UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE QUÍMICA DE SÃO CARLOS

PLANEJAMENTO DE QUINONAS COM ATIVIDADE
TRIPANOSSOMICIDA

Fábio Alberto de Molfetta

Tese apresentada ao Instituto de Química
de São Carlos, da Universidade de São
Paulo, para obtenção do título de Doutor
em Ciências (Físico-Química).

Orientador: Prof. Dr. Albérico Borges Ferreira da Silva

São Carlos
2007

AGRADECIMENTOS

- À Patrícia, por seu amor, carinho e incentivo em todos os momentos.
- À toda minha família, a minha mãe, e em especial aos meus irmãos Mazé, Henrique e Núbia, pelo incentivo e apoio dado durante todos os momentos.
- Ao Prof. Dr. Albérico, pela paciência, orientação e ensinamentos transmitidos.
- À Profa. Dra. Márcia M. C. Ferreira, pela grande ajuda.
- Ao Prof. Dr. Montanari, pelo auxílio em diversos momentos.
- À profa. Dra. Roseli, pela ajuda em Redes Neurais.
- Aos meus amigos Agnaldo, Moacyr, Haiduke, Luiz Guilherme, Chico, Bird, Flávia, Káthia, Karen, Ednilsom, Paula e Melissa, que muito colaboraram para a realização deste trabalho.
- Aos meus amigos da Química Medicinal: Profa. Gorete, Maíra, Helton, Renato e Adriana.
- Ao compadre Wagner, pelo auxílio e amizade.
- A todos os Professores e Funcionários do prédio da Química Estrutural, em especial Ângela e Vânia.
- Às bibliotecárias: Lia, Wilneide, Eliana, Sônia, Solange e Vitória, pela atenção e grande ajuda.
- À Silvia e Andréia, da seção de Pós-Graduação, por todos os serviços prestados.
- À CAPES, pela bolsa concedida.
- Enfim, a todos que de alguma forma colaboraram com esta tese, os meus sinceros agradecimentos.

ÍNDICE

ÍNDICE DE FIGURAS.....	i
ÍNDICE DE TABELAS	iii
RESUMO	iv
ABSTRACT	vi
1.1. DOENÇA DE CHAGAS	1
1.2. QUINONAS	5
1.3. TRIPANOTIONA REDUTASE	10
1.4. QUÍMICA MEDICINAL	15
2. MÉTODOS DE CÁLCULOS UTILIZADOS	21
2.1. A EQUAÇÃO DE SCHRÖDINGER	21
2.2. TEORIA DO FUNCIONAL DE DENSIDADE (DFT)	23
3. MÉTODOS PARA ANÁLISE DE DADOS	28
3.1. PRÉ-PROCESSAMENTO DOS DADOS	29
3.2. ESCALAMENTO DOS DADOS	30
3.3. MÉTODOS UTILIZADOS	31
4. MÉTODOS QUIMIOMÉTRICOS UTILIZADOS	35
4.1. ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	35
4.2. ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)	37
4.3. ANÁLISE DE DISCRIMINANTES POR PASSOS (SDA)	40
4.4. MÉTODO DOS K-ÉSIMOS VIZINHOS MAIS PRÓXIMOS (KNN)	41
4.5. MODELOS INDEPENDENTES DE SIMILARIDADE UTILIZANDO COMPONENTES PRINCIPAIS (SIMCA)	42
4.6. REDES NEURAS ARTIFICIAIS	43
4.6.1. RETROPROGAÇÃO (<i>BACKPROPAGATION</i>)	46
5. COMPOSTOS ESTUDADOS	48
6. RESULTADOS	51
6.1. RESULTADOS DA ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)	51
6.2. RESULTADOS DA ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)	58
6.3. RESULTADOS DA ANÁLISE DE DISCRIMINANTE POR PASSO (SDA)	59
6.4. RESULTADOS DO K-ÉSIMOS VIZINHOS MAIS PRÓXIMOS (KNN)	61
6.5. RESULTADOS DO MODELO INDEPENDENTE DE SIMILARIDADE UTILIZANDO COMPONENTES PRINCIPAIS (SIMCA)	62
7. DISCUSSÃO DOS RESULTADOS OBTIDOS COM A PCA, HCA, SDA, KNN E SIMCA	65
8. VALIDAÇÃO DOS RESULTADOS OBTIDOS COM PCA, HCA, SDA, KNN e SIMCA	69
9. RESULTADOS DAS REDES NEURAS ARTIFICIAIS	71
10. DOCAGEM	76
10.1. MÉTODOS UTILIZADOS NA DOCAGEM	79
10.2. RESULTADOS DA DOCAGEM	81
10.2.1. RESULTADOS OBTIDOS COM O PROGRAMA AUTODOCK	82
10.2.2. RESULTADOS OBTIDOS COM O FLEXX	91
11. CONCLUSÕES	100
12. BIBLIOGRAFIA	102

ÍNDICE DE FIGURAS

FIGURA 1 – CICLO DE VIDA DO PARASITO <i>T. CRUZI</i> ENVOLVENDO AS FORMAS EPIMASTIGOTA, TRIPOMASTIGOTA E AMASTIGOTA.....	2
FIGURA 2 - FÁRMACOS INDICADOS PARA O TRATAMENTO DA DOENÇA DE CHAGAS.	3
FIGURA 3 – FÁRMACOS TESTADOS NO TRATAMENTO DA DOENÇA DE CHAGAS.....	4
FIGURA 4 – ESTRUTURA DE ALGUMAS QUINONAS.....	5
FIGURA 5 – ESTRUTURAS DA (A) UBIQUINONA, (B) VITAMINA K E (C) 1,4-NAFTOQUINONA.	6
FIGURA 6 - QUINONAS COM ATIVIDADE FARMACOLÓGICA: (A) MITOMICINA C E (B) DAUNORUBICINA.	6
FIGURA 7 – FORMAS ISOMÉRICAS DAS FORMAS PARA (A) E ORTO (B) QUINONAS.	7
FIGURA 8 – ESTRUTURA DO (A) 1,4- BENZOQUINONA E (B) ÁCIDO QUÍNICO.	8
FIGURA 9 – BESOIRO BOMBARDEIRO NO PROCESSO DE BORRIFAÇÃO CONTRA O PREDADOR.	8
FIGURA 10 – ESTRUTURA DO (A) LAPACHOL, (B) PLUMBAGINA E (C) MENADIONA.	9
FIGURA 11 – ENZIMAS (A) GLUTATIONA REDUTASE E (B) TRIPANOTONA REDUTASE...	11
FIGURA 12 – REAÇÃO CATALISADA PELA ENZIMA GLUTATIONA REDUTASE.	12
FIGURA 13 – REAÇÃO CATALISADA PELA ENZIMA TRIPANOTONA REDUTASE.	12
FIGURA 14 - ESTRUTURA DA TRIPANOTONA REDUTASE COM OS QUATRO RESÍDUOS: EM VERMELHO É O DOMÍNIO DO FAD, AMARELO É O DOMÍNIO DO NADPH, AZUL É O DOMÍNIO CENTRAL E VERDE O DOMÍNIO DE INTERFACE.....	13
FIGURA 15 – SÍTIO Z DA TRIPANOTONA REDUTASE CONTENDO OS AMINOÁCIDOS PHE 396, PRO 398 E LEU 399.....	13
FIGURA 16 – INIBIDORES DA TRIPANOTONA REDUTASE.....	14
FIGURA 17 - PROJEÇÃO DAS COMPONENTES PRINCIPAIS.....	36
FIGURA 18 - GRÁFICOS DE ESCORES (A) E LOADINGS (B) DE UM SISTEMA DE DADOS BIDIMENSIONAL.....	37
FIGURA 19 - DENDROGRAMA DE CONEXÃO, UTILIZANDO A CONEXÃO SIMPLES.....	39
FIGURA 20 - ALGUNS MÉTODOS DE AGRUPAMENTO.....	39
FIGURA 21 – REPRESENTAÇÃO DO MÉTODO KNN PARA CLASSIFICAÇÃO DE UMA NOVA AMOSTRA.	41
FIGURA 22 – REPRESENTAÇÃO DOS MODELOS SIMCA COM DIFERENTES COMPONENTES PRINCIPAIS PARA CADA CLASSE.	42
FIGURA 23 – REPRESENTAÇÃO DE UM NEURÔNIO BIOLÓGICO.....	44
FIGURA 24– REPRESENTAÇÃO DA RETROPROPAGAÇÃO PARA UMA REDE COM UMA CAMADA DE ENTRADA (W1), UMA CAMADA INTERMEDIÁRIA (W2) E UMA CAMADA DE SAÍDA (W3).	46
FIGURA 25 – ESTRUTURA GERAL DAS QUINONAS ESTUDADAS E SISTEMA DE NUMERAÇÃO ADOTADO NESTE TRABALHO.	48
FIGURA 26 – ESTRUTURA E INDICAÇÃO DE ATIVIDADE DAS QUINONAS ESTUDADAS....	50
FIGURA 27 – SEPARAÇÃO DOS COMPOSTOS UTILIZANDO O CONJUNTO COMPLETO DE VARIÁVEIS:	51
FIGURA 28 – HISTOGRAMA COM OS VALORES CALCULADOS OBTIDOS PARA AS VARIÁVEIS (T5, QTS1, VOLS2 E HOMO-1) RESPONSÁVEIS PELA SEPARAÇÃO ENTRE OS COMPOSTOS ATIVOS E INATIVOS.	53
FIGURA 29 – GRÁFICO DOS ESCORES (PC1 E PC2) PARA AS 25 QUINONAS ESTUDADAS COM SEPARAÇÃO ENTRE AS QUINONAS ATIVAS (GRUPO A) E INATIVAS (GRUPO B).....	54
FIGURA 30 – GRÁFICO DO LOADINGS (PC1 E PC2) PARA AS VARIÁVEIS RESPONSÁVEIS PELA SEPARAÇÃO DOS COMPOSTOS EM ATIVOS E INATIVOS.....	55

FIGURA 31 - DENDROGRAMA DE CONEXÃO, UTILIZANDO A CONEXÃO INCREMENTAL..	59
FIGURA 32 – GRÁFICO DOS ESCORES EM PC1 x PC2 x PC3 COM OS LIMITES DAS HIPERCAIXAS.	63
FIGURA 33 – GRÁFICO DAS DISTÂNCIAS ENTRE AS CLASSES: CS1 (CLASSE 1 – COMPOSTOS INATIVOS) E CS2 (CLASSE 2 – COMPOSTOS ATIVOS).	64
FIGURA 34 - ÂNGULO DE TORÇÃO T5 FORMADO PELOS ÁTOMOS C1, C2, C3 E O4.	66
FIGURA 35 – SOMA DAS CARGAS FORMADAS PELOS ÁTOMOS C1, C2, C17, C18 E DOS SUBSTITUINTES NA REGIÃO A.....	66
FIGURA 36 – VOLUME TOTAL FORMADO PELOS ÁTOMOS C15, O16, C17, C18 E PELOS SUBSTITUINTES NA REGIÃO B.	67
FIGURA 37 – SUPERFÍCIE DE DENSIDADE ELETRÔNICA DO ORBITAL HOMO-1 PARA OS COMPOSTOS 18 E 3.....	67
FIGURA 38 - SUPERFÍCIE DE DENSIDADE ELETRÔNICA DO ORBITAL HOMO-1 PARA OS COMPOSTOS 12, 16, 21 E 22.....	68
FIGURA 39 – SUPERFÍCIE DE DENSIDADE ELETRÔNICA DO ORBITAL HOMO-1 PARA O COMPOSTO 7 (COMPOSTO INATIVO COM GRUPO METOXILA).	69
FIGURA 40 – ESTRUTURA QUÍMICA DOS 3 COMPOSTOS UTILIZADOS NO ESTUDO DE PREVISÃO.	70
FIGURA 41 - ERROS RMS CALCULADOS CONTRA O NÚMERO DE NEURÔNIOS NA CAMADA INTERMEDIÁRIA.....	73
FIGURA 42 - RESULTADOS OBTIDOS COM AS 28 QUINONAS ATRAVÉS DO PROGRAMA AUTODOCK.	82
FIGURA 43 – ESTRUTURA DOS AMINOÁCIDOS SERINA, LISINA E TREONINA.	83
FIGURA 44 – MODELO DE INTERAÇÃO DAS QUINONAS INATIVAS COM A ENZIMA TRIPANOTIONA REDUTASE DENTRO DO SÍTIO DE INTERAÇÃO OBTIDO COM O PROGRAMA AUTODOCK.	85
FIGURA 45 – MODELO DE INTERAÇÃO DAS QUINONAS ATIVAS COM A ENZIMA TRIPANOTIONA REDUTASE DENTRO DO SÍTIO DE INTERAÇÃO OBTIDO COM O PROGRAMA AUTODOCK.	88
FIGURA 46 – ENERGIAS DE DOCAGEM OBTIDAS COM O PROGRAMA AUTODOCK PARA AS 28 QUINONAS ESTUDADAS.	89
FIGURA 47 - COMPOSTOS 25, 11, 21 E 16 COM OS RESÍDUOS E AS DISTÂNCIAS OBTIDAS ATRAVÉS DO PROGRAMA AUTODOCK.	90
FIGURA 48 – RESULTADOS OBTIDOS COM AS 28 QUINONAS ATRAVÉS DO PROGRAMA FLEXX.	91
FIGURA 49 – ESTRUTURA DO RESÍDUO DE LEUCINA.	91
FIGURA 50 - MODELO DE INTERAÇÃO DAS QUINONAS INATIVAS COM A ENZIMA TRIPANOTIONA REDUTASE DENTRO DO SÍTIO DE INTERAÇÃO OBTIDO COM O PROGRAMA FLEXX.	93
FIGURA 51 - MODELO DE INTERAÇÃO DAS QUINONAS ATIVAS COM A ENZIMA TRIPANOTIONA REDUTASE DENTRO DO SÍTIO DE INTERAÇÃO OBTIDO COM O PROGRAMA FLEXX.	95
FIGURA 52 – ENERGIAS LIVRES DE INTERAÇÃO OBTIDAS COM O PROGRAMA FLEXX PARA AS 28 QUINONAS ESTUDADAS.	97
FIGURA 53 - COMPOSTOS 20, 28, 25 E 2 COM OS RESÍDUOS E AS DISTÂNCIAS OBTIDAS COM O PROGRAMA FLEXX.	98

ÍNDICE DE TABELAS

TABELA 1 - TABELA DE CORRELAÇÃO ENTRE AS VARIÁVEIS SELECIONADAS UTILIZANDO O PESO DE FISHER	34
TABELA 2 - VALORES OBTIDOS PARA AS QUATRO PROPRIEDADES (VARIÁVEIS) USADAS NA CLASSIFICAÇÃO DAS 28 QUINONAS ESTUDADAS COM SUA RESPECTIVA INDICAÇÃO DE ATIVIDADE TRIPANOSSOMICIDA	52
TABELA 3 - VALORES DAS COMPONENTES PRINCIPAIS PARA AS VARIÁVEIS RESPONSÁVEIS PELA SEPARAÇÃO DOS COMPOSTOS	56
TABELA 4 – VALORES DOS ESCORES NA PRIMEIRA COMPONENTE PRINCIPAL (PC1) PARA CADA COMPOSTO ESTUDADO	57
TABELA 5 - MATRIZ DE CLASSIFICAÇÃO	60
TABELA 6 - MATRIZ DE VALIDAÇÃO CRUZADA RETIRANDO-SE UMA AMOSTRA DE CADA VEZ	61
TABELA 7 - CLASSIFICAÇÃO OBTIDA COM O MÉTODO KNN PARA AS 25 QUINONAS ESTUDADAS	62
TABELA 8 - VALORES DAS DISTÂNCIAS E RESÍDUOS INTER-CLASSE	63
TABELA 9 – PODER DISCRIMINATÓRIO E PODER DE MODELAGEM DAS VARIÁVEIS T5, QTS1, VOLS2 E HOMO-1	64
TABELA 10 – VALORES OBTIDOS PARA AS VARIÁVEIS PARA OS 3 NOVOS COMPOSTOS ESTUDADOS	70
TABELA 11 – RESULTADOS DA PREVISÃO DA ATIVIDADE TRIPANOSSOMICIDA UTILIZANDO OS MÉTODOS PCA, HCA, SDA, KNN E SIMCA PARA AS 3 NOVAS QUINONAS ESTUDADAS: (+) ATIVO E (-) INATIVO.....	71
TABELA 12 – COMPOSTOS QUE FORAM UTILIZADOS PARA TESTE DAS REDES NEURAI COM OS ERROS DE TREINAMENTO (RMS1) E ERROS DO CONJUNTO TESTE (RMS2).....	74
TABELA 13 - RESULTADOS OBTIDOS COM O MODELO 1 UTILIZANDO REDES NEURAI PARA O CONJUNTO TESTE.....	75
TABELA 14 - RESULTADOS OBTIDOS COM O MODELO 6 UTILIZANDO REDES NEURAI PARA O CONJUNTO TESTE.....	75
TABELA 15 – VALORES DOS PARÂMETROS UTILIZADOS PELO ALGORITMO LAMARCKIANO NO PROCEDIMENTO DE DOCAGEM	80
TABELA 16 – ENERGIAS LIVRES DE INTERAÇÃO PARA AS 28 QUINONAS ESTUDADAS OBTIDAS COM O PROGRAMA AUTODOCK	88
TABELA 17 – ENERGIAS LIVRES DE INTERAÇÃO PARA AS 28 QUINONAS ESTUDADAS OBTIDAS COM O PROGRAMA FLEXX	96

RESUMO

Um conjunto de 25 quinonas com atividade tripanossomicida foram estudadas usando o método da Teoria do Funcional da Densidade (DFT) a fim de calcular propriedades eletrônicas atômicas e moleculares que estejam correlacionadas com a atividade biológica.

Métodos quimiométricos tais como Análise de Componentes Principais (PCA), Análise Hierárquica de Agrupamentos (HCA), Análise Discriminante por Passo (SDA), Método do vizinho mais próximo (KNN) e Modelos independentes de similaridade utilizando componentes principais (SIMCA), foram usados para obter possíveis relações entre os descritores calculados e a atividade biológica estudada e prever a atividade tripanossomicida de novas quinonas a partir do conjunto teste. Quatro descritores foram responsáveis pela separação entre os compostos ativos e inativos: T_5 (ângulo de torção formado pelos átomos 1, 2, 17 e 18), HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO), QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A), VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B). Estes descritores representam as informações do tipo de interação que ocorre entre os compostos e o receptor biológico. O estudo de predição foi feito com um conjunto de três novas quinonas através dos métodos PCA, HCA, SDA, KNN e SIMCA.

Além dos cinco métodos quimiométricos, também foi utilizado o método de redes neurais artificiais, utilizando o algoritmo de retropropagação de erros. As quatro variáveis (T_5 , QTS1, VOLS2 e HOMO-1) foram empregadas de forma a validar os modelos usados anteriormente. A arquitetura da rede com 4 neurônios na camada de entrada, 10 neurônios na camada intermediária e 2 neurônios na camada de saída foi adotada observando-se o erro quadrático médio entre o valor verdadeiro e o desejado sobre todo o conjunto de treinamento. A porcentagem de informação correta foi de 87,5%, e apenas um composto foi previsto incorretamente no conjunto teste, o que indica que o modelo é robusto e que pode ser capaz de fazer previsões.

Estudos de docagem foram realizados com dois programas diferentes na abordagem dos ligantes, o Autodock e o FlexX. Os resultados de docagem na enzima tripanotona redutase mostraram que todos os compostos estudados permanecem em uma cavidade hidrofóbica fora da região do sítio ativo chamada de sítio Z. Os resíduos que estão envolvidos na ligação dos ligantes são Lys62, Thr66, Thr397, Thr463, Leu399, Ser464, Glu466 e Glu467, onde os resíduos Thr66, Thr463 e Leu399 são conservados em diferentes

triptoninas e podem ser usados no desenvolvimento de inibidores seletivos contra a enzima do parasito.

ABSTRACT

A set of 25 quinone compounds with anti-trypanocidal activity was studied by using the Density Functional Theory (DFT) method in order to calculate electronic atomic and molecular properties to be correlated with the biological activity.

The chemometric methods Principal Component Analysis (PCA), Hierarchical Cluster Analysis (HCA), Stepwise Discriminant Analysis (SDA), *K*th nearest neighbor (KNN) and Soft Independent Modeling of Class Analogy (SIMCA) were used to obtain possible relationships between the calculated descriptors and the biological activity studied and predict the anti-trypanocidal activity of new quinone compounds from a test set. Four descriptors were responsible for the separation between the active and inactive compounds: T_5 (torsion angle), HOMO-1 (energy of the first molecular orbital below HOMO), QTS1 (sum of absolute values of the atomic charges) and VOLS2 (volume of the substituent at region B). These descriptors give information on the kind of interaction that occurs between the compounds and the biological receptor. The prediction study was done with a set of three new quinone compounds by using the PCA, HCA, SDA, KNN and SIMCA.

Beside the five chemometric methods, the neural network method was used by employing the backpropagation algorithm. The four variables (T_5 , QTS1, VOLS2 and HOMO-1) were employed to validate the models constructed previously. The architecture of networks consisting of four neurons at input layers, ten neurons at intermediary layers and two neurons at output layers was adopted to observe the root mean square error between the true and desired output over the entire training set. The percentage of correct classification was 87.5%, and only one compound was predicted wrong in the test set, which indicates that the model is robust and could be able to make predictions.

The docking studies were carried out with two different programs in the approach of ligands: the Autodock and FlexX. The docking results on trypanothione reductase enzyme showed that all studied compounds stay at second hydrophobic pocket in the outer region of the active site called the Z-site. The residues that could be specifically involved in the binding of ligands are Lys62, Thr66, Thr397, Thr463, Leu399, Ser464, Glu466 and Glu467, where the residues Thr66, Thr463 and Leu399 are conserved in different trypanothiones and could be used for the development of selective inhibitors against to the parasite enzyme.

1. INTRODUÇÃO

1.1. DOENÇA DE CHAGAS

A doença de Chagas é uma endemia prevalente na América Latina causada pelo protozoário flagelado *Trypanosoma cruzi* (parasito da ordem *Kinetoplastida*), o qual é transmitido para o hospedeiro humano, principalmente pelas fezes do inseto do vetor hematófago conhecido como “barbeiro” ou “chupão”. As espécies mais comuns de triatomíneos são o *Triatoma infestans*, *Panstrongylus megistus* e *Triatoma sordida*, entre outras. Estes insetos encontraram um habitat favorável nas frestas das paredes de barro ou de madeiras das chamadas casas de pau a pique, encontradas principalmente na zona rural e na periferia da zona urbana (favelas).

Estima-se que cerca de 20 milhões de pessoas estão infectadas e 100 milhões (25% da população da América Latina) correm o risco de adquiri-la^[1-2].

A transmissão da doença ocorre principalmente por três mecanismos: através das picadas do barbeiro, por transfusão de sangue contaminado e por transfusão congênita^[3]. A transmissão congênita é a mais preocupante devido aos efeitos colaterais causados nas mães e nos bebês pelos medicamentos disponíveis^[4].

O ciclo de vida do parasito é complexo (Figura 1), envolvendo um inseto como vetor de transmissão e mamíferos como hospedeiros. Nos insetos, o protozoário multiplica-se sob uma forma flagelada, conhecida como epimastigota, que é liberada nas fezes do inseto sob a forma metacíclica flagelada, conhecida como tripomastigota. Após a infecção, que se dá pelo contato da mucosa intacta do mamífero com as fezes do barbeiro, ou através de uma ferida do hospedeiro, o parasito multiplica-se intracelularmente numa forma aflagelada, conhecida como amastigota, que prevalece na forma crônica da doença. O protozoário é subsequentemente liberado na corrente sanguínea do mamífero sob a forma tripomastigota^[5].

No Brasil, esta endemia atinge cerca de oito milhões de habitantes, principalmente populações pobres que residem em condições precárias^[6]. É a única doença infecciosa, dentre as grandes endemias brasileiras, que não tem tratamento eficiente, fazendo parte das 6 doenças endêmicas parasitárias reconhecidas como prioridade pela Organização Mundial de Saúde (OMS)^[7, 8].

Figura 1 – Ciclo de vida do parasito *T. cruzi* envolvendo as formas epimastigota, tripomastigota e amastigota.

A doença se manifesta de forma não muito severa (entre duas e três semanas após a infecção) podendo levar a quadros de miocardite e meningoencefalite em crianças, esta última sendo fatal em aproximadamente 50% dos casos. As piores complicações da doença de Chagas desenvolvem-se de 10 a 20 anos após a infecção e são caracterizadas por cardiopatia crônica (27% dos infectados), complicações crônicas digestivas (6%) e neurológicas (3%). Os pacientes com graves doenças crônicas vão tornando-se cada vez mais debilitados até a morte, geralmente por falha do coração^[9]. A doença causa grande impacto socioeconômico já que acomete muitos indivíduos em idade produtiva^[10].

