



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE QUÍMICA
COLÉGIADO DE PÓS-GRADUAÇÃO EM QUÍMICA**

**ESTUDO DA RELAÇÃO ESTRUTURA-ATIVIDADE (SAR) DE
DERIVADOS DA ARTEMISININA SOLÚVEIS EM ÁGUA**

Ricardo Morais de Miranda

**BELÉM-PA
2008**

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.



**UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE QUÍMICA
COLÉGIO DE PÓS-GRADUAÇÃO EM QUÍMICA**

**ESTUDO DA RELAÇÃO ESTRUTURA-ATIVIDADE (SAR) DE
DERIVADOS DA ARTEMISININA SOLÚVEIS EM ÁGUA**

Ricardo Morais de Miranda

Dissertação apresentada, como requisito final, ao curso de Pós Graduação em química, para obtenção título de Mestre em Química. Orientador Prof. Dr. José Ciriaco Pinheiro.

**BELÉM-PA
2008**



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
FACULDADE DE QUÍMICA
COLÉGIO DE PÓS-GRADUAÇÃO EM QUÍMICA

**ESTUDO DA RELAÇÃO ESTRUTURA-ATIVIDADE (SAR) DE
DERIVADOS DA ARTEMISININA SOLÚVEIS EM ÁGUA**

Ricardo Morais de Miranda

Banca Examinadora

Prof. Dr. José Ciríaco Pinheiro
Orientador
(Faculdade de Química da UFPA)

Prof. Dr. Oscar Augusto Sanchez Romero
(Faculdade de Química da UFPA)

Prof. Dr. Edinaldo Teixeira
(Faculdade de Física da UFPA)

Julgado em: ____/____/____

Conceito: _____

**"O insucesso é apenas uma
oportunidade para recomeçar de
novo com mais inteligência."
*Henry Ford***

Com a mais sincera gratidão
dedico este trabalho a meu
avô José Paulo (*in memoriam*)
e minha avó Maria Natividade.
“Por todo ensinamento”

AGRADECIMENTO

A Deus por permitir alcançar tantas vitórias na vida, à Minha esposa Cláudia Gatinho de Miranda pelo amor e companheirismo, a meu filho Pedro Vinícius Gatinho de Miranda pela felicidade que me proporciona.

A minha mãe Maria do Socorro e a Armando Quadros por todo esforço que fizeram para que essa realizações acontecessem.

A meus irmãos: Rogério, Rebeca, Roseane e Maria.

Em especial ao Prof. Dr. José Ciríaco Pinheiro pela orientação, dedicação, amizade e pela demonstração de como se fazer pesquisa com seriedade, honestidade e muita força de vontade.

Prof. Msc. Antônio Florêncio Figueiredo pelas contribuições.

Aos meus colegas do LQTC, que me ajudaram conclusão de mais esta etapa de minha vida, Edson Barbosa, Fabio Gil, João Elias, Fábio Cardoso, Fábio Lira, Raimundo Dirceu, Ruth Catarine, Jardel Pinto, Bruno Malheiros, Elierge Costa e Marcos Antônio.

Ao CNPq pelo apoio financeiro.

A todos que direta ou indiretamente me ajudaram nesta realização.

SUMÁRIO

RESUMO	8
ABSTRACT	9
LISTA DE TABELAS	10
LISTA DE FIGURAS	11
ABREVIATURAS	12
1 – MALÁRIA	13
1.2 – HISTÓRICO DA MALÁRIA	14
1.3 – VETOR CAUSADORA DA MALÁRIA.....	16
1.4 – ATIVIDADE ANTI-ANTIMALÁRICA	18
1.5 – ARTEMISININA.....	19
2 – METODOLOGIA	20
2.1 - MÉTODOS QUÂNTICOS E PROCEDIMENTOS COMPUTACIONAIS.....	20
2.1.1 – Introdução	20
2.1.2 – A Equação de Schrödinger	21
2.1.3 – Aproximação de Born-Oppenheimer	22
2.1.4 – Aproximação de Hartree	24
2.1.5 – Aproximação Hartree-Fock.....	26
2.1.6 - As Equações de Roothaan.....	29
3 – MÉTODOS QUIMIOMÉTRICOS	32
3.1 – INTRODUÇÃO.....	32
3.2 - ANÁLISE DE COMPONENTES PRINCIPAIS – PCA	32
3.3 - ANÁLISE HIERARQUICA DE CLUSTER – HCA.....	34
3.4 – KNN.....	35
3.5 – SIMCA.....	36
4 – INTERAÇÕES ARTEMISININA E DERIVADOS COM RECEPTOR	
“DOCKING MOLECULAR	38
5 - RESULTADOS E DISCUSSÃO	41
6 – CONCLUSÃO	65
7 – REFERÊNCIAS	67

RESUMO

A artemisinina com um conjunto de 15 compostos de sua derivação, foi estudada com o intuito de se obter informação a respeito da atividade antimalarial. Tendo em vista que esta classe de compostos apresenta atividade antimalárica comprovada contra clones de malária *Plasmodium falciparum* w-2 provenientes da Indochina. Estes clones apresentam resistências a mefloquina, quinina e cloroquina. Métodos de química quântica (Método Hartree-Fock – 6-31G**) e multivariados (PCA, HCA, KNN e SIMCA) foram usados de modo a se obter um conjunto de informações pertinentes que pudessem fornecer subsídios para propor novos derivados antimaláricos. Com estudos multivariados de PCA obteve-se a seleção dos seguintes descritores Lumo+1, HE, MAXDN e ALOGPS-logs que foram responsáveis pela separação dos compostos com alta atividade (12 compostos) dos compostos de baixa atividade (4 compostos). A variância máxima obtida para três PCs foi de 96,85%. O método KNN foi usado para classificação inicial do conjunto de dados do conjunto treinamento. Os resultados obtidos com um (1KNN), três (3KNN) e cinco (5KNN) vizinhos próximos, onde obteve-se 100% de informações corretas. A utilização do método SIMCA também possibilitou a classificação dos compostos do conjunto treinamento em duas classes, sendo a classe 1 formada pelos compostos com alta atividade (apresentando 12 compostos) e a classe 2 formada pelos compostos com baixa atividade (apresentando 4 compostos). Os modelos construídos com os métodos PCA, HCA, KNN e SIMCA foram utilizados na predição da atividade antimalárica de 15 compostos (conjunto teste). Dentre os 15 compostos submetidos à predição obteve-se 7 compostos que foram considerados com alta atividade. Análise complementar do mapa de potencial eletrostático molecular (MEP) foi usada para aumentar as características responsáveis pela atividade biológica apresentada pelos compostos estudados, juntamente com a construção dos modelos de docking molecular para as moléculas identificadas com alta atividade no conjunto teste.

Palavras Chaves: Artemisinina. Relação Estrutura atividade. Mapas MEP. Docking Molecular.

ABSTRACT

The artemisinin with a set of fifteen composed of its derivation was studied with the intention of obtain information regarding the activity antimalarial. Tends in view that this class of compositions presents antimalarial activity proven against malaria *Plasmodium Falciparum* clones w-2 from Indochina. These clones present resistances the mefloquina, quinine and cloroquine. Quantum chemistry of methods (Método Hartree-Fock - 6-31G * *) and multivariates (PCA, HCA, KNN and SIMCA) they were used in way the if to obtain a group of pertinent information that could supply subsidies to propose derived new antimalárical. With studies multivariates of PCA it was obtained the selection of the following descriptores Lumo+1, HE, MAXDN and ALOGPS-logs that were responsible for the separation of the composed with high activity (12 composed) of the composed of low activity (4 composed). The maximum variance obtained for three PCs was of 96,85%, The KNN method was used for initial classification of the group of data of the group training. The results obtained with a (1KNN), three (3KNN) and five (5KNN) close neighbors, where it was obtained 100% of correct information. The use of the SIMCA method also made possible the classification of the composed of the group training in two classes, being the class 1 formed by the composed with high activity (presenting 12 composed) and the class 2 formed by the composed with low activity (presenting 4 composed). The models built with the PCA, HCA, KNN and SIMCA methods were used in the prediction of the activity antimalárical of 15 composed (group tests). Among the fifteen compositions submitted to the prediction was obtained 7 composed that were considered with high activity. Complemental analysis of the map of molecular electrostatic potential (MEP) it was used to increase the responsible characteristics for the biological activity presented on the studied compositions, together with the construction of the models of molecular docking for the identified molecules with high activity in the group tests.

Key Words: Artemisinin. Relationship Structures activity. Maps MEP. Docking Molecular.

LISTA DE TABELAS

Tabela 1 –	Parâmetros geométricos experimentais e teóricos e do anel 1, 2, 4-trioxano da artemisinina, obtidos com o método Hartree-Fock com diferentes conjuntos de base.....	44
Tabela 2 –	Descritores disponíveis pelo programa E-DRAGON SOFTWARE.....	46
Tabela 3 –	Descritores responsáveis pela separação e sua classificação.....	47
Tabela 4 –	Propriedades físico-químicas selecionadas por análise de componentes principais, log AR experimental e descritores responsáveis pela separação das estruturas de alta atividade e baixa atividade.....	48
Tabela 5 –	Informações descritas por cada PC.....	49
Tabela 6 –	Contribuição dos descritores para cada PC.....	49
Tabela 7 –	Método KNN para o conjunto treinamento para 16 compostos.....	52
Tabela 8 –	Classes formadas entre os compostos do conjunto treinamento através da aplicação do método SIMCA.....	53
Tabela 9 –	Resultados dos métodos multivariados para os compostos do conjunto teste. (HÁ – alta atividade e LA – baixa atividade).....	58
Tabela 10 –	Resultados do docking entre heme-PDB / artemisinina e derivados com HF/6-31G**.....	61

LISTA DE FIGURAS

Figura 1 –	Área endêmica de malária	13
Figura 2 –	<i>Plasmodium Falciparum</i>	17
Figura 3 –	Ciclo de desenvolvimento da malária.....	17
Figura 4 –	Drogas antimaláricas: quinino, cloroquina e mefloquina.....	18
Figura 5 –	Classificação dos K vizinhos próximos pertencentes a três classes.....	35
Figura 6 –	Representação gráfica de uma matriz com três colunas e dividida em duas classes. x_1, x_2 e x_3 representam as colunas da matriz; PC1 o primeiro componente principal, PC2 o segundo componente principal, v_1 a variância em PC1, v_2 a variância em PC2, o desvio padrão dos resíduos é representado por e	37
Figura 7 –	Estrutura da molécula heme.....	38
Figura 8 –	Mecanismo de ação artemisinina e heme proposto por Posner.....	39
Figura 9 –	Mecanismo de ação artemisinina e heme proposto por Jefford.....	40
Figura 10 –	Estruturas estudadas no conjunto treinamento, com atividade antimalarial contra <i>Plasmodium falciparum</i> do tipo W-2 proveniente da Indochina, artemisinina (1) e derivados (2-16).....	42
Figura 11 –	Gráfico dos primeiros dois “vetores escores” PCs (PC1 e PC2) para os derivados da dihidroartemisinina com atividade de antimalarial (Conjunto Treinando). A análise de PC conduz a uma separação em dois grupos: Atividade alta (HA) e Baixa atividade (LA).....	50
Figura 12 –	Gráfico dos primeiros dois “loadings vetores” PCs (PC1 e PC2) para os quatro os descritores responsáveis pela separação dos derivados da dihidroartemisinina com atividade de antimalarial (Conjunto Treinando).....	51
Figura 13 –	Dendograma para os 16 compostos com atividade antimalarial. Grupo com alta atividade (compostos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12) e grupo com baixa atividade (13, 14, 15, e 16).....	52
Figura 14 –	Mapas de potencial eletrostático (MEP) das estruturas do conjunto treinamento (1-16).....	55
Figura 15 –	Derivados da artemisinina com atividade antimalarial não conhecida (conjunto teste).....	57
Figura 16 –	Mapas MEP para os compostos do conjunto treinamento (17-31).....	59
Figura 17 –	Heme, “docking” da artemisinina, compostos 8 e 9 (mais ativos conjunto treinamento), compostos 17, 18, 19, (compostos classificados como mais ativos pelos métodos PCA, HCA, KNN e SIMCA).....	63
Figura 18 –	20, 24, 25 e 27 (compostos classificados como mais ativos pelos métodos PCA, HCA, KNN e SIMCA).....	64

ABREVIATURAS

AR	Atividade Relativa
ALOGPS-logs	Coeficiente de Partição (Ghose-Crippen) octanol-água ao quadrado ($\log P^2$)
D.D.T.	Dicloro-difenil-tricloroetano
\hat{H}	Hamiltoniano
HCA	Hierarchical Cluster Analysis
HE	Energia de Hidratação
HF	Hartree-Fock
HFR	Hartree-Fok-Roothaan
KNN	K Nearest-neighbor
LUMO+1	A energia do orbital molecular um nível acima do mais baixo desocupado
MAXDN	Máxima variação eletrotológica negativa
MEP	Molecular Electrostatic Potentials
PC	Principal Component
PCA	Principal Component Analysis
SAR	Structure-Activity Relationships
SIMCA	Soft independent modeling of class analogy
SVD	Decomposição singular dos valores

1 - MALÁRIA

No presente há aproximadamente 500 milhões de casos clínicos e mais de um milhão de mortes causadas por malária por ano em todo mundo (WHO - World Malaria Report 2005). A *malária* é uma doença infecciosa potencialmente grave, causada por parasitas (protozoários do gênero *Plasmodium*) que são transmitidos de uma pessoa para outra pela picada de mosquitos do gênero *Anopheles*. Os maiores índices de ocorrência desta doença estão na África, Ásia e América Latina (Figura 1), sendo a terceira doença infecciosa que mais mata em todo mundo, estando apenas atrás da tuberculose e da Síndrome imunodeficiência adquirida (AIDS) (WHO - World Malaria Report 2005).



Figura 1. Área endêmica de malária no mundo.

1.2 – HISTÓRICO DA MALÁRIA

Desde a pré-história existem registros da ação da malária no homem. Acredita-se que o Continente Africano, tenha sido o "Berço da Malária", e acompanhou a migração humana pelas regiões do Mediterrâneo, Mesopotâmia, Índia, e Sudeste Asiático. Existem registros de presença de malária entre os assírios, chineses e indianos que associavam à doença a punição de deuses e presença de maus espíritos. Hipócrates, no século V a.C., foi o primeiro médico a descartar a superstição e associar a doença aos locais freqüentados e as estações do ano. Também contribuiu com o detalhamento do quadro clínico apresentado pela malária.

Posteriormente, no século II D. C., uma variedade de médicos da Grécia e de Roma mencionaram a doença que ocorria na Grécia, Itália e em várias partes da Europa.

Padres jesuítas, no século XVII observaram a utilização, da casca de uma árvore nativa para o tratamento de alguns tipos de febre, por populações indígenas da América do Sul. Sua utilização se espalhou pela Europa e passou a ser conhecido como "pó dos jesuítas" e a árvore de onde era extraído o pó recebeu o nome de Cinchona em 1735. Seu princípio ativo, o quinino, foi isolado em 1820.

No século XVIII a doença recebeu o nome italiano de "mal aire", que significa mau ar ou ar insalubre. No final do século XIX, o conhecimento sobre a malária passou por um período de importantes descobertas. Em 1880, o médico do exército francês Charles Alphonse Laveran, trabalhando na Argélia, foi o primeiro a observar e descrever parasitas da malária no interior de glóbulos vermelhos humanos. Em 1897, o médico britânico Ronald Ross, trabalhando na Índia, elucidou o modo de transmissão, a do parasita da malária no interior de um mosquito que havia se alimentado em um portador da doença. Pesquisadores italianos Amico Bignami, Giuseppe Bastianelli e Batista Grassi, realizaram entre 1898 e 1899 estudos completos do ciclo de desenvolvimento do parasita da malária no homem e na fêmea do mosquito *Anopheles*.

Durante a primeira metade do século XX muitas pesquisas eram dedicadas ao controle da malária, especialmente no sentido de reduzir ou eliminar a presença de criadouros do inseto transmissor, o que se mostrou bastante eficiente em algumas

situações. Durante a Primeira Guerra Mundial pesquisadores alemães empenharam-se para a obtenção de antimaláricos sintéticos, culminando com o desenvolvimento da pamaquina (1924), mepacrina (1930) e cloroquina (1934), devido haver dificuldade no fornecimento do quinino.

Em 1942, com a obtenção por Paul Muller do composto dicloro-difenil-tricloroetano (D.D.T.), que apresentava grande atividade inseticida, muitos foram levados a crer na possibilidade de a doença ser erradicada, pela eliminação do vetor. Tendo sido este conceito adotado na Assembléia Mundial de Saúde em 1955, a Organização Mundial da Saúde lançou em 1957 as bases para a Campanha Mundial de Erradicação da Malária, que incluía uma fase preparatória, com a identificação de todas as áreas malarígenas e seus imóveis em condições de receber a aplicação do inseticida, outra fase seguida de um ataque, com a borrifação semestral de DDT nos imóveis, tratamento de todos os moradores com sintomas, coleta de amostras de sangue para confirmação laboratorial e aplicação de medidas de controle de criadouros dos mosquitos (drenagem, aplicação de larvicidas, etc). Nesta fase era obrigatória a observação do princípio de cobertura total.

