z do vetor \vec{ab} são definidas da seguinte forma:

$$\begin{aligned}\vec{ab} \cdot x &= b \cdot x - a \cdot x \\ \vec{ab} \cdot y &= b \cdot y - a \cdot y \\ \vec{ab} \cdot z &= b \cdot z - a \cdot z\end{aligned}$$

Nas equações acima a expressão $\vec{ab} \cdot x$ significa “a coordenada x que pertence ao vetor \vec{ab} ” e assim por diante. A distância interatômica entre os átomos a e b corresponde à norma ou comprimento do vetor \vec{ab} calculada conforme a equação:

$$|\vec{ab}| = \sqrt{(x^2 + y^2 + z^2)}$$

Dado dois vetores \vec{ab} e \vec{ac} , o produto interno entre os mesmos é definido como (O caractere “ \cdot ” indica o produto interno e o caractere “ $*$ ” indica multiplicação de escalares):

$$\vec{ab} \cdot \vec{ac} = \vec{ab} \cdot x * \vec{ac} \cdot x + \vec{ab} \cdot y * \vec{ac} \cdot y + \vec{ab} \cdot z * \vec{ac} \cdot z$$

O ângulo θ entre os vetores é calculado como segue, correspondendo ao ângulo entre os átomos a , b e c , sendo a o átomo central (FIGURA 49):

$$\cos \Theta = \frac{\vec{ab} \cdot \vec{ac}}{|\vec{ab}| * |\vec{ac}|}$$

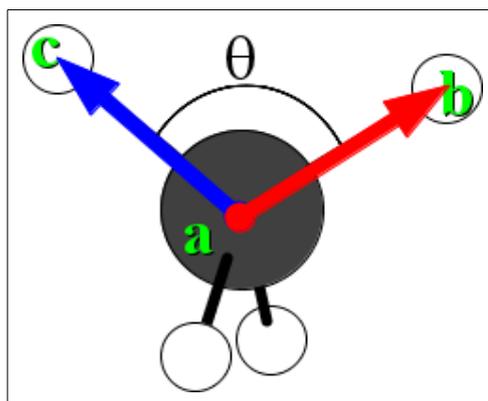


Figura 49 – Representação esquemática para o cálculo do ângulo entre três átomos.

Fonte: Do autor.

Outro cálculo importante na álgebra vetorial é o produto vetorial entre dois vetores, o qual é definido como (por simplificação as coordenadas de \vec{ab} estão identificadas com o

subscrito “1” e as de \vec{ac} com o subscrito “2”):

$$\vec{ab} \times \vec{ac} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \end{vmatrix}$$

$$\vec{n} = (y_1 * z_2 - z_1 * y_2) \vec{i} + (z_1 * x_2 - x_1 * z_2) \vec{j} + (x_1 * y_2 - y_1 * x_2) \vec{k}$$

$$\vec{n} \cdot x = y_1 * z_2 - z_1 * y_2$$

$$\vec{n} \cdot y = z_1 * x_2 - x_1 * z_2$$

$$\vec{n} \cdot z = x_1 * y_2 - y_1 * x_2$$

O resultado do produto vetorial é outro vetor \vec{n} , o qual é normal (forma um ângulo de 90°) a ambos \vec{ab} e \vec{ac} . Sendo assim, se os vetores \vec{ab} e \vec{ac} definem um plano entre três átomos, o produto vetorial é um modo eficaz de se identificar um vetor normal a esse plano.

O ângulo diedro é definido como o ângulo entre dois planos (FIGURA 50). Estando os átomos i, j, k e l covalentemente ligados em sequência, os átomos i, j e k definem o primeiro plano, enquanto que os átomos j, k e l definem o segundo plano. Cada plano define um vetor normal (vetores \mathbf{n}_{ab} e \mathbf{n}_{bc} na figura 50), o ângulo entre estes corresponde ao ângulo diedro.

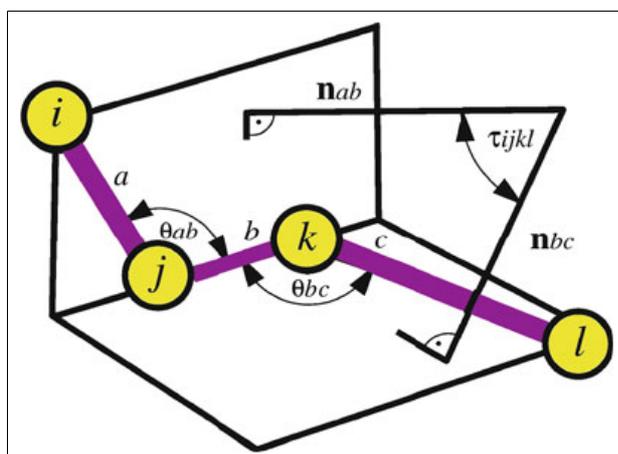


Figura 50 – Definição de ângulo diedro.
Fonte: Schlick (2010, p. 98).

Para se calcular o diedro, define-se o vetor \vec{ji} e o vetor \vec{jk} (plano i-j-k), a partir dos quais se calcula o vetor normal \vec{n}_{ijk} tal que $\vec{n}_{ijk} = \vec{ji} \times \vec{jk}$. Em seguida, define-se os vetores \vec{kj} e \vec{kl} (plano j-k-l), computa-se o vetor normal fazendo-se $\vec{n}_{jkl} = \vec{kj} \times \vec{kl}$. Então calcula-se o ângulo entre \vec{n}_{ijk} e \vec{n}_{jkl} conforme descrito anteriormente.

O ângulo de Wilson é definido para sistemas no qual um átomo está ligado a três outros átomos; calcula o quanto a organização espacial desses quatro átomos foge da planaridade, conforme se observa na figura 51. Na figura, B é o átomo central, o ângulo χ é definido como o ângulo entre a ligação $A-B$ e o plano $A-C-D$. Também são definidos os ângulos entre a ligação $B-C$ e o plano $A-C-D$, e entre a ligação $B-D$ e o plano $A-C-D$, de forma que para cada sistema de átomos trivalentes existirão três ângulos de Wilson.

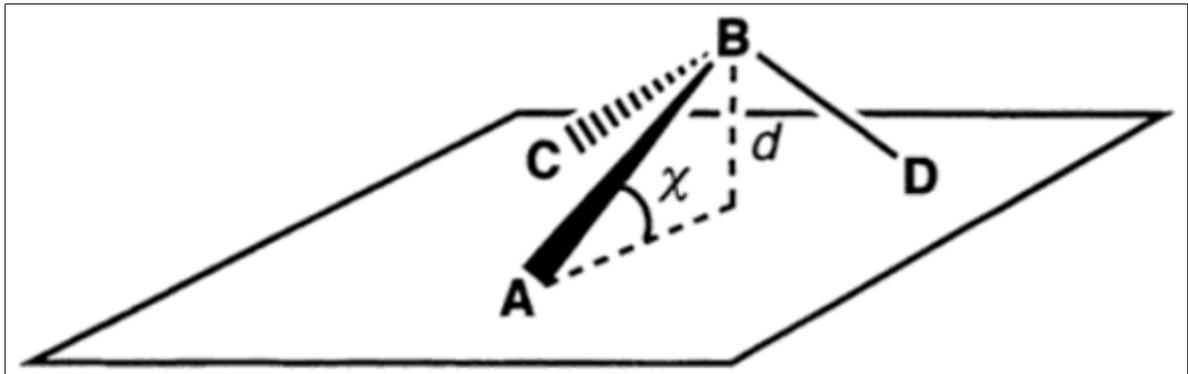


Figura 51 – Representação do ângulo de Wilson.
Fonte: Jensen (2007, p. 30).

Para se calcular o primeiro ângulo χ (ligação $A-B$), define-se os vetores \vec{BC} e \vec{BD} , a partir dos quais se computa o vetor normal $\vec{n} = \vec{BC} \times \vec{BD}$. Calcula-se então o ângulo entre os vetores \vec{AB} e \vec{n} , chegando ao resultado final. Os demais ângulos χ são calculados de maneira equivalente.

A distância covalente mede a menor distância entre dois átomos, definida como o número de ligações covalentes que se deve “atravessar” para se chegar de um átomo até o outro. Na figura 52, a distância covalente entre os átomos a e b é igual a 1, pois há apenas uma ligação covalente entre os mesmos (estão diretamente conectados); já entre os átomos a e c , a distância covalente é igual a 4, pois um caminho entre a e c envolve 4 ligações covalentes. Já a distância covalente entre b e c é igual a 3.


```

    Para cada Atom at ligado a root, Faça:
        Se at é o mesmo que endingAtom, Então:
            retorna d;
        Fim Se;

        Se at não está na lista markedAtoms, Então:
            adicionar at a markedAtoms;
            adicionar at a atomList;
            adicionar nextInt a distance;
        Fim Se;
    Fim Para;
Fim Enquanto;

{se a função chegar até aqui, startingAtom e endingAtom não
 estão na mesma molécula, retornar -1 para indicar isso}
retorna -1;
Fim Função;
Fim

```

3. 1. 5 Classe Molecule

Uma molécula completa é representada pela classe Molecule, a qual possui os atributos listados na tabela 6.

Tabela 6 – Atributos da classe Molecule

Atributo	Tipo	Descrição
atoms	Lista de Atom	Lista dos átomos da molécula
bonds	Lista de Bond	Lista das ligações covalentes da molécula
rings	Lista de Ring	Lista dos anéis da molécula
name	String	Nome da molécula
properties	Lista de String	Lista contendo propriedades da molécula
m2d	Molecule2D	Objeto para desenhar a molécula na tela
forceFieldParameters	ParameterTable	Tabela com parâmetros de campo de força
filterStatus	FilterStatus	Tipo enumerado que define se a molécula foi aprovada ou rejeitada em processo de filtragem de acordo com suas propriedades físico-químicas

Fonte: Do autor.

A molécula é construída adicionando-se à mesma seus átomos e suas ligações covalentes, geralmente lidos a partir de arquivos de texto, descrito na seção “3. 2 LEITURA E ESCRITA DE ARQUIVOS DE ESTRUTURAS MOLECULARES”. Os anéis são

identificados por meio dos algoritmos descritos anteriormente. A lista de propriedades é uma lista associativa, na qual o nome da propriedade está associada ao seu valor, o qual pode ser numérico ou uma string. Como exemplo, uma molécula de metano pode apresentar a propriedade “peso molecular”, cujo valor associado será 16,042. Algumas propriedades físico-químicas são computadas por algoritmos implementados no software, conforme será visto na seção “3. 6 CÁLCULO DE PROPRIEDADES *DRUG-LIKE*”. A tabela com parâmetros de campo de força é consultada toda vez que se calcula a energia potencial da molécula usando o campo de força MMFF94s. O tipo enumerado FilterStatus possui os valores “ACCEPTED”, “REJECTED”, e “NO_FILTERING”, indicando que a molécula foi aceita, rejeitada, ou ainda não foi filtrada, respectivamente. Regras empíricas, como as regras de Lipinski (LIPINSKI *et al.*, 1997) e as regras de Veber (VEBER *et al.*, 2002), são usadas para filtrar as moléculas (seção “3. 6 CÁLCULO DE PROPRIEDADES *DRUG-LIKE*”).

Alguns arquivos apresentam moléculas com hidrogênios implícitos, para adicionar os átomos de hidrogênio, emprega-se o seguinte algoritmo:

```
Início
Para cada Átomo atom da Molécula, Faça:

    inteiro : valency ← valência real do átomo, lida da lista de
                    propriedades do átomo;
    inteiro: aparentVal ← soma da ordem de todas as ligações
                    covalentes que incidem no átomo;
    inteiro : hydrogensToAdd ← valency - aparentVal;
    inteiro : i;

    Para i = 1 até hydrogensToAdd, Faça:
        criar novo átomo de hidrogênio hydrogen;
        adicionar átomo criado à molécula;
        criar ligação covalente entre atom e hydrogen;
        adicionar ligação covalente à molécula;
        determinar coordenadas cartesianas de hydrogen por
                    mecânica molecular;

    Fim Para;

Fim Para;
Fim
```

Em algumas aplicações é necessário se calcular o centro geométrico da molécula, como será visto na seção “3. 7 *VIRTUAL SCREENING*”. A função que computa o centro geométrico retorna este centro como um pseudo-átomo da seguinte forma:

```

Início
Função geometricCenter( ) : Atom

    real : mX ← 0;
    real : mY ← 0;
    real : mZ ← 0;

    Para cada Átomo atom da Molécula, Faça:
        mX ← mX + atom.x;
        mY ← mY + atom.y;
        mZ ← mZ + atom.z;
    Fim Para;

    int: size ← número total de átomos da molécula;
    mX ← mX / size;
    mY ← mY / size;
    mZ ← mZ / size;

    criar novo átomo newAtom;
    newAtom.x ← mX;
    newAtom.y ← mY;
    newAtom.z ← mZ;
    retorna newAtom;

Fim Função;
Fim

```

Também se define o raio molecular, que é a maior distância entre um átomo da molécula e seu centro geométrico:

```

Início
Função moleculeRadius( ) : real

    Atom : center ← geometricCenter( );
    real : radius ← 0;

    Para cada Átomo atom da Molécula, Faça:
        real : distance ← distância entre atom e center;
        Se radius < distance, Então:
            radius ← distance;
        Fim Se;
    Fim Para;

    retorna radius;

Fim Função;
Fim

```

3. 2 LEITURA E ESCRITA DE ARQUIVOS DE ESTRUTURAS MOLECULARES

Os dados necessários para se construir uma molécula em um computador são escritos em arquivos de texto com formatos padronizados. Optou-se por dar suporte a leitura e a escrita de formatos de arquivo comumente usados em química computacional, os quais são: mol (versões V2000 e V3000), sdf (versões V2000 e V3000), mol2 e pdb. Também é suportado o formato pdbqt apenas para escrita, a fim de ser usado em *virtual screening* (discutido na seção “3. 7 VIRTUAL SCREENING”). Embora diversos entre si, a maioria dos formatos possuem três blocos de dados em comum:

- a) Um cabeçalho contendo informações básicas da molécula, como nome e quantidade de átomos e ligações covalentes que possui;
- b) Uma tabela contendo dados sobre os átomos que constituem a molécula e suas coordenadas cartesianas;
- c) Uma tabela contendo as ligações covalentes presentes na molécula.

3. 2. 1 Formato mol

O formato de arquivo mol foi desenvolvido pela Accelrys para descrever a correlação estrutural e as propriedades de um conjunto de átomos (ACCELRYS, 2011), armazenando informações para uma única molécula. Possui duas versões: V2000 e V3000. A versão V2000 possui um cabeçalho de três linhas: a primeira linha contém o nome da molécula, a segunda o nome do software que gerou essa molécula, a terceira linha é para se escrever algum comentário. Qualquer uma dessas linhas ou até mesmo todas podem estar em branco, mas a presença dessas é indispensável. A próxima linha é chamada linha de contagem, traz como principais informações o número de átomos e ligações da molécula, além da versão em que o arquivo está escrito. Em seguida vem o bloco dos átomos, no qual cada linha contém a informação de um átomo dividida em dezesseis campos: os três primeiros campos representam as coordenadas x, y e z ocupadas pelo átomo, o quarto campo é o símbolo atômico; os demais campos contém dados não considerados durante a leitura do arquivo (a maioria dos arquivos mol possuem esses campos com valor *default zero*). Após o bloco dos átomos, há o bloco das ligações covalentes, cada linha representando uma ligação. Estas

possuem sete campos: os dois primeiros são os índices dos átomos conectados, o terceiro é o tipo de ligação (simples: 1, dupla: 2, tripla: 3, aromática: 4), os demais campos são ignorados (geralmente possuem o valor *default zero*). A última linha é escrita como “M END” e marca o fim do arquivo. A molécula de metano em formato mol V2000 é apresentada a seguir:

```
metano
5 4 0 0 0 0 0 0 0 0999 V2000
0.9831 -0.0750 0.0677 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.0742 -0.0283 0.0736 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.5813 0.8494 -0.3526 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.6181 -0.1998 1.0890 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.6585 -0.9226 -0.5411 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
1 3 1 0 0 0 0
1 4 1 0 0 0 0
1 5 1 0 0 0 0
M END
```

A versão V3000 possui o mesmo cabeçalho e linha de contagem encontrados na versão V2000, os blocos de átomos e ligações porém são diferentes. As demais linhas após o cabeçalho começam com a inscrição “M V30”. O bloco de átomos é delimitado pelas linhas “M V30 BEGIN ATOM” e “M V30 END ATOM”, enquanto que o bloco de ligações é delimitado pelas linhas “M V30 BEGIN BOND” e “M V30 END BOND”. Estes dois blocos são por sua vez conjuntamente delimitados pelas linhas “M V30 BEGIN CTAB” e “M V30 END CTAB”. No bloco dos átomos, cada átomo pode ter suas informações escritas em uma ou mais linhas; o primeiro campo contém o índice numérico do átomo iniciando em 1, em seguida estão as coordenadas x, y e z, logo após vem o elemento químico do átomo, outros campos complementares podem estar presentes, mas são ignorados. No bloco de ligações, uma ligação covalente pode ter suas informações em uma ou mais linhas, os campos consistem no índice numérico da ligação iniciando em 1, o tipo de ligação (simples: 1, dupla: 2, tripla: 3, aromática: 4) e os índices dos átomos participantes da ligação. O formato mol V3000 para o metano é o seguinte:

```
metano
0 0 0 0 0 999 V3000
M V30 BEGIN CTAB
M V30 COUNTS 5 4 0 0 0
M V30 BEGIN ATOM
M V30 1 C 0.9831 -0.0750 0.0677 0
```

```

M V30 2 H 2.0742 -0.0283 0.0736 0
M V30 3 H 0.5813 0.8494 -0.3526 0
M V30 4 H 0.6181 -0.1998 1.0890 0
M V30 5 H 0.6585 -0.9226 -0.5411 0
M V30 END ATOM
M V30 BEGIN BOND
M V30 1 1 1 2
M V30 2 1 1 3
M V30 3 1 1 4
M V30 4 1 1 5
M V30 END BOND
M V30 END CTAB
M END

```

3. 2. 2 Formato sdf

O formato sdf, acrônimo de *structure-data file* (arquivo de dados de estrutura), é uma extensão do formato mol. Possui em acréscimo um conjunto de propriedades não estruturais escritas como:

```

> <nome_da_propriedade>
valor_da_propriedade

```

Há uma liberdade quanto aos tipos e quantidade de propriedades reportadas. Outra característica do formato sdf é que o mesmo pode conter várias moléculas em um único arquivo, uma linha do tipo “\$\$\$\$” faz a separação entre as diferentes moléculas. A molécula de metano é apresentada a seguir em formato sdf V2000.

```

metano

5 4 0 0 0 0 0 0 0 0999 V2000
0.9831 -0.0750 0.0677 C 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
2.0742 -0.0283 0.0736 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.5813 0.8494 -0.3526 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.6181 -0.1998 1.0890 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
0.6585 -0.9226 -0.5411 H 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
1 2 1 0 0 0 0
1 3 1 0 0 0 0
1 4 1 0 0 0 0
1 5 1 0 0 0 0
M END
> <molecular_weight>
16.042

```

```

> <logP>
0.000

> <logS>
0.562

> <TPSA>
0.000

> <number_of_hydrogen_bond_donors>
0

> <number_of_hydrogen_bond_acceptors>
0

$$$$

```

3. 2. 3 Formato mol2

O formato mol2 foi desenvolvido pela SYBYL como uma representação molecular portátil e completa (SYBYL, 2005). Cada bloco de dados inicia-se com um indicador do tipo “@<TRIPOS>TIPO_DE_REGISTRO”, seguido pelos respectivos dados nas linhas subsequentes. O cabeçalho inicia-se após o indicador “@<TRIPOS>MOLECULE” e é composto por seis linhas: a primeira traz o nome da molécula, a segunda o número de átomos e ligações, a terceira o tipo de molécula (SMALL, BIOPOLYMER, PROTEIN, NUCLEIC_ACID, SACCHARIDE), a quarta o tipo de carga associada à molécula (NO_CHARGES, DEL_RE, GASTEIGER, GAST_HUCK, HUCKEL, PULLMAN, GAUSS80_CHARGES, AMPAC_CHARGES, MULLIKEN_CHARGES, DICT_CHARGES, MMFF94_CHARGES, USER_CHARGES), a quinta o status da molécula e a última é uma linha de comentário. As duas últimas linhas podem estar ausentes. O bloco de átomos inicia-se pelo indicador “@<TRIPOS>ATOM”, para cada linha que representa um átomo os dados lidos são: o índice do átomo, seu nome, suas coordenadas x, y, e z, e seu tipo conforme as convenções SYBYL (TABELA 7). Os demais dados, se presentes (índice da subestrutura que contém o átomo, nome da subestrutura que contém o átomo, carga do átomo e *status* do átomo) são ignorados na leitura. O bloco de ligações é marcado com o indicador “@<TRIPOS>BOND” e cada linha possui as seguintes informações: o índice da ligação covalente, o índice de cada átomo conectado por essa ligação e o tipo de ligação (simples: 1, dupla: 2, tripla: 3, aromática: ar, amídica: am).

Tabela 7 – Tipo de átomo conforme as especificações SYBYL

Definição	Código
Hidrogênio	H
Carbono sp ³	C.3
Carbono sp ²	C.2
Carbono sp	C.1
Carbono aromático	C.ar
Carbocátion (guanidínio)	C.cat
Nitrogênio sp ³	N.3
Nitrogênio sp ²	N.2
Nitrogênio sp	N.1
Nitrogênio aromático	N.ar
Nitrogênio amida	N.am
Nitrogênio trigonal planar	N.pl3
Nitrogênio sp ³ carregado	N.4
Oxigênio sp ³	O.3
Oxigênio sp ²	O.2
Oxigênio em carboxilatos e fosfatos	O.co2
Enxofre sp ³	S.3
Enxofre sp ²	S.2
Enxofre sulfóxido	S.o
Enxofre sulfona	S.o2
Fósforo sp ³	P.3
Demais elementos	O símbolo químico do elemento

Fonte: Adaptado de SYBYL, 2005.

O arquivo mol2 para o metano é mostrado a seguir:

```
@<TRIPOS>MOLECULE
metano
  5 4 0 0 0
SMALL
GASTEIGER

@<TRIPOS>ATOM
  1 C    0.9831  -0.0750   0.0677 C.3    1  LIG1    -0.0776
  2 H    2.0742  -0.0283   0.0736 H      1  LIG1     0.0194
  3 H    0.5813   0.8494  -0.3526 H      1  LIG1     0.0194
  4 H    0.6181  -0.1998   1.0890 H      1  LIG1     0.0194
  5 H    0.6585  -0.9226  -0.5411 H      1  LIG1     0.0194

@<TRIPOS>BOND
  1      1      2      1
```

2	1	3	1
3	1	4	1
4	1	5	1

3. 2. 4 Formato pdb

O formato pdb (de *Protein Data Bank*) contém informações de estruturas obtidas experimentalmente por métodos como cristalografia de raios X ou NMR (WWPDB, 2008). Armazena principalmente dados de macromoléculas biológicas como proteínas e ácidos nucleicos, mas pode conter também estruturas de pequenas moléculas cocrystalizadas. Além das coordenadas atômicas, também estão presentes informações como os parâmetros utilizados nos experimentos de raios X/NMR, a sequência primária e estruturas secundárias da proteína, dados de literatura e referência cruzada com bancos de dados. Cada linha inicia-se com uma *tag* ou etiqueta, que identifica o tipo de informação que a mesma contém. As *tags* de interesse para ler/escrever uma molécula genérica (que não uma biomacromolécula) são:

- HETATM: informa um átomo pertencente à molécula. Catorze campos estão contidos nessa tag e aparecem na seguinte ordem: índice do átomo, seu nome, sua localização alternada (se houver), código do resíduo ou molécula ao qual pertence, cadeia onde se encontra, número sequencial de seu resíduo ou molécula, código de inserção (se houver), coordenadas cartesianas x, y e z, ocupância, fator de temperatura, símbolo químico e carga.
- CONNECT: informa quais átomos estão conectados entre si. Diferente dos formatos mostrados anteriormente, em que cada linha de ligação química reporta apenas os dois átomos que estão conectados e qual o tipo da ligação, a *tag* CONNECT apresenta o índice de um átomo seguido de uma lista de índices de todos os átomos conectados com o primeiro. O tipo de ligação (simples, dupla, tripla, aromática) não é reportado.

A molécula anilina é mostrada no formato pdb, átomos de hidrogênio estão implícitos:

HETATM	1	C	UNL	1	-0.803	1.161	-0.014	1.00	0.00	C
HETATM	2	C	UNL	1	0.591	1.245	-0.007	1.00	0.00	C
HETATM	3	C	UNL	1	1.359	0.083	-0.015	1.00	0.00	C
HETATM	4	C	UNL	1	0.731	-1.160	-0.010	1.00	0.00	C
HETATM	5	C	UNL	1	-0.664	-1.238	-0.019	1.00	0.00	C
HETATM	6	C	UNL	1	-1.444	-0.080	-0.076	1.00	0.00	C
HETATM	7	N	UNL	1	-2.837	-0.161	0.039	1.00	0.00	N

CONNECT	1	2	6	
CONNECT	2	1	3	
CONNECT	3	2	4	
CONNECT	4	3	5	
CONNECT	5	4	6	
CONNECT	6	5	1	7
CONNECT	7	6		

Para se fazer a leitura das estruturas de biomacromoléculas (proteínas, DNA e RNA), busca-se pela *tag* ATOM, a qual possui os mesmos campos da *tag* HETATM. A *tag* CONNECT não é definida para biomacromoléculas, devendo as ligações covalentes serem determinadas internamente pelo programa. A leitura das biomacromoléculas será importante na extração de ligantes a partir de complexos receptor-ligante e na execução de *virtual screening* (seção “3. 5 ANÁLISE DE INTERAÇÕES RECEPTOR-LIGANTE” e seção “3. 7 VIRTUAL SCREENING”, respectivamente).

3. 2. 5 Formato pdbqt

O formato de arquivo pdbqt é uma extensão do formato pdb, o qual acrescenta para as *tags* ATOM/HETATM a carga do átomo e o tipo de átomo de acordo com as especificações AutoDock (THE SCRIPPS RESEARCH INSTITUTE, 2007a) (Tabela 8). É um formato de arquivo não padrão usado para *molecular docking* com os programas AutoDock 4, desenvolvido por Morris *et al.* (2009) e AutoDock Vina (TROTT E OLSON, 2010). Para as estruturas de ligante, acrescenta também informações quanto às ligações covalentes com liberdade de torção em seus diedros. O ligante é dividido em sub-regiões, com uma região *root*, geralmente a maior, a qual é rígida, e, caso existam, cadeias laterais com liberdade de torção. A região *root* é delimitada pelas *tags* ROOT e ENDROOT e as cadeias laterais pelas *tags* BRANCH e ENDBRANCH. Dois valores inteiros após a *tag* BRANCH indicam os índices dos átomos que pertencem à ligação covalente com liberdade de rotação. Um exemplo de arquivo pdbqt é mostrado a seguir (THE SCRIPPS RESEARCH INSTITUTE, 2007a):

COMPND	NSC7810
REMARK	3 active torsions:
REMARK	status: ('A' for Active; 'I' for Inactive)
REMARK	1 A between atoms: A7_7 and C22_23
REMARK	2 A between atoms: A9_9 and A11_11
REMARK	3 A between atoms: A17_17 and C21_21
ROOT	

ATOM	1	A1	INH	I	1.054	3.021	1.101	0.00	0.00	0.002	A
ATOM	2	A2	INH	I	1.150	1.704	0.764	0.00	0.00	0.012	A
ATOM	3	A3	INH	I	-0.006	0.975	0.431	0.00	0.00	-0.024	A
ATOM	4	A4	INH	I	0.070	-0.385	0.081	0.00	0.00	0.012	A
ATOM	5	A5	INH	I	-1.062	-1.073	-0.238	0.00	0.00	0.002	A
ATOM	6	A6	INH	I	-2.306	-0.456	-0.226	0.00	0.00	0.019	A
ATOM	7	A7	INH	I	-2.426	0.885	0.114	0.00	0.00	0.052	A
ATOM	8	A8	INH	I	-1.265	1.621	0.449	0.00	0.00	0.002	A
ATOM	9	A9	INH	I	-1.339	2.986	0.801	0.00	0.00	-0.013	A
ATOM	10	A10	INH	I	-0.176	3.667	1.128	0.00	0.00	0.013	A
ENDROOT											
BRANCH 9 11											
ATOM	11	A11	INH	I	-2.644	3.682	0.827	0.00	0.00	-0.013	A
ATOM	12	A16	INH	I	-3.007	4.557	-0.220	0.00	0.00	0.002	A
ATOM	13	A12	INH	I	-3.522	3.485	1.882	0.00	0.00	0.013	A
ATOM	14	A15	INH	I	-4.262	5.209	-0.177	0.00	0.00	-0.024	A
ATOM	15	A17	INH	I	-2.144	4.784	-1.319	0.00	0.00	0.052	A
ATOM	16	A14	INH	I	-5.122	4.981	0.910	0.00	0.00	0.012	A
ATOM	17	A20	INH	I	-4.627	6.077	-1.222	0.00	0.00	0.012	A
ATOM	18	A13	INH	I	-4.749	4.135	1.912	0.00	0.00	0.002	A
ATOM	19	A19	INH	I	-3.777	6.285	-2.267	0.00	0.00	0.002	A
ATOM	20	A18	INH	I	-2.543	5.650	-2.328	0.00	0.00	0.019	A
BRANCH 15 21											
ATOM	21	C21	INH	I	-0.834	4.113	-1.388	0.00	0.00	0.210	C
ATOM	22	01	INH	I	-0.774	2.915	-1.581	0.00	0.00	-0.644	OA
ATOM	23	03	INH	I	0.298	4.828	-1.237	0.00	0.00	-0.644	OA
ENDBRANCH 15 21											
ENDBRANCH 9 11											
BRANCH 7 24											
ATOM	24	C22	INH	I	-3.749	1.535	0.125	0.00	0.00	0.210	C
ATOM	25	02	INH	I	-4.019	2.378	-0.708	0.00	0.00	-0.644	OA
ATOM	26	04	INH	I	-4.659	1.196	1.059	0.00	0.00	-0.644	OA
ENDBRANCH 7 24											
TORSDOF 3											

Tabela 8 - Tipos de átomos de acordo com a especificação AutoDock.