Os fármacos disponíveis para o tratamento da doença de Chagas estão longe de serem satisfatórios (são pouco efetivos, tóxicos e com preço elevado), e muitos foram introduzidos nas décadas de 60 e 70^[3, 11].

Dois fármacos são indicados para o tratamento da doença de Chagas, o Nifurtimox (Lampit, da Bayer) e o Benzonidazol (Rochagan, da Roche) (Figura 2). Desde 1980, o Nifurtimox não é comercializado no Brasil, Argentina, Chile e Uruguai. O Nifurtimox e os nitrofuranos têm seu efeito parasitário devido a sua habilidade de gerar radicais nitros que promovem a formação de metabólitos de oxigênio tóxicos dentro do parasito^[12].

O Nifurtimox (Figura 2a) e o Benzonidazol (Figura 2b) apresentam sérios efeitos colaterais como: perda de peso, náuseas, vômitos, diarreia, alergia cutânea e neuropatia^[13]. Além disso, os resultados obtidos com ambos os fármacos variam de acordo com o estágio da

doença, período de tratamento, dose, idade e origem geográfica dos pacientes (uma vez que algumas cepas do *T. Cruzi* estudadas apresentam susceptibilidade diferente para esse tipo de fármaco) ^[14].

Figura 2 - Fármacos indicados para o tratamento da doença de Chagas.

Levando em consideração os vários efeitos colaterais do Nifurtimox e do Benzonidazol, foi testado o Megazol. Esse composto foi ativo contra diferentes cepas do parasito, o que representou uma alternativa promissora, apesar dos nitrocompostos serem suspeitos de provocar câncer ^[15]. Outros compostos tais como alopurinol (Figura 3a), fluconazol (Figura 3b), cetoconazol (Figura 3c), e itraconazol (Figura 3d) estão sendo testados^[3].

(d) - Itraconazol

Figura 3 – Fármacos testados no tratamento da doença de Chagas.

Para a eliminação do parasito no sangue contaminado em bancos de sangue, a única substância utilizada como agente quimioprolático é a violeta de genciana, que resulta em 100% de eliminação das formas infectantes. Porém, seu uso também é limitado pelos efeitos colaterais e pela cor transferida à pele e urina dos pacientes^[16].

O desenvolvimento no estudo da bioquímica do *T. cruzi* vem permitindo o desenvolvimento de compostos mais seletivos e eficazes a partir de alvos terapêuticos potenciais, tais como as enzimas cisteína protease, xantina fosforibosil transferase, gliceraldeído-3-fosfato desidrogenase, DNA topoisomerases, diidrofolato redutase e tripanotiona redutase^[3]. As enzimas são geralmente excelentes alvos de fármacos, visto que se podem planejar compostos que melhor se ajustem à cavidade do sítio ativo ou outro sítio que module sua função^[17].

Em 2000, apenas 1% do investimento global em pesquisas na área de saúde foi dedicado à descoberta de fármacos para doenças tropicais (malária, leishmaniose, tripanossomíase) e tuberculose^[18]. Estas doenças são totalmente negligenciadas pelas companhias farmacêuticas devido à baixa renda dos países afetados e ao baixo retorno financeiro esperado^[11]. Claramente, existe a necessidade de novos agentes terapêuticos mais seguros e eficazes contra a doença de Chagas^[19].

1.2. QUINONAS

Quinonas formam um grupo de compostos coloridos com dois grupamentos carbonílicos, que podem ser adjacentes ou separados, geralmente em um anel insaturado de seis átomos (Figura 1) ^[20].

Figura 4 – Estrutura de algumas quinonas.

As quinonas desempenham papel fundamental na bioquímica celular e exercem relevante atividade biológica. Devido ao fato de serem facilmente oxidadas e reduzidas, algumas quinonas participam da fosforilação oxidativa, transportando elétrons entre a coenzima da flavina e os citocromos, outras participam da fotossíntese e algumas participam dos processos enzimáticos onde atuam como co-fatores ^[20, 21].

As ubiquinonas (Figura 5a), também chamadas de coenzimas Q (encontradas dentro da membrana mitocondrial interna da célula), sendo importantes no transporte de elétrons. A vitamina do tipo K (Figura 5b), com estrutura da 1,4-naftoquinona (Figura 5c), é de irrestrita necessidade aos seres vivos, pois possuem ação controladora da coagulação sanguínea ^[22]. A Figura 5 mostra as estruturas dos compostos acima citados.

Figura 5 – Estruturas da (a) ubiquinona, (b) vitamina K e (c) 1,4-naftoquinona.

As quinonas apresentam propriedades microbicidas, virucidas, antitumorais, anticânceres e, como será mostrado no presente trabalho, também são estudadas no combate à doença de Chagas^[23].

Existem algumas classes de quinonas que apresentam atividades farmacológicas tais como as mitomicinas e as antraciclina. As mitomicinas, de uso clínico reconhecido, apresentam atividade antibiótica e antitumoral. A mitomicina C (Figura 6a) é a mais conhecida do grupo e está sendo utilizada na quimioterapia de certos tipos de tumores sólidos. As antraciclina apresentam atividade antibiótica, sendo que a daunorubicina (Figura 6b), que pertence a esta classe de antraciclina, também possui efeito terapêutico contra a leucemia humana^[23].

Figura 6 - Quinonas com atividade farmacológica: (a) mitomicina C e (b) daunorubicina.

Com base na sua estrutura molecular, as quinonas são divididas em diferentes grupos, utilizando-se como critério o tipo de sistema aromático que sustenta o anel das quinonas:

- ✓ Benzoquinonas – um anel benzênico
- ✓ Naftoquinonas – um anel naftalênico
- ✓ Antraquinonas – um anel antracênico linear ou angular

Em decorrência de diferentes arranjos (isomeria), com um mesmo tipo de anel, pode-se ter, dependendo das disposições relativas das carbonilas, diferentes quinonas. Por exemplo: no arranjo de base naftalênica, onde as carbonilas são vizinhas, tem-se a forma isomérica 1,2 ou *orto*-quinona (Figura 7a), e quando se tem dois carbonos entre as carbonilas, a forma isomérica é a 1,4 ou *para*-quinona (Figura 7b). Outros arranjos isoméricos são conhecidos, principalmente em sistemas policíclicos, nos quais os distanciamentos entre as carbonilas podem ser maiores. Estas formas isoméricas diferem muito em suas propriedades físicas, químicas e quanto a sua atuação biológica. Um exemplo típico é a *orto*-naftoquinona (β -lapachona) que é muito mais ativa contra o tripanossomatídeo do que seu isômero natural α -lapachona^[23].

(a) - β -Lapachona

(b) - α -Lapachona

Figura 7 – Formas isoméricas das formas para (a) e orto (b) quinonas.

Um exemplo conhecido da classe das quinonas é a 1,4-benzoquinona (Figura 8a), sintetizada pela primeira vez em 1838, e que envolveu a oxidação do ácido quínico (Figura 8b), Figura 8. O nome genérico quinona é derivado deste ácido^[24]. As 1,4-benzoquinonas são frequentemente produzidas por insetos para repelir predadores, conforme mostrado na Figura 9^[22].

Figura 8 – Estrutura do (a) 1,4- benzoquinona e (b) ácido quínico.

Figura 9 – Besouro bombardeiro no processo de borrifação contra o predador.

As naftoquinonas são compostos importantes biologicamente, especialmente por causa de suas ações farmacológicas e citotóxicas^[25]. Dentre as várias naftoquinonas naturais, destaca-se o lapachol (Figura 10a) que pode ser considerado um dos principais representantes do grupo de quinonas das tabebuias. O lapachol é de fácil extração da serragem da madeira de várias espécies de ipê^[23]. O lapachol e seus análogos são conhecidos por apresentarem atividade antitumoral, antibiótica e antimalárica. Foi observado também que o lapachol exibe atividade antiinflamatória, e foi altamente ativo contra carcinossarcoma (tumor maligno) de Walker 256^[26].

Muitas naftoquinonas tais como menadiona (Figura 10b), plumbagina (Figura 10c) e lapachol (Figura 10a), mostrados na Figura 10, mostraram atividade contra diferentes tripanossomatídeos e Leishmanias responsáveis pela doença do sono Africana (*Trypanossoma brucei rhodesiense* e *Trypanossoma brucei gambiense*), doença de Chagas (*Trypanossoma cruzi*) e Kala-azar (*Leishmania donovani*). Este protozoário parasito da ordem dos

cinetoplástideos são particularmente sensíveis ao estresse oxidativo e tem desenvolvido um sistema antioxidante específico^[27].

Figura 10 – Estrutura do (a) lapachol, (b) plumbagina e (c) menadiona.

O mecanismo de ação das naftoquinonas envolve a absorção pelo parasito e a redução da quinona em semiquinona ($Q^{\bullet-}$) e quinol (Q^{2-}). A formação da semiquinona é um processo muito rápido, com uma mistura quinol e quinona. Uma vez formada a espécie semiquinona ($Q^{\bullet-}$), esta reduz o oxigênio molecular ao ânion radical superóxido ($O_2^{\bullet-}$), conforme mostrado na reação 1. Na reação 2, a presença da enzima superóxido desmutase (SOD) transforma o ânion radical superóxido em peróxido de hidrogênio (H_2O_2). Embora o H_2O_2 não seja um radical livre, é uma substância bastante reativa, que pode promover a oxidação de biomoléculas. O ânion radical superóxido por catálise com metais de transição (reação do tipo Fenton), ou por reação com H_2O_2 (reação de Harber-Weiss), gera radicais hidroxilas ($OH^{\bullet}$) no interior da célula, conforme mostrado pelas reações 3 e 4. Em resumo, a produção de peróxido de hidrogênio, e dos radicais hidroxilas, são as principais espécies responsáveis pelo estresse oxidativo, levando a inativação da enzima e danificando a membrana celular. As reações químicas, que mostram as reações das quinonas com a formação dos radicais, são mostradas abaixo^[23, 27, 28, 29].

1.3. TRIPANOTIONA REDUTASE

O oxigênio pode ser reduzido apenas de modo parcial, produzindo espécies que reajam com uma variedade de componentes celulares. A espécie de oxigênio reativo mais conhecido é o radical superóxido, reação 5:

O radical superóxido é um precursor de outras espécies reativas. A protonação do $\text{O}_2^{\bullet -}$ produz $\text{OH}^{\bullet}$, um oxidante mais forte que o $\text{O}_2^{\bullet -}$. A espécie de oxigênio mais potente nos sistemas biológicos é o radical hidroxila, que forma o peróxido de hidrogênio (H_2O_2), reação 6:

Na reação 7, o radical hidroxila é formado por meio da reação do superóxido com H_2O_2 :

Embora a maioria dos radicais livres possuam uma vida extremamente curta (a meia vida do $\text{O}_2^{\bullet -}$ é de 1×10^{-6} s e a do $\text{OH}^{\bullet}$ é de 1×10^{-9} s), eles facilmente captam os elétrons de outras moléculas, convertendo-as em radicais livres, iniciando uma reação em cadeia.

A natureza aleatória dos ataques realizados pelos radicais livres dificulta a caracterização de seus produtos de reação, mas todas as classes de moléculas biológicas são suscetíveis às lesões oxidativas causadas pelos radicais livres. A oxidação dos lipídeos poliinsaturados nas células rompem a estrutura das membranas biológicas, e as lesões podem produzir mutações oxidativas no DNA. A função enzimática também pode ser comprometida devido à reação dos radicais com a cadeia lateral dos aminoácidos.

Os antioxidantes destroem os radicais livres, como o $\text{O}_2^{\bullet -}$ e o $\text{OH}^{\bullet}$. A principal enzima antioxidante é a superóxido dismutase, que está presente em todos os organismos aeróbios, e que catalisa a conversão do $\text{O}_2^{\bullet -}$ em H_2O_2 , conforme mostrado na reação 8:

Outras defesas celulares contra o dano oxidativo incluem as vitaminas antioxidantes C e E, além das enzimas catalase, glutathione peroxidase e glutathione reductase^[30, 31].

A glutathione reductase (Figura 11a) é uma flavoproteína oxirredutase dissulfito dependente da nicotinamida-adenina-dinucleotídeo-fosfato reduzida (NADPH), sendo essencial contra o estresse oxidativo. Sua capacidade redutora é determinada pelo grupamento SH presente na cisteína^[30, 32]. A enzima catalisa reações de oxidorredução que são extremamente importantes no processo metabólico^[30]. A forma tiol da glutathione funciona como agente protetor, mantendo o ambiente intracelular reduzido. A enzima glutathione reductase garante que altos níveis de tiol sejam preservados catalisando a redução dissulfito, como mostra a Figura 12^[33].

A tripanotona reductase (Figura 11b) desempenha uma função similar da glutathione reductase nos parasitos, pois mantém o ambiente intracelular reduzido e protege o parasito contra o estresse oxidativo^[34]. Em 1985, Fairlamb e colaboradores descobriram que os parasitos da família dos triatomíneos continham a N¹, N⁸-bis (glutathionil) espermidina, conhecida como tripanotona, sendo que os triatomíneos não possuem glutathione reductase^[12].

A enzima tripanotona reductase é responsável pela conversão da tripanotona dissulfito (T[S]₂) para a forma ditiol (dihidrotimpanotona) (T[SH]₂)^[33], como mostrado na Figura 13.

Figura 11 – Enzimas (a) glutathione reductase e (b) tripanotona reductase.

Figura 12 – Reação catalisada pela enzima glutathione redutase.

Figura 13 – Reação catalisada pela enzima tripanotiona redutase.

Comparação entre a tripanotiona e a glutathione revela notável similaridade, ambas são homodímeras de massa molecular variando de 100-110 kDa e ambas requerem dois cofatores, o NADPH e o FAD, que regulam o sítio ativo dissulfito. As duas enzimas apresentam 41% de similaridade e 70% dos resíduos críticos para a atividade da glutathione são conservados na tripanotiona, incluindo a ligação redox dissulfito^[35].

A atividade catalítica similar, e a especificidade com relação aos respectivos substratos, indicam que é possível inibir a tripanotiona do parasito, mas não a do hospedeiro humano^[33]. A tripanotiona redutase é essencial para o crescimento e desenvolvimento do parasito e representa um dos alvos para o desenvolvimento de um fármaco tripanossomicida,

e o entendimento desta enzima é necessário para facilitar a pesquisa por inibidores com as desejadas propriedades farmacológicas ^[33].

A estrutura do monômero da tripanotona redutase com a localização dos domínios está presente na Figura 14. Ambas as cadeias polipeptídicas estão subdivididas em quatro domínios, chamados de domínio do FAD (resíduos 1 até 160), domínio do NADPH (resíduos 161 até 290), domínio central (resíduos 291 até 360) e um domínio de interface (resíduos 361 até 492) ^[36]. Os resíduos envolvidos na catalise são os resíduos de cisteína 53 e 58 e da histidina 461 ^[33]. Estudos de modelagem na tripanotona sugerem uma segunda cavidade hidrofóbica em uma região fora do sítio ativo formado pelos aminoácidos Phe 396, Pro 398 e Leu 399, chamado sítio Z ^[37], conforme mostrado pela Figura 15.

Figura 14 - Estrutura da tripanotona redutase com os quatro resíduos: em vermelho é o domínio do FAD, amarelo é o domínio do NADPH, azul é o domínio central e verde o domínio de interface.

Figura 15 – Sítio Z da tripanotona redutase contendo os aminoácidos Phe 396, Pro 398 e Leu 399.

O amplo sítio da tripanotiona redutase é um excepcional desafio para a descoberta de fármacos baseados na estrutura do receptor. Esse extenso sítio ativo permite acomodar ligantes volumosos (como no caso da Kukuamina) de diferentes modos, ou que mais de uma molécula de inibidor possa se ligar simultaneamente na tripanotiona. A extensa cavidade hidrofóbica faz com que o planejamento de inibidores, usando interações direcionadas tais como ligações de hidrogênio, seja realmente difícil ^[37].

1.4. QUÍMICA MEDICINAL

A história do desenvolvimento das civilizações Oriental e Ocidental é rica em exemplos da utilização de recursos naturais na medicina, merecendo destaque as civilizações Egípcia, Greco-romana e Chinesa. A medicina tradicional chinesa desenvolveu-se com tal grandiosidade e eficiência que até hoje muitas espécies e preparados vegetais medicinais são estudados na busca pelo entendimento de seu mecanismo de ação e no isolamento dos princípios ativos ^[41].

As técnicas desenvolvidas e utilizadas no Egito para a conservação de múmias ainda são um desafio para a Química moderna. Na Idade Antiga, além de técnicas medicinais, muitos venenos foram descobertos na natureza e utilizados para fins de defesa, caça e ilícitos, como na utilização do veneno de Hemlock (*Conium maculatum*) na execução de prisioneiros, inclusive Sócrates, durante o Império Grego ^[41].

O ópio preparado dos bulbos de *Papaver somniferum* é conhecido há séculos por suas propriedades soporíferas¹ e analgésicas. Esta planta era utilizada desde a época dos Sumérios (4000 A.C.), havendo relatos da mitologia grega atribuindo à papoula do ópio o simbolismo de Morfeu, o deus do sono.

Durante a colonização espanhola do Peru, em 1630, os jesuítas tomaram conhecimento da utilização pelos índios das cascas secas de espécies de *Cinchona* para o tratamento de alguns tipos de febres ^[41].

A partir do século XIX, com o desenvolvimento da química orgânica em paralelo ao estudo de plantas, foram registrados, com base científica, os primeiros estudos que resultaram no isolamento de alguns princípios ativos já conhecidos como medicinais ^[42]. Pelletier e

¹ Que tem ou dá sabor, saboroso.

Caventou, em 1820, isolaram a quinina, que durante trezentos anos foi o único princípio ativo eficaz contra a malária ^[41].

Felix Hofmann descobriu o ácido acetil-salicílico, que foi lançado no mercado em 1897, com o nome de Aspirina®. O ácido acetil-salicílico foi o pioneiro dos fármacos sintéticos ^[41].

No século XX, logo depois da Grande Guerra, com o surgimento dos antibióticos, quando Alexander Fleming detectou a inibição do crescimento de placas de cultura semeadas com colônias de estafilococos contaminadas com fungos, houve o desenvolvimento de fármacos sintéticos produzidos pela indústria farmacêutica, o que marcou o declínio do uso de plantas medicinais e no investimento em fármacos de origem vegetal ^[42].

Por volta dos anos 80, a indústria farmacêutica redirecionou seu processo de seleção de produtos bioativos, processo denominado de busca racional de fármacos, através do uso de fracionamento guiado por bioensaios específicos para enzimas, receptores, DNA e também pela pesquisa para novas entidades químicas bioativas (bioNCEs- *Bioactives New Chemical Entity*) ^[42].

Nos anos 90, com o advento da química combinatória, era possível criar enormes coleções de moléculas que poderiam ser ensaiadas *en masse*. Com a produção de compostos diversos em coleções, houve um aumento da probabilidade de encontrar novos compostos de valor terapêutico e comercial ^[42, 43]. Paralelamente, ocorreu o desenvolvimento de métodos de ensaios biológicos automatizados (*high throughput screening* - HTS) que passaram a permitir a avaliação *in vitro* de milhares de substâncias por experimento. Estas técnicas, empregadas simultaneamente, permitem a identificação de novos compostos capazes de interagir com alvos terapêuticos ensaiados em escala, inicialmente, micromolar e, atualmente, nanomolar. Devido ao emprego destas técnicas combinadas surgiu o termo *hit*² ^[41].

A euforia inicial que cercava esses métodos que usavam processos sofisticados de automação em larga escala, onde seria gerado um número sem precedentes de novos compostos matrizes, resultando num aumento substancial de novas entidades lançadas no mercado por ano, diminuíram devido à baixa taxa de *hits* encontrados e aos elevados custos envolvidos ^[44]. Tal situação serviu como estímulo para o desenvolvimento de técnicas alternativas como o ensaio virtual (*virtual screening*).

O ensaio virtual envolve a análise de grandes coleções de compostos através de métodos computacionais com vista à identificação de uma pequena coleção que poderá ser

² *hit* – composto biologicamente ativo que excede certos limites de atividade em um ensaio biológico.

sintetizada para futuros testes biológicos ^[45]. O ensaio virtual é um método complementar ao HTS e acoplado com a biologia estrutural promove um aumento no sucesso de identificação de novos compostos matrizes (*lead compounds*) ^[46]. O alcance do ensaio virtual depende fortemente da quantidade e da qualidade de informação disponível sobre o sistema em consideração ^[47].

Em termos gerais, os métodos disponíveis para fazer as análises computacionais podem ser baseados na estrutura do ligante ou na estrutura do receptor ^[46].

Para os métodos baseados na estrutura do ligante, quando a estrutura do alvo é desconhecida, dados de atividade podem ser usados para construir um modelo farmacofórico através da identificação de características chaves tais como ligações de hidrogênio e grupos hidrofóbicos ^[48]. A estratégia é usar todas as informações fornecidas por um composto, ou conjunto de compostos, que são conhecidos por se ligarem a um alvo desejado e usar isto para identificar outros compostos que apresentem propriedades similares^[46]. Desse ensaio, evolverse-ão, muito provavelmente, várias moléculas candidatas. A transformação dessas moléculas candidatas em substâncias matrizes far-se-á, essencialmente, por compostos similares mais potentes. Os estudos de otimização por relações quantitativas estrutura-atividade (QSAR), absorção, distribuição, metabolismo, excreção e toxidez (ADME/Tox) serão então realizados e validados. Somente as estruturas das moléculas otimizadas prosseguirão para os estudos pré-clínicos com a inclusão de experimentos *in vivo*. As moléculas selecionadas dessa fase entram para os ensaios clínicos ^[49].

Quando a estrutura da proteína alvo é conhecida, normalmente por cristalografia de raio-X, RMN ou modelagem de homologia, métodos computacionais baseados na estrutura do receptor podem ser empregados. Os principais estágios, que devem ser seguidos na descoberta de novas substâncias químicas, iniciam-se com a obtenção da estrutura tridimensional do alvo. Após sua identificação, uma coleção virtual de compostos, em geral de domínio público, poderá ser avaliada frente a esses alvos e/ou modelos. O enfoque mais comum no ensaio virtual é a docagem molecular, que envolve a docagem de cada ligante no sítio ativo do alvo ou outro sítio que eventualmente module sua função. Esta docagem produz um modo de interação para cada composto do banco de dados, junto com a medida da qualidade do ajuste do composto no sítio ativo. As melhores moléculas candidatas serão, então, selecionadas e testadas contra o alvo especificado inicialmente, racionalizando as características estruturais mínimas que descrevam a potência ou afinidade ^[48, 49].

Ligantes podem ser planejados pelo método *de novo*, explorando características estruturais do sítio de ligação da proteína alvo. Novas estruturas moleculares podem ser

propostas, geralmente reunidas através de pequenas porções moleculares que se acomodam ao sítio de ligação ^[50]. Programas como o Ludi e o Pocket usam a estrutura da proteína para gerar sítios ou grades que caracterizam posições favoráveis que os átomos do ligante podem ocupar ^[51]. Embora seja possível o planejamento de novos compostos (não apenas em uma coleção pré-definida) através do emprego do método *de novo*, nem todos os compostos poderão ser sintetizados ^[52].

Os ligantes podem ser obtidos a partir de bancos de dados comerciais que possibilitam o uso de coleções de compostos, como Instituto Nacional do Câncer (NCI – *National Cancer Institute*), ACD (*Available Chemical Directory*) e o Spresi ^[51]. Recentemente, cerca de 1 milhão de moléculas comercialmente acessíveis tornou-se disponível através do banco de dados gratuito, o ZINC. Essas moléculas são utilizáveis por muitos programas populares de docagem ^[53].

O espaço químico pode ser comparado ao universo em sua extensão, onde os compostos químicos ocupam o espaço ao invés das estrelas. O espaço químico virtual contém aproximadamente 10^{100} possíveis moléculas, sendo que muitas dessas moléculas não podem ser facilmente sintetizadas, são indesejadas ou impraticáveis. Nem todos os compostos biologicamente ativos têm as propriedades físico-químicas desejadas para se tornarem fármacos. Um composto pode ser muito lipofílico para ser absorvido por via oral, ser muito polar para cruzar a parede gastrointestinal ou ter grupos funcionais que são atacados por enzimas no organismo. Na prática, a coleção virtual pode incluir cerca de 10^{15} possíveis compostos disponíveis para o planejamento de fármacos ^[54, 55].

Vários filtros podem ser aplicados no ensaio virtual com o objetivo de melhorar a qualidade das coleções de compostos. A aplicação dos filtros resulta na remoção de compostos indesejados de forma a maximizar a possibilidade de que um novo composto alcance os testes clínicos. Alguns filtros podem remover compostos quimicamente reativos, com grupos tóxicos, e compostos que não contenham átomos além de C, N, O, S, H, P, Cl, Br, F e I. ^[51, 56].