A terceira, ou de consolidação, tinha lugar após 3 ou 4 anos, desde que não mais fossem detectados casos autóctones ou a região apresentasse incidência muito baixa (menor que 0,1 casos/1000 habitantes). Também era necessário que a possibilidade de reativação de focos de transmissão fosse mantida níveis extremamente baixos.

A última fase, ou de manutenção, tinha início quando, após a completa cessação de uso de inseticidas por no mínimo três anos consecutivos, não havia evidências de transmissão natural da malária e a vigilância, visando manter a erradicação, passaria para os serviços gerais de saúde.

Como conseqüência, por volta de 1970, os programas de erradicação haviam livrado do risco da doença cerca de 53% da população residente em áreas malarígenas, evitando milhões de mortes e contribuindo para o desenvolvimento sócio-econômico de grandes áreas, especialmente na Ásia, sul e sudeste da Europa e nas Américas.

Crises econômicas, aumento dos custos dos inseticidas, surgimento de resistência dos anofelinos aos inseticidas e dos parasitas aos antimaláricos, foram

fatores que influenciaram para que ocorresse o aumento progressivo no número de casos na maioria dos países na década de 1980, exceção à Europa e América do Norte. Isto levou à mudança de estratégia global de erradicação da malária, sendo adotadas atividades de controle integradas a programas nacionais, visando reduzir os níveis de transmissão, contando com a participação da comunidade. Desta forma, o nível de mortalidade por malária sofre uma redução significativa, devido serem adotadas medidas de controle do vetor, acesso ao diagnóstico laboratorial e tratamento eficaz e imediato. Porém, ainda hoje a doença se faz presente em cerca de 100 países, mas é praticamente confinada as áreas tropicais pobres da África, Ásia e América Latina (Dutra, 2007).

1.3 – VETOR CAUSADAOR DA MALÁRIA

Existem 4 principais espécies de agentes causadores da malária que são (Suh; Kain; Keystone, 2004):

- ✓ ***Plasmódio falciparum***: encontrado em toda a África tropical, Ásia e América Latina.
- ✓ ***Plasmódio vivax***: encontrado em zonas tropicais e em algumas zonas temperadas do mundo.
- ✓ ***Plasmódio ovale***: encontrado principalmente no oeste tropical da África.
- ✓ ***Plasmódio malariae***: encontrado distribuído de forma esparsa em todo mundo.

A maioria das mortes é atribuída ao ***P. falciparum*** (Figura 2) que é inoculado através da picada da fêmea do mosquito anófeles. Uma vez inoculado, o parasita multiplica-se exponencialmente no fígado humano e infecta as células vermelhas do sangue (Suh; Kain; Keystone, 2004 / Nicholas, 2004). Os sintomas aparecem quando o parasita arrebenta os glóbulos vermelhos e são caracterizados por febre alta (com ou sem outras indicações tais como: calafrios, vômito, diarréia, tremores e dores musculares). O ciclo de desenvolvimento da doença é mostrado na Figura 3.



Figura 2. *Plasmodium Falciparum*

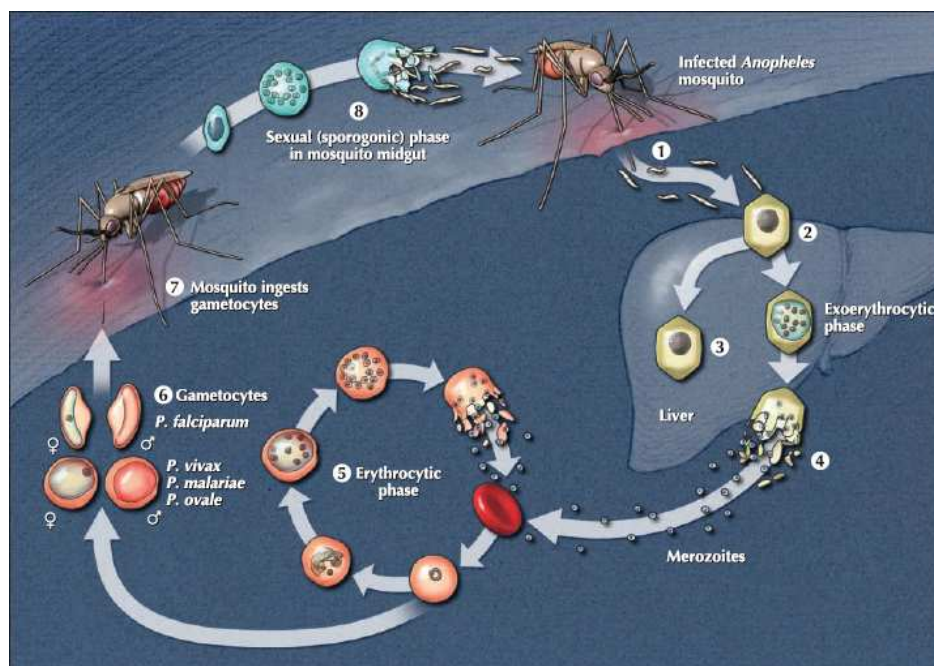


Figura 3. Ciclo de desenvolvimento da malária.

Aquecimento do clima, alta umidade e chuvas abundantes são condições favoráveis para aumentar a área de procriação do mosquito e prolongar seu tempo de vida, assim facilitando a transmissão da doença. A prevenção de picadas, o uso de medidas que possam controlar o mosquito transmissor da malária e a utilização de medicamentos profiláticos nos casos de provável exposição são estratégias que ajudam a combater a expansão da doença (Becke, 1993). Entretanto, nos casos de infecção, a escolha de antimaláricos a serem usados no tratamento é, em geral, determinada pelo padrão de resistência na área onde a infecção foi contraída (Becke, 1993).

No Brasil os casos de malária estão basicamente restritos à Amazônia Legal (Acre, Amapá, Amazonas, Maranhão, Mato Grosso, Pará, Rondônia, Roraima, Tocantins). Nas capitais dessa Região, em geral o risco é pequeno (em Belém é quase inexistente), mas pode haver transmissão significativa nos arredores das cidades. Em 2005, os municípios de Manaus e de Porto Velho foram responsáveis por 21% do total de casos de malária da Amazônia.

1.4 - ATIVIDADE ANTI-ANTIMALÁRICA

Os tratamentos usados no combate à malária, são principalmente a base de quinino, cloroquina e mefloquina (ANNE; ODILE; JÉRÔME; BERNARD, 2002), como mostradas na Figura 4.

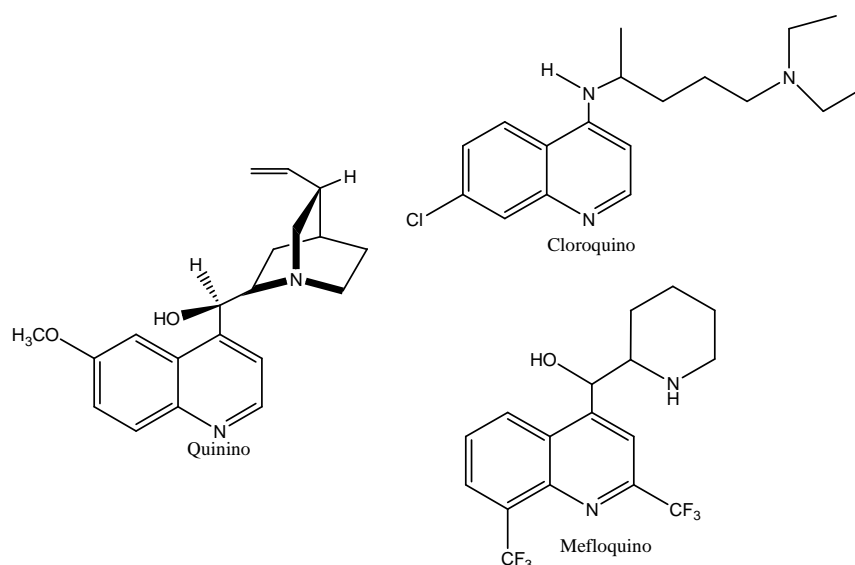


Figura 4. Drogas antimaláricas: quinino, cloroquina e mefloquina.

1.5 – ARTEMISININA

Na atualidade, dentre as drogas promissoras que vêm sendo estudadas, experimentalmente e computacionalmente, no combate da malária *falciparum*, situa-se a Artemisinina e seus derivados (KLAYMAN, 1985 / GUHA; JURIS, 2004). A Artemisinina é um produto natural isolado da planta *Artemisia Annuar L.* de origem chinesa que é usada na medicina tradicional daquele país no tratamento de aproximadamente 54 espécies de doenças (BUTLER; WU, 1992). Também vêm sendo realizados estudos do mecanismo de ação de artemisinina e seus derivados em busca de propiciar guias para síntese de novos derivados com melhor eficiência e estabilidade no combate a malária *falciparum*. Nesses estudos as atividades da Artemisinina e seus derivados parecem ser mediadas pelas suas interações com ferro da hemoglobina através da função endoperóxido (PANDEY; TEKWANI; SINGH; CHAUDAN, 1999. / PINHEIRO; KIRALJ; FERREIRA; ROMERO, 2003. / POSNER; CUMMING; PLOYPRADITH; OH, 1995. / RAFIEE; NADERI-MANESH, 2005).

Neste trabalho, realizou-se estudo a SAR de derivados da Artemisinina solúveis em água e usaram-se as informações resultantes como guia para obtenção de novos derivados.

2 - METODOLOGIA

2.1 – MÉTODOS QUÂNTICOS E PROCEDIMENTOS COMPUTACIONAIS

2.1.1 - INTRODUÇÃO

Com o surgimento da mecânica quântica, tornou-se possível obter informações da natureza microscópica da matéria. Desta forma é podemos conseguir qualquer resposta a nível atômico e molecular, bem como conhecer a natureza geral de todo um sistema de moléculas.

Simulações não relativísticas das propriedades eletrônicas e estruturais de sistemas contendo até cerca de mil átomos (BOWLER; GILLAN, 1998) tornaram-se rotina nos últimos anos em vários laboratórios. Estes sistemas requerem uma solução da equação de *Schrödinger*. A equação de *Schrödinger* em si é facilmente construída para um sistema de muitos corpos. Entretanto, é impossível resolvê-la diretamente para sistemas mais simples sem fazer algumas aproximações.

As aproximações Hartree-Fock para átomos e moléculas, servem como ponto de partida para se chegar à otimização desses tipos de sistemas atômicos e moléculares.

2.1.2 - A EQUAÇÃO DE SCHRÖDINGER

A equação de *Schrödinger* não relativística e independente do tempo para uma distribuição de elétrons e núcleos pode ser escrita de forma geral como:

$$\{H - E\}\Psi(r, R) = \{T + V_{e-e} + V_{e-n} + V_{n-n} - E\}\Psi(r, R) = 0 \quad (\text{Eq. 1})$$

com o Hamiltoniano H dado por

$$H = T + V_{e-e} + V_{e-n} + V_{n-n}$$

Na equação (Eq.1), E representa a energia total do sistema, T representa a energia cinética dos elétrons e dos núcleos, V_{e-e} é a energia potencial devido à interação elétron-elétron, V_{e-n} representa a energia potencial de interação elétron - núcleo e V_{n-n} representa a interação núcleo - núcleo. A função de onda $\Psi(\mathbf{r}, \mathbf{R})$ depende, portanto, das posições e coordenadas de *spin* de todos os N núcleos e n elétrons no sistema.

Ignorando os efeitos de acoplamento *spin-orbita* e as interações *spin-spin*, é possível escrever explicitamente todos os potenciais da equação (Eq.1). Em unidades atômicas, $\hbar = e = m_e = 4\pi\epsilon_0 = 1$, para uma distribuição normalizada dos elétrons com um conjunto de coordenadas \mathbf{r}_i e núcleos com cargas Z_a nas coordenadas \mathbf{R}_a , os termos da equação (Eq.1) podem ser escritos como:

$$\begin{aligned}
 T &= -\frac{1}{2} \left\{ \sum_{\mu} \nabla_{\mu}^2 + \sum_a \frac{1}{M_a} \nabla_a^2 \right\} \\
 V_{e-e} &= +\frac{1}{2} \sum_{\mu, \nu (\neq \mu)} \frac{1}{|r_{\mu} - r_{\nu}|} \\
 V_{e-n} &= -\sum_{\mu, a} \frac{Z_a}{|r_{\mu} - R_a|} \\
 V_{n-n} &= +\frac{1}{2} \sum_{a, b (\neq a)} \frac{Z_a Z_b}{|R_a - R_b|}
 \end{aligned}$$

A energia cinética (T) é somada sobre todos os elétrons ($\mu = 1 \rightarrow n$) e núcleos ($a = 1 \rightarrow N$). Os potenciais elétron-elétron e núcleo-núcleo são somados sobre todas as combinações de pares distintos. Para evitar que cada par seja contado duas vezes multiplica-se o potencial por 1/2.

2.1.3 - APROXIMAÇÃO DE BORN-OPPENHEIMER

Como foi escrita, a equação de *Schrödinger* é não tratável. Entretanto, a simplificação de Born e Oppenheimer (MCWEENY; SUTCLIFFE, 1969) é usada em quase todos os métodos para se resolver a equação de *Schrödinger*. A idéia central desta aproximação consiste no fato de que os núcleos são muito mais pesados do que os elétrons e, portanto, os elétrons se movem muito mais rapidamente do que os núcleos. Assim, como uma boa aproximação pode-se considerar os elétrons em uma molécula movendo em um campo fixo formado pelos núcleos. Desse modo, o termo correspondente à energia cinética dos núcleos pode ser desprezado e a energia potencial de repulsão núcleo-núcleo pode ser considerada como constante (SZABO; OSTLUND, 1996). Assim, o desacoplamento dos movimentos eletrônicos e nucleares pode ser usado para escrever a equação de *Schrödinger* de forma aproximada para o

movimento eletrônico das moléculas. Desse modo, $\Psi(\mathbf{r}, \mathbf{R})$ pode ser escrita como um produto:

$$\Psi(\mathbf{r}, \mathbf{R}) = \psi_{\mathbf{R}}(\mathbf{r})\chi(\mathbf{R}) \quad (\text{Eq.2})$$

onde $\psi_{\mathbf{R}}(\mathbf{r})$ é a função de distribuição dos elétrons para um arranjo nuclear fixo e $\chi(\mathbf{R})$ é a função de onda dependente das posições nucleares. A função $\psi_{\mathbf{R}}(\mathbf{r})$ satisfaz a equação de onda da seguinte forma:

$$\left\{ -\frac{1}{2} \sum_{\mu} \nabla_{\mu}^2 + V_{e-e} + V_{e-n} + V_{n-n} \right\} \psi_{\mathbf{R}}(\mathbf{r}) = U(\mathbf{R})\psi_{\mathbf{R}}(\mathbf{r}) \quad (\text{Eq. 3})$$

Assim, para cada distribuição nuclear, $U(\mathbf{R})$ é parte de um conjunto de autovalores para diferentes estados eletrônicos $\psi_{\mathbf{R}}$ e \mathbf{U} devem ser funções contínuas de \mathbf{R} . Substituindo a equação (Eq.2) na equação de *Schrödinger* e considerando a equação (Eq.3) tem-se:

$$\left\{ -\frac{1}{2} \sum_n \frac{1}{M_n} \nabla_n^2 + U(\mathbf{R}) \right\} \Psi(\mathbf{r}, \mathbf{R}) = E\Psi(\mathbf{r}, \mathbf{R}) \quad (\text{Eq.4})$$

a qual pode ser escrita como

$$\psi_{\mathbf{R}}(\mathbf{r}) \left\{ -\frac{1}{2} \sum_n \frac{1}{M_n} \nabla_n^2 + U(\mathbf{R}) - E \right\} \chi(\mathbf{R}) = \sum_n \frac{1}{2M_n} \left\{ \chi(\mathbf{R}) \nabla_n^2 \psi_{\mathbf{R}}(\mathbf{r}) + 2\nabla_n \chi(\mathbf{R}) \cdot \nabla_n \psi_{\mathbf{R}}(\mathbf{r}) \right\} \quad (\text{Eq.5})$$

onde o lado direito da equação (Eq.5) é zero se $\psi_{\mathbf{R}}(\mathbf{r})$ depende somente parametricamente de \mathbf{R} . Este resultado leva a aproximação da equação de *Schrödinger* para o movimento nuclear, a qual é desacoplada do movimento eletrônico, na forma

$$\left\{ -\sum_n \frac{1}{2M_n} \nabla_n^2 + U(\mathbf{R}) \right\} \chi(\mathbf{R}) = E\chi(\mathbf{R}) \quad (\text{Eq.6})$$

Considerando esta equação juntamente com a equação (Eq.3), a descrição dos movimentos nucleares e eletrônicos pode ser feita separadamente.