Tipo	Descrição	Tipo	Descrição
H	Hidrogênio apolar	Mg, MG	Magnésio
HD	Hidrogênio doador de interações-H*	P	Fósforo
HS	Hidrogênio doador S-esférico	SA	Enxofre acceptor de 2 interações-H
C	Carbono alifático	S	Enxofre não participante de interações-H
A	Carbono aromático	Cl, CL	Cloro
N	Nitrogênio não participante de interações-H	Ca, CA	Cálcio
NA	Nitrogênio acceptor de 1 interação-H	Mn, MN	Manganês
NS	Nitrogênio acceptor S-esférico	Fe, FE	Ferro
OA	Oxigênio acceptor de 2 interações-H	Zn, ZN	Zinco
OS	Oxigênio acceptor S-esférico	Br, BR	Bromo
F	Flúor	I	Iodo

Fonte: Adaptado de The Scripps Research Institute, 2007b

Nota: *Interações-H: Interações de hidrogênio

3. 2. 6 Algoritmos para leitura e edição de moléculas

De um modo geral, a leitura de arquivos moleculares segue os seguintes passos:

```

Início

criar Molécula; {inicialmente vazia}
abrir Arquivo para leitura;

Para cada Linha do Arquivo, Faça:
  Ler a Linha;

  Se Linha contém dados de um Átomo, Então:
    criar Átomo a partir dos dados de Linha;
    adicionar Átomo à Molécula;
  Fim Se;

  Se Linha contém dados de uma Ligação, Então:
    buscar em Molécula Átomo1 cujo índice é referenciado em Linha;
    buscar em Molécula Átomo2 cujo índice é referenciado em Linha;
    criar Ligação entre Átomo1 e Átomo2;
    adicionar Ligação em Molécula;
  Fim Se;
Fim Para;

fechar Arquivo;
identificar anéis em Molécula;
calcular parâmetros de campo de força da Molécula;

Fim

```

Quando se lê um arquivo em formato mol, sdf ou pdb e se pretende escrevê-lo em mol2, é necessário para cada átomo definir seu tipo de acordo com as especificações SYBYL (TABELA 7). Para realizar essa tarefa, implementou-se o algoritmo descrito pelo CCMS (*Computational Center for Macromolecular Structure*) obtido no endereço eletrônico http://www.sdsc.edu/CCMS/Packages/cambridge/pluto/atom_types.html, o qual é apresentado a seguir com algumas adaptações:

```

Início

Para cada Átomo da Molécula, Faça:

  string : element ← elemento químico do átomo;
  inteiro : num_bond ← número de átomos conectados ao Átomo;
  inteiro : num_nonmet ← número de átomos não-metais ligados ao Átomo;
  inteiro : singleBonds ← número de ligações simples do Átomo;
  inteiro : doubleBonds ← número de ligações duplas do Átomo;
  inteiro : tripleBonds ← número de ligações triplas do Átomo;
  inteiro : aromaticBonds ← número de ligações aromáticas do Átomo;

```

```

string : mol2Type; {tipo SYBYL do átomo que se quer determinar}

Se element = "D", Então:
    mol2Type ← "H";

Senão Se element = "P", Então:
    mol2Type ← "P.3";

Senão Se element = "Co", Então:
    mol2Type ← "Co.oh";

Senão Se element = "Ru", Então:
    mol2Type ← "Ru.oh";

Senão Se element = "C", Então:

    Se num_bond >= 4 E singleBonds = num_bond, Então:
        mol2Type ← "C.3";
    Senão Se num_bond = 3 E todas as ligações são com nitrogênios
        em um sistema guanidina, Então:
        mol2Type ← "C.cat";
    Senão Se num_bond >= 2 E aromaticBonds = 2, Então:
        mol2Type ← "C.ar";
    Senão Se num_bond <= 2 E tripleBonds = 1, Então:
        mol2Type ← "C.1";
    Senão:
        mol2Type ← "C.2";
    Fim Se;

Senão Se element = "O", Então:

    Se num_nonmet = 1, Então:
        Se a ligação ocorre em carboxilato, Então:
            mol2Type ← "O.co2";
        Senão Se a ligação ocorre em fosfato, Então:
            mol2Type ← "O.co2"
        Fim Se;
    Senão Se num_bond >= 2 E singleBonds = num_bond, Então:
        mol2Type ← "O.3";
    Senão:
        mol2Type ← "O.2";
    Fim Se;

Senão Se element = "N", Então:

    Se num_nonmet = 4 E singleBonds = num_bond, Então:
        mol2Type ← "N.4";
    Senão Se num_bond >= 2 E aromaticBonds = 2, Então:
        mol2Type ← "N.ar";
    Senão Se num_nonmet = 1 E tripleBonds = 1, Então:
        mol2Type ← "N.1";
    Senão Se num_nonmet = 2 E (doubleBonds = 2 OU
        singleBonds + tripleBonds = 2), Então:
        mol2Type ← "N.1";
    Senão Se num_nonmet = 3 E uma ligação é para C=O OU C=S, Então:
        mol2Type ← "N.am";
    Senão Se num_nonmet = 3, Então:
        Se singleBonds = 2, Então:
            mol2Type ← "N.pl3";

```

```

        Senão Se singleBonds = 3, Então:
            inteiro : HBond ← número de hidrogênios conectados
                ao Átomo;
            inteiro : delBond ← número de átomos com pelo menos
                uma ligação pi, conectados ao
                Átomo;
            real : angle ← soma dos ângulos ao redor do Átomo;
            Se delBond = 1 E HBond = 1, Então:
                mol2Type ← "N.pl3";
            Senão Se delBond = 1 E HBond = 0 E angle >= 350,
                Então:
                    mol2Type ← "N.pl3";
            Fim Se;
        Senão:
            mol2Type ← "N.3";
        Fim Se;
    Senão:
        mol2Type ← "N.2";
    Fim Se;

Senão Se element = "S", Então:

    inteiro : oxy_nonmet ← número de átomos de oxigênio com apenas
        uma ligação não-metal conectados ao Átomo;

    Se num_nonmet = 3 E oxy_nonmet = 1, Então:
        mol2Type ← "S.o";
    Senão Se num_nonmet = 4 E oxy_nonmet = 2, Então:
        mol2Type ← "S.o2";
    Senão Se num_bond >= 2 E singleBonds = num_bond, Então:
        mol2Type ← "S.3";
    Senão
        mol2Type ← "S.2";
    Fim Se;

Senão Se element = "Ti", Então:
    Se num_bond <= 4, Então:
        mol2Type ← "Ti.th";
    Senão:
        mol2Type ← "Ti.oh";
    Fim Se;

Senão Se element = "Cr", Então:
    Se num_bond <= 4, Então:
        mol2Type ← "Cr.th";
    Senão:
        mol2Type ← "Cr.oh";
    Fim Se;

Senão
    mol2Type ← element;
Fim Se;

Fim Para;

Fim

```

O formato mol2 também define a ligação covalente do tipo amida (am) não especificada nos outros formatos. A ligação amida é aquela que ocorre entre um nitrogênio e um carbono da carbonila. Para se identificar uma ligação do tipo amida, se adota o seguinte procedimento:

```

Início
Para cada Ligação da Molécula, Faça:

    Átomo : nitrogen;
    Átomo : carbon;

    string : element1 ← elemento químico do primeiro átomo de Ligação;
    string : element2 ← elemento químico do segundo átomo de Ligação;

    Se element1 = "N" E element2 = "C", Então:
        nitrogen ← primeiro átomo de Ligação;
        carbon ← segundo átomo de Ligação;
    Senão Se element1 = "C" E element2 = "N", Então:
        nitrogen ← segundo átomo de Ligação;
        carbon ← primeiro átomo de Ligação;
    Senão:
        próximo;
    Fim Se;

    real : order ← ordem da Ligação;

    Se order > 1, Então:
        próximo;
    Fim Se;

    Para cada Átomo conectado a carbon, Faça:
        string : element ← elemento químico de Átomo;
        order ← ordem de ligação entre carbon e Átomo;
        Se (element = "O" OU element = "S") E order = 2, Então:
            definir Ligação como do tipo amida;
        Fim Se;

    Fim Para;

Fim Para;
Fim

```

Arquivos pdb não possuem informação quanto à ordem de ligação das ligações covalentes. Estas são inferidas baseando-se na distância e na natureza dos átomos envolvidos na ligação, conforme foi visto na seção “3. 1. 2 Classe Bond”. Da mesma forma, é possível inferir as ligações se as *tags* CONECT estiverem ausentes.

3. 3 OTIMIZAÇÃO DA GEOMETRIA MOLECULAR

A energia potencial de uma molécula é determinada em função da posição relativa dos átomos que a constituem. A minimização da geometria molecular tem por objetivo encontrar as coordenadas cartesianas atômicas que correspondam à configuração mais estável da molécula, que por sua vez será aquela que possui o menor valor de energia potencial. Adotou-se o modelo da mecânica molecular, no qual átomos são tratados como esferas rígidas e ligações covalentes como molas conectando essas esferas (CRAMER, 2013). Nesse modelo, define-se uma geometria idealizada baseada nas distâncias de ligações covalentes em equilíbrio e nos ângulos e diedros em equilíbrio, a partir da qual a geometria é distorcida para acomodar interações não ligantes (van der Waals e eletrostática) (HEHRE, 2003). A equação genérica para a energia potencial molecular é apresentada a seguir:

$$E = \sum E_{\text{lig}} + \sum E_{\text{não lig}} + \sum E_{\text{cruz}}$$

em que E_{lig} é a soma das energias provenientes da interação de átomos diretamente conectados ou interações ligantes (ligações covalentes, ângulos, ângulos diedros, diedros impróprios); $E_{\text{não lig}}$ é a energia entre átomos não ligados diretamente ou interações não ligantes (interações de van der Waals e Coulombianas), E_{cruz} é um termo que expressa a interação que pode ocorrer entre os termos de interações ligantes (covalente-ângulo, ângulo-diedro, etc). Constantes de força são empregadas para controlar a distorção da geometria a partir da geometria de equilíbrio. Estas são específicas para cada par, trio ou quadra de átomos que interagem, dependendo da natureza destes (elemento químico, hibridização e ordem das ligações covalentes para interações ligantes). Ao conjunto de parâmetros calibrados e tabelados (constantes de força e distâncias e ângulos de equilíbrio) mais a equação que descreve a energia potencial molecular dá-se o nome de campo de força. Campos de força populares incluem UFF, CFF93, MM2, MM3, MM4, MMFF94, CVFF, AMBER, CHARMM, GROMOS e OPLS-AA (VANOMMESLAEGHE, GUVENCH E MACKERELL, 2014). Cada campo de força possui sua própria forma funcional e seus respectivos parâmetros, podendo ser de uso geral ou aplicável para determinado grupo de compostos.

3.3.1 Campo de força MMFF94s

O campo de força MMFF94 (*Merck Molecular Force Field*) foi desenvolvido nos laboratórios de pesquisa da Merck. É aplicado para simulações de dinâmica molecular em fase condensada, sendo otimizado para compostos de interesse da química orgânica e da química medicinal em combinação com simulação de proteínas (HALGREN, 1996). Pelo fato de ser parametrizado para estudos de dinâmica molecular, o campo de força MMFF94 não é aplicável para procedimentos de minimização de energia quando a molécula apresenta átomos de nitrogênio tricoordenados, cujos cálculos resultarão em uma geometria não-planar até mesmo para amidas; sendo assim, os laboratórios da Merck desenvolveram uma versão alternativa do campo de força, denominado MMFF94s, para estudos de minimização de geometria, o qual é implementado no presente trabalho. Os campos de força MMFF94 e MMFF94s são similares, apresentando a mesma forma funcional, a diferença consiste no valor das constantes de força para as interações de diedro impróprio e interações torcionais envolvendo átomos de nitrogênio mencionados acima; para as demais interações, as constantes de força são as mesmas (HALGREN, 1999).

A equação empregada pelo campo de força MMFF94 é da seguinte forma:

$$E_{\text{MMFF}} = \sum EB_{ij} + \sum EA_{ijk} + \sum EBA_{ijk} + \sum E\text{OOP}_{ijk;l} + \sum ET_{ijkl} + \sum E\text{vdW}_{ij} + \sum EQ_{ij}$$

O termo EB se refere ao estiramento de ligação entre dois átomos i e j ligados covalentemente, sendo:

$$EB_{ij} = 143,9325 \frac{kb_{ij}}{2} \Delta r_{ij}^2 \times \left(1 + cs \Delta r_{ij} + \frac{7}{12} cs^2 \Delta r_{ij}^2 \right)$$

em que kb_{ij} é uma constante de força que depende do par de átomos conectados, Δr_{ij} é a diferença entre o comprimento da ligação covalente e o comprimento de referência, e cs é uma constante, $cs = -2 \text{ \AA}^{-1}$.

O termo EA corresponde à aberta angular entre três átomos i , j e k diretamente conectados, escrita da seguinte forma:

$$EA_{ijk} = 0,043844 \frac{ka_{IJK}}{2} \Delta \vartheta_{ijk}^2 (1 + cb \Delta \vartheta_{ijk})$$

em que ka_{IJK} é uma constante de força, $\Delta \vartheta$ é a diferença entre o ângulo entre os átomos e o ângulo de referência e $cb = -0,4 \text{ rad}^{-1}$.

EBA é um termo cruzado para a interação entre o estiramento de ligação covalente e o ângulo entre três átomos:

$$EBA_{ijk} = 2,51210 (kba_{IJK} \Delta r_{ij} + kba_{KJI} \Delta r_{kj}) \Delta \vartheta_{ijk}$$

kba_{IJK} e kba_{KJI} são constantes de força para a interação covalente dos átomos $i-j$ e $k-j$ que formam um ângulo $i-j-k$, Δr_{ij} é a diferença entre o comprimento da ligação covalente e o comprimento de referência e $\Delta \vartheta$ é a diferença entre o ângulo entre os átomos e o ângulo de referência.

EOOP corresponde à interação de desvio de planaridade ou diedro impróprio, aplica-se a átomos com número de coordenação igual a três que formam um centro não planar, possui a seguinte forma:

$$EOOP_{ijk;l} = 0,043844 \frac{koop_{IJK:L}}{2} \chi_{ijk;l}^2$$

$koop_{IJK:L}$ é uma constante de força e $\chi_{ijk;l}$ é o ângulo de Wilson entre a ligação $j-l$ e o plano $i-j-k$. O átomo j corresponde ao átomo central. Sendo possível calcular três diferentes ângulos de Wilson, dependendo de qual átomo periférico é selecionado como átomo l , o termo EOOP é calculado para cada χ possível, o resultado final é o somatório de todos os cálculos efetuados.

ET corresponde à interação entre quatro átomos ligados em sequência devido ao ângulo diedro entre os mesmos, cuja equação é:

$$ET_{ijkl} = 0,5 (V_1 (1 + \cos \Phi) + V_2 (1 - \cos 2 \Phi) + V_3 (1 + \cos 3 \Phi))$$

V_1 , V_2 e V_3 são constantes que dependem dos tipos de átomos i , j , k e l , em que as ligações covalentes são entre os pares $i-j$, $j-k$ e $k-l$, e Φ é o ângulo diedro.

A interação de van der Waals é definida no termo EvdW, descrita pela equação:

$$E_{vdW_{ij}} = \varepsilon_{IJ} \left(\frac{1,07 R_{IJ}^*}{R_{ij} + 0,07 R_{IJ}^*} \right)^7 \left(\frac{1,12 R_{IJ}^{*7}}{R_{ij}^7 + 0,12 R_{IJ}^{*7}} - 2 \right)$$

em que as constantes ε_{IJ} e R_{IJ}^* são calculadas de acordo com os pares de átomos i e j interagindo. A interação de van der Waals é calculada apenas para pares de átomos separados por três ou mais ligações covalentes. Se os átomos envolvidos na interação são do mesmo elemento químico, a constante R_{IJ}^* é reescrita como R_{II}^* e calculada da seguinte forma:

$$R_{II}^* = A_I \alpha_I^{0,25}$$

em que A_I é um fator de escala e α_I é a polarizabilidade atômica. Se os átomos que interagem são de elementos químicos diferentes, se calcula R_{IJ}^* como uma média aritmética aumentada:

$$R_{IJ}^* = 0,5(R_{II}^* + R_{JJ}^*)(1 + B(1 - \exp(-\beta * \gamma_{IJ}^2)))$$

$$\gamma_{IJ} = \frac{(R_{II}^* - R_{JJ}^*)}{(R_{II}^* + R_{JJ}^*)}$$

em que B e β são constantes ($B = 0,2$, $\beta = 12$). A constante ε_{IJ} é calculada como segue:

$$\varepsilon_{IJ} = \frac{181,16 G_I G_J \alpha_I \alpha_J}{\sqrt{\frac{\alpha_I}{N_I} + \sqrt{\frac{\alpha_J}{N_J}}} R_{IJ}^{*6}}$$

em que N_I e N_J são números efetivos de Slater-Kirkwood da valência eletrônica, e G_I e G_J são fatores de escala.

Por fim, o termo EQ corresponde à interação eletrostática entre pares de átomos, definida como:

$$EQ_{ij} = \frac{332,0716 q_i q_j}{D(R_{ij} + \delta)^n}$$

para o qual q_i e q_j são cargas atômicas parciais dos átomos i e j , respectivamente; R_{ij} é a

distância interatômica, $\delta = 0,05 \text{ \AA}$, $n = 1$, e $D = 0,75$ para interações 1-4 (três ligações covalentes entre os átomos i e j) e $D = 1$ para demais casos. Assim como para as interações de van der Waals, as interações eletrostáticas não são computadas para pares de átomos com menos de três ligações covalentes entre si. A carga atômica parcial é calculada pela equação:

$$q_i = q_i^0 + \sum \omega_{KI}$$

em que q_i^0 é a carga inicial do átomo, geralmente igual a zero, mas para alguns átomos apresenta valor diferente, como o oxigênio em grupos carboxilatos para o qual $q_i^0 = -0,5$; ω_{KI} é a carga que o átomo i recebe do átomo j diretamente conectado, de forma que $\omega_{KI} = -\omega_{IK}$. Algumas espécies de átomos, como os oxigênios em grupos PO_4^{-3} , HPO_4^{-2} , H_2PO_4^- e H_3PO_4 , exigem uma equação mais elaborada a fim de se calcular valores mais acurados de carga parcial, a qual é escrita como:

$$q_i = (1 - M_i \vartheta_i) q_i^0 + \sum \vartheta_K q_K^0 + \sum \omega_{KI}$$

em que M_i se refere ao número de átomos conectados ao átomo i , e ϑ_i é um fator que depende do tipo de átomo; para os átomos de oxigênio descritos anteriormente, tem-se $\vartheta_i = 0,5$.

As constantes de força e os valores de referência estão definidas no arquivo MMFF-I_AppendixB.ascii disponível no endereço eletrônico <ftp://ftp.wiley.com/public/journals/jcc/suppmat/17/490/>. Este arquivo, escrito de forma a facilitar sua leitura por computadores, armazena arquivos específicos contendo parâmetros para cada um dos termos da equação de energia potencial do campo de força MMFF94. Como exemplo, é apresentado na figura 53 parte do arquivo MMFFBOND.PAR para as constantes de interação de estiramento de ligação covalente:

*	types		kb	r0	Source
0	1	1	4.258	1.508	C94
0	1	2	4.539	1.482	C94
0	1	3	4.190	1.492	C94
0	1	4	4.707	1.459	X94
0	1	5	4.766	1.093	C94
0	1	6	5.047	1.418	C94
0	1	8	5.084	1.451	C94
0	1	9	4.763	1.458	C94
0	1	10	4.664	1.436	C94

Figura 53 - Arquivo MMFFBOND.PAR do campo de força MMFF94.
Fonte: Do autor.

Na figura 53, a quarta coluna corresponde à constante Kb_{ij} e a quinta coluna é a distância de referência entre o par de átomos conectados. A segunda e a terceira colunas especificam os tipos de átomos. O campo de força MMFF94 define 95 tipos de átomos, os quais são caracterizados pelo elemento químico, hibridização e ambiente no qual está inserido, ou seja, a natureza dos átomos vizinhos e se está ou não inserido em um anel. A tabela 9 mostra alguns tipos de átomos definidos pelo campo de força MMFF94.

Tabela 9 - Alguns tipos de átomos definidos pelo campo de força MMFF94.

Tipo	Descrição
1	Carbono sp^3
2	Carbono sp^2 cuja ligação π é feita com outro carbono
3	Carbono sp^2 cuja ligação π é feita com átomo diferente de carbono
4	Carbono sp
5	Hidrogênio ligado a carbono ou silício
6	Oxigênio divalente
7	Oxigênio em carbonilas
8	Nitrogênio em aminas
9	Nitrogênio em iminas
10	Nitrogênio em amidas
20	Carbono sp^3 em anéis de quatro membros
21	Hidrogênio em hidroxilas
22	Carbono sp^3 em anéis de três membros
37	Carbono aromático

Fonte: Adaptado de Halgren, 1996.

O algoritmo que identifica os tipos de átomos é similar àquele usado para classificar os átomos de acordo com as especificações SYBYL (seção “3. 2. 6 Algoritmos para leitura e edição de moléculas”), embora seja mais complexo devido ao número de tipos de átomos definidos. De modo simplificado, a tipificação dos átomos é feita em três etapas:

- a) Inicia-se a identificação de átomos presentes em anéis, sendo que anéis menores são avaliados primeiro, de forma que se um carbono sp^3 pertence simultaneamente a anéis de três e quatro membros, define-se o átomo como tipo 22;
- b) Átomos não pertencentes a anéis e que sejam diferentes de hidrogênio são classificados em seguida;
- c) No final, tipifica-se os átomos de hidrogênio, de acordo com o tipo de átomo com o qual este esteja conectado. Consulta-se o arquivo MMFFHDEF.PAR presente no arquivo de parâmetros MMFF-I_AppendixB.ascii.

3. 3. 2 Validação do campo de força MMFF94s

A fim de se validar a implementação do campo de força MMFF94, utilizou-se a suíte de validação MMFF94 disponível no endereço eletrônico <http://www.ccl.net/cca/data/MMFF94/>. Esta consiste em um conjunto de 761 moléculas obtidas do *Cambridge Structural Database*, descrito em Groom *et al.* (2016), e minimizadas pelos programas OPTIMOL, desenvolvido por Holloway *et al.* (1995) e BatchMin (MOHAMADI *et al.*, 1990). Dentre as moléculas de validação, 235 consistem de estruturas que possuem um ou mais átomos de nitrogênio trigonal deslocalizado, para o qual os parâmetros torcionais e diedros impróprios são diferentes no campo de força alternativo MMFF94s. Dessa forma, há a suíte de validação para a versão MMFF94s disponível em <http://www.ccl.net/cca/data/MMFF94s/>, consistindo apenas das 235 moléculas citadas anteriormente. Os principais arquivos que se encontram na suíte MMFF94 são os seguintes:

- a) MMFF94_dative.mol2: contém as moléculas de validação no formato mol2 com representação de ligação dativa para compostos tetracoordenados de enxofre e fósforo;
- b) MMFF94_hypervalent.mol2: contém as mesmas moléculas de validação no formato mol2, mas com representação de ligação hipervalente para compostos tetracoordenados de enxofre e fósforo;

- c) MMFF94.mmd: contém as moléculas de validação no formato MacroModel, o qual usa representação de ligação hipervalente para compostos tetracoordenados de enxofre e fósforo;
- d) MMFF94.energies: lista as energias de cada molécula da suíte de validação computadas pelos programas OPTIMOL e BatchMin;
- e) MMFF94_bmin.log: arquivo de log detalhado gerado pelo programa BatchMin que lista a energia total das moléculas e as energias dos termos individuais (ligações covalentes, diedros, van der Waals, etc);
- f) MMFF94_opti.log: arquivo de log detalhado gerado pelo programa OPTIMOL que lista a energia total das moléculas e as energias dos termos individuais (ligações covalentes, diedros, van der Waals, etc), além de outras informações, como carga elétrica dos átomos, tipo de átomos, lista detalhada para cada par de interação.

A suíte de validação MMFF94s possui arquivos equivalentes. O processo de validação consiste basicamente em se calcular a energia das moléculas da suíte usando o programa em teste (no presente caso, o programa MB-Isoster) e comparar as energia obtidas com aquelas reportadas na suíte (arquivo MMFF94.energies). Caso divergências sejam encontradas, é possível verificar os dados detalhados em MMFF94_bmin.log ou MMFF94_opti.log para diagnosticar os problemas. De acordo com informações reportadas na própria suíte de validação, as energias obtidas pelo programa OPTIMOL são mais precisas que aquelas encontradas pelo programa BatchMin, dessa forma, optou-se por validar a implementação do campo de força apenas contra os resultados do OPTIMOL (arquivos MMFF94_opti.log e MMFF94s_opti.log). Como o programa MB-Isoster implementa a versão MMFF94s do campo de força, foram usadas as duas suítes de validação (as 235 moléculas da suíte MMFF94s e as remanescentes 526 moléculas da suíte MMFF94 não encontradas na anterior). Inicialmente fez-se a leitura das moléculas contidas nos arquivos MMFF94s_datave.mol2 e MMFF94_datave.mol2. Em seguida fez-se a leitura dos arquivos MMFF94_opti.log e MMFF94s_opti.log para se obter a energia total e os termos parciais de energia calculadas pelo programa OPTIMOL. A energia das moléculas foram então calculada pelo programa MB-Isoster. Calculou-se a diferença entre estas e as definidas pela suíte de validação. Foram definidos três níveis de tolerância de erro:

- a) tolerância A: a diferença entre a energia computada pelo programa MB-Isoster e a energia reportada na suíte de validação deve ser inferior a 0,0001 kcal/mol;

- b) tolerância B: a diferença entre a energia computada pelo programa MB-Isoster e a energia reportada na suíte de validação deve ser inferior a 0,001 kcal/mol;
- c) tolerância C: a diferença entre a energia computada pelo programa MB-Isoster e a energia reportada na suíte de validação deve ser inferior a 0,01 kcal/mol;

Para cada nível de tolerância, foi calculado o percentual de moléculas cuja energia obtida caiu dentro da tolerância definida. A tolerância tipo A, com diferença de energia de 0,0001 kcal/mol foi assim definida pois esta é a diferença de energia total reportada entre os programas OPTIMOL e BatchMin, logo, qualquer programa implementando o campo de força MMFF94 ou MMFF94s que retorne a energia molecular total dentro deste limite, terá implementado o campo de força corretamente. As tolerâncias tipo B e C foram definidas para se flexibilizar o nível de tolerância e ajudar a qualificar a implementação do campo de força.

3. 3. 3 Implementação da álgebra matricial

Os métodos de minimização de geometria molecular e construção de moléculas bioisostéricas, descritos nas seções “3. 3. 4 Algoritmo de minimização de energia” e “3. 4 CONSTRUÇÃO DE MOLÉCULAS ANÁLOGAS POR BIOISOSTERISMO”, dependem da aplicação de álgebra matricial, desse modo, construiu-se a classe Matrix para implementar as operações envolvendo matrizes.

Uma matriz é um conjunto numérico ordenado organizado em linhas e colunas (ROGERS, 2003). Uma matriz com n linhas e m colunas é definida como matriz $n \times m$. Abaixo é apresentado uma matriz genérica com 2 linhas e 5 colunas:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} \end{bmatrix} \text{ (matriz } 2 \times 5\text{)}$$

Cada elemento particular da matriz é referenciado como a_{ij} , em que i é o índice da linha ocupada pelo elemento, e j é o índice da coluna. Uma matriz cujo número de linhas é igual ao número de colunas é chamada matriz quadrada de $n \times n$.