Com base em métodos experimentais e computacionais, Lipinski anunciou um conjunto de regras para a absorção oral de substâncias ativas. De acordo com as essas regras, um fármaco é pouco absorvido provavelmente porque tem:

- Número de átomos capazes de participar em doações de ligações de hidrogênio maior do que 5.
- Massa molar maior que 500 Da.
- Logaritmo do coeficiente de partição ($\log P$) maior que 5.

- Mais de 10 receptores de ligações de hidrogênio.

A “regra dos cinco” (RoF) é assim chamada porque todos os parâmetros de classificação são múltiplos de cinco. Embora existam exceções, a regra dos cinco tem se mostrado útil porque ela fornece um caminho fácil para estimar a absorção oral de substâncias ativas^[57]. Não obstante, essa regra tem sido expandida^[51].

Considerando que a descoberta de novos fármacos é um processo longo, atualmente em torno de 12 anos, a química medicinal abrange três fases: descoberta, otimização e desenvolvimento do fármaco.

A escolha do alvo terapêutico (receptores, enzimas, DNA, receptores transmembrânicos, ácidos nucleicos, etc.) precede a fase de descoberta (ou identificação) de ligantes que interajam com o alvo. Após sua otimização, tais compostos são chamados de compostos matrizes (*lead compounds*), e eles podem originar-se da química orgânica sintética ou de produtos naturais.

A fase de otimização trata da melhora da estrutura matriz. Nesta fase, a identificação de grupos farmacofóricos presentes na estrutura matriz será de fundamental importância para que as modificações moleculares sejam realizadas de forma produtiva. O processo de otimização leva principalmente em consideração o aumento de potência, seletividade e toxidez. Suas características são o estabelecimento e análise da relação estrutura e atividade, pois permite o entendimento do modo de ação em nível molecular.

A avaliação dos parâmetros farmacocinéticos, aqueles que governam os fatores de absorção, distribuição, metabolismo e excreção (ADME) do fármaco, são sistematicamente praticados nas primeiras fases do desenvolvimento, de maneira que se eliminem candidatos insatisfatórios. Um outro parâmetro importante que deve ser levado em consideração é o farmacodinâmico, que é a fase de maior interesse dos químicos medicinais, pois regem as interações responsáveis pelo reconhecimento molecular do fármaco pelo biorreceptor. Dessas interações pode resultar a resposta terapêutica desejada. Essa fase aborda diretamente a natureza e a qualidade das interações do fármaco com seu alvo biológico.

A fase de desenvolvimento das substâncias bioativas tem como proposta a melhora das propriedades farmacocinéticas e o ajuste das propriedades farmacêuticas (como por exemplo a formulação química), para torná-las mais adequadas ao uso clínico. O composto passa por testes em animais e ensaios clínicos, onde fatores que ainda não foram considerados, tais como toxicidade, biodisponibilidade e permeabilidade podem ser considerados, no que diz respeito, principalmente, ao uso oral dos medicamentos ^[42, 58]. Em verdade, mais

recentemente essa etapa tem sido incluída o mais cedo possível na gênese planejada de fármacos ^[59].

A Síndrome da Imunodeficiência Adquirida (AIDS) foi a primeira grande epidemia causada por um patógeno, e surgiu justamente no período correspondente ao desenvolvimento das ciências farmacêuticas modernas. A ausência completa de agentes capazes de tratar a doença tornou necessário o desenvolvimento de agentes terapêuticos completamente novos. Esta necessidade, juntamente com o rápido progresso da estrutura e do ciclo de vida virais, levou a um desenvolvimento sem precedentes de novos fármacos agindo sobre diversas proteínas virais, como por exemplo a transcriptase reversa, integrase e protease. A descoberta e o desenvolvimento de inibidores de HIV-protease constituíram num sucesso da farmacologia e da biologia estrutural modernas, sendo um dos melhores exemplos de planejamento de fármacos baseados na estrutura do receptor (*structure-based drug design*) e utilizando-se de metodologias como cristalografia, RMN e modelagem molecular de forma a guiar a síntese de novos agentes terapêuticos ^[60].

A introdução de novas tecnologias tornou a química medicinal mais ampla em sua concepção, ampliando seu caráter interdisciplinar. Hoje a química medical engloba áreas como a farmacologia, a bioquímica, a química orgânica sintética e a química computacional, entre outras, na busca do desenvolvimento e/ou descoberta de novos fármacos ^[60].

Atualmente, as maiores esperanças no campo das ciências médicas estão centradas nos projetos Genoma e Proteoma. O maior desafio procedente do genoma humano, e de outros genomas em estudo, será a elucidação estrutural de novos alvos moleculares, principalmente proteínas, enzimas e receptores que contribuirão para o aparecimento de novos fármacos. Estas estruturas serão à base da revolução da “medicina do futuro”, em que a compreensão dos fenômenos biológicos, em nível molecular, terá um papel cada vez mais relevante, o que implicará em avanços significativos das técnicas de diagnóstico e tratamento. Com isso, espera-se que seja possível iniciar os tratamentos clínicos no feto ou na primeira infância, muito antes do surgimento dos primeiros sintomas ^[61].

2. MÉTODOS DE CÁLCULOS UTILIZADOS

2.1. A EQUAÇÃO DE SCHRÖDINGER

A compreensão da natureza das propriedades dos átomos, moléculas e sistemas mais complexos estão diretamente relacionadas à distribuição e o comportamento das partículas microscópicas que os constitui. Diversas evidências experimentais sugeriram que a estrutura dos átomos e moléculas não poderiam ser explicadas por princípios baseados na mecânica clássica. Experimentos como a radiação do corpo negro, capacidade calorífica e espectros atômicos e moleculares, sugeriam que os processos de absorção ou emissão de energia só poderiam ser explicados admitindo-se que a energia envolvida nesses processos fosse quantizada e não contínua, como prevê a mecânica clássica^[62]. Em 1925, houve um grande avanço na teoria quântica, quando lhe foi incorporada a natureza ondulatória do elétron. Essa teoria aperfeiçoada, que se chama mecânica quântica ou mecânica ondulatória, foi desenvolvida principalmente pelo físico alemão Werner Heisenberg, pelo físico austríaco Erwin Schrödinger e pelo físico inglês Paul Adrien Maurice Dirac^[63].

O referencial de toda descrição quântica de um sistema qualquer está na utilização da equação de Schrödinger. Na sua forma independente do tempo, esta equação é escrita como:

$$\hat{H} \Psi = E \Psi \quad (1)$$

A equação de Schrödinger é uma equação de autovalores: quando um operador atua sobre uma função e resulta na mesma função multiplicada por uma constante. A função de onda Ψ é a autofunção do operador hamiltoniano (H) correspondente ao autovalor E (energia do sistema). A obtenção da função de onda exata implica em uma descrição completa do sistema, ou seja, é possível calcular qualquer propriedade do mesmo, já que se tem toda informação necessária para a predição das propriedades atômicas^[64, 65].

Para átomos ou moléculas, a aparente simplicidade da equação desaparece, pois a equação acima corresponde a uma abreviação de diversos termos. O termo H corresponde a um operador que permite obter informações sobre a energia do sistema. Este operador é conhecido como operador hamiltoniano, dado por:

$$\hat{H} = T + V_{e-e} + V_{e-n} + V_{n-n} \quad (2)$$

onde T representa a energia cinética dos elétrons e dos núcleos, V_{e-e} é a energia potencial devido a interação elétron-elétron, V_{e-n} representa a energia potencial de interação elétron-núcleo e V_{n-n} representa a interação núcleo-núcleo.

Na prática, a equação de Schrödinger não é fácil ou mesmo possível de ser resolvida, possui solução exata somente para o átomo de hidrogênio ou cátions hidrogenóides. Para átomos polieletrônicos e nas moléculas não possui solução exata, necessitando de aproximações. A principal aproximação feita no sentido de resolver a equação de Schrödinger consiste na aproximação de Born e Oppenheimer^[66], que permite a separação dos movimentos nucleares e eletrônicos, simplificando, assim, o tratamento do sistema. Esta aproximação considera o fato que, sendo os núcleos mais pesados que os elétrons, movem-se devagar quando comparado ao movimento dos elétrons. Assim, como uma boa aproximação pode-se considerar os elétrons em uma molécula movendo-se em um campo fixo formado pelos núcleos. Desse modo, o termo correspondente a energia cinética dos núcleos pode ser desprezado e a energia potencial de repulsão núcleo-núcleo pode ser considerada como constante^[44]. O hamiltoniano para a energia eletrônica, então, é dado por

$$H_e \Psi_e = E_e \Psi_e \quad (3)$$

Mesmo considerando esta simplificação, a equação de Schrödinger só pode ser resolvida exatamente para sistemas monoelétrônicos. Então, recorreu-se a outros métodos de aproximação, o método de Hartree-Fock^[67], o método de Hartree-Fock-Roothaan^[68], e o método DFT (*Density Functional Theory* – Teoria do Funcional da Densidade), que possibilitam resolver satisfatoriamente, mesmo de forma aproximada, a equação de Schrödinger para sistemas atômicos e moleculares polieletrônicos.

Nesta tese foi utilizado o método DFT para a obtenção das propriedades moleculares (geométricas, eletrônicas e outras propriedades correlacionadas).

2.2. TEORIA DO FUNCIONAL DE DENSIDADE (DFT)

A entidade básica na Teoria do Funcional da Densidade é a densidade eletrônica - $\rho(r)$, que descreve a distribuição de carga em uma molécula. A solução formalmente exata da equação de Schrödinger para a densidade foi dada por Hohenberg e Kohn em 1964, e por este trabalho, Walter Kohn veio a ser agraciado com o prêmio Nobel de Química em 1998^[69].

A Teoria do Funcional da Densidade, desde os trabalhos de Hohenberg e Kohn, apresenta-se como alternativa para a resolução da equação de Schrödinger, e a metodologia tem sido empregada para sistemas moleculares encontrados em pesquisas farmacêuticas, agroquímicas e biotecnológicas; na ciência de materiais e polímeros; em pesquisas com complexos organometálicos e aglomerados típicos de catálise, superfície e estado sólido; em eletroquímica e microeletrônica^[70].

Com o DFT pode-se substituir vários métodos computacionais, até então baseados em cálculos Hartree-Fock ou pós Hartree-Fock. A grande vantagem da metodologia do Funcional da Densidade sobre os métodos *ab initio* padrão (métodos baseados nas equações de Hartree-Fock-Roothan (HFR) está no ganho em velocidade computacional e espaço em memória. Em contraste com métodos semi-empíricos, o hamiltoniano na DFT é bem definido e suas características não são obscurecidas por freqüentes e injustificáveis aproximações no procedimento computacional. Nas derivadas das equações do DFT, nenhum parâmetro em princípio necessita ser ajustado ou determinado empiricamente, sob este aspecto a DFT pode ser considerada como uma teoria de natureza *ab initio*^[70].

As primeiras aplicações da densidade como meio para obtenção da energia de sistemas com mais de um elétron resultaram nos trabalhos de Thomas e Fermi^[69]. Thomas em 1927 e Fermi em 1928 propuseram um novo esquema baseado na densidade eletrônica do sistema, $\rho(r)$. Esta idéia original substitui a complicada função de onda de N elétrons e a equação de Schrödinger pela densidade eletrônica muito mais simples. Este modelo assume que as propriedades eletrônicas dos sistemas reais são similares aquelas dos gases de férmions, *i.e.*, um ensemble clássico de partículas não interagentes. Apesar desta aproximação ter tido limitado sucesso em reproduzir as propriedades dos sistemas reais, ela foi o protótipo da DFT, que focaliza a densidade eletrônica como variável fundamental^[71].

Em 1964, Hohenberg e Kohn (HK), em seus trabalhos sobre gás de elétrons não-homogêneos, propuseram que a energia molecular no estado fundamental, e todas as propriedades eletrônicas moleculares, pudessem ser determinadas unicamente pela densidade

de probabilidade eletrônica $\rho(x, y, z)$. A energia do estado fundamental é um funcional de ρ e é escrita como $E_0 = E_0[\rho]$ ^[66].

O teorema de Hohenberg e Kohn não diz como é possível encontrar E_0 a partir de ρ ou como encontrar ρ sem encontrar Ψ . Este problema foi equacionado por Kohn e Sham nos anos de 1965 e 1966^[72]. Eles propuseram que no estado fundamental a energia puramente eletrônica E_0 de uma molécula com muitos elétrons no estado fundamental e com densidade de probabilidade eletrônica ρ , é dada pela equação 4:

$$E[\rho] = T[\rho] + V_{ee}[\rho] + V_{Ne}[\rho] = T[\rho] + V_{ee}[\rho] + V_{ext}[\rho] = F[\rho] + V_{ext}[\rho] \quad (4)$$

F é um funcional universal da densidade, que não depende do potencial externo, atua sobre os elétrons e possui as contribuições da energia cinética e da repulsão intereletrônica. O último termo da Equação (4) representa a energia de interação dos elétrons com o potencial externo v , o que inclui a interação com os núcleos.

$$V_{ext}[\rho] = \int v(r)\rho(r)dr \quad (5)$$

Uma importante idéia introduzida por Kohn e Sham é usar o conceito de um sistema de referência de partículas independentes, ou seja, que não interagem. Eles reescreveram a equação (4) tornando explícita a repulsão elétron-elétron de Coulomb, e definindo uma nova função universal $G[\rho]$ ^[73].

$$E[\rho] = G[\rho] + \frac{1}{2} \iint \frac{\rho(r)\rho(r')}{(r-r')} drdr' + \int v(r)\rho(r)dr \quad (6)$$

onde

$$G[\rho] = T[\rho] + E_{xc}[\rho] \quad (7)$$

O termo $T[\rho]$ na equação (7) é a energia cinética de um sistema de elétrons não interagentes com densidade $\rho(r)$ e $E_{XC}[\rho]$ contém a energia troca e correlação de um sistema interagente com densidade $\rho(r)$. A parte que contém a correção da energia cinética desse sistema está nesse funcional de correlação, $E_{XC}[\rho]^{[69]}$. O orbital de um elétron pode ser utilizado permitindo calcular precisamente a energia cinética via procedimento auto-consistente. O termo de energia cinética T é dado pela seguinte expressão:

$$T = \sum_{i=1}^N \left\langle \psi_i \left| -\frac{1}{2} \nabla^2 \right| \psi_i \right\rangle \quad (8)$$

Os orbitais são auto-funções do Hamiltoniano Kohn-Sham efetivo de um elétron, é e dado pela Eq. (9):

$$H_{KS} = -\frac{1}{2} \nabla^2 + v_{ef}(r) \quad (9)$$

onde o potencial efetivo, é definido como segue

$$v_{ef}(r) = v(r) + \int \frac{\rho(r')}{|r-r'|} dr' + v_{XC}(r) \quad (10)$$

onde o potencial v_{XC} é definido como a derivada funcional da energia total de troca e correlação:

$$v_{XC} = \frac{\partial E_{XC}(\rho)}{\partial \rho(r)} \quad (11)$$

No estado fundamental, a densidade eletrônica, ρ , pode ser encontrada a partir das Ψ_i 's por

$$\rho(r) = \sum_i^N |\psi_i(r)|^2 \quad (12)$$

Portanto, os orbitais de Kohn-Sham (KS) são obtidos a partir da equação de Schrödinger de um elétron

$$\left(\frac{1}{2} \nabla^2 + v_{\text{ef}} \right) \psi_i = \epsilon_{\text{KS}} \psi_i \quad (13)$$

As Equações (10), (11) e (12) representam o esquema Kohn-Sham auto-consistente, uma vez que o potencial efetivo, $v_{\text{ef}}(\mathbf{r})$, depende da densidade eletrônica $\rho(\mathbf{r})$ e as equações de Kohn-Sham são resolvidas através de um procedimento auto-consistente^[73].

O esquema Kohn-Sham permite calcular a densidade eletrônica do estado fundamental. Todas as outras propriedades do sistema podem ser calculadas desde que os funcionais da densidade eletrônica apropriados sejam conhecidos.

A parte mais difícil para resolver as equações de Kohn-Sham é determinar o funcional de troca e correlação, uma vez que a forma exata analítica do potencial de troca e correlação não é conhecida^[73].

A busca de uma expressão exata de E_{xc} tem encontrado grandes dificuldades e na atualidade segue sendo o objetivo da teoria da funcional densidade. Entre as várias aproximações que têm surgindo, uma muito simples foi proposta por Kohn e Sham em 1965 e se chama de aproximação da densidade local (LDA). Outro método é o chamado funcional de troca de Slater, X_{α} , proposto por Slater e que apareceu das simplificações das equações de Hartree-Fock, onde se despreza a correlação^[74-76]. Becke em 1988 propôs um funcional de troca do tipo Slater incluindo as correções de densidade ao gradiente da densidade denotado por B88^[77].

Também existem vários funcionais de correlação, tais como:

- Funcional de correlação de Vosko, Wilk e Nusair (VWN), de 1980, que se refere à correlação LSD (Local Spin Density)^[78]
- O funcional de correlação de Lee, Yang e Parr (LYP), que inclui o termo local e não local^[79-80]
- As correções ao gradiente de Perdew de 1981 e 1986 (P81 e P86)^[81]
- Perdew e Wang de 1991 (PW91)^[82]

Os funcionais de correlação devem combinar-se com os funcionais de troca, sendo estes largamente utilizados. Nesta tese, foi empregado o funcional híbrido B3LYP (ou Becke3LYP – o número 3 indica que este funcional possui três parâmetros). Este funcional é definido por:

$$E_{XC}^{B3LYP} = (1 - a_0 - a_x)E_X^{LSDA} + a_0E_X^{HF} + a_xE_X^{B88} + (1 - a_c)E_C^{VWN} + a_cE_C^{LYP} \quad (14)$$

onde o termo E_X^{LSDA} denota o funcional de troca baseado no modelo do gás uniforme de elétrons, E_X^{HF} é o termo de troca nas equações Hartree-Fock, o termo E_X^{B88} é o funcional de Becke de 1988, E_C^{VWN} denota o funcional de correlação desenvolvido por Vosko-Wilk-Nusair no procedimento LSDA e E_C^{LYP} utiliza o funcional de correlação de LYP. Os valores das constantes obtidas pelo ajuste com os valores experimentais das energias de atomizações moleculares são: $a_0=0,20$, $a_x=0,72$ e $a_c=0,81$ ^[66].

3. MÉTODOS PARA ANÁLISE DE DADOS

Um problema comum aos nossos dias refere-se ao acesso, manipulação e quantidade de informações disponíveis. Não obstante, mais importante do que seu conhecimento está a necessidade de sua organização racional^[43]. Até a década de 60, a maioria dos problemas químicos eram resolvidos através de métodos experimentais, tais como titulação, pesagem, precipitação e reações com reagentes específicos. Hoje, existem várias técnicas como ressonância magnética nuclear (RMN), infravermelho (IR), ultravioleta (UV), fluorescência de raio-X e várias técnicas cromatográficas, permitindo-nos uma rápida obtenção de um extenso conjunto de dados.

À medida que cresce a quantidade de dados e as correlações entre eles, cresce também a dificuldade de se obter informações interessantes, e a avaliação das variáveis que mais contribuem para se extrair a informação desejada torna-se uma tarefa cada vez mais árdua.

Para a manipulação apropriada dos dados multivariados, deve-se armazená-los numa matriz de dados, A_{ij} , onde i (índice de linhas) corresponde à amostra e j (índice das colunas) corresponde a variável^[83]. Também se deve levar em consideração que para a utilização da habilidade humana é necessário representar os dados em três dimensões ou duas.

Interpretar e analisar dados objetivando extrair o máximo de informações tem sido possível utilizando a análise multivariada. Desta forma, o emprego de técnicas multivariadas tem encontrado inúmeras aplicações no tratamento de dados químicos, como por exemplo, na determinação de teores de minerais em sucos de frutas^[84], em trabalhos de oceanografia^[85], até mesmo de outras áreas não ligadas à química como economia^[86] e psicologia^[87].

Na Química, o método que estuda a aplicação de métodos estatísticos multivariados^[88] é chamado Quimiometria. Sendo que suas aplicações enfocam dois temas centrais: a escolha das condições experimentais ótimas para uma investigação, também conhecido como planejamento experimental, e a análise de dados multivariados, de forma a extrair deles o máximo possível de informação química.

3.1. PRÉ-PROCESSAMENTO DOS DADOS

O processamento de um conjunto de dados inicia-se, geralmente, para um determinado conjunto de variáveis, pelo **pré-processamento** das mesmas. Esse estágio pode incluir a redução de variáveis e detecção de pontos deslocados (*outliers*). Para fazer um pré-processamento dos dados precisamos de grandezas estatísticas, as mais usadas são a média e o desvio padrão, respectivamente. A média indica a sua localização, ou tendência central, e é simplesmente a soma de todos os valores divididos pelo número total de elementos no conjunto. O valor médio numa amostra é normalmente indicado por uma barra colocada sobre o símbolo que representa. Já a variância é uma espécie de média dos quadrados dos desvios. As Equações (15) e (16) mostram os conceitos de média e variância amostral:

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (15)$$

$$V(x) = s^2 = (1/N - 1) \sum_{i=1}^N (x_i - \bar{x})^2 \quad (16)$$

onde N é o número total de medidas realizadas, x_i é a *i-ésima* medida, $\bar{x}$ é a média e $V(x) = s^2$ é a variância amostral.

Para que as medidas de dispersão e de posição tenham as mesmas unidades, costuma-se substituir a variância pela sua raiz quadrada, que é chamada de desvio padrão. O desvio padrão geralmente é utilizado para definir intervalos em torno da média, e pode ser definido por:

$$s = (s^2)^{1/2} \quad (17)$$

Outra grandeza utilizada em análises estatísticas é a covariância, que é uma tendência de duas variáveis se desviarem de forma mais ou menos conjunta em relação às respectivas médias^[89]. A covariância é definida por:

$$C = \frac{1}{N-1} \sum (x_i - \bar{x})(y_i - \bar{y}) \quad (18)$$

onde x_i e y_i correspondem as variáveis em estudo, $\bar{x}$ e $\bar{y}$ são as respectivas médias e N é o número de medidas.

O valor da covariância depende da escala utilizada para medir x e y , o que dificulta comparar o grau de associação estatística existente em diferentes pares de variáveis. Para eliminar esse problema, aplica-se um fator de escala dividindo-se cada desvio individual pelo desvio padrão da variável correspondente, obtendo-se assim uma espécie de covariância normalizada, chamada de coeficiente de correlação das duas variáveis. Em decorrência dessa definição, o coeficiente de correlação de qualquer par de variáveis fica restrito ao intervalo $[-1, 1]$ e correlações de diferentes pares de variáveis ficam na mesma escala, podendo ser comparadas^[89]. O coeficiente de correlação, r_{xy} , mede a associação linear existente entre duas variáveis e pode ser obtido pela seguinte expressão:

$$r_{xy} = [1 / (N - 1)] \sum [(x_i - \bar{x}) / s_x] [(y_i - \bar{y}) / s_y] \quad (19)$$

onde s_x e s_y são os respectivos desvios padrão das variáveis^[89].

3.2. ESCALAMENTO DOS DADOS

Geralmente, o primeiro passo do pré-processamento envolve o escalamento. O objetivo do escalamento dos dados é dar a todas as variáveis a mesma importância, ou seja, o mesmo peso na análise de dados. Isto se consegue por meio da transformação das variáveis de modo que o valor médio e a variância (o quadrado do desvio padrão) sejam 0 e 1, respectivamente. Esse tipo de escalamento assegura que as influências relativas das diferentes variáveis sobre os cálculos sejam independentes das unidades dessas variáveis. Alguns métodos de escalamento são^[83]:

- ✓ Dados centrados na média
- ✓ Escalamento pela variância
- ✓ Autoescalamento
- ✓ Escalamento pela variação (*range*)

Neste trabalho, o autoescalamiento foi o pré-processamento utilizado. Depois de autoescalar, faz-se a redução de variáveis. Esta etapa permite eliminar as variáveis que não sejam relevantes para a classificação desejada através de algoritmos que irão nos permitir reduzir a dimensão dos dados ou organizá-los numa estrutura que facilite a visualização do conjunto de forma global. A seguir faz-se uma análise exploratória de dados que visa agrupar os dados em conjuntos com propriedades semelhantes, e também visa manter o espaço dimensional acessível à visualização e compreensão humana.

3.3. MÉTODOS UTILIZADOS

A geração de uma determinada estrutura de forma tridimensional não está, necessariamente, na conformação mais estável, pois ocorrem distorções na molécula como a formação desfavorável de comprimentos e ângulos de ligações e ângulos diédricos. Átomos não ligados também podem interagir em uma mesma região do espaço e provocar repulsões estereoquímicas e eletrostáticas. Para corrigir estas distorções, as moléculas são otimizadas pelo processo de minimização de energia^[90].