2.1.4 - APROXIMAÇÃO DE HARTREE

Com a simplificação do hamiltoniano (H), escrevendo-o como uma função da distribuição fixa dos núcleos, a equação de *Schrödinger* simplifica bastante. Entretanto ela ainda continua não tratável devido à presença do potencial V_{e-e} . Claramente, alguma aproximação simplificadora é necessária para este termo. Em 1927, Hartree (PARR, 1963) propôs que a função de onda eletrônica poderia ser escrita como um produto de funções de onda monoelétrica e ortogonais de tal modo que cada elétron interage com o potencial médio formado pelos núcleos fixos e por todos os outros elétrons do sistema. Esta aproximação é equivalente a escrever a função de onda de muitos corpos como

$$\psi_R(r) = \prod_i \phi_i(r_i) \quad (\text{Eq.7})$$

onde $\phi(r_i)$'s representam as funções de onda monoelétrônicas e independentes, as quais são funções das coordenadas espaciais e de *spin* dos n elétrons. Para um conjunto de funções de onda normalizadas, o Hamiltoniano relacionado com a energia de estado ϕ_i é:

$$\left\{ -\frac{1}{2} \nabla_i^2 - \sum_n \frac{Z_n}{|r_i - R_n|} + \sum_{j(\neq i)} \int |\phi_j(r_j)|^2 \frac{dr_j}{|r_i - r_j|} \right\} \phi_i(r_i) = \varepsilon_i \phi_i(r_i) \quad (\text{Eq.8})$$

Este procedimento fornece um conjunto de n equações simultâneas integrodiferenciais, as quais devem ser resolvidas autoconsistentemente para o conjunto de n ϕ_i estados. A solução deste tipo de problema é, geralmente, feita de modo iterativo, no qual um potencial inicial é usado para calcular as funções de onda monoelétrônicas. Com este conjunto de funções de onda monoelétrica calcula-se um

novo potencial a partir da nova distribuição total de cargas. Com o novo potencial obtém-se um novo conjunto de funções monoelétrônicas. Este processo é repetido até que as funções de onda monoelétrônicas permaneçam inalteradas entre dois ciclos iterativos consecutivos (ou que se atinja certo limite de tolerância).

Não é difícil mostrar que a função de onda de Hartree corresponde à solução de mais baixa energia possível do Hamiltoniano, usando o princípio variacional de Raleigh-Ritz (SLATER, 1930), o qual afirma que para uma função de onda com energia dada por

$$E = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle} \quad (\text{Eq.9})$$

a energia mais baixa possível é obtida quando

$$\frac{\delta E}{\delta \psi} = 0$$

$$\frac{\delta^2 E}{\delta^2 \psi} > 0 \quad (\text{Eq.10})$$

As soluções das equações de Hartree apresentam uma falha muito óbvia, elas não são anti-simétricas com relação à permutação dos elétrons. Assim, essas soluções não podem representar sistemas fermiônicos sem alguns ajustes.

2.1.5 – APROXIMAÇÃO HARTREE-FOCK

Para sistemas fermiônicos, a função de onda de Hartree, equação (Eq.7), carece de um importante princípio físico, *i.e.*, o princípio da exclusão de *Pauli*. Tanto Fock (FOCK, 1930) quanto Slater (SLATER, 1930) sugeriram um método para incluir este princípio escrevendo a função de onda eletrônica total como um conjunto de funções monoelétrônicas que são explicitamente anti-simétricas com respeito à permutação dos elétrons.

Para um conjunto de n elétrons, a função de onda pode ser escrita como um determinante de *Slater* normalizado (SLATER, 1930) na forma

$$\psi(r_1 \dots r_n) = \frac{1}{\sqrt{n!}} \det[\phi_a(r_1) \phi_b(r_2) \dots \phi_z(r_n)] \quad (\text{Eq.11})$$

onde $\phi_a(\mathbf{r}_i)$ é o a -ésimo spin-orbital monoelétrônico e ortonormal nas coordenadas espaciais e de spin, r_i . Este determinante garante que a função de onda total seja anti-simétrica em relação às permutações dos elétrons.

As funções de onda monoelétrônicas são construídas para serem ortogonais com relação às coordenadas espaciais (i, j) e de spin (α, β), *i.e.*,

$$\langle \phi_i^\alpha | \phi_j^\beta \rangle = \delta_{ij} \delta_{\alpha\beta} \quad (\text{Eq.12})$$

Aplicando o princípio variacional à função de onda ψ , é possível derivar um Hamiltoniano apropriado para a aproximação Hartree – Fock.

Começando com

$$E = \left\langle \psi \left| \sum_i h_i + \frac{1}{2} \sum_{i,j(\neq i)} \frac{1}{|r_i - r_j|} \right| \psi \right\rangle \quad (\text{Eq.13})$$

Onde

$$h_i = -\frac{1}{2} \nabla_i^2 - \sum_n \frac{Z_n}{|r_i - R_n|} \quad (\text{Eq.14})$$

e usando o fato de que os elétrons são indistinguíveis, esta energia pode ser escrita como

$$\left\langle \psi \left| \sum_i^n h_i \right| \psi \right\rangle + \frac{1}{2} n(n-1) \left\langle \psi \left| \frac{1}{|r_1 - r_2|} \right| \psi \right\rangle \quad (\text{Eq.15})$$

Expandindo o segundo termo na equação (Eq.15) como um conjunto de spin-orbitais a partir do determinante HF, temos:

$$\frac{1}{2} \sum_{i,j(\neq i)} \left\langle \phi(1)\phi(2) \left| \frac{1}{|r_1 - r_2|} \left\{ \phi_i(1)\phi_j(2) - \phi_j(1)\phi_i(2) \right\} \right\rangle \quad (\text{Eq.16})$$

Esta expressão pode ser escrita como

$$\sum_i \langle \phi_i(1) | h_i | \phi_i(1) \rangle + \frac{1}{2} \sum_{i,j(\neq i)} \left\{ \left\langle \phi_i(1)^* \phi_i(1) \left| \frac{1}{|r_i - r_j|} \phi_j(2)^* \phi_j(2) \right\rangle - \left\langle \phi_i(1)^* \phi_j(1) \left| \frac{1}{|r_i - r_j|} \phi_j(2)^* \phi_i(2) \right\rangle \right\} \quad (\text{Eq.17})$$

Se a energia for minimizada e sujeita as condições de ortonormalidade das funções de onda monoelétrônicas e com a introdução dos multiplicadores indeterminados de Lagrange, tem-se:

$$\delta E - \sum_{i,j} \lambda_{ij} \left\{ \langle \delta \phi_i | \phi_j \rangle + \langle \phi_i | \delta \phi_j \rangle \right\} = 0 \quad (\text{Eq.18})$$

Esta condição é satisfeita quando

$$\left[h_1 + \sum_j \{ J_j(1) - K_j(1) \} \right] \phi_i(1) = \sum_j \lambda_{ji} \phi_j(1) \quad (\text{Eq.19})$$

Onde, os operadores de Coulomb, J_u , e de troca, K_u , são definidos como

$$J_u(1)\phi_a(1) = \left\langle \phi_u(2) \left| \frac{1}{r_1 - r_2} \right| \phi_u(2) \right\rangle \phi_a(1) \quad (\text{Eq.20})$$

$$k_u(1)\phi_a(1) = \left\langle \phi_u(2) \left| \frac{1}{r_1 - r_2} \right| \phi_u(2) \right\rangle \phi_a(1) \quad (\text{Eq.21})$$

Se o operador de Fock for definido como

$$\mathfrak{S}_1 = h_1 + \sum_u \{J_u(1) - K_u(1)\} \quad (\text{Eq.22})$$

Então

$$\mathfrak{S}_1\phi_i(1) = \sum_j \lambda_{ji}\phi_j(1) \quad (\text{Eq.23})$$

Se esta equação for transformada pela introdução de um novo conjunto de orbitais canônicos $\phi_i(n)$, de tal modo que a forma do \mathfrak{S}_1 permaneça inalterada e ϕ_{ji} se transforme em uma matriz diagonal, então resulta na forma familiar da equação de Hartree – Fock:

$$\mathfrak{S}_1\phi'_i(1) = \varepsilon'_i\phi'_i(1) \quad (\text{Eq.24})$$

Retirando as “linhas” tem-se:

$$\mathfrak{S}_1\phi_i(1) = \varepsilon_i\phi_i(1) \quad (\text{Eq.25})$$

que é a sua forma convencional.

O conjunto dos n spin-orbitais é avaliado autoconsistentemente, de modo similar ao feito para a função de onda de Hartree. Da definição do operador de Fock (equação Eq.22), a solução para um spin-orbital particular é dependente dos outros $n-1$ spin-orbitais ocupados do sistema. O termo de troca resultante da anti-simetrização da função de onda leva a um “buraco de troca” de carga unitária ao redor de cada elétron

no sistema, onde todos os outros elétrons de spins paralelos são repelidos. O potencial de troca para o a -ésimo elétron no sistema é da forma

$$V_x^a(r_1, r_2) = \frac{-1}{|r_1 - r_2|} \sum_{b(\neq a)} \langle \phi_b(r_1) | \phi_b(r_2) \rangle \quad (\text{Eq.26})$$

onde a soma é sobre todos os outros elétrons de spins paralelos. A energia de troca E_x pode ser definida para um gás uniforme de elétrons com uma média de spin em função da densidade de carga como

$$E_x = \int \rho(r) \varepsilon_x[\rho(r)] dr \quad (\text{Eq.27})$$

Com

$$\varepsilon_x(\rho) = -\frac{3}{2} \left(\frac{3\rho}{8\pi} \right)^{\frac{1}{3}} \quad (\text{Eq.28})$$

2.1.6 - AS EQUAÇÕES DE Roothaan

Quando aplicadas a moléculas ou a sólidos, em contraste a átomos isolados, soluções numéricas das equações de Hartree-Fock tornam-se impraticáveis. Roothaan e Hall sugeriram uma reformulação das equações de HF usando um conjunto de funções de base para expandir a parte espacial dos spin-orbitais. Desse modo, o problema pode ser resolvido numericamente por diagonalização matricial (ROOTHAAN, 1951). A parte espacial dos orbitais é representada como uma soma de M funções de base, *i.e.*,

$$\phi_i = \sum_j^M C_{ji} \theta_j \quad (\text{Eq.30})$$

onde θ_j é uma das M funções com coeficiente c_{ji} . Este procedimento permite expressar a parte espacial da função de onda como uma combinação linearmente independente das M funções do conjunto de base. Substituindo esta definição na equação (Eq.25) tem-se:

$$\mathfrak{S}_1 \sum_j^M c_{ji} \theta_j(1) = \varepsilon_a \sum_j^M C_{ji} \theta_j(1) \quad (\text{Eq.31})$$

a qual pode ser transformada em

$$\sum_j^M C_{ji} \langle \theta_i(1) | \mathfrak{S}_1 | \theta_j(1) \rangle = \varepsilon_a \sum_j^M C_{ji} \langle \theta_i(1) | \theta_j(1) \rangle \quad (\text{Eq.32})$$

Se a matriz de sobreposição (S) e a matriz de Fock (F) da forma

$$S_{ji} = \langle \theta_i(1) | \theta_j(1) \rangle \quad (\text{Eq.33})$$

$$F_{ji} = \langle \theta_i(1) | \mathfrak{S}_1 | \theta_j(1) \rangle \quad (\text{Eq.34})$$

forem introduzidas, então a equação de Roothaan na forma matricial pode escrita como

$$Fc = \varepsilon Sc \quad (\text{Eq.35})$$

onde c é uma matriz $M \times M$, e ε é uma matriz diagonal $M \times M$ das energias dos orbitais.

Esta equação possui solução não trivial somente no caso em que

$$\det|F - \varepsilon S| = 0 \quad (\text{Eq.36})$$

Novamente, a solução desta equação é obtida autoconsistentemente, em termos dos coeficientes ajustáveis c_{ji} . O conjunto das soluções da equação (Eq.35) resulta em n orbitais ocupados e $(M - n)$ orbitais virtuais. Se os elementos da forma F_{ji} forem escritos explicitamente, então:

$$F_{ji} = h_{ji} + \sum_{i,m} b_{lm} \{2(j|lm) - (im|lj)\} \quad (\text{Eq.37})$$

onde os elementos da matriz de densidade são definidos como

$$b_{lm} = \sum_u C_{lu}^* C_{mu} \quad (\text{Eq.38})$$

e os termos da forma $(ab|cd)$ são as integrais de dois elétrons e quatro centros sobre as funções de base, *i.e.*,

$$(ab|cd) = \int \theta_a^*(1)\theta_b(1) \frac{1}{|r_1 - r_2|} \theta_c^*(2)\theta_d(2) dr_1 dr_2 \quad (\text{Eq.39})$$

Estas equações foram estendidas para o caso de sistemas de camada aberta por Pople e Nesbet (POPLE; NESBET, 1954), onde os spins dos n elétrons são considerados explicitamente. Os cálculos das integrais da forma $(ab|cd)$, combinados com os requerimentos para a reavaliação dos elementos da matriz de densidade \mathbf{b} , constituem o principal gargalho computacional para este procedimento metodológico, no qual os cálculos são da ordem $O(n^4)$

3 – MÉTODOS QUIMIOMÉTRICOS

3.1 – INTRODUÇÃO

Quimiometria provem de um conjunto de métodos com a perspectiva de obter dados, e extrair informações relevantes. A típica estratégia quimiométrica consiste nas seguintes etapas:

- ✓ Coleção dos dados conhecidos.
- ✓ Geração do modelo matemático como é base usual em multivariada estatística ou rede neural.
- ✓ Interpretação dos parâmetros padrão em termos de semelhança química.
- ✓ Aplicação do modelo em casos novos.

Usualmente somente uma pequena parte do conhecimento que é disponível sobre o sistema é usada por uma aproximação quimiométrica, mas estes dados orientam o procedimento. Consequentemente, os resultados com métodos quimiométricos não devem ser mal interpretados, e devem sempre tentar explicar e entender os parâmetros padrões resultante.

Métodos quimiométricos provêm para reduzir as enormes quantidades de dados que podem hoje em dia ser facilmente produzidos por instrumentos automáticos e direcionar os fatores para descobrir um sistema químico quando só os conceitos teóricos não são prósperos. Esse tópico apresenta uma breve discussão dos métodos quimiométricos de reconhecimento padrão PCA, HCA, KNN e SIMCA que foram usados neste trabalho.

3.2 - ANÁLISE DE COMPONENTES PRINCIPAIS - PCA

A análise de componentes principais (PCA) é amplamente usada para simplificar grande conjunto de dados (BRERETON, 2003 / FERREIRA, 2002 / PINHEIRO; FERREIRA; ROMERO, 2001) a um caminho padrão que pode ser imediatamente relacionado, reconhecido e entendido. A proposta desta técnica é a redução da dimensão da matriz de dados. O método gera um novo conjunto de dados chamados

de componentes principais, PCs, sobre qual os dados são projetados. Estes PCs são completamente não relacionados e é construído como uma combinação linear simples de variáveis originais. Estes são importantes para mostrar que os PCs contêm a maioria da variabilidade no conjunto de dados. Embora muito abaixo do espaço bidimensional. A primeira componente principal, PC1, está definida na direção do máximo de variância para o conjunto total de dados. PC2 é o máximo de variância descrito na direção ortogonal do subespaço de PC1. Os componentes subsequentes são ortogonais e são levados para que descreva o máximo da variância restante. Uma vez que a redundância é removida, somente as primeiras menores componentes principais são requeridas para descrever a maioria das informações contidas no conjunto de dados original. Os dados da matriz X (i_xj) correspondentes às moléculas i e aos descritores j , são decompostos dentro de duas matrizes, T e L , tal que:

$$X = TL^T \quad (\text{Eq.40})$$

A matriz T , conhecida como a matriz dos “score”, representa a posição dos compostos no novo sistema de coordenadas considerando que os PCS são os eixos. L é a matriz dos “loading” cuja coluna descreve agora o novo eixo, i , e os PCS são construídos dos velhos eixos. É melhor fazer isso usando a técnica de decomposição singular dos valores, SVD, como decompor X dentro de três matrizes U , S e V , onde U e V são ortogonais (autovetor da matriz quadrada) e S é a matriz diagonal contendo os valores singulares (equivalente à raiz quadrada dos autovalores). O produto $U \times S$ é a matriz score T , enquanto V corresponde a matriz loading L . O quadrado de cada elemento diagonal de S , é igual ao valor da variância nos dados originais descritos pelo correspondente componente principal (FERREIRA, 2002).