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \quad (\text{matriz quadrada } 3 \times 3)$$

As operações com matrizes envolvem adição de matrizes, subtração de matrizes, multiplicação de matriz com escalar, multiplicação de matrizes, transposição de matrizes. Na adição e subtração de matrizes é necessário que as matrizes a serem somadas ou subtraídas possuam o mesmo número de linhas n e o mesmo número de colunas m . Na soma $\mathbf{A} + \mathbf{B} = \mathbf{C}$, cada elemento em \mathbf{C} é calculado como:

$$c_{ij} = a_{ij} + b_{ij}$$

Para uma matriz $\mathbf{A} 2 \times 2$ e $\mathbf{B} 2 \times 2$, o resultado será:

$$\mathbf{A} + \mathbf{B} = \mathbf{C}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} + \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} (a_{11} + b_{11}) & (a_{12} + b_{12}) \\ (a_{21} + b_{21}) & (a_{22} + b_{22}) \end{bmatrix}$$

A subtração segue o mesmo princípio, em que os elementos de \mathbf{C} são calculado como a diferença entre os correspondentes elementos de \mathbf{A} e \mathbf{B} :

$$c_{ij} = a_{ij} - b_{ij}$$

Para uma matriz $\mathbf{A} 2 \times 2$ e $\mathbf{B} 2 \times 2$, tem-se:

$$\mathbf{A} - \mathbf{B} = \mathbf{C}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} - \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} (a_{11} - b_{11}) & (a_{12} - b_{12}) \\ (a_{21} - b_{21}) & (a_{22} - b_{22}) \end{bmatrix}$$

Na multiplicação de escalar por matriz, cada elemento da matriz é multiplicado pelo escalar. Se $\mathbf{M} = e\mathbf{A}$, então:

$$m_{ij} = e \times a_{ij}$$

$$e \mathbf{A} = \mathbf{M}$$

$$e \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} ea_{11} & ea_{12} \\ ea_{21} & ea_{22} \end{bmatrix}$$

A multiplicação de matrizes envolve operações mais elaboradas e exige que o número de colunas m da primeira matriz \mathbf{A} seja igual ao número de linhas n da segunda matriz \mathbf{B} . A matriz resultante \mathbf{C} terá o mesmo número de linhas de \mathbf{A} e o mesmo número de colunas de \mathbf{B} . Cada linha de \mathbf{A} é tomada como um vetor e cada coluna de \mathbf{B} também é tomada como um vetor. O elemento c_{ij} será o produto escalar entre o vetor formado pela linha i de \mathbf{A} com o vetor formado pela coluna j de \mathbf{B} , da seguinte forma:

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

$$\mathbf{AB} = \mathbf{C}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} (a_{11}b_{11} + a_{12}b_{21}) & (a_{11}b_{12} + a_{12}b_{22}) \\ (a_{21}b_{11} + a_{22}b_{21}) & (a_{21}b_{12} + a_{22}b_{22}) \end{bmatrix}$$

A matriz transposta, denotada por \mathbf{A}^T , é obtida ao se trocar a posição das linhas pelas colunas, assim, se \mathbf{A} é definida como:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \\ 5 & 6 \end{bmatrix}$$

sua transposta \mathbf{A}^T será:

$$\mathbf{A}^T = \begin{bmatrix} 1 & 3 & 5 \\ 2 & 4 & 6 \end{bmatrix}$$

Uma matriz de transformação \mathbf{T} é aquela aplicada a um vetor \vec{v} a fim de transformá-lo em novo vetor \vec{v}' de forma que:

$$\vec{v}' = \mathbf{T} \vec{v}$$

A transformação será linear, se obedecer as seguintes propriedades (em que \vec{v} e \vec{w} são vetores, λ é um escalar, e $\vec{0}$ é o vetor nulo, cujos elementos são todos zero) (AMMERAAL E ZHANG, 2008):

$$\begin{aligned}\mathbf{T}(\vec{v} + \vec{w}) &= \mathbf{T}(\vec{v}) + \mathbf{T}(\vec{w}) \\ \mathbf{T}(\lambda \vec{v}) &= \lambda \mathbf{T}(\vec{v}) \\ \mathbf{T}(\vec{0}) &= \vec{0}\end{aligned}$$

O vetores \vec{v} que se pretende transformar correspondem às coordenadas cartesianas dos átomos da molécula, de forma que para um átomo a , temos

$$\vec{v} = [a.x \quad a.y \quad a.z \quad 1]$$

A dimensão extra do vetor, cujo quarto elemento é igual a 1, é adicionada para se formar uma coordenada homogênea, para a qual a matriz de transformação também recebe uma linha e uma coluna extras. A coordenada homogênea é usada para que uma transformação não linear (no caso a transformação de translação) possa ser convertida em linear. Duas matrizes de transformação são importantes para o presente trabalho: a matriz de translação e a matriz de rotação. A matriz de translação \mathbf{T} é construída a partir de um vetor deslocamento \vec{d} , tem por objetivo transladar um ponto p somando seu vetor posição \vec{v} a \vec{d} . Apresenta a seguinte forma:

$$\mathbf{T} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ d.x & d.y & d.z & 1 \end{bmatrix}$$

A matriz de rotação \mathbf{R} é construída a partir de um vetor \vec{e} que funciona como eixo de rotação e um ângulo φ que define o ângulo de rotação em radianos. O vetor \vec{e} deve possuir norma igual a 1. Assumindo que:

$$a = e.x;$$

$$b = e.y;$$

$$c = e.z;$$

a matriz \mathbf{R} é escrita como segue:

$$\mathbf{R} = \begin{bmatrix} a^2(1 - \cos \varphi) + \cos \varphi & ab(1 - \cos \varphi) - c \operatorname{sen} \varphi & ac(1 - \cos \varphi) + b \operatorname{sen} \varphi & 0 \\ ab(1 - \cos \varphi) + c \operatorname{sen} \varphi & b^2(1 - \cos \varphi) + \cos \varphi & bc(1 - \cos \varphi) - a \operatorname{sen} \varphi & 0 \\ ac(1 - \cos \varphi) - b \operatorname{sen} \varphi & bc(1 - \cos \varphi) + a \operatorname{sen} \varphi & c^2(1 - \cos \varphi) + \cos \varphi & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

A matriz \mathbf{R} irá rotacionar um ponto p de um ângulo φ sobre o eixo \vec{e} tomando como centro de rotação a origem do sistema de coordenadas cartesianas. Para rotacionar com centro em um ponto qualquer c diferente da origem, procede-se da seguinte forma:

- a) Cria-se uma matriz de translação \mathbf{T}_1 que irá transladar o ponto c até a origem:

$$\mathbf{T}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -c.x & -c.y & -c.z & 1 \end{bmatrix}$$

- b) Cria-se a matriz de rotação \mathbf{R} conforme descrito anteriormente;
 c) Cria-se outra matriz de translação \mathbf{T}_2 para transladar o ponto c da origem até sua posição original;

$$\mathbf{T}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ c.x & c.y & c.z & 1 \end{bmatrix}$$

- d) Multiplica-se as três matrizes na ordem dada, a matriz resultante fará a rotação com centro em c :

$$\mathbf{R}_{\text{res}} = \mathbf{T}_1 \mathbf{R} \mathbf{T}_2$$

3.3.4 Algoritmo de minimização de energia

O algoritmo de minimização de energia é implementado pela classe `Optimizer`, sendo constituído por cinco passos:

- a) Minimização de anéis: uma vez que sistemas de anéis são difíceis de minimizar devido à sua menor flexibilidade, são utilizados *templates* para determinar as coordenadas tridimensionais dos anéis presentes na molécula;
- b) Minimização das ligações covalentes: cada ligação covalente não-cíclica é modificada de forma a corresponder à distância ideal entre seus pares de átomos; a distância ideal é definida pela constante r_{ij}^0 do campo de força adotado;
- c) Minimização dos ângulos: cada ângulo formado por três átomos consecutivos é alterado baseado em coordenadas de *templates*; se três átomos consecutivos pertencem ao mesmo anel, a minimização de ângulo não é realizada;
- d) Minimização dos diedros: para ligações não-cíclicas que formam pelo menos um diedro, o diedro é rotacionado até se obter uma conformação de menor energia;
- e) Refinamento: por fim, é feito um refinamento na estrutura utilizando-se gradientes conjugados não lineares pelos métodos de Newton-Raphson e Fletcher-Reeves.

3.3.4.1 Minimização de anéis

Construiu-se uma biblioteca de *templates* de anéis a partir das moléculas contidas na suíte de validação do campo de força MMFF94. Cada anel possui uma representação em forma textual baseado nos índices numéricos do tipo de átomo (atributo `atomMMFFType`) de seus constituintes; dessa forma, o anel aromático, cujos átomos possuem todos `atomMMFFType = 37`, tem a representação “37-37-37-37-37-37”, enquanto que a piridina, cujos carbonos aromáticos possuem `atomMMFFType = 37`, e cujo nitrogênio aromático possui `atomMMFFType = 38`, tem a representação “37-37-37-37-37-38” (FIGURA 54). A representação é feita de forma que a sequência dos índices sucessivos correspondam a átomos ligados covalentemente no anel e que na medida do possível os índices estejam ordenados de forma crescente.

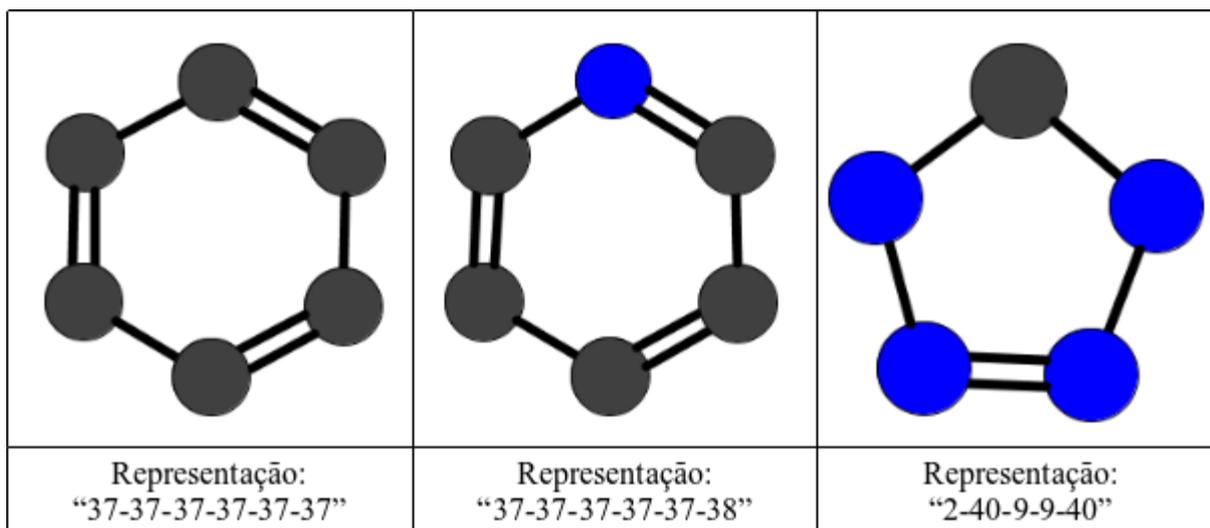


Figura 54 – Alguns *templates* de anéis usados na minimização da geometria molecular. Cada *template* possui uma representação textual baseado no índice de tipo de átomo de seus constituintes. Os índices de tipo de átomo são definidos pelo campo de força MMFF94. Em azul, átomos de nitrogênio, em cinza, átomos de carbono.

Fonte: Do autor.

A classe Ring possui métodos para gerar a representação textual dos anéis e para organizar seus átomos constituintes de acordo com os critérios estabelecidos para a representação textual (átomos em sequência ligados covalentemente, primeiros átomos com menor valor do atributo atomMMFFType). Para iniciar a minimização dos anéis, para cada anel encontrado na molécula busca-se o *template* correspondente, em seguida, o *template* é alinhado com o anel da molécula, por meio das seguintes etapas:

1. Calcula-se o centro geométrico do *template* e do anel da molécula;
2. O *template* é transladado até que seu centro coincida com o centro do anel da molécula;
3. São definidos três átomos de referência para o *template* e para o anel da molécula da seguinte forma:
 - a) define-se $r = (\text{número de átomos do anel}) - 3$;
 - b) define-se $db = 0$ se r é divisível por 3, ou $db = 1$ caso contrário;
 - c) define-se $dc = 1$ se o resto da divisão de r por 3 for 2, ou $dc = 0$ caso contrário;
 - d) átomo de referência 1 é aquele com índice de vetor $i = 0$;
 - e) átomo de referência 2 é aquele com índice de vetor $i = 1 + r/3 + db$;
 - f) átomo de referência 3 é aquele com índice de vetor $i = b + 1 + r/3 + dc$;
4. Rotaciona-se o *template* de forma que o plano formado pelos seus átomos de referência se alinhe com o plano formado pelos átomos de referência do anel da molécula;

5. Rotaciona-se o *template* até que seus átomos de referência fiquem alinhados com os átomos de referência do anel da molécula;
6. Por fim, as coordenadas cartesianas dos átomos do *template* são copiadas para os átomos correspondentes do anel da molécula.

3.3.4.2 Minimização das ligações covalentes

Cada ligação covalente não-cíclica da molécula terá seu comprimento aumentado ou reduzido até que a distância entre seus átomos corresponda à distância ideal para o par de átomos ligados. A distância ideal é definida pela constante r_{ij}^0 do campo de força MMFF94, cada par de átomos possui uma constante específica que leva em conta o índice numérico do tipo de átomo (atomMMFFType) e a ordem da ligação (simples, dupla, tripla, aromática). Um algoritmo recursivo divide a molécula em duas sub-moléculas como se a ligação covalente a ser minimizada fosse “rompida”. Os átomos pertencentes à menor sub-molécula são selecionados para serem movidos e os demais permanecem fixos. O algoritmo é o seguinte:

```

Função atomsToMove(bd : Bond) : Atom[]
Início
  Atom[] submolecule1, submolecule2; {listas de átomos para dividir a
                                         molécula em duas sub-moléculas}
  Atom atom1, atom2; {primeiro átomo a ser adicionado a submolecule1 e
                      submolecule2, respectivamente}

  inteiro n1, n2;
  atom1 ← primeiro átomo contido em bd;
  atom2 ← segundo átomo contido em bd;
  submolecule(atom1, atom2, submolecule1);
  submolecule(atom2, atom1, submolecule2);
  n1 ← número de átomos contidos em submolecule1;
  n2 ← número de átomos contidos em submolecule2;
  Se n1 < n2 Então
    retorne submolecule1;
  Senão
    retorne submolecule2;
  Fim Se;
Fim
Fim Função

Procedimento submolecule(a1 : Atom, a2: Atom, submoleculeList : Atom[])
Início
  Adicionar a1 na lista de átomos submoleculeList;
  Para cada Atom bondedAtom ligado a a1 Faça
    Se bondedAtom é o mesmo que a2 Então

```

```

        passar para próximo Atom;
    Fim Se;
    Se bondedAtom já está na lista submoleculeList Então
        passar para próximo Atom;
    Fim Se;
    submolecule(bondedAtom, a1, submoleculeList)
Fim Para;
Fim
Fim Procedimento

```

Dos dois átomos pertencentes à ligação covalente, um estará na sub-molécula menor e o outro na maior; o primeiro é definido como *tail* e o segundo *head*. Constrói-se um vetor \vec{v} com origem em *tail* e destino em *head*, redimensiona-se o comprimento do vetor subtraindo-se a distância ideal r_{ij}^0 da sua norma original. Por fim, os átomos da sub-molécula menor são transladados no sentido de \vec{v} até uma distância equivalente à norma de \vec{v} (FIGURA 55).

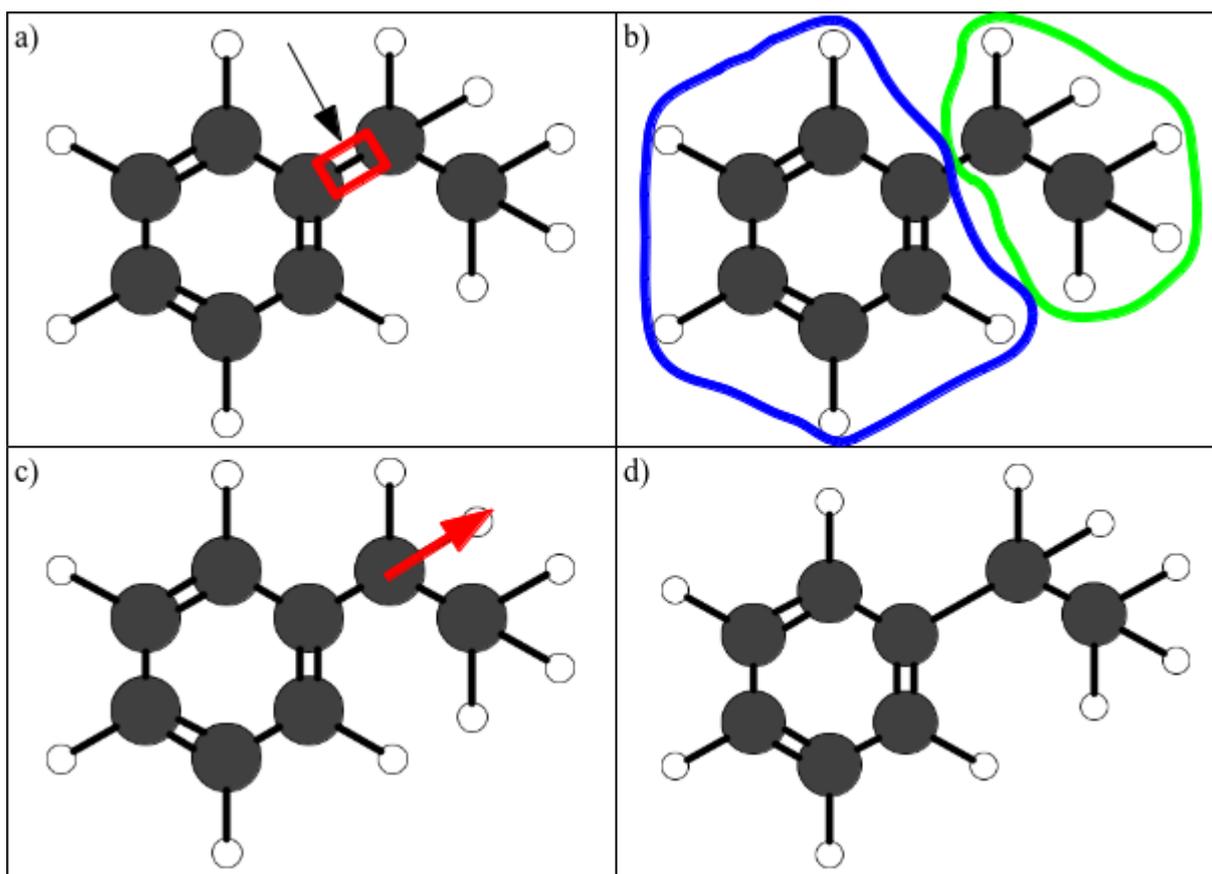


Figura 55 – Minimização de ligações covalentes.

Fonte: Do autor.

Legenda: a) Seleção da ligação covalente a ser minimizada;

b) Divisão da molécula em duas sub-moléculas; os átomos da sub-molécula menor são selecionados para mover;

c) Definição do vetor translação e modificação das coordenadas cartesianas dos átomos da sub-molécula menor;

d) molécula após minimização da ligação covalente.

3. 3. 4. 3 Minimização dos ângulos

De maneira similar à minimização de anéis, a minimização de ângulos se utiliza de *templates*. A biblioteca de *templates* angulares foi construída por meio do programa openbabel (BOYLE *et al.*, 2011). Para cada tipo de átomo definido pelo campo de força MMFF94 que tenha número de coordenação igual ou maior que 2 (átomos com número de coordenação 1, como o hidrogênio e o flúor, ou o oxigênio que realiza uma dupla ligação, não podem ser o átomo central em um ângulo) foi construído um *template* de forma que este seja a molécula mais simples possível capaz de armazenar as informações dos ângulos formados entre o átomo central e os átomos a ele ligados. Dessa forma, para o átomo com MMFFType = 1 (átomo de carbono realizando quatro ligações simples) foi definido como *template* a molécula de metano, para os átomos com MMFFType = 2 (átomo de carbono com uma ligação dupla a outro carbono e demais ligações simples) e MMFFType = 3 (átomo de carbono com uma ligação dupla a um heteroátomo e demais ligações simples) foi definido como *template* a molécula de formaldeído e assim por diante. Na minimização de ângulos, o algoritmo percorre cada átomo da molécula e verifica se este possui número de coordenação igual a 2 ou superior, em caso afirmativo, o átomo é definido como átomo central e os átomos a ele ligados são os átomos laterais. Cada combinação dois a dois dos átomos laterais mais o átomo central define um ângulo a ser minimizado. Procura-se na biblioteca de *templates* angulares um que seja adequado para o átomo central. Os átomos laterais são ordenados da seguinte forma: átomos pertencentes a anéis tem prioridade sobre átomos em cadeias abertas, átomos cuja ligação covalente ao átomo central possua ordem de ligação superior tem maior prioridade (ligação tripla > ligação dupla > ligação simples), átomos de hidrogênio possuem última prioridade. Essa ordenação é feita para que os átomos laterais do átomo central tenham uma correspondência exata aos átomos laterais do *template*, os quais já foram previamente ordenados quando se construiu a biblioteca. O *template* é transladado de forma que as coordenadas cartesianas de seu átomo central coincidam com as do átomo central da molécula. Os dois primeiros átomos laterais do átomo central são definidos como *b1* e *b2*, respectivamente, enquanto que os dois primeiros átomos laterais do *template* são definidos como *r1* e *r2*. O *template* é rotacionado de forma que *b1* se alinhe com *r1* e *b2* se alinhe com *r2* o mais próximo possível. Por fim, para cada átomo lateral do átomo central, a submolécula formada pelo átomo lateral e demais ligados a este, é rotacionada em relação ao átomo central até que haja o alinhamento entre o átomo lateral da molécula e o seu

correspondente átomo lateral do *template*. O algoritmo é apresentado no quadro abaixo:

Para cada Átomo central da Molécula

```

inteiro : coord;
Atom : t, r1, r2, c, b1, b2;
Vetor : v1, v2, w1, w2, nv1v2, nw1w2, gv1, gv2, gv;
real : theta, theta1, theta2, theta3, angle;
Molecule : template, toMove;

coord ← número de coordenação de central;
Se coord < 2; Então
    passar para próximo Átomo;
Fim Se;

template ← template específico para central;
lateral[] ← lista de átomos conectados a central;
Ordenar lateral[];

t ← átomo central do template;
r1 ← primeiro átomo ligado a t;
r2 ← segundo átomo ligado a t;
c ← átomo central da molécula;
b1 ← primeiro átomo ligado a c;
b2 ← segundo átomo ligado a c;
translate ← Vetor(t, c);
Transladar template na direção e módulo de translate;

v1 ← Vetor(t, r1);
v2 ← Vetor(t, r2);
w1 ← Vetor(c, b1);
w2 ← Vetor(c, b2);
nv1v2 ← v1 x v2; {produto vetorial entre v1 e v2}
nw1w2 ← w1 x w2; {produto vetorial entre w1 e w2}
theta ← ângulo entre nv1v2 e nw1w2;
gv1 ← nv1v2 x nw1w2; {produto vetorial entre nv1v2 e nw1w2}
Girar template de um ângulo theta em torno de gv1 com centro em t;

v1 ← Vetor(t, r1); {v1 é recalculado pois template foi girado}
v2 ← Vetor(t, r2); {v2 é recalculado pois template foi girado}
theta1 ← ângulo entre v1 e w1;
theta2 ← ângulo entre v2 e w2;
theta3 ← (theta1 + theta2) / 2;
gv2 ← v1 x w1; {produto vetorial entre v1 e w1}
Girar template de um ângulo theta3 em torno de gv2 com centro em t;

Para i de 1 até coord Faça
    bn ← átomo ligado a c, com índice i;
    rn ← átomo ligado a t, com índice i;

    bond ← ligação covalente entre bn e c;
    toMove ← menor sub-molécula obtida de Molecule pela ruptura de
    bond;

    v1 ← Vetor(t, rn);
    v2 ← Vetor(c, bn);

```

```

    angle ← ângulo entre v1 e v2;
    gv ← v1 x v2; {produto vetorial entre v1 e v2}
    Girar toMove de um ângulo angle em torno de gv centrado em c;
  Fim Para;
Fim Para;

```

3. 3. 4. 4 Minimização dos diedros

A minimização dos ângulos diedros é feita para ligações covalentes não cíclicas para as quais seja definido pelo menos um diedro, isto é, ambos os átomos envolvidos na ligação covalente devem possuir número de coordenação igual ou superior a 2. A molécula é dividida em duas sub-moléculas conforme descrito na seção “3. 3. 4. 2 Minimização de ligações covalentes”, sendo os átomos da sub-molécula menor selecionados para movimentação. Para cada ligação covalente cujo diedro será minimizado, constrói-se um vetor \vec{v} com origem em *tail* e destino em *head* (*tail* e *head* tem a mesma definição descrita em “3. 3. 4. 2 Minimização de ligações covalentes”). É feita então uma pré-minimização, a fim de se obter um ângulo diedro aproximado:

- a) Calcula-se a energia atual da molécula, que é salva em uma variável do tipo real denominada *energy*;
- b) Define-se uma variável do tipo real com o nome *angleI*, inicialmente com valor zero;
- c) O diedro é girado de 36° em torno de \vec{v} , calcula-se a energia da nova conformação;
- d) Se a energia for menor que a inicialmente gravada em *energy*, a variável *energy* recebe o valor da nova energia e o ângulo com menor energia é gravado em *angleI*;
- e) O diedro é girado em mais 36° (somando um ângulo de 72°) e os passos anteriores são repetidos, até o ângulo final de 360° ;
- f) Após a definição do ângulo diedro aproximado que confere à molécula uma menor energia, o diedro é girado neste ângulo específico.

Em seguida, aplica-se o método de gradientes conjugados não lineares (descritos na seção “3. 3. 4. 5 Refinamento”) a fim de otimizar o ângulo diedro a partir do ângulo aproximado encontrado anteriormente (FIGURA 56). O algoritmo descrito é a combinação de dois algoritmos; poder-se-ia não utilizar gradientes conjugados e fazer o giro do ângulo diedro

com passo de 1° ou menor, entretanto, haveria um custo de tempo computacional maior e dependendo do passo adotado, o resultado ainda seria aproximado; poder-se-ia utilizar diretamente gradientes conjugados, que é um método rápido, sem realizar uma pré-minimização, entretanto, o resultado de um cálculo por gradientes conjugados é dependente da estrutura de partida, e assim corre-se o risco de se ter um resultado ruim pelo fato da estrutura de partida estar em uma conformação ruim. A combinação dos dois algoritmos faz um balanço entre velocidade e exatidão.

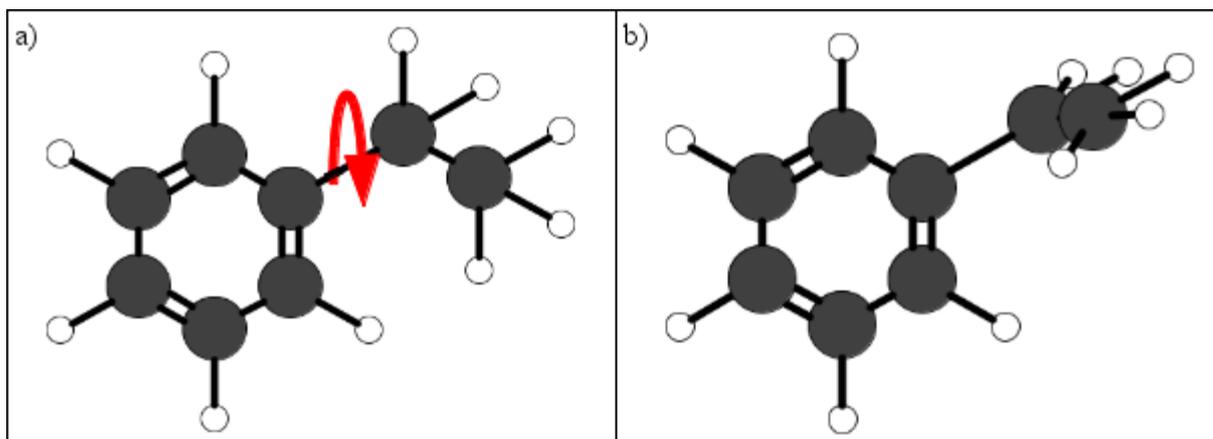


Figura 56 – Minimização dos ângulos diedro.

Fonte: Do autor.

Legenda: a) ângulo diedro a ser minimizado;

b) molécula após minimização do diedro especificado.

3. 3. 4. 5 Refinamento

Gradientes conjugados constituem um método numérico iterativo utilizado para se achar o mínimo de uma função quadrática que possa ser escrita da forma (SHEWCHUK, 1994):

$$f(x) = \frac{1}{2} x^T A x - b^T x + c$$

em que x é um vetor contendo as variáveis independentes da função, A é uma matriz simétrica positiva-definida, b é um vetor e c é um escalar. Para se minimizar $f(x)$ deve-se encontrar a solução de $Ax = b$. Entretanto, a forma funcional da equação de energia potencial do campo de força MMFF94 não pode ser escrita na forma quadrática apresentada acima. Neste caso, deve-

se utilizar a forma alternativa do gradiente conjugado não linear, o qual é aplicável a qualquer função contínua para a qual se possa calcular o gradiente. Partindo-se de um vetor inicial x_0 (o qual é obtido pelos algoritmos anteriores: minimização de anéis, ligações covalentes, ângulos e diedros), calcula-se o vetor d_0 que aponta para o mínimo da função, este vetor é escalonado por um valor a e em seguida adicionado a x_0 , gerando o novo vetor x_1 , o processo é repetido até se encontrar um valor de x que minimize a função dentro de um limite de tolerância.

Para o caso específico da minimização de geometria molecular, o vetor x é formado pelas coordenadas cartesianas dos átomos da molécula, tendo $3 \times n$ elementos, em que n é o número de átomos da molécula. Assim, o primeiro elemento do vetor x é a coordenada X do primeiro átomo, o segundo elemento é a coordenada Y do primeiro átomo, o terceiro elemento é a coordenada Z do primeiro átomo, o quarto elemento é a coordenada X do segundo átomo e assim por diante.

O cálculo do vetor d e do escalar a dependem do gradiente e da hessiana da função. O vetor gradiente ou $f'(x)$, que é definido como:

$$f'(x) = \begin{bmatrix} \frac{\partial}{\partial x_1} f(x) \\ \frac{\partial}{\partial x_2} f(x) \\ \vdots \\ \frac{\partial}{\partial x_n} f(x) \end{bmatrix}$$

Cada elemento do vetor gradiente, o qual tem o mesmo número de elementos que x , é obtido pela derivada parcial de primeira ordem do correspondente elemento em x . Uma vez que a derivação analítica da função do campo de força é algo complexo, resolveu-se aproximar a derivada por métodos numéricos. Inicialmente calcula-se a energia molecular, armazenando o resultado em uma variável $g(x)$; em seguida, a variável em relação a qual se deseja derivar é incrementada com um pequeno valor h (adotou-se $h = 0,00001$) e a energia é calculada novamente, com o resultado armazenado em $g(x + h)$, a derivada parcial é então calculada pela equação:

$$g'(x) = \frac{g(x + h) - g(x)}{h}$$

Após o cálculo da derivada, a variável para a qual se derivou é decrementada do valor h para restabelecer seu valor original e se realizar a derivação das próximas variáveis.

A hessiana substitui a matriz A vista nas equações quadráticas, sendo seus elementos formados pela segunda derivada parcial, como segue:

$$f''(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \dots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix}$$

Da mesma forma que para o gradiente, as derivadas parciais de segunda ordem são calculadas por métodos numéricos. Se a derivação é feita sobre uma única variável, como em

$\frac{\partial^2 f}{\partial x_1^2}$, utiliza-se a seguinte aproximação:

$$g''(x) = \frac{-(2g(x) - g(x+h) - g(x-h))}{h^2}$$

Se a derivação é feita sobre duas variáveis, como em $\frac{\partial^2 f}{\partial x_1 \partial x_2}$, a derivada é computada como:

$$g''(x) = \frac{g(x_1+h, x_2+h) - g(x_1+h, x_2-h) - g(x_1-h, x_2+h) + g(x_1-h, x_2-h)}{4h^2}$$

O algoritmo de gradientes conjugados não linear é então implementado conforme está definido no artigo de Shewchuk, 1994, página 52, adotando-se $i_{\max} = 100$, $j_{\max} = 10$, e $\varepsilon = 0,0001$. O algoritmo é apresentada na figura 57:

```

 $i \leftarrow 0$ 
 $k \leftarrow 0$ 
 $r \leftarrow -f'(x)$ 
 $d \leftarrow r$ 
 $\delta_{new} \leftarrow r^T r$ 
 $\delta_0 \leftarrow \delta_{new}$ 
While  $i < i_{max}$  and  $\delta_{new} > \epsilon^2 \delta_0$  do
   $j \leftarrow 0$ 
   $\delta_d \leftarrow d^T d$ 
  Do
     $\alpha \leftarrow -\frac{[f'(x)]^T d}{d^T f''(x) d}$ 
     $x \leftarrow x + \alpha d$ 
     $j \leftarrow j + 1$ 
  while  $j < j_{max}$  and  $\alpha^2 \delta_d > \epsilon^2$ 
   $r \leftarrow -f'(x)$ 
   $\delta_{old} \leftarrow \delta_{new}$ 
   $\delta_{new} \leftarrow r^T r$ 
   $\beta \leftarrow \frac{\delta_{new}}{\delta_{old}}$ 
   $d \leftarrow r + \beta d$ 
   $k \leftarrow k + 1$ 
  If  $k = n$  or  $r^T d \leq 0$ 
     $d \leftarrow r$ 
     $k \leftarrow 0$ 
   $i \leftarrow i + 1$ 

```

Figura 57 – Algoritmo de gradientes conjugados não lineares.