O processo de otimização de geometria envolve a minimização da energia potencial molecular. Esta minimização realiza uma busca pelo mínimo da função de energia potencial através de cálculos de derivadas desta função versus as coordenadas^[66]. As mudanças nas conformações moleculares podem ser estimadas como deslocamentos sobre a superfície multidimensional que descreve a relação entre a energia potencial e a geometria da molécula. Cada ponto sobre a superfície de energia potencial representa o potencial de energia de uma única conformação. As conformações estáveis da molécula correspondem ao mínimo local sobre esta superfície de energia^[91].

As energias conformacionais calculadas por métodos teóricos podem ser por:

- Mecânica molecular
- Química quântica

Mecânica molecular é o método para o cálculo das geometrias e das energias conformacionais das moléculas que usa os conceitos da mecânica clássica. A molécula é considerada como uma coleção de átomos mantidos por forças clássicas. As equações obtidas por mecânica molecular consideram apenas os núcleos dos átomos e não incluem os elétrons nos cálculos. O processo de mecânica molecular promove a modificação dos ângulos e

comprimentos de ligação originais e fornece novas conformações com os correspondentes cálculos de energia^[90-91].

Os métodos de química quântica são baseados no uso da Equação de Schrödinger, e podem ser subdivididos em métodos semi-empíricos e *ab initio*. Os métodos *ab initio* usam equações exatas, sem aproximações. Já nos semi-empíricos, as integrais são ajustadas a partir de dados experimentais de forma que simplificam os cálculos computacionais.

As otimizações das geometrias das 28 quinonas estudadas foram feitas usando o método *ab initio* da Teoria do Funcional da Densidade (DFT – *Density Functional Theory*) com o funcional de troca-correlação B3LYP. O B3LYP é um método com custo computacional similar ao Hartree-Fock e com resultados entre o método de perturbação de segunda (MP2) e terceira (MP3) ordens de Møller-Pleset^[71]. A escolha pelo método DFT foi feita porque esse método leva a excelentes resultados para a análise das geometrias e das energias^[92, 93]. O conjunto de base que foi utilizado foi o 6-31G*, implementado no programa Gaussian 98^[94].

As estruturas das moléculas mostradas na Figura 26 foram submetidas a uma busca conformacional e a partir da conformação de menor energia obtida para cada composto foram calculadas várias propriedades tais como energia total, energia do orbital molecular ocupado de maior energia (HOMO – *Highest Occupied Molecular Orbital*), energia do primeiro orbital molecular abaixo do HOMO (HOMO-1), energia do orbital molecular desocupado de menor energia (LUMO – *Lowest Unoccupied Molecular Orbital*), ordens de ligação, cargas atômicas sobre todos os átomos que incluem o esqueleto básico das quinonas estudadas (Figura 25), soma das cargas atômicas na região A e B, dureza, moleza, momento dipolar, volume e ângulos de torção formados pelos átomos 1-4.

As cargas atômicas foram calculadas usando o método CHELPG do programa GAUSSIAN 98^[94]. O método CHELPG permite avaliar o potencial eletrostático da molécula a partir de cargas atômicas pontuais de modo que represente o melhor potencial molecular para um conjunto de pontos definidos em volta da molécula^[95].

O programa utilizado para calcular o coeficiente de partição octanol/água para os compostos estudados foi o XlogP^[96].

O programa Dragon 2.1^[97] foi utilizado para o cálculo de uma grande quantidade de propriedades atômicas e moleculares. As propriedades calculadas pelo programa Dragon podem ser agrupadas em diversos tipos de descritores tais como geométricos e topológicos.

As análises estatísticas (PCA, HCA, KNN e SIMCA) foram realizadas utilizando o programa Pirouette 3.11^[98] e a análise de discriminante por passo (SDA) foi feita com o

programa MINITAB 10.1^[99]. A análise empregando Redes Neurais Artificiais (ANN) foi feita utilizando o algoritmo de retropropagação de erros (*backpropagation*) com o programa Matlab 6.1^[100].

Antes da realização das análises estatísticas (PCA, HCA, SDA, KNN, SIMCA e ANN), todas as variáveis foram autoescaladas de forma que cada variável tenha a mesma importância nas análises.

Visando à redução de variáveis e a detecção de pontos deslocados (*outliers*), foram utilizados o Peso de Fisher e a matriz de correlação, calculados com a planilha eletrônica do Microsoft Excel. A redução do número de variáveis se fez necessária, pois o conjunto inicial de dados apresentava 223 variáveis e esta quantidade de informação dificulta as análises de dados. A primeira redução do número de variáveis foi executada através do peso de Fisher^[83, 101], o qual fornece a importância relativa de cada variável.

Depois da análise dos valores de Peso de Fisher do conjunto inicial de 223 variáveis, somente 12 variáveis foram selecionadas, ou seja, aquelas que apresentaram valores de peso de Fisher mais significativos. O critério de seleção das variáveis mais significativas levou em consideração as variáveis que apresentaram Peso de Fisher acima de 0,8. Estas variáveis são aquelas que possuem maior habilidade de discriminação (ou separação) entre as moléculas ativas e inativas.

A partir das 12 variáveis selecionadas usando o Peso de Fisher, foi calculada a matriz de correlação entre essas variáveis (Tabela 1). A partir da Tabela 1, pode-se observar que algumas variáveis são altamente correlacionadas umas com as outras, isto é, variáveis consideradas altamente correlacionadas são aquelas que apresentam coeficientes de correlação acima de 0,7. De acordo com os resultados apresentados na Tabela 1, somente 8 variáveis podem ser consideradas importantes na separação entre os compostos que apresentam atividade tripanossomicida. As variáveis são: T₅ (ângulo de torção formado pelos átomos 1, 2, 17 e 18), QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A), VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B) e HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO), BIC (índice de Balaban), E_{2v} (variável relativa aos volumes de van der Waals atômicos), E_{1e} (variável relativa às eletronegatividades atômicas) e E_{2p} (variável relativa à polarizabilidades atômicas).

Tabela 1 - Tabela de correlação entre as variáveis selecionadas utilizando o Peso de Fisher

HOMO-1	T ₁	T ₂	T ₃	T ₅	OL3	BIC	E _{2v}	E1e	E2p	VOLS2	QTS1
HOMO-1	0,164	-0,368	0,402	-0,404	-0,466	0,533	-0,216	-0,56	-0,112	0,504	-0,612
T ₁		-0,851	0,566	-0,567	-0,387	0,372	-0,253	-0,429	-0,173	-0,37	0,023
T ₂			-0,886	0,881	0,671	-0,374	0,48	0,556	0,344	0,47	-0,405
T ₃				-0,981	-0,635	0,237	-0,437	-0,463	-0,297	-0,368	0,584
T ₅					0,718	-0,189	0,499	0,513	0,344	0,504	-0,612
OL3						-0,218	0,602	0,582	0,428	0,679	-0,523
BIC							-0,038	-0,219	0,042	0,151	-0,019
E _{2v}								0,456	0,959	0,448	-0,618
E1e									0,331	0,323	-0,566
E2p										0,323	-0,566
VOLS2											-0,381

Depois de realizadas as etapas de pré-processamento dos dados e de redução das variáveis, partiu-se para a análise estatística dos dados, que tem como objetivo extrair o máximo de informação dos dados. As técnicas utilizadas nesse trabalho foram:

- ✓ Análise de Componentes Principais (PCA)
- ✓ Análise Hierárquica de Agrupamentos (HCA)
- ✓ Análise Discriminante por Passo (SDA)
- ✓ Método dos k-ésimos vizinhos mais próximos (KNN)
- ✓ Modelos independentes de similaridade utilizando componentes principais (SIMCA)
- ✓ Redes Neurais Artificiais (ANN)

Estas técnicas, que foram empregadas neste trabalho, são utilizadas em conjunto, pois uma complementa a outra, o que possibilita uma melhor visualização das classes de compostos obtidos.

4. MÉTODOS QUIMIOMÉTRICOS UTILIZADOS

4.1. ANÁLISE DE COMPONENTES PRINCIPAIS (PCA)

A técnica da componente principal foi primeiro descrito por Karl Pearson em 1901^[102]. Ele acreditava que essa técnica poderia dar solução correta para alguns problemas da biomedicina. A descrição prática dos métodos de cálculo só veio mais tarde com Hotelling em 1933. Os cálculos, na época, eram extremamente exaustivos e eram feitos à mão. Somente com o advento dos computadores é que a técnica adquiriu uso generalizado. Na química, a PCA foi introduzido por Malinowski por volta de 1960 sob o nome de análise de fatores principais, e após 1970, um grande número de aplicações já tinham sido publicadas^[103].

A análise de componentes principais é uma técnica de estatística multivariada que tem por objetivo reduzir a dimensão dos dados originais permitindo a fácil visualização das informações mais importantes em um número menor de fatores, ou componentes principais.

A técnica da PCA agrupa variáveis altamente correlacionadas em um número menor de novas variáveis não correlacionadas chamadas de componentes principais (CP). Essas novas variáveis devem ser ortogonais entre si, ou seja, há um novo sistema de coordenadas (Figura 17). Cada componente principal é combinação linear das variáveis originais, de forma que este novo conjunto de variáveis é construído em ordem decrescente da quantidade de variância que descrevem (a primeira componente principal descreve a maior variância nos dados que a segunda, e assim por diante)^[104-106].

Figura 17 - Projeção das componentes principais.

De uma forma geral e bem simplificada, as combinações lineares das m -variáveis originais que geram cada componente principal podem ser representadas pela Equação (20):

$$PC_i = a_{i1}v_1 + a_{i2}v_2 + \dots + a_{im}v_m \quad (20)$$

Para esta equação, v_j (para $j=1, 2, \dots, m$) são variáveis originais e a_{ij} (para $j = 1, 2, \dots, m$) são os coeficientes que medem a importância de cada variável na i -ésima componente principal (PC_i), ou seja, o peso que cada variável tem naquela combinação linear. Este peso é o co-seno do ângulo entre o eixo da componente principal e o eixo da variável original, que sempre estará entre 1 e -1 . Se o coeficiente estiver próximo de zero, menor a influência da variável naquela componente principal, mas, se estiver próximo de ± 1 maior a influência que esta determinada variável tem na descrição desta componente principal^[107-108].

A PCA fornece uma aproximação de uma tabela de dados, a matriz $\mathbf{X}$, em termos do produto de duas pequenas matrizes, $\mathbf{T}$ e $\mathbf{P}'$. Estas duas matrizes, $\mathbf{T}$ e $\mathbf{P}'$, absorvem os dados essenciais de $\mathbf{X}$. A matriz $\mathbf{T}$ corresponde aos escores, os quais são definidos como as coordenadas das amostras no novo sistema de coordenadas, e a matriz $\mathbf{P}'$ corresponde aos

loadings, que indicam a importância (peso) de cada variável original no novo conjunto de variáveis, que podem ser organizados na forma de um gráfico conforme a Figura 18. Esses vetores são escritos usualmente em ordem decrescente de autovalores. Portanto, uma análise da PCA também estima a estrutura de correlação das variáveis. Na prática, as análises da PCA são feitas sempre observando as informações dos dois gráficos, *loadings* e escores, e há liberdade de se excluir variáveis, buscando melhorar resultados. A PCA permite que se encontre correlações entre amostras e/ou variáveis e distinguir categorias nas quais novas amostras possam ser adequadas^[106].

Figura 18 - Gráficos de escores (A) e loadings (B) de um sistema de dados bidimensional.

4.2. ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)

Essa técnica agrupa compostos que pareçam similares entre si, baseando-se em suas proximidades. Os valores de similaridade, S_{ij} , são calculados segundo a Equação (21):

$$S_{ij} = 1 - \frac{d_{ij}}{d_{ij}(\max)} \quad (21)$$

onde d_{ij} é a distância entre os pontos i e j e $d_{ij}(\max)$ é a maior distância na matriz distância. Assim, os dois pontos mais distantes no conjunto de dados têm $s=0$ e pontos idênticos apresentam similaridade de 1^[101].

Cada composto é, inicialmente, considerado como sendo um único agrupamento. A matriz de similaridade é explorada buscando-se o maior valor, cujos pontos correspondentes são agrupados e tratados como um único ponto. As respectivas linhas e colunas referentes aos pontos antigos são removidas e o novo agrupamento é adicionado, como sendo a média de dois pontos. Esse processo é repetido até que todos os pontos tenham formado um agrupamento. O resultado disto é conhecido como um dendrograma de conexão (Figura 19)^[101].

As distâncias entre as amostras e variáveis são calculadas pela Equação (22):

$$d_{ij} = \left[\sum_{j=1}^q (X_{ij} - X_{jj})^M \right] \frac{1}{M} \quad (22)$$

onde q é o número total de variáveis e M está relacionado com o tipo de distância calculada, por exemplo, quando M = 2 têm-se a distância no espaço euclidiano^[101].

Um outro ponto importante a ser salientado é a forma como os pontos são agrupados. Muitas maneiras de agrupar os pontos têm sido desenvolvidas e algumas dessas maneiras estão ilustradas na Figura 20^[101].

O método de conexão simples baseia-se nas menores distâncias calculadas entre os pontos mais próximos; o método de conexão completa utiliza a distância do ponto mais distante; a conexão por meio de centróides faz uso da distância entre os centros de gravidade; a conexão por meio de médias faz uso da distância entre as médias dos grupos de pontos; a conexão incremental baseia-se na média ponderada, etc^[101]. A conexão mais recomendada quando se tem uma interface entre uma classe de compostos e outra classe é a incremental, a qual foi utilizada neste trabalho para garantir a melhor separação entre os compostos.

Figura 19 - Dendrograma de conexão, utilizando a conexão simples.

Figura 20 - Alguns métodos de agrupamento.

4.3. ANÁLISE DE DISCRIMINANTES POR PASSOS (SDA)

Análise de discriminantes foi desenvolvida pelo estatístico R. A. Fisher, em 1930, para aplicações em problemas de biologia e medicina, sendo amplamente usada e aplicada em todas as áreas da ciência^[109]. É uma técnica multivariada que tem dois objetivos principais: (i) separar os objetos em populações bem definidas; (ii) alocar novos objetos dentro de populações previamente definidas^[110].

Na SDA, as variáveis são combinadas linearmente para se obter a função discriminante (FD)^[111]. Pesquisas usando funções discriminantes costumam utilizar o menor número possível de variáveis para classificação, pois o uso de poucas variáveis torna a função discriminante mais fácil de trabalhar^[112].

Na análise discriminante por passos iterativos, as variáveis independentes são colocadas dentro da função discriminante, uma por vez, com base no seu poder discriminante. A aproximação por passo começa escolhendo uma única variável com melhor discriminação. A variável inicial é então combinada com cada outra das variáveis independentes uma por vez, e a variável capaz de melhorar o poder de discriminação da função é combinada com a primeira variável escolhida. A terceira ou qualquer outra variável subsequente é selecionada de maneira similar. O processo termina quando o efeito da variável suplementar não seja mais significativo^[111-113].

A discriminação é significativa se o número de indivíduos classificados incorretamente é mínimo para cada classe. A SDA pode ser usada como um instrumento para prever e sintetizar compostos cuja projeção corresponderia a classe dos ativos^[111].

Um problema com a função da análise de discriminantes por passo é a tendência que este procedimento introduz em testes de significância. Dado um grande número de variáveis é quase certa que alguma combinação delas irá produzir uma função discriminante (significante) em apenas uma chance. Se a análise por passo é executada, é aconselhável conferir a validade, revendo a alocação dos indivíduos nos grupos e vendo o quanto significativo são os resultados obtidos^[102].

Para realizar a análise de discriminantes por passo foi utilizado o programa Minitab 10^[99].

4.4. MÉTODO DOS K-ÉSIMOS VIZINHOS MAIS PRÓXIMOS (KNN)

No método dos vizinhos mais próximos, calcula-se a distância entre todas as amostras do conjunto de treinamento no espaço N-dimensional. Pode-se usar a distância Euclidiana ou outras distâncias, como por exemplo a distância de Mahalanobis^[107].

Para classificar uma amostra desconhecida, calcula-se a distância dessa amostra e de todas as amostras do conjunto de treinamento. As amostras mais próximas são usadas para fazer a classificação e a atribuição à uma classe é feita por votos usando a classe de amostras mais próximas como critério. Os K vizinhos mais próximos de cada amostra são escolhidos para votar, onde cada amostra dá um voto para a sua classe. A classe que recebe mais votos ganha a amostra. Em caso de empate, a classe com a menor distância acumulada fica com a amostra^[107, 114].

Para selecionar o número ótimo de vizinhos, K (número de vizinhos que votam em uma classificação futura), o procedimento de validação cruzada é aplicado ao conjunto de dados. O processo de validação cruzada deixando uma amostra de fora (*leave one out*) é usado. Neste procedimento, cada amostra do conjunto de treinamento é excluída uma vez e então classificada usando o restante das amostras como conjunto de treinamento. O processo é repetido para diferentes valores de K e o número de vizinhos selecionados no modelo final é escolhido com base ao número de erros de classificação^[107, 114].

Um exemplo de classificação com o método KNN é mostrado na Figura 21, onde há duas classes: os triângulos (vermelhos), os quadrados (azuis) e uma nova amostra (o círculo roxo). Com três vizinhos mais próximos, a nova amostra seria atribuída à classe dos triângulos, conforme a Figura 21.

Figura 21 – Representação do método KNN para classificação de uma nova amostra.

4.5. MODELOS INDEPENDENTES DE SIMILARIDADE UTILIZANDO COMPONENTES PRINCIPAIS (SIMCA)

Ao contrário do método KNN, que é baseado somente na distância entre as amostras, no método SIMCA cada classe do conjunto de treinamento (conjunto de dados com as classes conhecidas) é submetida a uma análise de componentes principais (Figura 22). O número de componentes principais é determinado para cada classe e as amostras de cada classe ficam contidas em uma hiper-caixa onde os limites são definidos com o nível de confiança desejado^[107-114].

Uma vez obtidas as fronteiras de cada classe, é importante determinar se há superposição entre as mesmas. O poder de discriminação dos modelos SIMCA em distinguir as classes é dado pela distância entre classes e pelos resíduos entre classes. A atribuição de uma amostra teste à uma dada classe é baseada na projeção da mesma no espaço dos escores e sua distância das fronteiras da classe em questão. Isto é repetido para todas as classes e no final, se os modelos não tiverem poder de discriminação suficiente, a amostra pode ser membro de mais de uma classe. O outro extremo também pode acontecer, quando a amostra é anômala em relação ao conjunto de treinamento (suficientemente diferente), e neste caso ela não pertence a nenhuma das classes^[107-114].

Figura 22 – Representação dos modelos SIMCA com diferentes componentes principais para cada classe.

4.6. REDES NEURAIIS ARTIFICIAIS

Redes Neurais Artificiais (ANN – *Artificial Neural Networks*) ou simplesmente redes neurais, vêm sendo muito utilizadas na Química, demonstrando um grande potencial de aplicação em modelagem, simulação, controle e predição^[115].

Ao contrário dos métodos estatísticos clássicos tais como a análise de regressão ou o método de mínimos quadrados parciais (PLS), as redes neurais permitem a investigação de relações complexas não lineares. Entretanto, os resultados não são sempre fáceis de serem interpretados em termos químicos, devido à complexidade e a não linearidade do modelo de redes neurais, a habilidade de lidar com a relação não linear é uma característica atrativa das redes neurais, sendo apropriada para o uso de planejamento de fármacos^[116, 58].

As redes neurais podem ser usada nas seguintes aplicações no processo de planejamento de fármacos:

- ✓ Análises de dados multidimensionais
- ✓ Classificação e predição da atividade biológica e das propriedades da ADME/Tox (absorção, distribuição, metabolismo, excreção e toxicidade)
- ✓ Descoberta de compostos matrizes (*lead compounds*)
- ✓ Análise da similaridade e diversidade de coleções combinatórias
- ✓ Predição da solubilidade aquosa de fármacos^[116]

As redes neurais tem também produzido modelos preditivos que discriminam entre fármacos e não-fármacos, com aproximadamente 80% de sucesso na classificação^[117].

Um dos primeiros trabalhos sobre redes neurais artificiais foi fruto do trabalho pioneiro de Warren McCulloch e Walter Pitts, em 1943, e ficou conhecido como neurônio de McCulloch-Pitts. O modelo de processamento lógico proposto por eles tinha como objetivo simular os procedimentos executados pelos neurônios do cérebro humano^[118].

Redes neurais é o nome dado ao conjunto de métodos matemáticos e algoritmos computacionais desenvolvidos no sentido de simular o processamento de dados de maneira semelhante ao cérebro humano. As redes neurais imitam os neurônios humanos, recebendo e enviando sinais. Ela pode aceitar uma série de dados de entrada e produzir a partir destes uma ou mais soluções (saídas) através da busca de uma função que descreva o sistema. Esta analogia feita com os neurônios humanos é baseada na transmissão de sinais elétricos. De maneira simplificada, o neurônio humano pode ser entendido como um corpo celular com

duas ramificações: dendritos e axônios (Figura 23). A transferência de sinais elétricos é feita através de um axônio de um neurônio para os dendritos de um outro neurônio vizinho em um processo chamado sinapse^[119].

Figura 23 – Representação de um neurônio biológico.

As sinapses representam barreiras que modulam os sinais que passam através delas, e a quantidade de sinal trocado em uma sinapse depende de um parâmetro chamado de intensidade da sinapse. Em um neurônio artificial, a intensidade da sinapse é simulada por um fator de ponderação chamado peso da sinapse ou simplesmente peso.

O sinal total que entra no corpo de processamento de um neurônio artificial é chamado Net, cujo valor é calculado através da simples multiplicação do sinal que chega ao neurônio pelo peso da sinapse:

$$\text{Net} = \text{Sinal de entrada (input)} \times \text{Peso} \quad (23)$$

Neurônios possuem um grande número de dendritos e, portanto, podem estabelecer várias sinapses com outros neurônios da rede neural, simultaneamente. Assim, o cálculo da Net deve ser generalizado para n entradas (inputs). Chamando os sinais de entrada de x_i e os pesos de w_i , o valor da Net será dado por^[120]:

$$\text{Net} = \sum_{i=1}^n w_i x_i \quad (24)$$

A saída de um neurônio é obtida a partir de uma função de transferência aplicada sobre a Net. Esta função de transferência pode ser linear, sigmoideal, tangente hiperbólica ou qualquer outra função que possa ser criada, o que depende do comportamento dos dados.

Neste trabalho, foi utilizada a função de transferência sigmoideal, que é a que mais se aproxima da saída de um neurônio biológico e que tem um comportamento não-linear. Matematicamente, escreve-se a função sigmoideal como^[120]:

$$F = \frac{1}{1 + e^{-(\text{net})}} \quad (25)$$

As redes neurais possuem sempre uma camada de entrada e uma camada de saída. Entre a camada de entrada e a de saída existe um número de camadas intermediárias. Para esta disposição das camadas e o número de neurônios por camada, dá-se o nome de arquitetura das redes neurais^[120].

Para que as redes neurais executem tarefas de reconhecimento de padrões ou modelagem de dados, ela deve ser treinada. O treinamento pode ser dividido em duas fases: uma fase de treinamento ou aprendizagem e outra de validação ou generalização.

Na primeira fase de treinamento ou aprendizagem é essencial que se utilize um conjunto de calibração, no qual são ajustados os pesos em cada neurônio e em cada camada, até que a saída forneça um resultado estabelecido. Esse ajuste é realizado através de métodos de otimização dos pesos em função dos erros, onde os mais utilizados são o método de Retropropagação (*backpropagation*) e o de Levenberg-Marquardt, sendo que ambos diferem entre si na metodologia da correção dos pesos. O processo de treinamento pode ser supervisionado ou não supervisionado^[119]. No processo supervisionado, o algoritmo ajusta os pesos de conexão entre os neurônios de acordo com a diferença entre os resultados esperados e preditos pela rede. Já o não supervisionado, o algoritmo não necessita de dados de saída conhecidos, onde apenas os dados de entrada são apresentados à rede que automaticamente adapta os pesos das conexões entre os neurônios^[121].

As redes neurais irão apresentar uma boa capacidade de validação ou generalização quando for aplicada em amostras que não participaram do treinamento, deste modo pode-se

avaliar a saída da rede esperada através do modelo matemático desenvolvido na fase de treinamento.

4.6.1. RETROPROGAÇÃO (*BACKPROPAGATION*)

No método de retropropagação de erros, o processo de correção dos pesos é iniciado na última camada e prossegue em direção a primeira camada, daí o nome retropropagação. E este processo é representado na Figura 24.

Figura 24– Representação da retropropagação para uma rede com uma camada de entrada (W1), uma camada intermediária (W2) e uma camada de saída (W3).

Uma das principais aplicações do método de retropropagação é na modelagem de dados não-lineares, onde o comportamento matemático entre as variáveis não é representado por funções de ordem definida.