3.3 – ANÁLISE HIERARQUICA DE CLUSTER - HCA

Análise Hierarquica de Cluster (HCA) é outro importante método de análise multivariada de dados. Seu propósito primário é exibir os dados de tal modo que enfatize seus agrupamentos naturais e padrões (FERREIRA, 2002). Os resultados, como são de naturezas qualitativas, estão representados na forma de um dendograma permitem uma visualização de amostras ou variáveis no espaço bidimensional. São calculadas as distâncias entre amostras ou variáveis e transformadas em uma matriz de similaridade S (uma anotação habitual que não deve ser equivocada pela matriz de valores singulares de PCA) cujos elementos são os índices de similaridade. Para duas amostras quaisquer K e I , o índice de similaridade é definido como:

$$S_{kl} = 1.0 - \frac{d_{kl}}{d_{max}} \quad (\text{Eq.41})$$

Como S_{kl} é um elemento de S , d_{max} é a maior distancia para qualquer par de amostras num conjunto de dados. d_{kl} é a distancia Euclidiana entre as amostras k e l calculadas como:

$$d_{kl} = \sqrt{(x_{k1} - x_{l1})^2 + (x_{k2} - x_{l2})^2 + \dots + (x_{kj} - x_{lj})^2} \quad (\text{Eq.42})$$

Onde X_{ij} é um elemento da matriz de dados X .

A escala de similaridade varia de zero até um. Está claro que quanto maior o índice S_{kl} , menor é a distância entre k e l . Entretanto, S_{kl} reflete diretamente a semelhança deles (FERREIRA, 2002).

3.4 – KNN

Um objeto não conhecido pode ser classificado (BRERETON, 2003) por investigação de objetos de seus K vizinhos próximos de qual classe o grupo é conhecido. Para achar um vizinho próximo do objeto é necessário computar a distância (usualmente a distancia euclidiana) para todo objeto do conjunto de dados. Um objeto não conhecido é nomeado para a classe que tem a maioria entre os K vizinhos (Figura 6). Valores mais freqüentes usados para K estão entre um e dez. Esta aproximação é similar a uma pesquisa espectral em uma biblioteca, embora mais complicada (não linear). Medidas similares são usadas para pesquisas na biblioteca e a apontam usualmente na identificação do objeto no lugar da classificação. O método KNN pode também ser comparado como um produto para estimar a local densidade de probabilidade para a classe.

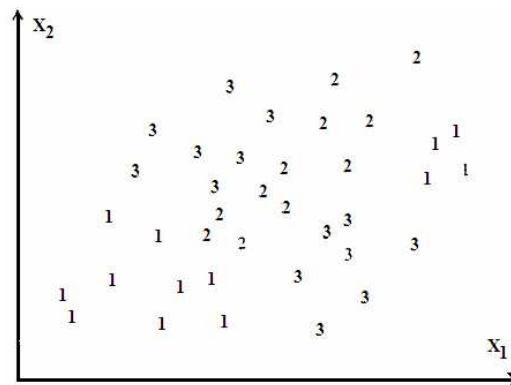


Figura 5. Classificação dos K vizinhos próximos pertencentes a três classes.

Além disso, esta simplicidade matemática da classificação KNN tem algumas vantagens: nenhuma ordenação sobre a distribuição dos dados é necessária; a classe de objetos não precisa ser linearmente separada; ele é um método de multiclasse; e não é necessário classificar o treinamento. Uma classificação KNN consiste para todo vetor objeto do conjunto treinamento; isto, porém, implica um tempo computacional bastante grande até para conjunto de dados de tamanho médio, a classificação KNN é usada frequentemente como método referencia.

3.5 – SIMCA

É um método para classificação (BRERETON, 2003) que considera informações da distribuição da população, estima um grau de confiança da classificação e pode prever novas amostras como pertencentes a uma ou mais classes ou nenhuma classe. Para fazer a classificação o SIMCA utiliza o espaço das componentes principais de cada classe. Desta forma, a classe n passa a ser representada pela equação (Eq.43).

$$X_n = \bar{T}_n \bar{P}_n^t + E_n \quad (\text{Eq.43})$$

onde X_n são os dados da classe, T_n a matriz contendo as coordenadas nas componentes principais da classe n , P_n a matriz de transformação linear e E_n a matriz de resíduos. Na construção do modelo de classificação o SIMCA calcula, para cada classe em separado, o desvio padrão dos resíduos. Para o espaço descrito pelas componentes principais, são calculadas as variâncias das amostras, em cada eixo. Estes dois parâmetros são usados na classificação de novas amostras. O objetivo do SIMCA é criar um espaço limitado para cada classe. Isto pode ser melhor compreendido para uma classe descrita por duas componentes principais. Em termos geométricos, os resíduos desta classe correspondem às distâncias das amostras ao plano dos componentes principais. Desta forma, o cálculo do desvio padrão dos resíduos dá origem a dois planos paralelos aos destes componentes, isto é, um acima e outro abaixo. Considerando a variância em cada componente principal e os planos, referentes ao desvio padrão dos resíduos, pode-se dizer que a classe está limitada por uma caixa, uma hipercaixa no caso de três ou mais componentes e um cilindro para uma componente principal.

A classificação de uma nova amostra é feita através de sua projeção nas componentes principais de cada classe, onde são calculadas as variâncias e seu resíduo. Estes dois são comparados por teste F com aqueles já determinados na construção do modelo. Assim, naquelas classes onde o resíduo é menor ou igual, o mesmo é válido para as variâncias, a amostra é classificada positivamente. Com Isto, a

amostra pode ser colocada em uma ou mais classes. Em caso contrário, desvio ou variâncias maiores, a amostra é classificada como não pertencente à classe.

A Figura 5 apresenta um exemplo de uma matriz, com três colunas, dividida em duas classes. A primeira é representada por duas componentes principais, onde v_1 é a variância na primeira componente principal PC1, v_2 a variância na segunda componente principal PC2 e o desvio padrão dos resíduos são representados por e :

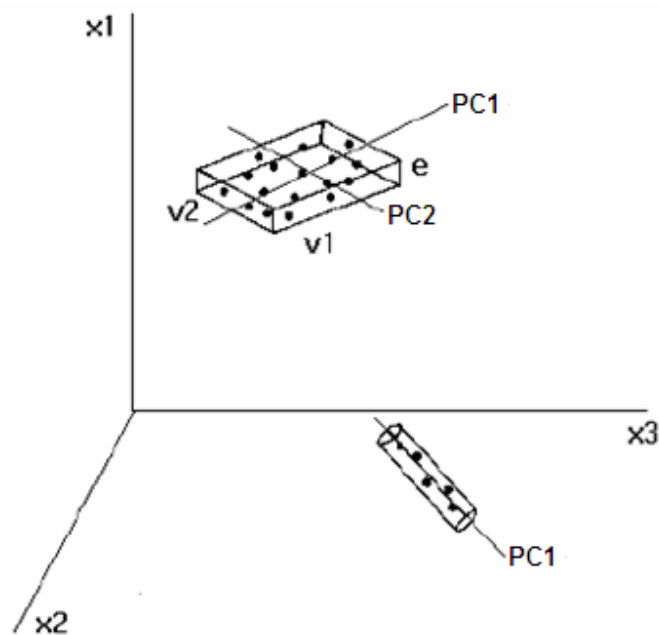


Figura 6. Representação gráfica de uma matriz com três colunas e dividida em duas classes. x_1 , x_2 e x_3 representam as colunas da matriz; PC1 o primeiro componente principal, PC2 o segundo componente principal, v_1 a variância em PC1, v_2 a variância em PC2, o desvio padrão dos resíduos é representado por e .

4 – DOCKING – O MECANISMO DE AÇÃO NA ARTEMISININA

O mecanismo para ação antimalárica da artemisinina e de seus derivados esta ainda sem conclusão, contudo investigações dão uma grande significância para o grupo endoperóxido na artemisinina para atividade biológica. Esta mesma atividade não é evidenciada nos compostos análogos da deoxiartemisinina (CHINA COOPERATIVE RESEARCH GROUP ON QINGHAOSU AND ITS DERIVATIVES AS ANTIMALARIALS, CHEMICAL STUDIES ON QINGHAOSU (AREMISININ), 1982). Experimentos *in vitro* sugerem que alguma quantia de ferro devem ser adicionadas a artemisinina para mostrar atividade biológica (MESHNICK; THOMAS; RANZ; XU; PAN, 1991 / POSNER; OH; WANG; GERENA; MILHOUS; MESHNICK; ASAWAMAHASAKDA, 1994). Nos seres humanos, a estrutura da heme (Figura 7), o qual é um produto do processo digestivo, é a fonte de ferro para artemisinina (HONG; YANG; MESHNICK, 1994).

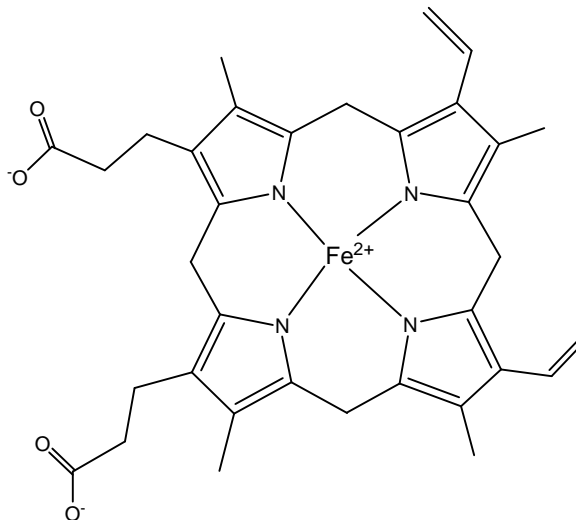


Figura 7. Estrutura da molécula heme.

Parasitas de malária em humanos degradam a hemoglobina e a célula vermelha do sangue em heme e globina. Subsequentemente, a globina é hidrolisada para produzir amino ácidos como fonte para síntese de proteínas. A porção tóxica da heme será principalmente desintoxicada pelo processo de polimerização a hemozoina.

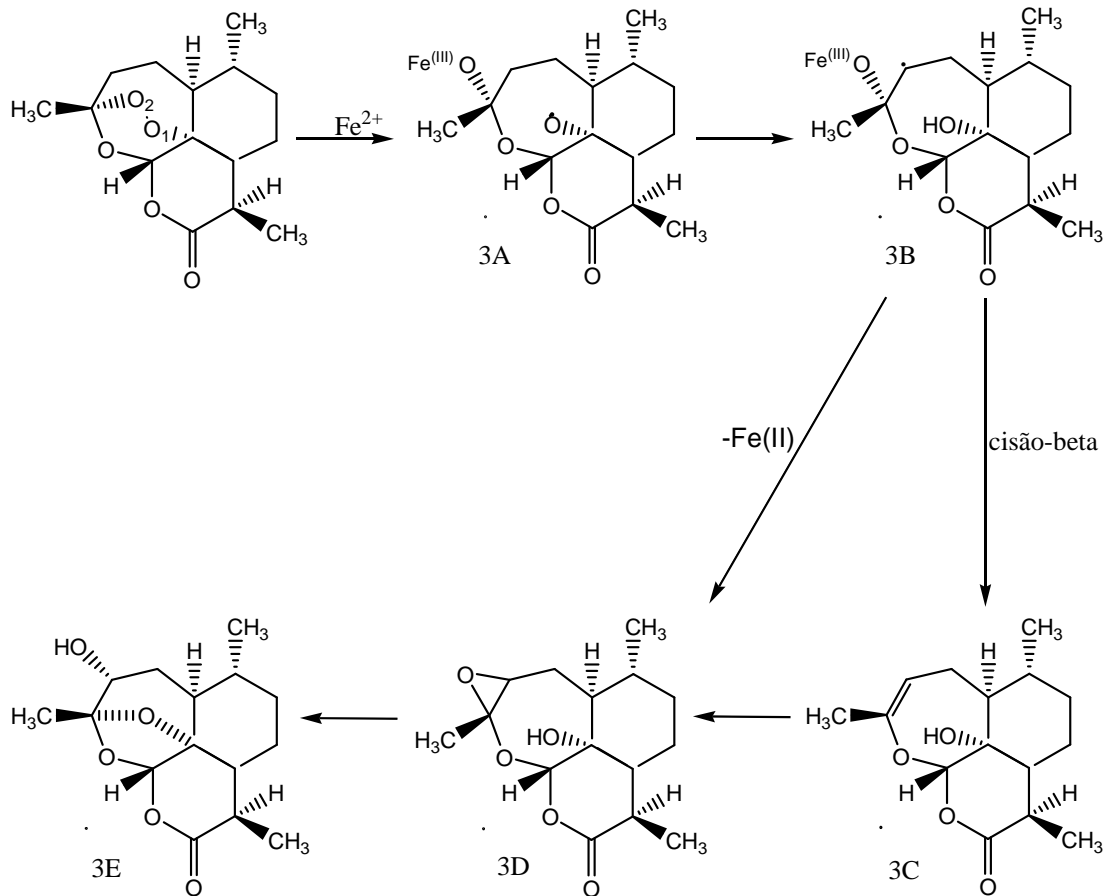


Figura 8. Mecanismo de ação artemisinina e heme proposto por Posner (POSNER; CUMMINGS; PLOYPRADITH; OH, 1995).

Existem duas possibilidades para o ferro da heme acoplar nos compostos artemisinicos. Postner e col. (POSNER; CUMMINGS; PLOYPRADITH; OH, 1995) propuseram que o ataque do ferro ao composto (artemisinicos) se dá pela posição O2 e produz o radical livre na posição O1 com posterior rearranjo para um radical livre em C4 (Figura 8). Este radical (3B) foi sugerido por ser uma importante substância para atividade biológica (POSNER; OH; WANG; GERENA; MILHOUS; MESHNICK; ASAWAMAHASAKDA, 1994). O composto 3B é mudado para o composto 3C por uma reação de cisão beta. Subsequentemente, o composto 3C é rearranjado para formar um composto epóxido (3D). Alternativamente, uma direta formação intramolecular para o radical 3B forma o composto 3D. Este composto é capaz de alquilar a proteína específica do parasita malarial e possivelmente causa danos aos parasitas (ASAWAMAHASAKDA; ITTARAT; PU; ZIFFER; MESHNICK, 1994). Por outro lado,

Jefford e col. (JEFFORD; VICENTE; JACQUIER; FAVARGER; MAREDA; MILLASSONSCHMIDT; BRUNNER; BURGER; 1996) acreditam que o ataque do ferro aos compostos se dá na posição O1 e produz um radical livre na posição O2 (Figura 9). Depois que a ligação C3-C4 é rompida para dar um radical no carbono C4 (4B). Este radical pode também ser muito prejudicial para o parasita.

Na descoberta e desenvolvimento de drogas, conhecer o mecanismo de ação pode ajudar o desenvolvimento de novas e mais efetivas drogas.

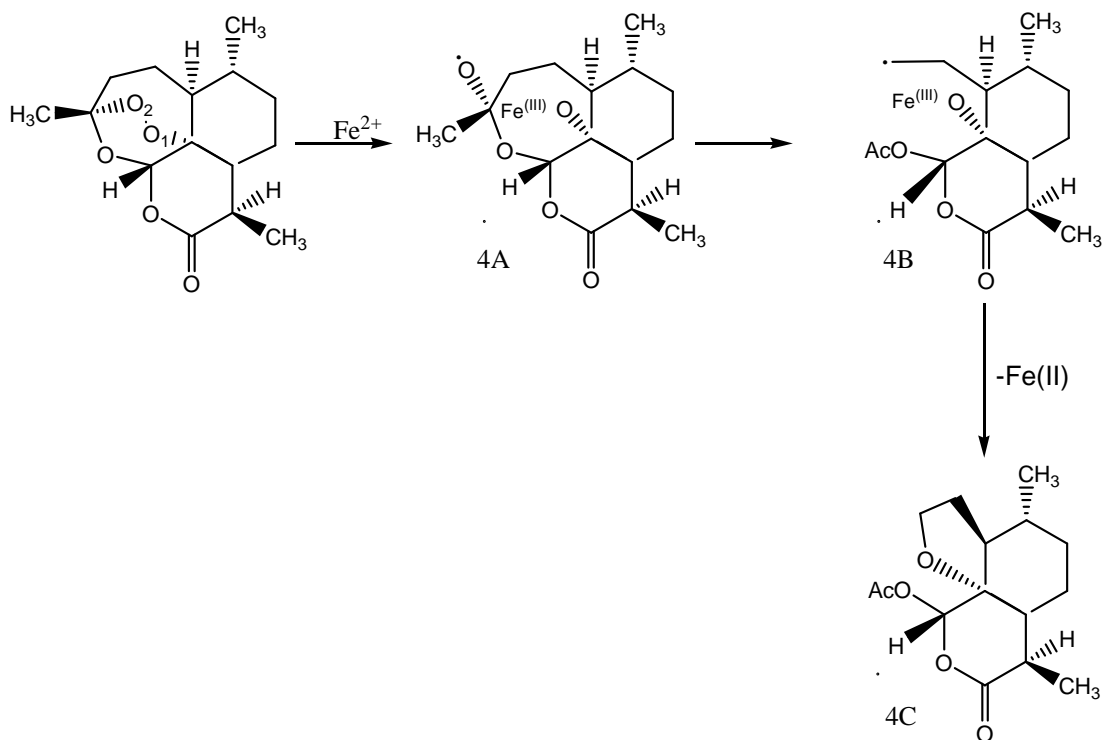


Figura 9. Mecanismo de ação artemisinina e heme proposto por Jefford (JEFFORD; VICENTE; JACQUIER; FAVARGER; MAREDA; MILLASSONSCHMIDT; BRUNNER; BURGER; 1996).