Fonte: Shewchuk (1994, p. 52).

3. 4 CONSTRUÇÃO DE MOLÉCULAS ANÁLOGAS POR BIOISOSTERISMO

Para iniciar a construção dos bioisósteros, a molécula de entrada definida pelo usuário é dividida em duas partes pela quebra de ligações covalentes específicas, gerando dois fragmentos. O fragmento maior é definido como *fragmento root* e é comum a todos os bioisósteros que serão construídos. O fragmento menor é denominado *fragmento*

bioisotérico e será substituído nos bioisósteros formados (FIGURA 58). Uma vez formado, o fragmento adquire um radical (representado como uma estrela na figura 58) conectado à ligação covalente que foi rompida. O radical corresponde a um pseudoátomo cujas coordenadas são as mesmas do átomo anteriormente conectado e que pertence ao outro fragmento. O átomo do fragmento diretamente conectado ao radical é denominado “átomo âncora”. A presença do radical é necessária para o processo de montagem dos bioisósteros a partir do *fragmento root* e de fragmentos alternativos relacionados ao *fragmento bioisotérico*.

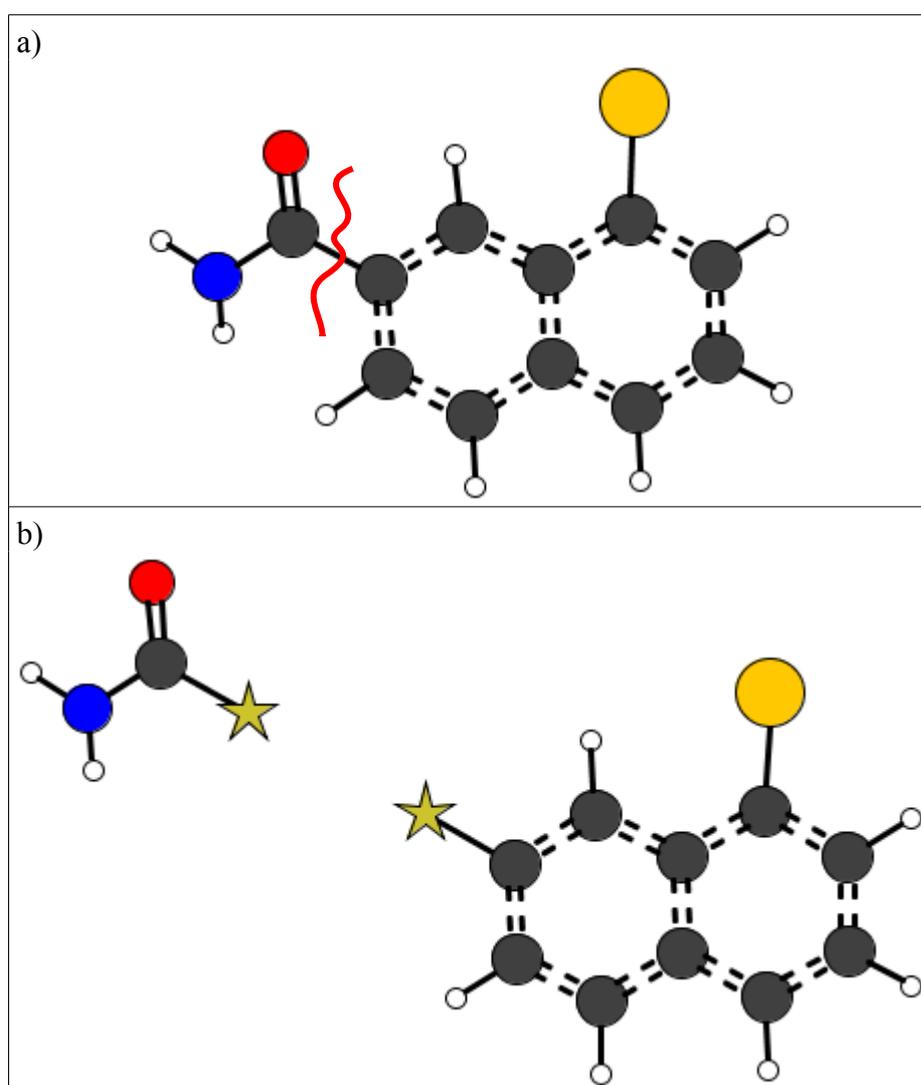


Figura 58 – Fragmentação molecular. Após a quebra da ligação covalente, geram-se dois fragmentos, o maior denominado *fragmento root*, e o menor denominado *fragmento bioisotérico*. Em cada fragmento aparece um radical, usado para auxiliar na construção de novas moléculas por meio da junção do *fragmento root* com fragmentos alternativos. Em cinza: átomos de carbono; em azul: átomos de nitrogênio, em vermelho: átomos de oxigênio; em laranja: átomos de cloro; em branco: átomos de hidrogênio. As estrelas representam radicais, locais em que a ligação covalente foi rompida.

Fonte: Do autor.

Um fragmento é representado pela classe *Fragment*, a qual deriva da classe *Molecule*, possuindo em adição uma lista de radicais presentes no fragmento. Por sua vez, um radical é representado pela classe *Radical*, cujos atributos são apresentados na tabela 10. Um fragmento pode ter até três radicais.

Tabela 10 – Atributos da classe *Radical*.

Atributo	Tipo	Descrição
radical	Atom	Pseudoátomo formado na ligação covalente quebrada na molécula original
linker	Atom	Átomo do fragmento ligado ao pseudoátomo
bondOrder	string	Ordem de ligação entre radical e linker
index	inteiro	Índice do radical, o qual pode ser 1, 2, ou 3

Fonte: Do autor.

Para se criar um novo fragmento a partir de uma molécula, informa-se qual é a molécula original juntamente com uma lista dos átomos da molécula que deverão ser removidos no fragmento, conforme é visto no algoritmo a seguir:

Início

Função criarFragmento(mol : *Molecule*, radicals : lista de *Atom*)

```

    fragAtom, at1, at2, cAt1, cAt2 : Atom;
    linker : Radical;
    bond : Bond;
    Para cada Atom atom presente em mol, Faça:
        Se atom não está em radicals, Então:
            fragAtom ← cópia de atom;
            adicionar fragAtom à lista de átomos do Fragmento;
        Fim Se;
    Fim Para;

    Para cada Bond lig presente em mol, Faça:
        at1 ← primeiro átomo de lig;
        at2 ← segundo átomo de lig;
        Se at1 está em radicals E at2 está em radicals, Então:
            próximo;
        Senão Se at1 está em radicals, Então:
            linker ← novo Radical(radical ← at1, ancora ← at2);
            adicionar linker à lista de radicais do Fragmento;
        Senão Se at2 está em radicals, Então:
            linker ← novo Radical(radical ← at2, ancora ← at1);
            adicionar linker à lista de radicais do Fragmento;
        Senão:
            cAt1 ← átomo do Fragmento equivalente a at1;
            cAt2 ← átomo do Fragmento equivalente a at2;
            bond ← novo Bond(atom1 ← cAt1, atom2 ← cAt2);

```

```

                                adicionar bond à lista de ligações de Fragmento;
                                Fim Se;
                                Fim Para;
                                Fim Função;
                                Fim

```

Foi construída uma biblioteca de fragmentos em formato sdf, tendo como base dados da literatura. Os pseudoátomos são nomeados como R1, R2, e R3, dependendo da quantidade de radicais presentes no fragmento. Cada fragmento, denominado *fragmento de biblioteca*, possui duas propriedades: a família à qual o fragmento pertence e a sua representação em forma de string. A família corresponde a grupos de fragmentos que podem ser substituídos entre si na estrutura de um composto protótipo. Como exemplo, os fragmentos listados na figura 59 podem substituir um grupamento amida, todos então são classificados na família “AmideLike”.

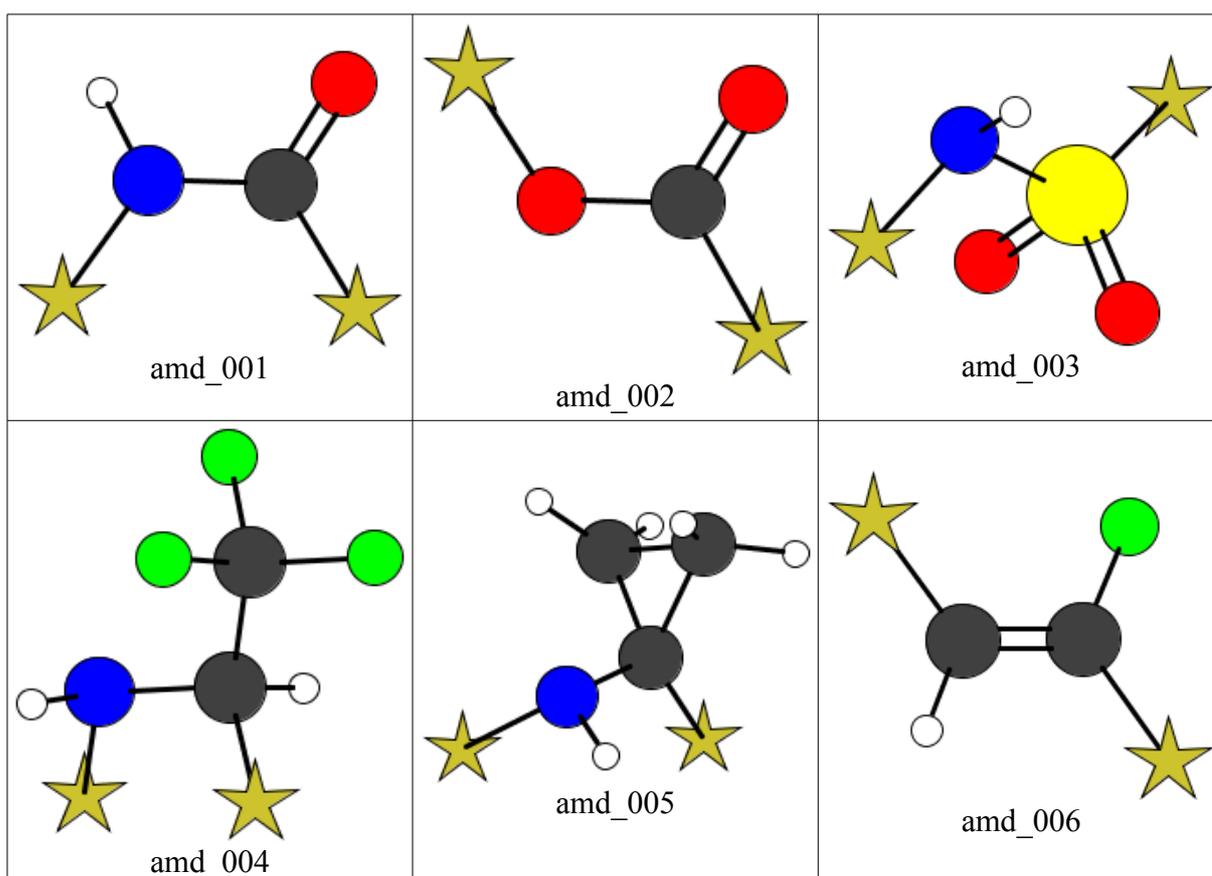
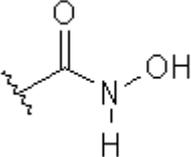
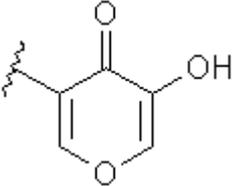
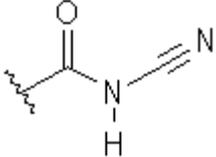
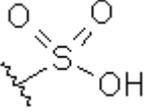


Figura 59 – Fragmentos bioisómeros do grupo amida presentes na biblioteca de fragmentos. Abaixo de cada fragmento é colocado um código que representa o nome do fragmento dentro da biblioteca. Em cinza: átomos de carbono; em azul: átomos de nitrogênio, em vermelho: átomos de oxigênio; em amarelo: átomos de enxofre; em verde: átomos de flúor; em branco: átomos de hidrogênio. As estrelas representam radicais, locais em que a ligação covalente foi rompida.

Fonte: Do autor.

A representação em string é uma forma conveniente de representar os fragmentos, para fácil consulta à biblioteca. É baseada no formato smile, mas sem apresentar explicitamente as ramificações e tendo em adição os pseudoátomos. A tabela 11 apresenta a representação em string para alguns fragmentos.

Tabela 11 – fragmentos bioisotéricos e sua representação linear

Fragmento	Representação
 ácido_hidroxâmico	R1-C-N-O=O
 hidróxi-cromona	R1-C=C-O-C=C-C=O-O
 acil-cianamida	R1-C-N-C#N=O
 ácido_sulfônico	R1-S-O=O=O

Fonte: Do autor.

Para selecionar *fragmentos de biblioteca* que possam substituir o *fragmento bioisotérico*, gera-se uma representação em string para este, em seguida, consulta-se um catálogo contendo uma relação entre os fragmentos (representados em string) e a família a qual pertencem. Uma vez encontrada a família do *fragmento bioisotérico*, todos os *fragmentos de biblioteca* que pertençam a esta família são selecionados para a construção dos novos bioisósteros.

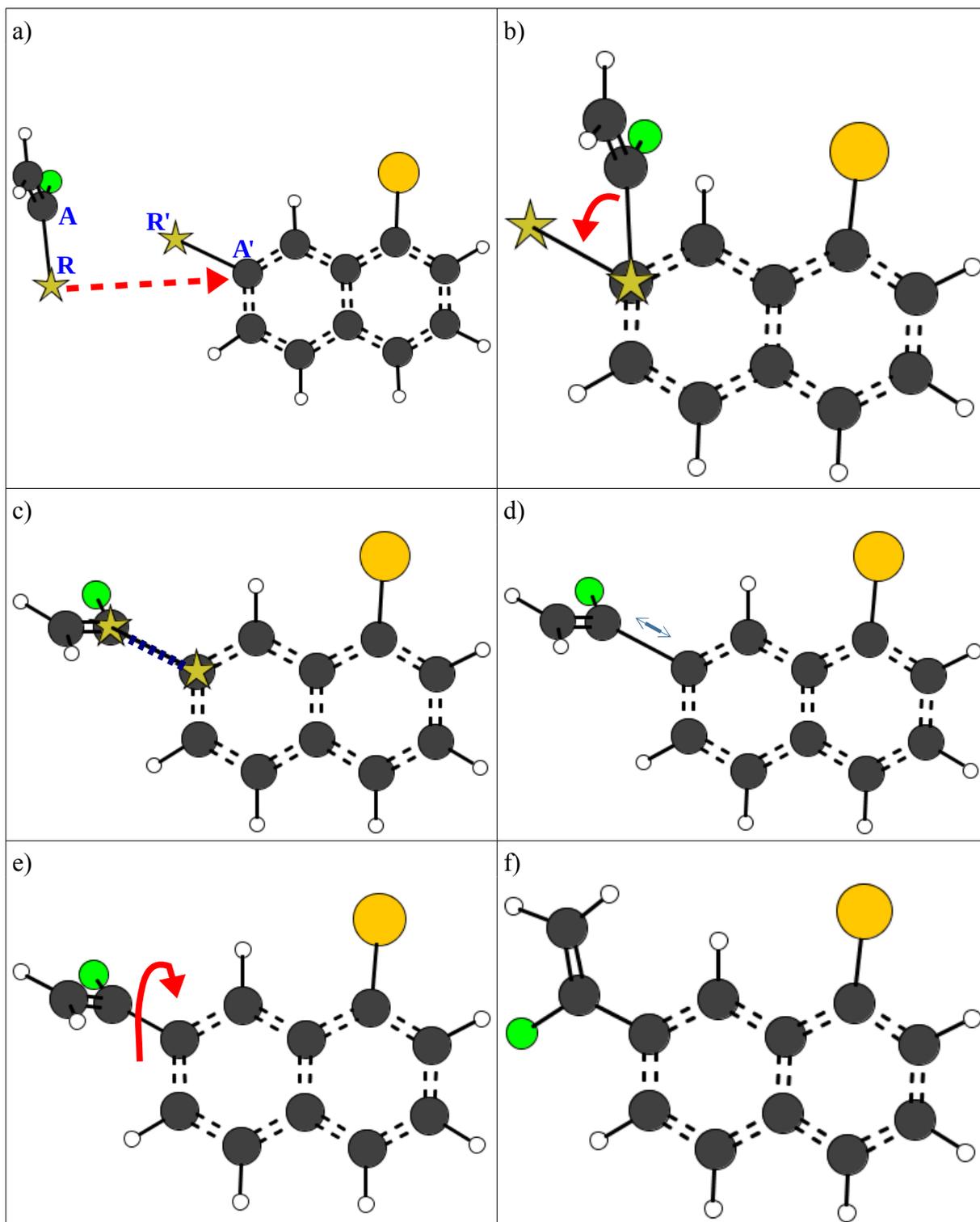


Figura 60 – Construção dos bioisómeros. Átomos **R** e **A** são o radical e o átomo âncora do *fragmento de biblioteca*. O *fragmento root* possui os átomos correspondentes **R'** e **A'**.

Fonte: Do autor.

- Legenda: a) Translação do fragmento de biblioteca até que coordenadas de **R** se igualem às coordenadas de **A'**;
 b) Rotação do fragmento de biblioteca em **R**, até que a ligação **R-A** se alinhe com **R'-A'**;
 c) Remoção de **R** e **R'**, formação da ligação **A-A'**;
 d) Minimização do comprimento de ligação **A-A'**;
 e) Minimização do diedro da ligação **A-A'**
 f) Bioisómero final.

A construção do bioisómero é feita nas seguintes etapas (FIGURA 60):

- a) O *fragmento de biblioteca* sofre uma translação até que seu pseudoátomo **R1** se sobreponha ao átomo âncora **A'** do *fragmento root*;
- b) O *fragmento de biblioteca* sofre uma rotação ao redor de R1 até que a ligação **R1-A** se alinhe com a ligação **R1'-A'**;
- c) Os átomos **R1** e **R1'** são apagados;
- d) Cria-se uma ligação covalente entre **A** e **A'**, unindo os dois fragmentos. A ordem de ligação é a mesma definida no atributo bondOrder do fragmento de biblioteca;
- e) A distância entre **A** e **A'** é minimizada conforme descrito na seção “3. 3. 4. 2 Minimização das ligações covalentes”;
- f) Caso a ligação **A-A'** defina pelo menos um ângulo diedro, esta ligação será rotacionada e minimizada conforme descrito na seção “3. 3. 4. 4 Minimização dos diedros”.

3. 4. 1 Construção da biblioteca de fragmentos

Para gerar as coordenadas cartesianas dos átomos dos fragmentos presentes na biblioteca, utilizou-se o programa openbabel (BOYLE *et al.*, 2011). Uma vez que o programa openbabel não reconhece pseudoátomos (R1, R2 e R3), estes foram inicialmente representados por átomos pesados não presentes em moléculas orgânicas, de acordo com a ordem de ligação do pseudoátomo (ligação covalente simples: átomo astato [At]; ligação dupla: átomo polônio [Po]; ligação tripla: átomo bismuto [Bi]). A linha de comando usada com o programa openbabel para converter uma estrutura em formato smile para sdf é a seguinte:

```
obabel -:"[At]C" -O temp.sdf --gen3D
```

em que a string entre aspas duplas é o fragmento em formato smile, a instrução -O temp.sdf gera um arquivo temporário em formato sdf e a instrução --gen3D instrui o programa openbabel a construir as coordenadas cartesianas em 3 dimensões. Para automatizar a construção da biblioteca, definiu-se o arquivo de entrada mostrado na figura 61.

amd_001	[At]C(=O)N[At]	amideLike
amd_002	[At]C(=O)O[At]	amideLike
amd_003	[At]S(=O)(=O)N[At]	amideLike
Amd_004	[At]C(C(F)(F)(F))N[At]	amideLike
amd_005	[At]C1(N[At])CC1	amideLike
amd_006	[At]C(F)=C[At]	amideLike

Figura 61 – Arquivo de configuração para construção da biblioteca de fragmentos.

Fonte: Do autor.

Cada linha do arquivo corresponde a um fragmento, a primeira coluna é o nome atribuído ao fragmento, a segunda coluna é a representação smile do fragmento e a terceira coluna é a família à qual o fragmento pertence. As etapas para a construção da biblioteca estão apresentadas no algoritmo a seguir:

Início

Função criarBiblioteca(input : Arquivo)

```

name, smile, family : string;
mol : Molecule;
Para cada linha de input, Faça:
  name ← primeiro token de linha;
  smile ← segundo token de linha;
  family ← terceiro token de linha;
  chamar programa obabel para criar molécula "tmp.mol2" a partir
    de smile;
  mol ← molécula lida do arquivo "tmp.mol2";
  apagar arquivo "tmp.mol2";
  mol.name ← name;
  adicionar propriedade a mol ("BioisosterFamily" => family);
  Para cada Atom at de mol, Faça:
    Se at.element = "At" OU at.elemet = "Po"
      OU at.element = "Bi", Então:
        transformar at em radical;
    Fim Se;
  Fim Para;
  adicionar mol à lista dos fragmentos de biblioteca;
Fim Para;

Fim Função;
Fim

```

Observação: no algoritmo acima, a palavra *token* corresponde a dados lidos de uma entrada genérica, cujos valores estão separados por espaços em branco ou outro delimitador que se especifique.

3. 4. 2 Construção de novas moléculas por substituição de átomos de hidrogênio

Em adição ao modo padrão de construção dos bioisómeros, em que o usuário seleciona o fragmento a ser substituído, criou-se um método alternativo e automático que produz transformações em toda a estrutura da molécula sem necessidade de se selecionar uma região em particular. Para cada átomo de hidrogênio constrói-se uma coleção de moléculas análogas em que o hidrogênio é substituído por um fragmento de uma biblioteca especial. Esta biblioteca consiste de 35 substituintes comumente encontrados em compostos orgânicos (metil, etil, fenil, benzil, hidroxil, carbonil, etc). Se um átomo de hidrogênio for equivalente a outro já utilizado como ponto de substituição (por exemplo, os três hidrogênios do grupamento metila), o mesmo não é considerado para a construção de novas moléculas. Com isso, evita-se a construção de moléculas idênticas. O algoritmo que seleciona os átomos de hidrogênio usados como ponto de substituição é apresentado a seguir:

Início

Função selecionarHidrogenios(mRoot : Molecule) : lista de inteiros:

hydrogenIndexes: lista de inteiros;

fragRepList: lista de string;

n: inteiro;

n ← número de átomos em mRoot;

at: Atom;

fragRep : string;

Para i = 0 até n, Faça:

at ← átomo de mRoot com índice i;

Se at.element = "H", Então:

fragRep ← representação em string de mRoot assumindo at
como radical;

Se fragRepList não contém uma cópia de fragRep, Então:

adicionar fragRep à fragRepList;

adicionar i à hydrogenIndexes;

Fim Se;

Fim Se;

Fim Para;

retorna hydrogenIndexes;

Fim função

Fim;

Para cada átomo de hidrogênio da molécula, constrói-se uma representação em string (TABELA 11) assumindo que o átomo de hidrogênio corresponde ao radical. Se para dois

átomos de hidrogênio a representação em string é idêntica, os átomos são considerados equivalentes e apenas o primeiro é usado como ponto de substituição para a construção das novas moléculas, cuja implementação é a seguinte:

```
Início
Função makeOverH(mRoot: Molecule):
    hIndexes: lista de inteiros;
    hIndexes ← selecionarHidrogenios(mRoot);

    molH : Molecule;
    rad : Atom;
    rootFrag, localRootFrag, fragToAdd : Fragment;

    Para cada inteiro i em hIndexes, Faça:
        molH ← cópia de mRoot;
        rad ← átomo de molH com índice i;
        rootFrag ← novo Fragment(mol => molH, radicals => (rad));
        Para cada Fragment frag contido na biblioteca, Faça:
            localRootFrag ← cópia de rootFrag;
            fragToAdd ← cópia de frag;
            juntar os fragmentos localRootFrag com fragToAdd;
            adicionar a nova molécula na lista de bioisósteros;
        Fim Para;
    Fim Para;

Fim Função;
Fim
```

3. 5 ANÁLISE DE INTERAÇÕES RECEPTOR-LIGANTE

Além das formas convencionais de se obter uma molécula de entrada lida de um arquivo com formato suportado (mol, sdf, mol2, pdb), é possível ler uma molécula em arquivo pdb que esteja complexada a uma biomacromolécula, geralmente uma proteína. Neste caso, é possível computar a energia de interação (van der Waals e eletrostática) entre cada átomo da molécula e o sítio de ligação da biomacromolécula (ou receptor) onde aquela esteja inserida. Átomos da molécula com interações menos favoráveis ao receptor (energia potencial calculada mais elevada) são indicados como potenciais pontos de substituição bioisostérica. Inicialmente, identifica-se o sítio ativo do receptor da seguinte forma:

- a) Calcula-se o centro geométrico da molécula (seção “3. 1. 5 Classe Molecule”);

- b) Calcula-se o raio da molécula, o qual corresponde à maior distância entre um átomo da molécula e o centro geométrico (seção “3. 1. 5 Classe Molecule”);
- c) Ao raio da molécula acrescenta-se o valor de 4 Å, gerando-se o “raio de sítio ativo”. O valor de 4 Å foi escolhido pois esta é a maior distância observada para uma interação de hidrogênio fraca (JEFFREY, 1997);
- d) Todos os átomos do receptor que estejam a uma distância do centro geométrico da molécula que seja menor ou igual ao “raio de sítio ativo” são incluídos no sítio ativo (FIGURA 62);
- e) Para cada átomo da molécula, computa-se o somatório das energias de interação não-ligante (van der Waals e eletrostática) entre este e os átomos do sítio ativo;
- f) No final, é exibida uma tabela em ordem decrescente das energias potenciais dos átomos da molécula para que o usuário faça sua seleção (seção “4. 1. 2 Leitura de moléculas a partir de complexos receptor-ligante”).

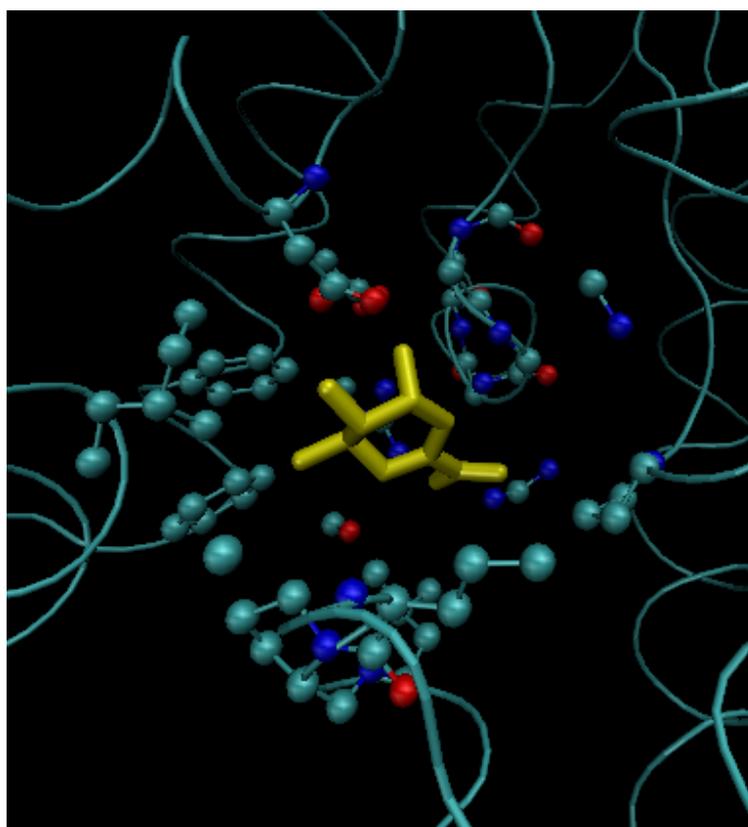


Figura 62 – Definição do sítio ativo para cálculo da interação entre átomos da molécula e do receptor. Em amarelo está a molécula cocrystalizada ao receptor; renderizados como esferas estão os átomos do receptor que formam o sítio ativo. Imagem obtida com a estrutura PDB 1WE2 (PEREIRA *et al.*, 2004).

Fonte: Do autor.

3. 6 CÁLCULO DE PROPRIEDADES *DRUG-LIKE*

Compostos *drug-like* são aqueles que além de possuir boa interação com seu alvo molecular, também apresentam propriedades físico-químicas que lhes garantam características ADME (absorção, distribuição, metabolismo, excreção) adequadas e toxicidade reduzida, garantindo sua aprovação ao menos nos testes de Fase Clínica I (LIPINSKI, 2000). Durante a etapa de descoberta de compostos candidatos a fármacos, é necessário promover modificações estruturais a partir do composto protótipo a fim de otimizar suas características *drug-like*; caso contrário, se apenas o aumento da potência for o foco das pesquisas, corre-se o risco de se desenvolver bons ligantes, mas que irão falhar nos testes clínicos subsequentes devido a propriedades ADME inadequadas, como baixa solubilidade, baixa permeabilidade e instabilidade química (KERNS E DI, 2008).

Propriedades estruturais da molécula são responsáveis pelas suas propriedades físico-químicas e bioquímicas, as quais em conjunto interagem causando as propriedades farmacocinéticas e toxicidade. Dessa forma, modificando-se a estrutura da molécula é possível controlar suas demais propriedades. As propriedades estruturais de interesse são as seguintes: interações de hidrogênio, lipofilicidade, peso molecular, pK_a , PSA (*Polar Surface Area*, Área de Superfície Polar), formato e reatividade (KERNS E DI, 2008).