Cerca de 90% das aplicações de redes neurais em Química empregam este método ^[119].

A principal característica desta rede é a equação da correção dos pesos, que pode ser escrita como:

$$\Delta w^k = \eta D + m \Delta w^{k-1} \quad (26)$$

onde:

- Δw^k representa a diferença entre os pesos inicial e final para a interação k
- η é chamado velocidade de treinamento (*learning rate*), sendo que sua função é ponderar as correções dos pesos de modo a evitar grandes correções, principalmente no final do processo quando boa parte dos pesos já foi corrigido
- D é a derivada do erro calculado através de transferências escolhidas, conhecidas como método do gradiente descendente, e fornece a direção que se deve caminhar na superfície de erros em direção ao mínimo
- μ é chamado *momentum*, uma constante que é multiplicada pela variação de pesos da interação anterior (Δw^{k-1}), o que evita a possibilidade de se achar um mínimo local
- Δw^{k-1} representa a diferença entre os pesos inicial e final para a interação k-1 ^[119]

5. COMPOSTOS ESTUDADOS

Goulart e colaboradores^[122] testaram uma série de quinonas *in vitro* contra a forma tripomastigota do *T. cruzi*, onde foi estabelecida uma correlação entre a atividade tripanossomicida e o primeiro potencial catódico (E_{pc1}). Este potencial foi medido por voltametria cíclica e representa uma indicação quantitativa da facilidade de redução e subsequente reoxidação para espécies reduzidas. A avaliação biológica desses compostos foi feita através do valor do primeiro potencial catódico^[122].

Quinonas que apresentaram valores de E_{pc1} acima de -0,72 V foram consideradas ativas, enquanto aquelas com valores abaixo de -0,72 V foram consideradas inativas^[122].

A estrutura geral das quinonas estudadas, juntamente com o sistema de numeração utilizado, está mostrada na Figura 25.

Figura 25 – Estrutura geral das quinonas estudadas e sistema de numeração adotado neste trabalho.

As 25 quinonas estudadas podem ser divididas em dois grupos: Grupo A, contém os compostos ativos (compostos 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 e 25) e o Grupo B que contém os compostos inativos (compostos 1, 2, 3, 4, 5, 6, 7, 8, 9 e 10). As 25 quinonas estudadas juntamente com sua indicação de atividade tripanossomicida estão mostradas na Figura 26.

1 - INATIVO **2 - INATIVO** **3 - INATIVO** **4 - INATIVO** **5 - INATIVO** **6 - INATIVO** **7 - INATIVO** **8 - INATIVO** **9 - INATIVO** **10 - INATIVO** **11 - ATIVO** **12 - ATIVO** **13 - ATIVO** **14 - ATIVO** **15 - ATIVO**

Figura 26 – Estrutura e indicação de atividade das quinonas estudadas.

Analisando a Figura 27a verifica-se que não houve inicialmente uma boa separação entre os compostos ativos (11-25) e inativos (1-10). Além disso, verifica-se que há muita informação no gráfico dos *loadings* (Figura 27b) (variáveis), o que dificulta a visualização e a análise de dados.

Várias análises da PCA foram realizadas posteriormente utilizando 8 variáveis selecionadas, mas a melhor discriminação entre as moléculas ativas e inativas foi obtida usando somente 4 das 223 variáveis que se tinha inicialmente. As 4 variáveis responsáveis pela separação foram (Tabela 2): T_5 (ângulo de torção formado pelos átomos 1, 2, 17 e 18), QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A), VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B) e HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO). Isto sugere que as outras 219 variáveis não são importantes para a classificação destes compostos. Na Figura 28, para melhor visualização das variáveis, são mostrados os histogramas com as quatro variáveis que foram responsáveis pela separação (T_5 , QTS1, VOLS2 e HOMO-1) com relação aos compostos estudados.

Tabela 2 - Valores obtidos para as quatro propriedades (variáveis) usadas na classificação das 28 quinonas estudadas com sua respectiva indicação de atividade tripanossomicida

COMPOSTO	T_5 (°)	QTS1	VOLS2 (Å ³)	HOMO-1 (eV)	ATIVIDADE
1	106,53	0,021	243,56	-0,3545	Inativo
2	106,47	0,002	243,09	-0,3522	Inativo
3	106,46	0,014	242,91	-0,3520	Inativo
4	112,93	-0,018	241,33	-0,3541	Inativo
5	112,84	-0,049	241,43	-0,3449	Inativo
6	112,96	-0,049	242,02	-0,3501	Inativo
7	113,17	-0,046	242,46	-0,3404	Inativo
8	113,07	-0,074	242,31	-0,3421	Inativo
9	112,80	-0,075	241,44	-0,3449	Inativo
10	120,80	-0,001	248,43	-0,3488	Inativo
11	118,35	-0,056	354,71	-0,3557	Ativo
12	112,84	-0,097	241,98	-0,3711	Ativo
13	132,55	-0,097	261,14	-0,3614	Ativo
14	132,66	-0,161	373,62	-0,3591	Ativo

15	132,60	-0,149	418,22	-0,3588	Ativo
16	121,75	-0,090	200,43	-0,3570	Ativo
17	123,09	-0,018	450,33	-0,3495	Ativo
18	123,80	-0,034	454,67	-0,3525	Ativo
19	123,82	-0,025	405,35	-0,3526	Ativo
20	118,19	-0,130	420,39	-0,3528	Ativo
21	122,40	-0,093	187,59	-0,3600	Ativo
22	121,49	-0,053	200,99	-0,3566	Ativo
23	123,24	-0,102	410,63	-0,3570	Ativo
24	123,21	-0,048	466,21	-0,3630	Ativo
25	123,31	-0,180	540,66	-0,3562	Ativo

Figura 28 – Histograma com os valores calculados obtidos para as variáveis (T₅, QTS1, VOLS2 e HOMO-1) responsáveis pela separação entre os compostos ativos e inativos.

O gráfico dos escores que são as novas coordenadas de cada composto no novo sistema de coordenadas (PCs) está representado na Figura 29. A partir da Figura 29, pode-se observar uma boa separação entre os dois grupos de compostos: grupo A (compostos ativos, 11-25) e B (compostos inativos, 1-10). Os resultados de PCA mostram que as três primeiras PCs descrevem 91,1% da informação total (variância) do conjunto de dados. As duas primeiras componentes principais (PC1 e PC2) expressam 77% do total de variância dos dados, sendo que PC1 é responsável por 56,2% da informação e a segunda componente principal (PC2) tem 20,8% de informação.

Figura 29 – Gráfico dos escores (PC1 e PC2) para as 25 quinonas estudadas com separação entre as quinonas ativas (Grupo A) e inativas (Grupo B).

Figura 30 – Gráfico do *loadings* (PC1 e PC2) para as variáveis responsáveis pela separação dos compostos em ativos e inativos.

A Figura 30 ilustra o gráfico de *loadings* para cada uma das variáveis utilizadas na análise de PCA. Na Tabela 3 são apresentados os valores numéricos dos *loadings* das 4 variáveis selecionadas para a construção das duas primeiras PCs (PC1 e PC2). Os resultados obtidos (Figura 30 e Tabela 3) mostram que as variáveis podem ser agrupadas em duas classes distintas: eletrônicas (HOMO-1 e QTS1) e estéricas (T₅ e VOLS2). De fato, pode-se dizer que as propriedades eletrônicas (HOMO-1 e QTS1) podem ser responsáveis por interações eletrostáticas que são estabelecidas entre os compostos e o receptor biológico, e as propriedades estéricas (T₅ e VOLS2) estão relacionadas com o ajuste tridimensional da molécula com o receptor biológico.

Tabela 3 - Valores das componentes principais para as variáveis responsáveis pela separação dos compostos

VARIÁVEL	PC1	PC2
T ₅	0,579	0,073
QTS1	-0,539	0,019
VOLS2	0,455	0,623
HOMO-1	-0,406	0,778

Como foi descrito anteriormente, cada componente principal é escrita como combinação linear das variáveis originais. Assim, a partir dos valores da Tabela 3, pode-se expressar a primeira componente principal pela seguinte equação:

$$PC1 = 0,579 [T_5] - 0,539 [QTS1] + 0,455 [VOLS2] - 0,406 [HOMO-1] \quad (27)$$

A partir dos valores de PC1 apresentados na Tabela 4, pode-se observar que as moléculas ativas apresentam $PC1 > 0$ (com exceção do composto 22) e as inativas apresentam $PC1 < 0$. Da Equação (27) e da Figura 30, verifica-se que QTS1 e HOMO-1 contribuem negativamente para PC1, enquanto T5 e VOLS2 contribuem positivamente para PC1. Portanto, para que um composto apresente atividade tripanossomicida, as variáveis T5 e VOLS2 precisam apresentar valores altos, combinadas com valores menores para as variáveis QTS1 e HOMO-1. Estas características podem ser usadas no planejamento de novos compostos com alta atividade tripanossomicida, visto que as variáveis podem estar relacionadas com a intensidade da associação molecular por interação eletrostática (QTS1 e HOMO-1) e interação estérea (T5 e VOLS2).

Tabela 4 – Valores dos escores na primeira componente principal (PC1) para cada composto estudado

Composto	PC1
1 (inativo)	-2.051
2 (inativo)	-2.003
3 (inativo)	-2.137
4 (inativo)	-1.211
5 (inativo)	-1.454
6 (inativo)	-1.132
7 (inativo)	-1.723
8 (inativo)	-1.349
9 (inativo)	-1.196
10 (inativo)	-1.075
11 (ativo)	0.1710
12 (ativo)	0.5924
13 (ativo)	1.5799
14 (ativo)	2.5876
15 (ativo)	2.6404
16 (ativo)	0.1687
17 (ativo)	0.1955
18 (ativo)	0.6076
19 (ativo)	0.3082
20 (ativo)	1.0178
21 (ativo)	0.3703
22 (ativo)	-0.243
23 (ativo)	1.3238
24 (ativo)	1.3807
25 (ativo)	2.6357

6.2. RESULTADOS DA ANÁLISE HIERÁRQUICA DE AGRUPAMENTOS (HCA)

No estudo da Análise Hierárquica de Agrupamentos (HCA), utilizou-se a conexão incremental, obtendo-se o dendrograma ilustrado na Figura 31.

Os resultados obtidos pela análise de HCA são similares com a PCA. De fato, o dendrograma pode ser usado para fornecer informações sobre o comportamento químico e verificar os resultados obtidos pela PCA. No dendrograma, as linhas verticais correspondem aos compostos e as linhas horizontais representam os valores da similaridade entre pares de compostos e entre grupos de compostos.

Na Figura 31, verifica-se a boa separação obtida entre as classes de compostos, observando-se dois grupos A e B. O grau de similaridade obtido entre as duas classes de compostos foi igual à zero. Os grupos A e B da Figura 31 correspondem aos mesmos grupos A e B da Figura 29 (análise de PCA). Ambos os métodos PCA e HCA classificaram as 25 quinonas estudadas em dois grupos A e B, exatamente da mesma maneira. Baseado na classificação obtida pela PCA e HCA, as variáveis T_5 (ângulo de torção formado pelos átomos 1, 2, 17 e 18), QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A), VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B) e HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO) são as responsáveis pela separação entre as quinonas ativas e inativas estudadas.

Figura 31 - Dendrograma de conexão, utilizando a conexão incremental.

6.3. RESULTADOS DA ANÁLISE DE DISCRIMINANTE POR PASSO (SDA)

Na análise de discriminante por passo (SDA) as variáveis foram também autoescaladas e quatro variáveis significantes foram extraídas das 223 que se tinha inicialmente, após três passos. As variáveis foram: T_5 , QTS1, VOLS2 e HOMO-1. As duas funções discriminantes obtidas para os compostos ativos e inativos foram:

$$\text{Grupo A (ativos)} = -0,959 + 1,335 [T_5] - 0,278 [QTS1] + 0,805 [VOLS2] - 1,438 [HOMO-1]$$

$$\text{Grupo A (inativos)} = -2,895 + 2,404 [T_5] + 0,325 [QTS1] - 1,454 [VOLS2] - 2,480 [HOMO-1]$$

As duas funções obtidas através da SDA têm as variáveis T5, QTS1, VOLS2 e HOMO-1, apresentando um elevado peso nesta metodologia de classificação.

Um outro modo de considerar a performance do critério de classificação através das funções discriminantes A e B, obtidas com a SDA, é calcular a matriz de classificação. Usando-se os coeficientes mostrados pela função discriminante, a matriz de classificação é dada pela Tabela 5.

Tabela 5 - Matriz de Classificação

Grupo	Grupos Correspondentes	
	A	B
A	15	0
B	0	10
Total	15	10
Porcentagem	100%	100%

O erro obtido com a matriz de classificação foi de 0% e a separação entre os dois grupos foi muito boa. Quando o grau de atividade tripanossomicida para novas quinonas é investigado, o critério para alocar estes compostos a partir do resultado de SDA é:

- ✓ Inicialmente, para uma nova quinona estudada, calculam-se os valores das quatro variáveis obtidas (T5, QTS1, VOLS2 e HOMO-1) com a metodologia SDA.
- ✓ Substituem-se os valores obtidos nas duas funções discriminantes.
- ✓ Verifica-se qual função discriminante (Grupo A – compostos ativos ou Grupo B – compostos inativos) apresenta maior valor. Se o maior valor está relacionado à função discriminante do grupo A, a nova quinona estudada é ativa; senão é inativa.

Além da matriz de classificação presente na Tabela 5, também foi usado o método de validação cruzada. Neste método uma amostra é retirada de cada vez (*leave-one-out*), a fim de determinar se o modelo obtido é confiável. Primeiro, se retira uma amostra do conjunto de dados, constrói-se as funções discriminantes com as amostras restantes e faz-se a previsão para estas amostras. Posteriormente, o composto retirado é classificado de acordo com a função discriminante gerada. A seguir, o composto retirado é incluído e um novo composto é

retirado, e se repete o processo até que todas as amostras sejam testadas. Os resultados obtidos com a validação cruzada são mostrados na Tabela 6, e a partir destes resultados pode-se ver que os modelos obtidos com a PCA, HCA e SDA são seguros, uma vez que o erro de validação cruzada foi de zero por cento.

Tabela 6 - Matriz de validação cruzada retirando-se uma amostra de cada vez

Grupo	Grupos Correspondentes	
	A	B
A	15	0
B	0	10
Total	15	10
Porcentagem	100%	100%

Comparando os resultados obtidos pela SDA com a PCA e a HCA, nota-se que as mesmas quatro variáveis (T5, QTS1, VOLS2 e HOMO-1) foram selecionadas nas três metodologias, e que os três métodos classificam as 28 quinonas estudadas em dois grupos, A e B, exatamente da mesma maneira. Assim, a partir dos resultados obtidos com a PCA, HCA e SDA pode-se dizer que as variáveis T5, QTS1, VOLS2 e HOMO-1 são variáveis chaves para entender a atividade tripanossomicida das quinonas estudadas e, conseqüente, elas podem ser usadas para o planejamento de novas quinonas com atividade tripanossomicida.

6.4. RESULTADOS DO K-ÉSIMOS VIZINHOS MAIS PRÓXIMOS (KNN)

Este método foi usado para validar as variáveis selecionadas nas análises de PCA, HCA e SDA, ou seja, T5, QTS1, VOLS2 e HOMO-1. A Tabela 7 apresenta os resultados do KNN obtidos com 1, 3, 5 e 7 vizinhos mais próximos. Para todos os casos, a porcentagem de informação correta foi de 100%. Portanto, decidiu-se usar 7 vizinhos próximos, ao invés de 1, 3 ou 5 vizinhos próximos, porque quanto maior o número de vizinhos mais próximos, melhor a confiabilidade do método KNN.

Tabela 7 - Classificação obtida com o método KNN para as 25 quinonas estudadas

Categoria	Número de compostos	Compostos incorretamente classificados			
		1NN	2NN	5NN	7NN
Ativo	15	0	0	0	0
Inativo	10	0	0	0	0
Total	25	0	0	0	0
% de informação correta		100	100	100	100

Os resultados obtidos com o método KNN foram similares àqueles obtidos com a PCA, HCA e SDA confirmando que os resultados obtidos com os quatro métodos quimiométricos (PCA, HCA, SDA e KNN) são bastante confiáveis, pois as quatro metodologias classificaram as quinonas estudadas em duas classes distintas (ativas e inativas) exatamente do mesmo modo.

6.5. RESULTADOS DO MODELO INDEPENDENTE DE SIMILARIDADE UTILIZANDO COMPONENTES PRINCIPAIS (SIMCA)

A análise de SIMCA também foi usada para validar as variáveis selecionadas (T5, QTS1, VOLS2 e HOMO-1) através dos quatro métodos quimiométricos PCA, HCA, SDA e KNN, usados para a discriminação entre as quinonas ativas e inativas.

O método SIMCA constrói um modelo de componentes principais para cada categoria. A Figura 32 mostra o gráfico dos escores em PC1 x PC2 x PC3 com os limites das hipercaixas para ambas as classes. Os resultados através da análise de SIMCA mostraram que as duas classes são bem modeladas, e que foram utilizadas duas componentes principais para modelar os compostos ativos (11-25) e três para os compostos inativos (1-10).

Figura 32 – Gráfico dos escores em PC1 x PC2 x PC3 com os limites das hipercaixas.

Pelo gráfico dos escores (Figura 32) observa-se que as classes possuem um bom poder discriminatório, o que pode ser observado pela distância e pelos resíduos inter-classes^[114]. Os valores das distâncias e dos resíduos inter-classes são mostrados na Tabela 8.

Tabela 8 - Valores das distâncias e resíduos inter-classe

	Classe 1		Classe 2	
	Resíduos	Distâncias	Resíduos	Distâncias
Classe 1	0,274	0	1,302	46,106
Classe 2	38,974	46,106	0,781	0

Da Tabela 8, pode-se ver que o modelo SIMCA tem um bom poder discriminatório, pois apresenta alto valor para a distância inter-classe (46,106) e para os resíduos. Como esperado, os resíduos inter-classes são significativamente menores quando as amostras são ajustadas na sua própria classe do que quando à outra classe^[114].

Outra informação obtida com o método SIMCA foi a importância das variáveis. Para este objetivo, dois conceitos são usados: poder de modelagem e poder discriminatório. O poder de modelagem é a contribuição de cada variável em construir o modelo, e o poder discriminatório é a capacidade das variáveis de diferenciar as duas classes (compostos ativos e inativos). A Tabela 9 mostra o poder de modelagem e o poder discriminatório das variáveis

T5, QTS1, VOLS2 e HOMO-1. De acordo com a Tabela 9, a variáveis HOMO-1 tem o maior valor de poder de modelagem e a variável QTS1 tem o maior valor do poder discriminatório.

Tabela 9 – Poder discriminatório e poder de modelagem das variáveis T5, QTS1, VOLS2 e HOMO-1

Variáveis	Poder discriminatório	Poder de modelagem
T ₅	652,21	0,528
QTS1	3363,77	0,609
VOLS2	1207,15	0,554
HOMO-1	851,94	0,666

A Figura 33 mostra o gráfico das distâncias entre as classes: amostras no quadrante azul pertencem à classe 1 (compostos inativos) e as amostras no quadrante amarelo pertencem à classe 2 (compostos ativos).

Figura 33 – Gráfico das distâncias entre as classes: CS1 (classe 1 – compostos inativos) e CS2 (classe 2 – compostos ativos).

Através do gráfico das distâncias entre as classes (Figura 33), observa-se que todos os compostos inativos (1-10) estão presentes no quadrante azul (CS1), enquanto que os ativos (11-25) estão presentes no quadrante amarelo (CS2), o que mostra que os resultados obtidos com o SIMCA foram similares àqueles obtidos com os quatro métodos quimiométricos anteriores (PCA, HCA, SDA e o KNN). Portanto, as cinco metodologias classificaram as quinonas estudadas em duas classes distintas (ativas e inativas) exatamente da mesma maneira. A partir desses resultados nós podemos dizer que é possível obter uma alta porcentagem de sucesso na classificação de uma nova quinona com atividade tripanossomicida, utilizando os modelos de PCA, HCA, SDA, KNN e SIMCA obtidos neste trabalho.

7. DISCUSSÃO DOS RESULTADOS OBTIDOS COM A PCA, HCA, SDA, KNN E SIMCA

O objetivo principal deste trabalho é encontrar as propriedades responsáveis pela atividade tripanossomicida a fim de auxiliar no desenvolvimento de novas moléculas baseadas na estrutura das quinonas estudadas, para que possam atuar como agentes terapêuticos. A partir das variáveis responsáveis pela separação dos compostos ativos e inativos, podem-se inferir algumas conclusões:

1) T_5 (ângulo de torção formado pelos átomos 1, 2, 17 e 18)

A partir da Tabela 2 e da Figura 28, verifica-se que os compostos ativos possuem valores mais elevados para T_5 do que os compostos inativos, indicando que os compostos ativos necessitam de um ângulo de torção (formada pelos átomos C1, C2, C3 e O4) acima de 115° para que possam interagir com o receptor biológico.

Figura 34 - Ângulo de torção T5 formado pelos átomos C1, C2, C3 e O4.

2) QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A)

Com relação à soma das cargas dos átomos C1, C2, C17, C18 e dos substituintes na região A, através da Figura 28 e da Tabela 2, pode-se observar que os compostos ativos precisariam apresentar valores mais baixos de QTS1, ou seja, ter átomos doadores de elétrons nessa região A.

Figura 35 – Soma das cargas formadas pelos átomos C1, C2, C17, C18 e dos substituintes na região A.

3) VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B)

Também a partir da Tabela 2 e Figura 28, verifica-se que os compostos ativos possuem valores mais elevados do que os inativos de VOLS2 formado pelos átomos C15,

O16, C17, C18 e dos substituintes na região B. A presença de grupos volumosos nessa região ajudaria no ajuste com o receptor biológico para a atividade tripanossomicida.

Figura 36 – Volume total formado pelos átomos C15, O16, C17, C18 e pelos substituintes na região B.

4) HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO)

As energias dos orbitais de fronteira são propriedades importantes em diversos processos químicos e farmacológicos. Estas propriedades oferecem informações a respeito do caráter doador ou receptor de elétrons dos compostos, isto é, a formação do complexo de transferência de carga.

Da Tabela 2 e da Figura 28, observa-se que a energia do orbital HOMO-1 para os compostos ativos é menor do que para os compostos inativos.

A fim de mostrar a importância do orbital HOMO-1 na discriminação entre os compostos ativos e inativos estudados, a superfície de densidade eletrônica deste orbital está mostrada na Figura 37 para os compostos 18 (Ativo) e 3 (Inativo).

Figura 37 – Superfície de densidade eletrônica do orbital HOMO-1 para os compostos 18 e 3.

A Figura 37 mostra a contribuição do orbital HOMO-1 para um composto ativo e inativo sem a presença do grupo metoxila em sua estrutura.

Na maioria dos compostos ativos e nos compostos inativos sem o grupo metoxila, a principal contribuição do orbital HOMO-1 está localizada nos átomos da região A e nos oxigênios das carbonilas, o que demonstra que essa região pode ser um possível sítio doador de elétrons.

As exceções foram para os compostos ativos 12, 16, 21 e 22 que são mostradas na Figura 38. Esses compostos além de apresentarem lóbulos nos átomos da região A e no oxigênio da carbonila, apresentam lóbulos no anel B, caso dos compostos 21 e 22. No composto 12 há também a presença de lóbulos no carbono ligado ao grupo nitro e nos oxigênios do grupo nitro. Já para o composto 16, ocorre a presença de um lóbulo no oxigênio da hidroxila.

Figura 38 - Superfície de densidade eletrônica do orbital HOMO-1 para os compostos 12, 16, 21 e 22.

Composto 7

Figura 39 – Superfície de densidade eletrônica do orbital HOMO-1 para o composto 7 (composto inativo com grupo metoxila).

No composto inativo 7 com a presença do grupo metoxila, a principal contribuição do orbital HOMO-1 está localizada nos átomos da região B e no anel com o átomo de enxofre, conforme mostrado pela Figura 39. Como exemplo, pode-se citar o composto 7 (inativo) e que apresenta um lóbulo grande na região B e no anel do átomo de enxofre, diferente do composto 12 (ativo) que não apresenta um lóbulo tão grande nessa região. Assim, somente os compostos inativos apresentam o orbital HOMO-1 significativamente afetado pela presença ou ausência dos grupos metoxilas.

Como conclusão, pode-se afirmar que as propriedades eletrônicas e estéreas são fatores importantes para o entendimento da interação entre as quinonas estudadas com atividade tripanossomicida e o receptor biológico, sendo que as propriedades eletrônicas estão relacionadas à intensidade de associação molecular por interações eletrônicas e as estereoquímicas estão relacionadas ao posicionamento da molécula durante a interação com o receptor biológico. De fato, estas características apresentadas pelas quinonas podem ser usadas para o planejamento de novas quinonas com atividade tripanossomicida.