5 - RESULTADO E DISCUSSÃO

As estruturas estudadas inicialmente neste trabalho (Figura 10) foram retiradas a partir de informações disponíveis nas literaturas (LIN; LEE; KLAYMAN, 1989) consideradas conjunto treinamento, sendo construídas com a ajuda do pacote computacional Gauss View 1.0 (GAUSSVIEW, 1997) e submetidas à otimização para obtenção da geometria de conformação mais estável (menor energia) com o programa computacional Gaussian 98 (FRISCH, 2001). A otimização das geometrias realizou-se com método Hartree-Fock utilizando um conjunto de base 6-31G**. As moléculas depois de otimizadas foram submetidas também a cálculos para geração de cargas (CHELPG) e geração de Mapas MEP de potencial eletrostático (Análise Populacional de Milikan).

As estruturas utilizadas neste trabalho (conjunto treinamento) encontram-se mostradas na Figura 10, essas estruturas foram testadas *in vitro* contra malária humana *Plasmodium falciparum* do tipo W-2 proveniente de Indochina, estruturas 1-16. Essas estruturas foram classificadas como alta atividade (HA) e baixa atividade (LA). Esta definição foi dada a partir da verificação da atividade relativa, onde se definiu que os compostos com alta atividade têm o logaritmo da atividade relativa maior ou igual à zero ($\log AR \geq 0$) e os compostos com baixa atividade relativa possuem o logaritmo da atividade relativa menor que zero ($\log AR < 0$). O log da atividade relativa foi definido de acordo com a seguinte equação:

$$\log AR = \frac{IC_{50}(W2)_{artemi\ sin\ ina}}{IC_{50}(W2)_{derivado}} \times \frac{MM_{derivado}}{MM_{artemi\ sin\ ina}}$$

Desta forma os compostos considerados com alta atividade são 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12 e os compostos considerados com baixa atividade são 13, 14, 15 e 16. Os valores dos log AR para os compostos de alta atividade encontram-se no intervalo de 0 a 1,81218 e os compostos de baixa atividade encontram-se no intervalo de -0,3682 a -0,09297.

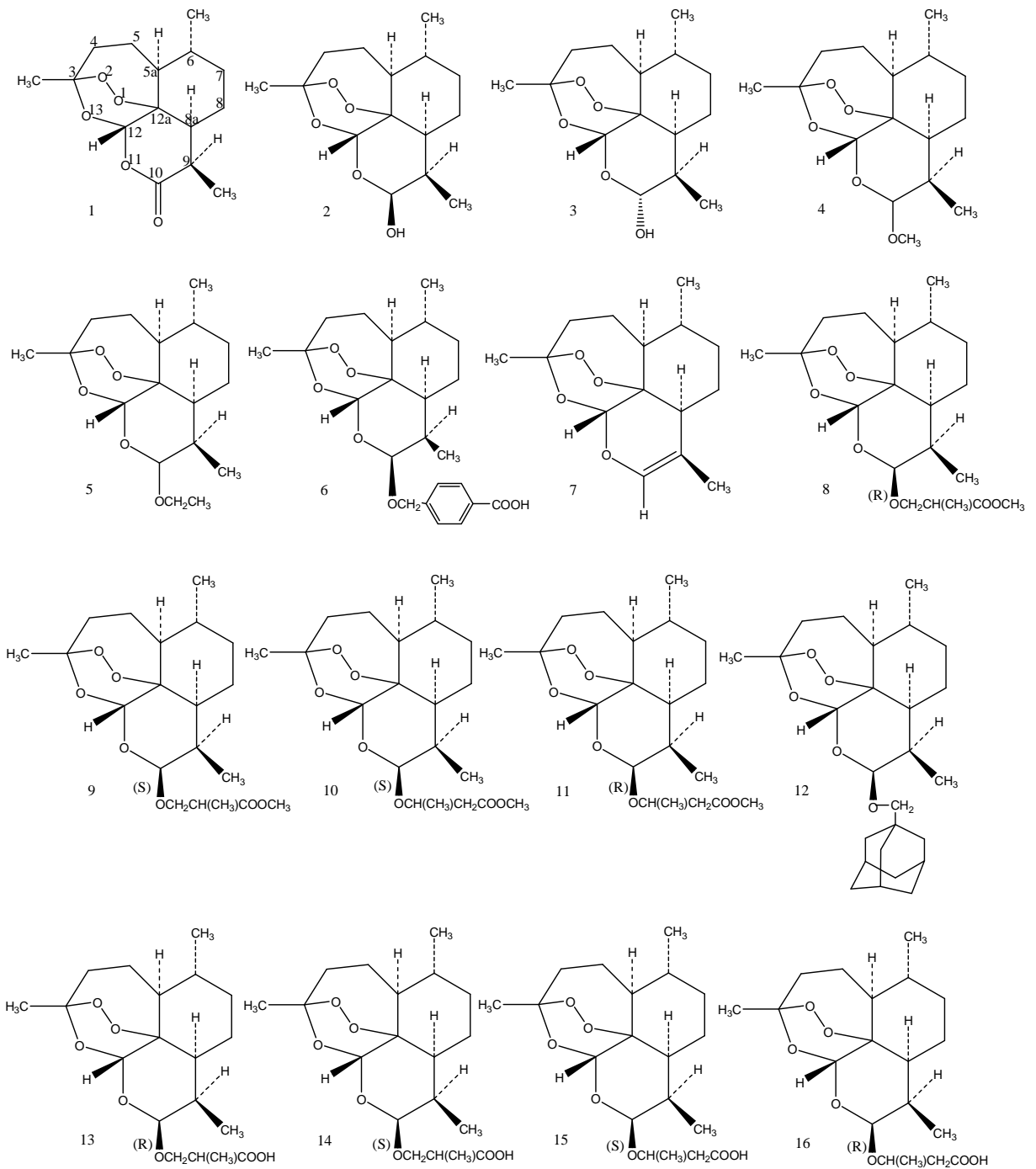


Figura 10. Estruturas estudadas no conjunto treinamento, com atividade antimalarial contra *Plasmodium falciparum* do tipo W-2 proveniente da Indochina, artemisinina (1) e derivados (2-16).

A partir da geometria da artemisinina obtida com a base 6-31G**, avaliou-se os parâmetros geométricos teóricos para o anel 1,2,4-trioxano, da artemisinina, esta avaliação foi dada com os seguintes parâmetros: comprimento de ligação, ângulo de ligação e ângulo de torção para os átomos do anel. Esta avaliação tem o intuito de observar a qualidade da base (6-31G**) comparando com os parâmetros geométricos obtidos neste trabalho com os experimentais, juntamente com os encontrados na literatura para o anel 1,2,4-trioxano da artemisinina. Tais resultados são mostrados na Tabela 1. Esta análise foi realizada a partir de cálculos a nível Hartree-Fock com diferentes conjuntos de base (3-21G, 6-31G, 6-31G* e 6-31G**) e a partir destes resultados observa-se que para o comprimento de ligação todas as bases a este nível de cálculo descrevem bem os resultados comparando-os com os experimentais, levantando destaque para a base 3-21G que mais se aproximou dos resultados experimentais bem como a base 6-31G** estudada neste trabalho que também mostrou bom desempenho na descrição dos resultados. Para os ângulos de ligação pode-se observar também uma boa descrição dos resultados, devido sua aproximação com os resultados experimentais. Destacando-se na observação deste parâmetro a base 3-21G, sendo que a base 6-31G** proporcionou um bom desempenho. Na avaliação dos ângulos diedros pode-se observar o bom desempenho dado pela base 6-31G, observa-se também que as bases 3-21G e 6-31G** mostraram bons resultados. De um modo geral todas as bases a este nível de cálculo demonstraram resultados dentro do limite aceitável para todos estes parâmetros geométricos.

Tabela 1. Parâmetros geométricos experimentais e teóricos e do anel 1, 2, 4-trioxano da artemisinina, obtidos com o método Hartree-Fock com diferentes conjuntos de base.

PARÂMETRO	3-21G ^a	6-31G ^b	6-31G ^{*c}	6-31G ^{**}	EXPER. ^d	EXPER. ^e
(este trabalho)						
Bond lengths (Å)						
O1-O2	1,4619	1,4467	1,390	1,39010	1.474	1.469
O2-C3	1,4405	1,4351	1,396	1,39630	1.418	1.416
C3-O13	1,4359	1,4347	1,408	1,40860	1.451	1.445
O13-C12	1,4074	1,4026	1,376	1,37630	1.388	1.380
C12-C12a	1,5294	1,5325	1,532	1,53180	1.528	1.523
C12a-O1	1,4772	1,4687	1,429	1,42960	1.450	1.462
Bond Angles (degree)						
O1-O2-C3	107,10	108,80	109,5	109,462	107.7	108.1
O2-C3-O13	107,28	106,76	107,8	107,818	107.1	106.6
C3-O13-C12	115,67	117,30	115,3	115,308	113.6	114.2
O13-C12-C12a	112,08	112,28	112,3	112,258	114.7	114.5
C12-C12a-O1	111,57	110,91	110,5	110,538	111.1	110.7
C12a-O1-O2	111,29	113,24	112,7	112,700	111.5	111.1
Torsion angles (degree)						
O1-O2-C3-O13	-74,67	-71,84	-73,4	-73,3700	-75.5	-75.5
O2-C3-O13-C12	32,30	33,29	31,1	31,0360	36.3	36.0
C3-O13-C12-C12a	28,29	25,32	27,4	27,4330	24.7	25.3
O13-C12-C12a-O1	-50,86	-49,41	-50,1	-50,1670	-50.8	51.3
C12-C12a-O1-O2	9,989	12,51	10,9	10,9330	12.2	12.6
C12a-O1-O2-C3	50,33	46,70	48,7	48,6740	47.7	47.8

a - Values from Ref. 1.

b - Values from Ref. 2.

c - Values from Ref. 3.

d - Values from Ref. 4.

e - Values from Ref. 5.

[1] - W.J. Hehre, R. Ditchfield, J. A., Pople, *J. Chem. Phys.* 56 (1972) 2257-2261.

[2] - J.S. Binkley, J. A. Pople, W.J. Hehre, *J. Am. Chem. Soc.* 102 (1980) 939-946.

[3] - J. C. Pinheiro, R. Kiralj, M. M. C. Ferreira, *QSAR Comb. Sci.* 22 (2003) 830-842.

[4] - I. Leban, L. Golicoe, M. Japelj, *Acta Pharm. Jugosl.* 1988, 38, 71 - 77.

[5] - J. N. Lisgarten, B. S. Potter, C. Bantuzeko, R. A. Palmer, *J. Chem. Cryst.* 1998, 28, 539-543.

Determinada a estrutura mais estável da artemisinina, realizou-se a construção de seus derivados como mostrado na figura 1, estruturas 2-16. O programa Gauss View 1.0 (GAUSSVIEW, 1997.) foi usado nesta etapa de construção das estruturas dos derivados, sendo que o programa utilizado nas otimizações das geometrias mais estáveis foi o Gaussian 98 [34]. As otimizações foram realizadas em nível de cálculo Ab Initil (Hartree Fock) com o conjunto de base 6-31G**.

Propriedades moleculares foram retiradas de todas as moléculas do conjunto treinamento, estruturas 1-16, caracterizadas como descritores químico quânticos. Com o auxílio do programa E-DRAGON SOFTWARE (TODESCHINI; CONSONNI, 2001 / TETKO; GASTEIGER; TODESCHINI; MAURI; LIVINGSTONE; ERTL; PALLYULIN; RADCHENKO; ZEFIROV; MAKARENKO; TANCHUK; PROKOPENKO, 2005 / TETKO, 2005), foram obtidos outros descritores importantes que podemos observar na Tabela 2 mostrada abaixo.

Tabela 2. Descritores disponíveis pelo programa E-DRAGON SOFTWARE.

ID do bloco	Descrição do Bloco	Número de Descritores
1	Descritores constitucionais	48
2	Descritores Topologicos	119
3	Quantidades de caminhos e trajetórias	47
4	Índices de conectividade	33
5	Índices de informações	47
6	Autocorrelações 2D	96
7	Índices de adjacência extrema	107
8	Descritor autovalor de Burden	64
9	Índices carga topologica	21
10	Índice baseado em auto valores	44
11	Perfis moleculares Randic	41
12	Descritores geometricos	74
13	Descritores RDF	150
14	Descritores MoRSE-3D	160
15	Descritores WHIM	99
16	Descritores GETAWAY	197
17	Quantidade de Grupos funcionais	154
18	Fragmentos centrado no átomo	120
19	Descritores de carga	14
20	Propriedades moleculares	29

Com esse conjunto de descritores determinados construiu-se a matriz de descritores que apresentou uma relação de 16 linhas com 811 colunas, que posteriormente foi submetida a estudos quimiométricos.

Na etapa de estudos quimiométricos, realizou-se a seleção de descritores iniciando com a análise de correlação entre os descritores e o logaritmo da atividade da atividade relativa. As variáveis que tiveram correlação menor que 0,2 foram descartadas, ressaltando que nenhum descritor químico quântico foi excluído mesmo apresentado correlação menor que o ponto de corte, desta forma a matriz de descritores passou a apresentar 210 descritores.

A este conjunto de descritores presentes na matriz aplicou-se a técnica de análise de componentes principais (PCA) que permite reduzir a matriz de descritores combinando com as estruturas e assim obter as informações levam a realização de estudos de relação entre estrutura e atividade.

Entre todos os descritores da matriz (210 descritores) os que melhor separaram as estruturas com alta atividade das de baixa atividade foram energia do orbital molecular mais baixo desocupado um nível acima (LUMO+1), energia de hidratação (HE), máxima variação eletrotológica negativa (MAXDN) e Coeficiente de Partição (Ghose-Crippen) octanol-água ao quadrado (ALOGPS-logs), a tabela 3 mostra os decritores selecionados bem com como sua classificação.

Tabela 3. Descritores responsáveis pela separação e sua classificação.

DESCRITOR	CLASSE	SIGNIFICADO
LUMO+1	Químico quântico	A energia do orbital molecular mais baixo desocupado um nível acima
HE	QSAR	Energia de Hidratação
MAXDN	Topológico	Máxima variação eletrotológica negativa
ALOGPS-logs	Propriedade molecular	Coeficiente de Partição (Ghose-Crippen) octanol-água ao quadrado ($\log P^2$)

A Tabela 4 mostra os valores dos descritores selecionados com análise de componentes principais (PCA) bem como o logaritmo da atividade relativa para todas as estruturas. Observa-se que as estruturas consideradas com alta atividade apresentam logaritmo da atividade relativa maior ou igual a zero e as estruturas consideradas de baixa atividade apresentam valores de logaritmo da atividade relativa menor que zero.

Tabela 4. Propriedades físico-químicas selecionadas por análise de componentes principais, log AR experimental e descritores responsáveis pela separação das estruturas de alta atividade e baixa atividade.

COMPOSTO	LUMO+1	HE	MAXDN	ALOGPS-logs	AR	logAR
1+	-0,0056	0,8393	0,1464	0,9007	1	0
2+	0,0053	-0,4152	-0,0285	1,3630	3.78388	0.57793
3+	0,0085	-1,9281	-0,0842	1,2995	3.78388	0.57793
4+	0,0038	0,2904	0,0545	0,7774	0.82078	0.30611
5+	0,0029	0,2998	0,0554	0,5361	0.9916	0.38824
6+	0,0060	-2,4458	0,0277	-1,1831	3.14535	0.49766
7+	-0,0017	-0,0503	-0,0912	0,4693	5.66702	0.75335
8+	-0,0009	0,3215	0,0461	-0,4368	64.8902	1.8121
9+	-0,0037	0,3409	-0,0159	-0,5074	246.58	2.391963
10+	-0,0085	0,9439	-0,0761	-0,5770	6.98047	0.84388
11+	-0,0029	-0,2115	-0,0444	-0,5410	47.1919	1.67386
12+	-0,0021	0,5826	0,1142	-1,2204	2.00043	0.30112
13-	-0,0008	0,5657	0,0496	-0,1883	0.67071	-0.1734
14-	-0,0003	0,3545	0,0450	-0,1935	0.42835	-0.3682
15-	-0,0056	2,2270	0,0875	-0,2031	0.48024	-0.3185
16-	0,0056	-1,7149	0,0066	-0,2952	0.80728	-0.0929

Os resultados obtidos na relação entre estrutura e atividade mencionados na tabela 4, descrevem o modelo construído com três componentes principais, e a descrição das informações de cada PC se dá na seguinte ordem: a primeira componente principal (PC1) descreve 55,1%, a segunda componente principal (PC2) descreve 26,7% e a terceira componente principal (PC3) descreve 14,9%. É possível observar também as informações originais das componentes, onde a primeira componente principal descreve 55,2%, avaliando as duas primeiras componentes (PC1 e PC2) observa-se a descrição de 81,89% e as três componentes (PC1, PC2 e PC3) descrevem 96,8%, estas observações encontram-se na tabela 5.

Tabela 5. Informações descritas por cada PC.