Uma vez calculadas as propriedades estruturais, deve-se definir um intervalo de valores aceitáveis para se caracterizar uma molécula como tendo boas propriedades *drug-like*. Regras empíricas foram desenvolvidas com esse propósito, dentre as quais pode-se citar as regras de Lipinski *et al.* (1997) e as regras de Veber (VEBER *et al.*, 2002). De acordo com as regras de Lipinski, a absorção e a permeabilidade é favorecida para moléculas que possuam:

- a) menos que 5 átomos doadores de interações de hidrogênio;
- b) menos que 10 átomos receptores de interação de hidrogênio;
- c) peso molecular menor que 500 Da;
- d) logP menor que 5;

As regras de Veber definem que uma boa biodisponibilidade oral é observada para moléculas em que:

- a) número de ligações rotativas ≤ 0 ;
- b) $PSA \leq 140 \text{ \AA}^2$;

Foram implementados algoritmos para calcular propriedades *drug-like* dos bioisómeros, a fim de se filtrar aqueles com características que potencializem suas

propriedades farmacocinéticas. As seguintes propriedades são calculadas: peso molecular, número de doadores de interação de hidrogênio, número de aceptores de interação de hidrogênio, logP, logS e TPSA.

O peso molecular é calculado como o somatório dos pesos atômicos dos átomos constituintes da molécula. Os pesos atômicos são consultados em uma tabela. O algoritmo é o seguinte:

```
Início
Função compPesoMolecular(mol : Molecule)
    pesoMolecular, pesoAtomico : real;
    pesoMolecular ← 0;

    Para cada Átomo atom de mol, Faça:
        pesoAtomico ← peso atômico de atom consultado na tabela;
        pesoMolecular ← pesoMolecular + pesoAtomico;
    Fim Para;

    adicionar propriedade a mol ("molecular_weight" => pesoMolecular);

Fim Função;
Fim
```

O número de doadores de interação de hidrogênio é computado contando-se quantos átomos da molécula possuem a propriedade DA = "D" (a qual é estabelecida quando o átomo é classificado pelo campo de força MMFF94), e o número de aceptores de interação de hidrogênio é computado contando-se quantos átomos da molécula possuem a propriedade DA = "A":

```
Início
Função compInteracaoH(mol : Molecule)
    hba, hbd : inteiro;
    hba ← 0; {número de aceptores de interação de hidrogênio}
    hbd ← 0; {número de doadores de interação de hidrogênio}

    Para cada Átomo atom de mol, Faça:
        Se atom.DA = "A", Então:
            hba ← hba + 1;
        Fim Se;
        Se atom.DA = "D", Então:
            hbd ← hbd + 1;
        Fim Se;

Fim Função;
```

```

Fim Para;

adicionar propriedade a mol ("hydrogen_bond_acceptors" => hba);
adicionar propriedade a mol ("hydrogen_bond_donors" => hbd);

Fim Função;

Fim

```

O logP é usado para calcular a lipofilicidade de um composto. Experimentalmente é definido da seguinte forma:

$$\log P = \log \left(\frac{[\text{composto}_{\text{fase orgânica}}]}{[\text{composto}_{\text{fase aquosa}}]} \right)$$

Tendo-se uma matriz composta por dois líquidos imiscíveis, sendo uma fase orgânica (geralmente octanol) e outra fase aquosa, adiciona-se o composto a este meio; após o equilíbrio, mede-se a concentração do composto na fase orgânica e na fase aquosa e utiliza-se esses valores na equação acima para se calcular o logP. Diversos algoritmos tem sido desenvolvidos para se calcular o logP computacionalmente. Escolheu-se implementar o algoritmo XLOP3-AA desenvolvido Cheng *et al.*, 2007. Este algoritmo apresenta um modelo aditivo em que cada átomo ou grupo de átomos na molécula contribui com uma parcela específica para o logP total, também são adicionados alguns fatores de correção. A equação para cálculo do logP é a seguinte:

$$\log P = \sum_{i=1}^M a_i A_i + \sum_{j=1}^N c_j C_j$$

em que a_i é a contribuição do átomo/grupo tipo i no cálculo do logP, A_i é o número de ocorrências do átomo/grupo tipo i , c_j é a contribuição do fator de correção j e C_j é o número de ocorrências do fator de correção j . São definidos 83 tipos de átomos no algoritmo XLOP3-AA, alguns sendo apresentados na tabela 12.

Tabela 12 – Exemplos de tipos de átomos definidos no algoritmo XLOP3-AA

Id	Símbolo	Descrição	Contribuição (a_i)
10	C.3.2h	Carbono tetravalente conectado com dois hidrogênios	0,3436
47	N.3	Nitrogênio trivalente não conectado a hidrogênio	0,1799
62	O.ar	Oxigênio em anel aromático	0,5238
76	F.pi	Flúor conectado a um sistema conjugado	0,4401
77	F	Flúor	0,5360
84	-C#N	Grupo ciano	0,0337
86	-NO2	Grupo nitro	1,2442

Fonte: Adaptado de Cheng *et al.*, 2007.

Dois fatores de correção são empregados: o primeiro contabiliza interações de hidrogênio internas, as quais reduzem a hidrofobicidade da molécula ($c_j = 0,6123$), e a segunda contabiliza a presença de ambientes aminoácidos, ou seja, moléculas que apresentam simultaneamente grupos carboxil e nitrogênios sp^3 alifáticos ($c_j = -2,4431$), as quais se encontram ionizadas mesmo em condições de pH neutro (FIGURA 63).

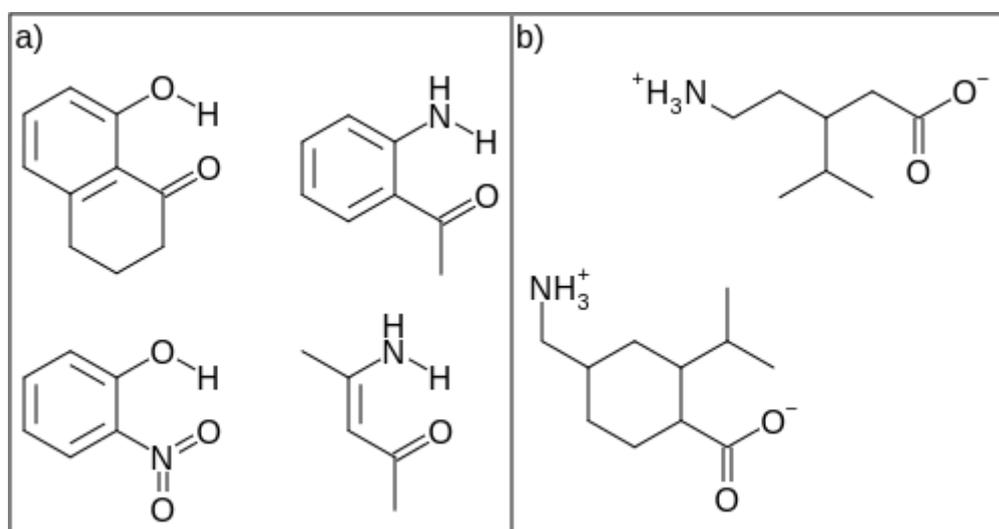


Figura 63 – Exemplos de moléculas para as quais se aplicam fatores de correção no cálculo do logP pelo algoritmo XLOP3-AA.

Fonte: a) Adaptado de Cheng *et al.* (Supporting Information) (2007, p. 5); b) Do autor.

Legenda: a) Interação de hidrogênio interna;

b) Ambientes do tipo aminoácido.

O logS é o logaritmo da solubilidade de um composto, a qual corresponde à máxima concentração do composto em solvente aquoso quando em equilíbrio com seu sólido. Para calcular o logS, adotou-se o algoritmo desenvolvido por Hou *et al.*, 2004. Este utiliza um modelo aditivo similar ao algoritmo de logP descrito anteriormente. A equação é a seguinte:

$$\log S = C_0 + \sum_i a_i n_i + \sum_j b_j B_j$$

em que C_0 é uma constante ($C_0 = 0,518$), a_i é a contribuição do átomo tipo i no cálculo do $\log S$, n_i é o número de ocorrências do átomo tipo i , b_j é a contribuição do fator de correção j e B_j é o número de ocorrências do fator de correção j . São definidos 76 tipos de átomos, alguns exemplos são mostrados na tabela 13.

Tabela 13 – Exemplos de tipos de átomos definidos no algoritmo de cálculo de $\log S$

nº	Representação SMART	Descrição	Contribuição (a_j)
1	[CX4;H4] [CX4;H3]	Carbono tetraédrico ligado a quatro hidrogênios, ou ligado a três hidrogênios	0,273
2	[CX4;H3][#6]	Carbono tetraédrico ligado a três hidrogênios, quarta ligação é feita com outro carbono	-0,099
26	[C;H0]=O	Carbono com ligação dupla a oxigênio e não ligado a hidrogênio	-0,950
45	[#8]=C	Oxigênio realizando ligação dupla com um carbono	0,535
50	[N;H2][C]	Nitrogênio ligado a dois hidrogênios e a um carbono	0,205
62	N#*	Nitrogênio com ligação tripla com qualquer outro átomo capaz de fazer ligação tripla	0,166
65	[S;H0]	Enxofre não ligado a hidrogênio	-1,118

Fonte: Adaptado de How *et al.*, 2004.

Utilizam-se dois fatores de correção para corrigir desvios devido às interações intramoleculares hidrofóbicas. No primeiro caso contabiliza-se a quantidade de carbonos hidrofóbicos, definidos como aqueles cuja hibridização é sp^3 ou sp^2 , não estão conectados a heteroátomos em uma relação 1-4, e não pertencem a sistemas de anéis ($b_j = -0,230$). O segundo fator de correção adiciona ao $\log S$ o quadrado do peso molecular do composto (MW^2) ($b_j = 0,00008$).

O PSA é definido como a superfície pertencente a átomos polares e está relacionado com o transporte passivo através da membrana celular. O método tradicional de se calcular o PSA consiste em se calcular a superfície molecular a partir da estrutura tridimensional e identificar quanto dessa superfície pertence a átomos polares; consistindo em um método demorado. A fim de se realizar um cálculo rápido dessa propriedade, adotou-se o método

TPSA (*Topological Polar Surface Area*, Área de Superfície Polar Topológica) desenvolvido por Ertl et al., 2000, o qual apresenta um modelo aditivo similar aos já vistos no cálculo de logP e logS. A equação para cálculo do TPSA é definida da seguinte forma:

$$TPSA = \sum_i n_i c_i$$

em que n_i é a contribuição do átomo tipo i no cálculo do TPSA e c_i é o número de ocorrências do átomo tipo i . 43 tipos de átomos são definidos, número reduzido quando comparado aos outros métodos (cálculo de logP e logS) uma vez que átomos de carbono não são considerados no cálculo de TPSA. Na tabela 14 estão alguns exemplos de tipos de átomos.

Tabela 14 – Exemplos de tipos de átomos definidos no algoritmo de cálculo de TPSA

Tipo de átomo	Descrição	Contribuição (n_i)
[N](-*)(-*)(-*)	Nitrogênio realizando três ligações simples com quaisquer átomos diferente de hidrogênio	3,24
[NH](-*)(-*)	Nitrogênio realizando três ligações simples, uma sendo com hidrogênio	12,03
[n](:*):*	Nitrogênio aromático, ligado a dois outros átomos	12,89
[O](-*)(-*)	Oxigênio realizando duas ligações simples com quaisquer átomos diferente de hidrogênio	9,23
[O]=*	Oxigênio realizando ligação dupla com qualquer átomo	17,07
[S]=*	Enxofre realizando ligação dupla com qualquer átomo	32,09
[P](-*)(-*)(-*)	Fósforo realizando três ligações simples com quaisquer átomos	13,59

Fonte: Adaptado de Ertl *et al.*, 2000.

3. 7 VIRTUAL SCREENING

Na pesquisa por novos fármacos, busca-se por moléculas capazes de interagir com um alvo molecular específico. Essa interação pode ser modelada computacionalmente utilizando-se *docking molecular*. O alvo molecular é também chamado de receptor e a molécula capaz de interagir com o receptor é definida como ligante. Ao se rodar um programa de *docking*, duas

questões são avaliadas: qual é a orientação relativa preferencial que o ligante adota ao se ligar ao sítio ativo do receptor, a chamada pose, e qual é a energia de interação resultante, que pode ser calculada por uma função de *score* (MENG *et al.*, 2011). Quando se tem uma coleção virtual de moléculas, outra questão pode ser levantada: dentre as moléculas da coleção, quais delas apresentam boa interação com o receptor, e deveriam ser avaliadas em laboratório, e quais deveriam ser descartadas por apresentar baixa afinidade. Neste caso, tem-se o *virtual screening*, em que se roda o *docking* de várias moléculas contra um receptor, sendo selecionadas aquelas cujas funções de *score* retornam os valores mais negativos (MUKESH E RAKESH, 2011).

A fim de se selecionar bioisómeros com melhor afinidade a um determinado receptor, implementou-se uma função para *virtual screening* entre os bioisómeros e um receptor escolhido pelo usuário. Utiliza-se o programa Autodock Vina (TROTT E OLSON, 2010) como *engine* de *docking molecular*, sendo assim, é necessário que o usuário tenha este programa instalado no computador, o qual pode ser obtido em <http://vina.scripps.edu/download.html>. O AutoDock Vina é um programa de código aberto usado em *docking molecular* desenvolvido por Dr. Oleg Trott (TROTT E OLSON, 2010). Utiliza uma metodologia de *docking* rígido-flexível, no qual o receptor é mantido como estrutura rígida (sem rotação, translação ou torção de diedros) e o ligante possui flexibilidade (liberdade de rotação, translação e rotação de diedros para ligações simples não-cíclicas); implementa o algoritmo otimizador global de Pesquisa Local Iterada para busca das poses favoráveis, e possui uma função de *score* escrita da seguinte maneira:

$$c = \sum_{i < j} f_{i,t_j}(r_{ij})$$

em que f_{i,t_j} é uma função descrevendo interações não-ligantes entre os pares de átomos i e j , estando ponderada por um fator obtido por regressão linear; e r_{ij} é a distância entre os átomos i e j . As interações 1-4 são excluídas da função. A tabela 15 apresenta os termos da função de *score* e seus respectivos pesos.

Tabela 15 – Termos da função de *score* empregada pelo programa AutoDock Vina.

Peso	Termo	Forma funcional
-0,0356	gauss ₁	$\text{gauss}_1(d) = e^{-(d/0,5\text{\AA})^2}$
-0,00516	gauss ₂	$\text{gauss}_2(d) = e^{-((d - 3\text{\AA})/2\text{\AA})^2}$
0,840	Repulsão	$\text{Repulsão}(d) = \begin{cases} d^2, & \text{se } d < 0 \\ 0, & \text{se } d \geq 0 \end{cases}$
-0,0351	Hidrofóbico	$\text{Hidrofóbico}(d) = \begin{cases} 1, & \text{se } d < 0,5 \\ -d + 1,5, & \text{se } 0,5 < d \leq 1,5 \\ 0, & \text{se } d \geq 1,5 \end{cases}$
-0,587	Interação de hidrogênio	$\text{IntHidro}(d) = \begin{cases} 1, & \text{se } d \leq -0,7 \\ \frac{-d}{0,7}, & \text{se } -0,7 < d < 0 \\ 0, & \text{se } d \geq 0 \end{cases}$
0,0585	N_{rot}	Número de ligações com liberdade de rotação em seus diedros presentes no ligante

Fonte: Adaptado de Trott e Olson, 2010.

Com o propósito de otimizar o tempo de busca e garantir que o ligante se ancore na região do sítio ativo do receptor (ou outra região de interesse, como um sítio alostérico), o *docking* é limitado a uma região do espaço definida como *grid box* (FIGURA 64), dentro da qual o algoritmo testará diferentes poses para o ligante.

O *grid box* é definido pelo programa AutoDock Vina como uma caixa cúbica por meio dos seguintes parâmetros:

- center_x: centro do *grid box* no eixo x, em angstroms;
- center_y: centro do *grid box* no eixo y, em angstroms;
- center_z: centro do *grid box* no eixo z, em angstroms;
- size_x: comprimento da aresta do *grid box* no eixo x, em angstroms;
- size_y: comprimento da aresta do *grid box* no eixo y, em angstroms;
- size_z: comprimento da aresta do *grid box* no eixo z, em angstroms;

Estes parâmetros podem ser informados na linha de argumentos ao se chamar o programa vina ou podem ser escritos em arquivo de texto, informando-se o nome do arquivo pela linha de argumentos na chamada do programa. O *grid box* apresentado na figura 64 é escrito em arquivo de texto conforme figura 65.

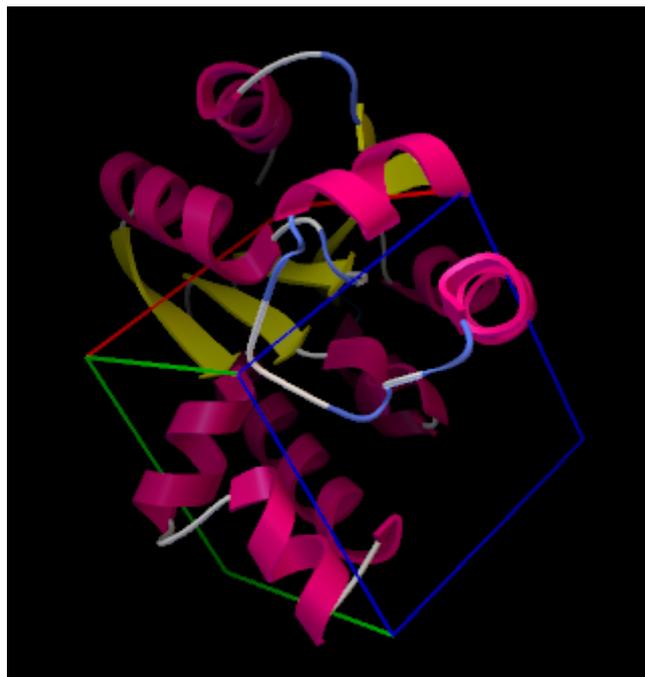


Figura 64 – *Grid box* empregado no *docking molecular*. O *grid box* é a região do espaço, geralmente centrado no sítio ativo do receptor, onde o ligante será ancorado. Imagem obtida com a estrutura PDB 1WE2 (PEREIRA et al., 2004).

Fonte: Do autor.

```
center_x = 25.474
center_y = 34.143
center_z = 30.904
size_x   = 20
size_y   = 20
size_z   = 20
```

Figura 65 – Arquivo de configuração contendo os parâmetros do *grid box* para rodar *docking molecular* com o programa AutoDock Vina.

Fonte: Do autor.

As etapas para se rodar o *virtual screening* são as seguintes:

- a) Inicialmente é feita uma busca automática pelo programa vina, sendo pesquisados os diretórios presentes no PATH (variável de ambiente que aponta os diretórios onde existem programas executáveis) e o diretório do usuário; no caso do sistema operacional Windows, adiciona-se à pesquisa os diretórios "C:\Program Files\The Scripps Research Institute\" e "C:\Program Files (x86)\The Scripps Research Institute\". Caso o programa vina não seja encontrado, o usuário deve informar em qual diretório se encontra o vina;

- b) Deve-se selecionar um receptor em formato pdb. O receptor deve possuir obrigatoriamente pelo menos um ligante cocrystalizado a fim de auxiliar na definição do sítio-ativo;
- c) Seleciona-se um dos ligantes do receptor para definir o sítio ativo ou *grid box*; o centro do grid corresponde ao centro geométrico do ligante cocrystalizado; as arestas do grid são calculadas da seguinte forma: calcula-se o raio molecular de cada um dos bioisómeros, o maior valor encontrado é multiplicado por 3 e este é atribuído às arestas de todos os eixos coordenados. Se o comprimento da aresta for menor que 22,5 Å, ajusta-se a aresta para 22,5 Å, conforme protocolo estipulado pelos desenvolvedores do AutoDock Vina;
- d) Escolhe-se um diretório onde serão armazenados os *inputs* e *outputs* do *virtual screening*, sendo criada a árvore de diretórios mostrada na figura 66;
- e) O receptor é salvo no formato pdbqt no diretório “receptor”, os bioisómeros são salvos no formato pdbqt no diretório “ligands”, especifica-se o sítio ativo no arquivo config.txt salvo no diretório “gridBox”;
- f) Para cada bioisómero roda-se o docking molecular, por meio da chamada ao seguinte comando:

```
vina      --receptor      receptor/receptor.pdbqt      --ligand
ligands/ligandName.pdbqt      --config      gridBox/config.txt      --out
out/receptor-ligandName.pdbqt --log log/receptor-ligandName.log
```

em que "ligandName" é substituído pelos respectivos nomes atribuídos a cada bioisómero. No diretório “out” são salvas as poses de *docking* e no diretório “log” as energias obtidas pela função de *score*;

- g) Ao final é gerada uma tabela em ordem decrescente da função de *score* obtida no *docking molecular* para cada bioisómero.



Figura 66 – Árvore de diretórios organizada para se rodar *virtual screening*.

Fonte: Do autor.

3. 8 ESTUDO DE CASO

O seguinte estudo de caso é proposto com o intuito de se ilustrar o uso da ferramenta MB-Isoster: a tuberculose é uma doença infecciosa transmitida por contato direto entre pessoas, tais quando um indivíduo sadio inala gotículas de saliva provocadas pelo espirro de alguém contaminado; afeta principalmente o pulmão, tendo como manifestações clínicas a tosse crônica, perda de apetite, perda de peso, febre e hemoptise (LAWN E ZUMLA, 2011). É causada pela bactéria *Mycobacterium tuberculosis*, um bacilo aeróbio, não encapsulado e imóvel; sendo um patógeno intracelular obrigatório que cresce preferencialmente em tecidos com alta taxa de oxigênio, como ocorre com os pulmões (GRANGE, 2009) (FIGURA 67a).

Dentre as vias metabólicas apresentadas por esse micro-organismo, a via bioissintética do chiquimato é apontada como potencial alvo terapêutico, por ser essencial para a sobrevivência de *M. tuberculosis* e não estar presente em humanos (CORACINI E AZEVEDO JR., 2014). Esta via é composta por sete enzimas, iniciando-se com os substratos eritrose-4-fosfato e fosfoenol piruvato, até a conversão final para corismato, o qual é precursor na síntese de aminoácidos aromáticos (HERRMANN E WEAVER, 1999). A enzima chiquimato quinase (EC: 2.7.1.71) (FIGURA 67b) é a quinta na sequência da via metabólica, cataliza a transferência de um grupo fosfato do ATP para o ácido chiquímico, produzindo chiquimato 3-fosfato e ADP (FIGURA 67c) (HARTMANN *et al.*, 2006).

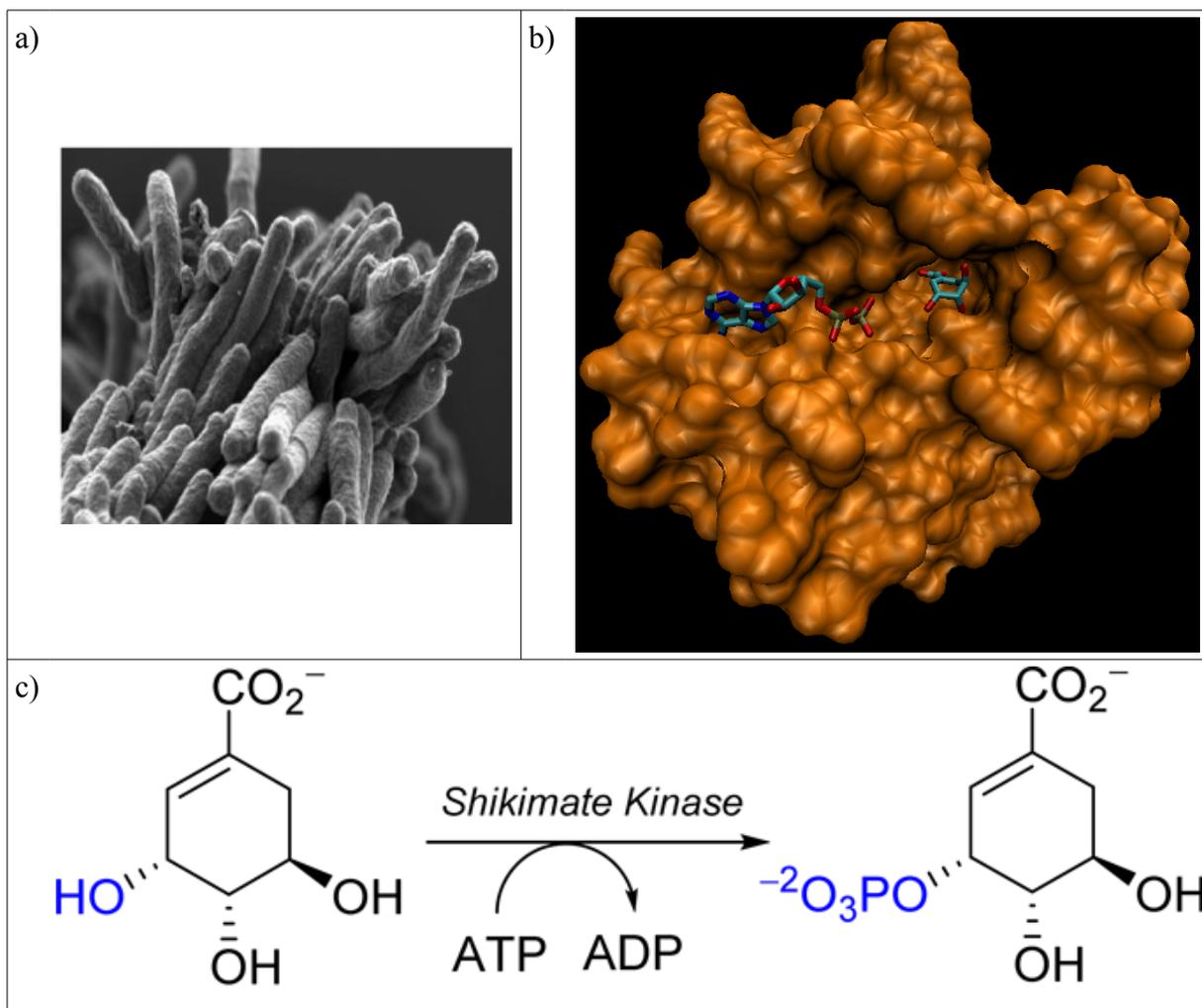


Figura 67 – Sistema modelo usado para ilustrar o uso do programa MB-Isoster.

Fonte: a) Koch e Mizrahi (2018, p. 555); b) Do autor; c) Prado *et al.* (2016, p. 5472).

Legenda: a) Bactéria *Mycobacterium tuberculosis*, agente causador da tuberculose.

b) Enzima chiquimato quinase em complexo com ácido chiquímico e ADP, código pdb: 2DFN (DIAS *et al.*, 2007).

c) Reação enzimática catalizada pela enzima chiquimato quinase.

Neste estudo de caso, a enzima chiquimato quinase foi selecionada como alvo terapêutico, e seu substrato natural, o ácido chiquímico, como composto *hit* para desenho de novas moléculas. Inicialmente, utilizou-se a ferramenta de análise de interação receptor-ligante (seção “3. 5 ANÁLISE DE INTERAÇÕES RECEPTOR-LIGANTE”), avaliando-se as interações na estrutura cristalizada de chiquimato quinase em complexo com ácido chiquímico obtido do PDB, código pdb: 2IYX (HARTMANN *et al.*, 2006). O grupo funcional cujos átomos apresentaram interação menos favorável com o receptor foi selecionado para ser substituído nos bioisómeros (seção “3. 4 CONSTRUÇÃO DE MOLÉCULAS ANÁLOGAS POR BIOISOSTERISMO”). Também foram construídas novas moléculas por substituição de átomos de hidrogênio a partir do composto *hit* (seção “3. 4. 2 Construção de novas moléculas por substituição de átomos de hidrogênio”). As moléculas formadas foram filtradas de acordo

com suas propriedades físico-químicas (seção “3. 6 CÁLCULO DE PROPRIEDADES *DRUG-LIKE*”). Os parâmetros adotados foram:

- a) $50 \leq \text{peso molecular} \leq 300$;
- b) $0 \leq \log P \leq 3$;
- c) $-4,5 \leq \log S \leq 1$;
- d) $30 \leq \text{TPSA} \leq 200$;
- e) $0 \leq \text{número de doadores de interação de hidrogênio} \leq 5$;
- f) $0 \leq \text{número de aceptores de interação de hidrogênio} \leq 10$.

Por fim, realizou-se *virtual screening* com as moléculas que passaram no filtro e a enzima chiquimato quinase (código PDB: 2iyx) (seção “3. 7 *VIRTUAL SCREENING*”). O *grid box* foi definido da seguinte forma:

- a) $\text{center_x} = 2,927$;
- b) $\text{center_y} = 34,475$;
- c) $\text{center_z} = 19,298$;
- d) $\text{size_x} = 22,5$;
- e) $\text{size_y} = 22,5$;
- f) $\text{size_z} = 22,5$.

4 RESULTADOS E DISCUSSÃO

O programa MB-Isoster foi desenvolvido em java e salvo no arquivo MB-Isoster.jar; está disponível para download na página <http://molmod-cs.unifal-mg.edu.br/tools>. Para iniciá-lo no sistema operacional Windows, basta dar duplo clique no ícone do programa; estando no Linux, deve-se abrir o terminal, mover-se para o diretório onde está o software e digitar:

```
java -jar MB-Isoster.jar
```

Opcionalmente, pode-se incluir na linha de argumentos o nome do arquivo contendo a molécula que se deseja abrir:

```
java -jar MB-Isoster.jar molecule.mol
```

em que “molecule.mol” deve ser substituído pelo nome do arquivo que será aberto, o mesmo devendo estar no diretório do executável ou deve-se especificar o caminho total do arquivo, como “/home/usuario/molecule.mol”. Pode-se também criar um script bash, com o nome “mbisoster” (FIGURA 68), salvá-lo em um diretório do PATH (como /usr/local/bin ou home/usuario/bin) e torná-lo executável com o comando:

```
chmod +x mbisoster
```

desse modo basta executar o comando:

```
mbisoster
```

ou

```
mbisoster molecule.mol
```

não sendo necessário estar no diretório do executável.

```
#!/bin/bash

#caminho onde está salvo o executável MB-Isoster.jar
#(alterar a variável se necessário)
executavel=/home/user/MB/MB-Isoster.jar

#caminho onde está instalado o java
#(usar essa variável caso o java não esteja no PATH)
java=/home/user/jdk/jdk1.8.0_121/bin/java

$java -jar $executavel $@

exit
```

Figura 68 – Script bash para rodar o programa MB-Isoster.
Fonte: Do autor.