8. VALIDAÇÃO DOS RESULTADOS OBTIDOS COM PCA, HCA, SDA, KNN e SIMCA

A fim de validar os resultados obtidos com os métodos PCA, HCA, SDA, KNN e SIMCA, obtidos com o conjunto de treinamento (25 quinonas estudadas), decidiu-se realizar um estudo de previsão com um conjunto de 3 novas quinonas que apresentam estrutura

química similar ao conjunto de treinamento e cuja atividade tripanossomicida é conhecida^[123]. Este estudo teve como objetivo principal testar os modelos obtidos para o conjunto de treinamento, empregando os métodos PCA, HCA, SDA, KNN e SIMCA. A estrutura química de cada uma das 3 novas quinonas utilizadas no estudo de previsão é representada na Figura 40.

Figura 40 – Estrutura química dos 3 compostos utilizados no estudo de previsão.

As variáveis selecionadas (T_5 , QTS1, VOLS2 e HOMO-1) para os 3 compostos utilizados são mostradas na Tabela 10. É importante ressaltar que todos os valores calculados para as variáveis T_5 , QTS1, VOLS2 e HOMO-1 foram autoescaladas para realizar as análises quimiométricas.

Tabela 10 – Valores obtidos para as variáveis para os 3 novos compostos estudados

Compostos	T_5	QTS1	VOLS2 ($\text{\AA}^3$)	HOMO-1 (eV)
27	122.93	-0.1150	411.17	-0.3423
28	120.07	-0.0480	465.14	-0.3622
29	128.48	-0.1820	380.90	-0.3576

Os resultados de previsão obtidos para o novo conjunto de compostos são mostrados na Tabela 11. Da Tabela 11 pode-se observar que o composto 26 foi classificado corretamente

como inativo e os compostos 27 e 28 foram classificados corretamente como ativos pelos cinco métodos (PCA, HCA, SDA, KNN e SIMCA). Assim, pode-se concluir que os modelos obtidos para o conjunto de treinamento através da PCA, HCA, SDA, KNN e SIMCA são confiáveis e podem ser aplicados em novas quinonas cuja atividade tripanossomicida seja desconhecida.

Tabela 11 – Resultados da previsão da atividade tripanossomicida utilizando os métodos PCA, HCA, SDA, KNN e SIMCA para as 3 novas quinonas estudadas: (+) ativo e (-) inativo

Composto	PCA	HCA	SDA	KNN	SIMCA
26	-	-	-	-	-
27	+	+	+	+	+
28	+	+	+	+	+

9. RESULTADOS DAS REDES NEURAIS ARTIFICIAIS

Como as redes neurais artificiais do tipo Multi Layer perceptrons (MLP) são aproximadores não lineares, resolveu-se também utilizá-las para validar o conjunto de variáveis selecionadas (T_5 , QTS1, VOLS2 e HOMO-1) para as 28 quinonas estudadas. As 28 quinonas estudadas foram separadas em conjunto de treinamento e teste. Para isso, foram escolhidas aleatoriamente 20 quinonas como conjunto de treinamento, para ensinar uma rede neural MLP (70% dos compostos), e o restante das quinonas foi utilizado como conjunto teste (30% dos compostos) para verificar a performance do modelo.

O programa utilizado para treinamento das redes neurais foi o Matlab 6.0^[100], e o algoritmo de aprendizagem utilizado foi o *backpropagation* (de retropropagação de erros). Esse algoritmo descreve o comportamento do erro entre a saída desejada e a saída real. A partir da estimativa deste erro, os pesos das conexões são corrigidos (ajustados) de forma a minimizar o erro.

A rede MLP utilizada para modelar os dados é constituída de três camadas (camada de entrada (com 4 neurônios), intermediária (de 1 a 15 neurônios) e camada de saída (com 2 neurônios)), o valor limite adotado para o erro de treinamento da rede foi igual a 5.10^{-4} . O limite de iterações para atingir convergência foi de 15000 iterações e as funções utilizadas na rede neural são do tipo sigmóide (não lineares).

Todos os resultados obtidos com as redes neurais MLP foram feitos em triplicatas, pois a escolha dos pesos iniciais para os neurônios em cada camada é feita pelo Toolbox do Matlab através de uma função aleatória. Devido a isto, o primeiro cálculo de erro de treinamento feito pela rede pode gerar um resultado ruim, então todo o procedimento é repetido mais duas vezes.

Os neurônios de entrada correspondem as variáveis responsáveis pela separação entre os compostos ativos e inativos, ou seja, T₅, QTS1, VOLS2 e HOMO-1; e os 2 neurônios na camada de saída correspondem as duas classes de compostos estudados (ativos e inativos).

O procedimento de otimização da arquitetura das redes neurais MLP consiste na observação do erro quadrático médio (RMS - *Root Mean Square*) no treinamento das redes neurais, em função da adição de neurônios à camada intermediária. As redes neurais foram treinadas adicionando-se neurônios à camada intermediária (de 1 a 15 neurônios) observando-se o erro RMS no treinamento das redes neurais. O ponto onde o erro é mais baixo considera-se como sendo a configuração ideal das redes. Este procedimento permite a melhor escolha de configuração evitando que ocorra o sub-ajuste (*underfitting*) ou super-ajuste (*overfitting*) das redes.

O erro quadrático médio (RMS - *Root Mean Square*) pode ser definido por:

$$RMS = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (28)$$

onde y_i é o valor real, $\hat{y}_i$ representa o valor estimado pelo método e n é o número de amostras de teste.

A partir da Figura 41, o menor erro RMS foi obtido com 10 neurônios na camada intermediária, e a arquitetura da rede adotada foi 4-10-2 (4 neurônios na camada de entrada, 10 neurônios na camada intermediária e 2 neurônios na camada de saída). A Figura 41 mostra o erro RMS contra a adição de neurônios na camada intermediária das redes neurais, no qual se observa que existe um padrão de comportamento em que o erro começa alto para um neurônio na camada intermediária, depois ele diminui até o número ótimo e então volta a crescer. Pode-se concluir que este comportamento mostra que as redes passam por estado de sub-ajuste e super-ajuste antes de atingir a melhor configuração.

Figura 41 - Erros RMS calculados contra o número de neurônios na camada intermediária.

Após a escolha da melhor arquitetura das redes, foi feito um processo de validação cruzada com o intuito de avaliar a capacidade de previsão das redes neurais. Para isto, utilizou-se o conjunto teste e a classe destes compostos foi prevista pelas redes. As duas classes de compostos foram definidas por dois códigos (rótulos), ou seja, os compostos ativos foram nomeados pelo código (1 0) e os compostos inativos com o código (0 1).

Foram construídos 8 modelos com o método de validação cruzada do tipo deixa um de fora (*leave-one-out*). Os 8 compostos do conjunto teste foram escolhidos aleatoriamente para previsão, e os resultados são mostrados nas Tabelas 12 e 13. De acordo com os dados das Tabelas 12 e 13, apenas o composto 26 apresentou rótulo diferente daquele correspondente a sua classe. Da Tabela 14, verifica-se que o composto 26 apresenta rótulo (0 1), mas o rótulo fornecido pelas redes foi (0,9198 0,0104), o que indica que esse composto foi classificado incorretamente. Assim, apenas do composto 26 não foi classificado corretamente no conjunto teste, o que mostra que os resultados obtidos com as redes neurais apresentam alta porcentagem de informação correta (87,5%). Outro fator importante foram os baixos erros de treinamento (RMS1) e teste (RMS2), o que demonstra novamente que as variáveis

selecionadas (T_5 , QTS1, VOLS2 e HOMO-1) podem ser consideradas bons descritores quando se deseja planejar novas quinonas que apresentem atividade tripanossomicida.

Deste modo, pode-se dizer que o modelo de redes neurais para as quinonas estudadas é um modelo robusto e capaz de gerar previsões confiáveis, e além dos métodos PCA, HCA, SDA, KNN e SIMCA, a redes neurais também pode ser utilizada para avaliar se as variáveis selecionadas são mesmo importantes na discriminação entre os compostos ativos e inativos.

Tabela 12 – Compostos que foram utilizados para teste das redes neurais com os erros de treinamento (RMS1) e erros do conjunto teste (RMS2)

	Compostos do conjunto teste	RMS1			Média	RMS2
		Treinamento			RMS1	Teste
		1	2	3		
Modelo 1	9, 10, 22, 23, 24, 25, 27 e 28	0,0097	0,0122	0,0101	0,0107	0
Modelo 2	1, 10, 22, 23, 24, 25, 27 e 28	0,006	0,0065	0,005	0,0058	0
Modelo 3	9, 2, 22, 23, 24, 25, 27 e 28	0,0062	0,005	0,0069	0,0060	0
Modelo 4	9, 10, 3, 23, 24, 25, 27 e 28	0,005	0,0037	0,0045	0,0044	0
Modelo 5	9, 10, 22, 4, 24, 25, 27 e 28	0,0019	0,0049	0,0066	0,0045	0
Modelo 6	5, 9, 10, 22, 23, 25, 26 e 27	0,0083	0,0055	0,0066	0,0068	0,125
Modelo 7	6, 9, 10, 22, 23, 25, 27 e 28	0,0056	0,0009	0,0060	0,0042	0
Modelo 8	1, 2, 7, 22, 23, 24, 25 e 26	0,0056	0,0057	0,0077	0,0063	0,125

Tabela 13 - Resultados obtidos com o modelo 1 utilizando redes neurais para o conjunto teste

Composto	Atividade	Classe Real		Classe Prevista	
9	Inativo	0	1	0,0129	0,9851
10	Inativo	0	1	0,0253	0,9731
22	Ativo	1	0	0,9530	0,0378
23	Ativo	1	0	0,9912	0,0224
24	Ativo	1	0	0,9413	0,0151
25	Ativo	1	0	0,9539	0,0136
27	Ativo	1	0	0,9510	0,0326
28	Ativo	1	0	0,9900	0,0128

Tabela 14 - Resultados obtidos com o modelo 6 utilizando redes neurais para o conjunto teste

Composto	Atividade	Classe Real		Classe Prevista	
5	Inativo	0	1	0,0171	0,9707
9	Inativo	0	1	0,0154	0,9626
10	Inativo	0	1	0,0412	0,9357
22	Ativo	1	0	0,8508	0,1025
23	Ativo	1	0	0,9967	0,0172
25	Ativo	1	0	0,9927	0,0192
26	Inativo	0	1	0,9198	0,0104
27	Ativo	1	0	0,9927	0,0269

10. DOCAGEM

O reconhecimento molecular é a base fundamental para a ação de fármacos. Em 1894, Emil Fisher foi o primeiro a propor o princípio da chave e fechadura entre proteínas/receptor e seus ligantes pelo processo de complementaridade estereoquímica e eletrônica^[124]. Esta noção de complementaridade foi subseqüentemente aplicada à reação antígeno e anticorpo e depois extrapolada para a interação entre fármacos e seus respectivos receptores. Esta complementaridade, em nível molecular, seria observada através de interações intermoleculares (interações hidrofóbicas, eletrostáticas, de van der Waals, entre nuvens π , ligações de hidrogênio, etc.) capazes de estabilizar a energia livre do complexo fármaco-receptor^[125].

O processo de docagem envolve a predição da conformação e a orientação (ou posicionamento) do ligante com o sítio de ligação do alvo. O processo começa com a aplicação de algoritmos de docagem que colocam as moléculas no sítio ativo^[126].

O primeiro programa usado para a docagem foi o DOCK, descrito em 1982 por Kuntz, depois surgiram outros programas que estão sendo empregados com sucesso tais como FlexX, GOLD, Glide, ICM, Fred e o Autodock^[47]. A maioria dos programas disponíveis trata o ligante como flexível e a proteína (com raras exceções) é tratada como rígida^[127].

Os protocolos de docagem podem ser descritos como a combinação de duas componentes: as funções de busca e de pontuação.

A função de busca pode gerar um número ótimo de configurações até incluir o modo de ligação determinado experimentalmente. A função de busca poderia elucidar todos os possíveis modos de ligação entre o ligante e o receptor se todos os seis graus de liberdade rotacional e translacional do ligante pudessem ser explorados junto com os graus de liberdade conformacional interna do ligante e da proteína. Porém, isto é impraticável devido ao tamanho do espaço de busca. Na prática, são aplicadas restrições e aproximações para reduzir a dimensionalidade do problema na tentativa de localizar o mínimo global. Os primeiros algoritmos de docagem tratavam o ligante e o alvo como corpos rígidos e somente seis graus de liberdade (rotacional e translacional) eram explorados^[128].

Durante o processo de busca do melhor modo de ligação entre o ligante e o receptor, um grande número de conformações é gerado para cada molécula, isso pode ser feito através de métodos aleatórios ou estocásticos, como por exemplo pelo uso de algoritmos genéticos; ou através do método incremental que explora a flexibilidade do ligante no sítio de ligação no

alvo. Este envolve a divisão da molécula em pequenos fragmentos e a reconstrução incremental da molécula no sítio de ligação ^[46].

Os programas utilizados para realizar os estudos descritos nesse trabalho foram: Autodock^[129] (com a função de busca baseada no algoritmo genético Lamarckiano) e o FlexX^[130] (com o algoritmo de busca incremental). Ambos os programas tem sido usados com sucesso na docagem de vários ligantes ^[131].

O Autodock^[129] utiliza um método estocástico inspirado no conceito de evolução. Inicialmente, uma população aleatória é gerada no qual cada membro corresponde a uma solução potencial do problema. Um membro da população é representado por um cromossomo, nos quais as variáveis a serem otimizadas são reunidas. Isto significa que cada cromossomo contém um número de genes, onde esses genes correspondem aos valores de certas variáveis ou conjunto de variáveis. No caso de docagem, as variáveis são os ângulos de translação, rotação e de torção do ligante que são reunidos no cromossomo. Operadores genéticos (cruzamento e a mutação) são aplicados à população inicial para gerar uma nova população. No cruzamento, os genes de dois cromossomos distintos são trocados para gerar dois novos indivíduos e na mutação o gene é modificado aleatoriamente. Indivíduos com melhores escores recebem uma melhor chance de serem selecionados como membros de uma nova população, e assim uma maior chance de sobreviver e reproduzir. Portanto, o ajuste médio aumenta de geração para geração até que em certo ponto o processo é terminado (alcançando um número fixo de gerações ou por ajuste constante da população).

A última versão do Autodock utiliza uma variante do algoritmo genético chamado de algoritmo genético modificado Lamarckiano, que é uma combinação do tradicional algoritmo genético com método de busca local (este realiza o processo de minimização de energia prevendo a conformação de docagem de ligantes flexíveis no alvo macromolecular rígido) ^[51,132].

O FlexX^[130] é o programa mais popular na docagem de fragmentos com a construção de um algoritmo incremental. Como todos os métodos baseados em fragmentos, a escolha do fragmento base é crucial para o algoritmo, que deve conter as interações predominantes com o receptor. O fragmento base é a primeira porção do ligante a ser colocado dentro do sítio de ligação. Neste estágio, o fragmento base é considerado rígido e três sítios no fragmento são mapeados dentro dos três sítios no receptor, e aqueles geometricamente acessíveis são agrupados e sobrepostos. As sobreposições são removidas e as energias são calculadas através do fragmento base. Começando com diferentes colocações do fragmento base, o ligante é construído por passo ligando os componentes de acordo com um banco de dados de ângulos

de torção³. Em seguida, conectam-se fragmentos adicionais e novas interações são buscadas usando-se a função de pontuação. Isso é feito até que o último fragmento tenha sido adicionado resultando no ligante completo. As posições do ligante são finalmente armazenadas e classificadas de acordo com o valor da afinidade de ligação prevista ^[128, 131].

Uma vez geradas todas as conformações possíveis do inibidor dentro do sítio de interação, é necessário avaliar cada uma dessas conformações e determinar qual dentre elas é a correta. Para tanto utilizam funções de pontuação, que podem ser divididas em funções de pontuações empíricas, funções de pontuações derivadas do campo de força e funções de pontuação baseadas no conhecimento^[46].

Funções de pontuação baseadas no campo de força levam em conta as interações intermoleculares descritas por termos clássicos da mecânica molecular tais como AMBER ou CHARMM. No caso das funções de pontuação empíricas, essas funções são ajustadas para reproduzir dados experimentais onde a energia livre de ligação total é composta por vários termos que correspondem a ligações de hidrogênio, interações hidrofóbicas, mudanças entrópicas e, em alguns casos de interações com metais. O planejamento das funções de pontuação empíricas baseia-se na idéia de que as energias de ligação podem ser aproximadas pela soma de termos individuais não correlacionados, e os coeficientes de vários termos serão obtidos através da análise de regressão derivados do ajuste das energias de ligação experimentais de diferentes complexos entre ligante e proteína.

Já as funções de pontuação baseadas no conhecimento são derivadas da análise estatística, sem referência a afinidade de ligação determinada experimentalmente. A função de energia chamada de potencial de força média relaciona a frequência da ocorrência de contatos individuais entre a proteína e os tipos de átomos do ligante como medida da sua contribuição energética para ligação. Se um contato específico ocorre mais frequentemente do que uma distribuição aleatória, isto indica uma interação atrativa. A distribuição de frequência para um conjunto de dados de moléculas que interajam pode ser convertido em um conjunto de potenciais de pares de átomos favorecendo contatos preferidos e penalizando interações repulsivas^[46, 126, 127, 133].

Dadas as limitações das atuais funções de pontuação (sendo a maioria ajustadas ou derivadas de dados experimentais), uma tendência tem sido a introdução das funções de pontuação de consenso. A função de pontuação de consenso combina a informação de uma

³ FlexX usa as aproximações baseadas nos fragmentos e um conjunto de ângulos de torção de baixa energia derivados do *Cambridge Structural Database* (CSD) para cada ligação simples.

série de funções de pontuação, reduzindo o número de falsos positivos e melhorando a probabilidade de se identificar verdadeiros ligantes ^[134].

Como um exemplo, podemos citar o sucesso baseado nos programas de docagem onde novos inibidores mais potentes e seletivos da caseína cínase II (CK2) foram obtidos a partir de um conjunto de 400.000 moléculas do banco de dados da Novartis, usando o programa DOCK. Pré-filtros relacionados à massa molecular, número de ligações rotacionais e subestruturas indesejadas foram utilizados antes do ensaio virtual para eliminar possíveis compostos indesejados. Os compostos com melhores escores no DOCK foram posteriormente refinados usando informações farmacofóricas conhecidas da enzima. Os melhores compostos foram finalmente inspecionados antes de submeter um total de 12 compostos a testes biológicos. Um inibidor com IC₅₀ de 80 nM foi encontrado ^[135].

10.1. MÉTODOS UTILIZADOS NA DOCAGEM

Foram realizados estudos de modelagem com dois programas de docagem, com o objetivo de encontrar o sítio e a forma de ligação das quinonas com a enzima e avaliar diferentes funções de pontuação e de busca. Os programas utilizados foram:

- ✓ Autodock 3.05 ^[129]
- ✓ FlexX 2.0 ^[130]

Das 28 quinonas estudadas, algumas apresentaram atividade inibitória ^[3, 4] contra a enzima tripanotona redutase e por essa razão foram selecionadas para estudos de modelagem molecular.

Após a otimização da geometria das 28 quinonas, através do método da Teoria do Funcional da Densidade (DFT – *Density Functional Theory*), com o funcional B3LYP, foram calculadas as cargas de Gasteiger ^[138] para todos os átomos das quinonas através do programa Vega ZZ ^[139].

As coordenadas cristalográficas da enzima tripanotona redutase foram obtidas através do banco de dados de proteínas PDB (*Protein Data Bank*, 1 BZL). A utilização dessa estrutura foi de boa resolução (2,4Å) e de baixo valor de R (20,9%).

Para as docagens com os programas Autodock 3.05 e FlexX 2.0, foram removidos o ligante e todas as moléculas cristalográficas de água do arquivo PDB. Depois, a estrutura da enzima tripanotona redutase (1 BZL) foi checada e todos os resíduos incluídos dentro de uma

esfera de raio de 20Å em volta da Cys58 foram definidos como o sítio de interação, através do programa Chimera^[140]. No Autodock 3.05, com a estrutura desse sítio, foram adicionados somente os hidrogênios polares, e depois foram calculadas as cargas atômicas de Kollman para todos os átomos definidos. Isto se fez necessário para resolver as ambigüidades nas posições dos hidrogênios polares da tirosina, treonina e serina^[141]. O resultado da estrutura tridimensional foi salva em um novo arquivo PDB.

A docagem das 28 quinonas na enzima tripanotona redutase foi feita usando-se o programa Autodock 3.05. As energias de docagem foram calculadas de uma série de grades de energia centradas no sítio da enzima. O algoritmo genético lamarckiano foi usado para a busca conformacional e a conformação do ligante em cada sítio de docagem, foi buscada através do método de busca local de Solis e Wets^[142] a fim de experimentar todas as possíveis conformações do ligante. Entre a série dos parâmetros, o tamanho da grade, no qual a enzima e o ligante estavam incluídos, foi fixado em 45Å x 45Å x 45Å (direções x, y e z) e com espaçamento de 0,375Å. O tamanho da grade foi grande o suficiente para cobrir todos os átomos da enzima próximos ao sítio de docagem e para calcular todas as interações eletrostáticas entre a enzima e as moléculas do ligante.

As coordenadas da enzima foram mantidas fixas durante o cálculo, enquanto que as quinonas foram flexíveis, isto é, se moviam dentro da grade. Assim, todos os 28 compostos foram avaliados e pontuados em termos das energias livres de interação usando a função de pontuação baseada no campo de força AMBER. O melhor composto que apresentou maior complementaridade química e estereoquímica pela enzima foi selecionado.

Outros valores foram usados como parâmetros no algoritmo genético lamarckiano, que são mostrados na Tabela 15:

Tabela 15 – Valores dos parâmetros utilizados pelo algoritmo lamarckiano no procedimento de docagem

Parâmetros	Valores Utilizados
Número de indivíduos na população	150
Número máximo de avaliações de energia	250000
Número máximo de gerações	27000
Número de indivíduos que sobrevivem	1
Taxa de mutação do gene	0,02

Taxa de cruzamento	0,8
Número de gerações para escolha do pior indivíduo	10
Varição da distribuição de Cauchy para mutação de um gene	1
Distribuição de Cauchy média para mutação de um gene	0
Número de interações de Solis e Wets durante a busca local	300
Probabilidade de se realizar a busca local para um indivíduo	0,06
Critério de terminação para a busca local	0,01
Número de execuções	20

O segundo programa de docagem utilizado foi o FlexX 2.0, disponível no programa Sybyl 7.2. O FlexX utiliza um algoritmo de construção incremental e é um algoritmo rápido para a docagem flexível de ligantes na enzima. A mesma região do sítio de interface foi utilizada, mas com uma segunda cavidade (*subpocket*), definida pelos resíduos Trp 22a, Met 114, Phe 396b, Pro 398b, Leu 399b, Glu 466b e Glu 467b. As cargas de Amber95 foram atribuídas para todos os átomos da enzima. A colocação final e a classificação foi baseada na função de pontuação do FlexX, desenvolvida por Böhm, que calcula a soma das contribuições de energia livre derivada das interações intermoleculares tais como o número de ligações rotacionais do ligante, ligações hidrogênio, interações entre pares iônicos, hidrofóbicas, entre nuvens π de grupos aromáticos e interações lipofílicas^[51].

10.2. RESULTADOS DA DOCAGEM

Derivados de naftoquinonas e nitrofuranos foram testados frente à enzima tripanotiona redutase, e foi mostrado que esses compostos são reduzidos pela enzima, e via reação do ciclo redox com oxigênio molecular produzem espécies de oxigênio reativo^[35]. Outro estudo feito com naftoquinonas por Salmon-Chemin e colaboradores^[27], mostrou ser possível planejar inibidores potentes e específicos da tripanotiona redutase a partir da estrutura matriz da menadiona e plumbagina (Figura 10), conhecidos por sua atividade tripanossomicida^[27]. A grande extensão do sítio da enzima tripanotiona redutase indica que muitas diferentes orientações de pequenos ligantes podem ser possíveis^[143]. A interação das quinonas na

tripanotiona pode ocorrer em um sítio diferente do sítio ativo formado pelas cisteínas 53 e 58^[27, 35], como foi o caso de compostos derivados de acridina^[144]. Estudos de modelagem com fenotiazinas encontraram uma cavidade hidrofóbica na tripanotiona redutase, chamado sítio Z, como um potencial novo sítio de ligação e que é formada pelos aminoácidos Phe396, Pro 398 e Leu399.

Vega-Tejido e colaboradores^[19] exploraram três possíveis sítios de ligação (sítio ativo, de interface e de ligação com o NADPH) das enzimas tripanotiona e glutathione redutase, com uma série de oito compostos derivados de nitrofuranos e nitrotiofenos. Os resultados desse trabalho mostraram que, para as duas enzimas, os compostos são mais prováveis de se ligarem no sítio de interface, o que está de acordo com os dados experimentais^[19].