Componente Principal	Variância (%)	Variância Acumulada (%)
PC1	55,1791	55,1791
PC2	26,7129	81,8920
PC3	14,9642	96,8562

Encontram-se na Tabela 6 a contribuição de cada descritor molecular responsável pela separação. Pode-se observar que a maior contribuição para PC1 foi dada por MAXDN e HE, e as maiores contribuições em PC2 foi dada por ALOGPS-logs e LUMO+1. Na contribuição para o PC3 LUMO+1 e HE estão prevalecendo.

Tabela 6. Contribuição dos descritores para cada PC.

	PC1	PC2	PC3
LUMO+1	0.467442	0.482949	-0.649677
HE	0.572979	-0.212625	0.560022
MAXDN	-0.640790	-0.109102	-0.133076
ALOGPS-logs	-0.206352	0.842406	0.496574

As expressões matemáticas da primeira e segunda componente principal estão descritas abaixo, e são determinadas através da combinação linear dos descritores.

$$\mathbf{PC1 = 0.4674LUMO+1 + 0.5729HE - 0.6407MAXDN - 0.2063ALOGPS-logs}$$

$$\mathbf{PC2 = 0.48294LUMOI+1 - 0.2126HE - 0.1091MAXDN - 0.8424ALOGPS-logs}$$

As Figuras 11 e 12 mostram os gráficos dos “scores” e “loadings”, respectivamente, onde estão presentes as 16 estruturas do conjunto treinamento. A separação dessas estruturas de alta atividade e de baixa atividade ocorre na primeira componente principal (PC1), onde as estruturas de baixa atividade se dispõem na região esquerda e as estruturas de alta atividade se dispõem na região direita.

Os descritores moleculares, LUMO+1, HE, MAXDN e ALOGPS-logs foram os responsáveis pela separação das 16 estruturas da figura 3, onde observa-se que os descritores LUMO+1 e HE foram responsáveis pelo deslocamento das estruturas de alta atividade para a região direita, e o descritor MAXDN foi responsável pelo deslocamento das estruturas de baixa atividade para a esquerda. O descritor ALOGPS-logs foi responsável pelo deslocamento dos compostos para a região inferior do gráfico.

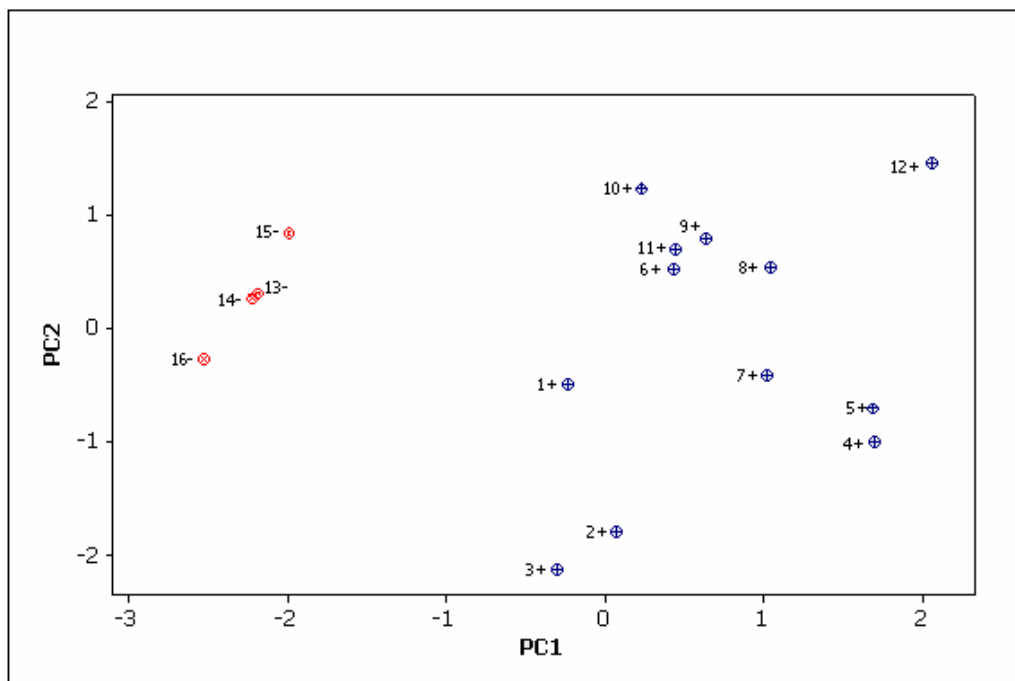


Figura 11. Gráfico dos primeiros dois “scores” vetores PCs (PC1 e PC2) para os derivados da dihidroartemisinina com atividade de antimalarial (Conjunto Treinando). A análise de PC conduz a uma separação em dois grupos: Atividade alta (HA) e Baixa atividade (LA).

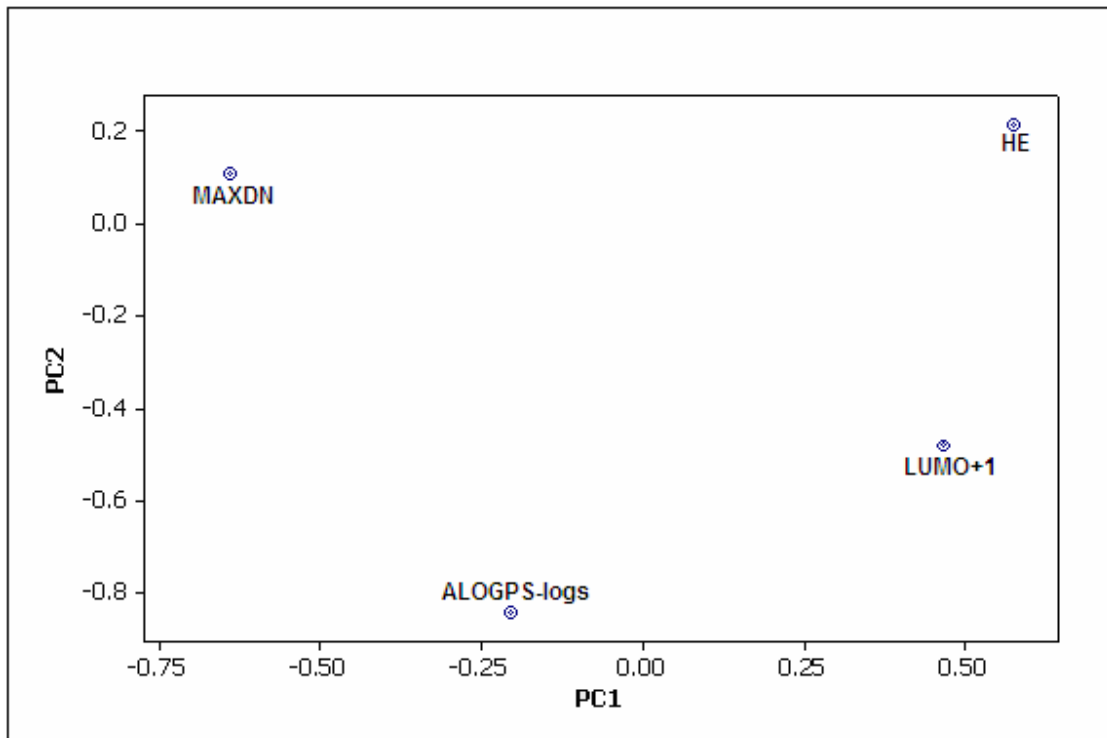


Figura12. Gráfico dos primeiros dois “loadings” PCs (PC1 e PC2) para os quatro descritores responsáveis pela separação dos derivados da dihidroartemisinina com atividade de antimalarial (Conjunto Treinamento).

O método da análise hierárquica de “cluster” usado neste estudo pode ser observado, através do dendograma mostrado na Figura 13. Observa-se que esta análise proporciona plena concordância com o método de componentes principais, onde se pode observar que houve a formação de pequenos “clusters” entre os compostos classificados com alta atividade. Destaca-se um “cluster” entre os compostos 1, 2 e 3, um outro “cluster” com os compostos 4, 5 e 7 e ainda os compostos 8, 9, 10, 11 e 12 formam um “cluster” maior entre os compostos de alta atividade e o composto 6 ficou entre os compostos de alta atividade integrando o “cluster” principal entre os mais ativos. No caso dos compostos de baixa atividade observa-se a formação de um “cluster” único entre os compostos 13, 14, 15 e 16.

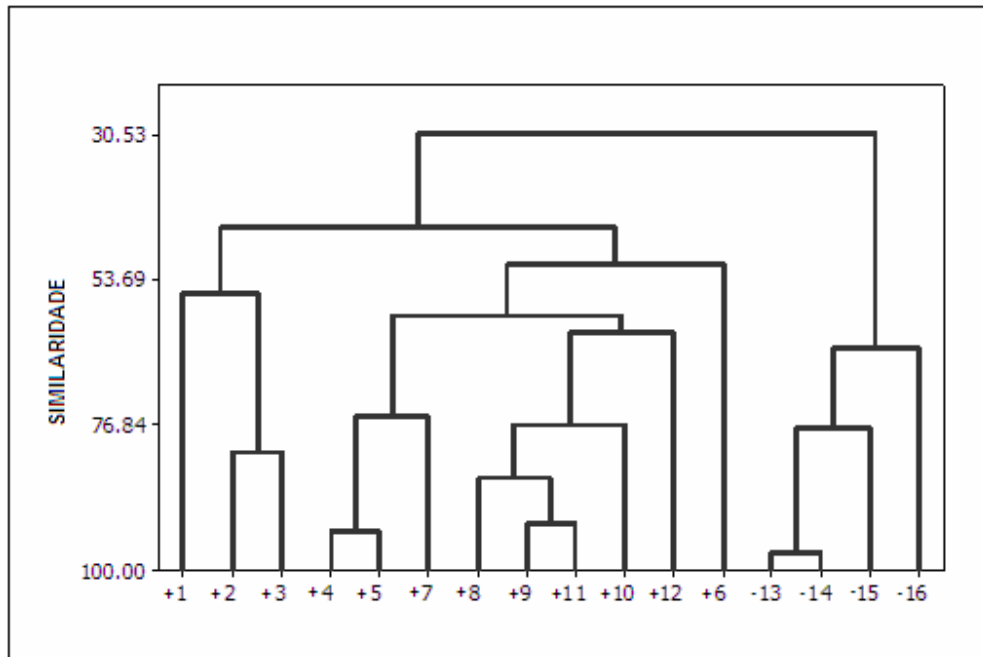


Figura 13. Dendrograma para os 16 compostos com atividade antimarial. Grupo com alta atividade (compostos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12) e grupo com baixa atividade (13, 14, 15, e 16).

O método KNN foi usado para validação inicial do conjunto de dados do conjunto treinamento (Tabela 7). Os resultados obtidos com um (1KNN), três (3KNN) e cinco (5KNN) vizinhos próximos. A porcentagem de informações corretas foi de 100% e usou-se cinco vizinhos próximos, pois quanto maior o número de vizinhos maior a confiabilidade do método (SOUZA; SANTOS; FERREIRA; MOLFETTA; CAMARGO; HONORIO; SILVA, 2003).

Tabela 7. Método KNN para o conjunto treinamento para 16 compostos.

CATEGORIA	NÚMERO DE COMPOSTOS	COMPOSTOS CLASSIFICADOS INCORRETAMENTE		
		1 NN	3 NN	5 NN
ATIVO	12	0	0	0
INATIVO	04	0	0	0
TOTAL	16	0	0	0
PORCENTAGEM DE INFORMAÇÃO CORRETA		100	100	100

A utilização do método SIMCA possibilitou a divisão dos compostos do conjunto treinamento em duas classes, sendo a classe 1 formada pelos compostos com alta atividade (apresentando os compostos 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12) e a classe 2 formada pelos compostos com baixa atividade (apresentando os compostos 13, 14, 15 e 16). As classes podem ser mostradas na Tabela 8.

Tabela 8. Classes formadas entre os compostos do conjunto treinamento através da aplicação do método SIMCA.

COMPOSTO	CLASSE
1+	1
2+	1
3+	1
4+	1
5+	1
6+	1
7+	1
8+	1
9+	1
10+	1
11+	1
12+	1
13-	2
14-	2
15-	2
16-	2

Na figura 14, pode-se observar os mapas de potencial eletrostático para as 16 estruturas estudadas, observa-se que as regiões de maior potencial negativo apresenta-se disposta sobre o anel 1,2,4-trioxano que de acordo com a literatura é a região responsável pela atividade antimalárica (TONMUNPHEAN; PARASUK; KOKPOL, 2009).

Desta forma para auxiliar e proporcionar maior consistência no estudo da relação estrutura atividade, realizou-se o estudo do mapa de potencial eletrostático das 16 estruturas do conjunto treinamento. Onde se observou a disposição de densidade de carga negativa sobre a região do anel trioxano na artemisinina. Esta mesma disposição negativa observou-se também nos derivados da artemisinina, e vem ser um suporte teórico e auxiliar no estudo da atividade das estruturas, pois se a artemisinina, é uma estrutura ativa contra malária, e apresentam região de potencial negativo sobre a região do anel, os derivados que se encontram disposto da mesma forma, isto é, com o potencial negativo também sobre a região do anel, deverão apresentar possíveis atividades antimalariais. Observa-se nesta avaliação da relação do mapa potencial eletrostático com a atividade que para as moléculas com menor log AR, isto é, de maior atividade antimalárica, a distribuição da densidade de cargas negativas no MEP é maior. Isto pode ser visualizado para as estruturas 4, 5, 6 e 12 que apresentam os seguintes log AR 0.30611, 0.38824, 0.49766 e 0.30112, respectivamente.

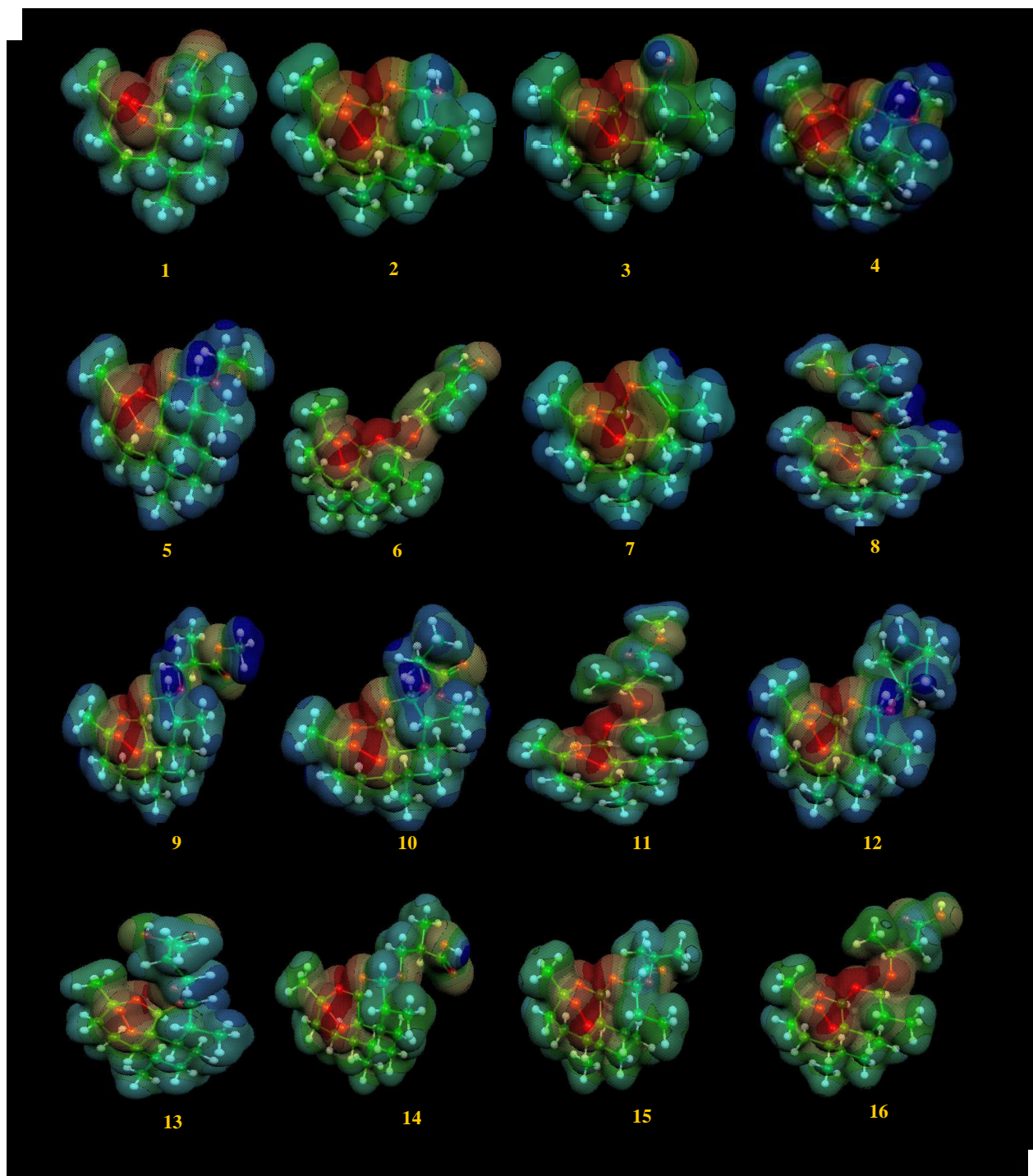


Figura 14. Mapas de potencial eletrostático (MEP) das estruturas do conjunto treinamento (1-16).