O ícone apresentado na figura 69 foi desenhado para o programa MB-Isoster.

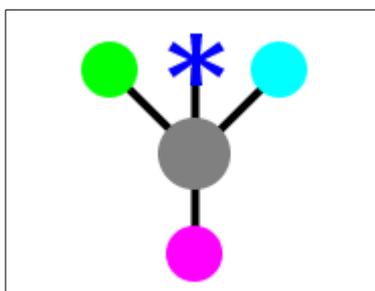


Figura 69 – Ícone do programa MB-Isoster.

Fonte: Do autor.

4. 1 INTERFACE GRÁFICA

Ao se iniciar o programa, três painéis são visíveis, conforme mostrado na figura 70. O painel superior à esquerda é o *Molecule Panel* ou “Painel da Molécula”, é onde a molécula aberta pelo usuário é exibida, podendo ser manipulada; o painel superior à direita é o *Biososteres Panel* ou “Painel dos Bioisósteros”, onde são exibidos os bioisósteros gerados pelo programa; e o painel inferior à esquerda é o *Log Panel* ou “Painel de Mensagens”, onde são exibidas mensagens para o usuário referentes aos cálculos e operações efetuadas. Cada um dos painéis pode ser movido e redimensionado dentro da tela principal, permitindo ao usuário alterar o *layout* do sistema.

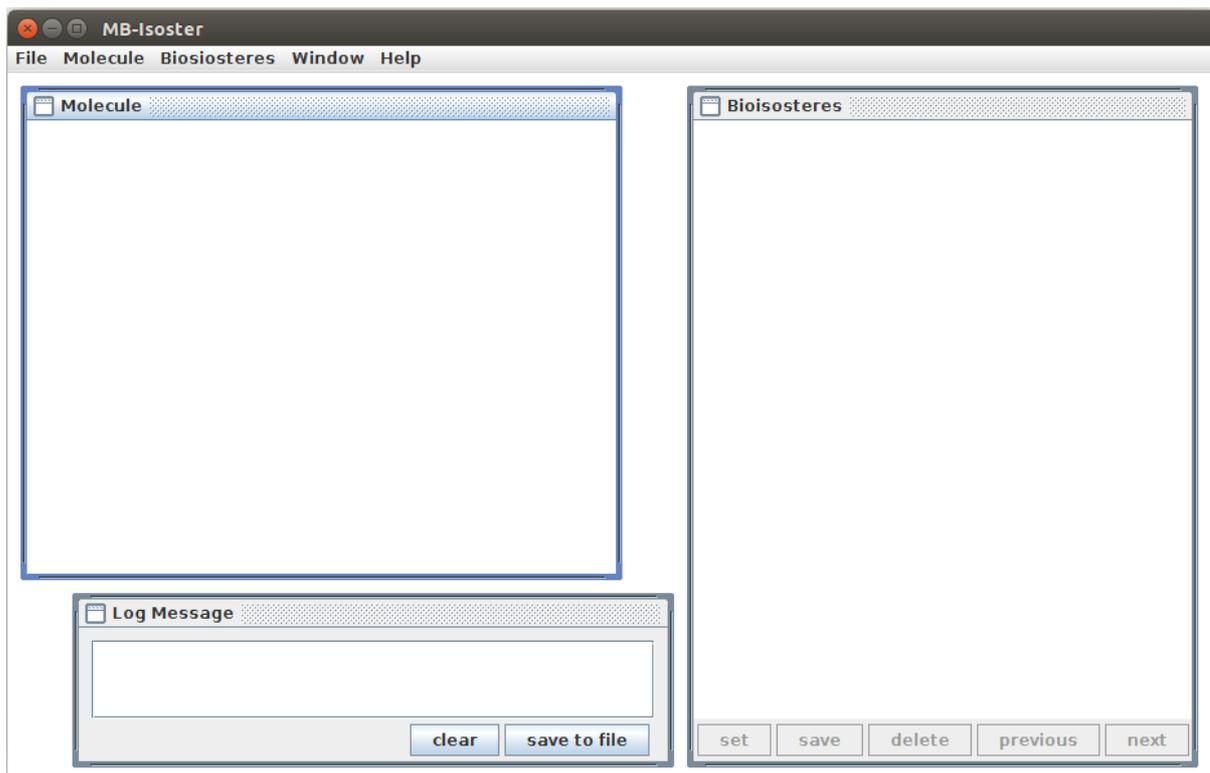


Figura 70 – Tela inicial do programa MB-Isoster.
Fonte: Do autor.

Para abrir uma molécula deve-se ir em **File** → **Open** e selecionar a estrutura de interesse. As teclas de direção do teclado são usadas para rotacionar a molécula nos eixos X e Y, e as setas “Z” e “X” rotacionam a molécula no eixo Z. A molécula é inicialmente exibida em um tamanho *default* otimizado em relação ao tamanho atual do *Molecule Panel*, para aumentar o zoom utiliza-se a tecla “V” e para diminuir usa-se a tecla “C”. A operação de zoom é limitada, o máximo zoom é o dobro do tamanho *default* e o mínimo zoom é a metade do tamanho *default*. Operações de translação não são implementadas. Clicando-se em um ou mais átomos da molécula, os mesmos são selecionados, pode-se então ir em **Molecule** → **Make biososteres** para iniciar a construção dos biosósteros. Uma vez concluídos, os biosósteros são exibidos no *Biososteres Panel*, podendo ser manipulados por cinco botões abaixo deste painel (FIGURA 71). Uma vez que são gerados vários biosósteros e apenas um por vez pode ser exibido no *Biososteres Panel*, os botões **next** e **previous** são usados para se navegar entre as diferentes estruturas, o botão **delete** apaga o biosósteros exibido (esta operação não pode ser desfeita), o botão **save** permite salvar o biosósteros exibido em um dos formatos de arquivo suportados ou em uma imagem png; por fim, o botão **set** cria uma cópia do biosósteros exibido e o envia para o *Molecule Panel*, de forma que novos biosósteros possam ser formados. Para salvar todos os biosósteros deve-se ir em **Biososteres** → **Save**

bioisósteres e selecionar um dos formatos de arquivo disponíveis, há opções para salvar todas as moléculas em um único arquivo ou cada molécula em arquivo separado.

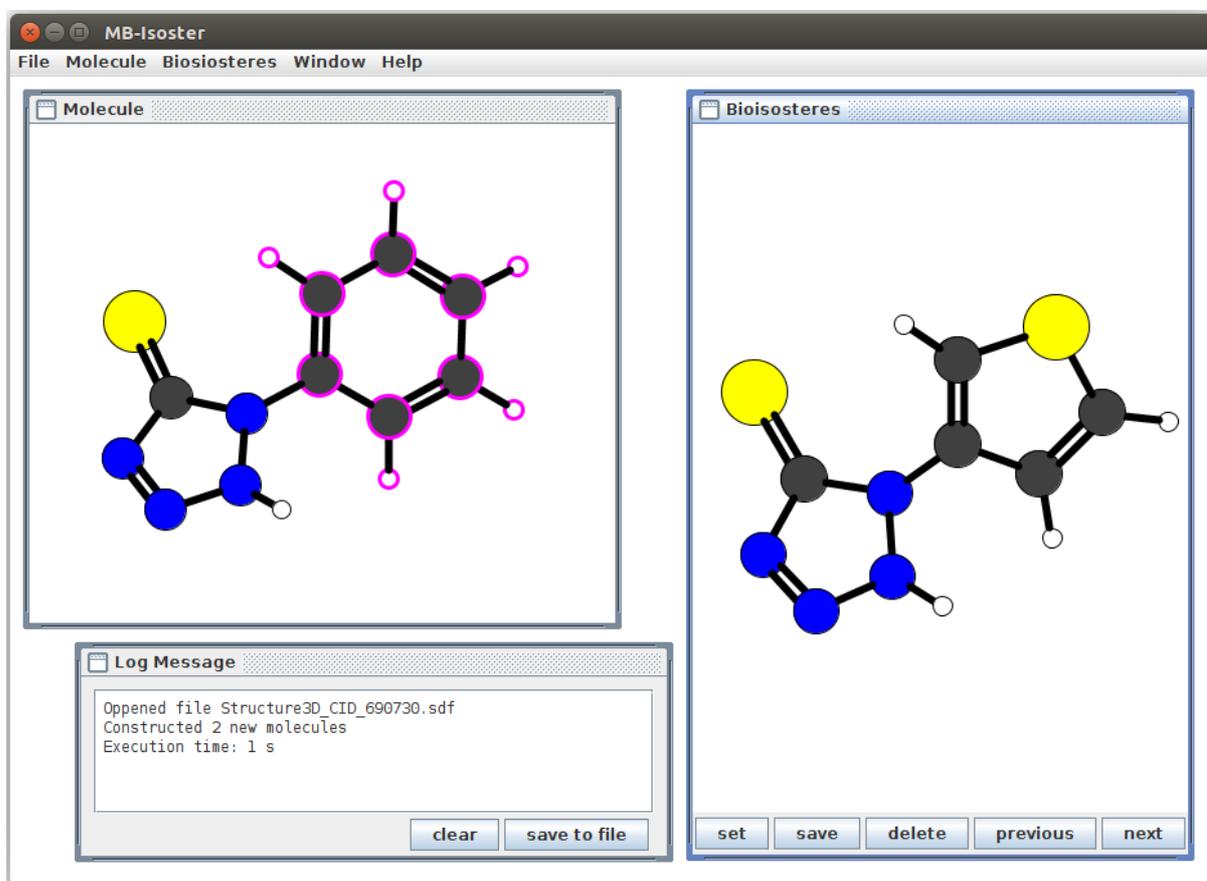


Figura 71 – Bioisósteros exibidos no *Bioisosteres Panel*. Os átomos selecionados para substituição bioisostérica estão realçados em magenta no *Molecule Panel*.

Fonte: Do autor.

4. 1. 1 Barra de menus

A barra de menus é formada por cinco menus: **File**, **Molecule**, **Bioisosteres**, **Window** e **Help**. **File** possui funções de entrada de dados, **Molecule** possui funções que operam sobre a molécula exibida em *Molecule Panel*, **Bioisosteres** tem opções para manipular os bioisósteros, **Window** é usado para modificar o *layout* do programa e **Help** possui funções de ajuda. Os submenus de cada menu são descritos a seguir:

- Menu **File**
 - Submenu **Open**: abre uma molécula e a exhibe em *Molecule Panel*;
 - Submenu **Extract from PDB-complex**: abre a janela de diálogo “Read Ligand

from PDB-Complex” para abrir uma molécula que esteja complexada com um receptor, podendo-se calcular as interações não ligante entre receptor e ligante;

– Submenu **Exit**: fecha o programa;

- Menu **Molecule**

– Submenu **Compute energy**: calcula a energia da molécula baseado no campo de força MMFF94s, sem minimização prévia da estrutura, o resultado é exibido em *Log Panel*;

– Submenu **Minimize geometry**: minimiza a geometria da molécula usando mecânica molecular com implementação do campo de força MMFF94s, o tempo de minimização e a energia da molécula após a minimização são exibidas em *Log Panel*;

– Submenu **Compute properties**: as propriedades *drug-like* da molécula (peso molecular, logP, logS, TPSA, número de doadores de interação de hidrogênio e número de aceptores de interação de hidrogênio) são computadas e exibidas em *Log Panel*. Se a molécula for salva em formato sdf, as propriedades calculadas estarão presentes no arquivo de saída;

– Submenu **Add hydrogens**: adiciona hidrogênios à molécula se os mesmos estiverem faltando;

– Submenu **Remove hydrogens**: remove átomos de hidrogênio da molécula;

– Submenu **Make bioisosteres**: irá construir bioisósteros modificando-se os átomos da molécula selecionados pelo usuário; se nenhum átomo estiver selecionado, é exibida uma mensagem de aviso;

– Submenu **Build over all H**: para cada hidrogênio da molécula, será criado um conjunto de novas moléculas em que o hidrogênio é substituído por um fragmento presente na biblioteca especial contendo substituintes representativos (metil, hidroxil, fenil, amina, etc), não é necessário que o usuário selecione átomos para esta operação;

– Submenu **Save molecule**: abre um novo submenu para salvar a molécula em um dos formatos de arquivo suportados (mol (V2000 e V3000), sdf (V2000 e V3000), mol2 e pdb) ou em imagem png;

– Submenu **Show atom's name**: usado para mostrar/esconder o nome dos átomos em *Molecule Panel*;

- Menu **Bioisosteres**

– Submenu **Remove bioisosteres**: apaga todos os bioisósteros contidos em

Bioisosteres Panel, esta operação não pode ser desfeita;

- Submenu **Filter bioisosteres**: abre a janela de diálogo “Molecule Filter” a fim de se filtrar os bioisósteros de acordo com suas propriedades *drug-like*;
- Submenu **Virtual screening**: abre a janela de diálogo “Virtual Screening” para rodar *virtual screening* entre um receptor selecionado pelo usuário e os bioisósteros formados;
- Submenu **Save bioisosteres**: abre um novo submenu para salvar todos os bioisósteros em um dos formatos de arquivo suportados, dependendo da opção escolhida, pode-se salvar cada bioisósteros em arquivo individual ou todos em um mesmo arquivo;
- Submenu **Show atom's name**: usado para mostrar/esconder o nome dos átomos em *Bioisosteres Panel*;
- Menu **Window**
 - Submenu **Show log panel**: usado para mostrar/esconder o *Log Panel*;
 - Submenu **Show bioisostere panel**: usado para mostrar/esconder o *Bioisosteres Panel*;
 - Submenu **Resize panels while resizing window**: se esta opção estiver ativa, ao se redimensionar a janela principal do programa, seus painéis internos serão redimensionados na mesma proporção, caso contrário, os painéis não serão redimensionados;
 - Submenu **Show default visualization**: a visualização dos painéis é retornada para a visualização *default* (*Molecule Panel* superior à esquerda, *Bioisosteres Panel* superior à direita, e *Log Panel* inferior à esquerda);
- Menu **Help**
 - Submenu **About**: mostra os créditos do programa;
 - Submenu **How to use**: abre uma janela de ajuda mostrando as funcionalidades do programa.

4. 1. 2 Leitura de moléculas a partir de complexos receptor-ligante

Para acessar a estrutura de uma molécula ligante contida em um arquivo pdb, sendo este um complexo receptor-ligante, deve-se ir em **File** → **Extract from PDB-complex**; será

aberta a janela de diálogo “Read Ligand from PDB-Complex” (FIGURA 72). Pode-se ler um arquivo pdb a partir da web ou de um arquivo salvo no computador do usuário. Para ler a partir da web, deve-se selecionar “Read from web”, digitar o código pdb de interesse na caixa de texto “PDB code:” e clicar no botão **Load**; para ler de um arquivo local, deve-se selecionar “Read from file” e clicar no botão **Browser** para selecionar o arquivo. Se não forem encontrados ligantes no arquivo lido, uma mensagem de aviso é exibida, caso contrário, é exibida uma tabela contendo todos os ligantes lidos com sua id, cadeia e número. Se forem encontrados os nomes e/ou sinônimos dos ligantes, estes são exibidos abaixo da tabela. Ao lado da tabela há um painel onde os ligantes são exibidos, para se exibir os diferentes ligantes deve-se clicar na sua entrada correspondente na tabela. Ao se clicar em **COMPUTE INTERACTIONS**, as interações não ligante (van der Waals e eletrostáticas) entre o receptor e o ligante exibido serão computadas e uma nova tabela é exibida mostrando as energias para cada átomo do ligante. Clicando-se em uma entrada da tabela de energias, seleciona-se o átomo correspondente, para salvar a tabela de energias em arquivo de texto deve-se clicar em **SAVE**, por fim, clicando-se em **OK**, fecha-se a janela de diálogo e a molécula selecionada é enviada ao *Molecule Panel*; também é possível selecionar uma molécula sem calcular as interações não ligante.

Read Ligand from PDB-Complex

1 ligand has found in 1OXR file.

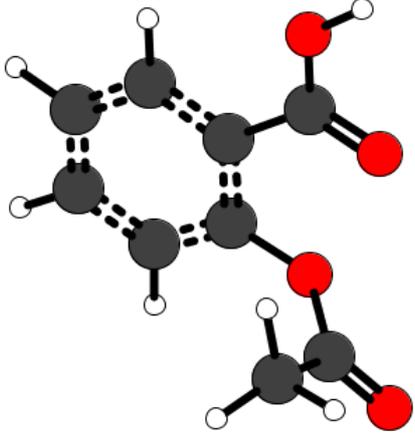
Read from web
 PDB code: 1OXR

Read from file
 Browser: no file

Id	Chain	Residue Number
AIN	A	141

Chemical Name
 AIN: 2-(acetyloxy)benzoic acid

Synonyms
 AIN: acetylsalicylic acid, aspirin



Computed Interactions

Ligand AIN_A_141 from file 1OXR

Atom	vdW	Electrostatic	Total
O1	-0.544	251.362	250.818
O2	-0.724	208.320	207.596
O4	0.009	171.493	171.503
O3	-0.556	70.120	69.564
C4	-0.790	46.638	45.848
C5	-0.912	42.133	41.221
C1	-0.706	40.586	39.880
C6	-0.930	39.630	38.700
H8	1.618	0.000	1.618
H6	-0.196	0.000	-0.196
H7	-0.235	0.000	-0.235
C9	-0.503	-20.439	-20.942
C2	-0.745	-24.438	-25.182
C3	-0.571	-27.932	-28.502
H4	-0.248	-36.700	-36.949
H5	-0.128	-38.354	-38.481
H3	-0.193	-40.223	-40.416
H2	-0.315	-47.728	-48.043
H1	-0.129	-205.057	-205.187
C8	-0.693	-206.772	-207.465
C7	-0.630	-227.838	-228.468

Figura 72 – Janela de diálogo “Read Ligand from PDB-Complex”. É utilizada para ler uma molécula a partir de estrutura receptor-ligante no formato pdb. Opcionalmente, é possível calcular as interações não-ligante entre os átomos do ligante e o sítio ativo do receptor, o que pode ser usado como critério de seleção dos átomos para substituições bioisotéricas.

Fonte: Do autor.

4. 1. 3 Seleção de bioisósteros com propriedades *drug-like*

Após a construção dos bioisósteros, suas propriedades *drug-like* (peso molecular, logP, logS, TPSA, número de doadores de interação de hidrogênio, número de aceptores de interação de hidrogênio) são automaticamente computadas. Para que os bioisósteros sejam filtrados baseado nessas propriedades, deve-se acessar **Bioisosteres** → **Filter bioisosteres**; abrindo-se a janela de diálogo “Molecule Filter” apresentada na figura 73.

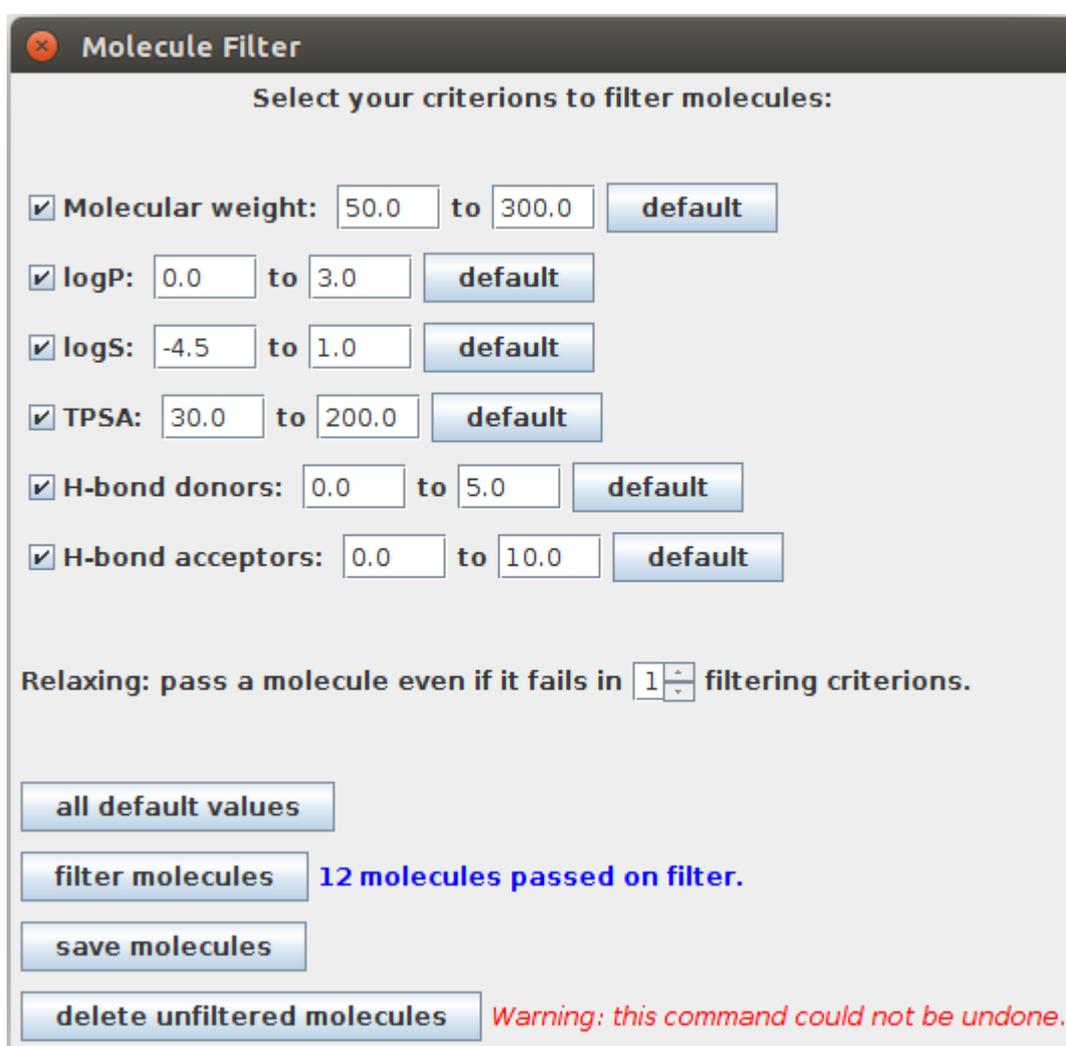


Figura 73 – Janela de diálogo “Molecule Filter”. É utilizada para selecionar os bioisósteros baseado em propriedades *drug-like*. O usuário pode definir quais propriedades usar como critério de filtragem e qual o intervalo de valores válidos.

Fonte: Do autor.

São apresentadas as propriedades e o intervalo válido de cada uma. O usuário deve marcar a caixa de seleção correspondente para selecionar as propriedades usadas no processo de filtragem, se uma propriedade não estiver selecionada, ela não será considerada na seleção

dos bioisósteros. O usuário pode alterar o limite de aceitação superior e inferior, de acordo com suas necessidades; clicando-se no botão **default**, os limites superior e inferior da propriedade correspondente voltam ao seu valor default. Clicando-se no botão **all default values**, os intervalos de valores válidos de todas as propriedades retornam para seu valor default. É possível relaxar o critério de filtragem, de forma que uma molécula seja aceita mesmo se esta estiver fora do intervalo válido de alguma das propriedades. Para isso, deve-se alterar o valor da caixa de seleção numérica contida em “Relaxing: pass a molecule even if it fails in  filtering criteria”; os critérios de relaxamento admitem até três propriedades fora do intervalo aceito. Clicando-se em **filter molecules**, a filtragem é realizada e o número de moléculas que passaram no filtro é apresentada ao lado do botão **filter molecules**. Essa operação não apaga as moléculas rejeitadas, ela apenas define o atributo filterStatus de cada molécula para “ACCEPTED” ou “REJECTED”, se os critérios de filtragem forem alterados (seja adicionando ou retirando propriedades, mudando o intervalo válido, ou definindo critérios de relaxamento), o número de moléculas aceitas e seus status (“ACCEPTED” ou “REJECTED”) poderão ser alterados. Para apagar as moléculas rejeitadas, deve-se clicar em **delete unfiltered molecules**, esta operação não pode ser desfeita. Clicando-se em **save molecules**, as moléculas selecionadas no processo de filtragem serão salvas em um dos formatos de arquivo suportados.

4. 1. 4 Execução de *virtual screening*

Selecionando-se **Bioisosteres** → **Virtual screening**, abre-se a janela de diálogo “Virtual Screening”, a partir da qual serão executadas as etapas para se rodar *virtual screening* (FIGURA 74). A primeira etapa é localizar o programa AutoDock Vina, o que é feito automaticamente quando o programa MB-Isoster é iniciado. Se o programa vina não for encontrado, este deverá ser manualmente localizado pelo usuário; se o vina não estiver instalado no computador, as demais funções na janela “Virtual Screening” ficam congeladas, não sendo possível executar o *virtual screening*. Em seguida seleciona-se o receptor, que pode ser obtido da web pelo banco de dados PDB ou lido em arquivo salvo no computador, de forma similar ao que ocorre na janela “Read Ligand from PDB-Complex”. Os ligantes cocrystalizados encontrados são exibidos abaixo, o usuário seleciona um para ser usado como

referência na definição do *grid box*. Clicando-se em **browser**, seleciona-se um diretório raiz para armazenar os *inputs* e *outputs* do *virtual screening*, o qual será iniciado ao se clicar em **start**. Marcando-se a caixa de seleção “Include starting molecule (showed in Molecule Panel) in virtual screening”, a molécula usada como *hit* para a construção dos bioisómeros também entrará no *virtual screening*. Ao final, será exibida uma tabela com os resultados obtidos. Se os bioisómeros forem filtrados pelas propriedades *drug-like*, porém as moléculas rejeitadas não forem apagadas, estas também serão consideradas no *virtual screening*, aparecendo na tabela abaixo das moléculas aceitas (FIGURA 75).

Virtual Screening

Follow instructions below to execute virtual screening campaign. This function uses vina program as molecular docking engine in order to perform virtual screening with biososters and a selected receptor. Auto Dock Vina program must be installed in your computer to properly use of this function.

- 1. Find vina program**
vina program is automatically found at:
- 2. Select a receptor**
Select a receptor from web or from your computer in pdb format. Receptor must contains at least one ligand in order to define active site.
 from web → **pdb code:** **load**
 from file →
- 3. Define active site**
Select a ligand from receptor to define active site. Active site will be centered on geometric center of selected ligand.
 AIN_A_141
- 4. Choose working directory**
Select a working directory to perform virtual screening.
- 5. Perform virtual screening**
Click start button to starts virtual screening.

Figura 74 – Janela de diálogo “Virtual Screening”. O usuário é conduzido passo a passo na execução de *virtual screening* usando o programa AutoDock Vina como *engine* de *docking molecular*.

Fonte: Do autor.

Virtual Screening Results

Table

Molecule	Energy (kcal/mol)	Molecular Weight	logP	logS	TPSA	H-bond donors	H-bond acceptors	Filter Status
bioisoster_004	-8,1	246,205	1,93	-5,215	72,83	1	5	ACCEPTED
bioisoster_012	-8	219,188	1,421	-3,214	72,56	1	5	ACCEPTED
bioisoster_002	-7,7	195,168	0,334	-3,645	75,63	2	5	ACCEPTED
bioisoster_003	-7,6	204,192	1,934	-2,554	80,76	0	6	ACCEPTED
bioisoster_009	-7,6	236,251	2,916	-5,348	93,06	1	6	ACCEPTED
bioisoster_008	-7,5	215,136	0,307	-2,842	99,43	3	5	ACCEPTED
bioisoster_010	-7,4	220,18	1,559	-4,563	76,99	1	6	ACCEPTED
bioisoster_007	-7,2	216,209	1,372	-2,608	89,05	1	5	ACCEPTED
bioisoster_001	-6,9	204,181	1,729	-3,939	79,19	1	5	ACCEPTED
bioisoster_006	-6,9	215,228	1,046	-2,863	94,84	0	5	ACCEPTED
bioisoster_014	-6,9	179,169	2,612	-3,565	66,4	2	4	ACCEPTED
bioisoster_005	-6,8	242,229	0,614	-4,395	89,54	1	7	ACCEPTED
bioisoster_015	-6,8	180,15	1,991	-3,757	63,6	1	4	ACCEPTED
bioisoster_019	-6,8	180,17	2,571	-3,44	37,3	1	3	ACCEPTED
bioisoster_016	-6,7	215,228	2,794	-2,505	95,45	2	6	ACCEPTED
bioisoster_013	-7,7	235,259	3,728	-3,946	87,66	1	5	REJECTED
bioisoster_017	-7,5	233,188	4,564	-4,286	49,33	2	6	REJECTED
bioisoster_018	-7,5	191,222	4,139	-3,395	49,33	2	3	REJECTED
bioisoster_011	-7,4	236,251	3,938	-3,562	100,55	1	6	REJECTED

Figura 75 – Tabela exibida após a conclusão do *virtual screening*.

Fonte: Do autor.

4. 2 VALIDAÇÃO DO CAMPO DE FORÇA

Os resultados para a validação do campo de força, adotando-se as tolerâncias A, B e C, encontram-se nas tabelas seguintes (TABELA 16, TABELA 17 E TABELA 18). Para alguns tipos de energia, a energia calculada é zero simplesmente porque a molécula não apresenta aquele tipo de interação, como é o caso de diedros impróprios para moléculas sem centro tricoordenado plano. Desse modo, acrescentou-se uma terceira coluna, no qual essas moléculas foram excluídas para o cálculo de percentual de eficácia. As seguintes siglas são usadas:

- a) EB – energia de estiramento de ligação covalente;
- b) EA – energia de estiramento de ângulos;
- c) EBA – termo cruzado das energias de estiramento de ligação covalente e ângulos;
- d) EOOP – energia de diedro impróprio;
- e) ET – energia de torção do diedro;
- f) EvdW – energia de Van der Waals;
- g) EQ – energia eletrostática;
- h) E_{total} – energia total.

Tabela 16 – Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,0001 kcal/mol.

Parcela de energia	Todas as moléculas	Excluindo-se moléculas de energia zero
EB	42,58%	42,27%
EA	11,43%	10,96%
EBA	42,97%	42,67%
EOOP	85,02%	75,80%
ET	63,34%	59,09%
EvdW	45,73%	44,93%
EQ	53,61%	52,30%
E_{total}	17,87%	17,44%

Fonte: Do autor.