Na tentativa de elucidar o sítio de ligação e o mecanismo de interação entre as 28 quinonas estudadas com a enzima tripanotiona redutase, foram realizados estudos de docagem dessas quinonas com a enzima tripanotiona redutase através dos programas Autodock 3.05 e FlexX 2.0.

10.2.1. RESULTADOS OBTIDOS COM O PROGRAMA AUTODOCK

Todas as 28 quinonas estudadas apresentaram uma alta afinidade por uma região hidrofóbica, formada pelos aminoácidos Lys62, Thr66, Phe396, Pro398 e Pro462, Leu399, His461, Ser464, Glu466 e Glu467, conforme mostrada na Figura 42.

Figura 42 - Resultados obtidos com as 28 quinonas através do programa Autodock.

Das 28 quinonas estudadas, 27 quinonas formaram ligações de hidrogênio com os resíduos da Ser464 (43a), Lys62 (43b) e com a Thr66 (43c). O único composto que não formou nenhuma ligação de hidrogênio foi o composto 14. As ligações de hidrogênio foram obtidas através do programa Chimera^[140].

Figura 43 – Estrutura dos aminoácidos serina, lisina e treonina.

Entre os compostos inativos (1-10), 9 compostos apresentaram duas ligações de hidrogênio entre os átomos de oxigênio das carbonilas (oxigênios 4 e 16, Figura 44) com os hidrogênios dos grupos amino e hidroxila da Ser464 e Lys62, e apenas o composto 8 apresentou uma ligação do oxigênio 4 da carbonila com o hidrogênio do grupo amino da Ser464. Para os compostos inativos sem a presença de grupos metoxilas (compostos 1, 2, 3, 4 e 10), além das ligações dos oxigênios 4 e 16 das carbonilas com os hidrogênios dos resíduos da Ser464 e Lys62, ocorre também uma ligação de hidrogênio com o átomo de oxigênio nos compostos 1, 2, 3 e 10 ou átomo de enxofre no composto 4 (posição 1) com os hidrogênios do grupo hidroxila da Ser464 e do grupo amino e da Lys62. Para os compostos inativos com os grupos metoxilas, compostos 5, 6, 7 e 9, quando o grupo metoxila está à direita da carbonila, ocorre uma ligação de hidrogênio entre o oxigênio do grupo metoxila com o hidrogênio do grupo amino da Lys62, o que não ocorre quando o grupo metoxila se encontra à esquerda das carbonilas (compostos 6 e 8).

Para o composto 26 (inativo), ocorre também uma ligação de hidrogênio entre o átomo de oxigênio 4 da carbonila com o hidrogênio da hidroxila da Ser464, e a outra ligação de hidrogênio do oxigênio 11 com o hidrogênio do grupo amino da Lys62.

1

2

3

4

5

6

7

8

Figura 44 – Modelo de interação das quinonas inativas com a enzima tripanotiona redutase dentro do sítio de interação obtido com o programa Autodock.

Os compostos ativos 11, 12, 21 e 22, mostrados na Figura 45, tiveram duas ligações de hidrogênio entre os átomos de oxigênios 4 e 16 das carbonilas com os hidrogênios dos grupos amino e hidroxila da Ser464 e Lys62. Os compostos 13, 23, 24 e 28 tiveram somente uma ligação de hidrogênio com os resíduos de Ser464 ou Lys62, entre o átomo de oxigênio da carbonila com o hidrogênio do grupo amino ou hidroxila.

Nos compostos 20, 25 e 27 ocorre a formação da ligação de hidrogênio entre o grupo hidroxila da Thr66 com as carbonilas, e também a ligação de hidrogênio com o oxigênio 16 com o hidrogênio do grupo amino da Ser464. Para os compostos 16, 17, 18, e 19, que apresentam um grupo hidroxila vizinho à carbonila, ocorre uma ligação de hidrogênio entre essas hidroxilas (posição 1) com o hidrogênio do grupo amino da Lys62.

Figura 45 – Modelo de interação das quinonas ativas com a enzima tripanotona redutase dentro do sítio de interação obtido com o programa Autodock.

Os resultados das energias de docagem previstas entre as quinonas e a enzima tripanotona redutase são mostrados na Tabela 16.

Tabela 16 – Energias livres de interação para as 28 quinonas estudadas obtidas com o programa Autodock

Compostos	ΔE (kcal/mol)
1	-6,79
2	-7,08
3	-7,45
4	-6,89
5	-7,35
6	-7,08
7	-7,57
8	-7,05
9	-7,90
10	-8,32
11	-8,51
12	-6,94
13	-7,29
14	-7,61
15	-7,70

16	-5,93
17	-8,46
18	-8,08
19	-8,27
20	-7,13
21	-5,74
22	-6,32
23	-7,03
24	-8,07
25	-8,65
26	-7,30
27	-8,06
28	-7,44

Figura 46 – Energias de docagem obtidas com o programa Autodock para as 28 quinonas estudadas.

Através da Tabela 16 e da Figura 46, observa-se que as moléculas que apresentaram as maiores energias de docagem dentro do sítio de interação foram para os compostos 25 e 11. Através da Figura 45, observa-se que compostos 25 e 11 apresentam um maior número de resíduos próximos (distâncias inferiores a 3,5 Å), que poderiam interagir com esses resíduos

através de ligações de hidrogênio fracas (do tipo C-H...O e C-H...N)^[145, 146]. O composto 25 está próximo dos resíduos de Leu399 (3,053 Å), Met400 (2,326 Å), Thr463 (1,767 Å), Ser464 (1,915Å) e os Glu466 (2,757Å) e Glu467 (3,100Å), e pode se ligar através de ligações de hidrogênio fracas com esses resíduos. O composto 11 pode se ligar através dessas ligações fracas com os resíduos de Lys62 (2,185 Å), Met400 (3,463 Å), Thr463 (3,421 Å) e com os Glu436 (2,489Å), Glu466 (3,101Å) e Glu467 (2,559Å). Já para os compostos que apresentaram as menores energias de docagem, compostos 21 e 16, verifica-se que o número de resíduos que podem se ligar através de ligações de hidrogênio fracas são os resíduos de Glu466 (2,648 Å e 2,909 Å) e Glu467 (2,331 Å e 2,314 Å) para os compostos 21 e 16. Esse número menor de interações explicaria as menores energias de docagem obtidas com o programa Autodock.

Na Figura 47, são mostrados os resíduos e as distâncias dos compostos com as maiores e menores energias de docagem obtidas através do programa Autodock.

Figura 47 - Compostos 25, 11, 21 e 16 com os resíduos e as distâncias obtidas através do programa Autodock.

10.2.2. RESULTADOS OBTIDOS COM O FLEXX

Como ocorreu no Autodock, todas as 28 quinonas estudadas apresentaram alta afinidade por uma região hidrofóbica formada pelos aminoácidos Lys62, Phe96, Pro398, Leu399, His461, Pro462, Ser464, Glu466 e Glu467, conforme mostrado na Figura 48.

Figura 48 – Resultados obtidos com as 28 quinonas através do programa FlexX.

No caso do programa FlexX, das 28 quinonas estudadas, 27 formaram ligações de hidrogênio com a Leu399 (Figura 49), e apenas a quinona 3 interagiu com os aminoácidos da Ser464 e Lys62, conforme mostrado na Figura 50.

Figura 49 – Estrutura do resíduo de leucina.

Nos compostos inativos, Figura 50, sem a presença de grupos metoxilas (compostos 1, 2 e 4), ocorre a ligação de hidrogênio entre o átomo de oxigênio 4 e 16 da carbonila do anel da quinona com o hidrogênio do grupo amino da Leu399. Nos compostos inativos com os

grupos metoxilas (compostos 5, 6, 7, 8 e 9), mostrados na Figura 50, ocorre a ligação de hidrogênio entre o átomo de oxigênio do grupo metoxila com o hidrogênio do grupo amino da Leu399. No composto 10 não houve a ligação de hidrogênio dos átomos de oxigênio das carbonilas, mas houve a ligação de hidrogênio entre o átomo de oxigênio 1 do anel aromático com o hidrogênio do grupo amino da Leu399.

O composto 3 foi o único composto da série que não apresentou ligação de hidrogênio com a Leu399, Esse composto, uma das carbonilas se ligou ao hidrogênio do grupo hidroxila da Ser464, e a outra ligação de hidrogênio foi com o hidrogênio do grupo amino da Lys62. No composto 26 ocorre ligação de hidrogênio do grupo hidroxila com o hidrogênio do grupo amino da Leu399.

Figura 50 - Modelo de interação das quinonas inativas com a enzima tripanotona redutase dentro do sítio de interação obtido com o programa FlexX.

Nos compostos ativos (11-25, 27 e 28), todas as quinonas fizeram ligações com o hidrogênio do grupo amino da Leu399, conforme mostrado na Figura 51.

Como as 28 quinonas estudadas ficaram mais próximas da leucina, haverá um sítio de interação hidrofóbica que irá reconhecer o anel aromático do ligante, havendo uma maior estabilização destes compostos.

11

12

13

14

15

16

17

18

19

20

Figura 51 - Modelo de interação das quinonas ativas com a enzima tripanotona redutase dentro do sítio de interação obtido com o programa FlexX.

Os resultados das energias livres de interação previstas com as quinonas e a enzima tripanotona redutase são mostrados na Tabela 17.

Tabela 17 – Energias livres de interação para as 28 quinonas estudadas obtidas com o programa FlexX

Compostos	ΔE (kJ/mol)
1	-9,058
2	-6,845
3	-7,485
4	-7,566
5	-9,91
6	-10,975
7	-9,577
8	-11,032
9	-10,158
10	-9,575
11	-9,945
12	-9,218
13	-9,103
14	-8,336
15	-8,623
16	-10,882
17	-9,718
18	-2,397
19	-8,528
20	-12,445
21	-9,295
22	-10,378
23	-11,162
24	-8,907
25	-7,315
26	-10,141
27	-8,296
28	-11,164

Figura 52 – Energias livres de interação obtidas com o programa FlexX para as 28 quinonas estudadas.

Através da Tabela 17 e da Figura 52, observa-se que as moléculas que apresentaram as maiores energias de docagem dentro do sítio de interação foram os compostos 20, 28, 23, 8 e 6, e são estes compostos que apresentam a melhor energia de interação com a enzima. Os compostos 25 e 28 apresentam um número maior de resíduos próximos (distâncias inferiores a 3,5 Å), que podem interagir com esses resíduos através de ligações de hidrogênio fracas (do tipo C-H...O e C-H...N)^[145-146]. O composto 20 (Figura 53a) está próximo dos resíduos de Thr397 (2,531 Å) e Thr463 (3,007 Å), Ser464 (2,792 Å) e Glu466 (3,240 Å), e pode se ligar através das ligações de hidrogênio fracas com esses resíduos. O composto 28 (Figura 53b) pode se ligar através das ligações de hidrogênio fracas com os resíduos de Thr397 (2,898 Å) e Thr463 (2,429 Å), His461 (3,236 Å) e Ser464 (2,962 Å). Para os compostos que apresentaram as menores energias de docagem, compostos 25 (Figura 53c) e o 2 (Figura 53d), o resíduo que está mais próximo e que poderia se ligar é apenas a Thr397 (1,696 Å para o composto 25 e 1,752 Å para o composto 2). Essa única interação explicaria sua menor energia de docagem obtida com o programa FlexX.

Figura 53 - Compostos 20, 28, 25 e 2 com os resíduos e as distâncias obtidas com o programa FlexX.

A tripanotona redutase é de fundamental importância no sistema de defesa antioxidante do parasito e, apesar de não haver informações da estrutura cristalográfica de raio-X da enzima tripanotona redutase com as 28 quinonas estudadas, foi feita uma inferência para que futuros testes experimentais comprovem os resultados obtidos, auxiliando na busca por novos e melhores inibidores da tripanotona redutase.

As quinonas estudadas apresentam características essenciais para serem consideradas como compostos matrizes, pois apresentam estrutura simples e pontos para modificações moleculares que podem ser otimizados. Cadeias laterais que são positivamente carregadas em meio fisiológico poderiam ser um bom ponto de partida para o desenvolvimento de novos derivados^[136].

Os programas de docagem Autodock e FlexX, que apresentam métodos diferentes na abordagem dos ligantes, com diferentes funções de busca e de pontuação, mostraram que para todas as 28 quinonas estudadas, a região de maior interação com a enzima tripanotona redutase foi formada pelos resíduos de aminoácidos Lys62, Thr66, Thr397 e Thr463, Leu399,

Ser464, Glu466 e Glu467. Os resíduos Thr66, Thr463 e Leu399 são conservados em diferentes tripanotionas redutases (*T. brucei*, *T. congolense* e *Leishmania donovani*)^[27] e poderiam ser importantes para o planejamento de inibidores dessas enzimas.

Com relação ao sítio de interação, Zani e colaboradores^[137] propuseram que as quinonas atuariam como inibidores não competitivos com relação à enzima tripanotiona redutase, isto é, os compostos se ligam em um sítio diferente do sítio ativo da enzima formado pelas cisteínas^[137]. Cenas e colaboradores^[35] propuseram que a redução das quinonas ocorre em um sítio diferente do sítio das cisteínas e do NADPH, chamando-o de terceiro sítio^[35]. Nesse sentido, os resultados apresentados neste trabalho vêm corroborar e estão de acordo com estudos prévios realizados com outras classes de compostos onde também se verificou que as moléculas apresentaram uma maior interação pelo sítio Z da enzima tripanotiona redutase^[144, 145].

11. CONCLUSÕES

Apesar do aparecimento de novos alvos potenciais para o desenvolvimento de novos agentes para a terapia farmacológica contra os tripanossomatídeos, até o momento não existe um fármaco capaz de tratar a doença. Os que existem atualmente apresentam alta incidência de desenvolvimento à resistência e elevados efeitos colaterais, o que aumenta a necessidade de busca e/ou desenvolvimento de novos agentes terapêuticos.

Os resultados obtidos nesta tese para as 25 quinonas estudadas mostraram para os cinco métodos quimiométricos utilizados (PCA, HCA, SDA, KNN e SIMCA) 100% de informação correta, e que somente 4 variáveis das 223 variáveis calculadas podem ser consideradas importantes na discriminação entre os compostos ativos e inativos. As variáveis mais relevantes foram T_5 (ângulo de torção formado pelos átomos 1, 2, 17 e 18), QTS1 (soma das cargas atômicas dos C1, C2, C17, C18 e dos substituintes da região A), VOLS2 (volume total formado pelos átomos C15, O16, C17, C18 e os substituintes na região B) e HOMO-1 (energia do primeiro orbital molecular abaixo do HOMO). Dentre essas 4 variáveis, duas podem ser classificadas como propriedades estruturais (T_5 e VOLS2) e duas como propriedades eletrônicas (HOMO-1 e (QTS1). Assim, quando se deseja planejar quinonas com atividade tripanossomicida deve-se observar algumas características importantes na estrutura do composto: valores altos para T_5 e VOLS2, combinados com valores menores para QTS1 e HOMO-1. Dos resultados obtidos, pode-se concluir que as propriedades eletrônicas e estruturais são fatores importantes na interação entre as quinonas que apresentam atividade tripanossomicida e o receptor biológico, onde as propriedades eletrônicas estão relacionadas à intensidade de associação molecular por interações eletrostáticas e as propriedades estruturais estão relacionadas ao posicionamento da molécula durante a interação com o receptor biológico.

A partir dos resultados obtidos para as 25 quinonas utilizadas no conjunto de treinamento, 3 novas quinonas com atividade conhecida foram utilizadas para previsão. Os cinco métodos quimiométricos PCA, HCA, SDA, KNN e SIMCA classificaram as 3 quinonas corretamente.

Além dos cinco métodos quimiométricos, foi utilizado o método de redes neurais artificiais, utilizando o algoritmo de retropropagação de erros. As quatro variáveis (T_5 , QTS1, VOLS2 e HOMO-1) foram empregadas de forma a validar os modelos usados anteriormente. A arquitetura da rede com 4 neurônios na camada de entrada, 10 neurônios na camada

intermediária e 2 neurônios na camada de saída foi adotada observando-se o erro quadrático médio. Os resultados obtidos com as redes neurais apresentaram alta porcentagem de informação correta (87,5%) e baixos erros de treinamento e teste, e apenas um composto foi classificado incorretamente.

Na docagem foram utilizados dois programas que apresentam métodos diferentes na abordagem dos ligantes, o Autodock e o FlexX. Os resultados de docagem na enzima tripanotiona redutase mostraram que todas as quinonas estudadas ficaram numa região hidrofóbica chamada de sítio Z.

Os resíduos que interagem com as quinonas são Lys62, Thr66, Thr397, Thr463, Leu399, Ser464, Glu466 e Glu467, sendo que os resíduos de Thr66, Thr463 e Leu399 são conservados em diferentes tripanotionas redutases e poderiam ser usados no planejamento de inibidores seletivos contra o parasito.

Por fim, como proposta para futuros trabalhos, pode-se sugerir a utilização dos modelos de classificação e de docagem, obtidos através das análises empregadas neste estudo, para propor novas quinonas com atividade tripanossomicida que poderão ser sintetizadas. Posteriormente, testes biológicos seriam realizados com o intuito de verificar a presença de atividade tripanossomicida nesses compostos e sua possível utilização como fármacos.

12. BIBLIOGRAFIA

- [1] THE DISEASE. Disponível em: <http://www.who.int/ctd/chagas/disease.htm>>. Acesso em: 15 jan. 2003.
- [2] LINDOSO, A. A. B.; YASUDA, M. A. S. Doença de Chagas crônica: do xenodiagnóstico e hemocultura à reação em cadeia da polimerase. **Rev. Saúde Pub.**, v. 37, n. 1, p. 107-115, 2003.
- [3] COURA, J. R.; DE CASTRO, S. L. A critical review on Chagas disease chemotherapy. **Mem. Inst. Oswaldo Cruz**, v. 97, n. 1, p. 3-24, 2002.
- [4] GÜRTLER, R. E.; SEGURA, E. L.; COHEN, J. E. Congenital transmission of *Trypanosoma cruzi* infection in Argentina. **Emerg. Infect. Dis.**, v. 9, n. 1, p. 29-32, 2003.
- [5] TRAPANI, Stefano. **Estudos estruturais d enzima fosfoenolpiruvato carboxiquinase (peck) de *Trypanosoma cruzi*, um potencial alvo para o planejamento de drogas antichagásicas.** 2001. 186 f. Tese (Doutorado em Físico-Química) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2001.
- [6] NEVES, D. P. **Parasitologia Humana.** 8.ed. São Paulo: Atheneu, 1991. p. 73.
- [7] CHUNG, M. C. **Planejamento e síntese de pró-fármacos recíprocos de nitrofural e primaquina potencialmente antichagásicos.** 1996. 196 f. Tese (Doutorado) – Faculdade de Ciências Farmacêuticas de São Paulo, Universidade de São Paulo. 1996.
- [8] PESSÔA, S. B.; MARTINS, A. V. **Parasitologia médica.** 11. ed. Rio de Janeiro: Ed. Guanabara Koogan, 1988. p. 872.
- [9] PANEPUCCI, E. H. **Modelagem Molecular da Enzima gliceraldeído-3-fosfato desidrogenase de *T. cruzi* e Análise de Potenciais Inibidores Específicos.** 1994. 72 f. Tese (Mestrado) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 1994.
- [10] ROCHA, A. L. L.; ROCHA, M. O. C.; TEIXEIRA, B. O. S.; LOMBARDI, F.; ABREU, C. D. G.; BITTENCOURT, R. J.; BARROS, M. V. L.; RIBEIRO, A. L. P. Índice cronotrópico-metabólico na doença de Chagas. **Rev. Soc. Bras. Med. Trop.**, v. 38, n. 5, p. 373-376, 2005.

- [11] DARDONVILLE, C. Recent advances in antitrypanosomal chemotherapy: patent literature 2002-2004. **Exp. Opin. Ther. Patents**, v. 15, n. 9, p. 1-17, 2005.
- [12] MANSOUR, T. E. **Chemotherapeutic targets in parasites: contemporary strategies**. New York: Cambridge University Press, 2002. p. 90.
- [13] VERONESI, R. **Doenças infecciosas e parasitárias**. 8. ed. Rio de Janeiro: Guanabara Koogan, 1991. p. 674-705.
- [14] URBINA, J. A.; DOCAMPO, R. Specific chemotherapy of Chagas disease: controversies and advances. **Trends in parasitology**, v. 19, n. 11, p. 495-501, 2003.
- [15] VIODÉ, C.; BETTACHE, N.; CENAS, N.; KRAUTH-SIEGEL, R. L.; CHAUVIÈRE, G.; BAKALARA, N.; PÉRIÉ, J. Enzymatic reduction studies of nitroheterocycles. **Biochem. Pharmacol.**, v. 57, p. 549-557, 1999.
- [16] DOCAMPO, R.; SCHMUÑIS, C. A. Sterol biosynthesis inhibitors potential chemotherapeutics against Chagas disease. **Parasitol. Today**, v. 13, p. 127-128, 1997.
- [17] ANDERSON, A. C. The process of Structure-based drug design. **Chem. & Biol.**, v. 10, p. 787-797, 2003.
- [18] PINK, R.; HUDSON, A.; MOURIÈS, M. A.; BENDIG, M. Opportunities and challenges in antiparasitic drug discovery. **Nature**, v. 4, p. 727, 2005.
- [19] VEGA-TEJIDO, M.; CARACELLI, I.; ZUKERMAN-SCHPECTOR, J. Conformational analyses and docking studies of a series of 5-nitrofuranyl- and 5-nitrothiophen-semicarbazone derivatives in three possible binding sites of trypanothione and glutathione reductases. **J. Mol. Grap. Model.**, v. 24, p. 349-355, 2006.
- [20] MALTA, V. R. S. **Estudo cristalográfico de naftoquinonas e seus derivados e cálculos teóricos de propriedades relevantes na relação estrutura-atividade**. 2000. 258 f. Tese (Doutorado em Físico-Química) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2000.
- [21] ABREU, F. C.; FERRAZ, P. A. F.; GOULART, M. O. F. Some applications of electrochemistry. Emphasis on the correlation of electrochemical and bioactive properties. **J. Braz. Chem. Soc.**, v.13, n. 1, p. 19-35, 2002.

- [22] SOLOMONS, T. W. G.; FRYHLE, C. B. **Química Orgânica**. Rio de Janeiro: Livros Técnicos e Científicos, 2006. v. 2.
- [23] SILVA, M. N.; FERREIRA, V. F.; SOUZA, M. C. B. V. Um panorama atual da química e da farmacologia de naftoquinonas, com ênfase na β -lapachona e derivados. **Quím. Nova**, v. 26, n. 3, p. 407-416, 2003.
- [24] ENCYCLOPEDIA of chemical technology. New York: John Wiley, 1982. v. 19, p. 573-599.
- [25] PORTELA, M. P. M.; VILLAMIL, S. H. F.; STOPPANI, A. O. M. Redox cycling of o-naphthoquinone in trypanosomatids. **Biochem. Pharmac.**, v. 52, p. 1875, 1996.
- [26] SUBRAMANIAN, S.; FERREIRA, M. M. C.; TRSIC, M. A structure-activity relationship of lapachol and some derivatives of 1,4-naphthoquinones against carcinosarcoma Walker 256. **Struc. Chem.**, v. 9, n. 1, p. 47-57, 1998.
- [27] SALMON-CHEMIN, L.; BUISINE, E.; YARDLEY, V.; KOHKER, S.; DEBREU, M. A.; LANDRY, V.; SERGHERAERT, C.; CROFT, S. L.; KRAUTH-SIEGEL, R. L.; DAVIOUD-CHARVET, E. 2- and 3-substituted 1,4-naphthoquinone derivatives as subversive substrates of trypanothione reductase and lipoamide dehydrogenase from *Trypanosoma Cruzi*: synthesis and correlation between redox cycling Activities and in vitro cytotoxicity. **J. Med. Chem.**, v. 44, p. 548-565, 2001.
- [28] RIBEIRO-RODRIGUES, R.; DOS SANTOS, W. G.; OLIVEIRA, A. B.; ZANI, C. L.; ROMANHA, A. J. Growth inhibitory effect of naphthofuran and naphthofuranquinone derivatives on *Trypanosoma Cruzi* epimastigotes. **Bioorg. Med. Chem. Lett.**, v. 5, n. 14, p. 1509-1512, 1995.
- [29] BOVERIS, A.; DOCAMPO, R.; TURRENS, J. F.; STOPPANI, A. Effect of β -lapachona on superoxide anion and hydrogen peroxide production in *Trypanosoma Cruzi*. **Biochem. J.**, v. 175, p. 431-439, 1978.
- [30] VOET, D.; VOET, J. G.; PRATT, C. W. **Fundamentos da bioquímica**. Porto Alegre: Artes Medicas Sul, 2000. p. 423-526.
- [31] BERG, J. M.; TYMOCZKO, J. L.; STRYER, L. **Bioquímica**. 5. ed. Rio de Janeiro: Guanabara Koogan, 2004. 525p.