Um conjunto de 15 compostos do conjunto teste (não sintetizados) foram planejados e são mostrados na Figura 15. A esse grupo de compostos selecionou-se os mesmos descritores responsáveis pela separação do conjuntos treinamento e aplicou-se os métodos de reconhecimento padrão (PCA, HCA, KNN e SIMCA) com o intuito de observar quais entre as 15 estruturas apresentam condições de serem consideradas mais ativas ou menos ativas.

Estas estruturas do conjunto teste foram planejadas com intuito de aumentar a lipofilicidade, pois se observou que quanto maior a lipofilicidade da artemisinina maior a atividade biológica (LIN; LEE; KLAYMAN, 2009).

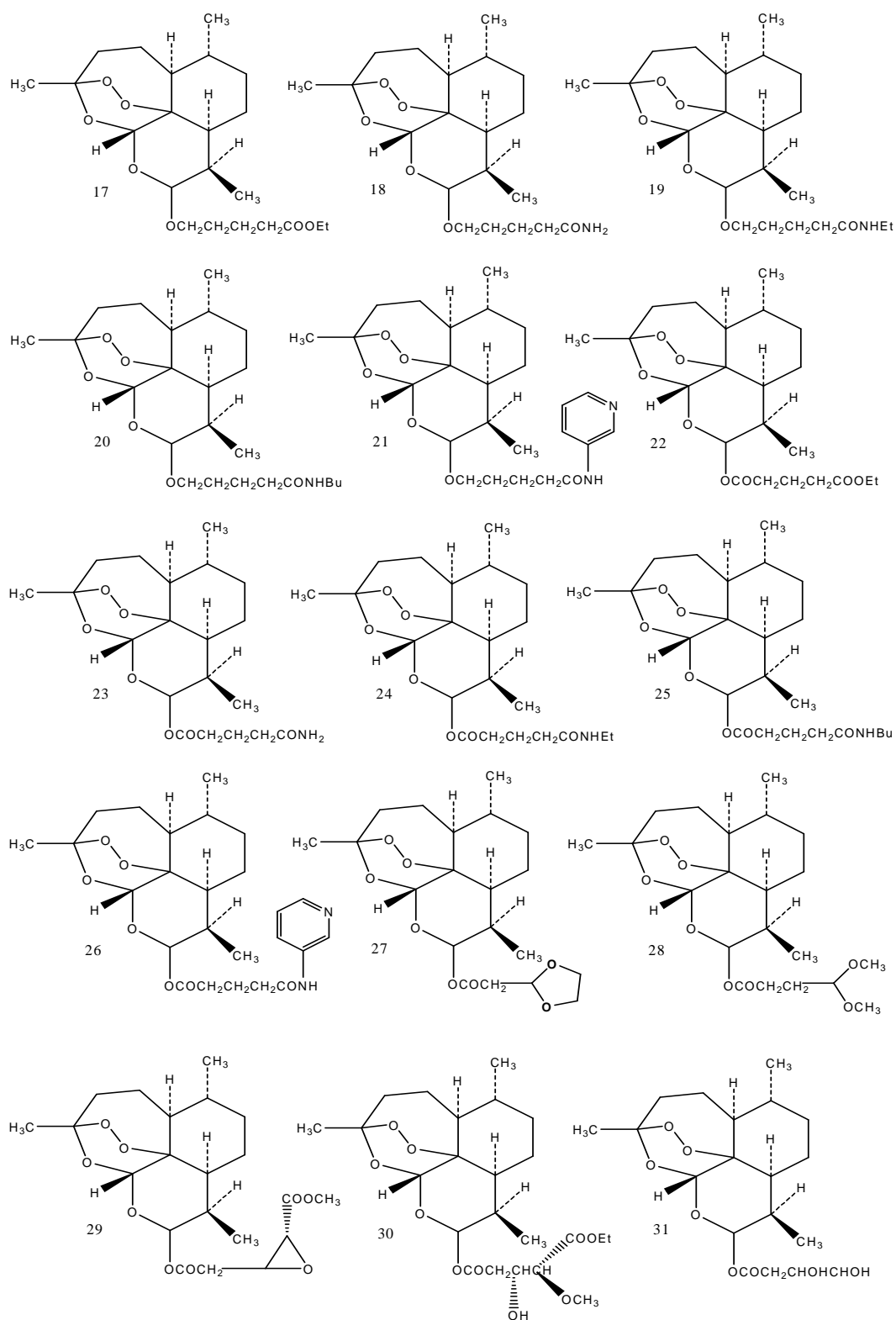


Figura 15. Derivados da artemisinina com atividade antimalarial não conhecida (conjunto teste).

A Tabela 9 apresenta os resultados obtidos com a avaliação dos métodos PCA, HCA, KNN e SIMCA para os 15 compostos do conjunto teste. Nesta avaliação observou-se que os compostos 17, 18, 19, 20, 24, 25 e 27 foram classificados por todos esses métodos com alta atividade. Somente os compostos 28 e 30 foram classificados com baixa atividade na maioria dos métodos. É interessante ressaltar que os compostos 21, 22, 26, 28, 29 e 30 não foram classificados com alta e nem com baixa atividade para o método SIMCA.

Com essa avaliação dos quatro métodos de reconhecimento padrão foi possível indicar os compostos 17, 18, 19, 20, 24, 25 e 27 para síntese, devido todos apresentarem atividade biológica de acordo com os métodos de reconhecimento padrão.

Tabela 9. Resultados dos métodos multivariados para os compostos do conjunto teste. (HÁ – alta atividade e LA – baixa atividade).

COMPOSTOS	PCA	HCA	KNN	SIMCA
17	HA	HA	HA	HA
18	HA	HA	HA	HA
19	HA	HA	HA	HA
20	HA	HA	HA	HA
21	HA	LA	LA	0
22	HA	HA	HA	0
23	HA	HA	LA	HA
24	HA	HA	HA	HA
25	HA	HA	HA	HA
26	HA	LA	LA	0
27	HA	HA	HA	HA
28	LA	LA	LA	0
29	HA	HA	LA	0
30	LA	LA	LA	0
31	LA	HA	LA	HA

Os mapas de potencial eletrostático MEP também foram construídos para os compostos do conjunto teste, de modo a observar a semelhança com a artemisinina e assim reforçar os resultados indicados pelos métodos multivariados. Esses mapas estão apresentados na Figura 16.

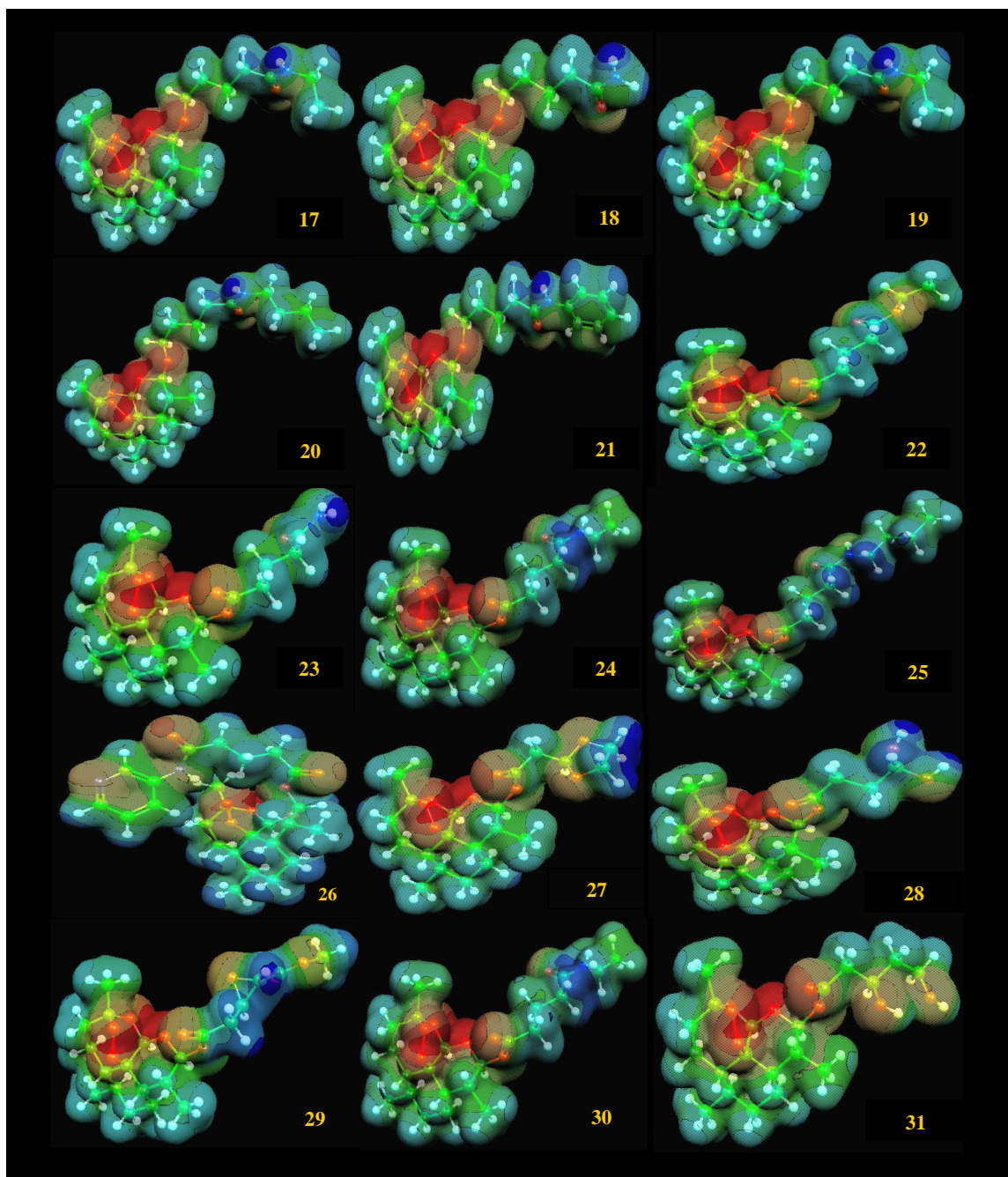


Figura 16. Mapas MEP para os compostos do conjunto treinamento (17-31).

Cálculos de “docking” foram realizados para artemisinina (composto 1) e para os compostos de maior atividade biológica do conjunto treinamento (compostos 8 e 9) bem como para os compostos que classificados com alta atividade pelos métodos de reconhecimento padrão (PCA, HCA, KNN e SIMCA) do conjunto teste (compostos 17, 18, 19, 20, 24, 25 e 27).

Os cálculos de docking foram realizados usando o programa AutoDock 2.4 (MORRIS; GOODSELL; HUEY; OLSON), em uma CPU Amd 4600+ 4gb ram rodando na plataforma Linux. Em adição a estrutura cristalográfica de raio-X, o docking da heme (obtida do Protein Data Bank-PDB (BERMAN; WESTBROOK; FENG; GILLILAND; BHAT; WEISSIG; SHINDYALOV; BOURNE, 2000) e a estrutura otimizada da artemisinina obtida com o nível de teoria HF/6-31G** foi investigada baseadas nos resultados mostrados nos trabalhos de Tonmunphean e Colaboradores (TONMUNPHEAN; PARASUK; KOKPOL, 2000).

Com os cálculos de docking obteve-se 20 possibilidades de interação entre artemisinina-heme, onde se retirou a interação de menor energia e a partir desta calculou-se o comprimento de ligação no complexo entre o ferro da heme e os seguintes oxigênios da artemisinina O1, O2, O11 e O13 estes resultados estão mostrados na Tabela 10.

Tabela 10. Resultados do docking entre heme-PDB / artemisinina e derivados com HF/6-31G**.

Moléculas	Energia (Kcal mol ⁻¹)	O1-Fe distance (Å)	O2-Fe distance (Å)	O13-Fe distance (Å)	O11-Fe distance (Å)
1	-5.84	2.759	3.866	5.332	5.750
8	-5.95	2.469	3.346	5.180	5.609
9	-5.48	2.634	3.607	5.234	5.475
17	-5.34	2.787	3.700	5.566	5.151
18	-5.87	2.669	3.249	5.035	5.558
19	-5.62	2.542	2.948	5.022	5.247
20	-5.49	2.482	3.189	5.202	4.897
24	-5.85	2.543	3.266	5.066	5.444
25	-5.18	3.072	3.644	5.244	5.988
27	-6.19	2.482	3.432	5.186	5.281

Trabalhos realizados por Tonmunphean e colaboradores (TONMUNPHEAN; PARASUK; KOKPOL, 2000) mostraram que o átomo de ferro da heme aproxima-se da ligação endoperóxido da droga, com distâncias de 1,93 a 2,73 Å para o Fe-O1 e no caso do Fe-O2 a distância fica entre 2,41 a 3,76 Å. Isto mostra a importância da interação entre o Fe do heme e o grupo endoperóxido dos derivados da artemisinina. Resultados obtidos experimentalmente para o comprimento da ligação do Fe-O1 (ferro da heme e oxigênio da oxihemoglobina A retirada do banco de dados cristalográficos) indicam a distância de 1,86 Å.

Os resultados de comprimento da ligação entre o Fe da heme e os oxigênios da artemisinina obtidos para estas interações estão mostrados na tabela 10. Com esses resultados é possível verificar certo grau de aproximação em comparação aos resultados obtidos por Tonmunphean (MORRIS; GOODSELL; HUEY; OLSON).

Os resultados obtidos neste trabalho mostram uma distância mínima de 2.469 Å para o comprimento entre Fe-O1 na estrutura 8 do conjunto teste e a distância máxima obtida foi pra a estrutura 25 com o comprimento de 3.072 Å para o Fe-O1. Os outros

comprimentos obtidos para as outras estruturas estão na faixa intermediária das estruturas 8 e 25. Avaliando o comprimento de ligação do Fe-O2 observa-se a distância mínima de 2.948 Å para a estrutura 19 enquanto que a distância máxima fica em 3.886 Å para a estrutura 1, sendo que os outros comprimentos estão na faixa intermediária das estruturas 19 e 1.

Para o comprimento de ligação obtido para o Fe-O13 observa-se a distância mínima de 5.022 Å para a estrutura 19 enquanto que a distância máxima encontra-se em 5.566 para a estrutura 17, sendo as outras estruturas apresentando distâncias na faixa intermediária das estruturas 19 e 17. Para o comprimento de ligação obtido para o Fe-O11 observa-se a distância mínima de 4.897 Å para a estrutura 20 enquanto que a distância máxima encontra-se em 5.988 Å para a estrutura 25, sendo as outras estruturas apresentando distâncias na faixa intermediária das estruturas 20 e 25.

As Figuras 17 e 18 mostram a heme e as interações entre a heme e a artemisinina (estrutura 1), estruturas 8 e 9 (mais ativos conjunto treinamento), estruturas 17, 18, 19, 20, 24, 25 e 27 (compostos classificados como mais ativos pelos métodos PCA, HCA, KNN e SIMCA).

Estes resultados de docking obtidos para estas estruturas do conjunto teste podem auxiliar no estudo da atividade biológica destas estruturas mencionadas neste trabalho. Tendo em vista a boa aproximação com resultados encontrados na literatura (TONMUNPHEAN; PARASUK; KOKPOL, 2000).