Tabela 17 – Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,001 kcal/mol.

Parcela de energia	Todas as moléculas	Excluindo-se moléculas de energia zero
EB	74,11%	73,98%
EA	59,66%	59,45%
EBA	86,20%	86,13%
EOOP	95,27%	92,36%
ET	87,12%	85,63%
EvdW	87,91%	87,73%
EQ	88,30%	87,97%
E _{total}	60,84%	60,63%

Fonte: Do autor.

Tabela 18 – Eficácia no cálculo de energia implementando-se o campo de força MMFF94s, assumindo tolerância de 0,01 kcal/mol.

Parcela de energia	Todas as moléculas	Excluindo-se moléculas de energia zero
EB	99,61%	99,60%
EA	98,55%	98,55%
EBA	97,63%	97,62%
EOOP	99,61%	99,36%
ET	94,48%	93,84%
EvdW	100,00%	100,00%
EQ	98,95%	98,92%
E _{total}	89,49%	89,43%

Fonte: Do autor.

17,44% das moléculas do pacote de validação do campo de força MMFF94s tiveram suas energias calculadas dentro da tolerância de 0,0001 kcal/mol, enquanto que 89,43% das energias foram calculadas com a precisão menor de 0,01 kcal/mol. 80 moléculas apresentaram erro no cálculo da energia potencial superior a 0,01 kcal/mol, o maior erro encontrado foi de 0,962 kcal/mol. Fazendo-se um diagnóstico mais detalhado para as diferentes parcelas de energia, foi constatado que os erros residem principalmente na atribuição incorreta de alguns parâmetros de campo de força, muitos dos quais não estão definidos nas tabelas de parâmetros e devem ser calculados por regras empíricas, as quais são cheias de detalhes e mais difíceis de implementar. Ainda deverão ser feitos mais testes e correções no algoritmo a fim de que todas

as moléculas do pacote de validação tenham suas energias calculadas ao menos dentro da tolerância de 0,01 kcal/mol.

4.3 ESTUDO DE CASO

O cálculo das interações não-ligante entre o ligante cocrystalizado chiquimato e a enzima chiquimato quinase foi realizado por meio da janela de diálogo “Read Ligand from PDB-Complex” (**File** → **Extract from PDB-complex**), fazendo-se o download da estrutura 2iyx (TABELA 19). Pelos resultados mostrados na tabela 19, observa-se que os átomos O2 e O3 possuem os maiores valores de energia para interações não-ligante com o receptor. Como estes fazem parte da função carboxilato da molécula chiquimato (FIGURA 76), esse grupamento foi selecionado para substituição bioisostérica.

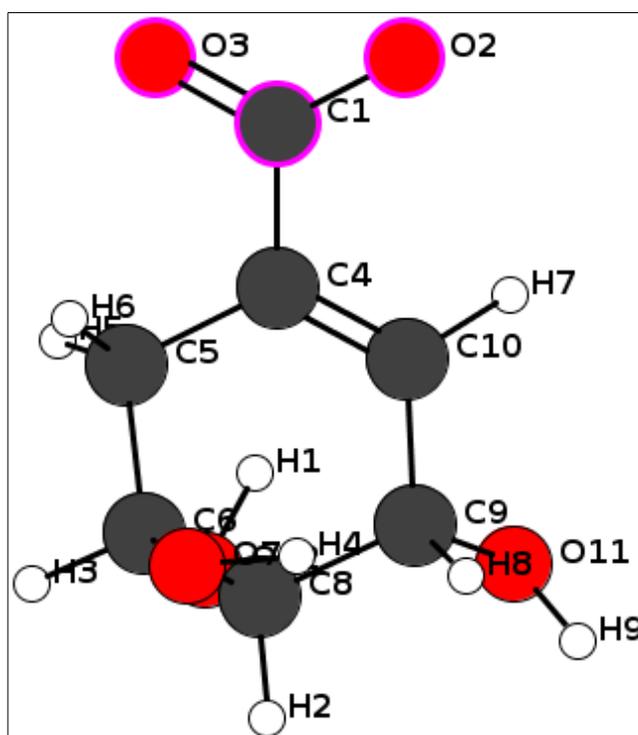


Figura 76 – Molécula chiquimato usada como protótipo para substituições bioisostérica. Selecionados estão os átomos do grupamento carboxilato (O2, O3 e C1), os quais foram substituídos nos bioisósteros gerados.

Fonte: Do autor.

Tabela 19 – Cálculo das interações não ligante entre os átomos do chiquimato e o sítio ativo da enzima chiquimato quinase, obtidos com a estrutura pdb 2iyx.

Átomo	Energia de Van der Waals (kcal/mol)	Energia Eletrostática (kcal/mol)	Energia Total (kcal/mol)
O3	1,800	309,473	311,272
O2	0,562	257,968	258,530
O12	0,405	255,918	256,323
O11	1,457	235,480	236,937
O7	-0,674	190,876	190,201
C4	-0,911	148,083	147,173
C10	-0,750	103,295	102,545
H8	-0,109	0,000	-0,109
H3	-0,127	0,000	-0,127
H2	-0,189	0,000	-0,189
H6	-0,251	0,000	-0,251
H5	-0,294	0,000	-0,294
C5	-0,929	-51,918	-52,847
H7	-0,228	-53,052	-53,280
C6	-0,853	-91,406	-92,259
C8	-0,687	-94,327	-95,014
H4	-0,083	-112,348	-112,432
H9	-0,085	-130,434	-130,519
C9	-0,548	-140,799	-141,348
H1	-0,148	-158,542	-158,690
C1	-0,861	-648,400	-649,261

Fonte: Do autor.

Foram construídos 13 bioisósteros, os quais foram submetidos aos critérios de filtragem de acordo com suas propriedades *drug-like* (**Bioisosteres** → **Filter bioisosteres**). Ao se fazer a filtragem sem nenhum relaxamento, apenas 3 moléculas foram aprovadas, para se aumentar o número de moléculas aceitas, permitiu-se a aprovação para moléculas que falhassem em no máximo uma das propriedades, sendo recuperadas 12 estruturas. Em seguida, rodou-se o *virtual screening* tendo-se como receptor a estrutura pdb 2iyx, cujos resultados se encontram na tabela 20.

Tabela 20 – *Virtual screening* entre os bioisósteros do chiquimato e a estrutura 2iyx.

(continua)

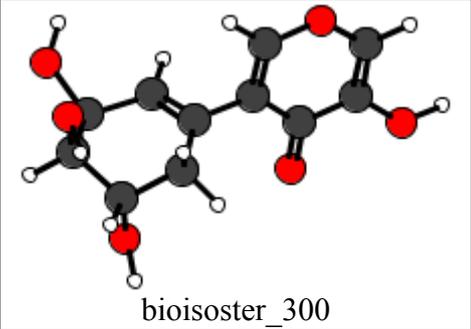
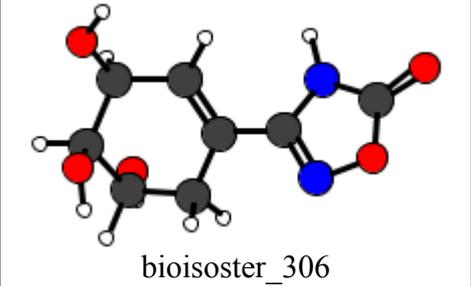
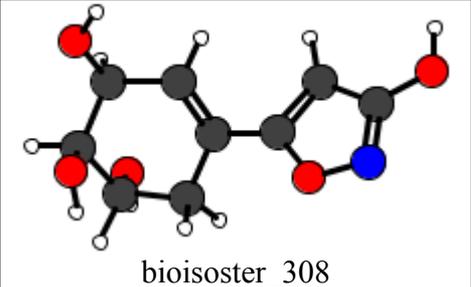
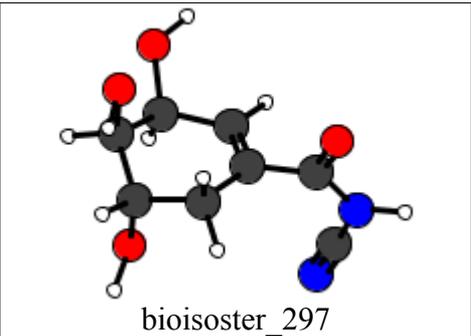
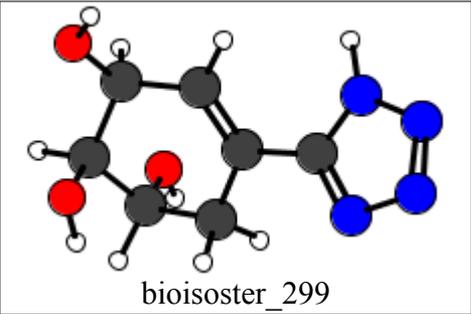
Bioisótero	DS	PM	logP	logS	TPSA	HD	HA
 <p>bioisoster_300</p>	-7,200	240,200	-0,601	-2,800	107,220	4	6
 <p>bioisoster_306</p>	-6,700	214,175	-0,972	-2,766	111,380	4	7
 <p>bioisoster_308</p>	-6,700	213,183	-1,110	-0,306	106,950	4	6
 <p>bioisoster_297</p>	-6,600	198,176	-0,801	-1,523	113,580	4	6
 <p>bioisoster_299</p>	-6,600	198,187	-0,596	-0,756	115,150	3	7

Tabela 20 – *Virtual screening* entre os bioisómeros do chiquimato e a estrutura 2iyx.

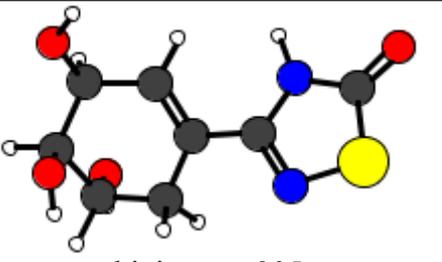
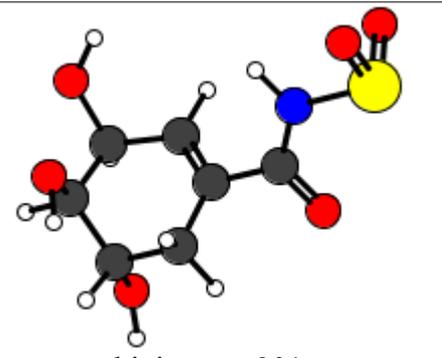
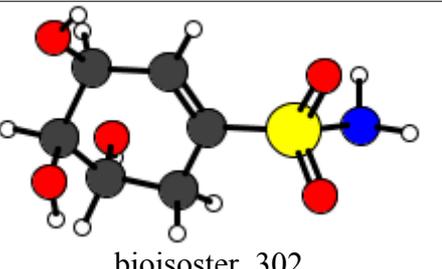
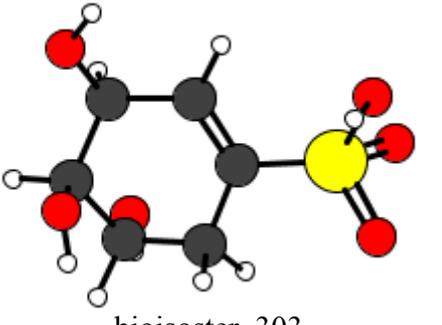
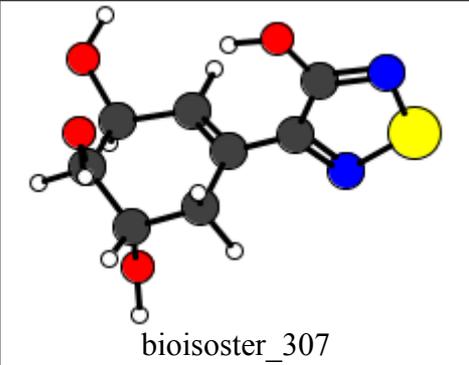
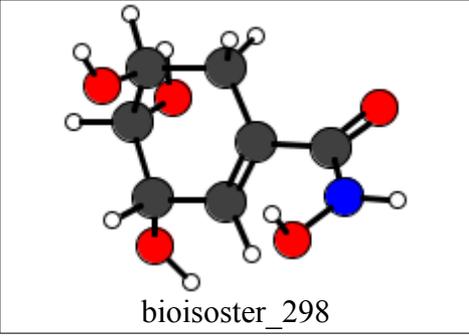
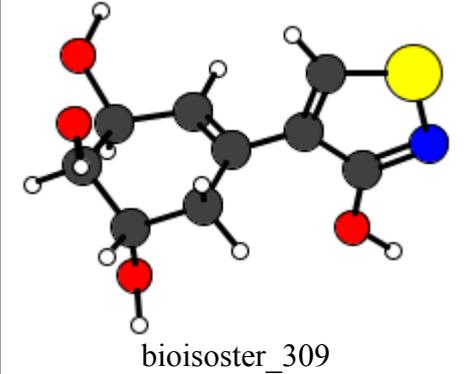
(continuação)							
Bioisótero	DS	PM	logP	logS	TPSA	HD	HA
 <p>bioisoster_305</p>	-6,600	230,246	0,386	-3,551	127,450	4	7
 <p>bioisoster_301</p>	-6,400	236,224	-1,917	-1,979	123,930	4	8
 <p>bioisoster_302</p>	-6,400	209,223	-1,280	-0,447	129,230	3	6
 <p>bioisoster_303</p>	-6,400	210,204	-0,954	-0,192	123,440	4	6

Tabela 20 – *Virtual screening* entre os bioisósteros do chiquimato e a estrutura 2iyx.

Bioisótero	DS	PM	logP	logS	TPSA	(conclusão)	
						HD	HA
 <p>bioisoster_307</p>	-6,400	230,246	1,408	-0,654	134,940	4	7
 <p>bioisoster_298</p>	-6,300	189,163	-2,196	-1,230	110,020	5	6
 <p>bioisoster_309</p>	-6,300	229,254	1,197	-0,859	122,050	4	6

Fonte: Do autor.

Legenda: DS: docking score (kcal/mol).

PM: peso molecular (Da).

HD: número de doadores de interação de hidrogênio.

HA: número de aceptores de interação de hidrogênio.

O bioisótero com melhor *docking score* foi aquele no qual o carboxilato foi substituído por uma hidroxí-cromona (DS = -7,2 kcal/mol), embora apresente logP fora do intervalo ótimo (logP = -0,601; intervalo ótimo: $0 \leq \log P \leq 3$). Todas as nove moléculas que falharam em uma propriedade *drug-like* tiveram logP abaixo de zero. A fim de se realizar uma comparação, as propriedades do composto *hit* chiquimato (FIGURA 76) foram computadas, assim como seu *docking score* por meio de um *redocking* (*molecular docking* no qual o

ligante utilizado é exatamente aquele cocristalizado com o receptor). Os valores obtidos foram os seguintes:

- a) *Docking score*: -6,2 kcal/mol;
- b) Rmsd do *redocking*: 1,023 Å;
- c) Peso molecular: 173,137 Da;
- d) logP: -0,501;
- e) logS: -1,883;
- f) TPSA: 100,820 Å²;
- g) N° doadores de interação de hidrogênio: 3;
- h) N° aceptores de interação de hidrogênio: 5;

O rmsd é definido como a distância média entre os átomos de duas moléculas sobrepostas, excluindo-se hidrogênios, obtido com a seguinte equação:

$$RMSD(\vec{v}, \vec{w}) = \sqrt{\frac{1}{n} \sum_{i=1}^n ((v_{ix} - w_{ix})^2 + (v_{iy} - w_{iy})^2 + (v_{iz} - w_{iz})^2)}$$

em que v e w são vetores posição contendo as coordenadas cartesianas das duas moléculas comparadas. Para que o *redocking* seja validado, o rmsd deve estar abaixo de 2 Å.

Nota-se que o logP do composto *hit* também está abaixo de zero, e que todos os bioisómeros apresentaram *docking score* equivalente ou superior ao *hit*. Dessa forma, fazendo-se um balanço entre afinidade de interação com o receptor e propriedades *drug-like*, os compostos cujo logP está acima de zero podem ser indicados como potenciais estruturas para síntese e análises *in vitro* e *in vivo*, sendo estes as moléculas bioisoster_305, bioisoster_307 e bioisoster_309.

Também se construíram moléculas análogas por substituição de átomos de hidrogênio (**Molecule** → **Build over all H**), em um total de 280 estruturas. Realizando-se a filtragem de acordo com as propriedades *drug-like*, sem nenhum relaxamento, foram aprovadas 109 moléculas, as quais foram submetidas a *virtual screening* conforme descrito anteriormente. Na tabela 21 estão os dez ligantes com melhor *docking score*. Como pode ser observado, todos os ligantes apresentados possuem como substituinte um grupo aromático, podendo ser fenil, fenol, ou anilina.

Tabela 21 – *Virtual screening* entre os análogos por substituição de hidrogênio do chiquimato e a estrutura 2iyx.

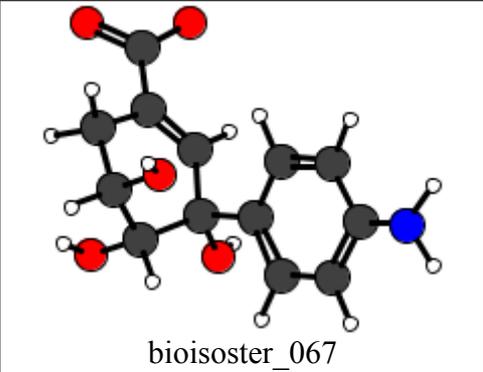
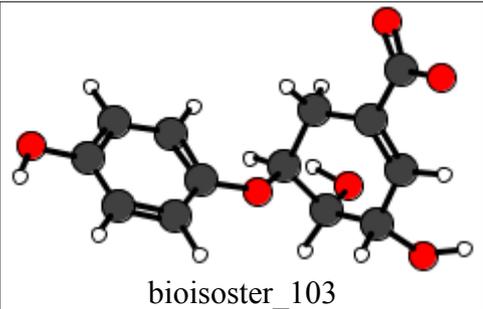
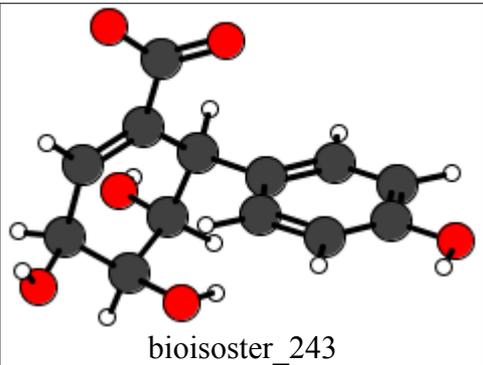
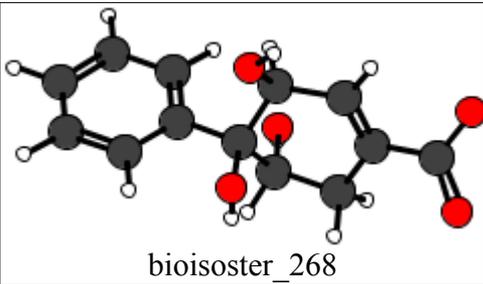
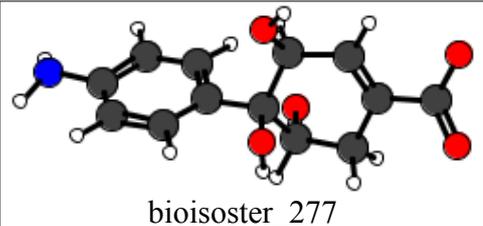
(continua)							
Bioisótero	DS	PM	logP	logS	TPSA	HD	HA
 <p>bioisoster_067</p>	-7,600	264,247	0,401	-3,817	126,840	5	6
 <p>bioisoster_103</p>	-7,600	265,228	0,227	-3,922	110,050	3	6
 <p>bioisoster_243</p>	-7,600	265,228	0,716	-3,118	121,050	4	6
 <p>bioisoster_268</p>	-7,600	249,229	0,863	-3,947	100,820	3	5
 <p>bioisoster_277</p>	-7,600	264,247	0,173	-3,817	126,840	5	6

Tabela 21 – *Virtual screening* entre os análogos por substituição de hidrogênio do chiquimato e a estrutura 2iyx.

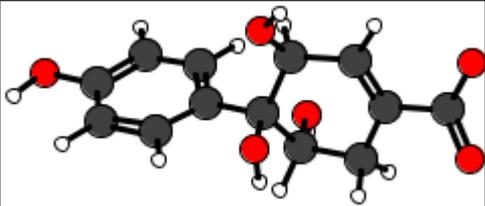
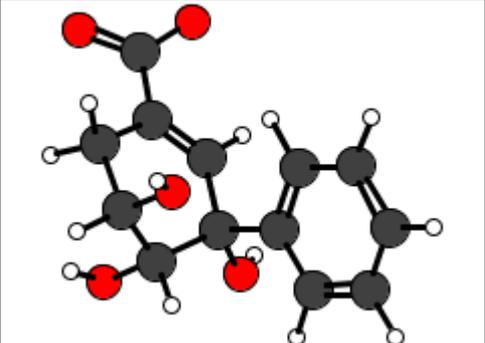
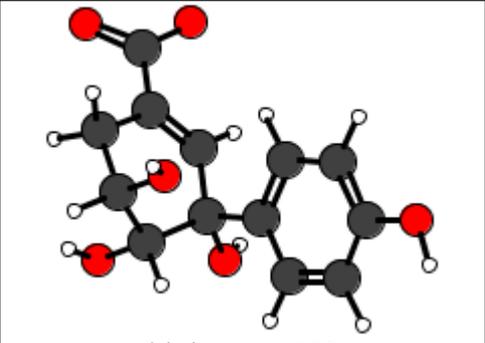
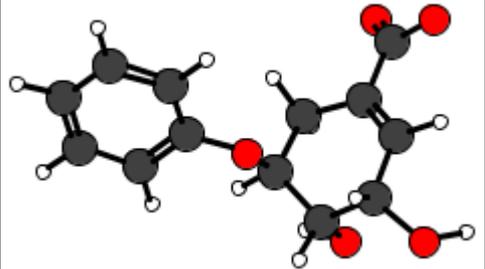
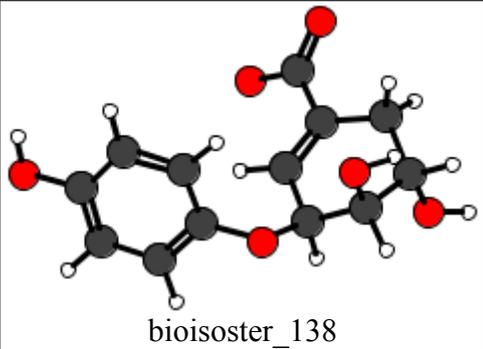
(continuação)							
Bioisótero	DS	PM	logP	logS	TPSA	HD	HA
 <p>bioisoster_278</p>	-7,600	265,228	0,498	-3,562	121,050	4	6
 <p>bioisoster_058</p>	-7,500	249,229	1,092	-3,947	100,820	3	5
 <p>bioisoster_068</p>	-7,500	265,228	0,727	-3,562	121,050	4	6
 <p>bioisoster_093</p>	-7,500	249,229	0,592	-4,307	89,820	2	5

Tabela 21 – *Virtual screening* entre os análogos por substituição de hidrogênio do chiquimato e a estrutura 2iyx.

Bioisótero	DS	PM	logP	logS	TPSA	(conclusão)	
						HD	HA
 <p style="text-align: center;">bioisoster_138</p>	-7,500	265,228	0,227	-3,922	110,050	3	6

Fonte: Do autor.

Legenda: DS: docking score (kcal/mol).

PM: peso molecular (Da).

HD: número de doadores de interação de hidrogênio.

HA: número de aceptores de interação de hidrogênio.

Com o intuito de se identificar os análogos cujo modo de interação com o receptor seja similar ao composto *hit*, calculou-se o rmsd entre a conformação de *docking* dos análogos e a molécula chiquimato cocristalizada na estrutura do receptor (TABELA 22). No cálculo do rmsd, os átomos dos substituintes não foram considerados.

Tabela 22 – Rmsd entre o composto chiquimato presente na estrutura pdb 2iyx e a conformação de *docking* de seus análogos.

Composto	Rmsd (Å)
bioisoster_067	2,107
bioisoster_103	3,384
bioisoster_243	3,563
bioisoster_268	1,295
bioisoster_277	1,312
bioisoster_278	1,315
bioisoster_058	2,246
bioisoster_068	2,085
bioisoster_093	3,304
bioisoster_138	1,351

Fonte: Do autor.

As estruturas bioisoster_268, bioisoster_277, bioisoster_278 e bioisoster_138 apresentaram rmsd inferior a 2 Å, as quais são exibidas na figura 77.

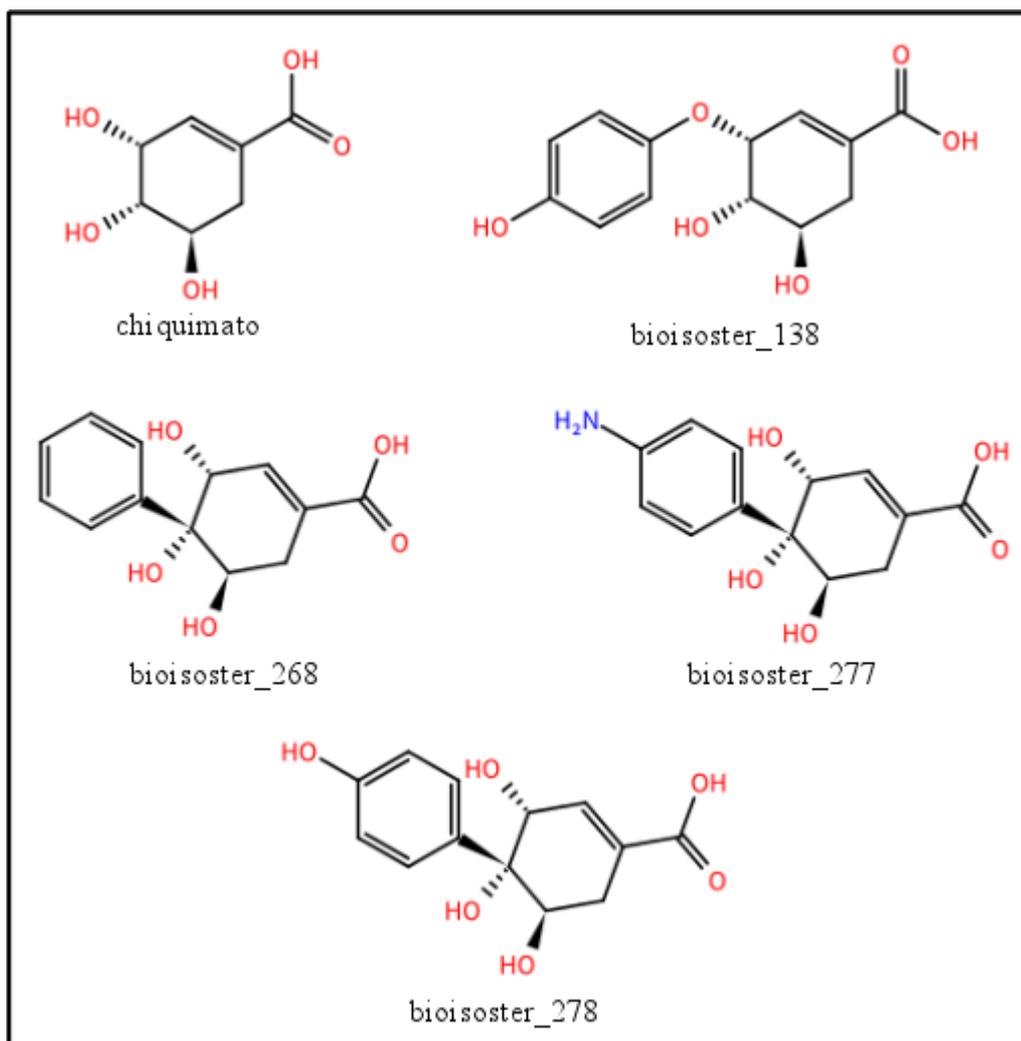


Figura 77 – Análogos por adição de substituintes do composto chiquimato. As estruturas apresentadas possuem conformação de docking com rmsd inferior a 2 Å em relação ao composto protótipo.

Fonte: Do autor.

Dos quatro análogos apresentados, três apresentam substituinte aromático na posição *para* em relação ao grupo carboxilato, enquanto o último apresenta o substituinte ligado ao oxigênio na posição *orto* em relação à ligação dupla do anel. Computou-se o gráfico 2D da interação entre os ligantes e o receptor por meio do programa LigPlot (WALLACE, LASKOWSKI, E THORNTON, 1996) (FIGURA 78). O chiquimato apresenta interações de hidrogênio entre seu grupo carboxilato e os resíduos ARG58 e ARG136, além de outra interação de hidrogênio entre dois de seus grupos hidroxilas e o resíduo ASP34.