[32] FERREIRA, A. L. A.; MATSUBARA, L. S. Radicais livres: conceitos, doenças relacionadas, sistema de defesa e estresse oxidativo. **Rev. Ass. Med. Brasil**, v. 43, n. 1, p. 61-68, 1997.

[33] BOND, C. S.; ZHANG, Y.; BERRIMAN, M.; CUNNINGHAM, M. L.; FAIRLAMB, A. H.; HUNTER, W. N. Crystal structure of *Trypanosoma Cruzi* Trypanothione Reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors. **Structure**, v. 7, n. 1, p. 81-89, 1999.

[34] GARRARD, E. A.; BORMAN, E. C.; COOK, B. N.; PIKE, E. J.; ALBERG, D. G. Inhibition of Trypanothione Reductase by substrate analogues. **Org. Letters**, v. 2, n. 23, p. 3639-3642, 2000.

[35] CENAS, N. K.; ARSCOTT, D.; WILLIAMS, C. H.; BLANCHARD, J. S. Mechanism of reduction of quinones by *Trypanosoma congolense* Trypanothione Reductase. **Biochemistry**, v. 33, p. 2509-2515, 1994.

[36] STOLL, V. S.; SIMPSON, S. J.; KRAUTH-SIEGEL, R. L.; WALSH, C. T.; PAI, E. F. Glutathione reductase turned into Trypanothione Reductase: structural analysis of an engineered change in substrate specificity. **Biochemistry**, v. 36, p. 6437-6447, 1997.

[37] MEIERING, S.; INHOFF, O.; MIES, J.; VINCEK, A.; GARCIA, G.; KRAMER, B.; DORMEYER, M.; KRAUTH-SIEGEL, R. L. Inhibitors of *Trypanosoma cruzi* Trypanothione Reductase revealed by virtual screening and parallel synthesis. **J. Med. Chem.**, v. 48, p. 4793-4802, 2005.

[38] DIXON, M. J.; MAURER, R. I.; BIGGI, C.; OYARZABAL, J.; ESSEX, J. W.; BRADLEY, M. Mechanism and structure-activity relationships of norspermidine-based peptidic inhibitors of trypanothione reductase. **Bioorg. Med. Chem.**, v. 13, p. 4513-4526, 2005.

[39] BONSE, S.; SANTELLI-ROUVIER, C.; BARBE, J.; KRAUTH-SIEGEL, R. L. Inhibitor of *Trypanosoma cruzi* Trypanothione Reductase by acridines: kinetic studies and structure-activity relationships. **J. Med. Chem.**, v. 42, p. 5448-5454, 1999.

[40] KRAUTH-SIEGEL, R. L.; INHOFF, O. Parasite-specific Trypanothione Reductase as drug a drug target molecule. **Parasitol. Res.**, v. 90, p. 77-85, 2003.

[41] VIEGAS JUNIOR, C.; BOLZANI, V. S.; BARREIRO, E. J. Os produtos naturais e a química medicinal moderna. **Quím. Nova**, v. 29, n. 2, p. 326-337, 2006.

[42] MONTANARI, C. A.; BOLZANI, V. S. Planejamento de fármacos baseado em produtos naturais. **Quím. Nova**, v. 24, n. 1, p. 105-111, 2001.

[43] LEITÃO, A; MONTANARI, C. A; DONNICI, C. Sobre o uso de métodos quimiométricos em química combinatória. **Quím. Nova**, v. 23, n. 2, p. 178-184, 2000.

[44] KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. **DDT**, v. 11, n. 13-14, p. 580-594, 2006.

[45] LYNE, P. D.; KENNY, P.W.; COSGROVE, D. A.; DENG, C.; ZABLUDOFF, S. WENDOLOSKI, J. J., ASHWELL, S. Identification of compounds with nanomolar binding affinity for checkpoint kinase-1 using knowledge based screening. **J. Med. Chem.**, v. 47, p. 1962-1968, 2004.

[46] LYNE, P. D. Structure-based virtual screening: an overview. **DDT**, v. 7, n. 20, p. 1047-1055, 2002.

[47] KLEBE, G. Virtual ligand screening: strategies, perspectives and limitations. **DDT**, v. 11, n. 13-14, p. 580-594, 2006.

[48] JORGENSEN, W. L. The many roles of computation in drug discovery. **Science**, v. 303, p. 1813-1818, 2004.

[49] AMARAL, A. T.; MONTANARI, C. A. Química Medicinal: 25 anos de planejamento racional de fármacos. **Quím. Nova**, v. 25, p. 39-44, 2002.

[50] KLEBE, G. Recent developments in structure-based drug design. **J. Mol. Med.**, v. 78, p. 269-281, 2000.

[51] ABRAHAM, D. J. **Burger's Medicinal Chemistry Drug Discovery**. 6. ed. New York: John Wiley, 2002. v. 1, p. 187-356.

[52] ORRY, A. J. W.; ABAGYAN, R. A.; CAVASSOTTO, C. N. Structure-based development of target-specific compound libraries. **DDT**, v. 11, n. 5/6, p. 261-266, 2006.

[53] IRWIN, J. J.; SHOICHET, B. K. ZINC – A free database of commercially available compounds for virtual screening. **J. Chem. Inf. Model**. v. 45, p. 177-182, 2005.

- [54] LIPINSKI, C.; HOPKINS A. Navigating chemical space for biology and medicine. **Nature**, v. 432, p. 855-861, 2004.
- [55] WALTERS, W. P.; STAHL, M. T.; MURCKO, M. A. Virtual screening –an overview. **DDT**, v. 3, n. 4, p. 160-177, 1998.
- [56] ROCHE, O.; SCNEIDER, P.; ZUEGGE, J.; GUBA, W.; KANSY, M.; ALANINE, A.; BLEICHER, K.; DANIEL, F.; GUTKNECHT, E. M.; ROGERS-EVANS, M.; NEIDHART, W.; STALDER, H.; SCHNEIDER, G. Development of virtual screening method for identification of frequent hitters in compound libraries. **J. Med. Chem.** v. 45, p. 137-142, 2002.
- [57] WESS, G.; URMANN, M.; SICKENBERGER, B. Medicinal chemistry: Challenges and opportunities. **Angew. Chem. Int. Ed.**, v. 40, n. 18, p. 3341-33450, 2001.
- [58] WERMUTH, C. G. **The practice of medicinal chemistry**. 2. ed. London: Academic Press, 2003. p. 351-386.
- [59] WATERBEEMD, H. V.; GIFFORD, E. ADME *in silico* modeling: towards prediction paradise. **Nature**, v. 2, p. 192-204, 2003.
- [60] VERLI, H.; BARREIRO, E.J. Um paradigma da química medicinal: flexibilidade dos ligantes e receptores. **Quím. Nova**, v. 28, n. 1, p. 95-102, 2005.
- [61] SANTOS FILHO, O. A.; ALENCASTRO, R. B. Modelagem de proteínas por homologia. **Quím. Nova**, v. 26, n. 2, p. 253-259, 2003.
- [62] ATKINS, P.W. **Physical Chemistry**. 6.ed. Oxford: Oxford University, 1997. 434p.
- [63] PAULING, L. **Química Geral**. 2.ed. Rio de Janeiro: Livro Tecnico, 1966. 201p.
- [64] BUNGE, A.V. **Introdução à Química Quântica**. São Paulo: Edgard Blucher, 1977. 410p.
- [65] McWEENY, R. **Quantum-mechanics**: methods and basic applications. New York: Pergamon Press, 1973. 243p.
- [66] LEVINE, I.N. **Quantum chemistry**. 4.ed. New York: Prentice Hall, 1991. 628p.

[67] SZABO, A.; OSTLUND, N.S. **Modern quantum chemistry: introduction advanced electronics structure theory**. New York: Macmillian, 1982. 446p.

[68] DA SILVA, Albérico Borges Ferreira **Bases Universais Geradas pelo Método da Coordenada Geradora Hartree-Fock e suas aplicações em cálculos de propriedades atômicas e moleculares**. 1990. 136f. Tese (Doutorado) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 1990.

[69] VIANA, J. D. M.; FAZZIO, A.; CANUTO, S. **Teoria quântica de moléculas e sólidos: simulação computacional**. São Paulo: Ed. Livraria da Física, 2004. 213p.

[70] MORGON, N. H.; CUSTODIO, R. Teoria do Funcional de Densidade. **Quím. Nova**, v. 18, n. 1, p. 44-55, 1995.

[71] CAMARGO, A. J. **Estudo químico-quântico *ab initio* e semi-empírico de compostos inorgânicos e orgânicos com possíveis aplicações tecnológicas**. 2001. 195f. Tese (Doutorado) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 2001.

[72] CAPELLE, K. A bird's-eye view of density functional theory. Disponível em: <<http://arXiv.org/archive/cond-mat>>. Acesso em: 04 ago. 2005.

[73] DUARTE, H. A. Índices de reatividade química a partir da teoria do funcional de densidade: Formalismo e perspectivas. **Quím. Nova**, v. 24, n. 4, p. 501-508, 2001.

[74] HOHEMBERG, P; KOHN, W. Inhomogeneous Electron Gas. **Phys. Rev.**, v.136, p. 864, 1964.

[75] KOHN, W.; SHAM, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. **Phys. Rev.**, v. 140, p. 1133, 1965.

[76] SLATER, J. C. **The Self-Consistent field for Molecular and Solids**. New York: McGraw-Hill, 1974. v. 4.

[77] BECKE, A. D. Density-functional exchange-energy approximation with correct asymptotic behavior. **Phys. Rev. A**, v. 38, p. 3098, 1988.

[78] VOSKO, S. H.; WILK, L.; NUSAIR, M. Accurate spin-dependent electron liquid correlation energies for local spin density calculations: a critical analysis. **Canadian J. Phys.**, v. 58, p. 1200, 1980.

[79] LEE, C.; YANG, W.; PARR, R.G. Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density. **Phys. Rev. B**, v. 37, p. 785, 1988.

[80] MIEHLICH, B.; SAVIN, A.; STOLL, H.; PREUSS, H. Results obtained with the correlation-energy density functionals of Becke and Lee, Yang and Parr. **Chem. Phys. Lett.**, v. 157, p. 200, 1989.

[81] PERDEW, J. P. Density-functional approximation for the correlation energy of the inhomogeneous electron gas. **Phys. Rev. B**, v. 33, p. 8822, 1986.

[82] PERDEW, J. P.; WANG, Y. Accurate and simple analytic representation of the electron-gas correlation energy. **Phys. Rev. B**, v. 45, p. 13244, 1992.

[83] BRUNS, R.E.; FAIGLE, J.F.G. Quimiometria. **Quim. Nova**, v. 8, p. 84, 1985.

[84] MORGANO, M. N.; QUEIROZ, S.C.N.; FERREIRA, M. M. C. Determinação dos teores de minerais em sucos de frutas por espectrometria de emissão ótica em plasma indutivamente acoplado (ICP-OES). **Cienc. Tecnol. Aliment.**, v. 19, n. 3, p. 344-348, 1999.

[85] FERREIRA, M.M.C.; FARIA, C.G.; PAES, E.T. Oceanographic Characterization of Northern São Paulo Coast: a Chemometric study. **Chemom. Intell. Lab. Syst.**, v. 47, p. 289-297, 1999.

[86] WELDSCHOLTE, C.M.; KROONENBERG, P.M.; ANTONIDES, G. Three-Mode Analysis of Perceptions of Economic Activities in Eastern and Western Europe. **J. Econ. Psyc.**, n. 19, p. 321-351, 1998.

[87] TAYLOR, S; COX, B.J. Anxiety Sensitivity: Multiple Dimensions and Hierarchical Structure. **Beh. Res. Ther.**, n. 36, p. 37-51, 1998.

[88] SENA, M. M.; COLLINS, C.C.; SCARMINIO, I.S. Aplicação de métodos quimiométricos na especificação de Cr(VI) em solução aquosa. **Quim. Nova**, v. 24, n. 3, p. 331-338, 2001.

[89] DE BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. **Planejamento e Otimização de experimentos**. Campinas: Ed. Unicamp, 1995. 133p.

[90] CARVALHO, I.; PUPO, M. T.; BORGES, A. D. L.; BERNARDES, L. S. C. Introdução a modelagem molecular de fármacos no curso experimental de química farmacêutica. **Quim. Nova**, v. 26, n. 3, p. 428-438, 2003.

[91] HOLTJE, H. D. **Molecular modeling**: basic principles and applications. New York: VCH, 1997. 194p.

[92] EL-AZHARY, A. A.; SUTTER, H.U. Comparison between Optimized Geometries and Vibrational Frequencies Calculated by the DFT Methods. **J. Phys. Chem.**, v. 100, p. 15056, 1996.

[93] TURECEK, F. Proton affinity of dimethyl sulfoxide and relative stabilities of C₂H₆OS molecules and C₂H₇OS⁺ ions. A comparative G2(MP2) ab initio and density functional theory study. **J. Phys. Chem. A**, v. 102, n. 34, p. 4703-4713, 1998.

[94] FRISCH, M. J.; TRUCKS, G.; HEAD-GORDON, M.; GILL, P. M. W.; JOHNSON, B. G.; ROBB, M. A.; CHEESXMAN, J. R.; KEITH, T. A.; PETERSON, G. A.; MONTGOMERY, J. A.; RAGHAVACHARI, K.; AL-LAHAN, M. A.; ZAKRZEWSKI, V. G.; ORTIZ, J. V.; FORESMAN, J. B.; CIOŚŁOWSKI, J.; STEFANOV, B. B.; NORAYAKKARA, A.; CHALLOCOMBE, M.; PENG, C. Y.; AYALA, P. Y.; CHEN, W.; WONG, M. W.; ANDES, J. L.; REPLOGLE, E. S.; GOMPERTS, R.; MARTIN, R. L.; FOX, D. J.; BINKLEY, J. S.; DEFREES, D. J.; BAKER, J.; STEWART, J. P.; HEAD-GORDON, M.; GONZALEZ, C.; POPLE, J. A. **GAUSSIAN 98**. Pittsburg: GAUSSIAN, 1998. [Programa de Computador].

[95] BRENEMAN, C. M.; WIBERG, K. B. Determining atom-centered monopoles from molecular electrostatic potentials - the need for high sampling density in formamide conformational-analysis. **J. Comput. Chem.**, v. 11, n. 3, p. 361-373, 1990.

[96] WANG, R.; FU, Y. LUHUA, L. A new atom-additive method for calculating partition coefficients. **J. Chem. Inf. Comput. Sci.**, v. 37, p. 615-621, 1997.

[97] MILANO CHEMOMETRICS. **Dragon**. [Programa de Computador]. Milano, 2002.

[98] INFOMETRIX. **Pirouette 3.10**. [Programa de Computador]. Woodinville, 2002.

[99] MINITAB 10. [Programa de Computador]. Wadsworth, 1994.

- [100] MATHWORKS. **MATLAB 6.1**. [Programa de computador]. Sherborn, 1991.
- [101] SHARAF, M.A.; ILLMAN, D.L.; KOWALSKI, B.R. **Chemometrics**. New York: John Wiley, 1986. 332p.
- [102] MANLY, B.F.J. **Multivariate statistical methods**. New York: Chapman and Hall, 1986. p. 154.
- [103] MASSART, D.L. **Chemometrics tutorials**. Amsterdam: Elsevier, 1990. 427p.
- [104] ESBENSEN, K.; GELADI, P. Principal component analysis. **Chem. Intell. Lab. Syst.**, v. 2, p. 37-52, 1987.
- [105] KOWALSKI, B.R. **Chemometrics: mathematics and statistics in chemistry**. Cosenza: Reidel: 1984. 485p.
- [106] MALINOWSKI, B.R. **Factor analysis in chemistry**. 3.ed. New York: John Wiley, 1991. 350p.
- [107] BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M.B. **Chemometrics: a practical guide**. New York: Wiley-Interscience, 1998. 348p.
- [108] THOMAS, E.V. A primer on a multivariate calibration. **Anal. Chem.**, v. 66, n. 15, p. 795-803, 1994.
- [109] HUBERTY, C.J. **Applied Discriminant Analysis**. New York: John Wiley, 1994. 466p.
- [110] JOHNSON, R.A.; WICHERN, D.W. **Applied Multivariate Statistical Analysis**. 4.ed. New Jersey: Prentice-Hall, 1992. p. 629.
- [111] CALAS, M; CORDINA, G.; BOMPART, J.; BARI, M.B.; JEI, T.; ANCELIN, M.L.; VIAL, H. Antimalarial Activity of Molecules Interfering with *Plasmodium falciparum* Phospholipid Metabolism. **J. Med. Chem.**, v. 40, n. 22, p. 3557-3566, 1997.
- [112] SRIVASTAVA, M.S.; CARTER, E.M. **An Introduction to applied multivariate statistics**. New York: Elsevier, 1983. 243p.

- [113] HAIR JUNIOR, J. F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Multivariate data analysis with readings**. 4.ed. New Jersey: Prentice-Hall, 1995. 197p.
- [114] FERREIRA, M.M.C. Multivariate QSAR. **J. Braz. Chem. Soc.**, v. 13, n. 6, p. 742-753, 2002.
- [115] LEMES, M. R.; PINO JUNIOR, A. D. Estudo do estado fundamental de aglomerados em silício via redes neurais. **Quím. Nova**, v. 24, n. 4, p. 539, 2002.
- [116] TERFLOTH, L.; GASTEIGER, J. Neural networks and genetic algorithms in drug design. **DDT**, v. 6, n. 15, s102-s108, 2001.
- [117] BAJORATH, J. Integration of virtual and high-throughput screening. **Nature**, v. 1, p. 882-894, 2002.
- [118] BRAGA, A. P.; CARVALHO, A. C. P. L. F.; LUDEMIR, T. B. **Redes neurais artificiais**. Rio de Janeiro: Livros Técnicos e científicos, 2000. 2.p.
- [119] SCHIMIDT, F.; BUENO, M. I. S.; POPPI, R. J. Aplicação de alguns modelos quimiométricos à espectroscopia de fluorescência de raio-X de energia dispersiva. **Quím. Nova**, v. 25, n. 6, p. 949-956, 2002
- [120] CERQUEIRA, E. O.; ANDRADE, J. C.; POPPI, R. J. Redes Neurais e suas aplicações em calibração multivariada. **Quím. Nova**, v. 24, n. 6, p. 864-873, 2001.
- [121] PHAM, D. T. **An introduction to artificial neural networks: Neural networks for chemical engineers**. New York: Elsevier, 1995. p. 1-19.
- [122] GOULART, M. O. F.; ZANI, C. L.; TONHOLO, J.; FREITAS, L. R.; DE ABREU, F. C.; OLIVEIRA, A. B.; RASLAN, D. S.; STARLING, S.; CHIARI, E. Trypanocidal activity and redox potential of heterocyclic and 2-hydroxy-naphthoquinones. **Bioorg. Med. Chem. Lett.**, v. 7, p. 2043, 1997.
- [123] SUBRAMANIAN, S. **Estudo da relação estrutura-atividade biológica do lapachol, das lapachonas e de alguns de seus derivados**. 1996. 253f. Tese (Doutorado) – Instituto de Química de São Carlos, Universidade de São Paulo, São Carlos, 1995.
- [124] ALVAREZ, J. C. High-throughput docking as a source of novel drug leads. **Curr. Opin. Chem. Biol.**, v. 8, p. 365-370, 2004.

- [125] VERLI, H.; BARREIRO, E.J. Um paradigma da química medicinal: flexibilidade dos ligantes e receptores. **Quím. Nova**, v. 28, n. 1, p. 95-102, 2005.
- [126] KITCHEN, D. B.; DECORNEZ, H.; FURR, J. R.; BAJORATH, J. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nat. Rev.**, v. 3, p. 935-949, 2004.
- [127] SCHNEIDER, G.; BÖHM, H. J. Virtual screening and fast automated docking methods. **DDT**, v. 7, n. 1, p. 64-70, 2002.
- [128] TAYLOR, R. D.; JEWSBURY, P. J.; ESSEX, J. W. A review of protein-small molecule docking. **J. Comp.Aided Mol. Des.**, v. 16, p. 151-166, 2002.
- [129] MORRIS, G. M.; GOODSSELL, D. S.; HALLIDAY, R. S.; HUEY, R.; HART, W. E.; BELEW, R.K.; OLSON, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. **J. Comput. Chem.**, v. 19, p. 1639-1662, 1998.
- [130] KRAMER, B.; RAREY, M.; LENGAUER, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. **Proteins**, v. 37, n. 2, p. 228-241, 1999.
- [131] ALONSO, H.; GILLIES, M. B.; CUMMINS, P. L.; BLIZNYUK, A. A.; GREARY, J. E. Multiple ligand-binding modes in bacterial r67 dihydrfolate reductase. **J. Comp.Aided Mol. Des.**, v. 19, p. 165-187, 2005.
- [132] HALPERIN, I.; MA, B.; WOLFSON, H. NUSSINOV, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. **Proteins**. v. 47, p. 409-443, 2002.
- [133] BRANSON, K. M.; SMITH, B. J. The role of virtual screening in computer aided structure-based drug desing. **Aust. J. Chem.**, v. 57, p. 1029-1037, 2004.
- [134] GHOSH, S., NIE, A.; AN, J.; HUANG, Z. Structure-based virtual screening of chemical libraries for drug discovery. **Curr. Opin. Chem. Biol.**, v. 10, p. 194-202, 2006.
- [135] OPREA, T. I.; MATTER, H. Integrating virtual screening in lead discovery. **Curr. Opin. Chem. Biol.**, v. 8, p. 349-358, 2004.

[136] ZANI, C. A.; CHIARI, E.; KRETTLI, A. U.; MURTA, S. M. F.; CUNNINGHAM, M. L.; FAIRLAMB, A. H.; ROMANHA, A. J. Anti-plasmodial and anti-trypanosomal activity of synthetic naphtho[2,3-b] thiophen-4-9-quinones. **Bioorg. Med. Chem.**, v. 5, n. 12, p. 2185-2192, 1997.

[137] ZANI, C. A.; FAIRLAMB, A. H. 8-methoxy-naphtho[2,3-b]thiophen-4,9-quinone, a non- competitive inhibitor of trypanothione reductase. **Mem. Inst. Oswaldo Cruz.**, v. 98, n. 4, p. 565-568, 2003.

[138] GASTEIGER, J.; MARSILI, M. Iterative Partial Equalization of orbital electronegativity. A rapid access to atomic charges. **Tetrahedron**, v. 36, p. 3219-3288, 1980.

[139] VEGA ZZ Release 2.0.5.52. [Programa de computador]. Disponível em:<<http://www.ddl.unimi.it/vega/index2.htm>> . Acesso em: 22 maio 2006.

[140] PETTERSEN, E.F., GODDARD, T.D., HUANG, C.C., COUCH, G.S., GREENBLATT, D.M., MENG, E.C., AND FERRIN, T.E. UCSF Chimera - A Visualization System for Exploratory Research and Analysis. **J. Comput. Chem.**, v. 25, n. 13, p. 1605-1612, 2004.

[141] MORRIS, G. M.; GOODSELL, D. S.; HUEY, R.; OLSON, A. J. Distributed automated docking of flexible ligands to proteins: parallel applications of Autodock 2.4. **J. Comp. Aided Mol. Des.**, v. 10, p. 293-304, 1996.

[142] SOLIS, F. J.; WETS, R. J. B. Minimization by random search techniques. **Mat. Oper. Res.**, v. 6, n. 1, p. 19-30, 1981.

[143] SARAVANAMUTHU, A.; VICKERS, T. J.; BOND, C. S.; PETERSON, M. R.; HUNTER, W. N.; FAIRLAMB, A. H. Two interacting binding sites for quinacrine derivatives in the active site of trypanothione reductase. **J. Biol. Chem.**, v. 279, n. 28, p. 29493-29500, 2004.

[144] BONSE, S.; SANTELLI-ROUVIER, C.; BARBE, J.; KRAUTH-SIEGEL, R. L. Inhibition of Trypanosoma Cruzi trypanothione reductase by acridines: kinetic and structure activity relationship. **J. Med. Chem.**, v. 42, p. 5448-5454, 1999.

[145] FREITAS, R. F.; GALEMBECK, S. E. Computational study of the interaction between TIBO inhibitors and Y181 (C181), K101, and Y188 amino acids. **J. Phys. Chem. B.**, 2006. In press.

[146] COUTURE, J. F.; HAUKE, G.; THOMPSON, M. J.; BLACKBURN, G. M.; TRIEVEL, R. C. Catalytic roles for carbon-oxygen hydrogen bonding in SET domain lysine methyltransferase. **J. Biol. Chem.**, v. 281, n. 28, p. 19280-19287, 2006.