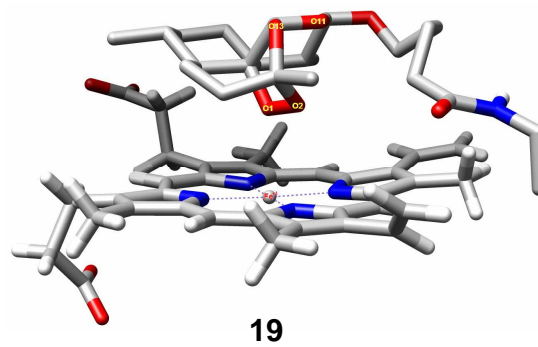
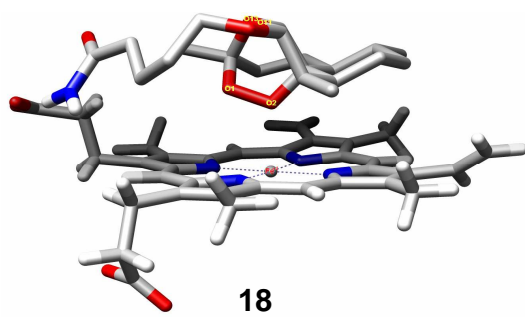
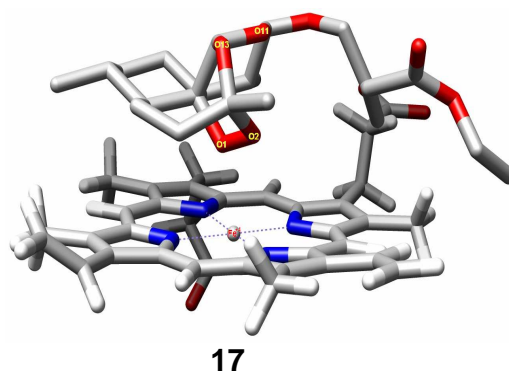
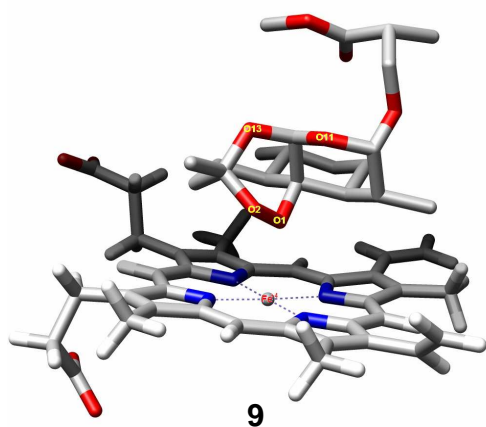
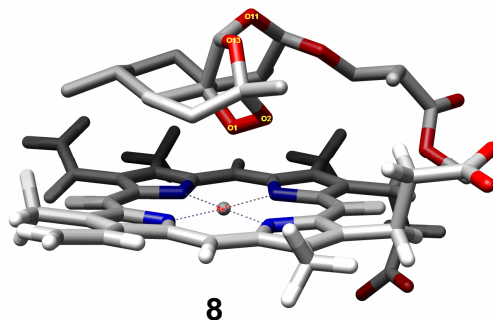
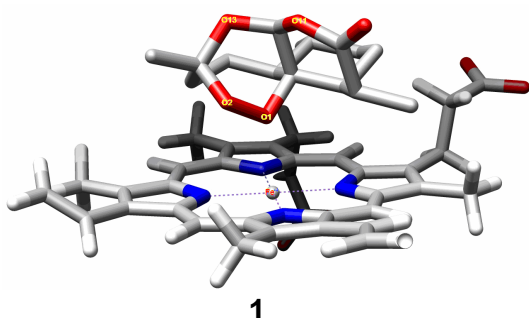
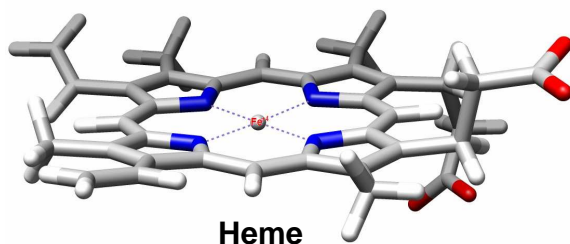


Figura 17. Heme, docking artemisinina, 8 e 9 (mais ativos conjunto treinamento), 17, 18, 19 (compostos classificados como mais ativos pelos métodos PCA, HCA, KNN e SIMCA).

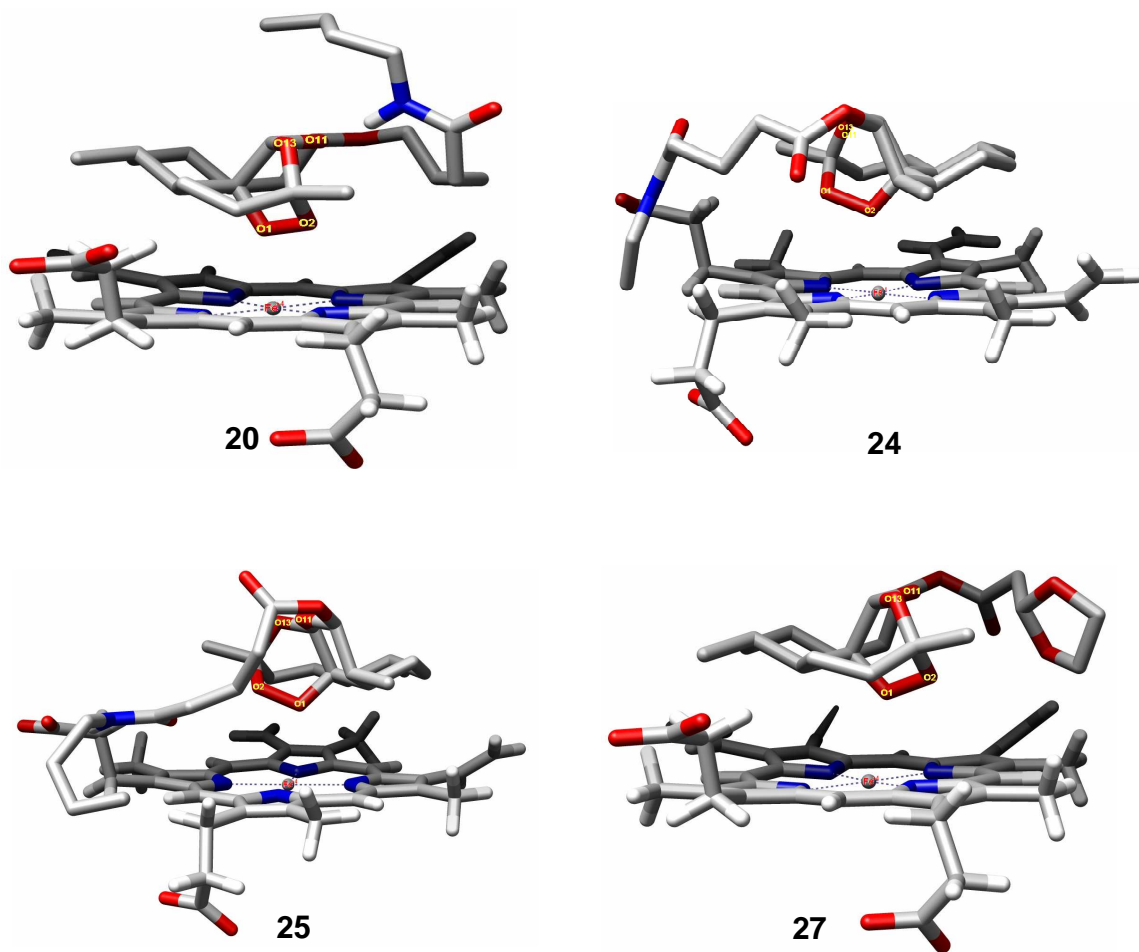


Figura 18. Docking dos compostos 20, 24, 25 e 27 (compostos classificados como mais ativos pelos métodos PCA, HCA, KNN e SIMCA).

6 - CONCLUSÃO

Neste trabalho realizaram-se o estudo SAR de alguns compostos derivados da artemisinina atividades. Com apoio de técnicas quimiométricas e da mecânica quântica, podem-se obter informações a respeito do comportamento molecular das estruturas usadas neste trabalho.

Com o estudo mecânico quântico obtiveram-se informações de grande importância como: propriedades moléculas, cargas e mapa de potencial eletrostático, entre outras. No caso das técnicas quimiométricas, este estudo possibilitou obter informações que possibilitaram agrupar as propriedades de maior importância e relaciona-las com as estruturas estudadas e assim propor um conjunto de moléculas com maiores chances de apresentarem atividades antimaláricas.

Entre as 16 estruturas do conjunto treinamento os compostos considerados com alta atividade são 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 e 12 e os compostos considerados com baixa atividade são 13, 14, 15 e 16. Os valores dos log AR para os compostos de alta atividade encontram-se no intervalo de 0 a 1,81218 e os compostos de baixa atividade encontram-se no intervalo de -0,3682 a -0,09297.

As 16 estruturas foram submetidas a otimização com auxílio do programa Gaussian 98 e em seguida retirou-se propriedades moleculares com a ajuda do programa E-dragon, que posteriormente foram usadas para realizar-se a seleção de variáveis com PCA e assim obter as propriedades de maior relevância para o estudo SAR.

Após a seleção de variáveis utilizaram-se também as técnicas de HCA, KNN e SIMCA que serviram como suporte para encaminhar as estruturas do conjunto teste (15 compostos) para possíveis presenças de atividade biológica. Nesta avaliação observou-se que os compostos 17, 18, 19, 20, 24, 25 e 27 foram classificados por todos esses métodos com alta atividade. Somente os compostos 28 e 30 foram classificados com baixa atividade na maioria dos métodos. É interessante ressaltar que os compostos 21, 22, 26, 28, 29 e 30 não foram classificados com alta e nem com baixa atividade para o método SIMCA.

Com essa avaliação dos quatro métodos de reconhecimento padrão foi possível indicar os compostos 17, 18, 19, 20, 24, 25 e 27 para síntese, devido todos apresentarem atividade biológica de acordo com os métodos de reconhecimento padrão.

Mapas de potencial eletrostático foram construídos para as estruturas treinamento e teste para se verificar a semelhança com o mapa da artemisinina de modo a aumentar confiabilidade das técnicas quimiométricas. Com estes mapas foi possível se verificar uma grande semelhança com o mapa da artemisinina, onde se encontra a região de maior densidade negativa sobre o anel trioxano.

Por fim realizou-se o “docking” dos compostos de maior atividade para se verificar como poderá ocorrer a interação do fármaco (artemisinina) com o receptor (heme). Neste trabalho foram usadas técnicas diferenciadas para se obter informações que dessem consistência a um posterior trabalho de síntese das estruturas estudadas e propostas com atividade antimalárica.

7 - REFERENCIAS

ANNE ROBERT, ODILE DECHY-CABARET, JÉRÔME CAZELLES; BERNARD MEUNIER, From Mechanistic Studies on Artemisinin Derivatives to New Modular Antimalarial Drugs, **Acc. Chem. Res.** 2002, 35, 167-174.

ASAWAMAHASAKDA, W.; ITTARAT, I.; PU, Y.-M.; ZIFFER, H.; MESHNICK, S.R., Reaction of Antimalarial Endoperoxides with Specific Parasite Proteins, **Antimicrob. Agents Chemother.** 1994, 38, 1854±1858.

BECKE, A. D., **J Chem Phys**, 1993, 98, 5648.

BERMAN, H.M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T.N.; WEISSIG, H.; SHINDYALOV, I.N.; BOURNE, P.E. **The Protein Data Bank. Nucleic Acids Research**, 2000, 28 pp. 235-242.

BOWLER, D. R.; GILLAN, M. J. Length-scale ill conditioning in linear-scaling DFT. **Computer Phys. Comm.**, 1998, v.112, p.103-111,.

BRERETON, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant.* **John Wiley & Sons, Ltd**, 2003.

BUTLER, A. R. ; WU , Y. L., **Chem. Soc. Rev.**, 1992, 21, 85.

China Cooperative Research Group on Qinghaosu and Its Derivatives as Antimalarials, Chemical Studies on Qinghaosu (Artemisinin), **J. Trad. Chin. Med.** 1982, 2, 3-8.

DUTRA, A. P., Malária. Disponível em: <http://www.opas.org.br /sistema/ arquivos/ livro.PDF>. Acesso em: 03/01/07.

FERREIRA, M. M. C. **Journal of the Brazilian Chemical Society**, 2002, 13, 742-753.

FOCK, V., Näherungsmethode Zur Lösung Des Quantenmechanischen Mehrkörperproblems. **Z. Phys.**, 1930, v.61, p.126-148.

FRISCH, M. J. et al. Gaussian 98-Revision A.11, Gaussian, Inc., **Pittsburgh, PA**, 2001.

GAUSSVIEW, 1.0, Gaussian, Inc., **Pittsburgh, PA**, 1997.

GUHA, R; JURIS, P. C., **J. Chem. Inf. Comput. Sci.** 2004, 44, 1400.

HONG, Y.-L.; YANG, Y.Z.; MESHNICK, S.R., The interaction of artemisinin with malarial hemozoin, **Mol. Biochem. Parasitol.** 1994, 63, 121±128.

JEFFORD, C.W., VICENTE, M.G.H., JACQUIER, Y., FAVARGER, F., MAREDA, J., MILLASSON-SCHMIDT, P., BRUNNER, G., BURGER, U., The Deoxygenation and Isomerization of Artemisinin and Artemether and Their Relevance to Antimalarial Action, **Helv. Chim. Acta.** 1996, 79, 1475±1487.

KLAYMAN, D. L. **Science**, 1985, 228, 1049.

LIN, A. J.; LEE, M.; KLAYMAN, D. L.. Antimalarial Activity of New Water-Soluble Dihydroartemisinin Derivatives. 2 Stereospecificity of the Ether Side Chain. **Journal of Medicinal Chemistry**, 1989, Vol. 32, No, 6.

MESHNICK, S.R.; THOMAS, A.; RANZ, A.; XU, C.-M.; PAN, H.-Z., Artemisinin (Qinghaosu): The role of Intracellular Hemin in its Mechanism of Antimalarial Action, **Mo. Biochem, Parasitol**, 1991, 49, 181-190.

MCWEENY, R.; SUTCLIFFE, B.T. Methods of Molecular Quantum Mechanics. **London, Academic Press**, 1969. 690p.

MORRIS, G.M., GOODSELL, D.S., HUEY, R., OLSON, A.J., AutoDock Version 2.4, The Scripps Research Institute, Department of Molecular Biology, MB-5, **La Jolla, California, USA**.

NICHOLAS J. White, Antimalarial drug resistance, **The Journal of Clinical Investigation**, 2004, 8, 113.

PANDEY, A . V.; TEKWANI, B. L.; SINGH, R. L.; CHAUDAN, V. S., **J. Biol. Chem.**, 1999, 274, 19383.

PARR, R.G. The Quantum Theory of Molecular Electronic Structure. **New York, Benjamin**, 1963. 530p.

PINHEIRO, J. C.; FERREIRA, M. M. C.; ROMERO, O. A. S. **Journal Molecular Structure (THEOCHEM)**, 2001, 572, 35-44.

PINHEIRO, J. C.; KIRALJ, R.; FERREIRA, M. M. C.; ROMERO, O. A. S., **QSAR & Comb. Sci.**, 2003, 22, 830.

POPLE, J.A; NESBET R.K. Self-consistent orbitals for radicals. **J. Chem. Phys.**, 1954, v.22, p. 571-572.

POSNER, G.H., CUMMINGS, J.N., PLOYPRADITH, P., OH, C.O., Evidence for Fe(IV).O in the Molecular Mechanism of Action of the Trioxane Antimalarial Artemisinin, **J. Am. Chem. Soc.** 1995, 117, 5885±5886.

POSNER, G.H.; OH, C.H.; WANG, D.; GERENA, L.; MILHOUS, W.K.; MESHNICK, S.R.; ASAWAMAHASAKDA, W., Mechanism-Based Design, Synthesis, and in vitro Antimalarial Testing of New 4-Methylated Trioxanes Structurally Related to Artemisinin: The Importance of a Carbon-Centered Radical for Antimalarial Activity, **J. Med. Chem.** 1994, 37, 1256±1258.

RAFIEE, M. A; NADERI-MANESH, N. L., H., **J. Chem. Inf. Model**, 2005, 45, 366.

ROOTHAAN, C.C.J. New developments in molecular orbital theory. **Rev. Mod. Phys.**, 1951, v.23, p.69-89.

SOUZA, J.; SANTOS, R. H. A.; FERREIRA, M. M. C.; MOLFETTA, F. A.; CAMARGO, A. J.; HONORIO, K. M.; SILVA, A. B. F. A quantum chemical and statistical study of flavonoid compounds (flavones) with anti-HIV activity. **European Journal of Medicinal Chemistry**, 2003, 38 929- 938.

SLATER, J.C. Note on Hartree's method. **Phys. Rev.**, 1930, v.35, p.210-211.

SUH, K. N.; KAIN, K. C.; KEYSTONE, J. S. Malaria *Synthèse Review*, **Canadian Medical Association or its licensors**, MAY 25, 2004; 170 (11)

SZABO, A.; OSTLUND, N.S. Modern Quantum Chemistry: introduction to advanced electronic structure theory. **New York, Dover**, 1996. 446p.

TETKO, I. V.; GASTEIGER, J.; TODESCHINI, R.; MAURI, A.; LIVINGSTONE, D.; ERTL, P.; PALYULIN, V. A.; RADCHENKO, E. V.; ZEFIROV, N. S.; MAKARENKO, A. S.; TANCHUK, V. Y.; PROKOPENKO, V. V., **J. Comput. Aid. Mol. Des.**, 2005, 19, 453-63. Virtual computational chemistry laboratory - design and description

TETKO, I. V. Computing chemistry on the web, **Drug Discov. Today**, 2005, 10, 1497-500.

TODESCHINI, R.; CONSONNI, D V.: "Handbook of Molecular Descriptors", **WILEY-VCH**, 2000.

TONMUNPHEAN, S.; PARASUK, V.; KOKPOL, S. QSAR Study of Antimalarial Activities and Artemisinin-Heme Binding Properties Obtained from Docking Calculations, **Quant. Struct.-Act. Relat.**, 2000 19.

WHO - **World Malaria Report**, 2005.

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)