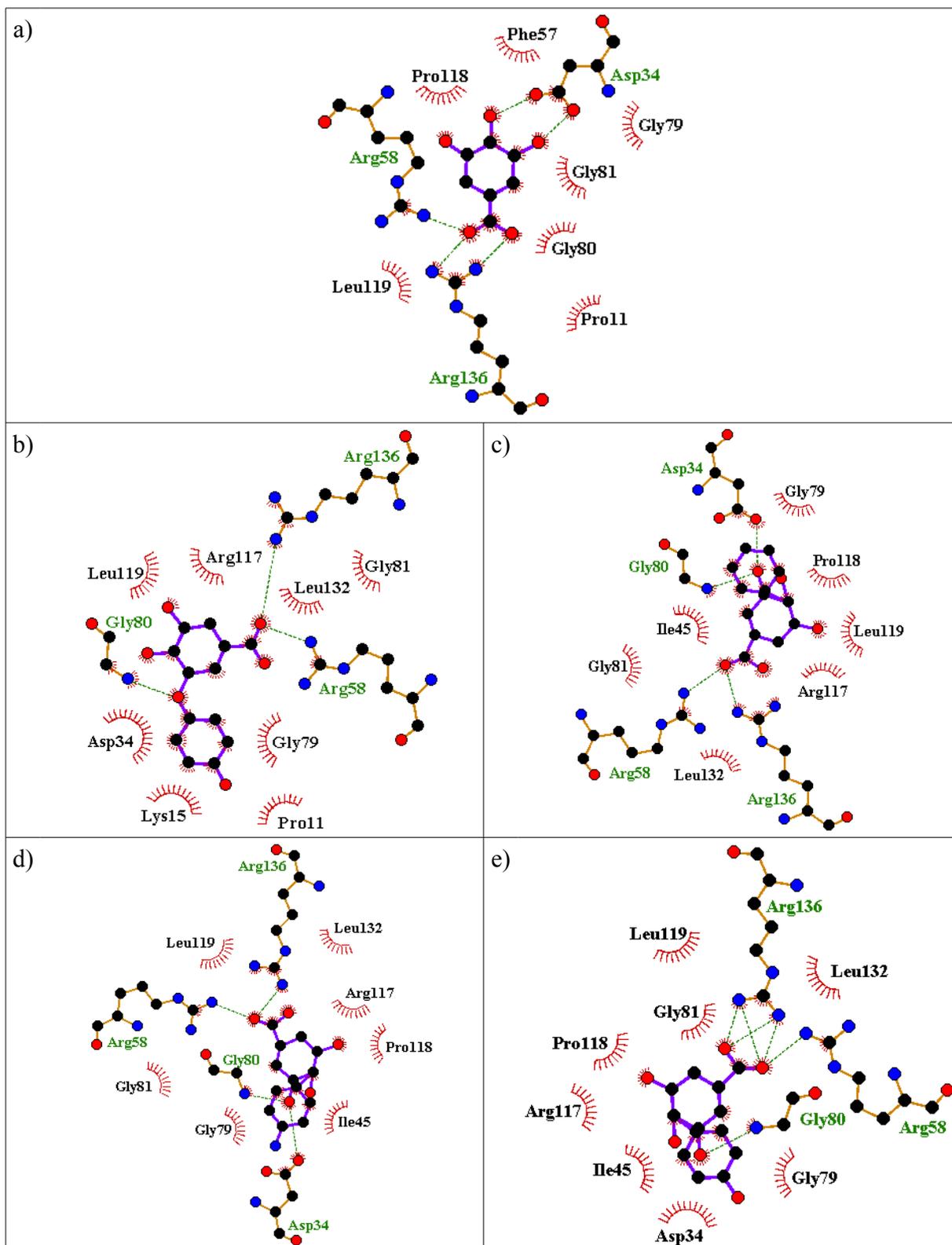


Figura 78 – Gráfico 2D da interação entre o chiquimato e análogos com a enzima chiquimato quinase.

Fonte: Do autor.

- Legenda: a) chiquimato;
 b) bioisoster_138;
 c) bioisoster_268;
 d) bioisoster_277;
 e) bioisoster_278.

Os análogos também apresentam interações de hidrogênio entre o grupo carboxilato e os resíduos ARG58 e ARG136, embora apenas os análogos bioisoster_268 e bioisoster_277 mantenham ainda interações de hidrogênio com o resíduo ASP34.

4. 4 COMPARAÇÃO COM SOFTWARES SIMILARES

Realizou-se uma busca a fim de se localizar softwares que fossem similares ao programa MB-Isoster, que tenham como propósito transformações em estruturas moleculares por meio do bioisosterismo. Foram identificados a base de dados SwissBioisostere (WIRTH *et al.*, 2013) e o servidor web BoBER (LESNIK *et al.*, 2017).

SwissBioisostere é uma base contendo dados sobre 5 586 462 diferentes tipos de transformações bioisostéricas disponível no endereço <http://www.swissbioisostere.ch>. Foi construída utilizando métodos MMP (*Matched Molecular Pairs*, Pares de Emparelhamento Molecular), no qual se buscam moléculas quase idênticas, exceto por uma pequena região, correlacionando suas estruturas com propriedades físico-químicas e farmacológicas. A base de dados ChEMBL (GAULTON *et al.*, 2011) foi usada como base primária para extração e análise dos dados, selecionando-se moléculas para as quais dados de atividade biológica, como IC_{50} , EC_{50} , K_i e K_d estivessem anotados. Para utilizar a base, o usuário deve desenhar o fragmento que deseja pesquisar, a base retorna os fragmentos bioisostéricos encontrados e uma relação percentual de quantas vezes determinada substituição bioisostérica resultou em em melhora ou piora da atividade biológica, ou resultou em um composto equivalente. Comparada ao programa MB-Isoster, a base de dados SwissBioisostere tem a vantagem de possuir uma gigantesca biblioteca de relações bioisostéricas, com mais de cinco milhões de pares de transformações, no entanto, ela trabalha apenas com o conceito de fragmento, não retornando estruturas completas; caso o usuário queira trabalhar com a estrutura tridimensional das moléculas, terá que desenhá-las manualmente em algum programa editor de moléculas. O programa MB-Isoster permite a construção automática de todas as moléculas para as quais um fragmento bioisostérico for encontrado. A biblioteca de fragmentos do programa MB-Isoster foi construída manualmente, resultando assim seu tamanho reduzido; no entanto, em um trabalho posterior será possível criar um algoritmo similar ao desenvolvido na construção da base SwissBioisoster a fim de se ampliar a biblioteca de fragmentos.

BoBER é um servidor web que implementa uma interface para substituições bioisostéricas, estando disponível em <http://bober.insilab.org>. Na construção dessa base, utilizou-se o PDB como base primária. Realizou-se um processo de sobreposição de proteínas pertencentes à mesma família e que possuíssem ligantes cocrystalizados; os ligantes sobrepostos foram tridimensionalmente comparados a fim de buscar correlações bioisostéricas. Na utilização do servidor, o usuário desenha a molécula completa para a qual deseja análogos bioisostéricos; o servidor fragmenta a molécula, podendo o usuário selecionar qual fragmento usar como ponto de substituição. Após a busca, retornam-se os fragmentos encontrados, o usuário escolhe o fragmento que pretende utilizar, a molécula é construída, podendo-se fazer o download da mesma. Como a biblioteca de fragmentos do servidor BoBER foi construída por algoritmos, ela também é maior quando comparada à biblioteca do programa MB-Isoster, embora não seja tão extensa quanto à base SwissBioisostere. Uma vez que o usuário desenha a molécula de entrada inteira e é o servidor o responsável por selecionar quais fragmentos são passíveis de transformação, pode acontecer que o fragmento que o usuário deseja substituir não esteja disponível, enquanto que no programa MB-Isoster o usuário é o total responsável pela seleção dos átomos a serem substituídos. Embora o servidor BoBER construa as moléculas resultantes das transformações bioisostéricas e permita que o usuário faça o download das mesmas, é necessário que os fragmentos sejam selecionados um por um, e as moléculas são construídas uma por uma, resultando em trabalho exaustivo caso o usuário queira fazer o download de todos os bioisósteros gerados.

Outro aspecto positivo do programa MB-Isoster em relação aos dois mencionados anteriormente, é a implementação de funções adicionais, as quais são a construção de análogos por adição de substituintes, a análise de interações receptor-ligante para identificação de pontos de substituição bioisostérica, o cálculo de propriedades *drug-like*, e a extensão para execução de *virtual screening* com o programa AutoDock Vina.

5 CONCLUSÃO

O bioisosterismo tem sido amplamente empregado em pesquisas de desenho de fármacos como forma de melhorar a atividade de compostos líderes. A substituição de um fragmento molecular por outro com propriedades físico-químicas similares promovem a modulação das propriedades farmacocinéticas e farmacodinâmicas. No presente trabalho apresentou-se o programa MB-Isoster, desenvolvido para construir moléculas análogas baseado nos conceitos do bioisosterismo. Foram incorporados procedimentos para criar substituições a partir de átomos selecionados pelo usuário; para criar substituições em todos os átomos de hidrogênio; para computar interações receptor-ligante a fim de se identificar átomos em potencial para substituição; computar propriedades *drug-like*; e executar *virtual screening*.

Sendo um software em desenvolvimento, existem possibilidades de aumentar as potencialidades do mesmo, dentre as quais:

- a) Ampliar a biblioteca de fragmentos;
- b) Implementar outros métodos de cálculo de propriedades *drug-like* e de predição de toxicidade;
- c) Implementar métodos de análise de interação receptor-ligante para os resultados obtidos pelo *virtual screening*;
- d) Incluir mecanismos de aprendizado de máquina para que o software possa sugerir alguma configuração a partir da experiência de uso;
- e) Incluir outros programas de *docking* para o *virtual screening*;
- f) Incorporar o conceito de espaço de funções escores.

Espera-se que o software possa contribuir nos estudos de planejamento racional de novos fármacos.

REFERÊNCIAS

- ACCELRY'S. **CTfile Formats**, 2011. San Diego: Accelrys Scientific and Technical Support, 2011.
- AMMERAAL, L.; ZHANG, K. **Computação Gráfica para Programadores Java**. Tradução de Acauan Fernandes. 2. ed. Rio de Janeiro: LTC, 2008.
- ANGELI, A. et al. Discovery of New Selenoureido Analogues of 4-(4-Fluorophenylureido) benzenesulfonamide as Carbonic Anhydrase Inhibitors. **ACS Medicinal Chemical Letters**, Washington, v. 8, n. 9, p. 963-968, 2017.
- BARREIRO, E. J.; FRAGA, C. A. M. **Química Medicinal: As bases moleculares da ação dos fármacos**. 2. ed. Porto Alegre: Artmed, 2008.
- BINDER, D. et al. Analogs and derivatives of tenoxicam. 1. Synthesis and antiinflammatory activities of analogs with different residues on the ring nitrogen and the amide nitrogen. **Journal of Medicinal Chemistry**, Nova Iorque, v. 30, n. 4, p. 678-682, 1987.
- BOLCHI, C. et al. Novel 5-substituted 3-hydroxyphenyl and 3-nitrophenyl ethers of S-propranolol as $\alpha 4\beta 2$ -nicotinic acetylcholine receptor ligands. **Bioorganic & Medicinal Chemistry Letters**, Oxford, v. 26, n. 23, p. 5613-5617, 2016.
- BURGER, A. Isosterism and Bioisosterism in Drug Design. **Progress in Drug Research**, Basel, v. 37, p. 287-371, 1991.
- CAPDEVILLE, R. et al. Glivec (STI571, imatinib), a rationally developed, targeted anticancer drug. **Nature Reviews Drug Discovery**, Londres, v. 1, n. 7, p. 493-502, 2002.
- CAPPARELLI, E. et al. SAR Studies of Tetrahydroisoquinoline Derivatives: The Role of Flexibility and Bioisosterism To Raise Potency and Selectivity toward P-glycoprotein. **Journal of Medicinal Chemistry**, Nova Iorque, v. 57, n. 23, p. 9983-9994, 2014.
- CASE, D. A. et al. The Amber biomolecular simulation programs. **Journal of Computational Chemistry**, Nova Iorque, v. 26, n. 16, p. 1668-1688, 2005.

CHENG, T. et al. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge. **Journal of Chemical Information and Modeling**, Washington, v. 47, n. 6, p. 2140-2148, 2007.

CORACINI, J. D.; AZEVEDO JR., W. F. Shikimate Kinase, a Protein Target for Drug Design. **Current Medicinal Chemistry**, Sharjah, v.21, n. 5, p.592-604, 2014.

CRAMER, C. J. **Essentials of Computational Chemistry: Theories and Models**, 2. ed.; New Jersey: John Wiley & Sons, 2013.

DIAS, M. V. B. et al. Effects of the magnesium and chloride ions and shikimate on the structure of shikimate kinase from *Mycobacterium tuberculosis*. **Acta Crystallographica Section F**, Oxford, v. 63, n. 1, p. 1-6, 2007.

DIXON, S. L. et al. PHASE: a new engine for pharmacophore perception, 3D QSAR model development, and 3D database screening: 1. Methodology and preliminary results. **Journal of Computer-Aided Molecular Design**, Amsterdam, v. 20, n. 10, p. 647-671, 2006.

DUSCHINSKY, R.; PLEVEN, E.; HEIDELBERGER, C. The synthesis of 5-fluoropyrimidines. **Journal of the American Chemical Society**, Washington, v. 79, n. 16, p. 4559-4560, 1957.

ERTL, P.; ROHDE, B.; SELZER, P. Fast Calculation of Molecular Polar Surface Area as a Sum of Fragment-Based Contributions and Its Application to the Prediction of Drug Transport Properties. **Journal of Medicinal Chemistry**, Nova Iorque, v. 43, n. 20, p. 3714-3717, 2000.

FRIEDMAN, H. L. Influence of Isosteric Replacements upon Biological Activity. **NASNRS**, Washington, v. 206, p. 295-358, 1951.

_____. Influence of isosteric replacements upon biological activity. Symposium on chemical-biological correlation. **Natural Research Council Publication n° 206**, Washington, DC. p. 295-395, 1951.

GAULTON, A. et al. ChEMBL: a large-scale bioactivity database for drug discovery. **Nucleic Acids Research**, Londres, v. 40, p. D1100-D1107, 2011.

GRANGE, J. M. The genus *Mycobacterium* and the *Mycobacterium tuberculosis* complex. In: SCHAAF, H. S.; ZUMLA, A. I. (editores). **Tuberculosis: a comprehensive clinical reference**. Philadelphia, PA: Saunders, 2009. cap. 6, p. 44-59.

GRIMM, H. G. Structure and size of the non-metallic hydrides. **Z. Electrochem**, Weinheim, v. 31, n. 9, p. 474-480, 1925.

GROOM, C. R. et al. The Cambridge Structural Database. **Acta Crystallographica Section B**, Oxford, v.72, n. Pt 2, p. 171-179, 2016.

HALGREN, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. **Journal of Computational Chemistry**, Nova Iorque, v. 17, n. 5, p. 490-519, 1996.

_____. MMFF VI. MMFF94s Option for Energy Minimization Studies. **Journal of Computational Chemistry**, Nova Iorque, v. 20, n. 7, p. 720-729, 1999.

HARTMANN, M. D. et al. Mechanism of Phosphoryl Transfer Catalyzed by Shikimate Kinase from *Mycobacterium tuberculosis*. **Journal of Molecular Biology**, Londres, v. 364, n. 3, p. 411-423, 2006.

HEHRE, W. J. **A Guide to Molecular Mechanics and Quantum Chemical Calculations**; Irvine: Wavefunction, 2003.

HEIDELBERGER, C. et al. Fluorinated Pyrimidines, a New Class of Tumor-Inhibitory Compounds. **Nature**, Londres, v. 179, p. 663-666, 1957.

HEINSTEIN, P. F.; CHANG, C. J. Taxol. **Annual Review of Plant Biology**, Palo Alto, v. 45, n. 1, p. 663-674, 1994.

HERRMANN, K. M.; WEAVER, L. M. The Shikimate Pathway. **Annual Review of Plant Physiology and Plant Molecular Biology**, Palo Alto, v. 50 ,p. 473-503, 1999.

HESS, B. et al. GROMACS 4: Algorithms for highly efficient, load-balanced and scalable molecular simulation. **Journal of Chemical Theory and Computation**, Washington, v. 4, n. 3, p. 435-447, 2008.

HOLLOWAY, M. K. et al. *A Priori* Prediction of Activity for HIV-1 Protease Inhibitors Employing Energy Minimization in the Active Site. **Journal of Medicinal Chemistry**, Nova Iorque, v. 38, n. 2, p. 305-317, 1995.

HORIEN, C.; YUAN, P. Drug Development. **Yale Journal of Biology and Medicine**, New Haven, v. 90, n. 1, p. 1-3, 2017.

HOU, T. J. et al. ADME Evaluation in Drug Discovery. 4. Prediction of Aqueous Solubility Based on Atom Contribution Approach. **Journal of Chemical Information and Computer Sciences**, Washington, v. 44, n. 1, p. 266-275, 2004.

IMMING, P. Medicinal Chemistry: Definitions and Objectives, Drug Activity Phases, Drug Classification Systems. In: Wermuth, C. G.; Aldous, D.; Raboisson, P.; Rognan, D. (editor) **The Practice of Medicinal Chemistry**. 4. ed. Londres: Elsevier, 2015. cap. 1, p. 3-13.

JEFFREY, G. A. **An introduction to hydrogen bonding**. Oxford: Oxford University Press, 1997.

JENSEN, F. **Introduction to Computational Chemistry**. 2. ed. Hoboken, New Jersey: John Wiley & Sons, 2007.

KAPETANOVIC, I. M. COMPUTER-AIDED DRUG DISCOVERY AND DEVELOPMENT (CADD): In silico-chemico-biological approach. **Chemico-Biological Interactions**, Limerick, v. 171, n.2, p. 165-176, 2008.

KERNS, E. H.; DI, L. **Drug-like Properties: Concepts, Structure Design and Methods: from ADME to Toxicity Optimization**. San Diego: Academic Press, 2008.

KHAN, M. O. F.; DEIMLING, M. J.; PHILIP, A. Medicinal Chemistry and the Pharmacy Curriculum. **American Journal of Pharmaceutical Education**, Arlington, v. 75, n. 8, p. 1-10, 2011.

KIM, K. S. et al. Discovery of aminothiazole inhibitors of cyclin-dependent kinase 2: synthesis, X-ray crystallographic analysis, and biological activities. **Journal of Medicinal Chemistry**, Nova Iorque, v. 45, n. 18, p. 3905-3927, 2002.

KIRK, K. L. Selective Fluorination in Drug Design and Development: An Overview of Biochemical Rationales. **Current Topics in Medicinal Chemistry**, Sharjah, v. 6, n.14, p. 1447-1456, 2006.

KOCH, A.; Mizrahi, V. *Mycobacterium tuberculosis*. **Trends in Microbiology**, Cambridge, v. 26, n. 6, p. 555-556, 2018.

KOTOKU, N. et al. Short-step synthesis and structure-activity relationship of cortistatin A analogs. **Tetrahedron**, Oxford, v. 73, n. 10, p. 1342-1349, 2017.

KRAUTWALD, S. et al. Bioisosteric Exchange of C_{sp3}-Chloro and Methyl Substituents: Synthesis and Initial Biological Studies of Atpenin A5 Analogues. **Angewandte Chemie - International Edition**, Weinheim, v. 55, n. 12, p. 4049-4053, 2016.

LABUTE, P. et al. Flexible alignment of small molecules. **Journal of Medicinal Chemistry**, Nova Iorque, v. 44, n. 10, p. 1483-1490, 2001.

LANGMUIR, I. Isomorphism, isosterism and covalence. **Journal of the American Chemical Society**, Washington, v. 41, n. 10, p. 1543-1559, 1919.

LARSEN, A. A.; LISH, P. M. A new Bio-isostere: alkylsulphonamidophenethanolamines. **Nature**, Londres, v. 203, n. 4851, p. 1283-1284, 1964.

LAWN, S. D.; ZUMLA, A. I. Tuberculosis. **The Lancet**, Nova Iorque, v. 378, n. 9785, p. 57-72, 2011.

LEI, M. et al. Synthesis and cytotoxicity evaluation of 4-amino-4-dehydroxylarctigenin derivatives in glucose-starved A549 tumor cells. **Bioorganic & Medicinal Chemistry Letters**, Oxford, v. 25, n. 3, p. 435-437, 2015.

LESK, A. M. **Introdução à bioinformática**. 2.ed. Porto Alegre: Artmed, 2008.

LESNIK, S. et al. BoBER: web interface to the base of bioisosterically exchangeable replacements. **Journal of Cheminformatics**, Londres, v. 9, n. 62, p. 1-8, 2017.

LIMA, L. M.; BARREIRO, E. J. Bioisosterism: A Useful Strategy for Molecular Modification and Drug Design. **Current Medicinal Chemistry**, Sharjah, v. 12, n. 1, p. 23-49, 2005.

LIN, P. et al. 2-(3,5-Dimethylphenyl)tryptamine derivatives that bind to the GnRH receptor. **Bioorganic & Medicinal Chemistry Letters**, Oxford, v. 11, n. 8, p. 1073-1076, 2001.

LIPINSKI, C. A. Drug-like properties and the causes of poor solubility and poor permeability. **Journal of Pharmacological and Toxicological Methods**, Nova Iorque, v. 44, n. 1, p. 235-249, 2000.

LIPINSKI, C. A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, Utah, v. 23, n. 1, p. 3-25, 1997.

LIU, Q. et al. Structure-based bioisosterism design, synthesis, insecticidal activity and structure-activity relationship (SAR) of anthranilic diamide analogues containing 1,2,4-oxadiazole rings. **Pest Management Science**, West Sussex, v. 73, n. 5, p. 917-924, 2017.

LIU, R.; LI, X.; LAM, K. S. Combinatorial chemistry in drug discovery. **Current Opinion in Chemical Biology**, Londres, v. 38, p. 117-126, 2017.

LOFTSSON, T. **Drug Stability for Pharmaceutical Scientists**. Oxford: Academic Press, 2014.

LOMBARDINO, J. G.; LOWE, J. A. The role of the Medicinal Chemistry in drug discovery – then and now. **Nature Reviews Drug Discovery**, Londres, v. 3, n. 10, p. 853-862, 2004.

LU, X. et al. Arylazolyl(azinyl)thioacetanilides. Part 20: Discovery of novel purinylthioacetanilides derivatives as potent HIV-1 NNRTIs via a structure-based bioisosterism approach. **Bioorganic & Medicinal Chemistry**, Oxford, v. 24, n. 18, p. 4424-4433, 2016.

LUTY, B.; ROSE, P. W. The need for scientific software engineering in the pharmaceutical industry. **Journal of Computer-Aided Molecular Design**, Amsterdam, v. 31, n. 3, p. 301-304, 2017.

MCKINNEY, J. D. et al. The Practice of Structure Activity Relationships (SAR) in Toxicology. **Toxicological Sciences**, Oxford, v. 56, n. 1, p. 8-17, 2000.

MENG, W. D.; QING, F. L. Fluorinated Nucleosides as Antiviral and Antitumor Agents. **Current Topics in Medicinal Chemistry**, Sharjah, v. 6, n. 14, p. 1499-1528, 2006.

MENG, X. Y. et al. Molecular Docking: A powerful approach for structure-based drug discovery. **Current Computer-Aided Drug Design**, Sharjah, v. 7, n. 2, p. 146-157, 2011.

MOHAMADI, F. et al. MacroModel – An Integrated Software System for Modeling Organic and Bioorganic Molecules Using Molecular Mechanics. **Journal of Computational Chemistry**, Nova Iorque, v. 11, n. 4, p. 440-467, 1990.

MORRIS, G. M. et al. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. **Journal of Computational Chemistry**, Nova Iorque, v. 30, n. 16, p. 2785–2791, 2009.

MUGNAINI, C.; PASQUINI, S.; CORELLI, F. The Bioisosteric Concept Applied to Cannabinoid Ligands. **Current Medicinal Chemistry**, Sharjah, v. 19, n. 28, p. 4794-4815, 2012.

MUKESH, B.; RAKESH, K. Molecular Docking: A Review. **International Journal of Research in Ayurveda & Pharmacy**, Rampur, v. 2, n. 6, p. 1746-1751, 2011.

MYERS, S.; BAKER, A. Drug discovery – an operating model for a new era. **Nature Biotechnology**, Nova Iorque, v. 19, n. 8, p. 727-730, 2001.

NETTO, O. B.; JURKIEWICZ, S. **Grafos: Introdução e Prática**. São Paulo: Editora Blücher, 2009.

NEWMAN, D. J.; CRAGG, G. M. Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. **Journal of Natural Products**, Gorakhpur, v. 73, n. 3, p. 311-335, 2012.

O'BOYLE, N. M. et al. Open Babel: An open chemical toolbox. **Journal of Cheminformatics**, Londres, v. 3, n. 1, p. 1-14, 2011.

OLIVEIRA FILHO, G. B. et al. Structural design, synthesis and pharmacological evaluation of 4-thiazolidinones against *Trypanosoma cruzi*. **Bioorganic & Medicinal Chemistry**, Oxford, v. 23, n. 23, p. 7478-7486, 2015.

PANDEYA, S. N.; THAKKAR, D. Combinatorial chemistry: A novel method in drug discovery and its application. **Indian Journal of Chemistry**, Nova Deli, v. 44B, p. 335-348, 2005.

PATANI, G. A.; LAVOIE, E. J. Bioisosterism: A Rational Approach in Drug Design. **Chemical Reviews**, Washington, v. 96, n. 8, p. 3147-3176, 1996.

PEREIRA, J. H. et al. Structure of shikimate kinase from *Mycobacterium tuberculosis* reveals the binding of shikimic acid. **Acta Crystallographica Section D**, Oxford, v. 60, n. 12, p. 2310-2319, 2004.

PIEMONTESE, L. et al. Design, synthesis and biological evaluation of a class of bioisosteric oximes of the novel dual peroxisome proliferator-activated receptor α/γ ligand LT175. **European Journal of Medicinal Chemistry**, Paris, v. 90, p. 583-594, 2015.

PRADO, V. et al. Targeting the Motion of Shikimate Kinase: Development of Competitive Inhibitors that Stabilize an Inactive Open Conformation of the Enzyme. **Journal of Medicinal Chemistry**, Nova Iorque, v. 59, p. 5471-5487, 2016.

ROGERS, D. W. **Computational Chemistry Using the PC**. 3. ed. New Jersey: John Wiley & Sons, 2003.

RUETSCH, Y. A.; BÖNI, T.; BORGEAT, A. From Cocaine to Ropivacaine: The History of Local Anesthetic Drugs. **Current Topics in Medicinal Chemistry**, Sharjah, v. 1, n. 3, p. 175-182, 2001.

SCHLICK, T. **Molecular Modeling and Simulation: An Interdisciplinary Guide**. 2. ed. New York: Springer, 2010.

SEDGEWICK, R.; WAYNE, K. **Algorithms**. 4. ed. Boston: Addison-Wesley, 2011.

SHAHLAEI, M. Descriptor Selection Methods in Quantitative Structure-Activity Relationship Studies: A Review Study. **Chemical Reviews**, Washington, v. 113, n. 10, p. 8093-8103, 2013.

SHEN, B. A New Golden Age of Natural Products Drug Discovery. **Cell**, Cambridge, v. 163, n. 6, p. 1297-1300, 2015.

SHEWCHUK, J. R. **An Introduction to the Conjugate Gradient Method Without the Agonizing Pain**. Edição 1 1/4, School of Computer Science, Carnegie Mellon University, Pittsburgh, 1994.

STEINBECK, C. et al. The Chemistry Development Kit (CDK): An Open-Source Java Library for Chemo- and Bioinformatics. **Journal of Chemical Information and Modeling**, Washington, v. 43, n. 2, p. 493-500, 2003.

SWAPNA, G. V. T et al. NMR investigation on the structure and conformation of 3'-azido-2',3'-dideoxyribosylthymine (AZT), an inhibitor of the HIV (AIDS virus). **Biochemical and Biophysical Research Communications**, San Diego, v. 164, n. 3, p. 1086-1092, 1989.

SYBYL. **Tripos Mol2 File Format**. Princeton: SYBYL, 2005.

TEOTICO, D. G. et al. Docking for fragment inhibitors of AmpC beta-lactamase. **Proceedings of the National Academy of Sciences of the United States of America**, Washington, v. 106, n. 18, p. 7455-7460, 2009.

THE SCRIPPS RESEARCH INSTITUTE. **What is the format of a PDBQT file?** 2007a. Disponível em: <<http://autodock.scripps.edu/faqs-help/faq/what-is-the-format-of-a-pdbqt-file>>. Acesso em 28 ago. 2018.

THE SCRIPPS RESEARCH INSTITUTE. **Where do I set the AutoDock 4 force field parameters?** 2007b. Disponível em: <<http://autodock.scripps.edu/faqs-help/faq/where-do-i-set-the-autodock-4-force-field-parameters>>. Acesso em 28 ago. 2018.

THORNER, C. W. Isosterism and molecular modification in drug design. **Chemical Society Reviews**, Letchworth, v. 8, n. 4, p. 563-580, 1979.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, Nova Iorque, v. 31, n. 2, p. 455–461, 2010.

TSUI, V.; ORTWINE, D. F.; BLANEY, J. M. Enabling drug discovery project decisions with integrated computational chemistry and informatics. **Journal of Computer-Aided Molecular Design**, Amsterdam, v. 31, n. 3, p. 287-291, 2017.

TU, Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. **Nature Medicine**, Nova Iorque, v. 17, n. 10, p. 1217-1220, 2011.

VAN-WIJNGAARDEN, I. et al. 2-Phenylpyrroles as conformationally restricted benzamide analogs: a new class of potential antipsychotics. 1. **Journal of Medicinal Chemistry**, Nova Iorque, v. 30, n. 11, p. 2099-2104, 1987.

VANOMMESLAEGHE, K.; GUVENCH, O.; MACKERELL JR., A. D. Molecular Mechanics. **Current Pharmaceutical Design**, Sharjah, v. 20, n. 20, p. 3281-3292, 2014.

VEBER, D. F. et al. Molecular properties that influence the oral bioavailability of drug candidates. **Journal of Medicinal Chemistry**, Nova Iorque, v. 45, n. 12, p. 2615-2623, 2002.

WALLACE, A. C.; LASKOWSKI, R. A.; THORNTON, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. **Protein Engineering**, Oxford, v. 8, n. 2, p. 127-134, 1996.

WERMUTH, C. G. Medicinal Chemistry: Definition and Objectives, the Three Main Phases of Drug Activity, Drug and Disease Classifications. In: Wermuth, C. G. (editor) **The Practice of Medicinal Chemistry**. 2. ed. Cambridge: Academic Press, 2003. cap. 2, p. 29-39.

WERMUTH, C. G. et al. Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1997). In: Bristol, J. A. (editor). **Annual reports in medicinal chemistry**. San Diego: Academic Press, 1998. cap. 36, p.385-395.

WIRTH, M. et al. SwissBioisostere: a database of molecular replacements for ligand design. **Nucleic Acids Research**, Londres, v. 41, p. D1137-D1143, 2013.

wwPDB. **PDB File Format v 3.3**, Nova Jersey: Worldwide Protein Data Bank, 2008.

YU, S. et al. New Scaffold for Angiogenesis Inhibitors Discovered by Targeted Chemical Transformations of Wondonin Natural Products. **ACS Medicinal Chemistry Letters**, Washington, v. 8, n. 10, p. 1066-1071, 2017.

YU, W.; MACKERELL JR., A. D. Computer-Aided Drug Design Methods. Antibiotics. In: Sass, P. (Ed.). **Antibiotics: Methods and Protocols**. New York: Springer, 2017. cap. 5, p. 85-106.

ZHANG, Q. et al. MORT: a powerful foundational library for computational biology and CADD. **Journal of Cheminformatics**, Londres, v. 6, n. 36, p. 1-9, 2014.